Vol. 4, No.1-2, 2005

Journal of Global Positioning Systems

ISSN 1446-3156 (Print Version) ISSN 1446-3164 (CD-ROM Version)

International Association of Chinese Professionals in Global Positioning Systems (CPGPS)

Journal of Global Positioning Systems

Aims and Scope

The Journal of Global Positioning Systems is a peer-reviewed international journal for the publication of new information, knowledge, scientific developments and applications of the global navigation satellite systems as well as other positioning, location and navigation technologies. The Journal publishes original research papers, review articles and invited contributions. Short research and technical notes, book reviews, and commercial advertisements are also welcome. Specific questions about the suitability of prospective manuscripts may be directed to the Editor-in-Chief.

Publication and Copyright

The Journal of Global Positioning Systems is an official publication of the International Association of Chinese Professionals in Global Positioning Systems (CPGPS). It is published twice a year, in June and December. The Journal is available in both print version (ISSN 1446-3156) and CD-ROM version (ISSN 1446-3164), which can be accessed through the CPGPS website at http://www.cpgps.org/journal.php. Whilst CPGPS owns all the copyright of all text material published in the Journal, the authors are responsible for the views and statements expressed in their papers and articles. Neither the authors, the editors nor CPGPS can accept any legal responsibility for the contents published in the journal.

Subscriptions and Advertising

Membership with CPGPS includes subscription to the Journal during the period of membership. Subscriptions from non-members and advertising inquiries should be directed to:

CPGPS Headquarters
Department of Geomatics Engineering
The University of Calgary
Calgary, Alberta, Canada T2N 1N4
Fax: +1(403) 284-1980
E-mail: journal@cpgps.org

CPGPS Logo Design:
Peng Fang
University of California, San Diego, United States

Cover Design and Layout:
Satellite Navigation & Positioning Group
The University of New South Wales, Australia

Printing:
Satellite Navigation & Positioning Group
The University of New South Wales, Australia

Journal of Global Positioning Systems

Vol. 4, No.1-2, 2005

Table of Contents

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 1

Journal of Global Positioning Systems
2005
© CPGPS

Letter From the Guest Editor

Professor Chris Rizos, Guest Editor

School of Surveying and Spatial Information Systems, University of New South Wales, Sydney, NSW 2052, Australia.
e-mail: c.rizos@unsw.edu.au, Tel: +61 2 9385 4205; Fax: +61 2 9313 7493

I am pleased to be the guest editor for these special issues of the Journal of Global Positioning Systems to publish the selected papers from the 2004 International Symposium on GPS/GNSS (GNSS-2004), 6-8 December 2004, organised by the School of Surveying and Spatial Information Systems, University of New South Wales, Australia.

The GNSS-2004, attended by over 340 delegates from 29 countries, was the largest symposium in the Asia/Pacific region in 2004 dedicated to GNSS and wireless positioning. On the first day of the symposium the plenary session consisted of several presentations by invited speakers. The Civil GPS Service Interface Committee (CGSIC) convened a public meeting where representatives from the US Dept of Transport and US Coast Guard informed participants of recent developments in GPS policies and modernization. Presentations by several European speakers introduced the future GNSS "Galileo", scheduled for deployment by the end of the decade. The Symposium also featured technical workshops and around 200 oral and 'flashing poster' presentations.

Over 130 papers presented at the symposium were submitted to the special issues of the Journal of Global Positioning Systems (JGPS). However, given the limited space pages in these special issues many high quality papers could not be selected for publication. This has been a considerable administrative challenge and I wish to thank the reviewers for their assistance in reading and selecting the papers.

The 2004 and 2005 volumes of the JGPS feature the selected papers from the symposium on various GNSS topics, as well as research and development into a variety of other positioning technologies.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 2-11

Journal of Global Positioning Systems
2005
© CPGPS

Using Multiple Reference Station GPS Networks for Aircraft Precision Approach and Airport Surface Navigation

Ahmed El-Mowafy

Department of Civil and Environmental Engineering, UAE University, P.O.Box 17555, Al Ain, UAE
e-mail:Ahmed.Mowafy@uaeu.ac.ae; Tel: + 97150 7535840; Fax:+9713 7632382

Received: 6 December 2004 / Accepted: 16 October 2005

Abstract. The use of multiple real-time reference stations (RTK Networks) for positioning during the aircraft's precision approach and airport surface navigation is investigated. These existing networks can replace the proposed airport LAAS systems and have the advantage of improving coverage area. Real-time testing of the proposed technique was carried out in Dubai, UAE, with a helicopter and a small fixed-wing aircraft using a network known as the Dubai Virtual Reference System (DVRS). Results proved the feasibility of the proposed approach as they showed that cm to sub-meter positioning accuracy was achieved most of the time. For some periods, only meter-level positioning accuracy was available due to temporary breaks in reception of the network carrier-phase corrections. Some solutions to improve availability of the corrections are discussed. It is also proposed to integrate the GPS with an IMU. The inertial system aids positioning during periods when the corrections are lost, as well as providing attitude information. The GPS and IMU systems were integrated using a decentralized adaptive Kalman filtering technique. The measurement noise covariance matrix and the system noise matrix are adaptively estimated, taking the aircraft dynamics changes into account. Tests of the integrated system show that it has a good overall performance, and navigation at categories III and II can be achieved during short outages of RTK-GPS network corrections.

Key words: Global Positioning, Airborne Navigation, Wide Area Networks, Adoptive Systems, INS

1 Introduction

Interest in the use of Global Navigation Satellite Systems (GNSS) as a main source of navigation reference is increasing. The system employed for such a purpose should be capable of meeting the strict requirements of air navigation in terms of accuracy, availability, integrity, and reliability. At present, the accuracy requirements for all flight categories up to precision approach are summarized in Table 1 (Whelan, 2001). The accuracy requirement for Category I can be achieved most of the time using wide area differential systems such as the American "WAAS", the European "EGNOS", and the Japanese "MSAS". To meet the most demanding accuracy for categories II and III, which involve the final and precision approach phases of flight, more accurate systems are needed. The American Federal Aviation Authority (FAA) has undertaken the development of a navigation augmentation system based on GPS in the form of a Local Area Augmentation System (LAAS). This LAAS is designed to enable precision approach navigation within the airport area. It includes at least four reference GPS receivers located at each airport. GPS measurements are collected from the four reference stations and processed in real time in a control computer. Next, GPS differential corrections are sent to aircraft to compute their locations for navigation at the sub-meter level of accuracy. Corrections are provided via a Very High Frequency (VHF) radio link from a ground-based transmitter. LAAS preliminary test results have generally demonstrated accuracy of less than 1 meter in the horizontal and vertical axes. However, the percentage of system availability is still under evaluation to see if it can meet the FAA requirements. The cost of establishing LAAS for major airports is also expected to be significant.

Tab. 1 Positioning accuracy requirements for all flight categories

	horizontal	**vertical**
Category I	17.1 m	4.1 m
Category II	5.2 m	1.7 m
Category III (precision approach)	4.1 m	0.6 m

This study proposes the use of Real-Time Kinematic (RTK) multi-station reference networks, as an alternative to the airport LAAS, to aid accurate positioning of aircraft during precision approach, takeoff and airport surface navigation. The feasibility of this approach, the problems experienced in practice and possible methods for improving the overall performance are investigated in this paper. The investigation is carried out on a typical RTK network, bearing in mind that the presented results are dependent, to some extent, on the network design and operational features.

2 Using the multi-station RTK networks for airborne navigation

GPS RTK multi-station reference systems were originally developed for surveying applications. The basics of this type of reference systems are discussed in Enge et al. (2000), Raquet & Lachapelle (2001), Hu et al. (2003) and El-Mowafy et al. (2003). In principle, observations from multiple reference stations covering a large area are gathered and processed in a common network adjustment at a central processing facility and measurement corrections are computed. The corrections are optimized for the coverage area to account for distance dependent errors. A single rover GPS receiver receives these measurement corrections from the control centre of the network and uses the corrections to estimate its position in real-time accurate to the cm-level with fixed integer carrier-phase ambiguity resolution, or to the sub-meter level with a float solution. Currently, the RTK network approach is mostly used in static or kinematic ground applications. In this study, the use of these networks in airborne navigation is considered, where the rover receiving the network corrections is mounted on the aircraft to determine its positions during flight.

2.1 Advantages of Using RTK Networks in Airborne Navigation

The main advantages of using multi-station reference RTK networks for precise airborne navigation can be summarized as follows:

- Due to the fact that multi-station reference networks usually have an area of coverage that extends to several tens or hundreds of kilometres, each network can cover more than one airport, including small airports, unlike the airport LAAS. In addition to airport navigation, the system can be used in search and rescue operations, emergency landing, road traffic control from the air, as well as emergency response.

- RTK networks provide cm to decimetre positioning accuracy even in the case of malfunctioning of some stations, particularly for dense networks. This situation is however more critical in airport LAAS due to the low number of receivers used.

- Compared to LAAS, no significant additional infrastructure cost is involved as the hardware and software of the GPS-RTK networks are available in most developed countries and the establishment of new networks is currently underway (or planned) in different regions worldwide.

- RTK networks can give better runway utilization by improving airport surface navigation. They can also enhance air traffic management by increasing dynamic flight planning.

2.2 The DVRS Network as an Example

The feasibility of using real-time reference networks to provide precise positioning navigation information for aircrafts is examined in this study. A network known as the Dubai Virtual Reference System (DVRS), located in Duabi, UAE, was used for this investigation. The focus was on various aspects of aircraft navigation including precision approach, takeoff and airport surface navigation. For accurate determination of aircraft heights from the ground using GPS-derived ellipsoidal heights, a recently established accurate geoid model for Dubai was utilized. The Dubai geoid model was developed from varying data sources, mainly: gravity measurements, a digital elevation model (DEM), orthometric heights and GPS-observations at levelling benchmarks.

The DVRS network consists of five active reference stations, with baseline lengths varying between 23.4 km and 90.8 km. The main software used in the processing of the DVRS data utilizes the area parameter method (FKP) to estimate and represent the state of individual GPS errors in real time. All stations of the network are processed simultaneously using un-differenced observables. Therefore, all error components including clock errors are estimated. The state vector ($\vec{X}$) used in the Kalman filtering process can be given as:

$$\vec{X} = (\vec{x}_i,\ N_i^s,\ \delta t_i,\ \delta t^s,\ \delta\vec{O}^s,\ \delta T_i^s,\ \delta I_i^s,\ \delta M_i^s)^T \quad (1)$$

where $\vec{x}_i$ is the position vector, δt_i and δt^s denote the receiver and clock errors, N_i^s is the ambiguity, $\delta\vec{O}^s$, δT_i^s and δI_i^s represent the distance dependent errors (the orbital, tropospheric, and ionospheric errors respectively), and δM_i^s is the multipath error. To compute its position, the rover receiver sends its approximate position via a cellular modem to the network control centre where computations are carried out for each user. The estimated network measurement corrections, mainly the distance dependent errors, are interpolated for a virtual reference station (VRS) close to the rover position and instantly sent to it. The predicted distance dependent error term (δ_i) at the VRS position (i) from the reference station (j) with respect to the satellite (s) can be expressed in the functional form:

$$\delta_i = f(\mathrm{FKP}_j^s, \Delta\phi_{ji}, \Delta\lambda_{ji}, \Delta h_{ji}) \qquad (2)$$

where Δ is the differential operator, ϕ, λ, and h denote the latitude, longitude and height respectively, and FKP_j^s represents the FKP computed error. The corrections at the VRS station (VRS_{ji}^s), which are used to correct the observations at the rover receiver, can be expressed as follows:

$$\mathrm{VRS}_{ji}^s = \mathrm{CR}_{ji}^s + \delta_i + \Delta T_{ji} \qquad (3)$$

where CR_{ji}^s denotes the corrected carrier phase observations of the reference station computed from the network solution, and ΔT_{ji} represents the difference in tropospheric modeling between processing of the network at the reference station and processing of the virtual reference station. For in-depth mathematical formulation of this method, interested readers may refer to Wübbena et al. (2001). Previous testing of the DVRS system for kinematic ground surveying showed that system positioning accuracy was typically 1-2 cm in planimetry and 3-5 cm in altimetry (El-Mowafy et al., 2003).

2.3 Concerns and Recommendations in Using the DVRS Network for Airborne Navigation

When applying the VRS technique to airborne navigation, the aircraft rover receiver uses a ground VRS station. The drawback is that continuously updated approximate coordinates have to be used for the VRS computation. This is similar to having a moving reference station. A system reset should thus be frequently performed when the VRS coordinates are changing, which will result in frequent initialization of the carrier-phase ambiguities. Therefore, it is preferable to keep the VRS location for the longest possible range and apply appropriate extrapolation. This can, however, affect the performance of the system. In addition, the duplex communication approach used for the DVRS network puts a restriction on the number of users, as this number is limited by the ability of the control centre to simultaneously perform calculations for different users. As this number grows, extended latency in receiving the corrections may result.

These problems can however be alleviated in the implementation phase of the system in aviation by using a one-directional communication method. In this case, one or two ground transmitters (repeaters) at the airport will be established; they will receive the reference-station measurement corrections from the control centre on-line and send them to the aircraft by means of, for instance, VHF modems. The receiver on board the aircraft will then be responsible for interpolating the corrections at its location and processing the measurements to estimate its position. Thus, the rover can independently use its own interpolation and processing models, and no restrictions exist on the number of users. For faster and continuous prediction of the corrections at the rover location, it is recommended that the software computes a particular set of aviation corrections sent to the airport transmitters, emphasizing the airport area with a preset radius (e.g. 30-40 km). The current architecture of the DVRS communication with the user can, however, be kept to serve ground-based surveying applications. Hence, both types of communication can be used to serve different applications (aviation and surveying), using the same infrastructure of the real-time reference stations.

The establishment of ground transmitters at the airport can also improve the current availability of the corrections to the rover receiver, as breaks in receiving such corrections frequently take place. Another recommendation is to integrate GPS with an inertial system. More details will be given in a following section.

Since the proposed system is based on the use of satellite measurements, the integrity of the system and continuous reception of the corrections are primary concerns. These items should be continuously monitored, and methods such as RAIM should be implemented to warn the pilot against any deficiency in the system. Other concerns in the use of RTK networks in airborne navigation include:

- Due to the high dynamics involved in airborne navigation, a high update rate of sending the corrections is needed compared with that implemented for land applications, which is usually 5-70 seconds for current networks. This rate has a direct impact on the Time-To-First-Fix of phase ambiguities, and thus on the overall positioning feasibility and accuracy (El-Mowafy, 2004).

- The format of GPS measurement corrections should be standardized to ensure that the system is independent of any single receiver manufacturer. The use of the RTCM standard for RTK multiple reference stations v3.0 is thus recommended, see Euler et al. (2004).

- The need to ensure the security of the reference station locations: these stations should be safe and unreachable by the public in order to prevent possible tampering.

- The possibility that the airport authorities share control of the system with surveying authorities is recommended.

3. Testing the DVRS system for Airborne Navigation

3.1 Test Description

The use of the DVRS network for aircraft navigation in the airport area was evaluated by conducting several flight tests. Two types of aircraft were used for this purpose, a helicopter and a small fixed-wing airplane. In these tests, aircraft positions (planimetric + height) were determined using a dual-frequency GPS rover receiver (Leica SR530) equipped with a DVRS GSM modem capable of receiving the DVRS corrections. The GPS/DVRS rover receiver collected both the GPS and the correction data during the aircraft takeoff, enroute flying, landing and airport surface navigation. The data were processed in real time at one-second intervals. On the other hand, the DVRS reference stations collected data at five-second intervals. Processing of their RTK network corrections was also carried out at this interval. Thus, the corrections were interpolated in time for the rover receiver to compute the measurement corrections at the one-second interval. The satellite window during testing was generally normal, and 6 to 8 satellites were observed at any moment, giving Dilution of Precision (PDOP) values ranging from 1.4 to 3.7 at the rover receiver locations. The GPS data were also stored in the receiver's internal memory for further post processing testing and analysis after being integrated with the data from the DVRS reference stations.

In the helicopter test, the GPS and GSM antennae were rigidly mounted on an arm approximately 0.9 m long extending outside the helicopter. The arm was attached to a frame rigidly fixed inside the helicopter. No arm vibration was experienced during flight testing. For better GPS as well as GSM signal reception, the GPS antenna was mounted high on the arm for better visibility of the sky, while the GSM antenna faced down. This architecture was designed only for testing purposes. Figures 1 and 2 show the system installation on the helicopter, and the test trajectory. For the fixed-wing aircraft test, the GSM antenna was installed inside the aircraft, which is acceptable for GSM communication. Figures 3 and 4 show the fixed-wing aircraft and the test trajectory respectively. Both tests were carried out over the city of Dubai.

Fig. 1 System instillation for the helicopter test

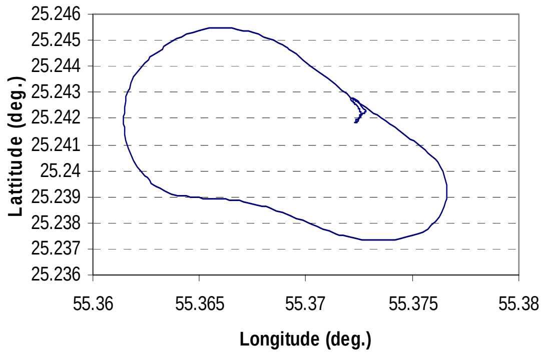


Fig. 2 Trajectory of the helicopter test

Fig. 3 Testing using the fixed-wing aircraft

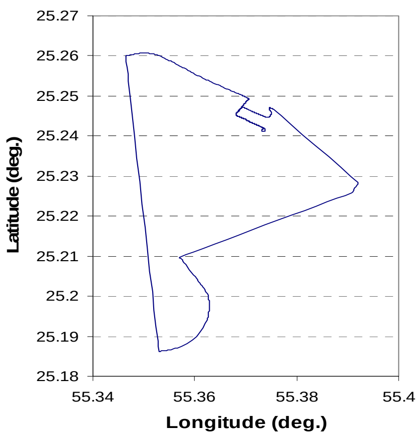


Fig. 4 Trajectory of the fixed-wing aircraft test

3.2 Test Results

Figure 5 shows the helicopter test results, illustrating the flying height and the 2-D and height positioning accuracies achieved during testing. At the beginning of the test and after an initial warming up period of less than 20 seconds, the phase ambiguities were successfully fixed. Thus, positioning accuracy was feasible at the cm level before starting the engine. The DVRS corrections were continuously received during takeoff until reaching the required height (first dashed region in Figure 5), which was approximately 145m. During the major part of the enroute flying time, the DVRS corrections were continuously received. However, during most of the landing phase the DVRS corrections were lost, but were regained after the helicopter landed (second dashed region in Figure 5). This can be mainly attributed to the use of GSM signals in sending the DVRS data, and partially to changes in the helicopter dynamics. In addition, due to changes in the VRS positions, initialization of the phase-ambiguities was often carried out. The change in error values from the decimetre to the cm level, which can be observed at some instances in the figure, can be ascribed to reaching a fixed ambiguity solution after a float solution. In general, during the two marked periods, the ambiguities were resolved as integers and the average positioning accuracy, represented by coordinate standard deviations, was 0.022 m in the planimetric 2-D positions and 0.034 m in height.

One can, however, note that the highest accuracies needed in airborne navigation corresponding to category III, which are 4.1 m for 2-D positioning and 0.6m for height determination, can generally only be achieved with a float ambiguity resolution. For instance, for the helicopter test, during the periods where the DVRS corrections were received but the ambiguities were resolved in a float solution, the positioning accuracy was at the sub-meter level, as shown in Table 2. This accuracy was on average 0.322 m for the 2-D positioning and 0.539 m for height determination. However, when the DVRS signals were not received, the 2-D positioning and the height errors were increased to more than 3.5 m, which are only suitable for navigation at category I, i.e. during the enroute flying.

Tab. 2 Average positioning accuracies (m)

	Helicopter test		Fixed-wing aircraft test	
	2D (E&N)	Height	2D (E&N)	Height
fixed solution	0.022	0.034	0.016	0.028
float solution	0.322	0.539	0.263	0.525
all test periods	0.484	0.642	1.107	0.831

Figure 6 shows the results of the fixed-wing aircraft test. In this test, the DVRS corrections were available during airport surface navigation and manoeuvring to the runway, in addition to the periods of takeoff, reaching the designated height, landing and parking, which are shown in the dashed areas in the figure. The DVRS corrections were, however, lost during flying for some periods. This can also be attributed to the use of GSM signals as the means of communication with the DVRS centre, which might result from changes in the aircraft dynamics as clearly seen from the top figure. This result, together with the outcome of the helicopter test, shows that accurate positioning using RTK network corrections in real-time is feasible during the critical phases of takeoff, landing, and airport surface navigation. However, the use of GSM signals for sending the RTK network corrections is not efficient and other methods are needed. The average 2-D and height positioning accuracies achieved during the fixed-wing aircraft test are given in Table 2. As with the helicopter test, when receiving the DVRS corrections and initializing fixed phase measurement ambiguities, the average 2-D and height positioning accuracies were at the cm level, as they were 0.016 m and 0.028 m respectively. When phase ambiguities were solved in a float solution, these accuracies were 0.263 m and 0.525 m. Both cases are sufficient for positioning in all phases of flight, including category III. When the DVRS corrections were lost, the 2-D and height positioning errors were more than 4 m, which can be only used for category I of navigation.

To compare results of the RTK network approach for this particular application with the standard double differencing technique, the phase data of the rover receiver stored in its internal memory were processed in a post-mission mode referenced to one of the DVRS network reference stations. This was possible since the tests were carried out at a distance of approximately 6 km from this station and the flight paths were within a range of a few kilometres. Precise IGS orbits were used. When comparing the results of positioning obtained by the DVRS real-time multi-station reference network with the post-mission positions, the 3D differences were within the range of a few millimetres to a few centimetres when the phase ambiguities were fixed. In general, the discrepancies were less than 7 cm.

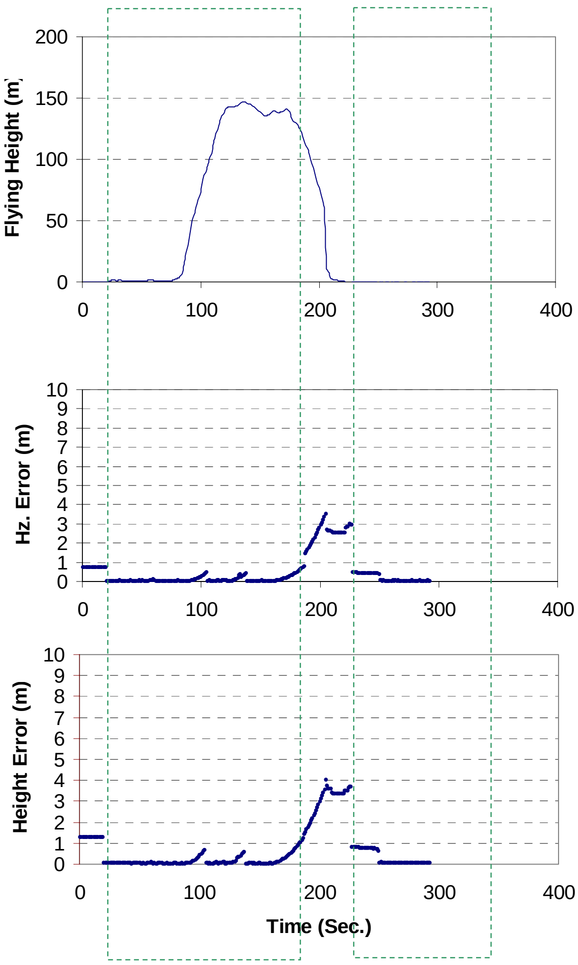


Fig. 5 Helicopter test results

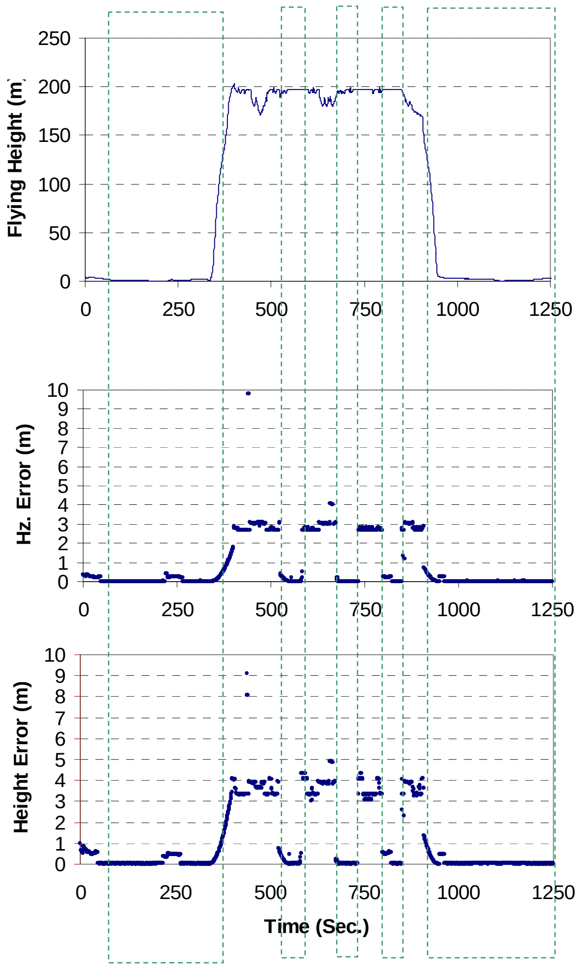


Fig. **6** Fixed-wing aircraft test results

4 Integration with the INertial system

4.1 Integration and Estimation Methodology

One method to increase the availability of the positioning accuracy at the required level is to integrate GPS with an Inertial Measuring Unit (IMU). Thus, for the same tests given above, the GPS/DVRS system was integrated with an IMU running simultaneously, with the purpose of bridging positioning outage by the Inertial Navigation System (INS) of the IMU during short breaks in reception of network corrections. The data of both systems were recorded for post mission processing and analysis. For testing purposes and due to hardware availability, a Honeywell tactical-grade (medium accuracy) IMU system of approximately 1-10 degrees/hour gyro drift was used. For simplicity, the GPS/INS integration was carried out in a decentralized loose coupling scheme. In this approach, the GPS and IMU (INS) filters ran independently in parallel. The GPS filter used the rover data and the corrections received from the multiple-reference station network as an input to the filter. The states are given in Equation (1), which includes the rover's position, phase ambiguities and measurement errors. The state vector of the INS comprises the misalignment, position, velocity, gyro drifts and accelerator biases. Positions determined from the GPS filter and velocity estimates were used as an update to the INS filter. Since real-time processing was required, no bridging algorithms such as backward smoothing were applied.

For the purposes of the test, positioning by the INS was mainly investigated during bridging of the GPS positioning outages for the short breaks in reception of network corrections. Figure 7 shows a flowchart of the integration scheme of the GPS/INS adopted during testing. To externally evaluate the performance of the INS positioning during the network correction outages, the rover receiver kept on collecting GPS observations, which were processed in a post-mission mode referenced to one of the DVRS network reference stations to compute position data that were compared with the INS results. Apart from this testing purpose, positioning information can be generally acquired from the IMU in an integrated GPS/INS system to benefit from its high frequency output. In addition, the INS is useful for determination of the attitude information of the aircraft, as well as cycle slip detection and repair, and ambiguity resolution, if a centralized filtering scheme is used.

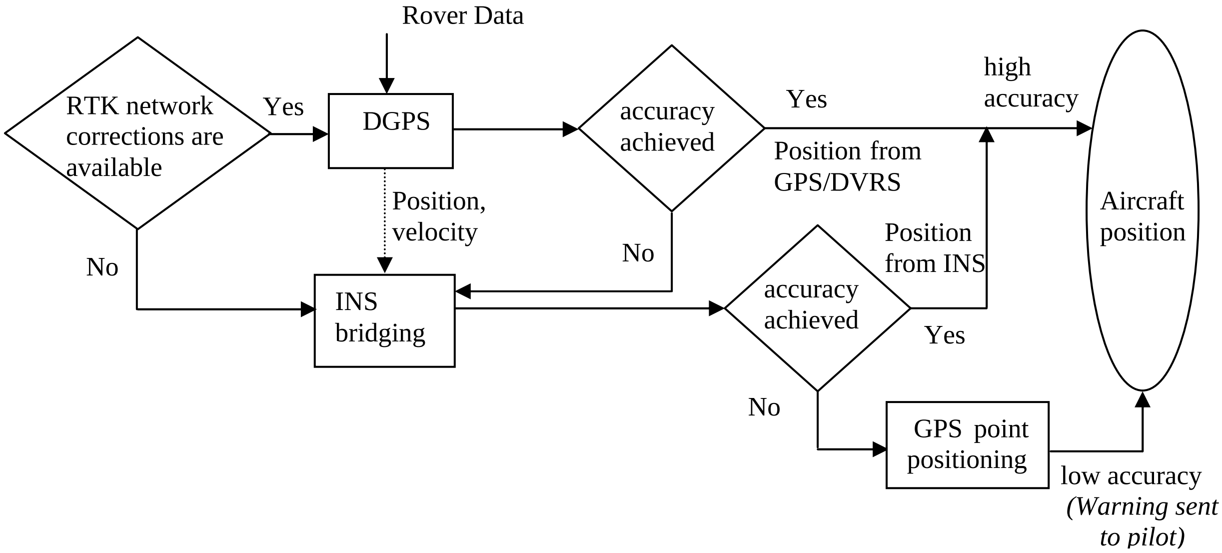


Fig. 7 Flowchart of GPS/INS integration for testing purposes

The mathematical models used in the filtering estimation approach can be written in matrix form as:

Dynamics model: $\dot{S}_i = F_{i,i-1}\, S_i + G\, u_i$ (4)

taking, for simplicity, $\Phi_{i,i-1} = F_{i,i-1}\, dt + I$ (5)

Observation model: $M_i = h(S_i) + e_i$ (6)

where S_i denotes the state vector, M_i is the measurement vector, $\Phi_{i,i-1}$ is the transition matrix, F represents the dynamics matrix, dt is the prediction time interval ($t_i - t_{i-1}$), and e_i represents the measurement noise. G is the design matrix and u_i denotes a forcing vector function, such that the term ($G\ u_i$) represents the noise of the dynamics model. This model for the INS is described using a first order Gauss-Markov process.

An extended Kalman filtering approach was used to represent the non-linear observation equations, where the filter states become estimated corrections (δ) to an approximate state (S_o) represented as a nominal time varying state updated using filter estimation, such that:

$$\delta = \hat{S} - S_o \tag{7}$$

The time update (prediction) equations take the form:

$$\delta_{i,i-1} = \Phi_{i,i-1}\, \delta_{i-1} \tag{8}$$

$$P_{i,i-1} = \Phi_{i,i-1}\, P_{i-1}\, \Phi^{T}_{i,i-1} + Q_{i-1} \tag{9}$$

$$K_i = P_{i,i-1}\, H_i^T\, (H_i\, P_{i,i-1}\, H_i^T + R_{i-1})^{-1} \tag{10}$$

The measurement update (information) equations can be formulated as:

$$\delta_i = \delta_{i,i-1} + K_i\, (\omega_i - H_i\, \delta_{i,i-1}) \tag{11}$$

$$P_i = (I - K_i\, H_i)\, P_{i,i-1} \tag{12}$$

where Q and R denote the covariance matrices of the dynamics model and the measurement model respectively. P is the covariance matrix of the filter states, H_i represents the partial derivatives (linearized design matrix) derived from the observation equation, K is the Kalman gain matrix, and ω symbolizes the measurement misclosure.

For the INS filter, the measurement equations can be formulated as follows:

$$M_i = \begin{Bmatrix} (\varsigma + h)(\phi_{IMU} - \phi_{GPS}) \\ (\xi + h)\cos\phi(\lambda_{IMU} - \lambda_{GPS}) \\ h_{IMU} - h_{GPS} \\ \hdashline v^n_{IMU} - v^n_{GPS} \\ v^e_{IMU} - v^e_{GPS} \\ v^d_{IMU} - v^d_{GPS} \end{Bmatrix} + e_i \tag{13}$$

where ζ and ξ are the radii of curvature for the meridian and prime vertical. v^n, v^e and v^d are the velocity components in the navigation frame axes (north, east, down).

The measurement noise matrix can be estimated from:

$$R = \mathrm{diag}\,(\ \sigma_{\phi}^2\ \ \sigma_{\lambda}^2\ \sigma_{h}^2\ \ \sigma_{v^n}^2\ \ \sigma_{v^e}^2\ \sigma_{v^d}^2\ ) \tag{14}$$

where σ_{v^n} , σ_{v^e} and σ_{v^d} denote the standard deviations of velocity. The initial position and velocity standard deviations are taken from the GPS solution.

The Q matrix can be calculated from (Shin, 2001):

$$Q = \Phi\ G\ q\ G^T\ \Phi^T\ dt \tag{15}$$

where q is the spectral density matrix computed as:

$$q = \mathrm{diag}\,(\ \sigma_{ax}^2\ \ \sigma_{ay}^2\ \sigma_{az}^2\ \ \sigma_{\psi x}^2\ \ \sigma_{\psi y}^2\ \sigma_{\psi z}^2\ ) \tag{16}$$

the σ_a and σ_ψ are the standard deviations of the accelerometers and gyroscopes, respectively.

Both the Q and R matrices play a main role in determining the quality of the estimated states owing to the fact that the predicted states covariance is affected by the Q matrix, while the update measurements covariance is R. The change of these covariance matrices reflects changes in the system dynamics, which represent a major factor affecting the performance of the tactical-grade IMU system. Thus, for medium accuracy IMU, tuning of the Q matrix is crucial to achieve filter stability. Hence, arrangement of the Q and R matrices in an adaptive manner can improve estimation, as they would dynamically reflect the actual situation. Prior field-testing results for a kinematic ground survey showed that the adaptive Kalman filter approach outperformed the conventional approach, both on internal and external bases (El-Mowafy and Mohamed, 2005). It was also shown that the track ability of the adaptive filter for the filter states was much better than that of the conventional filter.

For the above reasons, an adaptive Kalman filtering approach was employed in the processing of the test data. In this approach, the residual sequence η_i was first computed as:

$$\eta_i = M_i - h(S_i) \tag{17}$$

Then, the adaptive formulation of the R and Q matrices followed the following formulations (Mohammed and Schwarz, 1999):

$$C_\eta = \frac{1}{N} \sum_{k=k_0}^{i} \eta_k\ \ \eta_k^T \tag{18}$$

$$R_i = C_\eta\ + H_i\ P_i\ H_i^T \tag{19}$$

$$Q_i = K_i\ C_\eta\ K_i^T \tag{20}$$

where C_η is the covariance matrix of the residual sequence, and using $k_o = i - N + 1$ as the first epoch inside the estimation of a moving time window of the size (N), which can be taken as 20-30 epochs.

4.2 GPS/INS Integration Results

When breaks in reception of network corrections take place, an extrapolation of these corrections continues for a few seconds; after that GPS solution accuracy deteriorates. As a result, the GPS positions and velocity input are de-weighted in the filter, and the INS works in a stand-alone mode. Thus, the acceptable period of outage in reception of the network corrections is the summation of the extrapolation period of the network corrections, during which the GPS still provides positioning accuracy at the cm to decimetre level, and the time through which the INS positioning accuracy in a stand-alone mode without GPS updates is within the accuracy required for navigation. In the case of regaining GPS observations with network corrections, the time needed to resolve the ambiguities should be included in the GPS positioning outage period. For the tests in hand, an outage in receiving the network corrections occurred after the dashed areas, illustrated in Figures 5 and 6. The INS stand-alone positioning errors grew very rapidly with time in a non-linear form. However, unlike GPS, the tested IMU system has a better height determination accuracy than its horizontal accuracy. This is advantageous for airborne navigation, which is more restricted by the height accuracy. For instance, the maximum allowable height error for category III is 0.6m, while it is 4.1m for the horizontal error.

The test results showed that the accuracy requirements for precision approach (category III) were generally achieved within 25-31 seconds of the GPS data outages. This was dependent to some extent on the aircraft dynamics. During enroute flying, the aircraft generally had uniform dynamics, which resulted in a longer positioning outage bridging, while during takeoff and landing, more changes in dynamics took place, which resulted in shorter coverage of outages. In addition, during curved parts of the course, the INS performed less well than during straight flying. Thus, shorter position bridging periods were recorded during curved flying. Overall, for data outages up to 43 seconds, the positioning accuracy achieved was suitable for category II. After that, the vertical positioning error was several meters, which is only suitable for category I of airborne navigation. These results, however, correspond to the used system and may differ for other IMU systems. The performance of the tested INS in the stand-alone mode during positioning bridging of the GPS data outages is shown in Table 3 for the helicopter and the fixed-wing aircraft tests. The average and maximum standard deviations of the planimetric (horizontal) and height components are given for GPS network data outages of 20 seconds and 40 seconds.

Tab. 3 Standard deviations of the INS positioning results during GPS data outages (m)

Period of GPS data outages	Helicopter test				Fixed-wing aircraft test			
	2D (E&N)		Height		2D (E&N)		Height	
	Avg.	Max.	Avg.	Max.	Avg.	Max.	Avg.	Max.
20 seconds	2.635	3.725	0.354	0.582	2.140	3.971	0.310	0.534
40 seconds	4.161	5.103	0.915	1.662	3.651	4.837	0.832	1.388

Although the system hardware used and their integration processing schemes still have room for improvement, this configuration was tested to investigate the feasibility of the presented concepts, namely: using the multiple reference station RTK GPS networks for precision airborne navigation, and the ability of an integrated GPS/INS system to bridge positioning during short breaks in reception of network corrections. Other GPS/INS integration types, including tight coupling with centralized filtering, are currently under investigation. Tight coupling, as compared with loose coupling, is expected to provide a solution for longer data outage periods. Partial GPS data of less than 4 satellites, which only give under-determined solution, can also be used. In addition, the INS can help in detecting and correcting cycle slips, and aiding ambiguity resolution, see for instance Wu (2003). Other studies (e.g. Petovello, 2003) have also shown that the tight integration approach outperforms loose integration approaches in terms of the overall system accuracy, due to the reduced amount of process noise in the tight integration. The duration range of the INS positioning bridging under different operational conditions is also under investigation. However, this will vary according to the quality of the IMU used. For instance, better results can be achieved with higher accuracy systems (navigation grade IMU) compared with the medium accuracy IMU system used in this test.

5 Conclusions and Recommendations

The test results show that the use of RTK multi-station reference networks (e.g. the DVRS network) in precise aircraft navigation is feasible, particularly for the airport area. This new technology can increase the coverage area compared with other GPS-navigation systems, such as airport LAAS, with significant cost reduction. Small airports can thus benefit from this service. For the test flights conducted, the DVRS GSM signals were received during most of the testing periods. The loss of the signals, which took place for some periods, was expected due to the use of GSM signals and changes in aircraft dynamics. During the majority of periods of receiving the DVRS corrections, the phase measurement ambiguities were fixed and the average positioning accuracies were less than 4 cm. The accuracy needed for category III was achieved even with a float ambiguity resolution.

One can see from these results that to achieve the accuracy requirement of all phases of flight using the DVRS system, it is necessary to guarantee continuous transmission of the DVRS corrections in a suitable form for civil aviation. One suggestion to achieve this goal is by establishing ground transmitters at the airport. These transmitters will receive the corrections from the network control centre on-line and send them to the aircraft using, for instance, VHF modems instead of the currently used GSM modems. This can be implemented in the update phase of the DVRS network. A one-way direction of communication from the ground transmitter to the aircraft is recommended. In addition, it is advisable to add one reference station in the vicinity of the airport to enhance correction estimation in this area and act as a backup for the system, such that its corrections can be readily applied in case of any interruption in reception of the signals from the network control centre.

In the final implementation phase, the integrity of the system should be fully ascertained, with systems that can warn the pilot in case of system accuracy and availability deficiency being added as necessary. The security of the reference station locations should also be maintained at the highest levels. In addition, the format of the GPS corrections sent should be standardized so as to be independent of any single receiver manufacturer. This can be achieved by adopting the upcoming RTCM version 3.0 multi-station reference RTK standards.

One way to increase the availability of the positioning data is to integrate GPS with an Inertial Measuring Unit (IMU). Test results with medium accuracy IMU integrated using an adaptive Kalman filtering showed that positioning bridging can give acceptable results for category III and II if breaks in GPS solution availability are less than 40 seconds on average. After this period, without having new accurate GPS position updates, positioning errors grow and reach several meters. This accuracy is only suitable for category I of airborne navigation. However, better results can be achieved if navigation-grade IMU systems are used.

Acknowledgements

Dr. H. Fashir, Mr. A. Al Habbai, and Mr. Y. Al Marzooqi from the Survey Section, Planning & Survey Department, Dubai Municipality are sincerely acknowledged for providing the DVRS system, the logistics support and help in testing the system. The Dubai Air Wing is gratefully acknowledged for providing the test airplanes, airport facilities, and air flight crew, with special thanks to Captain G. Waddington. This study was partially funded by a grant from the Research Affairs Sector, the UAE University.

References

El-Mowafy A., Fashir H.; Al Marzooqi Y.; Al Habbai A.; Babiker T. (2003): ***Testing of the DVRS National GPS-RTK Network***, In: Proceedings of the 8th ISU International Symposium, Strasbourg, France, 25-28 May, 2003.

El-Mowafy A. (2004): ***Analysis of the Design Parameters of Multi-Station Reference RTK GPS Networks***, Journal of Satellite and Land Information Science, ACSM, Vol. 65, No. 1, 17-26.

El-Mowafy A.; Mohamed A. (2005): ***Attitude Determination from GNSS Using Adaptive Kalman Filtering***, Journal of Navigation, Cambridge Press, UK, Vol. 58, No. 1, 135-148.

Enge P.; Fan R.; Tiwari A. (2000): ***GPS Reference Network - New Role Providing Continuity and Coverage***, GPS World, Vol. 11, No. 1, 56-62.

Euler H-J.; Seeger S.; Zelzer O.; Takac F.; and Zebhauser B.E. (2004): ***Improvement of Positioning Performance Using Standardized Network RTK Messages***, In: Proceedings of ION NTM, San Diego, CA, January 26-28, 2004.

Hu G.R.; Khoo H.S.; Goh P.C.; and Law C.L. (2003): ***Development and Assessment of GPS Virtual Reference Stations for RTK Positioning***, Journal of Geodesy, Vol. 77, 292–302.

Petovello M. (2003): ***Real-Time Integration of a Tactical-Grade IMU and GPS for High-Accuracy Positioning and Navigation***, UCGE Report No. 20173, Dept. of Geomatics Engineering, University of Calgary, Canada.

Raquet J.; Lachapelle G. (2001): ***RTK Positioning with Multiple Reference Stations***, GPS World, Vol. 12, No. 4, 48-53.

Shin E-H. (2001): ***Accuracy Improvement of Low Cost INS/GPS for Land Applications***, UCGE Report No. 20156, Dept. of Geomatics Engineering, University of Calgary, Canada.

Whelan C. (2001): ***Evolution of Air Navigation Services during this Decade***, Journal of Navigation, Vol. 54, No. 3, 447-462.

Wübbena G.; Bagge A.; Schmitz M. (2001): ***Network-Based Techniques for RTK Applications***, Presented at the GPS JIN 2001 Symposium, GPS Society, Japan Institute of Navigation, Tokyo, Japan, November 14-16.

Wu H. (2003): ***On-The-Fly GPS Ambiguity Resolution with Inertial Aiding***, UCGE Report No. 20186, Dept. of Geomatics Engineering, University of Calgary, Canada.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 12-26

Journal of Global Positioning Systems
2005

An Integrated Position and Attitude Determination System to Support Real-Time, Mobile, Augmented Reality Applications

Allison Kealy & Stephen Scott-Young

Department of Geomatics, The University of Melbourne, Victoria, Australia
Tel: +61 (0)3 8344 6804 Fax: + 61 (0)3 9347 2916, Email: akealy@unimelb.edu.au

Received: 6 December 2004 / Accepted: 26 October 2005

Abstract. Augmented reality (AR) technologies enable digitally stored information (virtual objects) to be overlaid graphically on views of the real world. As such, they are able to significantly enhance decision-making and operational efficiency in complex environments. AR technologies typically comprise a fusion of positioning and attitude sensors with visualisation capability and an information processing system. The decreasing size and cost of visualisation and positioning hardware and the increasing portable processing power of laptop and handheld computers now offer enormous potential for the development of intelligent solutions based around real-time, mobile AR technologies.

For any application built around AR technologies, its effectiveness lies in the accuracy to which the virtual objects can be aligned with views of the real world. For many of these applications, this is directly a function of the accuracy to which the position and orientation of the operation platform can be determined. This paper presents an integrated positioning system that combines an array of dual frequency GPS receivers, a fibre optic gyroscope and vehicle odometer within a centralised Kalman filter. It assesses the accuracy of the filter outputs of position and attitude as appropriate to supporting real-time, mobile AR applications. The design and testing of an AR prototype that combines the Kalman filter state with real-time imagery containing augmented objects will also be presented. Finally, approaches adopted to tune the filter and reduce inherent sensor noise, as well as results from a case study undertaken within the land mobile environment will be described.

Key words: Augmented Reality, GPS, Inertial Sensors, Kalman Filter, Integrated Systems

1 Introduction

Over the last decade there has been an increasing trend towards the development of complex decision making systems that utilise spatial information. Whilst this has largely been attributed to developments in the fundamental technologies used to acquire, analyse and visualise spatial data, parallel developments in enabling technologies such as mobile computing and wired and wireless communications have also contributed significantly to increasing the diversity of applications and users that rely on spatial information.

This research investigates the use of Augmented Reality (AR) technologies as an innovative approach to presenting spatial information in an understandable, user-friendly way through an enhancement of a user's real-world perspective view. AR technology is not new and has already experienced some success in many areas such as powerplant maintenance procedures (Klinker *et al.*, 2001) and cardiac surgery (Devernay *et al.*, 2001), However, current generation AR systems suffer from many limitations. These include display systems that are often difficult to view in a wide range of environments (particularly outdoors); delays in displaying augmented information at the appropriate time or position caused by the time required to process data from the AR sensors and its databases; a time consuming calibration process of the AR sensors; a lack of interactivity between the user and the AR system; and difficulties in determining the location of the user in outdoor environments without the prior preparation of placing markers with known locations around the area (Azuma *et al.*, 2001).

The limitations of AR operation in unprepared environments forms the basis of the research problem addressed in this paper, that is, to develop a position and attitude determination component for AR systems capable of operation in unprepared environments. In particular, it investigates the integration of measurements from the Global Positioning System (GPS) and Dead Reckoning (DR) with intelligent information obtained from map matching techniques to enable the continuous and accurate real-time visual alignment of three-dimensional objects within the perspective view of a user operating in outdoor unprepared environments.

This paper presents the details of the Kalman filtering algorithm developed to calculate the position and attitude parameters, as well as the approaches adopted to "tune" and "constrain" the Kalman filter solution for operation within the land mobile environment. Practical test results using an AR prototype developed within this research are also presented to validate the performance of the integration algorithms.

2 Position and attitude determination

AR systems rely on position and attitude parameters to register augmented objects with the "real world" environment. The accuracy with which these parameters can be determined, as well as the availability of the solution, can have a significant effect on the success of the AR system as a whole.

To determine accurate and continuous outputs of position and attitude parameters (heading, pitch and roll), this research investigates the performance of an integrated system comprising an array of three Leica GPS 530® receivers operating in a Real Time Kinematic (RTK) mode, a fibre optic gyroscope and an odometer, see Figure 1.

Each GPS receiver is configured to transmit the GGA NMEA string 10 times per second to a processor (laptop computer). The fibre optic gyroscope is configured to output the rate of change of direction of the platform. It is also possible to connect the gyroscope to the engine computer management system of the land mobile platform used in this research and thereby obtain outputs of the distance travelled. The on-board processor was used to synchronise the measurements from all sensors, to collate the data and to calculate in real-time the position and attitude parameters.

To obtain optimal estimates of the position and attitude parameters required for the AR system developed in this research, a loosely coupled Kalman filter was used to integrate the measurements obtained from all available sensors with the geometric distances of the fixed antenna geometry and the spatial intelligence of a road network database accessed through map matching techniques.

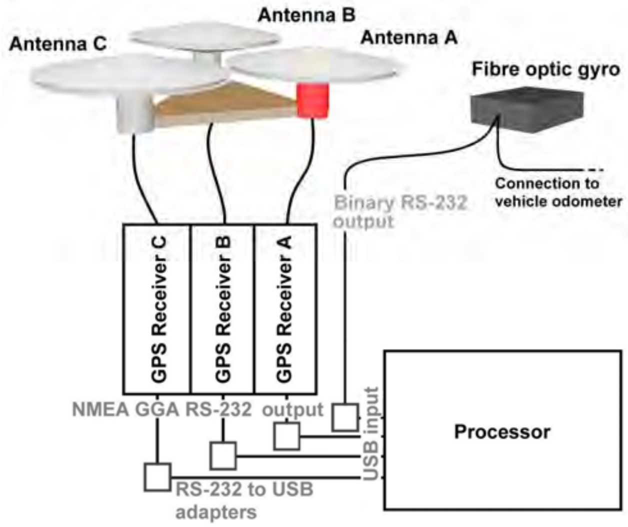


Figure 1 Schematic diagram of the hardware and the flow of data of the integrated positioning and attitude determination system

2.1 Reference Frames

To compute the attitude parameters for the mobile platform, the three GPS antennae (A, B, C) are used to define a platform reference frame. Figure 2 illustrates the GPS antennae configuration and the platform reference frame defined. The vector BC defines the pitch axis, the vector AD defines the roll axis, and the vector through D and perpendicular to the plane defined by the points A, B and C defines the heading axis. Antenna A acts as the platform origin and provides the position of the platform. A positive pitch rotation occurs when the platform tilts back (i.e., the front of the platform rises). A positive roll rotation occurs when the platform tilts to the right side, and a positive heading rotation occurs when the platform rotates to the right. For notation purposes, the platform reference frame axes are labelled with a subscript 'A' (originating from the fact that the platform reference frame is defined by the antennae). For computational efficiency the Geocentric Cartesian coordinates obtained from the GPS receivers are converted to the model reference frame (which is a local reference frame) with components referred to East, North and Up (EM, NM, UM) via a rotation matrix.

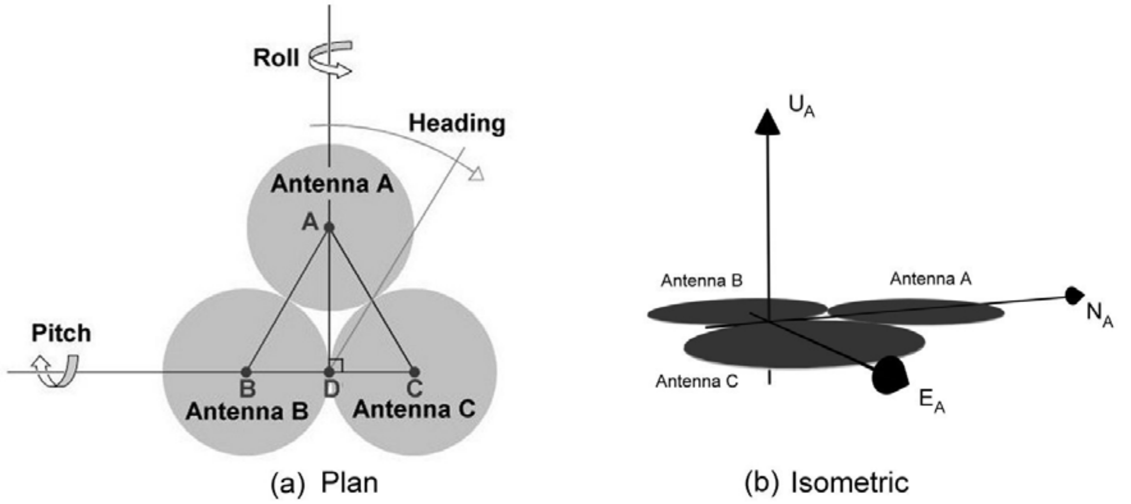


Figure 2 The platform reference frame: (a) roll, pitch and heading as defined by the fixed relative position of the three antennae A, B, and C, and (b) antennae referenced within an East, North and Up platform reference frame

2.2 Kalman filter integration models

To compute the position and attitude parameters for the mobile AR system, a Kalman filter was used to integrate the measurements obtained from the array of GPS antennae with those from the gyroscope and the odometer. As specific details on Kalman filtering algorithms and their implementations can be found in many references (eg Cross, 1990 and Logan, 2000), this paper focuses on the specification of the Kalman filter models generated for the AR prototype developed in this research.

2.2.1 Kalman filter functional model for the GPS observations

The approach taken in this research was to define the unknowns (or state, $\overline{x}_i$) to be solved, as the platform position in the model reference frame (i.e. the position of Antenna A ($\begin{bmatrix}\overline{E} & \overline{N} & \overline{U}\end{bmatrix}$, the master antenna) and the attitude parameters (i.e. heading, pitch and roll, $\begin{bmatrix}h & p & r\end{bmatrix}$), as presented in Equation (1).

$$\overline{x}_i = \begin{bmatrix}\overline{E} & \overline{N} & \overline{U} & \overline{h} & \overline{p} & \overline{r}\end{bmatrix}^T \tag{1}$$

The observations are the coordinates of the each of the GPS antennae in the model reference frame. To constrain the solution, the fixed spatial relationships between the GPS antennae are also included and are defined as offsets in Easting, Northing and Up of Antennae B and C from Antenna A, see Figure 3. Note that the platform reference frame has been defined in such a way that the offsets in altitude (i.e. the Up component in the platform reference frame) between the antennae are zero. Due to the high precision (approximately 0.2 millimetres) with which these offsets can be calculated using photogrammetric techniques and, since the antennae are rigidly fixed to the platform, the offsets are treated as constants.

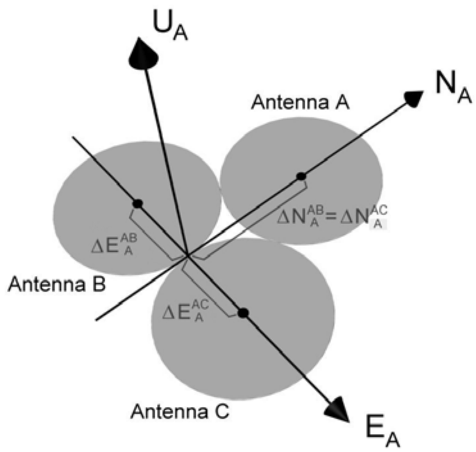


Figure 3 The spatial relationship between the antennae in the platform reference frame

The coordinates of Antennae A, B and C in the model reference frame can be defined in terms of the coordinates of Antenna A, the offsets between the antennae (defined in the platform reference frame), and the attitude (heading, pitch and roll) of the platform within the model reference frame. Hence:

$$A_M = \begin{bmatrix} E_M^A \\ N_M^A \\ U_M^A \end{bmatrix}$$

$$B_M = A_M + R_A^M \begin{bmatrix} \Delta E_A^{AB} \\ \Delta N_A^{AB} \\ \Delta U_A^{AB} \end{bmatrix} = \begin{bmatrix} E_M^A \\ N_M^A \\ U_M^A \end{bmatrix} + R_A^M \begin{bmatrix} \Delta E_A^{AB} \\ \Delta N_A^{AB} \\ \Delta U_A^{AB} \end{bmatrix} \qquad (2)$$

$$C_M = A_M + R_A^M \begin{bmatrix} \Delta E_A^{AC} \\ \Delta N_A^{AC} \\ \Delta U_A^{AC} \end{bmatrix} = \begin{bmatrix} E_M^A \\ N_M^A \\ U_M^A \end{bmatrix} + R_A^M \begin{bmatrix} \Delta E_A^{AC} \\ \Delta N_A^{AC} \\ \Delta U_A^{AC} \end{bmatrix}$$

where R_A^M is a rotation matrix that defines the rotation between the platform reference frame and the model reference frame. The rotation angles are simply the attitude parameters.

From Equations (2), the observation equations for the GPS measurements can be derived in the standard least squares observation equation form $F(\bar{x}) = \bar{l}$ (Equations (5)), where $\bar{x}$ is the vector of unknowns and $\bar{l}$ is the vector of observations. Note that from this point onwards unless otherwise indicated, values in the least squares adjustment equations without a subscript 'M' are in the model reference frame, allowing clearer use of subscripts to identify Kalman filter epoch values and predicted quantities.

2.2.1 Kalman filter dynamic model

For implementation of the Kalman filter, the vector of unknowns $\bar{x}$, must also include sufficient parameters to enable prediction of the platform state from one epoch to the next by modelling the mobile platform dynamics. Hence, 13 additional parameters (see Equation (3)) are introduced.

In this research, a polynomial model is used to predict heading, pitch, roll and velocity from one epoch to the next, while position is predicted using standard three-dimensional dead reckoning (DR) equations. The full dynamic model used in this research is shown in Equation (4).

The polynomial dynamic model for both heading and velocity are two orders higher than that of roll and pitch to cater for more frequent and larger magnitude changes in these parameters as would be expected from a land vehicle. Noise in the dynamic model is assigned within the Kalman filter through variances for the parameters $\overset{.....}{h}$, $\ddot{p}$, $\ddot{r}$ and $\overset{.....}{v}$.

2.2.3 Kalman filter functional model for the gyroscope and odometer

Similar to the GPS outputs, to develop the gyroscope and odometer observations equations, the measurements from these instruments are modelled in terms of the unknowns. For this project, the gyroscope is used to measure change in heading, while the odometer measures the distance travelled between measurement epochs. Thus, the observation equation model for the Kalman filter now contains two additional observations, and two additional observation equations. The unknown parameters in the Kalman filter are also increased by two to account for the inherent sensor biases of gyro drift rate and the odometer scale factor.

Equations (5) present the full functional models for all observations in the integrated positioning and attitude determination system developed in this research.

Where, $f_1(\bar{x},\bar{l})$ to $f_9(\bar{x},\bar{l})$ are the observation equations from the GPS antenna array. $f_{10}(\bar{x},\bar{l})$ to $f_{28}(\bar{x},\bar{l})$ are derived from the predictions within the Kalman filter and $f_{29}(\bar{x},\bar{l})$ to $f_{32}(\bar{x},\bar{l})$ are the observation equations derived from the gyroscope and odometer measurements, where;

$\dot{\beta}$ Gyro drift rate error (deg/s)

ε Odometer scale factor error

$\dot{h}$ Change in heading as measured by the gyro (deg)

d Distance travelled as measured by the odometer (m)

2.3 Tuning the Kalman filter

When implementing the Kalman filter, information about observation precisions, as well as the magnitude of noise in the dynamic model is required. In defining the dynamic model, estimates of variances for $\overset{.....}{h}$, $\ddot{p}$, $\ddot{r}$ and $\overset{.....}{v}$ are required. If the sizes of these estimates are large compared to the variances given to the observations, a slow reaction to sharp manoeuvres occurs. If the sizes of these estimates are small compared to the variances given to the observations, a quick reaction to sharp manoeuvres occurs. In an operational environment neither of these situations is ideal, and a model that applies increased smoothing when the dynamics of the antennae are relatively static and a quick reaction time when the dynamics of the antennae change rapidly would be more suitable. Within this research, this methodology of "smart stochastic modelling" is achieved through the implementation of a standard least squares unit variance confidence test at each epoch.

$$\bar{x}_i = \begin{bmatrix} \bar{E} & \bar{N} & \bar{U} & \bar{h} & \bar{p} & \bar{r} & \bar{v} & \bar{\dot{h}} & \bar{\dot{p}} & \bar{\dot{r}} & \bar{\dot{v}} & \bar{\ddot{h}} & \bar{\ddot{p}} & \bar{\ddot{r}} & \bar{\ddot{v}} & \bar{\dddot{h}} & \bar{\dddot{p}} & \bar{\ddddot{h}} & \bar{\ddddot{v}} \end{bmatrix}^T \tag{3}$$

$$\begin{aligned}
{}_p h_i &= h_{i-1} + \dot{h}_{i-1}\Delta t + \frac{1}{2}\ddot{h}_{i-1}\Delta t^2 + \frac{1}{6}\dddot{h}_{i-1}\Delta t^3 + \frac{1}{24}\ddddot{h}_{i-1}\Delta t^4 \\
{}_p \dot{h}_i &= +\dot{h}_{i-1} + \ddot{h}_{i-1}\Delta t + \frac{1}{2}\dddot{h}_{i-1}\Delta t^2 + \frac{1}{6}\ddddot{h}_{i-1}\Delta t^3 \\
{}_p \ddot{h}_i &= \ddot{h}_{i-1} + \dddot{h}_{i-1}\Delta t + \ddddot{h}_{i-1}\frac{1}{2}\Delta t^2 \\
{}_p \dddot{h}_i &= \dddot{h}_{i-1} + \ddddot{h}_{i-1}\Delta t \\
{}_p \ddddot{h}_i &= \ddddot{h}_{i-1} \\
{}_p p_i &= p_{i-1} + \dot{p}_{i-1}\Delta t + \frac{1}{2}\ddot{p}_{i-1}\Delta t^2 \\
{}_p \dot{p}_i &= \dot{p}_{i-1} + \ddot{p}_{i-1}\Delta t \\
{}_p \ddot{p}_i &= \ddot{p}_{i-1} \\
{}_p r_i &= r_{i-1} + \dot{r}_{i-1}\Delta t + \frac{1}{2}\ddot{r}_{i-1}\Delta t^2 \\
{}_p \dot{r}_i &= \dot{r}_{i-1} + \ddot{r}_{i-1}\Delta t \\
{}_p \ddot{r}_i &= \ddot{r}_{i-1} \\
{}_p v_i &= v_{i-1} + \dot{v}_{i-1}\Delta t + \frac{1}{2}\ddot{v}_{i-1}\Delta t^2 + \frac{1}{6}\dddot{v}_{i-1}\Delta t^3 + \frac{1}{24}\ddddot{v}_{i-1}\Delta t^4 \\
{}_p \dot{v}_i &= \dot{v}_{i-1} + \ddot{v}_{i-1}\Delta t + \frac{1}{2}\dddot{v}_{i-1}\Delta t^2 + \frac{1}{6}\ddddot{v}_{i-1}\Delta t^3 \\
{}_p \ddot{v}_i &= \ddot{v}_{i-1} + \dddot{v}_{i-1}\Delta t + \frac{1}{2}\ddddot{v}_{i-1}\Delta t^2 \\
{}_p \dddot{v}_i &= \dddot{v}_{i-1} + \ddddot{v}_{i-1}\Delta t \\
{}_p \ddddot{v}_i &= \ddddot{v}_{i-1} \\
{}_p E_i &= E_{i-1} + {}_p v_i \Delta t \cos({}_p p_i)\sin({}_p h_i) \\
{}_p N_i &= N_{i-1} + {}_p v_i \Delta t \cos({}_p p_i)\cos({}_p h_i) \\
{}_p U_i &= U_{i-1} + {}_p v_i \Delta t \sin({}_p p_i)
\end{aligned} \tag{4}$$

where:

- E Easting coordinate (m)
- N Northing coordinate (m)
- U Up coordinate (m)
- h Heading (deg)
- p Pitch (deg)
- r Roll (deg)
- v Velocity (m/s)
- $\dot{h}$ Change in h (deg/s)
- $\dot{p}$ Change in p (deg/s)
- $\dot{r}$ Change in r (deg/s)
- $\dot{v}$ Change in v (deg/s^2) (otherwise known as acceleration)
- $\ddot{h}$ Change in $\dot{h}$ (deg/s^2)
- $\ddot{p}$ Change in $\dot{p}$ (deg/s^2)
- $\ddot{r}$ Change in $\dot{r}$ (deg/s^2)
- $\ddot{v}$ Change in $\dot{v}$ (deg/s^3) (otherwise known as jerk)
- $\dddot{h}$ Change in $\ddot{h}$ (deg/s^3)
- $\dddot{v}$ Change in $\ddot{v}$ (deg/s^4)
- $\ddddot{h}$ Change in $\dddot{h}$ (deg/s^4)
- $\ddddot{v}$ Change in $\dddot{v}$ (deg/s^5)

Note that for these equations, a subscript ‘p’ denotes a predicted quantity, while a subscript ‘i’ denotes the current measurement epoch.

2.3.1 Unit variance confidence test

The unit variance $\left(\sigma_0^2\right)$ is a function of the least squares residuals ($\hat{v}$) (the difference between the measurements and the updated measurements after the Kalman filter process), the variance of the measurements (V_1), and the degrees of freedom of the adjustment (df) (Equation (6)).

$$\sigma_0{}^2 = \frac{\hat{v}^T V_1 \hat{v}}{df} \tag{6}$$

Given the correct selection of measurement variances and the absence of gross errors, the unit variance should not differ significantly from unity. In this research, the unit variance confidence test was applied at each epoch to distinguish actual movement of the vehicle platform from apparent movement caused by noise in the observations. In particular, the test seeks to determine when the dynamic model correctly predicts the motion of the vehicle. If the unit variance confidence test fails, this is used as an indication that the variances of the dynamic model predicted observations are overestimated. It can then be assumed that:

- the dynamic model is not successfully predicting the movement of the platform to the precision indicated by the predicted observations variances, and therefore;
- the Kalman filter has failed to respond to a movement of the platform.

Consequently, the weights in the Kalman filter are significantly reduced to allow the filter to catch up to the current platform attitude. On the other hand, if the unit variance confidence test does not differ significantly from unity:

- the dynamic model is successfully predicting the movement of the platform to the precision indicated by the predicted observation variances, and
- the Kalman filter is sufficiently reacting to vehicle dynamics.

Hence, the filter variances remain high (or are increased if they were previously lowered) to provide maximum filtering of the data and removal of noise.

$$\bar{x}_i = \begin{bmatrix} \bar{E} & \bar{N} & \bar{U} & \bar{h} & \bar{p} & \bar{r} & \bar{v} & \bar{\dot{h}} & \bar{\dot{p}} & \bar{\dot{r}} & \bar{\dot{v}} & \bar{\ddot{h}} & \bar{\ddot{p}} & \bar{\ddot{r}} & \bar{\ddot{v}} & \bar{\dddot{h}} & \bar{\dddot{p}} & \bar{\ddddot{h}} & \bar{\ddddot{v}} & \bar{\dot{\beta}} & \bar{\varepsilon} \end{bmatrix}^T$$

$$l_i = [P_i \quad e_i]^T$$

$$P_i = \begin{bmatrix} {}_p\bar{E} & {}_p\bar{N} & {}_p\bar{U} & {}_p\bar{h} & {}_p\bar{p} & {}_p\bar{r} & {}_p\bar{v} & {}_p\bar{\dot{h}} & {}_p\bar{\dot{p}} & {}_p\bar{\dot{r}} & {}_p\bar{\dot{v}} & {}_p\bar{\ddot{h}} & {}_p\bar{\ddot{p}} & {}_p\bar{\ddot{r}} & {}_p\bar{\ddot{v}} & {}_p\bar{\dddot{h}} & {}_p\bar{\dddot{p}} & {}_p\bar{\ddddot{h}} & {}_p\bar{\ddddot{v}} & {}_p\bar{\dot{\beta}} & {}_p\bar{\varepsilon} \end{bmatrix}^T$$

$$e_i = [E^A \quad N^A \quad U^A \quad E^B \quad N^B \quad U^B \quad E^C \quad N^C \quad U^C \quad \dot{h} \quad d]^T$$

$$\begin{aligned}
f_1(\bar{x},\bar{l}) &= \bar{E} \\
f_2(\bar{x},\bar{l}) &= \bar{N} \\
f_3(\bar{x},\bar{l}) &= \bar{U} \\
f_4(\bar{x},\bar{l}) &= \bar{E} + \left(\cos(\bar{r})\cos(\bar{h}) + \sin(\bar{r})\sin(\bar{p})\sin(\bar{h})\right)\Delta E_A^{AB} + \\
&\quad \left(\cos(\bar{p})\sin(\bar{h})\right)\Delta N_A^{AB} + \\
&\quad \left(\sin(\bar{r})\cos(\bar{h}) - \cos(\bar{r})\sin(\bar{p})\sin(\bar{h})\right)\Delta U_A^{AB} \\
f_5(\bar{x},\bar{l}) &= \bar{N} + \left(-\cos(\bar{r})\sin(\bar{h}) + \sin(\bar{r})\sin(\bar{p})\cos(\bar{h})\right)\Delta E_A^{AB} + \\
&\quad \left(\cos(\bar{p})\cos(\bar{h})\right)\Delta N_A^{AB} + \\
&\quad \left(-\sin(\bar{r})\sin(\bar{h}) - \cos(\bar{h})\cos(\bar{r})\sin(\bar{p})\right)\Delta U_A^{AB} \\
f_6(\bar{x},\bar{l}) &= \bar{U} + (-\sin(\bar{r})\cos(\bar{p}))\Delta E_A^{AB} + \\
&\quad \sin(\bar{p})\Delta N_A^{AB} + \\
&\quad \cos(\bar{r})\cos(\bar{p})\Delta U_A^{AB} \\
f_7(\bar{x},\bar{l}) &= \bar{E} + \left(\cos(\bar{r})\cos(\bar{h}) + \sin(\bar{r})\sin(p)\sin(\bar{h})\right)\Delta E_A^{AC} + \\
&\quad \left(\cos(\bar{p})\sin(\bar{h})\right)\Delta N_A^{AC} + \\
&\quad \left(\sin(\bar{r})\cos(\bar{h}) - \cos(\bar{r})\sin(\bar{p})\sin(\bar{h})\right)\Delta U_A^{AC} \\
f_8(\bar{x},\bar{l}) &= \bar{N} + \left(-\cos(\bar{r})\sin(\bar{h}) + \sin(\bar{r})\sin(\bar{p})\cos(\bar{h})\right)\Delta E_A^{AC} + \\
&\quad \left(\cos(\bar{p})\cos(\bar{h})\right)\Delta N_A^{AC} + \\
&\quad \left(-\sin(\bar{r})\sin(\bar{h}) - \cos(\bar{h})\cos(\bar{r})\sin(\bar{p})\right)\Delta U_A^{AC} \\
f_9(\bar{x},\bar{l}) &= \bar{U} + (-\sin(\bar{r})\cos(\bar{p}))\Delta E_A^{AC} + \\
&\quad \sin(\bar{p})\Delta N_A^{AC} + \\
&\quad \cos(\bar{r})\cos(\bar{p})\Delta U_A^{AC}
\end{aligned} \tag{5}$$

$$f_{10}(\bar{x},\bar{l}) = \bar{E} \quad f_{11}(\bar{x},\bar{l}) = \bar{N} \quad f_{12}(\bar{x},\bar{l}) = \bar{U} \quad f_{13}(\bar{x},\bar{l}) = \bar{h} \quad f_{14}(\bar{x},\bar{l}) = \bar{p} \quad f_{15}(\bar{x},\bar{l}) = \bar{r}$$

$$f_{16}(\bar{x},\bar{l}) = \bar{v} \quad f_{17}(\bar{x},\bar{l}) = \bar{\dot{h}} \quad f_{18}(\bar{x},\bar{l}) = \bar{\dot{p}} \quad f_{19}(\bar{x},\bar{l}) = \bar{\dot{r}} \quad f_{20}(\bar{x},\bar{l}) = \bar{\dot{v}} \quad f_{21}(\bar{x},\bar{l}) = \bar{\ddot{h}}$$

$$f_{22}(\bar{x},\bar{l}) = \bar{\ddot{p}} \quad f_{23}(\bar{x},\bar{l}) = \bar{\ddot{r}} \quad f_{24}(\bar{x},\bar{l}) = \bar{\ddot{v}} \quad f_{25}(\bar{x},\bar{l}) = \bar{\dddot{h}} \quad f_{26}(\bar{x},\bar{l}) = \bar{\dddot{v}} \quad f_{27}(\bar{x},\bar{l}) = \bar{\ddddot{h}}$$

$$f_{28}(\bar{x},\bar{l}) = \bar{\ddddot{v}}$$

$$f_{29}(\bar{x},\bar{l}) = \bar{\dot{\beta}} \quad f_{30}(\bar{x},\bar{l}) = \bar{\varepsilon} \quad f_{31}(\bar{x},\bar{l}) = \bar{\dot{h}}\Delta t + \bar{\dot{\beta}}\Delta t \quad f_{32}(\bar{x},\bar{l}) = \bar{\varepsilon}\bar{d}\Delta t$$

(3)

2.3.2 Intelligent navigation

Spatial information databases are now a standard component of many mobile navigation systems. This is directly due to their ability to provide detailed information about the location and inter-relationship of geographically defined features. Map matching techniques traditionally use this information in an attempt to improve navigation accuracy. This research proposes the integration of map matching techniques, (like other navigation instruments such as gyroscopes, odometers and GPS), within the Kalman filter, as a means of providing additional measurements that can be used to improve position and attitude determination. The map matching technique implemented in this research has been termed "Intelligent Navigation" (IN).

The IN algorithm developed in this research is modelled on the simple rules of navigation that humans use on a day-to-day basis, and in doing so incorporates both geometric and topological map matching techniques. This algorithm has several advantages; it consists of a simple, yet effective set of four rules (closest road, bearing matching, access only, distance in direction); it relies on the short term precision of the navigation sensors (in particular DR when GPS is unavailable); when implemented in the Kalman filter, IN has the unique advantage that it no longer assumes the navigator is on the road centreline (i.e. the road segments stored in the database), but instead is 'following' the road network. This is particularly important for high precision applications. No matter how accurate the database information, the navigator will not always travel directly on the network, but will travel to the left or right of the network to varying degrees (unless the navigator is fixed to the network as in the case of trains and trams).

The "closest road" rule of IN makes the assumption that the vehicle is travelling along a road (which is typically the case). This constraint can be included in the location solution, thus improving the accuracy of the computed position of the vehicle. This algorithm is most effective when the nearest road (according to the navigation instrument's computed position) is in fact the road being travelled. However, when approaching intersections or when two roads are close to each other, the nearest road may not be the road being travelled. In such cases, constraining the solution to fall on the nearest road actually downgrades the calculated position. To avoid such errors, the bearing matching rule is required. This rule requires that the nearest road to which the vehicle's position is corrected must have a bearing similar to the measured direction of travel. This corrects the problem previously described. The threshold of similarity between the vehicle's bearing and the bearing of the surrounding roads may be adjusted to suit the accuracy of the navigation instruments. However, the larger the threshold, the more likely it becomes that roads will be incorrectly matched as having the same bearing as that of the vehicle. The access only rule is designed to identify and prevent this error from occurring. Take, for example, a vehicle travelling along road A in the road layout diagram shown in Figure 4. Assuming the only route to road C is via road B, logic dictates that for the vehicle to be travelling along road C it must previously have travelled along road B. By logging previously travelled roads, the navigation system can prevent the vehicle from being located on a road that it could not possibly be on.

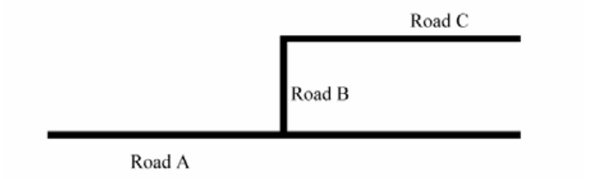


Figure 4 Road layout scenario

The distance in direction rule reduces the accumulation of distance error by calculating the distance travelled by the vehicle in the direction of the road rather than the direction measured by the navigation device. This is particularly important when navigation instruments of low accuracy are employed. For example, if a vehicle travels 1000 metres along a road of bearing 60 degrees while measuring the road to have a bearing of 65 degrees (i.e. 5 degrees in error), an error in distance of 4 metres will occur. Although this may seem insignificant, over several kilometres, or with lower accuracy navigation instruments, larger errors can accumulate. This error is avoided by calculating the distance travelled independently from the bearing of the vehicle and then applying this distance in the direction of the road being travelled.

2.3.3 Deriving the intelligent navigation observation equations

Incorporating IN into the Kalman filter requires the development of observation equations from the IN rules. This procedure also allows for additional parameters to be estimated by the filter, such as offset from the centreline. Furthermore, precisions can be associated with the information obtained from the map data to allow the Kalman filter to optimally estimate the position and attitude of the vehicle from all available measurements, rather than 'correcting' the vehicle's position to a point on the centreline.

The IN observation equations are derived from:

- the IN estimate of the vehicle's 'corrected' position (which lies on a road segment), and

- an estimate of the vehicle's heading (i.e. the heading of the road segment at the IN 'corrected' position).

An additional parameter is also added to the Kalman filter that estimates the shortest distance to the road segment that the vehicle is following (referred to as the Euclidean distance). By including this parameter, the IN rules no longer assume that the vehicle is directly on the road centreline, but instead follows a path parallel to the road centreline. The variance of this parameter dictates how closely the vehicle must follow the road centreline. The updated parameters for the state of the Kalman filter, including observations equations, are shown in Equations (7) and (8).

$$
\begin{aligned}
\bar{x}_i &= \begin{bmatrix} \bar{E} & \bar{N} & \bar{U} & \bar{h} & \bar{p} & \bar{r} & \bar{v} & \bar{\dot{h}} & \bar{\dot{p}} & \bar{\dot{r}} & \bar{\dot{v}} & \bar{\ddot{h}} & \bar{\ddot{p}} & \bar{\ddot{r}} & \bar{\ddot{v}} & \bar{\dddot{h}} & \bar{\dddot{p}} & \bar{\ddddot{h}} & \bar{\ddddot{v}} & \bar{\dot{\beta}} & \bar{\varepsilon} & \bar{O} \end{bmatrix}^T \\
l_i &= [P_i \quad e_i]^T \\
P_i &= \begin{bmatrix} {}_PE & {}_PN & {}_PU & {}_Ph & {}_Pp & {}_Pr & {}_Pv & {}_P\dot{h} & {}_P\dot{p} & {}_P\dot{r} & {}_P\dot{v} & {}_P\ddot{h} & {}_P\ddot{p} & {}_P\ddot{r} & {}_P\ddot{v} & {}_P\dddot{h} & {}_P\dddot{p} & {}_P\ddddot{h} & {}_P\ddddot{v} & {}_P\dot{\beta} & {}_P\varepsilon & {}_PO \end{bmatrix}^T \\
e_i &= \begin{bmatrix} E^A & N^A & U^A & E^B & N^B & U^B & E^C & N^C & U^C & \dot{h} & d & E^{IN} & N^{IN} & h^{IN} \end{bmatrix}^T
\end{aligned}
\tag{7}
$$

where:

O	Euclidean distance (m) from the road centreline (a prediction in P_i and a parameter in the Kalman filter $\bar{x}_i$)
E^{IN}	Easting coordinate (m) as measured from the road database
N^{IN}	Northing coordinate (m) as measured from the road database
h^{IN}	Heading (deg) as measured from the road database

The resultant observation equations are as follows:

$$
\begin{aligned}
f_{32}(\bar{x}, \bar{l}) &= \bar{O} \\
f_{33}(\bar{x}, \bar{l}) &= \bar{E} + \bar{O}\sin(\bar{h} + 90) \\
f_{34}(\bar{x}, \bar{l}) &= \bar{N} + \bar{O}\cos(\bar{h} + 90) \\
f_{35}(\bar{x}, \bar{l}) &= \bar{h}
\end{aligned}
\tag{8}
$$

where:

f_{32}	Observation equation for the Euclidean distance from the road centreline as predicted by the Kalman filter
f_{33}	Observation equation for the easting coordinate as measured from the road database
f_{34}	Observation equation for the northing coordinate as measured from the road database
f_{35}	Observation equation for the heading as measured from the road database

2.3.4 Additional intelligent navigation parameters

There are three important additional parameters that can affect the operation of IN. These are:

- search radius,
- inner intersection exclusion radius, and
- angular similarity.

The search radius is the radial distance that the IN algorithm uses to search the surroundings for geospatial features. For example, given a search radius of 50 metres, at each estimated vehicle location, IN will search the surrounding area up to a radius of 50 metres for features such as roads and road intersections. The larger this distance, the larger the error that IN is able to correct for. If the nearest road is up to 50 metres away, IN (using a 50 metre search radius) is able to locate the road and update its estimated position to the road (if it meets the suitable criteria).

When approaching intersections, IN is switched off and the short term precision of the navigation instruments (assuming DR navigation) is relied upon to avoid ambiguous situations. The proximity of intersections before IN is switched off is determined by the second parameter, inner intersection exclusion radius. The larger the radius, the more frequently intersections will be detected and IN corrections will therefore not occur.

The third parameter is the angular similarity. This refers to a tolerance within which the actual road bearing and calculated bearing (by the external navigation devices) are considered to be the same. As before, a larger value allows for larger errors to be corrected, but also increases the uncertainty in areas where roads intersect at oblique or acute angles.

In addition to these IN parameters, supplementary features have been added with the aim of further improving the accuracy of the navigation system. An additional parameter (the outer intersection exclusion radius) is used to constrain the operation of the IN algorithm, while a cornering algorithm allows for the use of IN on corners. IN could not previously implement the cornering algorithm, as prior to the inclusion of IN in the Kalman filter, an optimal estimate for the Euclidean distance to the road centreline was not available. This Euclidean distance is required to compute the trajectory of the vehicle through the corner.

The outer intersection exclusion radius parameter limits IN operation to the specified ranges of the inner intersection exclusion and outer intersection exclusion radii. For example, an inner intersection exclusion radius of 20m and outer intersection exclusion radius of 100m would limit IN operation to areas that are less than 100m but no closer than 20m to an intersection. This parameter is implemented in such a way that it also allows partial limitation of IN, for example, when the 'position exception' option is invoked, the position observations from the IN are included in the Kalman filter regardless of the outer intersection exclusion radius value. Similarly, using the 'heading exception' option allows heading observations to be included in the Kalman filter regardless of the outer intersection exclusion radius value.

All previous implementations of IN have not operated within the proximity of intersections (according to the inner intersection exclusion radius). As the vehicle travels through the intersection, IN does not provide any observations to the Kalman filter. The circular cornering algorithm overcomes this gap of observations by attempting to predict the vehicle's trajectory based on the available road centreline information. A circular curve is used as the prediction trajectory. Figure illustrates this procedure. As the vehicle nears an intersection, the rate of turn of the vehicle is monitored. If the turn rate increases above a specified threshold, the intersection is examined to determine which road segment the vehicle is turning into. A circular curve is fitted to the vehicle's current position and predicted end of turn location. This curve is then used as the road centreline required by IN to provide observations. If the vehicle turns away from or continues to turn past the road the vehicle was predicted to follow, a new road prediction is made (again based on the vehicle's current position, heading, turn rate and data available in the road centreline database).

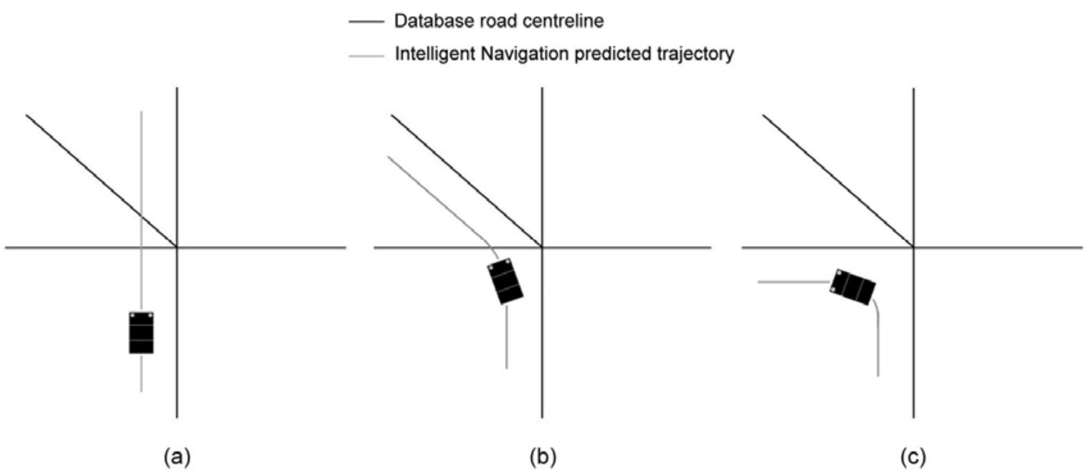


Figure 5 Circular cornering algorithm: (a) original trajectory prediction, (b) first updated trajectory prediction, and (c) second updated trajectory prediction

The only additional parameter required for the operation of the cornering algorithm is the turn rate threshold. This value indicates when the vehicle's turn rate is of a large enough magnitude to suggest that the vehicle is turning. All other information to predict the vehicle's trajectory is available either from the previous epoch of the Kalman filter or from the road centreline database. The method for calculating the IN position $\left(E^{IN}, N^{IN}\right)$ and heading $\left(h^{IN}\right)$ observables is shown in Figure and Equations 9. In order to compute the trajectory of the vehicle from one road to the next, an assumption is made that the Euclidean distance from the road centreline prior to the turn is the same once the turn has been completed.

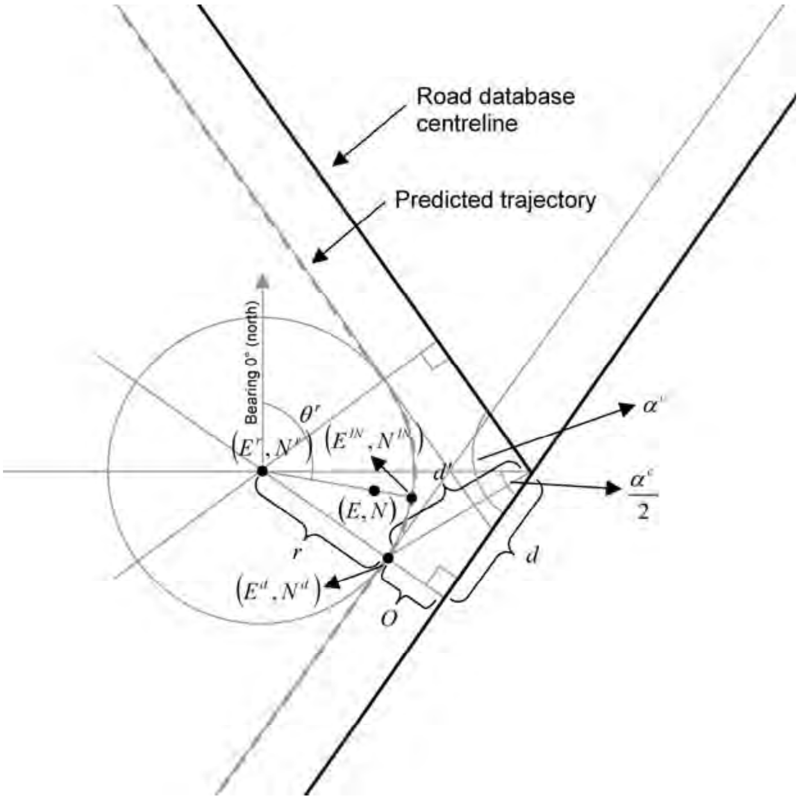


Figure6 Computing the turning radius and centre of rotation

$$
\begin{aligned}
d &= \sqrt{d'^2 - O^2} \\
r &= d \tan\left(\frac{\alpha^c}{2}\right) - O \\
\theta^r &= \arctan\left(\frac{E - E^r}{N - N^r}\right) \\
E^{IN} &= E^r + r\sin\left(\theta^r\right) \\
N^{IN} &= E^r + r\cos\left(\theta^r\right) \\
h^{IN} &= \left(\theta^r - 90°\right)
\end{aligned}
\tag{9}
$$

where

O	Euclidean distance between Kalman filtered estimated position and IN estimated centreline position (provided by the Kalman filter)
$\left(E^d, N^d\right)$	Point of deviation from the current road centreline (as determined by the turn rate threshold)
d'	Distance from the road centreline intersection to the point of deviation (computed from the road centreline database)

(E, N) Vehicle position as computed by the navigation instruments

The additional parameters and algorithm detailed in this section significantly enhance the complexity of the resultant Kalman filter. The different modes that are now available within the IN module alone are summarised in **Error! Reference source not found.**1.

Table 1 Summary of Intelligent Navigation modes

Mode	Description
Continuous	Provides position and heading estimations as inputs for the Kalman filter. Searches the database for road centrelines within the specified search radius, and with bearing matching the angular similarity constraint. Does not operate in areas where road intersections are within the intersection exclusion radius, but operates at all other times where road data is available. This mode includes the centreline offset that estimates the distance of the vehicle from the road centreline. Without this offset, an incorrect assumption would be made that the vehicle is travelling along the road centreline.
Outer intersection exclusion radius	Limits Intelligent Navigation operation to the specified ranges of the intersection inclusion and exclusion radii. For example, an inner intersection exclusion radius of 20m and outer intersection exclusion radius of 100m would limit Intelligent Navigation operation to areas that are less than 100m but no closer than 20m to an intersection.
Position exception	Allows input of Intelligent Navigation position estimation, overriding the outer intersection exclusion mode.
Heading exception	Allows input of Intelligent Navigation heading estimation, overriding the outer intersection exclusion mode.
Circular cornering algorithm	Estimates the turning path of the vehicle at road intersections and provides position and heading estimations throughout the curve. Operates only within the intersection exclusion radius.

The processes for including IN information in the Kalman filter are shown in Figure 7. Using data from the GPS and DR instruments, the position and attitude of the vehicle are estimated. This information provides input for the IN algorithms. The results from IN are then combined with the GPS/DR measurements and filtered to provide an optimal solution using all available information. Note that there is only one Kalman filter, although it must be run twice. The first run provides the input for the IN algorithms. The second run computes the optimal state of the mobile platform using all available measurements (GPS, DR and IN).

The estimates of precisions used for the observations and the dynamic model are shown in Table 2.

Table 2 Precision estimates for the Kalman filter parameters

Dynamic model	**Standard deviation**
$\overset{\cdot\cdot\cdot\cdot\cdot}{h}$	0.001 deg/s^5
$\dddot{p}$	0.01 deg/s^3
$\dddot{r}$	0.01 deg/s^3
$\overset{\cdot\cdot\cdot\cdot\cdot}{v}$	0.001 m/s^6
Instrument measurement	
Horizontal RTK GPS	0.01m
Vertical RTK GPS	0.02m
Gyro rate	0.008deg/s
Odometer	2% of distance measured
Intelligent Navigation	
E^{IN} N^{IN}	0.1 metres
h^{IN}	1 degree

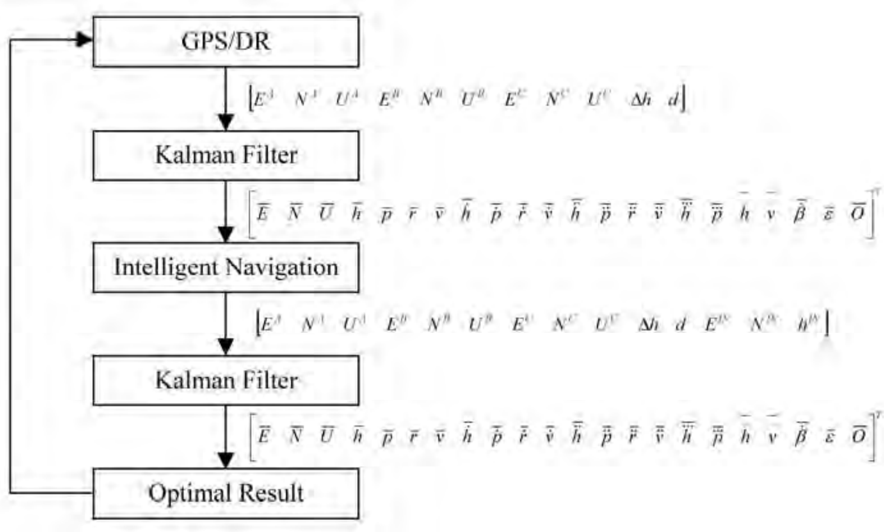


Figure 7 Kalman filter process with Intelligent Navigation

3 Testing and evaluation of Kalman filtering algorithms

To evaluate the integration algorithms developed in this research an AR prototype, ***i*ARM** (**I**ntelligent **A**ugmented **R**eality **M**apper) was constructed. The ***i*ARM** consists of the integrated positioning and attitude determination system described previously, combined with a digital video camera and a database containing three dimensional objects used for augmentation.

The ***i*ARM** was installed on a typical land mobile vehicle and a 1 kilometre road circuit located within the Melbourne General Cemetery was used as the test bed for this research. With an extensive network of roads, the cemetery offered many challenges to the integrated position and attitude determination system, as it contained GPS obstructions, as well as sharp corners and curving roads that continuously change the dynamics of the vehicle travelling the circuit. The location of road boundaries within the cemetery were predetermined and used to generate the objects used in the augmentation process.

The test vehicle was driven around the test circuit a number of times, with the AR prototype operating in real-time at 10 frames per second with VGA (640 × 480 pixels) resolution, and all navigation instruments operating. Figure 8 shows the AR prototype in operation with all sensors operating and the augmentation of the road boundaries.

Figure 8 Augmented image from the prototype

From Figure 8 it is clearly visible that when all observations are available the integrated position and attitude determination system is successful in accurately aligning the augmented objects (the road boundaries) with the real world images of the driver as captured by the digital video camera. The results presented in Figure 8 are typical of the visual registration accuracy of augmented data and the real world images captured in this research.

In order to further explore and quantify the effects of errors in the position and attitude determination system, a single epoch of data from the vehicle travelling around the cemetery test circuit was selected. A 2 degree error was added to the heading as computed at that epoch. The augmented data was then rendered (using the heading with the error). The result is shown in Figure 9.

Figure 9 Misalignment of the augmented road boundaries caused by a 2 degree error in heading

As theoretically determined, the misalignment caused by an error in heading is more clearly obvious in the distance. However, the effect of foreshortening causes both the object being augmented as well as the augmented model to become smaller in the distance. Hence, misalignment of the most distant road boundary and augmented model cannot be visibly detected. In comparison, an error in position of 2 metres has the results shown in Figures 10 and 11.

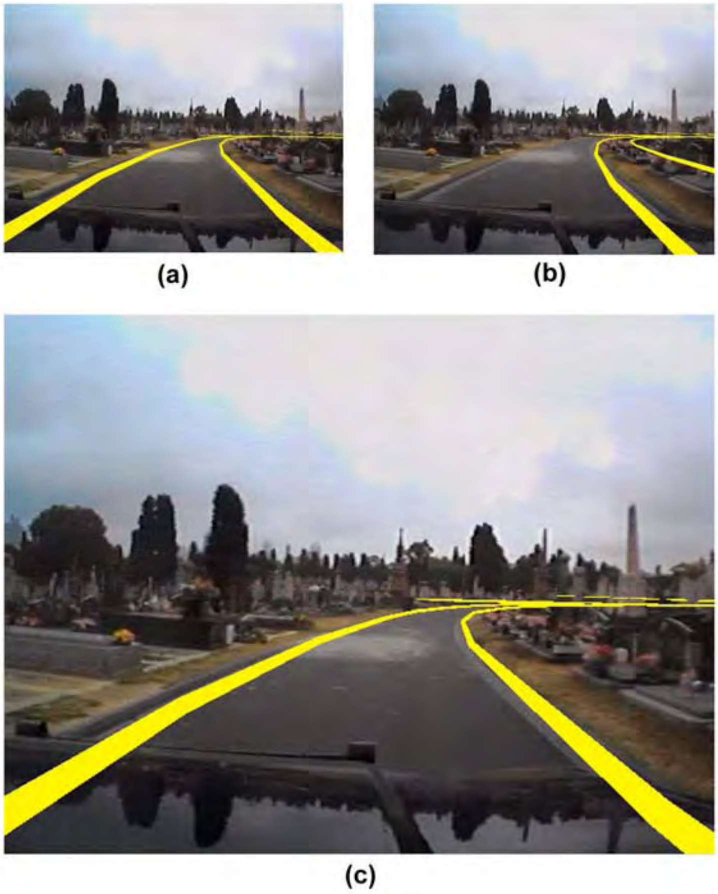
Figure 10. Misalignment of the augmented road boundaries caused by a 2 metre position error that is perpendicular to the direction of travel

Figure 11. Misalignment of the augmented road boundaries caused by a 2 metre position error that occurs in the direction of travel

Despite the position errors shown in Figures 8 and 9 having exactly the same magnitude, the direction of the error has a profound effect. Where the error occurred in the direction of travel, very little visual effect could be identified. On the other hand, when the direction of error was perpendicular to the direction of travel, the misalignment was clearly visible.

In order to evaluate the performance of the integrated position and attitude determination system during periods of GPS outages, a simulation test is conducted using the RTK GPS observations as a measure of the 'true' trajectory of the mobile platform. A portion of the data collected while travelling the test circuit was selected where RTK GPS observations were available. The GPS observations were then removed from the data, hence simulating a GPS outage. The simulated outage spanned 60 seconds. The Kalman filter was used to process the available DR measurements both with and without Intelligent Navigation. The alignment of augmented road boundaries exhibited by the AR prototype using only RTK GPS (i.e. the truth alignment), using only DR measurements (no IN), and using DR with IN are displayed in Figure 12.

Although Figure 12(b) shows a clear misalignment between the augmented and real road boundaries, the addition of the IN "observations" significantly reduces this misalignment as shown in Figure 12(c). In fact, as illustrated in Figure 13, the position error of the vehicle (calculated using the RTK GPS observations as truth) was reduced from 4.35 metres to 1.54 metres, a 65% reduction. It is important to note that despite a remaining position error of 1.54 metres after a 60 second outage, the alignment of the real and augmented boundaries appear visually correct as shown in Figure 12(c). The nature of the IN rules shift much of the remaining error in position into the direction of travel, hence minimising the visual misalignment caused by that error. This phenomenon is ultimately due to the linear nature of the road centrelines. This feature allows IN to readily detect and correct for errors perpendicular to the centreline, and without a change in heading (i.e. turning a corner), error in the direction of travel cannot be detected.



(a) (b) (c)

Figure 12 Augmentation of road boundaries using: (a) RTK GPS only, (b) integrated gyroscope and odometer after a 60 second GPS outage, and (c) integrated gyroscope, odometer and IN after a 60 second GPS outage

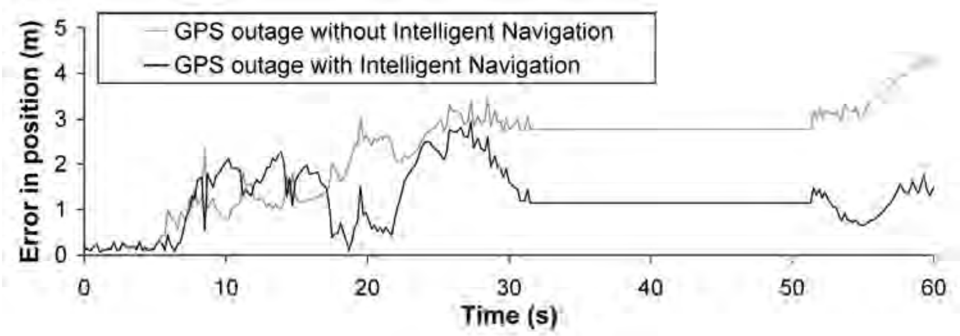


Figure 13 Position error over the 60 second GPS outage with and without IN

4 Conclusions

The aim of this research was to investigate the performance of an integrated position and attitude determination system to support AR applications in outdoor unprepared environments. To achieve this aim, multiple sensors and data sources were integrated within a Kalman filter. The position and attitude determination system was evaluated within a land mobile AR prototype developed within this research. Following are some of the conclusions drawn from this research.

- The Kalman filter developed was able to provide updates of position and attitude to enable accurate and continuous registration of the augmented objects with real world perspective views.
- The development of the unit variance test allowed for tuning of the Kalman filter to rapidly adapt to changes in the mobile platform dynamics. Hence the resulting filter provided improved results during times of low dynamics, but swift reaction times during high dynamics.
- The integration of the IN rules into the Kalman filter provided a significant improvement to position and attitude determination during GPS outages.
- The effect of errors in the position and attitude determination component on the alignment of augmented information with real world views was minimised when the error was contained primarily to the direction of travel. This occurrence is particularly relevant when considering the operation of IN. IN was found not only to reduce the magnitude of error, but to also constrain remaining error in the position solution to the direction of travel.

References

Azuma R., Baillot Y., Reinhold B., Feiner S., Julier S., MacIntyre B. (2001) ***Recent Advances in Augmented Reality***, IEEE Computer Graphics and Applications, Vol. 21, No. 6, pp. 34-47.

Cross P.A. (1990) ***Advanced least squares applied to position-fixing***, Working Paper No. 6, Second Edition, North East London Polytechnic, Department of Land Surveying, London, United Kingdom.

Devernay F., Mourgues G., Coste-Maniere E. (2001) ***Towards endoscopic augmented reality for robotically assisted minimally invasive cardiac surgery***, Paper presented at the International Workshop on Medical Imaging and Augmented Reality, Shatin, Hong Kong, China, 10-12 July, 2001, 16-20.

Klinker G., Creighton O., Dutoit A., Kobylinski R., Vilsmeier C., Brugge B. (2001) ***Augmented maintenance of powerplants: A prototyping case study of a mobile AR system***, Paper presented at the IEEE and ACM International Symposium on Augmented Reality, New York, NY, USA, 124-133.

Logan S.A. (2000) ***Integration of GPS phase and other measurements for kinematic mapping***, Doctor of Philosophy Thesis, The University of Melbourne, Melbourne, Australia.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 27-35

Utilizing Kriging to Generate a NLOS Error Correction Map for Network Based Mobile Positioning

Binghao Li, Chris Rizos

School of Surveying and Spatial Information System, UNSW, Australia
e-mail: binghao_li@student.unsw.edu.au Tel: 61-2-93854189; Fax: 61-2-93137493

Hyung Keun Lee

School of Avionics and Telecommunication, Hankuk Aviation University, Korea

Received: 27 November 2004 / Accepted: 12 July 2005

Abstract. Network mobilephone-based positioning experiences degradation of location accuracy due to localised non-line-of-sight (NLOS) signal propagation. This is well known to be a major source of error in network-based mobilephone positioning. NLOS error systematically causes the Mobile Station (MS) to appear further away from the base station than it actually is, thereby increasing the positioning error. One method to mitigate the effect of NLOS error is to generate a NLOS error correction map, and then use the correction map to rectify the distorted MS location. The correction map can be generated using the following procedure: (1) estimating the NLOS errors at points where the real positions can be obtained utilising other information such as the points very near BTS (Base Transceiver Station) and the intersections of streets, or the location where the measurement has been made; and (2) interpolating or extrapolating the errors to specific points that we are interested in. Assuming some reference points have been obtained, this paper utilises kriging, an estimation technique that is widely used in mining, to generate the correction map. Theoretically kriging can also be used wherever a continuous measure is made on a sample at a particular location in space or time. Using simulations with a typical dense urban environment assumption, the feature of the NLOS error variogram is analysed and different models of the variogram are compared. The correction map of NLOS error is generated using some 'sampled' points, and compared with the 'true' NLOS error map to show the efficiency of kriging.

Key words: NLOS, Positioning, Kriging

1 Introduction

Over the past decade, Mobile phone positioning techniques (PT) have received considerable attention (Rappaport TS, 1996) and great number of new value-added services have been proposed or developed, such as the location based information service, navigation assistant, resource management, gaming, emergency service etc (D'Roza and Bilchev, 2003; Wilde G, 2002). The requirements set forth by the US FCC E911 is one of the major forces to push the mobile phone PT moving forwards. It requires wireless carriers to provide precise location information within 50 to 100 meters in most cases (Http://www.fcc.gov/911/enhanced/).

The most popular and simple method is Cell ID. But the accuracy of Cell ID is dependent on the density of the BTS, and is relatively poor especially in rural areas (Dru and Saada, 2001). Another basic method is based on the received signal strength. However, since the signal propagation suffers rapid deep fading and long term fading (Lee, 1991), no model can describe the feature of signal propagation very well in many environments. The accuracy of this method is better than Cell ID (Yamamoto *et al*., 2001). Much attention has been focused on the signal time delay and angle of arrival measurements. No matter what kind of approach such as angle of arrival (AOA) (Sakagami *et al*., 1992), time of arrival (TOA) (Hashemi, 1991) or time difference of arrival (TDOA) (Drane *et al*., 1998) is utilized, line-of-sight (LOS) propagation is necessary for accurate location estimates. In other words, non-line-of-sight (NLOS) error is the dominant error in location estimation (Jr. JC and Stüber GL, 1998). NLOS errors are always positive, and range from a small number to thousand meters (Silventoinen and Rantalainen, 1996), depending on the propagation environment.

To protect location estimates from NLOS error corruption, many approaches have been investigated. In (Morley and Grover, 1995), an algorithm based on probability density function (pdf) model is utilized to reduce the NLOS error. However, it is very difficult to formulate the pdf, and this pdf should vary greatly with changing of the environment. A widely used idea to mitigate the NLOS error is NLOS error identification and reconstruction. The method in (Wylie and Holtzman, 1996) and (Woo SS *et al*., 2000) reconstruct LOS TOA measurements from a time history of LOS and NLOS TOA measurements, and assumes knowledge of the NLOS standard deviation for identifying NLOS BTS. While in (Cong and Zhuang, 2001), the NLOS BTS detection is based on TDOA residuals. The algorithm in (Wang, 2003) shows a constrained optimization method. Unfortunately, none of these methods can solve the NLOS problem well, since too many elements affect the signal propagation, and the propagation environment varies from place to place. In fact the problem could be solved using a basic data base method. In this method, the NLOS error can be directly extracted from the reference measurements at the reference points. Jayaraman et al. (2000) and Gunnarsdottir and Hole (2001) describe methods for collecting data to create the database. But the data collection and data base maintenance is quite a costly process. In order to make this work easier, a wireless signal map matching method (WSMM) is proposed (Lee and Rizos, 2003). Generally, more reference points can help to generate a database that can describe the real situation more precisely. But no matter how many reference points have been measured, the area of interest cannot be completely covered. Interpolation and extrapolation is absolutely necessary.

Now the question arises: how to efficiently estimate the NLOS error at the location (say x_0), which is not reference point, only knowing the limited reference points (say x_1, …, x_n). In this paper, we are interested in both estimation of the NLOS error and the confidence on the estimate. After a brief introduction of Geostatistics and kriging, the simulated NLOS error in a Manhattan-like urban environment is generated, and the features of NLOS error is checked using a simulation; this is followed by the different variogram models fitting. Finally, the results utilizing universal kriging (UK) are shown.

2 Geostatistics and Kriging

Geostatistics was first used by the mining industry, as high costs of drillings made the analysis of the data extremely important. The prediction of the ore grade in a mining block from observed samples at irregularly spaced locations is one of the most important problems. The basic tool in geostatistics, the variogram, is used to quantify spatial correlations between observations. The estimation procedure is called kriging after D. Krige, who and his colleagues started to apply statistical techniques to ore reserve estimation in 1950s. As there many advantages of kriging, and especially with the advent of powerful computers, application of kriging can be found in very different disciplines ranging from the classical fields mining and geology to soil science, hydrology, meteorology, environmental sciences, agriculture etc. Theoretically, kriging can also be used wherever a continuous measure is made on a sample at a particular location in space or time (Cressie, 1991; Armstrong , 1998). In spatial information system, it is necessary to use the limited data to describe the real nature as precisely as possible, kriging is a good candidate to choose.

In geostatistics, the geological phenomenon is described in terms of fluctuations around a fixed surface ("drift" or "trend"). The fluctuations are not error but rather fully-fledged features of the phenomenon, with a structure of their own. The observed value at each data point x is considered as the outcome, $z(x)$, of a random variable, $Z(x)$. Its mean is the drift, $m(x)$. A classical assumption in geostatistics is the second order stationarity, but in practice, a slightly weaker assumption is more widely used, that is the intrinsic hypotheses. It consists of two conditions:

- The expected value of the random variable $Z(x)$ is constant all over the domain D.
- The variance of the increment corresponding to two different locations depends only on the vector separating them

This condition can be formulated as:

$$E[Z(x)] = \mu \tag{1}$$

for all $x \in D$

$$\frac{1}{2} Var[Z(x+h) - Z(x)] = \frac{1}{2} E\left[(Z(x+h) - Z(x))^2\right] = \gamma(h) \tag{2}$$

where $\gamma(h)$, called variogram, depends only on the vector h and not on the locations x and $x+h$.

kriging provides a solution to the problem of estimation based on the knowledge of the variogram and the above assumption. Here is the simple case that the mean is constant across the entire region of study. Unfortunately, in reality it is common that the mean is not constant. The simulation shows that a drift of NLOS error exists. Assume the mean is a function of the site coordinates:

$$Z(x) = f_0(x)\beta_0 + f_1(x)\beta_1 + \ldots + f_p(x)\beta_p + \delta(x)$$

where $\beta_0, \ldots, \beta_p$ are unknown parameters; $\delta(x)$ is intrinsic and $E[\delta(x)]=0$.

In matrix notation, the above expression can be written as:

$$Z = X\beta + Y \tag{3}$$

In order to deal with the drift, UK (universal kriging) is proposed (The ordinary kriging can be treated as a subset of UK when $f_0(x)=1$, $\beta_1= \ldots = \beta_p=0$). To predict $Z(x_0)$ the UK predictor is a linear combination of values of the sample $Z(x_i)$.

$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i)$$

where λ_i is the weighting factor.

For the purpose of making this predictor to be unbiased for all possible vectors β, the following conditions need to be satisfied.

$$E\left[\sum_{i=1}^{n} \lambda_i Z(x_i) - Z(x_0)\right] = 0 \tag{4}$$

As the estimation variance is:

$$\begin{aligned}\sigma_K{}^2(x_0) &= Var\left[Z(x_0) - \hat{Z}(x_0)\right] = \\ &-\sum_{j=1}^{n}\sum_{i=1}^{n} \lambda_j \lambda_i \gamma\left(x_i - x_j\right) + 2\sum_{i=1}^{n} \lambda_i \gamma\left(x_i - x_0\right)\end{aligned} \tag{5}$$

the best unbiased linear estimator is the one which minimizes $\sigma^2_k(x_0)$ under the constraint on the sum of the coefficients in (4). Introducing the Lagrange multipliers this leads to a straightforward linear equation.

$$\begin{bmatrix} \Gamma & X \\ X' & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ m \end{bmatrix} = \begin{bmatrix} \gamma \\ f \end{bmatrix} \tag{6}$$

where

$$\Gamma = \begin{bmatrix} \gamma(x_1 - x_1) & \cdots & \gamma(x_1 - x_n) \\ \vdots & \gamma(x_i - x_j) & \vdots \\ \gamma(x_n - x_1) & \cdots & \gamma(x_n - x_n) \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & f_1(x_1) & \cdots & f_p(x_1) \\ \vdots & \vdots & f_i(x_j) & \vdots \\ 1 & f_1(x_n) & \cdots & f_p(x_n) \end{bmatrix}$$

$$\gamma = \left[\gamma(x_0 - x_1) \quad \cdots \quad \gamma(x_0 - x_n)\right]'$$

$$f = \left[1 \quad f_1(x_0) \quad \cdots \quad f_n(x_0)\right]'$$

So, the result is

$$\lambda' = \left[\gamma + X\left(X'\Gamma^{-1}X\right)^{-1}\left(f - X'\Gamma^{-1}\gamma\right)\right]' \Gamma^{-1}$$

$$m' = -\left(f - X'\Gamma^{-1}\gamma\right)'\left(X'\Gamma^{-1}X\right)^{-1}$$

and

$$\begin{aligned}&\sigma_k{}^2(x_0) = \\ &\gamma'\Gamma^{-1}\gamma - \left(f - X'\Gamma^{-1}\gamma\right)'\left(X'\Gamma^{-1}\gamma\right)^{-1}\left(f - X'\Gamma^{-1}\gamma\right)\end{aligned} \tag{7}$$

kriging is the best linear unbiased estimation (BLUE) that has the following features: (a) this estimator is a linear function of the data with weights calculated according to the specifications of unbiasedness and minimum variance. (b) The weights are determined by solving a system of linear equations with coefficients that depend only on the variogram that describes the structure of a family of functions. A major advantage of kriging is that it is more flexible than other interpolation methods. The weights are not selected on the basis of some arbitrary rule that may be applicable in some cases but not in others, but depend on how the function varies in space. Another advantage of kriging is that it provides the means to evaluate the magnitude of the estimation error. The mean square error is a useful rational measure of the reliability of the estimate; it depends only on the variogram and the location of the measurements.

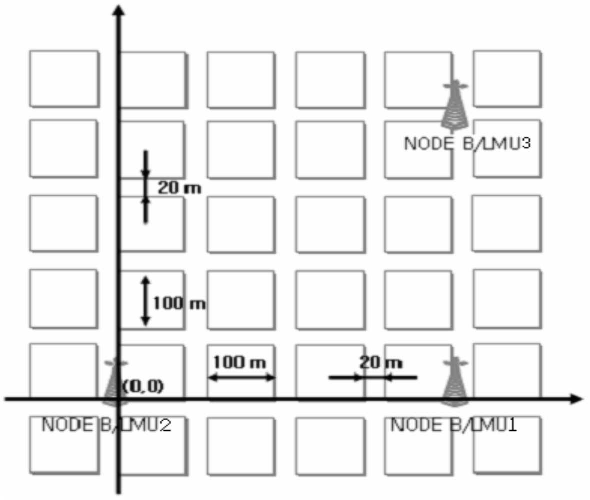


Fig. 1 Manhattan-like urban environment

3 NLOS error simulation

3.1 NLOS error

As a random function, NLOS error can be modeled by a deterministic part and a random part. Basically, there are three types of methods to generate the NLOS error. In most of the papers, simplified models such as deterministic, Gaussian or other distribution models (Cong and Zhuang, 2001) are utilized. Though this method is convenient, nevertheless, it can hardly describe the real NLOS error. On the contrast, 3D ray tracing plus Poisson or Rician model (Aguado *et al.*, 1997) can accurately generate the NLOS error in a special environment, but it is a very complex method. It is time consuming and also costly. The chosen method in this paper is the medium accuracy model; deterministic part is generated by Dijkstra algorithm (2D only), and the random part is represented by Gaussian model. Assuming in a Manhattan-like urban environment, only 3 BTSs are arranged as shown in Figure 1. By adding NLOS error and random measurement error, the TOA measurements are generated. Multiplying TOA with the speed of light, the real propagation distance can be obtained. Compare the propagation distance and the distance between the MS and BTS, the NLOS error (plus random measurement error) can be derived. Figure 2 shows the NLOS error in Manhattan-like environment. Streets 1 to 5 are the streets of south-north direction from east to west respectively (because of the symmetric environment, the NLOS error in the streets of west-east direction are similar). Two features can be noticed: first, there is a shift of NLOS error; second, the NLOS error is continuous. It is well known that using TOA measurements to compute the MS location, the MS's clock bias should be considered. However, the TDOA measurements can get rid of the clock bias automatically. In this paper, the object will be analyzed is the injected NLOS error rather than the NLOS error directly. The TDOA measurements can be generated by simply subtracting TOA measurement of one BTS from TOA measurement of reference BTS (here BTS1 is the reference). Two TDOA measurements representing TDOA21 and TDOA31 for each point are available. Figure 3 depicts the injected NLOS error on ideal domain and distorted domain. In the ideal domain, the injected NLOS error is more regular. The figure implies that there is a strong intrinsic relationship between the NLOS error and the location. Unfortunately, what should be dealt with is the NLOS error in the distorted domain. Since the determination of the MS' location is contaminated by the errors, the relationship between the NLOS error and the location is weaker and this makes the approach difficult.

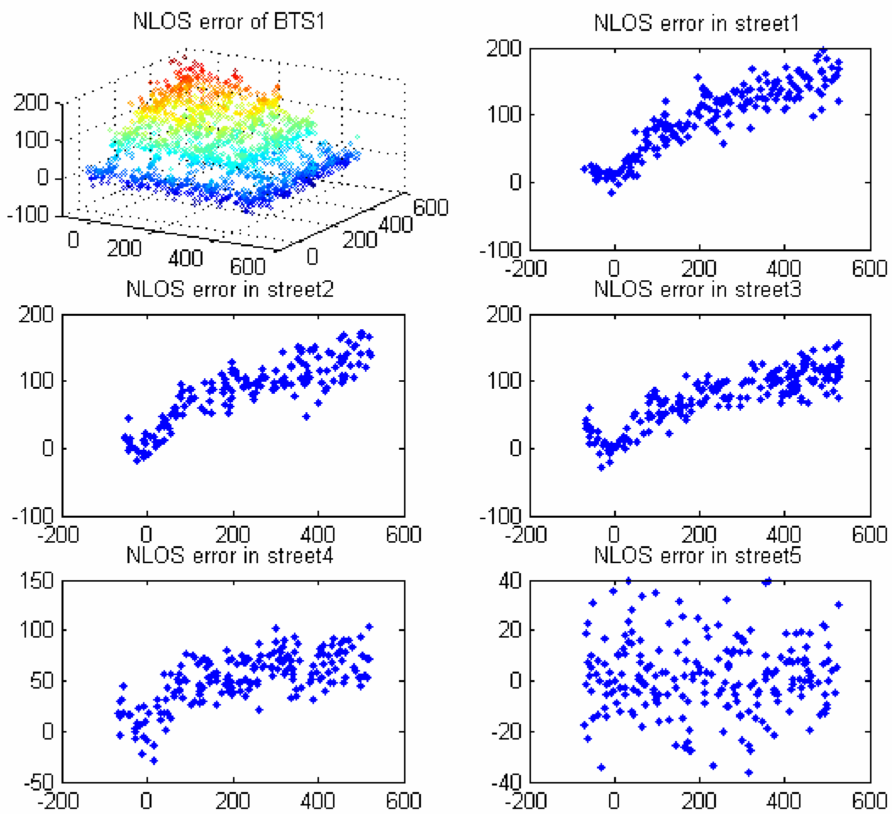


Fig. 2 NLOS error in Manhattan-like urban environment

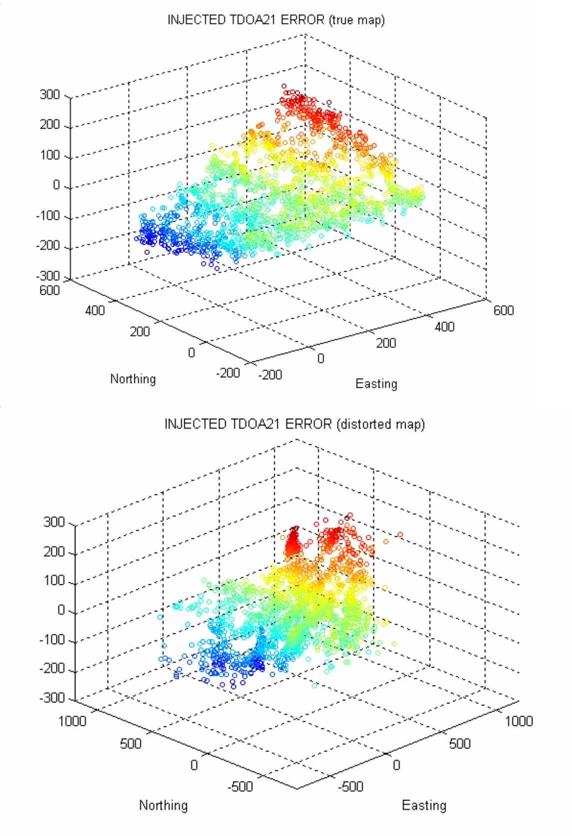


Fig. 3 Injected TDOA21 error in idea domain and distorted domain

3.2 Hyperbolic equation solving algorithms

Once the TDOA estimates have been obtained, they are converted into range difference measurements and these measurements can be converted into nonlinear hyperbolic equations. Assuming BTS1 is the reference BTS, let (x, y) be the MS location and (X_i, Y_i) be the known location of the ith BTS. The TDOA measurement is

$$R_{i,1} = cd_{i,1} = R_i - R_1 = \sqrt{(X_i - x)^2 + (Y_i - y)^2} - \sqrt{(X_1 - x)^2 + (Y_1 - y)^2} \tag{8}$$

Several algorithms have been proposed, such as Friedlander's method (Friedlander, 1987), Taylor-Series method (Foy, 1976), Fang's method (Fang, 1990) and Chan's method (Chan and Ho, 1994) etc. Each method has advantages and disadvantages. For example, Chan's method can provide exact solution; it also takes advantage of redundant measurements and further more it approaches CRLB (Cramer-Rao lower bound). However, occasionally, Chan's method has ambiguities in the solution, and if the measurement has large errors (including NLOS error), it cannot work efficiently. Taylor-Series method is an iterative method, and the redundant measurements can be used as well. The provisional value and GDOP can significantly affect the proceeding solution and convergence is not guaranteed. However, in the condition of large errors, it provides more freedom for tuning; so it works more efficiently than other methods. The Taylor-Series method is chosen to solve the hyperbolic equation in this paper.

With a set of TDOA estimates, the method starts with a provisional value (x_0, y_0) and computes the deviations of the location estimation dx and dy.

$$\begin{bmatrix} dx \\ dy \end{bmatrix} = \left(A'W^{-1}A\right)^{-1} A'Wb \tag{9}$$

where

$$A = \begin{bmatrix} \left[\dfrac{X_1 - x}{R_1}\right] - \left[\dfrac{X_2 - x}{R_2}\right] & \left[\dfrac{Y_1 - y}{R_1}\right] - \left[\dfrac{Y_2 - y}{R_2}\right] \\ \vdots & \vdots \\ \left[\dfrac{X_1 - x}{R_1}\right] - \left[\dfrac{X_M - x}{R_M}\right] & \left[\dfrac{Y_1 - y}{R_1}\right] - \left[\dfrac{Y_M - y}{R_M}\right] \end{bmatrix}$$

$$b = \begin{bmatrix} R_{2,1} - (R_2(x_0) - R_1(x_0)) \\ \vdots \\ R_{M,1} - (R_M(x_0) - R_1(x_0)) \end{bmatrix}$$

and W is the covariance matrix of the estimated TDOAs. The whole process is repeated until dx and dy are sufficiently small.

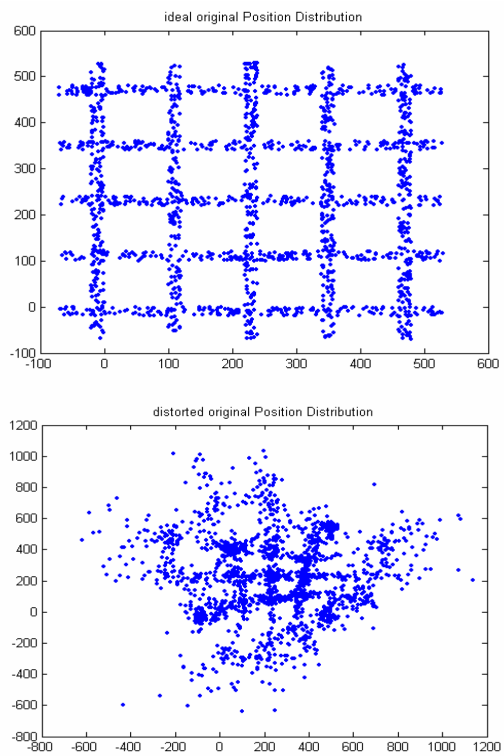


Fig. 4 Ideal MS position distribution and distorted MS position distribution

Due to the injected NLOS error and noise, the position estimates based on the TDOA do not coincide with the true position. The ideal MS position distribution and the distorted MS position distribution are shown in Figure 4.

4 Choosing a variogram model

Variogram γ is the basic tool for the structural interpretation of phenomena as well as for estimation. It is defined in (1). Normally, γ is not known and needs to be estimated from the TDOA measurements. There are several ways to estimate the variogram (Cressie, 1991). The classical formula is:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{x_i - x_j = h} \left(Z(x_i) - Z(x_j)\right)^2 \qquad (10)$$

Most of the time, the points are irregularly spaced. In order to have more pairs, the summation x_i-x_j=h has to be weakened.

$$\left\| x_i - x_j \right| - h \right| \le \varepsilon \;\; Angle(x_i - x_j, h) \le \delta$$

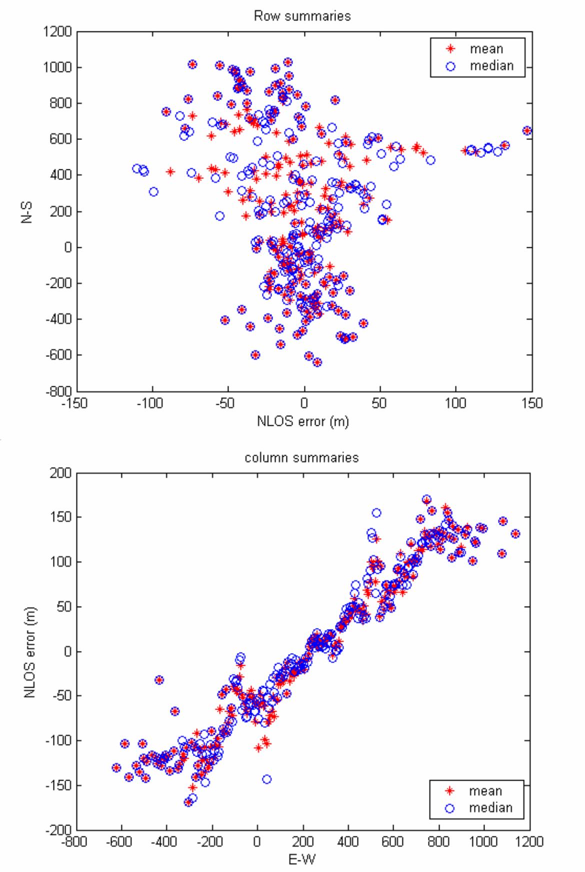


Fig. 5 Mean and median summaries of nonstationarity (top: row summaries, bottom: column summaries)

In order to make the processing easier, the data is adjusted slightly to a grided map. Before computing the γ, the nonstationarity of the injected NLOS error should be checked. Figure 5 is an attempt to summarize the possible nonstatioarity in the mean using the sample median and sample mean across rows and down columns. There appears to be a shift in the east-west direction but little or no shift in north-south direction. Computing the variogram in this direction first is a good choice.

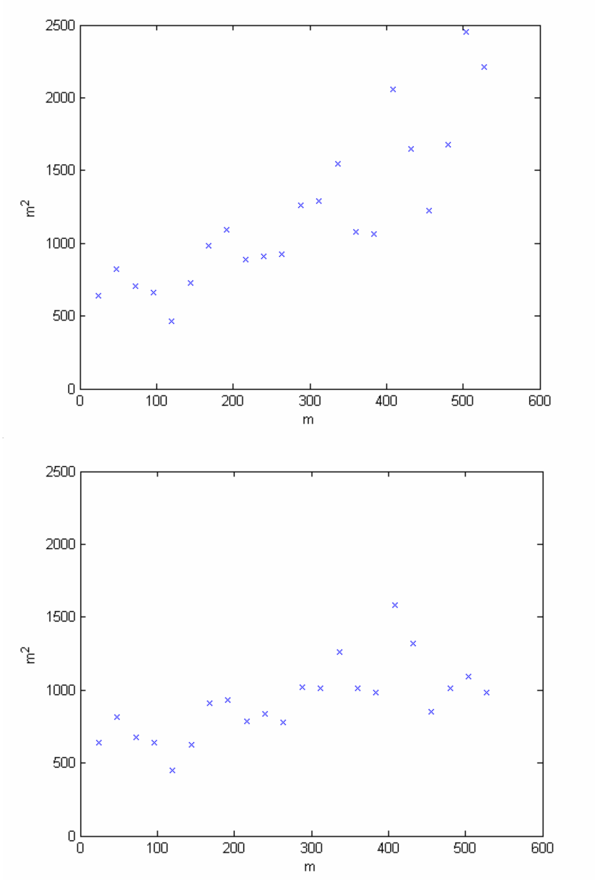
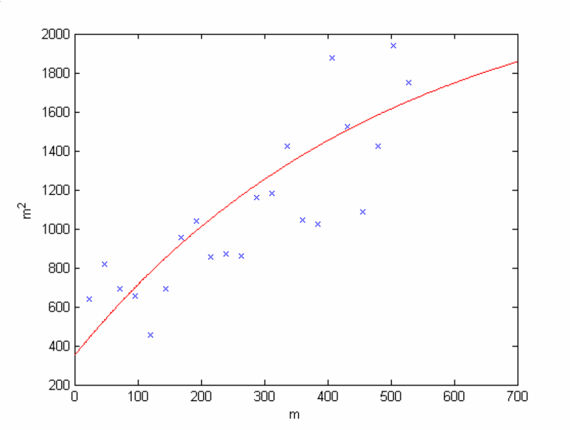


Fig. 6 (a) Estimated north-south variogram of injected TDOA21 error (top) (b) Estimated north-south variogram of injected TDOA21 error residuals (using o.l.s to get β) (bottom)

Unfortunately, there is still a little shift in north-south direction. Figure 6(a) depicts the variogram computed using the formula above, taking the width of each bin to be 24 meters. The effect of the trend is plainly obvious, leading to a steadily increasing parabolic-like curve. How to decompose the data with shift is a classical problem. Some approaches have been proposed. One of them is to start with o.l.s. (ordinary-least-squares) estimator of β in (3), compute a variogram estimator from the residuals, fit a variogram model, then obtain a g.l.s (general-least-squares) estimator of β based on the fitted model, and so forth. In this paper, this iterative approach is used, although this approach suffers from a bias problem (Cressie, 1991). Figure 6(b) shows the estimated north-south variogram of injected TDOA21 error residuals after o.l.s. The reason not to use the experimental variogram directly is because most of the experimental variograms are not admissible, as it needs to be conditionally negative definite. Only after few iterations, the result is converged. Finally, two models are chosen: exponential (11) and a spherical model (12).

$$\begin{cases} \gamma(0)=0 \\ \gamma(h)=C0+C\left(1-e^{-\frac{h}{a}}\right) \end{cases} \tag{11}$$

$$\begin{cases} \gamma(0)=0 \\ \gamma(0<h<a)=C0+C\left(\frac{3}{2}\frac{h}{a}+\frac{1}{2}\left(\frac{h}{a}\right)^3\right) \\ \gamma(h\geq a)=C0+C \end{cases} \tag{12}$$



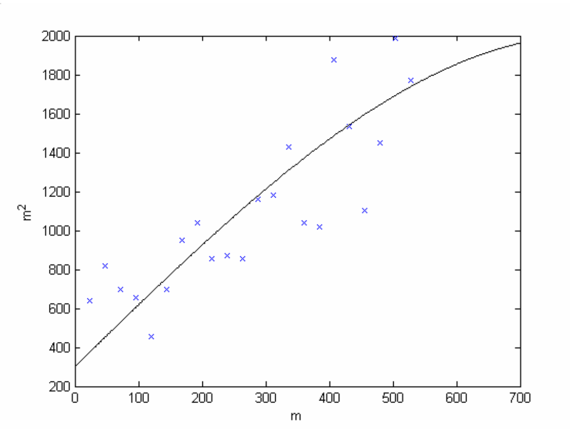


Fig. 7 The chosen variogram models: exponential model (top) and spherical model (bottom)

The parameters retained were: for exponential model, $C0$=350m^2, C=2000m^2, a=500m; and for spherical model, $C0$=300m^2, C=1700m^2, a=800m. These two models are shown in Figure 7. $\gamma(0)=0$ shows a discontinuity at the origin, which is called the nugget effect. This is caused by the unknown micro scale variation. The chosen models have a sill, which means that when the distance between two points is large enough, they are independent. This means the model can be described by covariance. In UK, some times the covariance matrix is needed (Cressie, 1991). The model is chosen visually. Some of the automatic methods such as least squares can be used, but it is not suggested by most of the professionals (Armstrong, 1998).

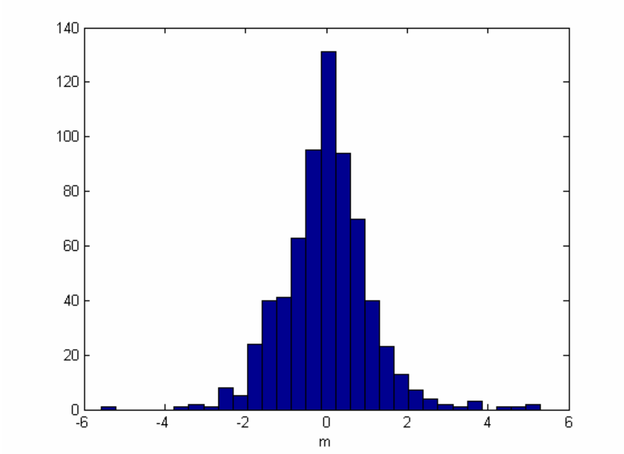


Fig. 8 Histogram of the normalized residuals based on exponential model

5 Results

The variogram model should be checked before the application. The evaluation of it was done through cross-validation (Isaaks and Srivastava, 1989; Armstrong, 1998). For each measurement location x_i the values are estimated as if they were unknown. The kriging variances in (5) are also computed. Then the normalized residual can be formed:

$$S=\frac{\hat{Z}-Z}{\sigma_K} \tag{13}$$

It should be normally distributed with 0 mean and 1 as standard deviation ($N(0,1)$). Figure 8 shows such a distribution using exponential model. The mean, standard deviation and max of the normalized residual using different model are compared in table 1. Here, the exponential model is slightly better than spherical model.

Tab. 1 Results of cross-validation

	Exponential model	Spherical model
Mean normalized residuals	0.0005	0.0006
Std. normalized residuals	1.0931	1.1882
Max normalized residuals	5.3041	5.7611

As mentioned before, in real application, obtaining the measurement of reference point is not trivial. WSMM can aid to get some references automatically. In the simulated Manhattan-like environment, 25 points can be derived automatically. A correction map should be generated by these reference points. Figure 9 plots the injected TDOA21 NLOS error correction map. It is similar with Figure 3 (bottom). Normalized residuals (the true data are known since they are generated by simulation) can be computed using (13). The results are listed in table 2.

Tab. 2 Results using 25 references

Mean of normalized residuals	0.1259
Std. of normalized residuals	1.0007
Max of normalized residuals	3.6893

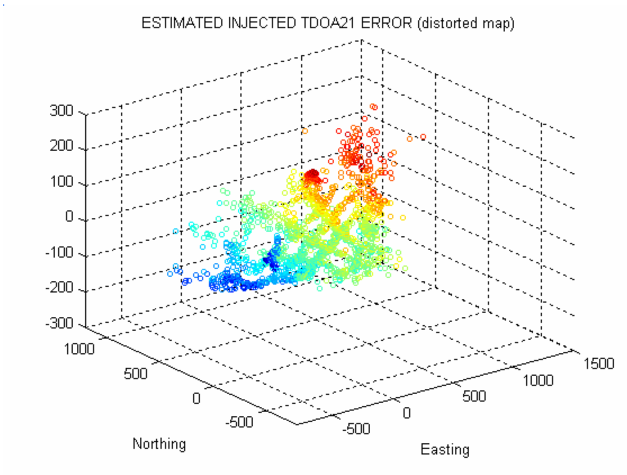


Fig. 9 Estimated injected TDOA21 error in distorted domain

Using the same algorithm, the injected TDOA31 NLOS error correction map can be generated. The corrected TDOA measurements can be derived by simply subtracting NLOS error from TDOA measurements. Then Taylor-Series method is applied again, and a new position distribution can be generated. Because of the reason mentioned in previous section, the NLOS error cannot be removed completely, but a part of it should be mitigated. The new distorted position exposes better relationship with the NLOS error. The processing can be done again. After two iterations, Figure 10 illustrates the final result. It is very clear after UK, the new distorted MS position distribution is closer to the ideal distribution.

6 Conclusions

The algorithm using kriging can efficiently estimate the NLOS error with the knowledge of some references and a reasonable variogram model. Normally, kriging is used in the situation where the exact location of measurement is known. In the application to mitigate NLOS error, however, the true location is the answer we are looking for. The distorted location covers some real specifics of the NLOS error and location. It makes the application harder. There are still some questions to be answered. First of all, the real data is necessary to verify this algorithm. Secondly, how to extract NLOS error at reference points efficiently. Thirdly how many references are needed to get rid of the NLOS error to a fix percentage. Finally, how to determine the variogram model for a specific environment, although some evidences show this application is not very sensitive with variogram model. Future work is focussed on optimising the algorithm for real world application.

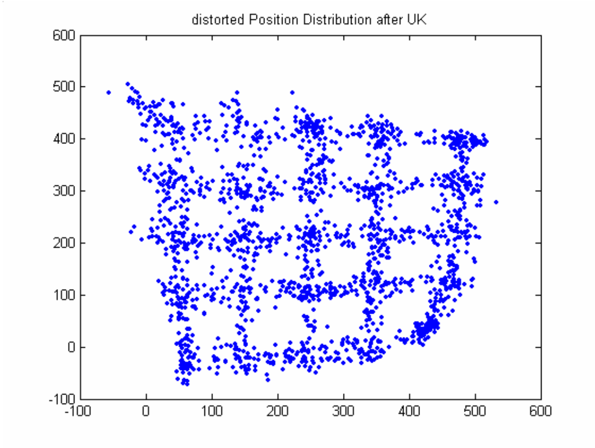


Fig. 10 The final distorted MS position distribution

References

Rappaport T.S.; Reed J.H.; Woerner B.D. (1996): ***Position location using wireless communications on highways of the future***, IEEE Communications Magazine, vol. 34, no. 10, Oct. 1996, pp. 33-41.

D'roza T.; Bilchev G. (2003): ***An overview of location-based services***, BT Technology Journal, January, Vol 21 No 1.

Wilde G. (2002): ***Making the most of legacy mobiles: examining the relationship between technology choices and location revenues***, BWCS white paper.

Dru M.A.; Saada S. (2001): ***Location-based mobile services: the essentials***, Alcatel telecommunications review, 1st quarter.

FCC-Enhanced 911 (E911) Home Page, http://www.fcc.gov/911/enhanced/

Lee W.C.Y. (1991): ***Overview of cellular CDMA***, IEEE Trans. on Vehicular Technology, May, Vol 40 No 2.

Yamamoto R.; Matsutani H.; Matsuki H.; Oono T.; Ohtsuka H. (2001): ***Position location technologies using signal strength in cellular systems***, IEEE VTC, 2570-2574.

Sakagami S.; Aoyama S.; Kuboi K.; Shirota S.; Akeyama A. (1992): ***Vehicle position estimates by multibeam antennas in multipath environments***, IEEE Transactions on Vehicular Technology, Volume: 41, Issue: 1, Feb. 63-68.

Hashemi H. (1991): ***Pulse ranging radiolocation technique and its application to channel assignment in digital cellular radio***, IEEE VTC, 675-680.

Drane C.; Macnaughtan M.; Scott C. (1998): ***Positioning GSM Telephones***, IEEE Communications Magazine, April, 46-59.

Jr. J.C.; Stüber G.L. (1998): ***Subscriber location in CDMA cellular networks***, IEEE Trans. Veh. Technol., May, vol. 47, 406-416.

Silventoinen M.I.; Rantalainen T. (1996): ***Mobile station emergency locating in GSM***, IEEE International conference on Personal Wireless Communication, February, pp. 232-38.

Morley G.; Grover W. (1995): ***Improved location estimation with pulse ranging in presence of shadowing and multipath excess-delay effects***, Electron. Lett., vol. 31, no.18, 1609-1610.

Wylie M.P.; Holtzman J. (1996): ***The non-line of sight problem in mobile location estimation***, 5th IEEE International Conference on Universal Personal Communications, vol. 2, 827–831.

Woo S.S.; You H.R.; Koh J.S. (2000): ***The NLOS mitigation technique for position location using IS-95 CDMA networks***, in Proc. IEEE VTC, September, vol. 6, 2556–2560.

Cong L.; Zhuang W. (2001): ***Non-Line-Of-Sight error mitigation in TDOA mobile location***, IEEE Global Telecommunications Conference, November, Vol. 1, 680-684.

Wang W.; Wang Z.; O'Dea B. (2003): ***A TOA-based location algorithm reducing the errors due to Non-Line-of-Sight (NLOS) propagation***, IEEE Trans.Veh. Technol., vol. 52,112-116.

Jayaraman S.; Wax M.; Hilsenrath O.A. (2000): ***Calibration table generation for wireless location determination***, US patent 6,101,390.

Gunnarsdottir H.M.; Hole D. (2001): ***Location detection technologies for cellular users***, EE359 Final Project, Department of Electrical Engineering, Stanford University, USA.

Lee H.K.; Rizos C. (2003): ***A frame work for calibrating NLOS error to support LBS in urban environment***, Int. Symp. on GPS/GNSS, 15-18 November, 69-77.

Aguado F.; Fontan F.P.; Formella A. (1997): ***Indoor and outdoor channel simulator based on ray tracing***, IEEE 47th VTC, Vol. 3, 2065-2069.

Friedlander B. (1987): ***A Passive Localization Algorithm and Its Accuracy Analysis***, IEEE Journal of Oceanic Engineering, January, vol. OE-12, no. 1, 234-244.

Foy W.H. (1976): ***Position-Location Solutions by Taylor-Series Estimation***, IEEE Transactions on Aerospace and Electronic Systems, March, vol. AES-12, 187-194.

Fang B.T. (1990): ***Simple Solutions for Hyperbolic and Related Fixes***, IEEE Transactions on Aerospace and Electronic Systems, September, vol. 26, no. 5, 748-753.

Chan Y.T.; Ho K.C. (1994): ***A Simple and Efficient Estimator for Hyperbolic Location***, IEEE Transactions on Signal Processing, August, Vol. 42, no. 8, 1905-1915.

Cressie N. (1991): ***Statistics for spatial data***, John Wiley & Sons, INC., New York.

Armstrong M. (1998): ***Basic linear geostatistics***, Springer, Berlin

Isaaks E.H.; Srivastava R.M. (1989): ***Applied Geostatistics***, Oxford university press, New York.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 36-40

Journal of Global Positioning Systems
2005
© CPGPS

Impact of Different Tropospheric Models on GPS Baseline Accuracy: Case Study in Thailand

Chalermchon Satirapod and Prapod Chalermwattanachai

Geo-Image Technology Research Unit, Department of Survey Engineering, Chulalongkorn University, Bangkok, Thailand
Email: chalermchon.s@chula.ac.th Tel: +66-2-2186662 Fax:+66-2-2186653

Received: 10 December 2004 / Accepted: 13 July 2005

Abstract. It is generally known that the atmospheric effects on the GPS signals are the most dominant spatially correlated biases. The atmosphere causing the delay in GPS signals consists of two main layers, ionosphere and troposphere. The ionospheric bias can be mitigated using dual frequency receivers. Unlike the ionospheric bias, the tropospheric bias cannot be removed using the same procedure. Compensation for the tropospheric bias is often carried out using a standard tropospheric model. Most standard tropospheric models were experimentally derived using available radiosonde data, which were mostly observed on the European and North American continents. In order to determine the best-fit standard tropospheric model with the GPS data collected in Thailand, investigations on the impact of different standard tropospheric models on GPS baseline accuracy are therefore needed. This paper aims to compare the GPS positioning results derived from the use of three different standard tropospheric models, namely the Saastamoinen model, Hopfield model and Simplified Hopfield model. In this study, both short and medium length baseline data sets were tested. In addition, each baseline data set is further divided into two scenarios, flat terrain and rough terrain. Overall results indicate that there are no statistically significant differences in the performance of the three tropospheric models. However, the use of the Saastamoinen and the Hopfield models tends to produce more reliable results than the use of the Simplified Hopfield model.

Key words: Tropospheric effect, Hopfield model, Saastamoinen model, Simplified Hopfield model, GPS baseline accuracy

1 Introduction

One of the factors limiting the GPS baseline accuracy is due to the atmospheric delay. The atmosphere causing the delay in GPS signals consists of two main layers, ionosphere and troposphere. The ionosphere is the band of the atmosphere from around 50km to 1000km above the earth's surface (Hofmann-Wellenhof et al., 1997; Langley 1998; Rizos, 1997). The ionospheric delay is a function of the total electron content along the signal path, and the frequency of the propagated signal. With regard to the dual-frequency user, the ionospheric delay is frequency-dependent and the ionosphere-free combination can be formed in order to eliminate this delay (Hofmann-Wellenhof et al., 1997; Leick, 1995; Rizos, 1997). The troposphere is the band of the atmosphere from the earth's surface to about 8km over the poles and 16km over the equator (Langley 1998; Rizos, 1997). The tropospheric delay is a function of elevation and altitude of the receiver, and is dependent on many factors such as atmospheric pressure, temperature and relative humidity. Unlike the ionospheric delay, the tropospheric delay is not frequency-dependent. It cannot therefore be eliminated through linear combinations of L1 and L2 observations. Several standard tropospheric models (e.g. Saastamoinen model, Hopfield model, etc.) are generally used to correct for the tropospheric delay.

All standard tropospheric models are empirically derived from available radiosonde data, which were mostly obtained in the European and North American continents. Global constants within some standard models take no account of latitudinal and seasonal variations of parameters in the atmosphere (Roberts and Rizos, 2001). Furthermore, daily variations of temperature and humidity may cause the tropospheric effects derived from standard models to be in error especially in the height component (Rühnöbl et al., 1998). The high and variable water vapor content, particularly in equatorial regions, may exaggerate this effect further (Mendes, 1999).

Gurtner et al. (1989) also states that tropospheric modelling is only valid for a flat terrain. A large height difference for the baseline points can introduce a bias of the order of 2-5 mm per 100m height difference. Roberts (2002) recommends that the effects of differential troposphere on the height component should be estimated as an additional parameter during a baseline estimation step.

In Thailand, an investigation on the impact of tropospheric delay is still very limited. *What is of particular interest to the GPS surveyors in Thailand is which standard tropospheric model should be used in the baseline processing.* In order to determine the best-fit tropospheric model for processing of the data collected in Thailand, investigations on the impact of different global tropospheric models on GPS baseline accuracy are therefore needed. This paper aims to emphasise an impact of the tropospheric delay on GPS baseline accuracy as well as to compare the GPS positioning results derived from the use of the three tropospheric models, namely the Saastamoinen model (Saastamoinen, 1973), Hopfield model (Hopfield, 1969) and Simplified Hopfield model (Wells, 1977). These models are available in most GPS software packages. This paper is organised as follows. The second section describes data sets used in a subsequent analysis. The third section explains how the data sets are processed. The fourth section presents an analysis of the results, followed by some concluding remarks in the final section.

2 Test data

In this study, both short and medium length baseline data sets were collected. Each baseline length data set is further divided into two scenarios, flat terrain and rough terrain. The details of data sets are given in this section. It should be noted that ground meteorological data (i.e. temperature and air pressure) at each station were also observed every hour.

2.1 Short Length Baseline Case

The short length baseline data were collected in static mode for 24 hours starting from 10:00am on 7th June 2003 to 10:00am on 8th June 2003 with three dual-frequency receivers (Leica SR530) at a 15-second data rate. In order to investigate an impact of standard tropospheric models on different terrains, the first receiver was set up at Station ‘A’ situated on top of the Pra Baht Pluang mountain, while the other two receivers were set up at Station ‘B’ and ‘C’ situated in a flat area. The baseline length between A and B is approximately 17 km and this baseline represents a rough terrain scenario. The baseline length between B and C is about 11 km and this baseline represents a flat terrain scenario. The height difference between A and B is about 950 m while the height difference between B and C is only about 35 m. Figure 1 illustrates the configuration of the short baseline case.

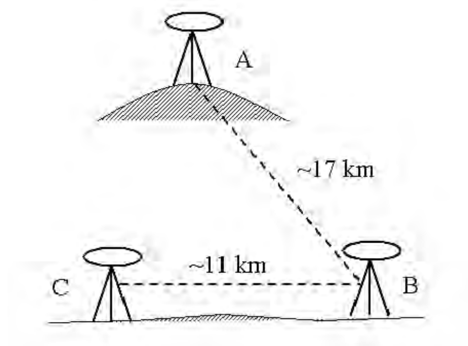


Fig. 1 Configuration of short baseline case

2.2 Medium Length Baseline Case

The medium length baseline data were collected in static mode for 24 hours starting from 8:00am on 14th June 2003 to 8:00am on 15th June 2003 with the same receivers and data rate. The first receiver was set up at Station ‘D’ which is close to station ‘A’, while the other two receivers were set up at Station ‘E’ and ‘F’ situated in a flat area.

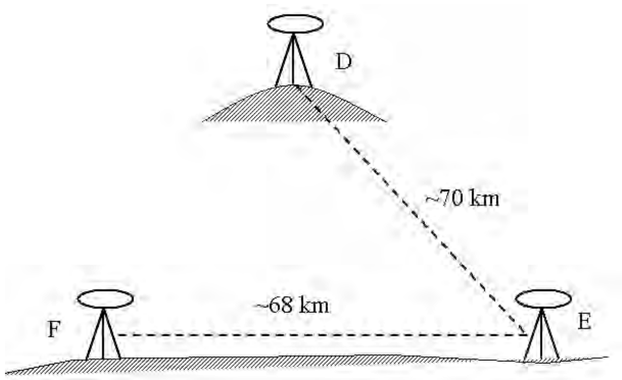


Fig. 2 Configuration of medium length baseline case

The baseline length between D and E is approximately 70 km and this baseline represents a rough terrain scenario. The baseline length between E and F is about 68 km and this baseline represents a flat terrain scenario. The height difference between D and E is about 970 m while the height difference between E and F is only about 25 m. Figure 2 illustrates the configuration of the medium length baseline case.

3 Data processing

3.1 Establishment of Reference Coordinates

Since the coordinates of stations 'B' and 'E' are known, these stations are held fixed in the baseline estimation step for the short and medium length baseline cases respectively. In order to obtain accurate coordinates for stations A, C, D and F, the University of Bern precise GPS data processing software, referred to simply as the 'Bernese software', was used to compute the coordinates of these stations. Table 1 gives a summary of options selected in a baseline estimation step.

Tab. 1 Processing options used in the Bernese software

Baseline	Orbit used	Tropospheric model applied	Solutions
Short length	Broadcast	Estimate as additional parameter	Ionosphere-free fixed double difference
Medium length	Broadcast	Estimate as additional parameter	Ionosphere-free float double difference

The 24-hr data sets were then processed with the Bernese software version 4.2 using the options presented in Table 1. The obtained coordinates are subsequently converted to UTM coordinates and presented in Table 2. These UTM coordinates will be used as references for subsequent analyses.

Tab. 2 Reference coordinates obtained from the Bernese software

Station	Northing (m)	Easting (m)	Height (m)
A	827650.831	1397561.694	-1.855
C	843951.136	1421202.169	980.850
D	735044.696	1487880.450	-17.903
F	843952.151	1421201.949	980.411

3.2 Baseline Processing

For convenience, all data sets are processed with the SKI software version 2.5. Processing options used in the SKI software are the same as the options used the Bernese software except that tropospheric modelling was used rather than the more rigorous parameter estimation approach using Bernese. Data processing strategies for each baseline length are described in this section.

3.2.1 Processing of short length baseline data

For the short length baseline case, the data sets were divided into 12 batches, each of 2 hours length. Each batch was treated as an individual session and processed using the following tropospheric models:

- Saastamoinen model
- Hopfield model
- Simplified Hopfield model
- No model applied

3.2.2 Processing of medium length baseline data

As the baseline length becomes longer, a minimum of 3-hr per observation session is needed. Thus, the data sets were divided into 8 batches, each of 3 hours length. Each batch was again treated as an individual session and processed using the same procedure as in the short length baseline case.

4. Analysis of results

In the following analyses, the discrepancies in the three coordinate components compared to the reference coordinates were firstly calculated. The performance of each standard tropospheric model can be characterised by the Root Mean Square Error (RMSE). Therefore, the RMSE values in both horizontal and vertical components for the stations A, C, D and F were computed and presented in Table 3. It can be seen from Table 3 that by applying any standard tropospheric model in the baseline estimation step, accuracies of coordinates in both horizontal and vertical components are improved. In addition, all RMSE values indicate that the Saastamoinen and the Hopfield models tend to produce more reliable baseline results than the Simplified Hopfield model.

In a further investigation, the hypothesis test was carried out to find out if the differences in performance of each standard tropospheric model are statistically significant. These differences are individually tested for horizontal and vertical components. For each station, the smallest RMSE value in each component was selected as a reference RMSE value. It should be noted that numbers highlighted with red color in the Table 3 indicate the smallest RMSE value for each case. The commonly used two-tailed F-test was chosen to test if the reference RMSE value and the other RMSE values are equal. The F hypothesis test is defined as (Snedecor and Cochran, 1989):

$H_0 : \sigma_1 = \sigma_n$ (Null hypothesis)

$H_a : \sigma_1 \neq \sigma_n$ (Alternative hypothesis)

σ_1 denotes the reference RMSE value calculated from the best-fit tropospheric model, while σ_n is the RMSE value calculated from the other tropospheric model. 5% significance level was used for the hypothesis testing. Table 4 shows a summary of the results obtained from the hypothesis testing.

Tab. 3 Summary of RMSE values of stations A, C ,D and F in horizontal and vertical components

Baseline	Terrain	Station	Tropospheric model Applied	RMSE (m)	
				Horizontal	Vertical
Short Length	Rough	A	Saastamoinen	0.014	0.064
			Hopfield	0.014	0.064
			Simplified Hopfield	0.012	0.083
			No model	0.135	0.835
	Flat	C	Saastamoinen	0.007	0.019
			Hopfield	0.007	0.019
			Simplified Hopfield	0.007	0.023
			No model	0.017	0.020
Medium Length	Rough	D	Saastamoinen	0.058	0.068
			Hopfield	0.058	0.068
			Simplified Hopfield	0.076	0.078
			No model	0.204	0.899
	Flat	F	Saastamoinen	0.060	0.060
			Hopfield	0.060	0.060
			Simplified Hopfield	0.060	0.062
			No model	0.188	0.097

Tab. 4 Summary of results using F-test at 5% significance level

Baseline	Terrain	Station	Tropospheric model Applied	Null hypothesis (H_0)	
				Horizontal	Vertical
Short Length	Rough	A	Saastamoinen	Accept	Reference
			Hopfield	Accept	Reference
			Simplified Hopfield	Reference	Accept
			No model	Reject	Reject
	Flat	C	Saastamoinen	Reference	Reference
			Hopfield	Reference	Reference
			Simplified Hopfield	Reference	Accept
			No model	Reject	Accept
Medium Length	Rough	D	Saastamoinen	Reference	Reference
			Hopfield	Reference	Reference
			Simplified Hopfield	Accept	Accept
			No model	Reject	Reject
	Flat	F	Saastamoinen	Reference	Reference
			Hopfield	Reference	Reference
			Simplified Hopfield	Reference	Accept
			No model	Reject	Accept

In relation to the results presented in Table 4 the following can be noted

- Neglecting the use of a standard tropospheric model in the baseline estimation step leads to unreliable baseline results especially in the case of the rough terrain.
- The three standard tropospheric models produce baseline results that are not statistically different.

An analysis of the relationship between the obtainable baseline accuracies and the ground meteorological data was further carried out. This analysis aims to find out if correlations between the obtainable baseline accuracies and the ground meteorological data are statistically significant. The results revealed that there is no statistically significant correlation between the obtainable baseline accuracies and the ground meteorological data. Hence, it implies that the weather conditions have no significant impact on the baseline results.

5. Concluding remarks

This paper has demonstrated the impact of tropospheric effect on GPS baseline accuracy. The omission of applying a standard tropospheric model in the baseline processing leads to unreliable baseline results especially in the case of large changes in height between stations. The testing procedure for determining the best-fit tropospheric model has also been presented in this paper. Based on the F hypothesis test performed in this study, the three standard tropospheric models, Saastamoinen, Hopfield and Simplified Hopfield models, produce baseline results that are not statistically different. However, the use of the Saastamoinen model and the Hopfield model tends to produce the most reliable baseline results, and hence they are recommended to be used in the baseline processing.

Acknowledgements

This research is supported by the Ratchadaphiseksomphot grant from Chulalongkorn University. The authors would like to thank to Dr. Craig Roberts, a lecturer at the School of Surveying and Spatial Information Systems, University of New South Wales, for his valuable suggestions on this paper.

References

Gurtner, W.; Beutler, G.; Botton, S.; Rothacher, M.; Geiger, A.; Kahle, H.-G.; Schneider, D.; Wiget, A. (1989): ***The use of the Global Positioning System in mountainous areas***, Manuscripta Geodetica, 14, 53-60.

Hopfield, H.S. (1969): ***Two–quadratic tropospheric refractivity profile for correction satellite data***, Journal of Geophysical Research, 74(18), 4487 – 4499.

Hofmann-Wellenhof, B.; Lichtenegger, H.; Collins, J. (1997): ***Global Positioning System: Theory and Practice***, 4th edition, Springer-Verlag, Berlin Heidelberg New York, 389pp.

Langley, R.B. (1998): ***Propagation of the GPS Signals***, In: Kleusberg, A. and Teunissen, P.J.G. (eds), GPS for Geodesy (2nd edition), Springer-Verlag, Berlin Heidelberg New York, 111-150.

Leick, A. (1995): ***GPS Satellite Surveying***, 2nd edition, John Wiley & Sons, Inc., New York, 560pp.

Mendes, V.B. (1999): ***Modeling the neutral-atmosphere propagation delay in radiometric space techniques***, Ph.D. dissertation, Dept. of Geodesy & Geomatics Engineering Technical Report No.199, University of New Brunswick, Canada, 353pp.

Rizos, C. (1997): ***Principles and Practice of GPS Surveying***, Monograph 17, School of Geomatic Engineering, The University of New South Wales, 555pp.

Roberts, C. (2002): ***A Continuous Low-Cost GPS-Based Volcano Deformation Monitoring System in Indonesia***, PhD Thesis, School of Surveying and Spatial Information Systems, University of New South Wales, Australia, 271pp.

Roberts, C.; Rizos, C. (2001): ***Mitigating differential troposphere for GPS-based volcano monitoring***. Proceedings of 5th Int. Symp. on Satellite Navigation Technology & Applications, Canberra, Australia, 24-27 July, paper 38.

Rührnößl, H.; Brunner, F.K.; Rothacher, M. (1998): ***Modellierung der Troposphärischen Korrektur für Deformationsmessungen mit GPS im alpinen Raum***, Allg. Verm. Nach., 1/98, 14-20.

Saastamoinen, J. (1973): ***Contribution to the theory of atmospheric refraction***, Bulletin Geodesique, 107,13-34.

Snedecor, G.W.; Cochran, W.G. (1989): ***Statistical Methods***, 8th Edition, Iowa State University Press.

Wells, D.E. (1977): ***Doppler Satellite Control***, Dept. of Geodesy & Geomatics Engineering Technical Report No. 29, University of New Brunswick, Canada.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 41-47

Journal of Global Positioning Systems
2005

Syren - A Ship Based Location-Aware Audio Experience

Daniel Woo[1], Nick Mariette[1], Nigel Helyer[2] and Chris Rizos[3]

[1]School of Computer Science and Engineering, The University of New South Wales, UNSW Sydney NSW 2052, Australia
[2]Sonic Objects, P.O. Box 195, Balmain NSW 2041, Australia.
[3]School of Surveying and Spatial Information Systems, The University of New South Wales, UNSW Sydney NSW 2052, Australia
e-mail: danielw@cse.unsw.edu.au Tel: + 61 2 9385 6495; Fax: +61 2 9385 5995

Received: 6 December 2004 / Accepted: 20 October 2005

Abstract. Syren, a location-based, multi-speaker augmented audio reality installation was presented as a shipboard exhibit at the 12th International Symposium on Electronic Art in August 2004. It was conceived as a continuous 3-day spatial audio experience that augments the landscape through the Baltic archipelago with location-based audio media, spatialised through a 12-channel speaker array. As the ship tracks between Helsinki, Mariehamn, Stockholm and Tallinn, listeners on the upper deck hear sounds that are perceived to originate from geographic features. Our custom GIS is derived from electronic nautical charting information that includes coastlines, buoys and beacons. A handheld GPS provides both position and direction data that was used by a software system to drive parameters of the spatial audio presentation. The sound production for the artwork was created using the custom application that enabled the artist to place sound media in relation to a real-world map. An important component to this software was the ability to audition the audio experience without ever taking the journey.

Keywords: Augmented audio reality, GPS applications, location-based media, Geographical Information Systems (GIS)

1 Introduction

Syren is an artwork that uses location-based spatial sound as the central user experience. It was presented onboard a passenger cruise ship on the Baltic Sea as part of the International Symposium on Electronic Art 2004. The journey was between Helsinki, Mariehamn, Stockholm and Tallinn. The nautical setting provided a unique opportunity to deploy a novel augmented audio reality artwork, presenting listeners on the upper deck with a spatial sound experience. As the ship moved through the Baltic archipelago sounds placed by the artist would be presented to listeners through a multi-channel speaker array. The visual landscape was augmented with audio media that had been attached to physical landmarks. As the ship approached an island, the sounds attached to that island would become louder as if the island was the source of sound. As the island moved from bow to stern, the augmented sound moved and attenuated in concert with the visual stimulus of reality.

Syren is the first project to be produced by an art / science / engineering collaboration between sound artist Nigel Helyer, Daniel Woo (User Interface Design, Usability and Implementation) and Chris Rizos (GPS and Spatial Information Systems). The focus of the project is to better understand the user interface tools that will support the creation of artistically motivated location-based deployments. In this study, a ship-based setting is explored but in the future, the tools should work with a range of emerging mobile devices and wearable audio interfaces such as those described by Cohen (2002).

1.1 Prior Art

Not long after consumer handheld GPS receivers became available in the early 1990s, artists utilised GPS in a multitude of ways. Early GPS art explored the implications of the technology and its capability to record the position traces of movements in the world. Kurgan (1995) displayed points derived from a fixed GPS to illustrate the random nature of the GPS signal, and also produced images of the lines and the letters “MUSEU” recorded using a handheld GPS receiver. The Field-Works series of projects by Fujihata (1992) used raw GPS data to create digital images such as ‘Impressing Velocity’ in 1992-1994, and has explored the use of situated photographic images recorded with position captured on a GPS device (eg. 'Tsumari' a workshop carried out in 2000). Many other artists have similarly used GPS in projects that annotate locations in the world with electronic media objects. This activity has encouraged the emergence of online artist communities concerned with situated media, such as the 'Locative Network' . Work by the group 'C5 Corp' is also notable with an aim to utilise "GIS and GPS technology for research into the epistemic function of human cooperation".

Another unique and pertinent example of GPS art is 'Sound Mappings' by Mott & Sosnin (1997), which is one of the earliest GPS art works using the medium of audio. In Sound Mappings, the urban landscape is used as the foundation of a real-time musical composition that responds to the GPS-tracked movements of participants.

Augmented reality (AR) superimposes synthetic representations of location-based information upon what we naturally perceive of the real world using our senses. Augmented audio reality specifically presents information via the sense of hearing using spatial audio. Positioning technology is a vital component of any AR system in order to locate the user in the world. Cohen (1994) identified GPS would be an appropriate positioning technology for audio augmented reality applications.

Location-based audio can be distinguished from audio AR in that it does not require content to be presented in a spatialised manner.

Another variant of audio AR is the project 'Hear and There' that used GPS as the outdoors positioning technology in a system that enabled users to annotate a courtyard space with sound samples (Rozier, Karahalios & Donath, 2000). The user can participate by generating or receiving audio content. A range of other audio AR projects and research exists that use positioning technologies other than GPS, such as radio frequency positioning and infrared.

Pedestrian user, location-based audio artworks have been implemented by Perry (2002), in 'Invisible Ideas', and Knowlton, Spellman & Hight, (2002) in '34 North, 118 West'. 'Invisible Ideas' is a collaborative site-specific, non-spatialised audio and visual piece set in the Boston Public Garden. It runs on a on a GPS-equipped Pocket PC handheld device. '34 North, 118 West' uses site-specific, non-spatialised audio presented in the streets of Los Angeles to tell a story about the local history of the railroad network.

GPS and spatial audio projects have appeared in Helyer's sound art pieces. Helyer and Rizos had worked together in 1999 when Helyer, the artist in residence at Lake Technology had developed Sonic Landscapes, a mobile virtual audio reality experience that was presented in St. Stephens graveyard in Newtown, Australia (Helyer 2003). As listeners moved through the graveyard they could hear spatialised sounds emerge from landmarks in a very realistic manner (Fig. 2). This included content relating to the occupants of graves. Dynamic content, such as a plane flying overhead and a ghost that followed the listener were included in the artwork. This was achieved using spatial audio rendering technologies developed by Lake Technology. The hardware consisted of a backpack mounted notebook computer, GPS with external antenna and digital compass head-mounted on stereo headphones (Fig. 1).

Fig. 1 Backpack worn by St Stephens visitor

Fig. 2 Visitor viewing a landmark in the graveyard

1.2 ISEA2004

The sea journey component proposed for the International Symposium on Electronic Art 2004 (ISEA2004) provided a novel arena to deploy a new location-based artwork. ISEA is a conference predominantly attended by practising artists, cultural and social theorists with a strong interest in electronic arts. Syren was designed for the helideck of the ship where listeners could walk within a region surrounded by an array of 12 speakers (Fig. 3). The speaker array provided the opportunity for many people to experience the spatial audio effect simultaneously without having to wear headphones.

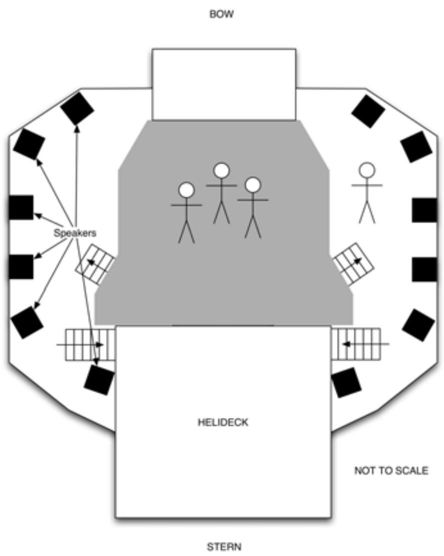


Fig. 3 Plan of the upper forward deck alongside the helideck.

From the forward, upper deck vantage point, the audience could see the panoramic view of the ocean and hear an augmented audio environment in which sounds were perceived to originate from locations in the landscape (Fig. 4).

Fig. 4 Listeners would hear a sound to originate from distant landmarks

Being sea-based, the GPS-based artwork did not suffer from the "urban canyon" issues that commonly arise in city-based GPS applications due to trees, awnings and buildings. The open water setting provided a clear view of the available satellites. The most significant unknown was how well the digital compass module in the handheld GPS unit would operate when placed on board a large metal object such as a ship. Fortunately, the bearing could be calculated within the handheld GPS device based on the actual course of the ship when travelling above 10 knots. The digital compass proved to be inaccurate at slow speeds onboard the ship.The journey took place between Helsinki, Mariehamn, Stockholm, Mariehamn and Tallinn, taking approximately 41 hours, covering 520 nautical miles (963 kms). The artistic design challenge was to develop enough content to span the journey. Some aspects of the content were scripted but other aspects used collections of audio files that could be played randomly or in sequence. A granular synthesis effect (Roads, 2001) was used to create ambient background sounds that gave a rich but unusual presence, requiring only a small section of audio.

2 Technology

2.1 Hardware System

The hardware system (Fig. 5) consisted of a G5 2.0GHz dual processor computer running Mac OS X 10.3.4. A Garmin Etrex system was connected serially via a USB RS-232 interface to provide location data in standard NMEA format. A separate power supply was used to power the Garmin to avoid the need to change batteries over the 3-day journey.

Total analog audio output of up to 16 channels was provided using two external audio devices. A MOTU828 MkII audio output device provided 8 analog audio channels. An additional 8 analog channels was produced by optically connecting a Behringer ADA-8000 to the MOTU828. The analog channels from these devices were directly connected to 12 amplified speakers. Software was used to dynamically mix the audio over a number of speakers to achieve the spatial audio experience.

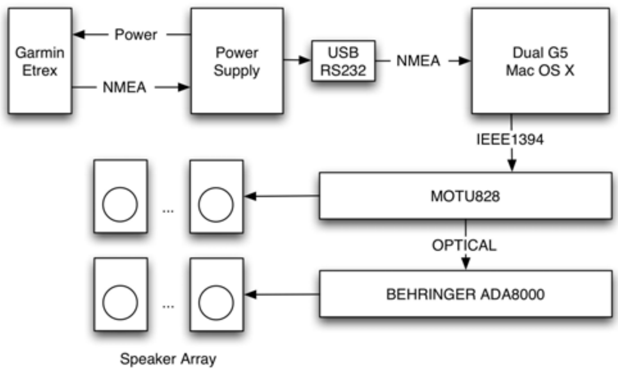


Fig. 5 Hardware Setup

2.2 Software System

The software system provided: 1) an accurate vector map representation of the Baltic Sea region; 2) an editing system for the artist to place and audition sounds (and monitor the real time operation of the system); 3) a simple GIS, consisting of a database of audio regions, the sounds and audio effects associated with each region; 4) a GPS interface to provide both position and bearing information; 5) audio algorithms to determine loudness and angle of a sound source relative to the listener updated each second; and 6) audio rendering process to render the sound source so that it is mixed correctly for the speaker array.

2.2.1 Map representation

The map data representing coastlines and navigation features was derived from electronic nautical chart (ENC) data. ENC data was obtained from the Finnish, Swedish and Estonian maritime authorities and was made available in standard S57 format (International Hydrographic Bureau, 2000). This data is based on the ISO8211 file format which provides a method for accessing structured data. An open source library (Warmerdam 2004) was used to implement access to the ISO8211 file information. Two strategies were used to extract the data: 1) a simple extraction of the necessary features (eg. coastline outlines); and 2) a more complicated method based around ISO8211 data structures. The former turned out to be the most practical within the time constraints but reduced the total number of features we could immediately access from the data set.

2.2.2 Editing software

For editing purposes, a user interface tool, VectorMap (Fig. 6) was developed on Mac OS X using the Objective-C programming language and the Cocoa environment. VectorMap provided the artist with a digital map of the Baltic Sea with scales ranging from (1:5000 to 1:1500000) and a method for laying out circular regions containing one or more audio files. Each region and associated audio files were referred to as a sound stack.

A user-centred design approach was adopted to encourage a close working relationship with the sound artist. This was done to identify how the sound artist would use the tools rather than insisting that the artist learn how to use tools developed by engineers, for engineers. Many computer-based tools for audio design are based on the software developer's perspective, not an artist's. Our aim is to avoid this pitfall in design.

The artist needed to access a large audio library, audition the audio files and define regions in which one or more audio files could be heard. Listening is an important part of the sound artist's workflow and the ability to simulate the position and bearing of the ship a key feature of the auditioning process. Previewing the audio content that was stored in the library was a requirement but there was also a need to understand the context of sounds presented in-situ. A map-based auditioning feature was mouse or keyboard driven allowing the ship position to be dragged with the mouse and the bearing modified with the keyboard. A variant on moving the ship was to use a metaphor of the ship being on the end of a string.

Circular regions were used over arbitrary polygon regions for reasons of simplicity. In physical terms it is straightforward to conceptualise a point source of audio radiating out over a circular region. Future versions of the software will support a range of tools for defining audio sound regions.

Each sound stack could be assigned a name, a display colour and the radius of the region could be adjusted. Variable opacity was used to clearly indicate to the designer the location of overlapping audio regions. When overlapping regions occurred, a context menu was provided to allow the selection of any specific region (by sound stack name) that contained the current mouse position.

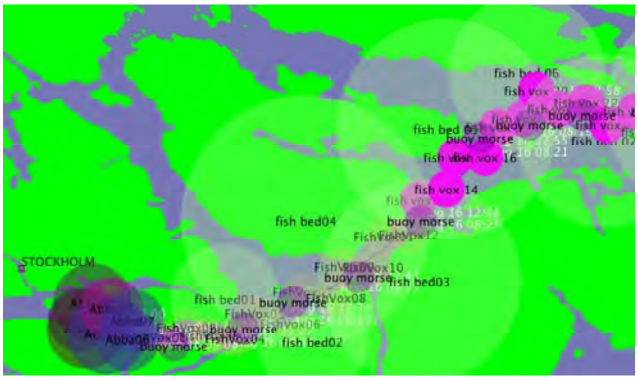


Fig. 6 Digital map user interface showing the audio design into Stockholm harbour.

Features that were requested but not implemented in the first version included support for audio layers, similar to image layers provided by graphical applications such as Adobe Photoshop, that could be enabled or disabled providing a mechanism for grouping audio regions with similar properties. The ability to search for regions by name was also suggested.

GPS track data from the shipping company was provided as a guide to indicate the path the ship was likely to take. The path was superimposed on the map indicating the expected times to reach the waypoints. Knowing the time of day that the ship was in a specific location was an important piece of information to assess whether people would be on deck or asleep. Design effort could be prioritised accordingly. Ideally, track data could be contained in a separate layer.

2.2.3 Database

The entire set of audio regions was defined in a custom database that essentially implemented a simple GIS appropriate for the project needs. The database was interrogated every second to compare whether the current position of the ship (the current position of the audition cursor or the actual position of the ship) was within one or more audio regions. The entire set of audio regions was contained in memory. No optimisation or caching was considered necessary for this prototype. This however provided rapid response when hit testing whether the region was to be processed or not.

The sound library audio content was not stored in this database, instead each audio region keeps a reference to the audio file locations on the hard drive.

2.2.4 GPS interface

The Garmin Etrex presented NMEA data serially at regular intervals of approximately one second. The device also provided compass information that was either calculated based on the current course (NMEA tag $GPHDG) or directly determined via a magnetic digital compass module (Garmin NMEA extension tag $HCHDG) embedded in the Garmin device. The Garmin simultaneously presented both pieces of direction data in the NMEA data stream. At slow speeds, the calculated heading by the device is inaccurate and the Garmin default threshold of 10 knots was used to determine whether to use the calculated or magnetic compass bearing.

The positioning aspect of the software system was designed so that it would be a simple task to switch between auditioning mode (where the user controls the boat position) and GPS mode (based on the NMEA data) without the underlying subsystems needing to know the source. This feature proved to be highly useful at slow speeds when the digital compass became inaccurate, and hence manual steering of the boat cursor was needed. It was useful for providing live demonstrations during the cruise, allowing manual positioning or rotating of the boat.

Code for GPS interfacing was based on the open source project Four Coordinates (Rosellini, 2003).

2.2.5 Audio lookup

In the VectorMap software, audio lookup is a process that occurs at regular intervals each time the GPS receiver generates a new position. The process manages a list of audible sounds that is based on their proximity to the current position. A search of the memory resident database is carried out to identify the top twenty sound stacks that should be audible at any point in time. A region is audible if the position of the ship falls within the radius of the sound stack region. For each audible sound, the relative bearing and volume level is calculated and passed to the rendering process.

The audio lookup process is also responsible for managing the start and stop triggering of sounds when the ship has entered or exited a sound region.

2.2.6 Audio rendering

Audio signal processing was implemented using Pure Data (Pd), Miller Puckette's real-time, music and multimedia graphical programming environment (Puckette, 1996).. The rendering of the positional audio through the speaker array utilised vector-based amplitude panning (Pulkki, 1997). The number of simultaneous sound sources was arbitrarily set to 20.

Several audio effects were implemented to provide the artist with alternative ways to enhance the audio content. A single granular synthesis (Roads, 2001) processor was developed to generate spatialised, unrecognisable, random-sounding audio effects from given audio samples. This effect was used as ambient audio, and provided an easy method for creating passages of constantly varying, random sound within controlled timbrel bounds. This provided a means to create large tracts of ambient sound content using relatively short sound files.

Sound source audio playback uses the Pd object 'readanysf~' by Black (2003). The audio system was controlled in real-time from the editing application via internal UDP packets containing data coded using the Open Sound Control protocol (Wright & Freed, 1997).

The original audio system included a 12.4 loudspeaker system that consisted of 12 loudspeakers arranged approximately in a circle, but were constrained by the layout of the deck. For editing purposes, a stereo audio output system was provided for headphone or desktop speaker use.

3 Discussion

3.1 Flexible Interfaces

Flexibility of the software tools was an important consideration for the project since there were many potential unknowns: this was the first time that such a project had been developed by this team in a geographic region that was unfamiliar to all of us. The artist was concerned that we may overlook visible landmarks that whilst not obvious from the map data were glaringly obvious from upper deck of the boat. The system needed to support live updating of the audio database. This proved to be very useful when the ship approached Stockholm, docking at an area that was not expected from the preliminary data provided by the shipping company. The interface allowed rapid reconfiguration of the audio content to geographical area that had not been considered in the initial designs.

3.2 Artist On Board

Having an artist as an integral member of the team who was not fully aware of the underlying technical constraints of the system helped expose design constraints that would not have been found if this project was focussed only on engineering deliverables. As an example, we only ever anticipated 20-30 audio files being associated with one audio region. However, since we created an audio effect that could randomly play files from a selection of files it seemed quite plausible to place an entire collection of 400 audio files into a region and copy and paste multiple copies of the sound region where it was required. Whilst we did not hard code a maximum value, the additional time to access the data structure exposed subtle timing problems elsewhere in the application that indicated the need for threading.

As part of the VectorMap interface colour and opacity were used to help identify sound regions. In the final version of the piece, the artist had used colour in a highly artistic way producing graduations of colour to also show the progress along the journey. Throughout the exhibition, other artists commented positively on the visual aesthetics of the display. Hence, providing a simple feature like colour enabled another level of creativity and expression.

Artists, and in essence real users, push the boundaries of an application, which can produce a positive benefit to the overall outcomes of a project.

3.3 Tools not artworks

Another impression gained from art-centric conference attendees was the difference between tool design and desire to build an exhibit. In many cases, the art community follows the latter and purpose-built exhibits are created for display, literally as one-off productions. In the engineering community, tools and systems may be created but their applicability to artistic applications is often limited since the motivations for use are quite different. In Syren, the aim was to develop reusable tools that can be applied to a range of spatial audio and location-aware artworks and demonstrators.

3.4 Location-based Databases

Whilst the project has developed a custom database for storing the audio regions and references to audio media, the approach indicates what might be needed from future location-based services that are media focussed. From this project it is recognised that the designer may want to define arbitrary shaped regions for audio content not be limited to the basic shape of a circle. The database must support proximity queries allowing hit testing of a variety of regions and if real-time audio presentation is required, this process must respond quickly. A ship based project will not need to be as responsive as one in which the user is mobile and listening over headphones. Another issue for portable handheld devices is that of caching nearby regions and making predictions about what content might be required in the short term. Our current work has not focussed on this issue but it is identified as an important consideration for small, mobile memory devices that download media from a network system.

3.4 Future Directions

Our current course of investigation is to take what has been learned in deploying Syren to understand: 1) how the user interface design tools can be applied to a broader range of contexts such as campus guides and other forms of artworks; 2) the issues of platform scalability when considering handheld devices; and 3) how more robust database backend technologies can be applied to enable multi-user and collaborative applications.

Whilst the first version of the mapping interface is far from optimal, observation of the system in use by an artist has indicated plenty of opportunities for exploring how to solve navigation and searching issues that are related to the use of vector-based maps. Essentially, the editing system is a front-end interface to a GIS allowing the definition and interpretation of location-based information. In future work, this user interface will provide a test bed application to investigate navigation, search and usability issues that are relevant to all location-based information systems.

The map-based interface provides fundamental infrastructure to support a wide range of audio experiences that are location aware. Once the dependency on the physical speaker matrix is removed and a mobile solution is created, we can design and deliver audio to the device within any definable region where position can be sensed. The software tools have the potential to create pedestrian and vehicle audio experiences. The outcomes could include new forms of artworks, navigation, tourist guides and more generally, the contextual delivery of media depending on where people are currently located.

4 Conclusions

Syren is the first project milestone for the team and much of what we have learned from the Baltic deployment will be incorporated into the next iteration of our tool base. The ocean setting of the ISEA2004 conference provided a unique opportunity to undertake an ambitious, location-based spatial sound installation, which covered a journey of almost 1000km in around 41 hours. In comparison to other location-based artworks, Syren is vehicle- rather than pedestrian-based, covering a larger area than city-based pieces. The artwork is also an example of audio AR in that it uses spatialised audio to augment the environment. It did not enable audience input, as it did not support annotation of the piece with additional audio whilst in transit.

User-centred design helped focus our attention to the needs of the audio artist and assisted in the development of reusable tools, rather than a one-off exhibit. Artist involvement early in the development phase and throughout was a key factor to produce a highly productive editing environment. The approach that artists take when exploring tools can be highly creative and lateral in ways that engineers would not have readily foreseen, hence pushing the boundaries of possibility. Designers of augmented audio reality systems need to be aware of the needs of creative users who will be significant stakeholders in the future of location-based services.

Acknowledgements

This project is supported by an Australian Research Council Linkage Project in conjunction with the New Media Board of the Australia Council as part of the Synapse Initiative.

We wish to acknowledge the people and organisations whose contribution proved to be a valuable part of the Syren Project. ISEA Conference Staff and Technical Crew, Apple Computer Europe, Apple Developer Relations, Australia, Finnish Maritime Authority, Swedish Maritime Authority and Estonian Maritime Authority, Peter Mumford, Joel Barnes (SNAP Lab), Doug Kinlyside (Department of Lands), Natasha Leow, Bill Pepper (NIDA) and students from the Voice Studies course (2004) – Alistar, Emma, Geoffrey, Linda, Loretta and Tetsuya. Anthony Knittel and Ben Hudson.

References

Azuma R. (1997): ***A survey of augmented reality*** In Presence: Teleoperators and Virtual Environments, Vol. 6, No. 4 (August 1997), pp. 355-385.

Black A. (2003): ***readanysf~*** http://aug.ment.org/readanysf/, accessed May 2004

Cohen M. (1994): ***Augmented Audio Reality: Design for a Spatial Sound GPS PGS***. Center on Disabilities Virtual Reality Conference, 1994. California State University, Northridgehttp://www.csun.edu/cod/conf/1994/proceedings/Aar~1.htm, accessed November 13, 2004

Cohen M. (2002): ***A Survey of Emerging and Exotic Auditory Interfaces,*** Proceedings of the 2002 International Conference on Auditory Display, Kyoto , Japan, 02-1 – 02-16.

Fujihata M. (1992): ***Field-Works projects website***, http://www.field-works.net/, accessed November 13, 2004

Helyer N. (2003): ***Sonic Landscapes***, http://www. sonicobjects.com/sonic2/projectarchive/SONICLAND/ SONICLAND.HTM, accessed November 13, 2004

International Hydrographic Bureau (2000): ***IHO Transfer Standard For Digital Hydrographic Data***, Edition 3.1 – Monaco, November 2000

Knowlton J.; Spellman N.; Hight J. (2002): ***34 North, 118 West,*** http://34n118w.net/, accessed November 13, 2004 from

Kurgan L. (1995): ***You Are Here: Museu*** http://www.princeton.edu/%7Ekurgan/urhere/html/intro.htm, accessed November 13, 2004

Mott I.; Sosnin J. (1997): ***Sound Mapping: an assertion of place***. Proceedings of Interface 1997. http://www.reverberant.com/SM/paper.htm, accessed November 15, 2004

Perry B. (2002): ***The Invisible Ideas Project: Using Mobile Devices, Macromedia Flash, and Global Positioning Systems***. Macromedia Developer Center article, http://www.macromedia.com/devnet/devices/articles/invisible_ideas.html, accessed November 13, 2004

Puckette M. (1996): ***Pure Data***, Proceedings, International Computer Music Conference. San Francisco: International Computer Music Association, 269-272.

Pulkki V. (1997): ***Virtual sound source positioning using vector base amplitude panning***, Journal of the Audio Engineering Society, 45(6), 456-466.

Roads C. (2001): ***Microsound***. Cambridge, Massachusetts: MIT Press.

Rosellini M. (2003): ***Four Coordinates Framework***, http://freshmeat.net/projects/fourcoordinates/, accessed November 13, 2004

Rozier J.; Karahalios K.; Donath J. (2000): ***Hear & There: An Augmented Reality System of Linked Audio***. http://citeseer.ist.psu.edu/rozier00hearthere.html, accessed October 29, 2004

Warmerdam F. (2004): ***ISO8211Lib,*** http://gdal.velocet.ca/projects/iso8211/, accessed November 13, 2004

Wright M.; Freed A. (1997): Open Sound Control: A New Protocol for Communicating with Sound Synthesizers. ICMC 1997

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 48-55

Journal of Global Positioning Systems
2005
© CPGPS

Efficient RTK Positioning by Integrating Virtual Reference Stations with WCDMA Network

Dinesh Kumar-Mills

School of Information Technology & Electrical Engineering, University of Queensland, Australia
Tel: (+617) 3365 2875; Email: dinesh@itee.uq.edu.au

John Homer

School of Information Technology & Electrical Engineering, University of Queensland, Australia
Tel: (+617) 3365 4139; Fax: (+617) 3365 4999; Email: homerj@itee.uq.edu.au

Kurt Kubik

School of Information Technology & Electrical Engineering, University of Queensland, Australia
Tel: (+617) 3365 8328; Email: kubik@itee.uq.edu.au

Matt Higgins

Department of Natural Resources and Mines, Locked Bag 40, Coorparoo DC, Queensland, Australia
Tel: (+617) 3896 3754; Fax: (+617) 3891 5168; Email: matt.higgins@nrme.qld.gov.au

Received: 6 December 2004/ Accepted: 26 October 2005

Abstract. The most advanced and proven implementation of the networked RTK is the VRS network concept. Its requirement of bi-directional communications is a critical disadvantage as this limits the robustness of the system. High costs and coverage limitations are also associated with the types of technology (i.e. UHF, GSM and GPRS) required for VRS communications. The Virtual Reference Cell (VRC) approach can be used to mitigate the disadvantages of the VRS network. This approach generates corrections for a fixed number of cells that are broadcast to the rovers. The drawback of the VRC system is a lower positioning accuracy due to the use of DGPS corrections instead of RTK. This paper proposes an RTK-VRC system whereby advantages of the VRC are maintained while achieving RTK level accuracy, mitigating high communications costs and supporting kinematic applications. The RTK-VRC system is an integration of the VRS network, to provide RTK positioning, and the WCDMA wireless network, to provide the cell structure and communications. For this system a novel communications link will be implemented using the pilot channel of the WCDMA network to minimise the communications costs. The results of a field experiment that utilises the NR&M VRS network in Australia shows that RTK positioning accuracy is achievable for VRS baselines of up to 2 km. This supports the idea of using the WCDMA cells with the RTK-VRC system.

Keywords: GPS, Network RTK, VRS, WCDMA, VRC

1 Introduction

The use of multiple reference station networks with real time kinematic (RTK) positioning provides a high precision, centimetre level, satellite positioning service that is extremely reliable and accessible (Retscher, 2002, Fotopoulos and Cannon, 2001). This technology has gained wide acceptance in the geodetic, engineering, earthmoving and public works communities. One proven implementation of the networked RTK is the Virtual Reference Station (VRS) network concept. The VRS concept is widely accepted as the most advanced approach for increased spatial separation of reference stations and error modelling (Retscher, 2002). This system calculates network corrections for systematic errors based on real-time data from all reference stations and simulates a local reference station (or VRS) near a

GPS rover station (Vollath et al., 2002, Vollath et al., 2001). Corrections for the VRS are transmitted through a communications link. This approach eliminates the need for actual reference stations on site as VRS data can be generated for any location within the network coverage area. The reduction in systematic errors allows increased spatial separation between the reference stations while increasing the reliability of the system and reducing the initialisation time (Retscher, 2002).

The underlying requirement for any successful RTK operation, including the VRS network, is the ability to communicate timely and reliable reference station corrections to the rover (Liu, 2004). The VRS network requires a bi-directional communication or data link between the rover and the control centre. To create the VRS data for a rover, its approximate location is initially transmitted uplink to the VRS network control centre. The control centre then generates corrections for that approximate location and transmits downlink to the rover. The downlink corrections are updated at a frequency of 1Hz (Landau et al., 2003b). These corrections are then employed with standard RTK GPS algorithms to obtain a precise position fix. For the rover, the data transmitted uplink is independent of the processing of the downlink corrections.

Retscher (2002) and Zhang and Roberts (2003) assert that the requirement of bi-directional communications is a critical disadvantage of the VRS concept. The generation of corrections specific for each rover limits the robustness of the system; especially when there are a large number of rovers (Retscher, 2002, Zhang and Roberts, 2003); (Petrovski et al., 2001). High costs (Zhang and Roberts, 2003) and coverage limitations are also associated with the type of technology required for VRS communications. The communications technology currently being employed are privately owned UHF radio networks, Global System for Mobile Communications (GSM) networks and General Packet Radio Service (GPRS) networks. The communications technology that is used must be carefully selected due to their inherent limitations. The UHF radio network has a limited coverage area and there are large costs associated with spectrum licensing. The GSM network was primarily designed for voice traffic and does not meet the affordability required for VRS network data transmission. GPRS on the other hand, has been introduced recently and is an affordable data transmission alternative to GSM. However for users that require continuous correction updates over long intervals the GPRS technology is still quite costly.

Another disadvantage of the VRS approach is that it poses a problem for kinematic applications with rovers moving over large network areas (Landau et al., 2003a). Distance dependant errors will be present in the solution as the rover moves away from the VRS (Wübbena et al., 2001).

According to Retscher (2002) the limitations associated with system robustness and the bi-directional communications can be mitigated by using the Virtual Reference Cell (VRC) approach. This is similar to the 'gridded corrections' concept discussed by Wanninger (2002). The VRC approach uses the VRS network to estimate correction models for a cell or gridded DGPS service area. Rovers within a particular VRC are assigned corrections associated with that cell. When the rover leaves a VRC it is assigned to another VRC. The tracking of the rover is thus achieved. This approach eliminates the need for bi-directional communications and a broadcast approach is sufficient. The limitation on the number of users is also eliminated as the corrections need to be calculated for only a fixed number of cells. The drawback of this system however, according to Retscher (2002), is a generally lower positioning accuracy than the VRS. This is due to the use of DGPS corrections and algorithms instead of RTK to obtain a position fix. It is also due to distance dependant errors that become more evident for rovers that are further away from the centre of the VRC. In order to achieve RTK level accuracy using the VRC approach, high communications costs will still be incurred.

In this paper an RTK-VRC system is proposed whereby the advantages of the VRC are maintained while achieving RTK level positioning accuracy. The RTK-VRC system will also mitigate the high communication costs and support kinematic applications.

The basis of this new system will be the integration of the VRS network with the Wideband Code Division Multiple Access (WCDMA) wireless network and infrastructure. This involves using the VRS network to provide RTK positioning and the WCDMA network to provide the cell structure and communications. The radius for WCDMA network cells ranges from 500m to 2 km (Norgaard, 2003). This small radius is ideal for the proposed system. The VRS corrections are generated by the VRS network control centre specific for each WCDMA base station within the VRS coverage area. The corrections are then streamed to the WCDMA base station and then communicated to all rovers within the coverage cell.

The communication of corrections to the rover is achieved through a novel approach that uses the Pilot Channel (CPICH) on the physical layer of the WCDMA base station (3GPP-R1, 2003). This approach hides the communication link within the WCDMA network and thus minimises the costs associated with the data transfer. The total coverage of this system is large and only limited by the VRS network coverage and the WCDMA network coverage.

On the user's end, each rover will be attached to a WCDMA wireless device. This device will be able to distinguish between coverage cells and will be used to receive corrections from the WCDMA base station. These corrections are used by the rover in the same manner as the VRS concept to obtain a precise position fix. When the rover moves from one cell to the other, the VRS corrections for the new cell are automatically adopted. This process also supports kinematic applications as the VRS baselines will not be more than the radius of a cell, thus limiting the distance dependant errors.

The objective of this paper is to demonstrate that RTK positioning accuracy is achievable for VRS baselines of up to 2 km. This supports the proposal of using the WCDMA network cell as the VRC. In the following sections background information on the WCDMA technology is presented followed by an outline of the RTK-VRC system. This will be followed by background information and discussion on field tests that were performed using an existing VRS network. These field tests show that RTK level accuracy is achievable with VRS baselines of up to 2 km. The paper then ends with some concluding remarks.

2 WCDMA

Wideband Code Division Multiple Access (WCDMA) is a third generation (3G) wireless communication system. The 3G systems extend the capabilities of the second generation (2G) wireless systems (only voice and low rate data) to include multimedia capabilities with high bit rates and packet data. WCDMA is based on the robust and well proven code division multiple access (CDMA) technology.

The deployment of WCDMA networks is gradually increasing throughout the world. The WCDMA standards have been designed to naturally supersede the aging 2G GSM networks. Consequently, the widespread use of WCDMA in the near future is imminent.

WCDMA uses direct sequence spread spectrum (Rappaport, 2002) on a 5 MHz bandwidth and operates in the frequency division duplex (FDD) mode and the time division duplex (TDD) mode. This paper will focus on the FDD mode. The WCDMA system features are detailed by Karim and Sarraf (2002) and Rappaport (2002). Some of the main features are listed in Table 1.

Table 1 Main Features of the WCDMA FDD Physical Layer

Channel Bandwidth	5 MHz
Chip Rate	3.84 Mcps
Frame Length	10 ms
No. of slots/frame	15
No. of chips/slot	2560 chips (Max 2560 bits)
Uplink Spreading Factor	4 – 256
Downlink Spreading Factor	4 – 512
Channel Rate	7.5 kbps – 960 kbps

The WCDMA system uses a layered protocol architecture at different interface points, each layer performing a set of specific functions (Karim and Sarraf, 2002). The three layers are the physical layer (L1), the data link layer (L2) and the network layer (L3) (3GPP-R1, 2003). The interrelationship between these layers is described by Wesolowski (2002). The purpose of the physical layer is to condition the digital data from higher layers so that it can be transmitted over a mobile radio channel reliably (Karim and Sarraf, 2002). This conditioning implements signal processing functions, channel coding, interleaving, modulation, spreading and synchronisation to user data (or signalling data). The signal conditioning is performed as part of the process of mapping data received from the higher layers through the transport channels to a physical channel.

Some of the specifications of the physical channels such as the chip rate, frame length, time slots per frame and chips per slot are given in Table 1. The types of physical channels available differ between the uplinks and downlinks. These channels are listed in Table 2 and Table 3 respectively. The functions of these channels are described by Karim and Sarraf (2002).

Table 2 Uplink Physical Channels

Dedicated Physical Data Channel	DPDCH
Dedicated Physical Control Channel	DPCCH
Physical Random Access Channel	PRACH
Physical Common Packet Channel	PCPCH

Table 3 Downlink Physical Channels

Dedicated Physical Channel	DPCH
Common Pilot Channel (Primary & Secondary)	CPICH
Common Control Physical Channel	CCPCH
Synchronisation Channel (Primary & Secondary)	SCH
Acquisition Indicator Channel	AIC

The spreading applied to the physical channels consists of two operations in succession – first the channelisation operation followed by the scrambling operation (3GPP-R1, 2001). The channelisation operation transforms every data symbol into a number of chips, increasing the bandwidth of the signal. The scrambling operation then applies a scrambling code to the spread signal. These successive processes spread the signal energy over a large bandwidth (Yang et al., 2000).

The channelisation operation utilises orthogonal channelisation codes. The codes used for WCDMA are variable-length Walsh codes, also known as Orthogonal Variable Spreading Factor (OVSF) codes (Karim and Sarraf, 2002). Orthogonality here implies that different codes, within a family of codes, are mutually orthogonal. The cross-correlation values between these different codes are zero and therefore there is no interference between channels. This concept allows these orthogonal codes to be used as signature codes to distinguish between the signals from different channels or users (Yang et al., 2000). In the uplink the OVSF codes are used to distinguish between channels transmitted by a user. In the downlink these codes are used to distinguish between users within a WCDMA cell.

The scrambling operation utilises pseudonoise (PN) scrambling codes (3GPP-R1, 2001). The aim of the scrambling is to differentiate between cells and thus reduce inter cell interference. WCDMA base stations are each assigned a scrambling code in the cell planning process. WCDMA wireless devices or mobile stations (MS) must know these scrambling codes in order to synchronise and recover data from the WCDMA base station. The scrambling operation and code generation are further described by Wesolowski (2002). The scrambling operation, together with suitable receiver processing is also important for reducing multipath induced distortion.

WCDMA base stations are called Nodes B and perform the physical layer processing (Wesolowski, 2002). The current cell radii range from about 500 m to 2 km (Norgaard, 2003) for micro and macro cells. As the uptake to WCDMA increases, the cell radii will naturally be reduced to handle the increased load. The identifiable nature of the WCDMA cell and its relatively small cell radius makes it well suited to the RTK-VRC system. The ability to differentiate between cells enables MSs to track their movements from one cell to another. This means that if the Nodes B are used to broadcast RTK corrections for the cell, the rovers (that are connected to the MSs) will be able to distinguish between corrections from one Node B to another. The small cell radius implies that the maximum distance between the rover and the Node B will be about 2 km. This will ensure that the inherent distance dependant errors are relatively low for rovers that use the broadcasted RTK corrections.

The primary Common Pilot Channel (CPICH) for the downlink is the basis of the novel communications link for the RTK-VRC system. The primary CPICH is broadcast over the entire cell and is always present. It is currently used for channel estimation. The channel estimate, in conjunction with the Rake receiver (Rappaport, 2002), is used to recover transmitted signals that have been distorted by the multipath phenomenon. The primary CPICH is also transmitted at a higher power level than the other physical channels. The power of the primary CPICH determines the cell coverage area and the capacity of a cell. This is because the handover of the MS from one cell to another is triggered by a measurement of the primary CPICH strength. Increasing or decreasing the power of the primary CPICH makes the cell larger or smaller (Valkealahti et al., 2002). The channelisation code for the primary CPICH is also currently fixed for all Nodes B (3GPP-R1, 2001).

The proposed communications link takes advantage of the fact that there is currently no logic associated with the primary CPICH; as it is not mapped to channels in the higher layers (i.e. L2 and L3). By modifying the layered protocol architecture, broadcast data will be mapped to the primary CPICH. However, in order to continue its use for channel estimation and cell size determination, and at the same time use it to broadcast data a new encoding method will have to be implemented. This will be achieved through the dynamic use of channelisation codes for the primary CPICH. Each of the channelisation codes that are used will be associated with a specific code word and a conversion table will be used to store these associations. The sequence of channelisation codes will then be used to generate a stream of data to the MS. The MS will also need to be modified to be able to process the dynamic channelisation codes. This includes having the correct conversion table. Without the correct conversion table the MS will not be able to correctly decode the data stream. This process will act like an encryption tool for the communications link.

The disadvantage of this method is that the channelisation codes that will be used with the proposed communications link cannot be reused for any other channel. This restriction combined with the fact that there are a fixed number of OVSF codes available, limits the feasibility of such a method. To overcome this limitation, Quasi-Orthogonal Sequences (QOS) will be introduced as additional channelisation codes. QOS are used as a means of increasing the number of physical channels that are available (Yang et al., 2000). These QOSs will be used in conjunction with the OVSF codes.

QOSs are codes that have low cross-correlation with the OVSF codes. The QOSs provide additional capacity at the expense of some channel interference. This constant and minimal interference is generally outweighed by the operational needs of the WCDMA network. According to Jalloul and Shanbhag (2002) the QOS can be generated by applying a mask to the existing family of OVSF codes.

For the proposed communications link the QOSs will be used as the channelisation codes for the primary CPICH. This means that only one QOS will be used at any period. In this manner, the level of channel interference generated from the QOSs will be minimised.

Applying this simple encoding method in conjunction with the implementation of the QOSs allows the novel communications link to be generated. This link requires the Node B and the MS to have the same conversion table to decode the data stream. This communications link is unavailable to MSs that do not have the correct conversion table. The communications link will be capable of broadcasting at a rate of 12 Kbits per second. This bit rate exceeds the required transmission rate for RTK communications of about 9.6 Kbits per second.

3 RTK-VRC

The proposed RTK-VRC system will exploit the cellular infrastructure and primary CPICH of the WCDMA network to provide a satellite positioning system that will be capable of achieving RTK level accuracy. The application of the RTK-VRC system will consist of three main areas – the correction streaming (from the VRS network control centre), the broadcasting of corrections (by the WCDMA base station or Node B) and the receiving of corrections (by the GPS rover station).

Correction data for the RTK-VRC system will be generated based on the location of each WCDMA cell. The correction data will be streamed from the VRS network control centre to the Node B. This streaming will be achieved using the Networked Transport of RTCM via Internet Protocol (NTRIP) technology. This technology is currently available with some VRS networks. The NTRIP technology is discussed by Lenz (2004).

The Node B then encodes the data using the QOSs. These QOSs are used as channelisation codes for the primary CPICH and broadcast to the WCDMA cell. The MSs within this cell are then responsible for receiving this broadcast and decoding it back to a stream of RTK correction data. This stream of data will then be fed to the rover that is attached to the MS. The rover will then use these corrections to help it obtain a precise position fix.

4 Field Experiments using NR&M VRS Network

In order to demonstrate the availability of RTK level positioning for VRS baselines of up to 2 km an experiment was carried out using the VRS network belonging to the Department of Natural Resources and Mines (NR&M). As the accuracy level plays a crucial role in high precision RTK applications the data collected during the experiment focussed on the correction availability and positioning accuracy.

4.1 NR&M VRS Network

The Department of Natural Resources and Mines (NR&M) is responsible for the surveying and geodetic infrastructure within the state of Queensland, Australia. The NR&M is a department of the State Government. It plays a critical role in the stewardship of natural resources. It also manages and allocates the State's land, water, mineral and petroleum resources, and manages native vegetation and the control of pest plants and animals (NRME, 2004).

In 2000 NR&M, in partnership with Trimble Australia and Ultimate Positioning, established a pilot VRS network. The location of the VRS network is shown in Figure 1. The pilot network was designed to investigate the viability of the VRS concept as a future element of the surveying and geodetic infrastructure in Queensland. The goals for the pilot network and the preliminary test results are set out by Higgins (2001). The VRS network is now setup as a production system. The main service provided by the NR&M VRS network is high precision positioning for surveying and earthmoving equipment.

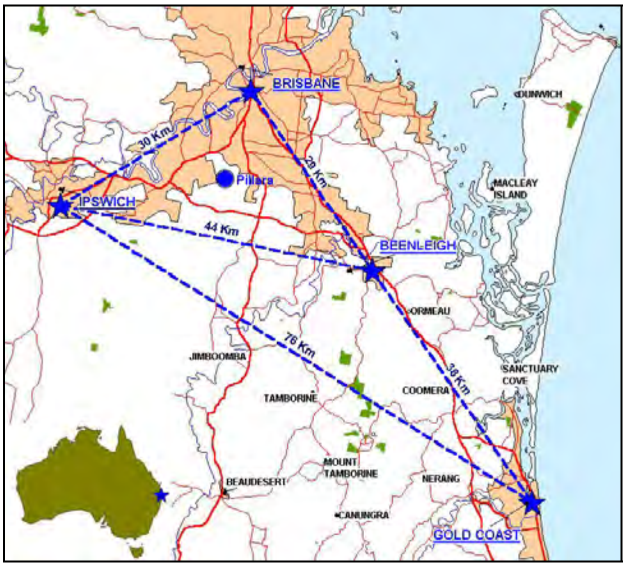


Figure 1 Location of VRS Base Stations

The Gold Coast, Ipswich and Brisbane VRS base stations currently utilise a Trimble 4700 receiver with a choke ring antenna. The Beenleigh VRS base station utilises a Trimble 5700 receiver with a Zephyr Geodetic antenna. The VRS base stations are all linked by a high-speed Wide Area Network (WAN). The VRS network control centre is situated in Brisbane and utilises the Trimble GPSNet VRS software to provide the VRS service. The VRS network is capable of utilising GSM, UHF and GPRS to establish a communication link.

The RTK rover consisted of a Trimble R8 receiver, a Trimble TSCE handheld controller and a standard data enabled GSM mobile phone. The GSM phone was used to establish the communications link between the rover and the VRS control centre.

4.2 Field Experiment

The field experiment was carried out on the 7th of October 2004 commencing at 2:00 pm and concluded at 4:00 pm. The tests were carried out using three markers along Ritchie Road and Gooderham Road in Pillara, Queensland (see Figure 2). Pillara is located at the centre of the VRS network coverage area as shown in Figure 1. The centre of the VRS network is generally the area where the VRS performance is lowest. This means that the worst case scenario was tested. Pillara is a flat semi-urban farming region. The three markers are located in relatively clean multipath environments. The distance from markers A to B is 1.3 km and the distance from markers A to C is 2.0 km. The locations of the surveyed markers are shown in Figure 2. The coordinates of the markers are given in Table 4.

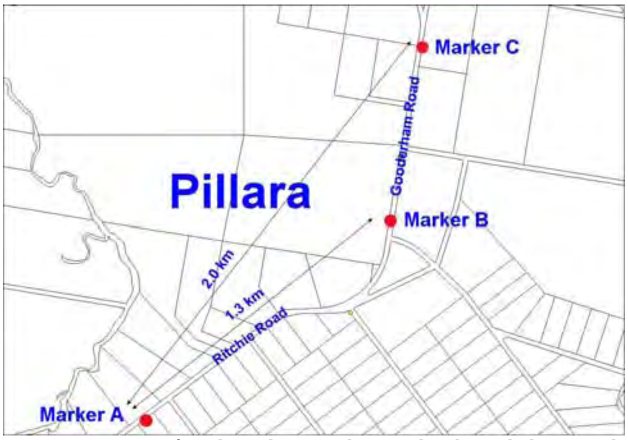


Figure 2 Position of markers along Ritchie Road and Gooderham Road in Pillara

Table 4 Surveyed Markers

Marker	Survey Code	Location (m)		
		Easting	Northing	Height
A	28113	499899.442	6945004.486	34.399
B	Star Picket 1	500938.476	6945844.184	17.377
C	71400	501072.328	6946576.259	13.372

For the first run, the VRS was created at the position of marker A. Then sixty solved position measurements were recorded at a rate of 1 Hz for marker A. Using the same VRS, sixty position measurements were also recorded at markers B and C respectively. This process was repeated twice to establish the repeatability of the recorded results.

The correction availability is an important performance indicator for VRS RTK positioning (Hu et al., 2003). During the time that the measurements were recorded there were at least six satellites with corrections. This is an important threshold for the implementation of RTK positioning (Edwards et al., 1999).

The accuracy and precision of the results was determined by comparing the differences between the three coordinate components of the solved positions and the mean of the positions. The horizontal position scatter plots for the three markers are displayed in

Figure 3, and respectively. The standard deviations for the northing, easting and height component for marker A were 4.7, 3.9 and 14.0 mm respectively. The distribution of deviation from the mean position at marker A for the three runs were graphed and found to follow a Gaussian distribution. This indicates that there were no systematic errors in the recorded data. The position accuracies at a 95% confidence level for the northing, easting and height components were 18.3, 15.4 and 54.9 mm respectively.

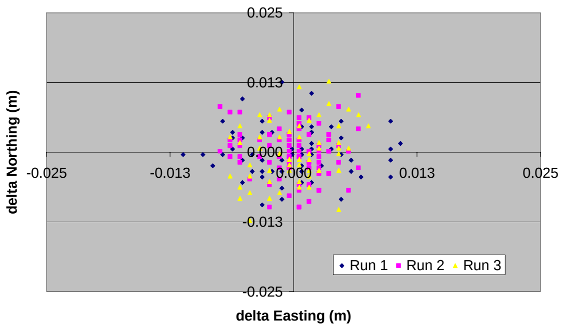


Figure 3 Horizontal Position Plot at Marker A (0 km away from VRS)

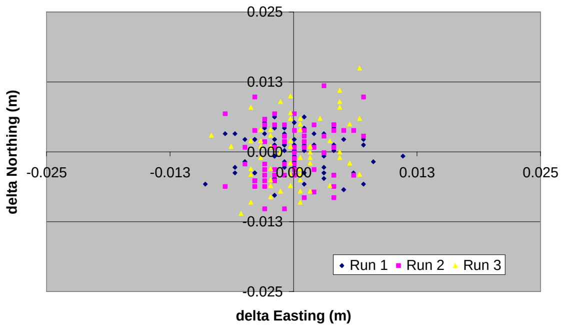


Figure 4 Horizontal Position Plot at Marker B (1.3 km away from VRS)

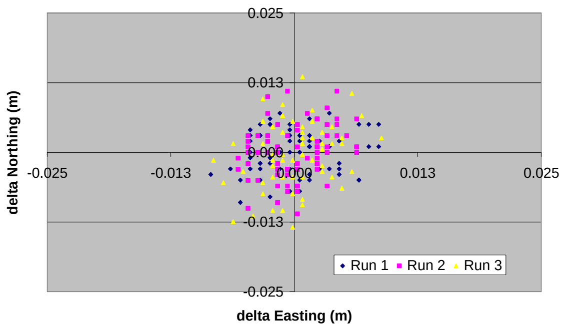


Figure 5 Horizontal Position Plot at Marker C (2.0 km away from VRS)

For marker B the standard deviations for the northing, easting and height components are 4.7, 3.5 and 13.9 mm respectively. The distribution of deviation from the mean position for the three runs were graphed and found to follow a Gaussian distribution. The position accuracies at a 95% confidence level for the northing, easting and height components were 18.5, 13.8 and 54.4 mm respectively.

For marker C the standard deviations for the northing, easting and height component are 5.0, 3.5 and 13.6 mm respectively. The distribution of the deviation from the mean position for the three runs were graphed and found to follow a Gaussian distribution. The position accuracies at a 95% confidence level for the northing, easting and height components were 19.5, 13.6 and 53.3 mm respectively.

A summary of the statistics for the three markers is given in Table 5. The results indicate that the accuracy of the horizontal position for all three markers is generally better than 2 cm and the height accuracy is better than 6 cm. The variation in accuracy between marker A and the other two markers, B and C, is almost negligible. This indicates that for VRS baselines of up to 2 km, RTK level (i.e. centimetre level) accuracy is achievable using the VRS network. This supports the use of the WCDMA base-stations with cell radii of up to 2 km for the RTK-VRC system.

Table 5 Statistical Results for Markers A, B & C in millimetres (where N, E & H are the Northing, Easting and Height)

Marker	Standard Deviation			Confidence Level 95%		
	N	E	H	N	E	H
A	4.7	3.9	14.0	18.3	15.4	54.9
B	4.7	3.5	13.9	18.5	13.8	54.4
C	5.0	3.5	13.6	19.5	13.6	53.3

5 Concluding Remarks

This paper provides a basic outline of the RTK-VRC system. The proposed system will integrate the VRS network with the WCDMA network and infrastructure and also mitigate the high communications costs (associated with current VRS systems). This system will provide RTK level (or centimetre level) positioning accuracy.

The field tests have demonstrated that RTK positioning accuracy is achievable for VRS baselines of up to 2 km. This is the typically maximum radius of the WCDMA cell supporting the proposal of using the WCDMA network cell as the VRC.

References

3GPP-R1 (2001): ***Technical specification group radio access network; spreading and modulation (FDD) (Release 4)***. 3rd Generation Partnership Project (www.3gpp.org)

3GPP-R1 (2003): ***Technical specification group radio access network; physical layer - general description (Release 6)***. 3rd Generation Partnership Project (www.3gpp.org)

Edwards S.J.; Cross P.A.; Barnes J.B.; Betaille D. (1999): ***A methodology for benchmarking Real Time Kinematic GPS.*** Survey Review, 35(273): 163-174

Fotopoulos G.; Cannon M.E. (2001): ***An overview of multi-reference station methods for cm-level positioning.*** GPS Solutions, 4(3): 1-10

Higgins M.B. (2001): ***An Australian pilot project for a real time kinematic GPS network using the virtual reference station concept.*** New Technology for a New Century International Conference, Seoul, South Korea, May

Hu G.R.; Khoo H.S.; Goh P.C.; Law C.L. (2003): ***Development and assessment of GPS virtual reference stations for RTK positioning.*** Journal of Geodesy, 77(5-6): 292-302

Jalloul L.M.A.; Shanbhag A. (2002): ***Enhancing data throughput using quasi-orthogonal functions aggregation for 3G CDMA systems.*** 55th Vehicular Technology Conference, May 6-9 2002, Birmingham, AL, United States, 2008-2012

Karim M.R.; Sarraf M. (2002): ***W-CDMA and cdma2000 for 3G Mobile Networks***. McGraw-Hill, New York, 211

Landau H.; Vollath U.; Chen X. (2003a): ***Virtual reference stations versus broadcast solutions in network RTK - Advantages and Limitations.*** GNSS, Graz, Austria, Apr.

Landau H.; Vollath U.; Deking A.; Pagels C. (2003b): ***Virtual reference station networks - recent innovations by Trimble***. Trimble, Terrasat GmbH

Lenz E. (2004): ***Networked transport of RTCM via internet protocol (NTRIP) application and benefit in modern surveying systems.*** FIG Working Week 2004, Athens, Greece, 22-27 May

Liu G.C. (2004): ***GPS RTK positioning via internet-based 3G CDMA2000/1X wireless technology.*** GPS Solutions, 7(4): 222-229

Norgaard J. (2003): ***Utilizing the inherent advantages of lower frequency bands for advanced communications systems.*** KKRRIT, Poland, 26 June

NRME. (2004): ***http://www.nrme.qld.gov.au/about/index.html***. The State of Queensland (Department of Natural Resources, Mines and Energy)

Petrovski I et al. (2001): ***Practical Issues of Virtual Reference Station implementation for nationwide RTK network.*** GNSS International Symposium, Seville, Spain, May

Rappaport T.S. (2002): ***Wireless Communications Principles and Practices.*** Prentice Hall Communications Engineering and Emerging Technologies Series. Prentice Hall, Upper Saddle River

Retscher G. (2002): ***Accuracy performance of virtual reference station (VRS) networks.*** Journal of Global Positioning Systems, 1(1): 40-47

Valkealahti K.; Hoglund A.; Parkkinen J.; Hamalainen A. (2002): ***WCDMA common pilot power control for load and coverage balancing.*** Proceedings of 13th IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC 20020, 15-18 Sept. 2002, Pavilhao Altantico, Lisboa, Portugal, 2002, 1412-16

Vollath U.; Deking A.; Landau H.; Pagels C. (2001): ***Long-range RTK positioning using virtual reference station.*** KIS-2001

Vollath U.; Landau H.; Chen X. (2002): ***Network RTK - concept and performance***. Trimble (www.trimble.com)

Wanninger L. (2002): ***Virtual Reference Stations for centimeter-level kinematic positioning.*** ION GPS, Portland, Oregon, Sept., 1400-1407

Wesolowski K. (2002): ***Mobile Communication Systems***. John Wiley & Sons, West Sussex

Wübbena G.; Bagge A.; Schmitz M. (2001): ***Network-based techniques for RTK Applications.*** GPS Symposium, Tokyo, Japan, 14 Nov.

Yang K.; Kim Y-K.; Kumar P.V. (2000): ***Quasi-orthogonal sequences for code-division multiple-access systems.*** IEEE Transactions on Information Theory, 46(3): 982-993

Zhang K. and Roberts C. (2003): ***Network-based real-time kinematic positioning system: current development in Australia.*** Geoinformatics and Surveying Conference, Malaysia, 7 Apr.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 56-64

Journal of Global Positioning Systems
2005
© CPGPS

Time Estimation of Superimposed Coherent Multipath Signals Using the EM Algorithm for Global Positioning System

Francis Chan

The University of New South Wales, Sydney NSW 2052, Australia
email: francis@student.unsw.edu.au Tel: +(612) 9385 5250; Fax: +(612) 9385 5993

Jinho Choi

The University of New South Wales, Sydney NSW 2052, Australia
email: j.choi@unsw.edu.au

Gyu-In Jee

Konkuk University, Hwayang-Dong, Gwangjin-gu, Seoul, 143-701, Korea
email: gijee@konkuk.ac.kr

Received: 11 November 2004 / Accepted: 13 July 2005

Abstract. A novel multipath mitigation technique for Global Positioning System (GPS) receivers using the Expectation-Maximization (EM) algorithm is proposed. It is well-known that conventional propagation delay estimation using parallel sliding correlators is only optimal in additive white Gaussian noise channel. In practical positioning systems, the weak GPS line-of-sight signal is generally embedded in the multipath signals and other source of interference. Although the GPS direct sequence spread spectrum (DS-SS) signal has inherent resistance to interference, the received superimposed multipath signals, which are possibly coherent, are the dominant source of the propagation delay estimation errors. From the parameter estimation point of view, the problem of multipath mitigation is equivalent to estimating the unknown phases, propagation delays and amplitudes of the superimposed multipath signals. The joint maximum likelihood (ML) estimation of all the unknown parameters is optimal and asymptotically efficient. However it involves multi-dimensional search which is computationally expensive. The proposed coarse/acquisition (C/A) code acquisition system using the EM algorithm is an iterative maximum likelihood estimator which decomposes the multi-parameter estimation problem into a number of separate ML optimizations. The performance of the proposed EM algorithm has been tested by simulations. We have observed that the proposed acquisition system is significantly superior to the conventional correlating receiver in a multipath fading channel.

Keywords: Multipath Mitigation, Expectation-Maximization (EM) Algorithm, Time Estimation, Code Acquisition

1 Introduction

The focus of this paper is on multipath mitigation of the received superimposed signals in Global Positioning System (GPS). The operating principle of radio navigation systems is based on the propagation time estimation of the broadcast signals from the constellation satellites. In practical wireless channel, the transmitted signals are propagated along various reflected paths to the receive antenna. These replicas of the received signals are known as the multipath signals. For Global Navigation Satellite Systems, the prime information in concern is the exact propagation time of the direct signal from the satellite to the receiver. The reflected signals arrived at the antenna not only convey no information about for the pseudorange measurement, but they also induce errors for the geometric distance determination. Although the GPS direct sequence spread spectrum (DS-SS) signal has inherent resistance to interference, however, the received superimposed multipath signals, which are possibly coherent, are the dominant source of the propagation delay estimation errors.

In this paper, the problem of multipath cancellation is tackled in terms of the parameter estimation point of view. Specifically, by estimating all the unknown phases, amplitudes and the time delays of the reflected signals, the effect of the undesirable multipath signals can be minimized. It is well-known from the estimation theory that, the maximum likelihood (ML) estimators are optimal and asymptotically efficient. However, the cost function of the ML estimator is a nonlinear function of the unknown time delays of the reflected signals. In addition, the joint estimation of all the unknown parameters is a multi-dimensional minimization problem which is computationally expensive.

In the literature, the Multipath Estimating Delay Lock Loop (MEDLL) (van Nee 1992; van Nee *et al.* 1994) is a maximum likelihood estimator which is tailored to a multipath propagation environment. The weak power of the received GPS signals is highly susceptible to jamming signals and unintended interferences. The MEDLL is similar to a conventional correlator, except that it is optimized with respect to a multipath fading channel rather than a Gaussian noise channel. By estimating the amplitudes, delays and phases of all the D identified multipath signals simultaneously, the MEDLL can significantly reduce the measurement errors induced by the multipaths. The iterative ML estimator proposed in this paper is similar to the MEDLL in the sense that, it also has the multipath suppression capabilities.

A novel GPS receiver endowed with the multipath mitigation capabilities is proposed in this paper. Essentially, the proposed acquisition system decomposes the multi-parameter estimation problem into a number of separate ML optimizations, and hence, the computational cost is reduced. The propagation time of the line-of-sight signal and other unknown parameters are estimated in an iterative manner. The code acquisition system adopted the well-known Expectation-Maximization (EM) algorithm (Dempster *et al.*, 1977), which is an iterative maximum likelihood estimator. In Section 2, the theoretical outline of the EM algorithm is described. The proposed code acquisition architecture for the GPS using the EM algorithm is presented in Section 3. The simulation results are presented in Section 4. It is shown that the performance of the proposed receiver architecture is significantly improved in a multipath fading environment.

2 The Expectation-Maximization (EM) algorithm

The problem of GPS C/A code acquisition in a multipath fading environment using the Expectation-Maximization (EM) Algorithm is considered in this paper. The EM algorithm is an iterative computation routine of maximum likelihood estimates. The EM algorithm is closely related to estimation with missing data; it is particularly applicable to incomplete data problems. In the literature, the term complete data **X** generally refers to the parameter bearing data which lies on the sample space Ω. On the other hand, the term incomplete data **Y** refers to the sampled vector that lies on the observation space Φ. The complete data **X** depends on a set of unknown parameter θ, which lies on a measurable parameter space, denoted as Θ. Suppose that there exists a many-to-one transformation $\mathrm{T} : \Omega \to \Phi$. The parameter embedded data **X** is not directly observed, but instead, it is observed through the transformation T.

2.1 Outline of the EM algorithm

Let $L(Y;\theta)$ and $L_c(X;\theta)$ be the likelihood function of the observed (incomplete) data and the complete data, respectively. The maximum likelihood estimate $\hat{\theta}_{ML}$ is the parameter which maximizes $L_c(X;\theta)$, i.e.

$$\hat{\theta}_{ML} = \arg\max_{\theta} L_c(X;\theta) \tag{1}$$

However, the complete data **X** is not observable under the transformation T and its probability density function $p(X;\theta)$ is unknown, and hence the ML estimate $\hat{\theta}_{ML}$ in (1) cannot be evaluated directly. The approach taken by the EM algorithm is to evaluate the expected likelihood function $L_c(X;\theta)$ instead. Specifically, the estimate is obtained by maximizing the expected value of the complete data log likelihood function, given the observation and a preliminary estimate of the unknown parameter, which is denoted as θ'. Mathematically, the EM estimate $\hat{\theta}_{EM}$ is given by

$$\hat{\theta}_{EM} = \arg\max_{\theta} E\left[\log L_c(X;\theta) \mid Y, \theta'\right] \tag{2}$$

Due to the fact that the EM estimate $\hat{\theta}_{EM}$ is dependent on the preliminary estimate θ', hence $\hat{\theta}_{EM}$ can be considered as an improved estimate over θ'. Subsequently, the updated estimate can be used as a preliminary estimate and equation (2) can be evaluated repeatedly with an improved estimate at each iteration. The EM algorithm is therefore an iterated estimator, in the sense that, given an initial guess θ_0, the estimate $\hat{\theta}_{EM}$ can be obtained by evaluating (2) iteratively. Essentially, the EM algorithm involves two main steps,

for k = 1,2,…

E-Step

$$U(\theta;\hat{\theta}^{(k)}) = E\left[\log L_c(X;\theta) \mid Y,\hat{\theta}^{(k)}\right] \quad (3)$$

M-Step

$$\hat{\theta}^{(k+1)} = \arg\max_{\theta} U(\theta;\hat{\theta}^{(k)}) \quad (4)$$

Under normal conditions, the estimate $\hat{\theta}_{EM}$ will eventually converge to the maximizer of the complete data log likelihood function in equation (1). Although the convergence to the global maximizer is not guaranteed for multimodal cost functions, but if the initial guess is sufficiently close to the global maximum, in most cases, convergence to the global maximum can be achieved. The issue of convergence of the EM Algorithm will be discussed in the next section.

2.2 Convergence of the EM algorithm

In this section we will outline the proof which shows that the EM Algorithm estimate $\hat{\theta}_{EM}$ would converge to the maximum likelihood estimate $\hat{\theta}_{ML}$ as the number of iterations increases. For a more detailed discussion, the readers are recommended to refer to the seminal paper by Dempster et al. (Dempster *et al.*, 1977). The conditional density probability function is given by $p(X \mid Y;\theta) = p(X \mid \theta)/p(Y \mid \theta)$, the log likelihood of the incomplete data can be written as

$$\log L(Y;\theta) = \log L_c(X;\theta) - \log p(X \mid Y;\theta) \quad (5)$$

By taking the conditional expectation of equation (5) given the observation Y with a preliminary estimate of the unknown parameter θ', we have

$$\begin{aligned} &\log L(Y;\theta) \\ &= E\left[\log L_c(X;\theta) \mid Y,\theta'\right] - E\left[\log p(X \mid Y;\theta) \mid Y,\theta'\right] \\ &= U(\theta;\theta') - H(\theta;\theta') \end{aligned}$$

where

$$\begin{aligned} H(\theta;\theta') &= E\left[\log p(X \mid Y;\theta) \mid Y,\theta'\right] \\ U(\theta;\theta') &= E\left[\log L_c(X;\theta) \mid Y,\theta'\right] \end{aligned}$$

By performing the E-step and M-step iteratively, we obtain a sequence of estimates $\{\hat{\theta}_k\}$ (k=1,2,…). The difference of the log likelihood $L(Y;\theta)$ between two successive estimate can be written as

$$\begin{aligned} &\log L(Y;\theta^{(k+1)}) - \log L(Y;\theta^{(k)}) \\ &\quad = \left\{U(\theta^{(k+1)};\theta^{(k)}) - U(\theta^{(k)};\theta^{(k)})\right\} - \\ &\qquad \left\{H(\theta^{(k+1)};\theta^{(k)}) - H(\theta^{(k)};\theta^{(k)})\right\} \end{aligned} \quad (6)$$

The first bracket in (6) is the difference of conditional expectation of the complete data. It is chosen to be greater or equal to zero, this is essentially the M-Step of the EM algorithm as stated in (4). The second bracket in (6), on the other hand, is the difference of the conditional expectation of the conditional probability density. It can be shown by using the Jensen's inequality that the difference is less than or equal to zero. For any estimate θ',

$$\begin{aligned} &H(\theta';\theta^{(k)}) - H(\theta^{(k)};\theta^{(k)}) \\ &= E\left[\log\left\{p(X \mid Y;\theta')/p(X \mid Y;\theta^{(k)})\right\} \mid Y,\theta^{(k)}\right] \\ &\le \log\left[E\left\{p(X \mid Y;\theta')/p(X \mid Y;\theta^{(k)})\right\} \mid Y,\theta^{(k)}\right] \\ &= \log\int p(X \mid Y,\theta)dx \\ &= 0 \end{aligned}$$

The above integral is over the complete data **X** which lies on the range of the transformation mapping T . By combining the above results, we have shown that $\log L(Y;\theta^{(k+1)}) \ge \log L(Y;\theta^{(k)})$, hence the estimation sequence $\{\hat{\theta}_k\}$ progressively increases the log likelihood function. In other words, the EM estimate is approaching to the maximum likelihood estimate as the number of iteration increases.

3 GPS signal acquisition using the EM algorithm

The weak received GPS signal is vulnerable to interfering signals. Generally, the received signal is a superimposed of the distorted multipath signals, intentional or unintentional interference and channel noise. In particular, the close-in coherent multipath reflections of the direct signal are difficult to distinguish from the desired signal. It can cause serious contamination of the observable measurements. Conventional correlator-based detectors provide no protection from multipath and interference; these disturbances are collectively regarded as Gaussian channel noise in general. In a wireless multipath environment, the navigation accuracy can be degraded significantly if the detector fails to combat against the multipath reflections and interferences. There have been efforts devoted to improve the jamming immunity and apply special multipath rejection techniques at the GPS receivers, e.g. MEDLL (van Nee 1992; van Nee *et al.* 1994). The proposed C/A code acquisition system using the EM Algorithm will be described in this section.

3.1 System model

The standard GPS C/A code has a chip rate of *1/Tc* = 1.023MHz, one period of the spreading code consists of P = 1023 chips, hence it spans in 1 millisecond. Suppose that a total number of *D* superimposed signals are detected at the receive antenna. Let T be the symbol transmission time, the C/A spreading waveform of the k^{th} received superimposed signal during the l^{th} signaling interval [(l-1)T, lT] can be mathematically represented as

$$c_k(l) = \sum_{i=0}^{P-1} q_k(i) p(l - iT_c),$$

where $p(\cdot)$ is the chip waveform with duration T_c and $\{q_k(i)\}_{i=0}^{P-1} \in \{-1,1\}^P$ represents the chip sequence of the pseudo-random (PRN) code. The message transmission rate in GPS is $1/T$ = 50Hz, so that the spreading factor is T/T_c = 20460. Suppose that the received waveform is sampled at a rate Q/T_c, where Q is an integer which represents the oversampling factor.

Let τ_i be the unknown propagation time delay of the i^{th} multipath signal. Without loss of generality, we assume $\tau_0 \le \tau_1 \le ... \le \tau_{D-1}$ so that τ_0 represents the propagation time of the line-of-sight signal. Since the propagation delays of the transmitted waveforms are unknown, the received sequence is not synchronized. In other words, during the m^{th} sampling bit interval at the receiver, it may span across the boundary of two transmitted data bits. In order to account for this fact, let us define $q_k^{(j)}$ and s_k to be the augmented zero-padding sampled vectors of the C/A code waveform of the k^{th} superimposed signal. Both $q_k^{(j)}$ and s_k are $2PQ\times 1$ real vectors, specifically they are defined as follows

$$q_k^{(j)} = [\underbrace{0,...,0}_{j}, \underbrace{q_k(0),...,q_k(0)}_{Q},...,$$
$$\underbrace{q_k(P-1),...,q_k(P-1)}_{Q}, \underbrace{0,...,0}_{PQ-j}]^t,$$
$$s_k = (1-\delta_k Q/T_c) q_k^{(i_k)} + (\delta_k Q/T_c) q_k^{(i_k+1)},$$

where

$$i_k = \left\lfloor \frac{Q\tau_k}{T_c} \right\rfloor, \quad \text{and} \quad \delta_k = \tau_k - \frac{i_k T_c}{Q}, \qquad k = 0,...,D-1$$

The two scalars i_k and δ_k account for the delay of the sampled signal. Let $d_k(m) \in \{-1,1\}$ be the data bit of the k^{th} superimposed signal during the m^{th} symbol interval. Let

$y(m) = [y(mPQ+PQ),..., y(mPQ+1)]^t \in C^{PQ}$ and $w(m) = [w(mPQ+PQ),..., w(mPQ+1)]^t \in C^{PQ}$, be the discrete received signal vector and the noise vector during the m^{th} symbol interval, respectively. The $PQ\times 1$ sampled received vector $y(m)$ during the m^{th} symbol interval can be represented as

$$y(m) = \sum_{k=0}^{D-1} \alpha_k (d_k(m-1)s_k^l + d_k(m)s_k^u) + w(m), \qquad (7)$$

where $s_k^l = [0_{PQ} \quad I_{PQ}]s_k$, $s_k^u = [I_{PQ} \quad 0_{PQ}]s_k$. $\alpha_k \in C$ is the unknown complex channel coefficient of the k^{th} superimposed signal detected at the sensor. If the m^{th} symbol interval spans across two message data bits, the vector s_k^l corresponds to the chip sequence of the first data bit, i.e. $d_k(m-1)$, and s_k^u corresponds to the chip sequence of the second data bit $d_k(m)$. The graphical illustration of the received C/A code sequence is depicted in Figure 1. The black box which spans for a time T, which corresponds to the received C/A code sequence with the same data bit. The C/A code sequence inside the dotted box corresponds to the received coded sequence during the m^{th} symbol interval at the receiver. The channel is assumed to be slowly varying in the sense that the channel coefficient, α_k, is an unknown deterministic constant during the observation time. The noise vector $w(m)$ is assumed to be Gaussian distributed with covariance matrix Σ_w.

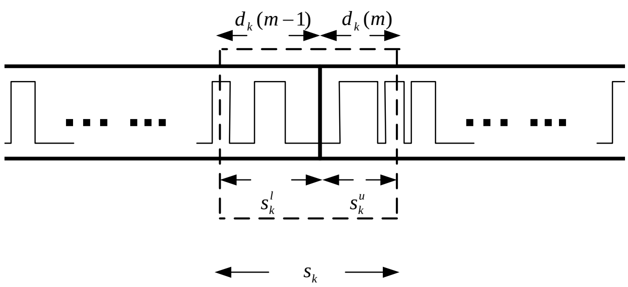


Fig. 1 Sampled received C/A code of the kth multipath signal during the mth symbol interval

3.2 Time delay estimation of superimposed multipath signals

We assumed that the Doppler shift is known perfectly, so that the problem of signal acquisition is equivalent to estimating the code shift. The system model of the superimposed GPS spreading signals is shown in (7). In order to mitigate multipath signals, the time delays $\tau = [\tau_0, \tau_1,..., \tau_{D-1}]^t$ and the channel coefficients

$\alpha = [\alpha_0,...,\alpha_{D-1}]^t$ for all D superimposed signals are required to be estimated.

The ML estimate of the time delay τ is a nonlinear multi-dimensional optimization problem. In order to formulate the EM algorithm, the equation in (7) is rewritten as follows

$$\begin{aligned} y(m) &= \sum_{k=0}^{D-1} \alpha_k (d_k(m-1)s_k^l + d_k(m)s_k^u) + w(m) \\ &= \sum_{k=0}^{D-1} \alpha_k u_k(m;\tau) + w(m) \\ &= \sum_{k=0}^{D-1} (\alpha_k u_k(m;\tau) + w_k(m)) \\ &= T \cdot x(m;\theta), \end{aligned}$$

where

$$\begin{aligned} &u_k(m;\tau) = d_k(m-1)s_k^l + d_k(m)s_k^u, \\ &w(m) = \sum_{k=0}^{D-1} w_k(m), \\ &x_k(m;\theta) = \alpha_k u_k(m;\theta) + w_k(m), \\ &x(m;\theta) = [x_0(m;\theta), x_1(m;\theta),..., x_{D-1}(m;\theta)]^t, \\ &T = \left[\underbrace{I_{PQ} ... I_{PQ}}_{D} \right]. \end{aligned} \tag{8}$$

Recall that Σ_w is the covariance matrix of the noise vector $w(m)$. We assume that the noise vectors $w_k(m)$, $k = 1,2,..., D-1$ are statistically independent, zero-mean and Gaussian distributed with covariance matrix $E(w_k(m)w_k^H(n)) = \Sigma^{(k)}\delta(m-n)$, such that $\Sigma^{(k)} = \beta_k \Sigma_w$. We denote $\delta(\cdot)$ as the Kronecker delta function. The parameters β_k can be adjusted under the constraint

$$\sum_{k=0}^{D-1} \beta_k = 1.$$

The matrix **T** is a many-to-one transformation matrix. The parameter bearing data vector $x(\theta)$ is referred as the complete data and the observation vector $y(m)$ is considered as the incomplete data. The navigation data message $d_k(m-1)$ and $d_k(m)$ are unknown to the receiver prior to code acquisition. In this paper, two assumptions have been made for the formulation of the EM algorithm.

Time delay estimation in CDMA multiple access channel has been widely investigated in the mobile communication research community, e.g. Strom *et al.*, 1996. The data bits for each superimposed signal is generally modelled as independently distributed, so that the propagation delay can be estimated by using subspace methods, e.g. MUSIC (Schmidt, 1986). However, the superimposed GPS multipath signals are possibly coherent, i.e. identical spreading code and data bits. It is particularly true for the close-in scattered signals which are reflected in the vicinity of the receiver. These close-in multipath signals which arrive only slightly later than the line-of-sight signal, usually also possess comparable power as the line-of-sight signal. These are particularly harmful for positioning systems. Due to the coherence of the reflected signals, the data bits of the detected multipath signals are generally identical, i.e. $d_0(m) = d_1(m) = ... = d_{D-1}(m)$. Hence the subspace methods are not applicable in this case, and ML estimation is adopted instead.

Secondly, one period of C/A code spans a time of 1ms, the dwell time used for code searching using the conventional correlators is therefore a multiple of 1ms. A long dwell time can be chosen to improve the acquisition of the weak satellite signals. A high probability of detection and small probability of false alarm can also be achieved simultaneously by using a longer dwell time. However, the GPS navigation message is transmitted at a rate of 50Hz, hence there is a sign reversal at most once every 20ms. This sets the limit of the data length for acquisition. In practice, the dwell time ranges from 1ms to 4ms depending on the SNR of the received signal. Since the close-in multipath signals arrive only slightly later that the desired line-of-sight signal, we assume that there is no sign reversal occurs for all received multipath signals during the dwell time.

3.3 Successive multipath suppression using the EM algorithm

Let us define $u(m;\tau) = [u_0(m;\tau), u_1(m;\tau),..., u_{D-1}(m;\tau)]^t$ and $\alpha = [\alpha_0,...,\alpha_{D-1}]^t$. By using (8), the log-likelihood of the complete data $x(m;\theta)$ is given by

$$\log f_x(x;\theta) = C - (x(m;\theta) - \alpha^H u(m;\tau))^H \Lambda^{-1} (x(m;\theta) - \alpha^H u(m;\tau)) \tag{9}$$

where $\Lambda = Diag(\Sigma^{(0)},...,\Sigma^{(D-1)})$ is a block diagonal matrix and C is a constant which is independent on the parameter θ. However, the complete data $x(\theta)$ is unobserved; the actual distribution cannot be analytically derived. With an initial estimate of the unknown parameters $\hat{\theta}^{(0)} = [\hat{\alpha}^{(0)}, \hat{\tau}^{(0)}]$, the EM estimate iterates until the current estimate $\hat{\theta}^{(k)} = [\hat{\alpha}^{(k)}, \hat{\tau}^{(k)}]$ is sufficiently close to the previous estimate $\hat{\theta}^{(k-1)}$, i.e.

$\left\|\hat{\theta}^{(k)} - \hat{\theta}^{(k-1)}\right\| < \varepsilon$, for an arbitrary small value ε. Let us denote the conditional expectation of the complete data as

$$U(\theta;\hat{\theta}^{(k)}) = E\left\{\log f_x(x;\theta) \mid y, \hat{\theta}^{(k)}\right\} \tag{10}$$

From (8), it is clear that the received vector $y(m)$ and the complete data $x(m;\theta)$ are jointly Gaussian distributed. The complete and the incomplete data are related by the transformation **T**. In particular, the parameter bearing data $x(m;\theta)$ has a mean $\alpha^H u(m;\tau)$ with covariance matrix $\Sigma^{(k)} = \beta_k \Sigma_w$.

The E-Step of the EM algorithm requires taking the conditional expectation of equation (9) stated above. Hence, it is necessity to evaluate $\hat{x}(m;\theta) = E[x(m;\theta) \mid y(m)]$. From the classical estimation theory, the conditional expectation $\hat{x}(m;\theta)$ is given by

$$\hat{x}(m;\theta) = E[x(m;\theta)] + \Sigma_{xy}\Sigma_y^{-1}(y(m) - E[y(m)]) \tag{11}$$

where Σ_{xy} is the cross covariance matrix of $x(m;\theta)$ and $y(m)$, Σ_y is the covariance matrix of $y(m)$. Due to the diagonal structure of the covariance matrix Λ, and by using (11), the conditional expectation of the j^{th} multipath data signal can be written as

$$\hat{x}_j(m;\hat{\theta}^{(k)}) = \hat{\alpha}_j^{(k)} u_j(m;\hat{\tau}^{(k)}) + \beta_j (y(m) - \sum_{n=0}^{D-1} \hat{\alpha}_n^{(k)} u_n(m;\hat{\tau}^{(k)})), \quad j = 0,1,\ldots,D-1 \tag{12}$$

The E-Step of the EM algorithm involves taking the conditional expectation $U(\theta;\hat{\theta}^{(k)})$ in (10), it can be expressed as

$$U(\theta;\hat{\theta}^{(k)}) = C - (\hat{x}(m;\hat{\theta}^{(k)}) - \alpha^H u(m;\tau))^H \Lambda^{-1} (\hat{x}(m;\hat{\theta}^{(k)}) - \alpha^H u(m;\tau)) \tag{13}$$

The conditional expectation $\hat{x}(m;\theta)$ is given in equation (12). On the other hand, the M-Step of the EM algorithm is to locate the parameter $\hat{\theta}$ which maximizes (13), i.e.

$$\hat{\theta}^{(k+1)} = [\hat{\alpha}^{(k+1)}, \hat{\tau}^{(k+1)}] = \arg\max_{\theta} U(\theta;\hat{\theta}^{(k)})$$

Recall that $\Lambda = Diag(\Sigma^{(0)},\ldots,\Sigma^{(D-1)})$, where $\Sigma^{(k)} = \beta_k \Sigma_w$. By using (12) and (13), the j^{th} multipath complex channel coefficient and its EM time delay estimation during the k^{th} iteration is given by

$$\hat{\tau}_j^{(k+1)} = \arg\max_{\tau} \left| \hat{x}_j(m;\hat{\theta}^{(k)})^H u_j(m;\tau) \right| \tag{14}$$

and

$$\hat{\alpha}_j^{(k+1)} = \frac{\hat{x}_j(m;\hat{\theta}^{(k)})^H u_j(m;\hat{\tau}^{(k+1)})}{u_j(m;\hat{\tau}^{(k+1)})^H u_j(m;\hat{\tau}^{(k+1)})} \tag{15}$$

We note from (14) that the time delay of the j^{th} multipath signal $\hat{\tau}_j^{(k+1)}$ is obtained by correlating the estimated multipath signal $\hat{x}_j(m;\hat{\theta}^{(k)})$ with the C/A spreading code $u_j(m;\tau)$. This can be done by using the conventional non-coherent combining method (Kaplan, 1996) or the subspace method (Schmidt, 1986). The channel coefficient estimate in (15) mimics to the Linear Minimum Mean Square Error (LMMSE) estimate. Note also that the denominator of (15) is the energy of the spreading code $u_j(m;\hat{\tau}^{(k+1)})$. It can be considered as a constant. The channel coefficient $\hat{\alpha}_j^{(k+1)}$ is evaluated by using the updated time delay $\hat{\tau}^{(k+1)}$ obtained in (14).

To summarize, the multipath time delay estimation using the EM algorithm involves the following steps:

Initialize $\hat{\theta}_0$ with an initial estimate.

For k = 1,2,…

E-Step

$$\hat{x}_j(m;\hat{\theta}^{(k)}) = \hat{\alpha}_j^{(k)} u_j(m;\hat{\tau}^{(k)}) + \beta_j (y(m) - \sum_{n=0}^{D-1} \hat{\alpha}_n^{(k)} u_n(m;\hat{\tau}^{(k)})), \quad j = 0,1,\ldots,D-1$$

M-Step

$$\hat{\tau}_j^{(k+1)} = \arg\max_{\tau} \left| \hat{x}_j(m;\hat{\theta}^{(k)})^H u_j(m;\tau) \right|$$

$$\hat{\alpha}_j^{(k+1)} = \frac{\hat{x}_j(m;\hat{\theta}^{(k)})^H u_j(m;\hat{\tau}^{(k+1)})}{u_j(m;\hat{\tau}^{(k+1)})^H u_j(m;\hat{\tau}^{(k+1)})}$$

A block diagram of the time delay EM estimate is given in Figure 2. The block "Correlator" in the figure corresponds to the non-coherent correlation in (14). The block "LMMSE" refers to the LMMSE channel coefficient estimation as given in (15). The block "SMS" corresponds to the estimated multipath signals decomposition in equation (12). The resultant signal $\hat{x}_j(m;\hat{\theta}^{(k)})$ is the j^{th} multipath signal estimated at the k^{th} iteration.

Remark 1: The initial estimate can be chosen randomly, for instance, in our simulation, the initial time delay $\hat{\tau}_0$ is randomly chosen in the range [0ms 1ms]. The initial channel coefficient $\hat{\alpha}$ can be chosen arbitrary, the choice of $\hat{\alpha}$ depends highly on the actual SNR of the channel. However, a wise choice of $\hat{\alpha}$ significantly improves the convergence rate of the algorithm.

Remark 2: There are a total number of D distinct multipath time delays $\hat{\tau}_j$ ($j = 0,1...,D-1$) to be estimated. The cost function in (14) is multimodal, convergence to the global maximum is not guaranteed. In some cases, we observed that the time delay estimate of two or more multipath signals give the same values (i.e. $\hat{\theta}_n^{(k)} = \hat{\theta}_m^{(k)}$, $m \neq n$), so that less than D multipath delay estimates are given by the algorithm. This is due to the fact that the resultant estimate falls into the same local stationary value. However, this can be alleviated by re-initializing the intermediate estimate $\hat{\theta}_j^{(k)}$ to an arbitrary value, so that it can follow another iterative path to converge to the desired time delay.

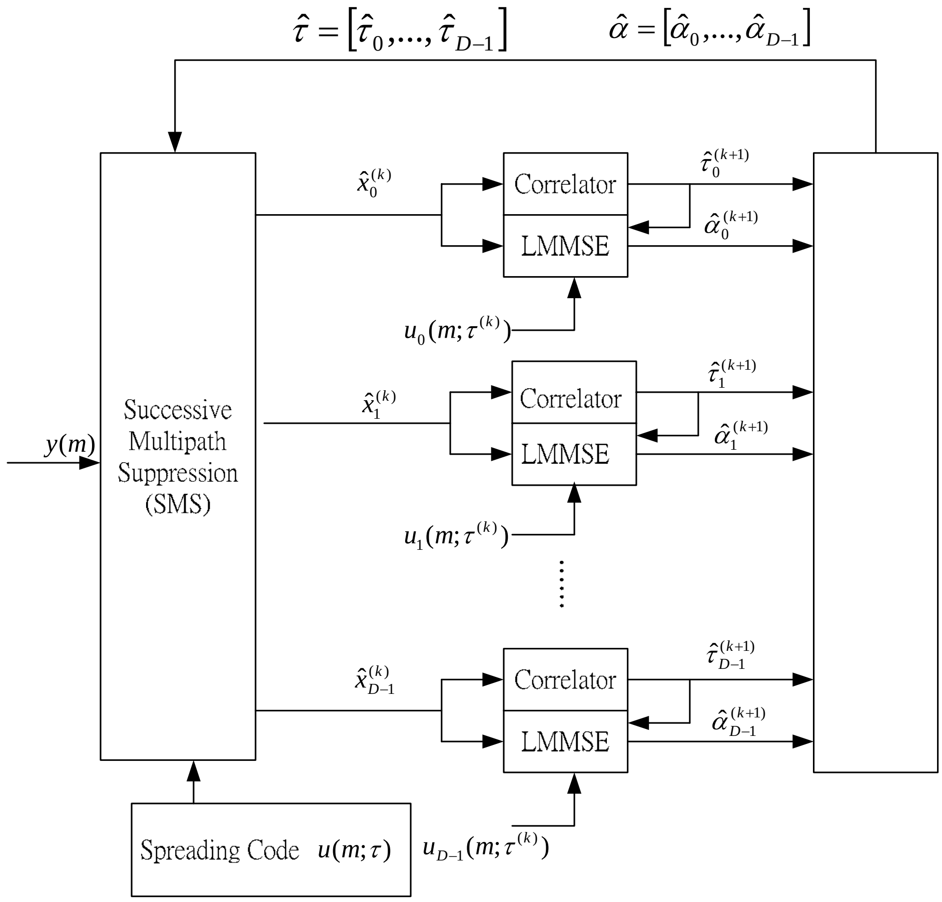


Fig. 2 Block diagram of the successive multipath suppression using the EM algorithm

Remark 3: We can see from (12) that $\hat{x}_j(m;\hat{\theta}^{(k)})$ is simply the estimated j^{th} multipath signal obtained by subtracting all other contributing multipath signals estimated at the previous iteration.

4 Simulation results and discussions

In order to demonstrate and evaluate the proposed iterative signal acquisition algorithm, a number of simulations have been performed. Suppose that two close-in multipath signals along with the line-of-sight signal superimposed signals are detected at the receiver. The SNR of the received line-of-sight signal is -23dB, while the ratio of the line-of-sight signal to each multipath signal is 3dB. For each experiment, the observation time and the acquisition dwell time of the correlator is 2ms, the signal is oversampled by a factor of five (i.e. Q = 5), so that a total number of 10000 samples are received at the GPS receiver during the observation time.

For the sake of illustrating the performance of the proposed algorithm, one of the experimental data is shown in Table 1. The actual time delays for the direct signal and each multipath signal are $\tau_0 = 0.3412ms$, $\tau_1 = 0.5344ms$ and $\tau_2 = 0.7272ms$, respectively. The real channel coefficients have values $\alpha_0 = 0.1001$ for the line-of-sight signal, and $\alpha_1 = \alpha_2 = 0.0709$ for the two multipath signals. The initial estimate of the unknown parameters, α and τ are chosen randomly as:

$\alpha_0 = [0.0136,\ 0.0104,\ 0.0099]^t$

$\tau_0 = [0.4112,\ 0.3093,\ 0.8398]^t$

Five iterations are performed and non-coherent combining is performed in parallel for each superimposed signal. It shows that both the estimation errors for the

Tab. 1 Experiment data of time delay estimates of two multipath signals with a direct signal.

Iteration	$\hat{\tau}_0$	$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\lVert\tau-\hat{\tau}\rVert^2$	$\lVert\alpha-\hat{\alpha}\rVert^2$
1	**0.341200**	**0.341200**	**0.341200**	**0.013602**	**0.010420**	**0.009897**	**0.18630**	**0.01486**
2	**0.341200**	**0.534400**	**0.534400**	**0.048723**	**0.010420**	**0.009897**	**0.03720**	**0.01002**
3	**0.341200**	**0.534400**	**0.698800**	**0.072219**	**0.027579**	**0.009897**	**0.00081**	**0.00638**
4	**0.341200**	**0.534400**	**0.698800**	**0.087977**	**0.045961**	**0.016689**	**0.00081**	**0.00371**
5	**0.341200**	**0.534400**	**0.698800**	**0.098534**	**0.058260**	**0.027845**	**0.00081**	**0.00202**

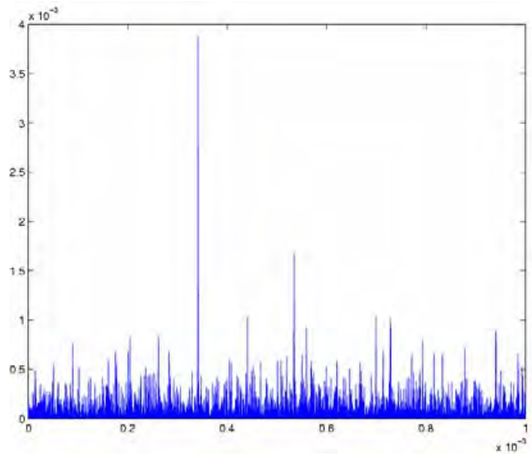


Fig. 3 Non-coherent combining with conventional sliding correlator

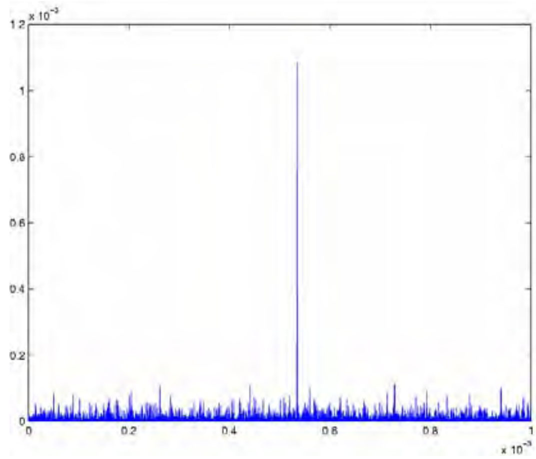


Fig. 5 Non-coherent combining using EM algorithm. The peak corresponds to the time delay estimate of the first multipath signal

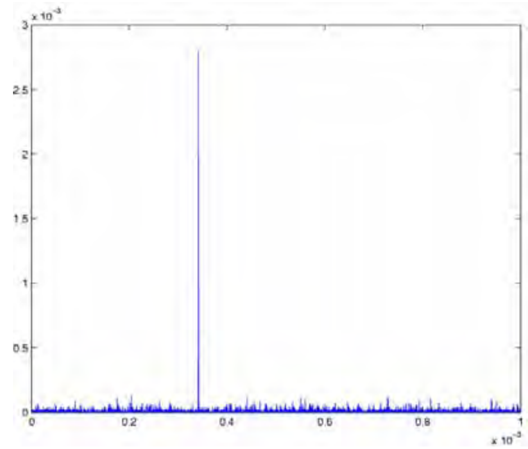


Fig. 4 Non-coherent combining using EM algorithm. The peak corresponds to the time delay estimate of the direct signal

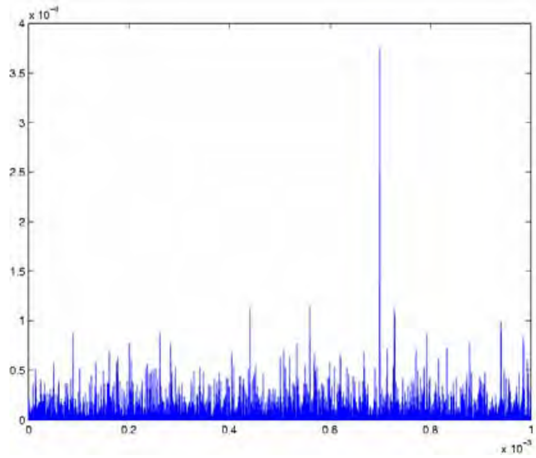


Fig. 6 Non-coherent combining using EM algorithm. The peak corresponds to the time delay estimate of the second multipath signal

time delay estimate and the channel coefficient estimate gradually decreases with the number of iterations performed.

Note that during the first and second iterations, two or more delay estimates give the same values (e.g. $\hat{\tau}_1^{(2)} = \hat{\tau}_2^{(2)} = 0.5344$), it is due to the fact the estimates converge to the same stationary points. This can be remedied by re-initializing the trapped values (e.g. $\hat{\tau}_2^{(2)}$ is set to a random value in the range [0ms 1ms]) at the next iteration, so that the next estimate relies on a new initial estimate, rather than the previous stationary estimate $\hat{\tau}_2^{(2)}$. This has been discussed in Remark 2 of the previous Section.

The correlator output with the conventional sliding correlator is shown in Figure 3. The peaks correspond to the time delays of the two multipath signals and the line-of-sight signal. The correlator output of the direct signal (i.e. $\hat{x}_0(\theta)$ in Figure 2) using the proposed signal acquisition technique with the EM algorithm is shown in Figure 4, we can see that the multipath signals have been suppressed substantially. Hence the probability of false alarm can be significantly reduced and the probability of detection can be increased simultaneously. Figure 5 and Figure 6 show the correlator outputs for the other two multipath signals. Our simulation results show that the proposed GPS signal acquisition system using the EM algorithm provide a multipath suppression capability which is highly attractive in multipath propagation environment.

5 Conclusions

A novel GPS C/A code acquisition system tailored to the multipath propagation environment is proposed in this paper. The proposed GPS receiver jointly estimates the time delays of all the multipath signals simultaneously by using the well-known EM algorithm. The time delays are estimated in an iterative manner, and the multipath signals can be subsequently suppressed. The iterative maximum likelihood estimator is asymptotically efficient. In particular, the pseudorange accuracy can be improved substantially in the presence of coherent multipath signals.

References

Dempster A.P.; Laird N.M.; Rubin D.B. (1977): ***Maximum likelihood from incomplete data via the EM algorithm,*** Journal of the Royal Statistical Society, Series B, vol. 39, no.1, 1977, pp.1-38.

Feder M.; Weinstein E. (1988): ***Parameter Estimation of Superimposed Signals Using the EM Algorithm,*** IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 36, April 1988, pp.477 - 489.

Iltis R.A.; Kim S. (2003): ***Geometric Derivation of Expectation-Maximization and Generalized Successive Interference Cancellation Algorithms with Applications to CDMA Channel Estimation,*** IEEE Transactions on Signal Processing, vol. 51, May 2003, pp.1367 - 1377.

Kaplan E.D. (1996): ***Understanding GPS Principles and Applications,*** Artech House.

Schmidt R.O. (1986): ***Multiple Emitter Location and Signal Parameter Estimation,*** IEEE Transactions on Antennas and Propagation, vol. 34, Mar 1986, pp.276 - 280.

Stoica P.; Nehorai A. (1989): ***MUSIC, Maximum Likelihood, and Cramer-Rao Bound,*** IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 37, May 1989, pp.720 - 741.

Strom E.G.; Parkvall S.; Miller S.L.; Ottersten B.E. (1996): ***Propagation Delay Estimation in Asynchronous Direct-Sequence Code-Division Multiple Access Systems,*** IEEE Transactions on Communications, vol. 44, Jan. 1996, pp.84 - 93.

Van Nee, R. D.J. (1992): ***The Multipath Estimating Delay Lock Loop,*** IEEE Second International Symposium on Spread Spectrum Techniques and Applications, ISSTA 1992, pp.39 - 42.

Van Nee, R. D.J., Siereveld, J., Fenton, P.C. and Townsend, B.R. (1994): ***The Multipath Estimating Delay Lock Loop: Approaching Theoretical Accuracy Limits,*** IEEE Position Location and Navigation Symposium, 1994, pp.246 – 251

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 65-75

Wide-lane Assisted Long Baseline High Precision Kinematic Positioning by GNSS

Hiroshi Isshiki

IMA (Institute of Mathematical Analysis), B507, 3-11, Nishiyamadai 6-chome, Osaka-Sayama, Osaka 589-0022, Japan
e-mail: isshiki@dab.hi-ho.ne.jp Tel: +81-072-367-7139; Fax: +81-072-367-7139

Received: 8 November 2004 / Accepted: 13 July 2005

Abstract. In the previous report (Isshiki, 2004b), theory and algorithm of a new dual frequency long baseline kinematic positioning method were discussed. In the theory, the wide-lane coordinates are used as a constraint for obtaining the correct L1 ambiguities by solving the ionosphere free equations. The effectiveness was verified by some numerical examples. A precise positioning for baseline of several hundred kilometeres are possible. In the present report, the effect of epoch interval and observation length is investigated by numerical calculations. The relationship between the positioning error and the baseline length is also discussed. Furthermore, an algorithm for the real time application is shown.

Key words: wide-lane, long baseline, high precision, kinematic positioning, GNSS

1 Introduction

In the previous report (Isshiki, 2004b), new theory and algorithm for dual frequency long baseline kinematic positioning were discussed. The theory is based on the facts that the initial phase ambiguities of the wide-lane combinations can be obtained correctly irrespective of the baseline length by using HMW (Hatch-Melbourne-Wübbena) combinations (Hatch, 1982; Melbourne, 1985; Wübbena, 1985), the slowly varying part of the ionospheric delays can be estimated with an appropriate accuracy from an external information source such as IONEX (Hugeltobler *et al.*, 2001), and the rapidly varying part can be obtained correctly by using the geometry-free combinations. Specifically, the coordinates of the observation receiver can be obtained rather precisely by solving the wide-lane combinations by using the above-mentioned ambiguities and the ionospheric delays. And the coordinates are used to impose an constraint to the least squares solution of the ionosphere free combinations. This corresponds to the constraint in the case of static positioning where the coordinates are constant.

In the previous report, the validity of the new solution was verified by using data with the epoch interval of thirty seconds downloaded from the homepage of GSI (Geographical Survey Institute, http://terras.gsi.go.jp /inet_NEW/). This new solution makes kinematic positioning for the baseline length of several hundred kilometers possible.

In the present report, the effects of the number of the epoch and observation length are discussed by using data with the epoch interval of one second which can't be downloaded directly from the above-mentioned homepage. The relationship between the baseline length and the positioning error is also investigated. And the algorithm for the real time processing is discussed too.

2 Theory

Double difference observation equations of the pseudo range $P_{\kappa\alpha}^{\,i}(t)$ and the phase range $\Phi_{\kappa\alpha}^{\,i}(t)$ are given as (Isshiki, 2003a-c, 2004a; Hugeltobler et al., 2001)

$$P_{\kappa\alpha\beta}^{\,i\,j} = \rho_{\alpha\beta}^{\,i\,j} + \left(\frac{f_1}{f_\kappa}\right)^2 I_{\alpha\beta}^{\,i\,j} + T_{\alpha\beta}^{\,i\,j} + e_{\kappa\alpha\beta}^{\,i\,j}, \tag{1a}$$

$$\Phi_{\kappa\alpha\beta}^{\,i\,j} = \rho_{\alpha\beta}^{\,i\,j} - \left(\frac{f_1}{f_\kappa}\right)^2 I_{\alpha\beta}^{\,i\,j} + T_{\alpha\beta}^{\,i\,j} + \lambda_\kappa N_{\kappa\alpha\beta}^{\,i\,j} + \varepsilon_{\kappa\alpha\beta}^{\,i\,j}, \tag{1b}$$

where

$$(\bullet)_{\alpha\beta}^{i} = (\bullet)_{\alpha}^{i} - (\bullet)_{\beta}^{i},$$

$$(\bullet)^{ij}_{\alpha} = (\bullet)^{i}_{\alpha} - (\bullet)^{j}_{\alpha}, \tag{2}$$

$$(\bullet)^{ij}_{\alpha\beta} = (\bullet)^{i}_{\alpha\beta} - (\bullet)^{j}_{\alpha\beta} = (\bullet)^{ij}_{\alpha} - (\bullet)^{ij}_{\beta}.$$

The subscripts $\kappa = 1, 2$ refer to the L1 and L2 signals, and the superscript i and the subscript α refer to the satellite and the receiver. f_κ denotes the frequency of κ signal. $I^{ij}_{\alpha\beta}$ and $T^{ij}_{\alpha\beta}$ are the ionospheric and tropospheric delays. $N_{\kappa}{}^{ij}_{\alpha\beta}$ is the initial phase ambiguity.

The initial phase ambiguities of the wide-lane combination can be determined easily for each combination of the satellite and receiver by using HMW (Hatch-Melbourne-Wübbena) combination irrespective of the baseline length. Specifically, if HMW combination:

$$N_{W\kappa\lambda}{}^{ij}_{\alpha\beta} \equiv N_{\kappa}{}^{ij}_{\alpha\beta} - N_{\lambda}{}^{ij}_{\alpha\beta}$$
$$= \frac{\Phi_{\kappa}{}^{ij}_{\alpha\beta}(t)}{\lambda_\kappa} - \frac{\Phi_{\lambda}{}^{ij}_{\alpha\beta}(t)}{\lambda_\lambda} - \frac{f_\kappa - f_\lambda}{f_\kappa + f_\lambda}\left(\frac{P_{\kappa}{}^{ij}_{\alpha\beta}(t)}{\lambda_\kappa} + \frac{P_{\lambda}{}^{ij}_{\alpha\beta}(t)}{\lambda_\lambda}\right) \tag{3}$$

is used, the wide-lane ambiguity $N_{W\kappa\lambda}{}^{ij}_{\alpha\beta}$ is obtained, where λ_κ and λ_λ refer to the wave lengths of κ and λ signals (→ Eq. (10)). This equation uses not only the phase ranges but also pseudo ranges, but the coefficient multiplied to the pseudo ranges is small. So, the noise in the pseudo range is suppressed.

If the geometry free combination $\Phi_{G\kappa\lambda}{}^{ij}_{\alpha\beta}(t)$:

$$\Phi_{G\kappa\lambda}{}^{ij}_{\alpha\beta}(t) \equiv \Phi_{\kappa}{}^{ij}_{\alpha\beta}(t) - \Phi_{W\kappa\lambda}{}^{ij}_{\alpha\beta}(t)$$
$$= -\frac{f_1}{f_\kappa}\left(\frac{f_1}{f_\kappa} + \frac{f_1}{f_\lambda}\right) I^{ij}_{\alpha\beta}(t) + \lambda_\kappa N_{\kappa}{}^{ij}_{\alpha\beta} - \lambda_{W\kappa\lambda} N_{W\kappa\lambda}{}^{ij}_{\alpha\beta} \tag{4}$$

is used, the rapidly varying part of the ionospheric delay can be determined correctly even when the initial phase ambiguity is unknown (Isshiki, 2003c, 2004a). Hence, if the slowly varying part is estimated by external information such as IONEX, a fairly correct estimate of the ionospheric delay may be possible.

The ionospheric delay $I^{ij}_{\alpha\beta}(t)$ is then decomposed into the slowly varying component $\bar{I}^{ij}_{\alpha\beta}$ and the rapidly varying component $\tilde{I}^{ij}_{\alpha\beta}(t)$:

$$I^{ij}_{\alpha\beta}(t) = \bar{I}^{ij}_{\alpha\beta} + \tilde{I}^{ij}_{\alpha\beta}(t), \tag{5}$$

$$\bar{I}^{ij}_{\alpha\beta} = \frac{1}{T}\int_0^T I^{ij}_{\alpha\beta}(t)dt, \tag{6a}$$

$$\tilde{I}^{ij}_{\alpha\beta}(t) = I^{ij}_{\alpha\beta}(t) - \bar{I}^{ij}_{\alpha\beta}, \tag{6b}$$

where T is the measuring time. If $N_{\kappa}{}^{ij}_{\alpha\beta}$ is known, from Eqs. (4) and (5)

$$\bar{I}^{ij}_{\alpha\beta} = \left\{ \begin{array}{l} -\dfrac{1}{T}\displaystyle\int_0^T \left[\Phi_{\kappa}{}^{ij}_{\alpha\beta}(t) - \Phi_{W\kappa\lambda}{}^{ij}_{\alpha\beta}(t)\right]dt \\ + \lambda_\kappa N_{\kappa}{}^{ij}_{\alpha\beta} - \lambda_{W\kappa\lambda} N_{W\kappa\lambda}{}^{ij}_{\alpha\beta} \end{array} \right\} \Bigg/ \left\{\frac{f_1}{f_\kappa}\left(\frac{f_1}{f_\kappa} + \frac{f_1}{f_\lambda}\right)\right\}. \tag{7}$$

The rapidly varying part is written as

$$\tilde{I}^{ij}_{\alpha\beta}(t) = \left\{ \begin{array}{l} -\left[\Phi_{\kappa}{}^{ij}_{\alpha\beta}(t) - \Phi_{W\kappa\lambda}{}^{ij}_{\alpha\beta}(t)\right] \\ + \dfrac{1}{T}\displaystyle\int_0^T \left[\Phi_{\kappa}{}^{ij}_{\alpha\beta}(t) - \Phi_{W\kappa\lambda}{}^{ij}_{\alpha\beta}(t)\right]dt \end{array} \right\} \Bigg/ \left\{\frac{f_1}{f_\kappa}\left(\frac{f_1}{f_\kappa} + \frac{f_1}{f_\lambda}\right)\right\}. \tag{8}$$

Since Eq. (8) uses only the phase ranges, the rapidly varying component $\tilde{I}^{ij}_{\alpha\beta}(t)$ can be estimated very precisely.

For the estimation of the mean component $\bar{I}^{ij}_{\alpha\beta}$, the use of the pseudo range or IONEX may be considered. In thecase of using IONEX, the measuring time T should be longer than the periods of the rapidly varying components. And much attention should also be paid to the reliability of the external information such as IONEX.

If the kinematic positioning using wide-lane combination is conducted together with the correctly obtained ambiguities and the reasonably estimated ionospheric delays, the receiver coordinates may be obtained rather precisely.

The WL (wide-lane) combination $\Phi_{W\kappa\lambda}{}^{ij}_{\alpha\beta}(t)$ is given as

$$\Phi_{W\kappa\lambda}{}^{ij}_{\alpha\beta}(t) \equiv \frac{\lambda_{W\kappa\lambda}}{\lambda_\kappa}\Phi_{\kappa}{}^{ij}_{\alpha\beta}(t) - \frac{\lambda_{W\kappa\lambda}}{\lambda_\lambda}\Phi_{\lambda}{}^{ij}_{\alpha\beta}(t)$$
$$= \rho^{ij}_{\alpha\beta} - \left[\frac{\lambda_{W\kappa\lambda}}{\lambda_\kappa}\left(\frac{f_1}{f_\kappa}\right)^2 - \frac{\lambda_{W\kappa\lambda}}{\lambda_\lambda}\left(\frac{f_1}{f_\lambda}\right)^2\right] I^{ij}_{\alpha\beta}$$
$$+ \lambda_{W\kappa\lambda}\left(N_{\kappa}{}^{ij}_{\alpha\beta} - N_{\lambda}{}^{ij}_{\alpha\beta}\right) + T^{ij}_{\alpha\beta} + \varepsilon_{W\kappa\lambda}{}^{ij}_{\alpha\beta}$$
$$= \rho^{ij}_{\alpha\beta} + \frac{f_1^2}{f_\kappa f_\lambda} I^{ij}_{\alpha\beta} + \lambda_{W\kappa\lambda} N_{W\kappa\lambda}{}^{ij}_{\alpha\beta} + T^{ij}_{\alpha\beta} + \varepsilon_{W\kappa\lambda}{}^{ij}_{\alpha\beta}, \tag{9}$$

where $f_{W\kappa\lambda}$ and $\lambda_{W\kappa\lambda}$ are the virtual frequency and the wave length of the wide-lane signal and given as

$$\frac{f_{W\kappa\lambda}}{c} = \frac{f_\kappa - f_\lambda}{c} = \frac{1}{\lambda_\kappa} - \frac{1}{\lambda_\lambda}, \tag{10a}$$

$$\lambda_{W\kappa\lambda} = \frac{c}{f_\kappa - f_\lambda} = 1\Big/\left(\frac{1}{\lambda_\kappa} - \frac{1}{\lambda_\lambda}\right). \tag{10b}$$

When the above-mentioned receiver coordinates of the observation point is substituted into the IF (Ionosphere free) combination $\Phi_{I\kappa\lambda\,\alpha\beta}^{\;ij}(t)$:

$$\begin{aligned}
&\Phi_{I\kappa\lambda\,\alpha\beta}^{\;ij}(t) \\
&\equiv \frac{1}{2}\left[\frac{\lambda_{N\kappa\lambda}+\lambda_{W\kappa\lambda}}{\lambda_\kappa}\Phi_{\kappa\,\alpha\beta}^{\;ij}(t) + \frac{\lambda_{N\kappa\lambda}-\lambda_{W\kappa\lambda}}{\lambda_\lambda}\Phi_{\lambda\,\alpha\beta}^{\;ij}(t)\right] \\
&= \rho_{\alpha\beta}^{ij} + \left(\lambda_{N\kappa\lambda} N_{\kappa\,\alpha\beta}^{\;ij} + \frac{cf_\lambda}{f_\kappa^{\;2} - f_\lambda^{\;2}} N_{W\kappa\lambda\,\alpha\beta}^{\;ij}\right) + T_{\alpha\beta}^{ij} + \varepsilon_{I\kappa\lambda\,\alpha\beta}^{\;ij}
\end{aligned} \tag{11}$$

a close approximation of the initial phase ambiguity $N_{\kappa\alpha\beta}^{\;ij}$ of the L1 signal is obtained. In the IF combination, the ionospheric delay $I_{\alpha\beta}^{ij}(t)$ is eliminated. The L1 ambiguities are rounded to integers. An integer grid is made around a point consisting of the set of the above-mentioned L1 integer ambiguities. A point on the grid which makes the product of the residuals of the least squares solution of the IF combinations and the distance between the above-mentioned coordinates obtained by the LW combinations and those by the IF combinations minimum is searched on the integer grids. As the result, very precise L1 ambiguities and receiver coordinates for the baseline length of several hundred kilometers are obtained, if the sufficient length is secured for observation time.

In Fig. 4, a flow of algorithm in case of real-time processing is shown. For offline processing, the repetition of calculation for each epoch is not necessary, and the averaging should be conducted in the whole epochs.

3 Observation data

Observation data for fixed stations used in the following numerical examples are shown in Tables 1a and 1b. They are obtained by GEONET operated by GSI (Geographical Survey Institute) and are similar to those used in the previous report (Isshiki, 2004b). The date of the observation data is June 4, 2004. In the previous report, the epoch interval was thirty seconds alone, but, in the present report, data of one second in epoch interval are also used. The thirty second data were downloaded from the homepage of GSI (http://terras.gsi.go.jp /inet_NEW/). The one second data can't be downloaded from the homepage. They were obtained from a source different from GSI. The both data are available for the stations with * in Tables 1a and 1b, and only the thirty second data for other stations. The observation data between GPS time 09:00:00 and 11:00:00 were used for the numerical calculations. The station coordinates in Table 1a are downloaded from the above-mentioned homepage, and the accuracy is very high. The baseline lengths are very close to those shown in Table 1b of the previous report, but the difference of baseline length between Sapporo and Hakodate is a little bit big. In the following calculations, the temporary data are used for the orbits and the ionospheric delays instead of the final data. For reference, some comparisons are shown on the difference between the results obtained by the temporary data and those by the final data.

4 Effects of epoch interval and observation length

4.1 Effects of epoch interval

First. the effects of the epoch interval are studied where the observation length is two hours. The results for SpprMrrn baseline (72209.202 m) are shown below. Table 2a shows the correct values of $N_{W12\,\alpha\beta}^{\;ij}$ and $N_{1\alpha\beta}^{\;ij}$.

Tab. 1a Stations and coordinates (GEONET: 04.06.04)

Stn Name	Stn.ID	Abribiation	$x \quad (m)$	$y \quad (m)$	$z \quad (m)$
Sapporo*	950128	Sppr	-3647450.0190	2923169.1696	4325315.2989
Chitose	960523	Chts	-3665179.0969	2925131.3183	4309263.7688
Tomakomai*	950136	Tmkm	-3681909.6939	2918009.3245	4299470.4964
Muroran*	940018	Mrrn	-3664512.7967	2973568.1063	4276499.5476
Sawara*	960528	Sawr	-3664617.9412	3002995.5330	4255789.3878
Hakodate*	940022	Hkdm	-3685967.8639	3011795.3488	4231283.2112
Iwaizumi*	950164	Iwa2	-3853718.7701	3032158.1371	4065250.0274
Kesennuma	950172	Ksnm	-3893613.1909	3089073.8262	3983982.4123
Kitaibaraki	950214	Ktib	-3959969.0560	3234979.5092	3799716.9881
Ichikawa	93023	Ichk	-3967874.2600	3340981.7196	3699025.1130

Tab. 1b Relative coordinates (dx, dy, dz) between stations and baseline length dr (GEONET: 04.06.04)

BL Name	dx (m)	dy (m)	dz (m)	dr (m)
SpprChts	-17729.0779	1962.1487	-16051.5301	23996.2882
HkdmSawr*	-21349.9227	8799.8158	-24506.1766	33672.0752
SpprTmkm*	-34459.6749	-5159.8451	-25844.8025	43382.5658
TmkmMrrn*	17396.8972	55558.7818	-22970.9488	62586.6979
SpprMrrn*	-17062.7777	50398.9367	-48815.7513	72209.2015
SpprSawr*	-17167.9222	79826.3634	-69525.9111	107241.9608
SpprHkdm*	-38517.8449	88626.1792	-94032.0877	134834.1853
HkdmIwa2*	-167750.9062	20362.7883	-166033.1838	236900.8818
MrrnIwa2*	-189205.9734	-58590.0308	211249.5202	289582.5476
SpprIwa2*	-206268.7511	108988.9675	-260065.2715	349369.9159
SpprKsnm	-246163.1719	165904.6566	-341332.8866	452359.1513
SpprKtib	-312519.0370	311810.3396	-525598.3108	686401.7925
SpprIchk	-320424.2410	417812.5500	-626290.1859	818216.6083

$N_{W12}{}_{\alpha\beta}^{ij}$ are obtained by HMW combinations and $N_1{}_{\alpha\beta}^{ij}$ by the static positioning of the ionosphere free combinations. Table 2b shows a relationship between the epoch interval and the L1 ambiguity $N_1{}_{\alpha\beta}^{ij}$ estimated by the present method. For example, the epoch interval 90 sec corresponds to having 80 data in 2 hours. The same results are obtained irrespective of the epoch interval. The results of the estimated baseline length are shown in Table 2c. The values in Tables 2a and 2b show very small difference, but they give the almost same positioning results as shown in Figs. 1a and 1b. So, the difference may be considered allowable. From the above-mentioned results, if the observation length is equal, the effects of the epoch interval may be small. This may correspond to the fact that the number of the observation equations per unknown does not increase in case of the kinematic positioning.

Tab. 2a Correct double differences DDNW and DDN1 of LW and L1 ambiguities
(SpprMrrn, 04.06.04; 09:00:00-11:00:00; IGR12735, ESAG1560)

	((09)-(05))	((14)-(05))	((22)-(05))	((30)-(05))
DDNW	5174360	-2110139	1387665	-2380484
DDN1	23350963	-10737766	6488635	-13439396

Tab. 2b Estimated double difference DDN1 of L1 ambiguity
(Obt1Ion1: precise orbit, IONEX; measuring time: 2 hours)

1 epoch	((09)-(05))	((14)-(05))	((22)-(05))	((30)-(05))
90 sec	23350963	-10737767	6488634	-13439396
60 sec	23350963	-10737767	6488634	-13439396
45 sec	23350963	-10737767	6488634	-13439396
30 sec	23350963	-10737767	6488634	-13439396
15 sec	23350963	-10737767	6488634	-13439396

Tab. 2c Estimated baseline length

1 epoch	Avg of BL by LI (m)	Sgm of BL by LI (m)	Avg of BL by LW (m)	Sgm of BL by LW (m)
90 sec	72209.219	0.014	72209.267	0.014
60 sec	72209.216	0.013	72209.264	0.014
45 sec	72209.215	0.014	72209.264	0.014
30 sec	72209.215	0.014	72209.263	0.014
15 sec	72209.214	0.014	72209.262	0.014

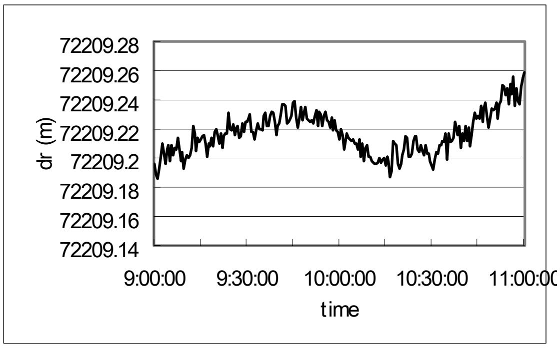


Fig. 1a Baseline length calculated by using L1 ambiguities in Table 2a

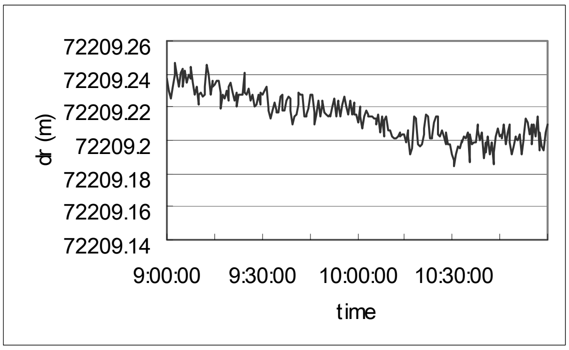


Fig. 1b Baseline length calculated by using L1 ambiguities in Table 2b

Tab. 3a Estimated double difference DDN1 of L1 ambiguity (Obt1Ion1: precise orbit, IONEX; 240 epochs)

Obs. Length	((09)-(05))	((14)-(05))	((22)-(05))	((30)-(05))
20 min	23350965	-10737767	6488633	-13439397
32 min	23350964	-10737766	6488634	-13439396
40 min	23350963	-10737767	6488634	-13439396
60 min	23350963	-10737767	6488634	-13439396
80 min	23350963	-10737767	6488634	-13439396
100 min	23350963	-10737767	6488634	-13439396
120 min	23350963	-10737767	6488634	-13439396

Tab. 3b Estimated baseline length of SpprMrrn (Obt1Ion1: precise orbit, IONEX; 240 epochs)

Obs. Length	Avg of BL by LI (m)	Sgm of BL by LI (m)	Avg of BL by LW (m)	Sgm of BL by LW (m)
20 min	72209.541	0.028	72209.235	0.007
32 min	72209.360	0.022	72209.276	0.008
40 min	72209.228	0.008	72209.280	0.009
60 min	72209.224	0.008	72209.280	0.009
80 min	72209.221	0.011	72209.274	0.012
100 min	72209.217	0.014	72209.265	0.015
120 min	72209.215	0.014	72209.263	0.014

Tab 4 Effects due to accuracy of orbit and ionospheric delay (Temporary File: IGR, ESAG; Final File: IGS, CODG)

	Avg of BL by LI (m)	Sgm of BL by LI (m)	Avg of BL by LW (m)	Sgm of BL by LW (m)
Temp.	72209.215	0.014	72209.263	0.014
Final	72209.215	0.014	72209.280	0.016

4.2 Effects of observation length

Next, the effects of the observation length are studied. The observation length is varied by changing the epoch interval of the 240 epochs. The results are shown in Tables 3a and 3b.

The orbits and ionospheric delays are not the final data (IGS, CODG) but the temporary data (IGR, ESAG). In Table 4 and Fig. 2, the baseline length obtained by using the final data is compared with the case obtained by using the temporary data. The temporary data seem sufficient at least for the baseline SpprMrrn whose baseline length is 72209.202 m .

The results where the orbits are approximated by the broadcast orbits and the ionospheric delays are neglected are shown in Table 5 for reference. The comparison between Tables 3 and 5 suggests that the observation length required for the convergence of the calculation seems shortened.

As another example, a case where the baseline is relatively short is discussed. In applications on land, the short baseline positioning is considered important.

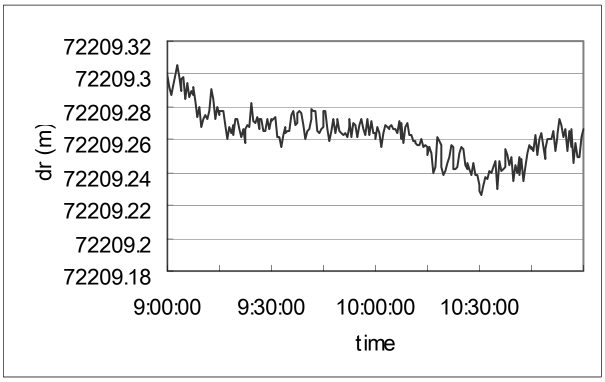


Fig. 2a Baseline length by LW where temporary orbits IGR and ionospheric delay estimates ESAG are used

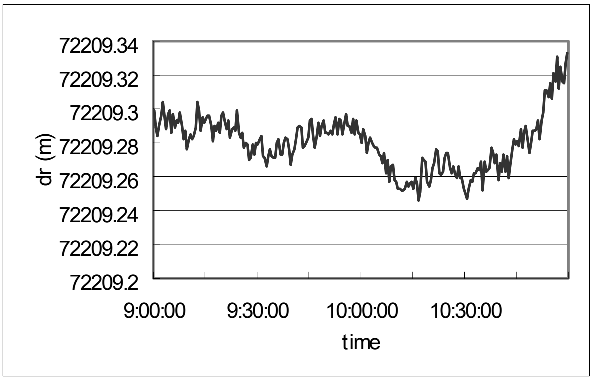


Fig. 2b Baseline length by LW where final orbits IGS and ionospheric delay estimates CODG are used

Tab. 5a Estimated double difference DDN1 of L1 ambiguity (Obt0Ion0: broadcast orbit, IONEX not used; 240 epochs)

Obs. Length	((09)-(05))	((14)-(05))	((22)-(05))	((30)-(05))
20 min	23350961	-10737767	6488635	-13439394
32 min	23350962	-10737768	6488634	-13439395
40 min	23350962	-10737768	6488634	-13439395
60 min	23350963	-10737767	6488634	-13439396
80 min	23350963	-10737767	6488634	-13439396
100 min	23350963	-10737767	6488634	-13439396
120 min	23350963	-10737767	6488634	-13439396

Tab. 5b Estimated baseline length of SpprMrrn (Obt0Ion0: broadcast orbit, IONEX not used; 240 epochs)

Obs. Length	Avg of BL by LI	Sgm of BL by LI	Avg of BL by LW	Sgm of BL by LW
20 min	72208.886	0.018	72209.304	0.008
32 min	72209.058	0.008	72209.342	0.010
40 min	72209.056	0.008	72209.344	0.011
60 min	72209.230	0.010	72209.345	0.010
80 min	72209.226	0.012	72209.338	0.015
100 min	72209.222	0.015	72209.328	0.017
120 min	72209.218	0.016	72209.328	0.017

Tab. 6a Correct double difference DDNW and DDN1 of LW and L1 ambiguities
(SpprTmkm, 04.06.04; 09:00:00-11:00:00; IGR12735, ESAG1560)

	((09)-(05))	((14)-(05))	((22)-(05))	((30)-(05))
DDNW	5228643	-1292258	2607851	-2630956
DDN1	24045828	-6856447	12167118	-13860004

Tab. 6b Estimated double difference DDN1 of L1 ambiguity
(Obt1Ion1: precise orbit, IONEX; 240 epochs)

Obs. Length	((09)-(05))	((14)-(05))	((22)-(05))	((30)-(05))
16 min	24045828	-6856447	12167118	-13860004
20 min	24045828	-6856447	12167118	-13860004
32 min	24045828	-6856447	12167118	-13860004
40 min	24045828	-6856447	12167118	-13860004
60 min	24045828	-6856447	12167118	-13860004
120 min	24045828	-6856447	12167118	-13860004

Tab. 6c Estimated baseline length of SpprTmkm
(Obt1Ion1: precise orbit, IONEX; 240 epochs)

Obs. length	Avg of BL by LI (m)	Sgm of BL by LI (m)	Avg of BL by LW (m)	Sgm of BL by LW (m)
16 min	43382.572	0.007	43382.558	0.007
20 min	43382.571	0.006	43382.561	0.007
32 min	43382.571	0.007	43382.563	0.010
40 min	43382.571	0.006	43382.566	0.011
60 min	43382.572	0.007	43382.574	0.013
120 min	43382.570	0.009	43382.588	0.023

Tab. 7a Estimated double difference DDN1 of L1 ambiguity
(Obt0Ion0: broadcast orbit, IONEX not used; 240 epochs)

Obs. Length	((09)-(05))	((14)-(05))	((22)-(05))	((30)-(05))
16 min	24045829	-6856446	12167118	-13860004
20 min	24045829	-6856445	12167119	-13860004
32 min	24045829	-6856445	12167119	-13860004
40 min	24045828	-6856447	12167118	-13860004
60 min	24045828	-6856447	12167118	-13860004
120 min	24045828	-6856447	12167118	-13860004

Tab. 7b Estimated baseline length of SpprTmkm
(Obt0Ion0: broadcast orbit, IONEX not used; 240 epochs)

Obs. length	Avg of BL by LI (m)	Sgm of BL by LI (m)	Avg of BL by LW (m)	Sgm of BL by LW (m)
16 min	43382.612	0.007	43382.607	0.007
20 min	43382.704	0.007	43382.610	0.007
32 min	43382.709	0.011	43382.607	0.007
40 min	43382.566	0.006	43382.607	0.006
60 min	43382.568	0.007	43382.611	0.006
120 min	43382.565	0.010	43382.608	0.013

Tab. 8 Accuracy of baseline length measured by using LI (Obt1Ion1: precise orbit, IONEX; Epoch=30 sec)

Baseline	Meas. Length	Correct (040604)	Estimated by LI (m)		Estimated by LW (m)	
			Average	Sigma	Average	Sigma
SpprChts	120 min	23996.2882	23996.295	0.009	23996.292	0.015
SawrHkdm	90 min	33672.0752	33672.083	0.009	33672.052	0.016
SpprTmkm	120 min	43382.5658	43382.570	0.009	43382.588	0.023
TmkmMrrn	120 min	62586.6979	62586.706	0.009	62586.783	0.042
SpprMrrn	120 min	72209.2015	72209.215	0.014	72209.263	0.015
SpprSawar	90 min	107241.9608	107241.973	0.014	107242.043	0.025
SpprHkdm	120 min	134834.1853	134834.195	0.021	134834.247	0.022
HkdmIwa2	120 min	236900.8818	236900.989	0.019	236900.852	0.202
MrrnIwa2	120 min	289582.5476	289582.622	0.013	289582.475	0.202
SpprIwa2	120 min	349369.9159	349369.976	0.007	349369.884	0.198
SpprKsnm	120 min	452359.1513	452359.201	0.014	452359.020	0.220
SpprKtib	120 min	686401.7925	686401.794	0.063	686401.428	0.285
SpprIchk	120 min	818216.6083	818216.529	0.111	818216.191	0.262

The results for SpprTmkm whose baseline length is 43382.566 m are shown in Tables 6 and 7. Te results in Table 6 are obtained by using the precise orbits and the ionospheric delays by IONEX, and those in Table 7 by using the broadcast orbits and by neglecting the ionospheric delays. When the baseline becomes short, the correct results are obtained by using the short observation length. In case of the short baseline length, the correct data are obtained even if the ionospheric delays are neglected. However, the ionospheric data are useful in shortening the observation length.

5 Effects of baseline length on precision of measurements

The effects of the baseline length on the positioning accuracy are shown in Tables 8. As the baseline length becomes longer, the accuracy becomes lower. The results tell that the estimation error is less than 10 cm, even if the baseline length exceeds 200 km.

For reference the results obtained by using the broadcast orbits and neglecting the ionospheric delays are shown in Table 9. For example, in case of SpprIwa2 whose baseline is 350 km, there seem to exist no big difference between the results in Table 8 and those in Table 9. However, if the variations are compared, the results obtained by utilizing the ionospheric delays give the higher precision as shown in Figs. 3a and 3b.

Tab. 9 Accuracy of baseline length measured by using LI (Obt0Ion0: broadcast orbit, IONEX not used; Epoch=30 sec)

Baseline	Meas. Length	Correct (040604)	Estimated by LI (m)		Estimated by LW (m)	
			Average	Sigma	Average	Sigma
SpprChts	120 min	23996.2882	23996.292	0.009	23996.309	0.011
SawrHkdm	90 min	33672.0752	33672.082	0.009	33672.096	0.009
SpprTmkm	120 min	43382.5658	43382.565	0.010	43382.608	0.013
TmkmMrrn	120 min	62586.6979	62586.710	0.011	62586.815	0.034
SpprMrrn	120 min	72209.2015	72209.218	0.016	72209.328	0.017
SpprSawar	90 min	107241.9608	107241.981	0.017	107242.131	0.025
SpprHkdm	120 min	134834.1853	134834.245	0.052	134834.381	0.025
HkdmIwa2	120 min	236900.8818	236900.969	0.018	236901.143	0.114
MrrnIwa2	120 min	289582.5476	289582.609	0.015	289582.846	0.112
SpprIwa2	120 min	349369.9159	349369.976	0.028	349370.329	0.118
SpprKsnm	120 min	452359.1513	452359.224	0.063	452359.671	0.125
SpprKtib	120 min	686401.7925	686401.948	0.140	686402.606	0.141
SpprIchk	120 min	818216.6083	818216.961	0.143	818217.630	0.130

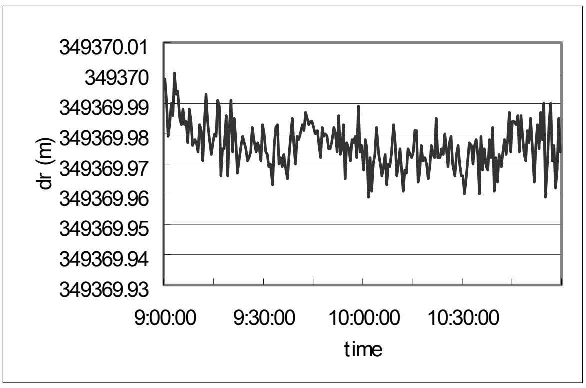


Fig. 3a Baseline length of SpprIwa2 by LI
(Obt1Ion1: precise orbit, IONEX; Epoch=30 sec)

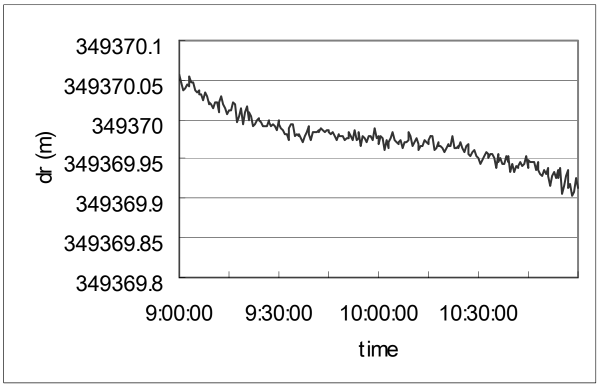


Fig. 3b Baseline length of SpprIwa2 by LI
(Obt0Ion0: broadcast orbit, IONEX not used; Epoch=30 sec)

6 Algorithm for real time processing

The results discussed above were obtained by the offline processing. An algorithm for the real time processing can be made easily. Instead of taking average over the whole epochs, the average is taken over specified epochs in the nearest past. An algorithm is shown in Fig. 4. Kalman filters may also be used instead of taking average over the past epochs. For the estimation of the LW ambiguity

$N_{W12\alpha\beta}^{ij}$, a Kalman filter may be effectively used, where $N_{W12\alpha\beta}^{ij}$, Eq. (3) and the constant nature of $N_{W12\alpha\beta}^{ij}$ are the state variable, observation equation and system transition equation.

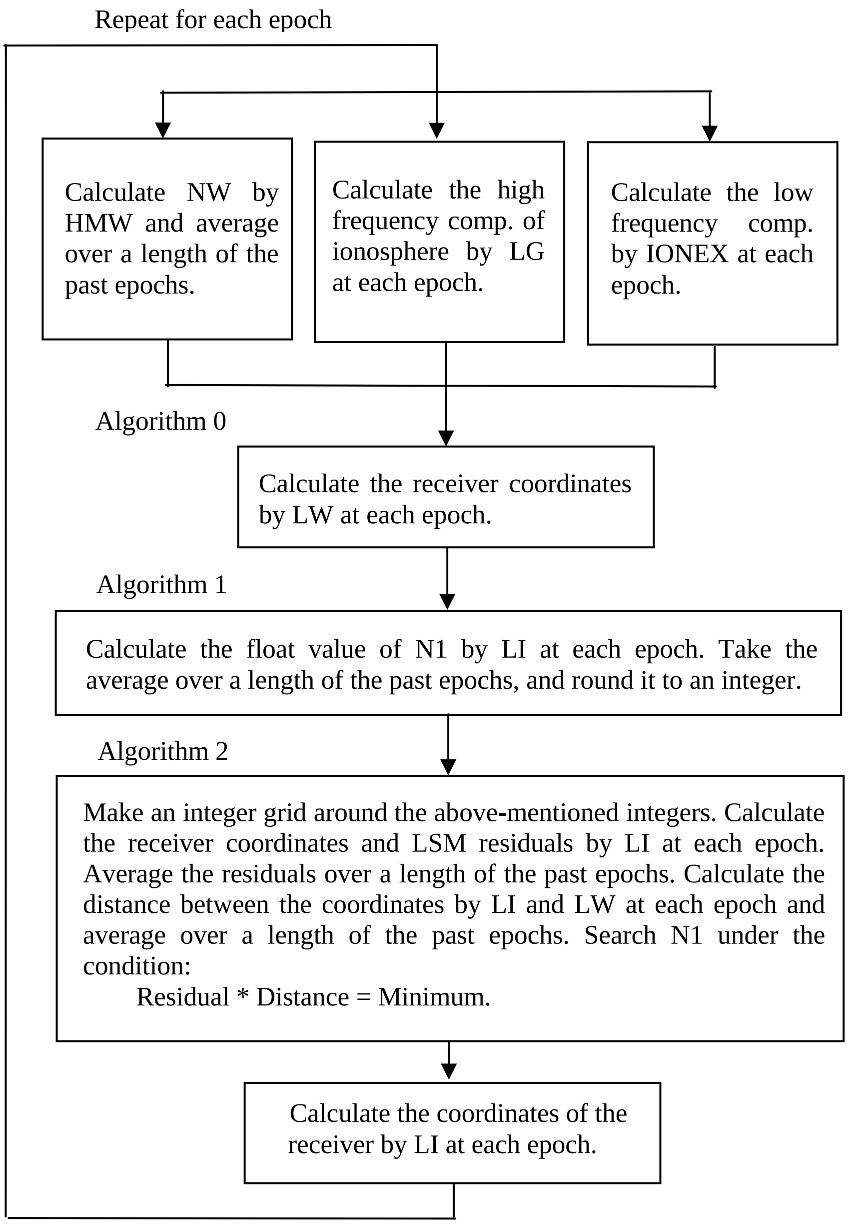


Fig. 4 Algorithm for positioning (Real-time algorithm)

7 Epoch interval of a base station

The epoch interval of the GEONET data downloaded from the homepage of GSI at free of charge is 30 seconds. On the other hand, that of the kinematic positioning is usually 1 second. It may be very useful, if the kinematic positioning of 1 second epoch is possible by combining the rover data of 1 second epoch with the base data of 30 second epoch supplied by GEONET. Furthermore, the merits in the data saving and transmission may be big, if the epoch interval of the base station data can be taken long.

Since the base station is fixed, it does not make unpredictable motions. The ionospheric and tropospheric delays may make small changes within 30 seconds and may be assumed continuous. So, errors included in the 1-second epoch data generated by interpolation of the 30-second epoch data may be small.

In the following numerical examples, the epoch interval of the rover is 5 seconds. Fourth order polynomials are obtained by using Least Square Fitting of nine point data of 30 second epoch. Assuming the offline or real time processing, the data of 5 second epoch are generated by interpolation or extrapolation by using the polynomials.

Figs. 5 and 6 show comparisons of the baseline length calculated by using the raw data with that by the interpolated and extrapolated data respectively. In case of the interpolation, the difference is small, about one centimeter. The difference in case of the extrapolation is bigger, several centimeters. The efforts to lessen the error of the extrapolation should be made.

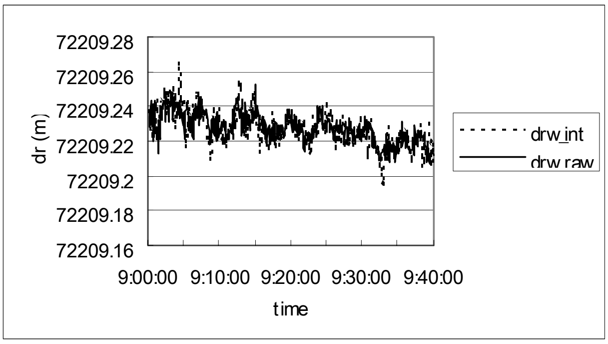


Fig. 5 Baseline length of SpprTmkm obtained by LI, where base station data are generated by interpolation (2004.06.04)

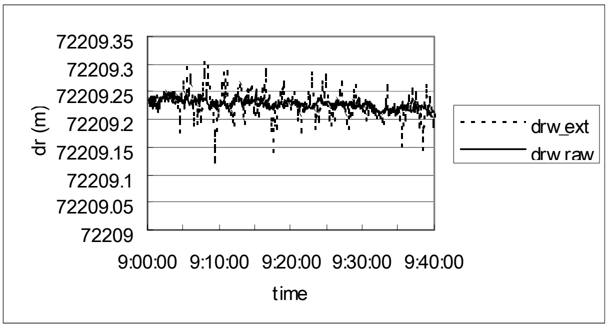


Fig. 6 Baseline length of SpprTmkm obtained by LI, where base station data are generated by extrapolation (2004.06.04)

8 Conclusion

In the previous report (Isshiki, 2004b), a new theory and algorithm of the dual frequency long baseline kinematic positioning were discussed, and the validity was proved by conducting some numerical calculations using the observation data of 30 sec in epoch interval. According to this new solution, the rather precise kinematic positioning was possible for the baseline whose length is several hundred kilometers.

In the present report, the effects of the epoch interval and the observation length are studied by using the observation data of 1 sec in epoch interval. As a result, it was clarified that the long enough observation length is required for the accurate positioning, and the epoch interval itself is not so important. However, the necessary observation length is a function of the baseline length and the ionospheric delays. In order to make this solution useful in practical applications, more data should be analyzed, and a guideline on this point should be given.

The relationship between the baseline length and the positioning error is also investigated. And it was confirmed that a rather accurate positioning is possible by the present solution for the baseline whose length is several hundred kilometers.

An algorithm for the real time positioning is also shown.

Acknowledgements

Nippon GPS Data Service supplied the observation data of one second in epoch interval with cooperation of Mr. M. Kanzaki of Nippon GPS Solutions Corporation. It was very helpful for the verification of the algorithm to be able to select freely the observation stations in GEONET. The author would like to express his sincere thanks to Mr. Kanzaki and both companies.

References

Isshiki H. (2003a): ***An application of wide-lane to long baseline GPS measurements (3)***, ION GPS/GNSS 2003, The Institute of Navigation

Isshiki H. (2003b): ***An approach to ambiguity resolution in multi frequency kinematic positioning***, Proceedings of 2003 International Symposium on GPS/GNSS, pp. 545-552

Isshiki,H. (2003c): ***Long baseline technology for floating body motion measurements in ocean by GPS –Possibility of dual frequency system–***, Conference Proceedings The Society of Naval Architects of Japan, Vol. 2 No.2003A-GS2-2, in Japanese

Isshiki H. (2004a): ***Long baseline GPS kinematic positioning by wide-lane combination***, Conference Proceedings The Society of Naval Architects of Japan, Vol. 3 No.2004S-G2-10, in Japanese.

Isshiki H. (2004b): ***A long baseline kinematic GPS solution of ionosphere-free combination constrained by widelane combination***, OCEANS'04, Kobe, Japan

Hatch R. (1982): ***Synergism of GPS code and carrier measurements***, Proceedings of the Third International Geodetic Symposium on Satellite Doppler Positioning, New Mexico State University, pp.1213-1232

Melbourne W. G. (1985): ***The case for ranging in GPS based geodetic systems***, Proceedings 1st International Symposium on Precise Positioning with the Global

Positioning System, edited by Clyde Goad, pp. 403-412, U. S. Department of Commerce, Rockville, Maryland

Wübbena G. (1985): ***Software developments for geodetic positioning with GPS using TI 4100 code and carrier measurements***, Proceedings 1st International Symp. on Precise Positioning with the Global Positioning System, edited by Clyde Goad, pp. 403-412, U. S. Dept. of Commerce, Rockville, Maryland

Hugeltobler H.; Schaer S.; Fridez P. (2001): *Bernese GPS Software Version 4.2*, Astronomical Institute, University of Berne

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 76-81

Journal of Global Positioning Systems
2005
© CPGPS

Gaussian Random Process and Its Application for Detecting the Ionospheric Disturbances Using GPS

H.. Zhang[1,2], J. Wang[3], W. Y. Zhu[1], C. Huang[1]

(1) Shanghai Astronomical Observatory, Chinese Academy of Sciences, 80# Nandan R.d, Shanghai, China
(2) Graduate School of Chinese Academy of Sciences, Beijing 100039, China
(3) School of Surveying and Spatial Information Systems, University of New South Wales, Sydney, Australia
e-mail: hpzhang@shao.ac.cn Tel: +86-21-64386191-666; Fax: 86-21-64876658

Received: 10 December 2004 / Accepted: 13 July 2005

Abstract. Usually, ionospheric Total Electron Content (TEC) variation with time can be viewed as a stationary random process under quiet conditions. However, sudden events of the Sun and the Earth such as solar flare and sudden commencement of geomagnetic storms may induce the disturbances of the ionosphere, so that the stationary random process is broken; the statistical model parameters change much. Based on this fact, here we make use of the time series of TEC and the auto-covariance function of the stationary process to construct independent identical distribution Gauss sample so that the χ^2 test can be used to detect the abnormity hidden in the sequence. In addition, GPS data from several IGS sites in China during the severe solar flare occurred on 14th July, 2000 are used to verify the method. The results indicate that the disturbances caused by the solar flare can be effectively detected.

Key words: GPS, Total Electron Content, Gaussian Random Process, Ionospheric disturbances

1 Introduction

Over the post two decades, GPS has been widely used in studying of the phenomena of Sudden Increase of Total Electron Content (SITEC) caused by the solar flare (Wan. et al., 2000; Zhang et al., 2000 and 2002; Edward et al., 2000), monitoring the large, middle and small scale of the Travelling Ionosphere Disturbance (TID) (Zhang et al., 2002; Ho., et al., 1996; Saito et al., 1998), verifying the theory of Chapman Ionization, analysing the effects of magnetic storm on the ionosphere(Saito et al., 1998; Ho., et al., 1998), monitoring the ionospheric irregularities (Pi et al., 1997). Due to the high spatial-temporal resolution of Total Electron Content (TEC) data provided by the globally or regionally covered GPS continuously operating reference stations, GPS is helping us further the understanding of the characteristics, the principles and the law of the ionospheric activities at the global, regional or local scale, which greatly promote the development of the high-air atmosphere science and space weather studies. However, most of these researches are accomplished by means of postprocessing, while some special service, such as space weather prediction, wire-less communication and high precision GPS geodetic surveying, needs to detect and deal with the disturbing of the ionosphere so that the effects of the ionosphere on them can be best controlled. Therefore, it is necessary that the theory and methods of detecting the ionospheric disturbances using GPS should be studied comprehensively and systematically. As to this subject, Yuan et al. (2001) offered the random ionospheric disturbance detecting theory and scheme for practical operation. The preliminary testing results based on such a theory using the auto-covariance estimation of variable samples (ACEVS) provided by Yuan (et al. 2001) suggest that the ionospheric anomaly can be best detected. Here we will construct the Independent Identical Gauss Distribution (IIDN(0,1)) samples based on the related characteristics of the stationary random process of the variation of the ionosphere, then the hypothesis testing of the chi-square is involved in analyzing the time series of TEC so that the anomaly can be checked out. On the other hand, to validate such a method, the ACEVS method is introduced and the results produced using the two methods are compared.

2 Constructing IIDN(0,1) sample with stationary random series

Assume a realization of the ergodic Gaussian stationary random process { x_t } with zero expectation value

$$\tilde{x}_i = x_i + e_i \qquad (i = 1, 2, ..., N) \tag{1}$$

where e_i is ergodic Gaussian white noise with zero expectation value (independent of x); N is the number of samples. For simplicity, the stochastic model and other relevant properties of { x_t } and { e_i } are written as

$$E(e_i) = E(x_i) = 0$$
$$COV(x_i, x_{i-r}) = E(x_i x_{i-r}) = \gamma(r)$$
$$COV(e_i, e_{i-r}) = E(e_i e_{i-r}) = \gamma_e(r)$$
$$COV(e, x) = E(ex) = 0$$
$$COV(\tilde{x}_i, \tilde{x}_{i-r}) = \gamma_e(r) + \gamma(r) \tag{2}$$

where COV is the covariance; γ and γ_e are the auto-covariance function of { x_t } and { e_i }, respectively. Since e_i is ergodic Gaussian white noise with zero expectation value, then the γ_e owns the following characteristics

$$\begin{cases} \gamma_e(y) = 0, & y > 0 \\ \gamma_e(y) = D_e, & y = 0 \end{cases} \tag{3}$$

Therefore the auto-covariance function of sample series { $\tilde{x}_i$ } can be expressed as:

$$COV(\tilde{x}_i, \tilde{x}_{i-r}) = E[(x_i + e)(x_{i-r} + e)]$$
$$= \gamma(r) + \gamma_e(r)$$
$$= \gamma(r) \qquad r > 0$$
$$COV(\tilde{x}_i, \tilde{x}_i) = \gamma(0) + D_e \qquad r = 0 \tag{4}$$

where D_e is the variance of the ergodic Gaussian white noise e_i . Since series { x_t } is an ergodic Gaussian random process, then $Y = (x_1, x_2, x_3, ..., x_N)$ can be viewed as N-variate random vector according to the properties of ergodic random process (Liu, 2000). Here $E(Y) = 0$. When variance D_e of e_i is known, covariance matrix Σ_{YY} of N-variate random vector Y can be determineed by using the autovariance function of series { $\tilde{x}_t$ }. Then vector Y follows the N-variate normal distribution with zero expectation and covariance matrix Σ_{YY} . That is: $Y \sim N(0, \Sigma_{YY})$

where covariance matrix Σ_{YY} is non-negative definite. If $\det \Sigma_{YY} > 0$, then random vector

$$Z = \Sigma^{-1/2}(Y - 0) \tag{5}$$

is N-variate random vector with $E(Z) = 0$, $\Sigma_{ZZ} = I_N$, where I_N is n-variate unit matrix. Then random vector Z follows IIDN(0,1).

However, the transferring process described above needs to use the autocovariance function of the stationary random process { x_t }, which can not be accurately known in practice. Then the estimates of autocovariance $\gamma(r)$ can be derived using samples with the following formula (Peter et al., 1991):

$$\tilde{\gamma}(r) = N^{-1} \sum_{i=1}^{N-r} \tilde{x}_i \tilde{x}_{i-r}, \qquad 0 < r \le N - 1 \tag{6}$$

Here Equation (2) is not the unbiased estimates of $\gamma(r)$, but in the condition that $Y = \Sigma^{1/2} \cdot Z$, $\{Z_t\} \sim IID(0, \sigma^2)$, when $N \to \infty$, the asymptotical distribution $\tilde{\gamma}(r)$ is $\gamma(r)$. Then the estimate $\tilde{\gamma}(r), r = 0, 1, ..., N-1$ owns the autocovariance matrix

$$\Sigma_n = \begin{bmatrix} \tilde{\gamma}(0) & \tilde{\gamma}(1) & \dots & \tilde{\gamma}(n-1) \\ \tilde{\gamma}(1) & \tilde{\gamma}(0) & \dots & \tilde{\gamma}(n-2) \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{\gamma}(n-1) & \tilde{\gamma}(n-2) & \dots & \tilde{\gamma}(0) \end{bmatrix} \tag{7}$$

Which is non-negative definite when $n \ge 1$.

As discussed above, the discrete series of Gussian stationary random process can be transferred to a multivariate random vector, then the analysis of the statistical characteristic parameters of stationary random process can be substituted by the corresponding analysis of a multivariate random vector. Therefore, when it happens to be anomaly at the ith observation of { x_t }, the corresponding ith of { Z_t } should be anomaly. Then the statistical tools can be used to analyze it.

3 Determining stationary Ionospheric TEC series with GPS

The GPS TEC observations include the deterministic I part (such as a trend and a period) and stochastic (δI) part due to the ionosphere activity. Usually, for short time scales, for deterministic effects, only the trend variation can be considered, and I can be written as a polynomial $I_t = \sum_{i=0}^{m} a_i t^i$

The stochastic effects δI may be considered as a Gaussian random process with zero expectation value. When random disturbances of the ionosphere happen, their effects on TEC will usually destroy the steady state of δI. Therefore, it is possible to test the change of state of δI using statistical tools.

Assume that $\tilde{I}_t$ is the ionosphere TEC observation at an arbitrary epoch t and ε_i is its Gaussian white noise $\{E(\varepsilon_i) = 0\}$, independent of δI [i.e. $E(\delta I_t \varepsilon_{t+i}) = 0$. Further, $\{\delta I + \varepsilon\}$ is a Gaussian random stationary process with zero expectation value. Thus the ionospheric TEC observation model can be expressed as

$$\tilde{I}_t = I_t + \delta I_t + \varepsilon_i = \sum_{i=0}^{m} a_i t^i + \delta I_t + \varepsilon_i \qquad (8)$$

Define the difference operator as

$$\nabla \tilde{I}_t = \tilde{I}_{t+1} - \tilde{I}_t$$

$$\nabla^k \tilde{I}_t = \nabla(\nabla^{k-1} \tilde{I}_t) = \sum_{i=0}^{k} (-1)^i C_k^i \tilde{I}_{t+k-1} \qquad (9)$$

where C_k^i is the combination operator.

To reduce the trend term I_t, a $q = m+1$ -order difference operation can be used for Eq. (8)

$$\nabla^q \tilde{I}_t = \nabla^q \delta I_t + \nabla^q \varepsilon_i = \nabla^q (\delta I_t + \varepsilon_i)$$

$$E(\nabla^q \tilde{I}_t) = \nabla^q E(\delta I_t + \varepsilon_i)$$

$$= \nabla^q E(\delta I_t) + \nabla^q E(\varepsilon_i) = 0 \qquad (10)$$

Similarly

$$\nabla^q \tilde{I}_{t+h} = \nabla^q \delta I_{t+h} + \nabla^q \varepsilon_{i+h}$$

$$E(\nabla^q \tilde{I}_{t+h}) = 0 \qquad (11)$$

From the above, it can be seen that $\nabla^q \tilde{I}_t$ is a linear combination of $\delta I_{t+q-i} + \varepsilon_{i+q-i}, (i = 0,1,2,...,q)$, while $\{ \delta I_{t+q-i} + \varepsilon_{i+q-i} \}$ is a Gaussian random variable with zero expectation value. So according to the invariance property of linear transformations of Gaussian distributions, $\nabla^q \tilde{I}_t$ is a Gaussian random variable with zero expectation value as well. Then it can be proved that $\{ \nabla^q \tilde{I}_t \}$ is stationary and obviously is an ergodic process as well [Yuan, et al., 2001].

Because $\{ \nabla^q \tilde{I}_t \}$ is an ergodic Gaussian process with zero expectation value, and if $\tilde{x}_t = \nabla^q \tilde{I}_{t+h,t}$ and $x_t = \nabla^q I_{t+h,t}$, then, under normal observation conditions, the series $\{ \tilde{x}_t = \nabla^q \tilde{I}_{t+h,t} \}$ may be considered as the approximate series of $\{ x_t = \nabla^q I_{t+h,t} \}$ and can be transformed to IIDN(0,1) samples according to the method discussed in Sect. 2. In the time series of TEC observations of GPS, the change of the statistical properties of the random ionospheric TEC $\{ x_t = \nabla^q I_{t+h,t} \}$ from a status of stability to one of disturbance can be distinguished by the change of its transformed IIDN(0,1) samples. So it is possible to test by using the GPS time series.

4 Application and analysis

4.1 Scheme for detecting the Anomaly

According to what have been discussed above, we can construct the scheme for testing the anomaly using the transformed *IIDN(0,1)* samples. Here we briefly describe the scheme as following:

1) Get the differenced TEC series from GPS observations so that it is stationary. Usually, second-order differencing is enough.

2) Calculate the estimates of the auto-covariance of the differenced stationary TEC series using formula (6), construct the auto-covariance matrix with formula (7). Here we set 40-order matrix so that the fixed-length samples can be sliding along the time series with time.

3) According to the fixed-length samples windows (40 samples used here) sliding with time, construct the *IIDN(0,1)* samples $\{ Z_t, t = 1,2,...,N \}$ using formula (5). Then χ^2 hypothesis testing can be used to detect the anomaly.

4) If $\chi^2(N) = Z_1^2 + Z_2^2 + ... + Z_N^2 \le \chi_\alpha{}^2(N)$, then the state of ionosphere is stable, otherwise an anomaly happens.

To validate the scheme above, here we also apply the ACEVS method to the sample example. The core principle of ACEVS is to construct an asymptotically independent normal Gaussian sequence $\{\rho_\nabla(r) \sim N(0,1)\}_M^N$ so that χ^2 hypothesis testing can be used. Here M is the minimum samples used to get $\tilde{\gamma}(r)$ using formula (6),

$$\rho_\nabla(r) = \frac{\Delta\tilde{\gamma}_{N+1,N}(r)}{\sqrt{\tilde{\gamma}^2(0)/N(N+1)}}$$

$$\Delta\tilde{\gamma}_{N+1,N}(r) = \tilde{\gamma}_{N+1}(r) - \tilde{\gamma}_N(r)\,. \qquad (12)$$

To simplify the ACEVS method in application, we briefly describe the scheme for testing an anomaly using ACEVS method as following:

1) Get the differenced TEC series from GPS observations so that it is stationary. Usually, second-order differencing is enough.

2) According to the N samples we obtained (here we set r=10, M=100, N>100), calculate $\gamma_N(r)$ and construct sequence $\left\{\tilde{\gamma}_i(r), i = 1,2,\ldots,N-M\right\}$.

3) Get the series

$$\left\{\Delta\tilde{\gamma}_{i,i+1}(r) = \tilde{\gamma}_{i+1}(r) - \tilde{\gamma}_i(r), i = 1,2,\ldots,N-M+1\right\},$$

then construct new sequence

$$\left\{\rho_\nabla(r) = \Delta\tilde{\gamma}_{i,i+1}(r) \Big/ \sqrt{\frac{\tilde{\gamma}^2(0)}{N(N+1)}}, i = 1,2,\ldots,N-M+1\right\}\,.$$

This sequence is an asymptotically independent normal Gaussian sequence.

4) Using the fixed-length sliding sample window (here we set it as 40, so N-M must be more than 40), constructs the statistical quantity $\chi^2(k) = \sum_{i=1}^{k} \rho_\nabla^2(r)$. Then slides the window with time and test the status of ionosphere, if $\chi^2(k) \le \chi_\alpha{}^2(k)$, the ionosphere is in good condition.

4.2 Example and analysis

The increase and decrease of the ionospheric TEC with time are the main phenomena that reflect the status of the ionosphere. Usually, solar flare and the irregular activities of the atmosphere may cause the phenomena of SITEC so that the quiet status of the ionosphere is broken. Here we will analyze the anomaly of the ionosphere caused by the strong solar flare happened on 14th, July, 2000. The two testing schemes described above are used in the following section.

Fig. 1 provides the first-order differenced TEC curves derived from the observations of GPS satellites PRN29 and PRN21 observed at IGS sites WUHN and BJFS respectively on 14th, July, 2000. Obviously, solar flare caused the phenomena of SITEC during the period of UTC10:00~10:30. The amplitude of such sudden increase was up to 0.76TECU/min and lasted almost half an hour. Such ionospheric anomaly could be seen in the TEC observations of other satellites observed at other GPS tracking sites, which means that the solar flare affected the ionosphere in the global scale.

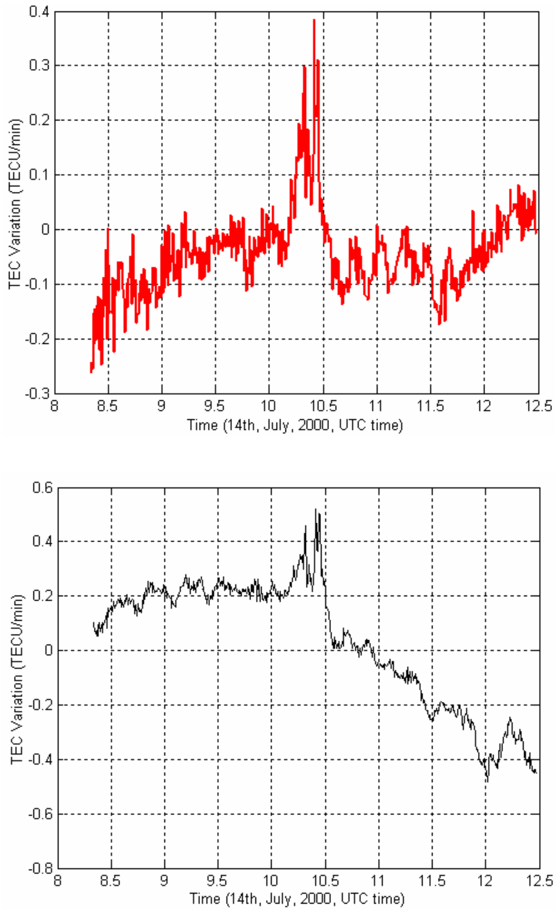


Figure 1. The first-order differenced TEC series observed by PRN29 (up panel) and PRN21(down panel) at WUHN and BJFS respectively on 14th , July, 2000

In Fig. 1 we can see that the trend of the first-order differenced TEC curves is obvious, which indicates that the series are not an stationary process. So the second-order differenced TEC series of PRN29 observed at WUHN site is derived, which is shown in Fig. 2(a). Fig. 2(b) provides the value of the auto-correlation function of R(10) varying with time (the anomaly samples are kicked out), which is a constant and indicates that it is independent of time. Here R(10) is as a example, other values, which is similar to that of Fig. 2(b), of the auto-correlation function varying with time at different intervals have been derived and not shown here. So the second-order differenced TEC series is stationary.

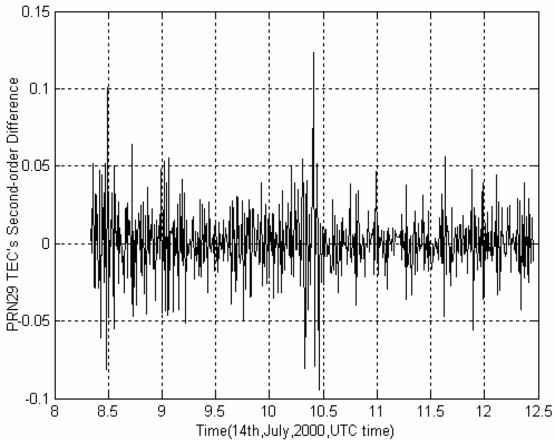


Figure 2(a). The second-order differenced TEC series observed by PRN29 at WUHN

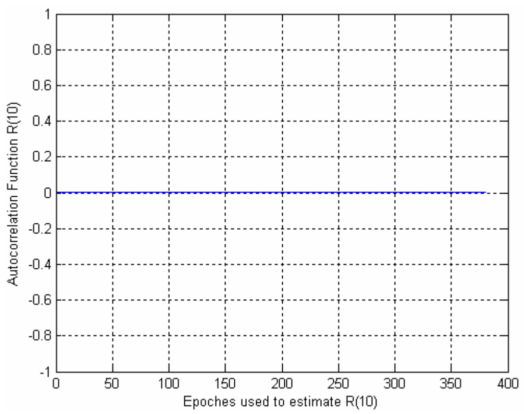


Figure 2(b). The Auto-correlation Function of R(10) (the interval is10 epoches) of the second-ordered differenced TEC series

Fig. 3 is what's obtained by the χ^2 hypothesis test using the IIDN(0,1) samples transformed from the second-order differenced TEC series with the approximate autocorrelation matrix derived from such a random process, while Fig. 4 is the result of the χ^2 hypothesis test using ACEVS method with the same realization of such a random process. In Fig. 3 and Fig. 4, we can see that the two figures are in the similar shape, what's more, the peaks appears and corresponds to the anomaly period reflected in Fig. 1. In addition, the value of χ^2 is more than the threshold value 20.72 before the peak value appear, which indicate that the ionosphere begun to be unstable before the solar flare break out, while the ionosphere recovered quiet soon after the solar flare. This reflects the characteristic of the events of that solar flare on the ionosphere. The similarity of such curves in Fig. 3 and Fig.4 show that the two schemes described above can achieve the same purpose of monitoring and detecting the anomaly of the ionosphere.

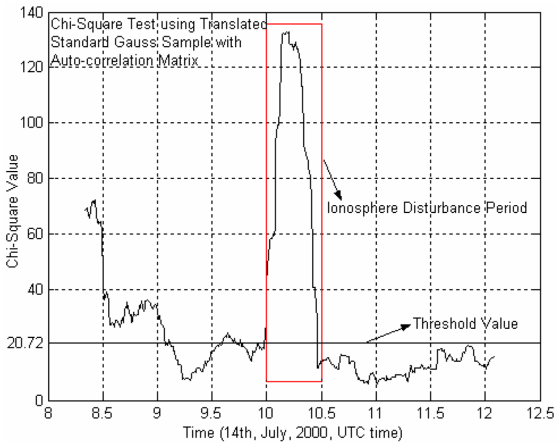


Figure 3. The result of χ^2 testing with the transformed *IIDN(0,1)* Samples

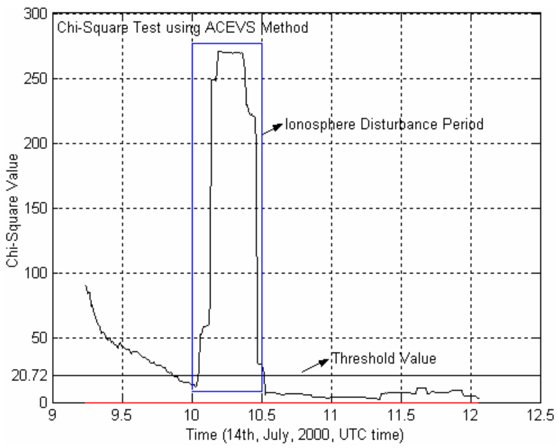


Figure 4. The result of χ^2 hypothesis test using the ACEVS method

4. Summary

The breakage of the quiet ionosphere corresponds to the change of the statistical parameters of the TEC time series, which can be used to monitor the activities of the ionosphere so that the disturbance of the ionosphere can be detected. Monitoring and detecting the ionospheric disturbance is important for the research and prediction of the space weather, as well as GPS surveying, satellite navigation, satellite communication and so on. So the theory and methods of real-time monitoring of the ionosphere with GPS can lead us to know the status of the ionosphere accurately and in real time, so that we can take good steps to avoid great loss. Therefore, the methods used to detect the disturbance of the ionosphere with the stationary random process in this paper provide a good alternative choice.

References

Wan WX, Yuan H, et al (2001) ***The Sudden Increase of the Total Electron Content caused by the great Solar Flare occurred on 7th, July, 2000.*** Science In China (A), 31(supp.), 120-125

Zhang DH, Xiao Z (2002) ***Correlative ionospheric disturbances in the sunlit hemisphere during the flare on July 14, 2000***, Chinese Science Bulletin, 47(12), 975-977

Zhang DH, Xiao Z (2000) ***The study of ionospheric TEC during the flare on Nov. 22, 1998 by means of 4 GPS receivers over China***, Acta Scientiarum Naturalium, Universitatis Pekinensis, 36(3), 414-421

Edward L. Afraimovich, Eugene A. Kosogorov, Ludmila A. Leonovich (2000) ***The use of international GPS network as the global detector (GLOBDET) simultaneously observing sudden ionospheric disturbances***, Earth Planes Space, 52, 1077-1082

Zhang DH, Igarashi K, Xiao Z, Ma GY (2002), ***The Observation of Large Scale Travelling Ionospheric Disturbances Based on GPS Network***, Chinese Journal of Geophysics, 45(4), 469-475

Ho CM, Mannucci A, Lindwister U, Pi X, Tsurutani B (1996) ***Global ionosphere perturbations monitored by the worldwide GPS network***, Geophysical Research Letters, 23(22), 3219-3222

Saito A, Fukao S (1998) ***High resolution mapping of TEC perturbations with the GSI GPS network over Japan***, Geophysical Research Letters, 25(16), 3079-3082

C.M. Ho, Mannucci A, Lindqwister U, Pi X, Tsurutani B, Sparks L, Iijima B, Wilson B, Harris I, and Reyes M (1998) ***Global ionospheric TEC variation during January 10, 1997 Storm***, Geophysical Research Letters, 25(14), 2589-2592

Pi X, Mannucci A, Lindqwister U, Ho C (1997) ***Monitoring of global ionospheric irregularities using the worldwide GPS network***, Geophysical Research Letters, 24(18), 2283-2286

Yuan Y, Ou J (2001) ***Auto-covariance estimation of variable samples (ACEVS) and its application for monitoring random ionospheric disturbances using GPS***, Journal of Geodesy, 75:438-447

Liu CH (2000) ***Stochastic Process,*** Huazhong University of Science and Technology Press

Peter J. Brockwell, Richard A. Davis (1991) ***Time Series: Theory and Methods***, Springer-Verlag Press, New York

Davis K, Hartmann GK (1999) ***Studying the ionosphere with the Global Positioning System***, Radio Science, 32(4), 1695-1703

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2:82-87

Journal of Global Positioning Systems
2005
© CPGPS

GPS Campaigns for Validation of InSAR Derived DEMs

InSu Lee, Hsing-Chung Chang, Linlin Ge

School of Surveying and Information Systems, University of New South Wales, Sydney, Australia
e-mail: i.lee@unsw.edu.au Tel: + 61 2 9385 4190: Fax: +61 2 9313 7493

Received: 6 December 2004 / Accepted: 26 October 2005

Abstract. Interferometric Synthetic Aperture Radar (InSAR) is a rapidly evolving technique. Spectacular results that are obtained in various fields, such as the monitoring of earthquakes, volcanoes, land subsidence and glacier dynamics, as well as in the construction of Digital Elevation Models (DEMs) of the Earth's surface and the classification of different land types, have demonstrated its strength.

As InSAR is a remote sensing technique, it has various error sources due to the satellite positions and attitude, atmosphere, and others, so it is important to validate its accuracy, especially for the DEM derived from SAR images before it can be used for various applications such as disaster prevention, flood mapping, and emergency map.

In this study, Real Time Kinematic (RTK) GPS positioning and Kinematic GPS positioning were chosen as tools for the validation of InSAR derived DEM. The results showed that Kinematic GPS positioning had greater coverage at field test, i.e. larger number of usable sampling points than RTK GPS. However, tracking satellites and transmitting a data between reference-rover, under trees are still pending tasks to be overcome in GPS positioning techniques. Additionally, Airborne Laser Scanning (ALS) is expected to be an alternative as an effective tool for the validation of DEMs.

Key words: Interferometric Synthetic Aperture Radar (InSAR), Digital Elevation Model (DEM), Real Time Kinematic (RTK) GPS positioning, Kinematic GPS positioning

1. Introduction

A DEM measures the height of terrain above a reference datum. DEM as a term is in widespread use and generally refers to the creation of a regular array of elevations, normally squares or hexagon pattern, over the terrain (El–Sheimy, 1999).

Nowadays DEMs can be generated with several methods such as ground surveys, photogrammetry (e.g., analytic photogrammetry and digital photogrammetry), InSAR technique and Airborne Laser Scanning (ALS).

The ground surveys (GPS positioning, levelling, etc) provide height information to a high degree of accuracy, but are time-consuming, laborious and costly, and provide information on point basis only. The point information on height may not be sufficient for conducting an engineering study on regional basis that requires dense spatial information. The spatial extent of height can be obtained from DEM.

The photogrammetric DEMs can be stereo-compilation methods, automatic collection of elevation data by digital correlation from digitized film or digital imagery, and hybrid approaches (Molander, 2004).

Recently Shuttle Radar Topography Mission (SRTM) obtained elevation data on a near-global scale to generate the most complete high-resolution digital topographic database of the Earth. SRTM consisted of a specially modified radar system that flew onboard the Space Shuttle Endeavor during an 11-day mission in February of 2000. This configuration produced the single-pass interferometry and during this period, SRTM mission imaged the Earth's entire land surface between 60 degrees north and 50 degrees south. The C-band SRTM data is being processed into DEMs on a continent-by-continent basis (Peltzer, 1999)

With the advent of InSAR, it may now be possible to obtain height information on regional basis thereby producing DEM up to meter level accuracy. Due to this, the technology is gaining its momentum in many application areas such as lithospheric movements in geology, crustal deformation studies in seismology, global volcano monitoring, landslide monitoring, ice and glacial studies (Arora et al, 2002).

The main aim of this paper is to provide the availability overview of GPS positioning for assessment of DEMs and to reveal the related problems.

2. Basic Concept of InSAR and GPS positioning

2.1 InSAR Overview

Synthetic Aperture Radars (SAR) produce all weather, day and night, high resolution images of the Earth's surface providing useful information about the physical characteristics of the ground and of the vegetation canopy, such as surface roughness, soil moisture, tree height and bio-mass estimates. By combining two or more SAR images of the same area, it is also possible to generate elevation maps and surface change maps with unprecedented precision and resolution. This technique is called “SAR interferometry”. With the advent of spaceborne radars, SAR interferometry has been applied to the study of a number of natural processes including earthquakes, volcanoes, glacier flow, landslides, and ground subsidence (Peltzer, 1999).

Fig.1 presents imaging geometry for a repeat-pass interferometer. One interferogram is formed with images acquired from positions A1 and A2. Assume two identical antennas, A1 and A2, are receiving radar echo signals from a single source. The path length difference, $\Delta\rho$, of the signals received by the two antennas is approximately given by

$$\Delta\rho = \left|\vec{\rho}_2\right| - \left|\vec{\rho}_1\right| \approx B\sin(\theta - \alpha) \tag{1}$$

where $\vec{\rho}_i$ indicates the vector from antenna i to the target, B is the length of the baseline vector which is the vector pointing from antenna 1 to antenna 2, θ is the desired elevation (or) look angle and the baseline orientation angle, α is the angle the baseline vector makes with respect to the horizontal. If a ground resolution element scatters identically for each observation, then the difference of the two phases depends only on the path length difference. The range difference, $\Delta\rho$, may be obtained by measuring, ϕ, the phase between two interferometer signals, using the relation

$$\phi = -\frac{2\pi m \Delta\rho}{\lambda},\ m = 1, 2 \tag{2}$$

where λ is the radar wavelength and m equals to 1 when the path length difference is associated with one way difference, or 2 for the two-way path difference. Using the simplified geometry of Fig. 1, the height of a target, h_t is given by

$$h_t = h - \rho\cos(\theta) \tag{3}$$

where h is the altitude of the radar antenna and ρ is the slant range from the antenna to the target. Generation of accurate topographic maps using radar interferometry places stringent requirements on the knowledge of the platform and baseline vectors (Hensley et al., 2001).

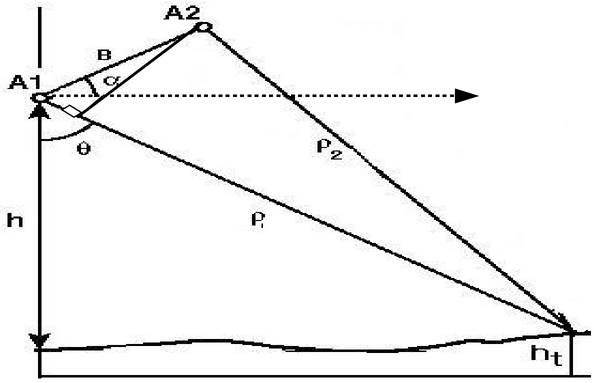


Fig. 1 Radar Interferometric geometry

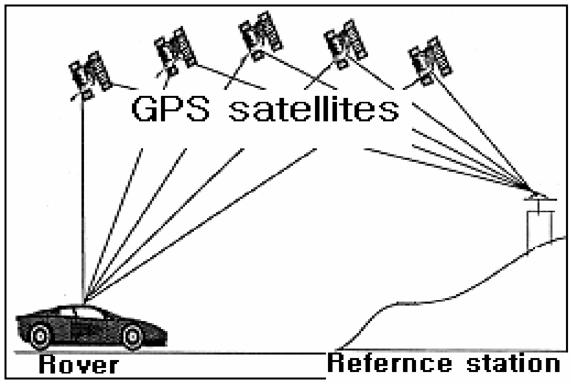


Fig. 2 Overview of Kinematic GPS positioning

2.2 GPS positioning techniques

Kinematic GPS positioning is productive in that the greatest number of points can be determined in the least time. In kinematic GPS positioning, the unknown rover was positioned 'relative' to a reference station that occupied a point of known 3-D coordinates. Fig. 2 presents the graphic presentation of Kinematic GPS positioning.

The Kinematic technique requires the resolution of the phase ambiguities. There are lots of ambiguity resolution techniques for the kinematic case. One of them is called On the Fly (OTF).This solution required an instantaneous positioning (i.e., for a single epoch). The main problem is to find the positions as fast and accurate as possible. This is achieved by starting with approximations for the positions and improving them using least squares adjustments or search techniques (Hoff-mann et al, 1997).

RTK GPS is the dynamic GPS positioning technique available. Using short observation times, this system provides precise results instantaneously whenever continuous four-satellite tracking is available. Nowadays kinematic carrier phase-based positioning can be carried out in real-time if an appropriate communications link is provided over which the carrier phase data collected at a static base receiver can be made available to the rover receiver's onboard computer; to generate the double-differences, resolve the ambiguities and perform the position calculations (Rizos, 1999). This is termed as Real Time Kinematic (RTK) GPS positioning.

3. Generation of InSAR DEM and GPS campaigns

3.1 InSAR DEM

Interferometric SAR is now established as a method for generating DEM from complex SAR data. Validation of such InSAR derived DEMs is still in progress and some results are founded in literature (Balan and Mather, 1999). Interferometry is a technique that interprets the phase difference between two identical SAR images of a single area taken one or more repeat orbit cycles apart. The two ERS satellites operated in tandem for a time, and this allowed for the collection of excellent interferometric pairs.

In this paper, the InSAR DEM was derived using the images acquired during tandem mission of the ERS-1 (20/10/1995) and ERS-2 (30/10/1995) satellites, where there was only one-day difference between the acquisitions of two radar images. Fig. 3 and 4 presents the procedure for DEM generation from both SAR Images and InSAR derived DEM, respectively.

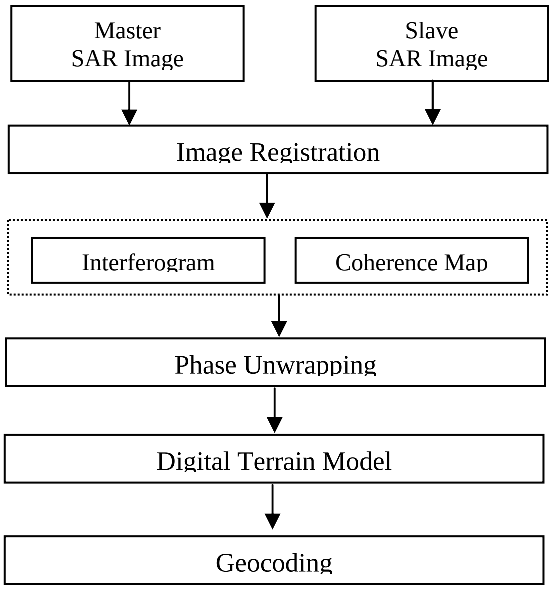


Fig. 3 Generation of InSAR DEM

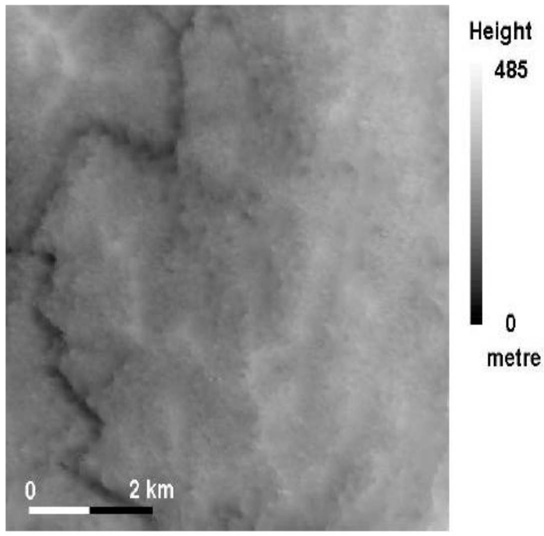


Fig. 4 InSAR Derived DEM

3.2 GPS Field Observation

In theses GPS campaigns, a pair of Leica SR530 receivers with firmware allowing dual-frequency and OTF technique, essential for RTK GPS, and a pair of AT502 antennas, L1/L2 microstrip built-in ground-plane, and a pair of radio modem for transmitting data between a reference station and a rover were employed. This campaign was conducted at the Mining site, Appin, in Australia.

For RTK GPS and Kinematic GPS positioning, a reference station was set up at a site that had a good view to track satellites during the period of test, and a rover moved along the motorway of test field. Positions of a rover antenna were recorded every 1 second in the receiver in real-time with accuracy in several centimeters. At the same time, raw data of both antennas were also stored in the receiver for post processing. With these data, Kinematic GPS positioning was processed. A reference station and a rover set up on the roof of vehicle are shown In Fig. 5.

Fig. 5 A reference station and a rover

4. Analysis of GPS Observable and Assessment of DEM accuracy

4.1 Kinematic GPS positioning and RTK GPS

First of all, the coverage of test area between Kinematic GPS positioning and RTK GPS and was evaluated according to the number of usable sampled points. Fig. 6 (a) and (b) indicate the display map of points acquired from Kinematic GPS positioning and RTK GPS, respectively and Fig. 6(c) presents the overlaid points of Kinematic GPS positioning and RTK GPS with an aerial photograph as background.

(a)

(b)

(c)

Fig. 6 Points of Kinematic GPS positioning (a), RTK GPS (b), and the overlaid map of both (c)

It seems that there is no much difference of point coverage between Kinematic GPS positioning and RTK GPS in Fig. 6, because some measurements recorded in receivers while a vehicle was stationary were already excluded in statistical analysis. However, in actuality, there is a wide difference of data coverage between these two methods.

Especially, some areas marked as circle and square in Fig. 6(c), showed the different data coverage between Kinematic GPS Positioning and RTK GPS. Kinematic GPS positioning has about two times as many usable sampling points as RTK GPS. This may be due to the interference of radio linkage between reference-rover, leading no position solution (e.g. in area, marked as square in Fig.6(c)), and the initialisation problem, leading no solution in RTK GPS (e.g., in area, marked as circle in Fig.6(c)). There is also some probability of both aspects in some areas.

The RMSE of height differences between Kinematic GPS positioning and RTK GPS is within several centimeters. This error value might be good as is the case with both methods.

4.2 Assessment of DEMs' accuracy

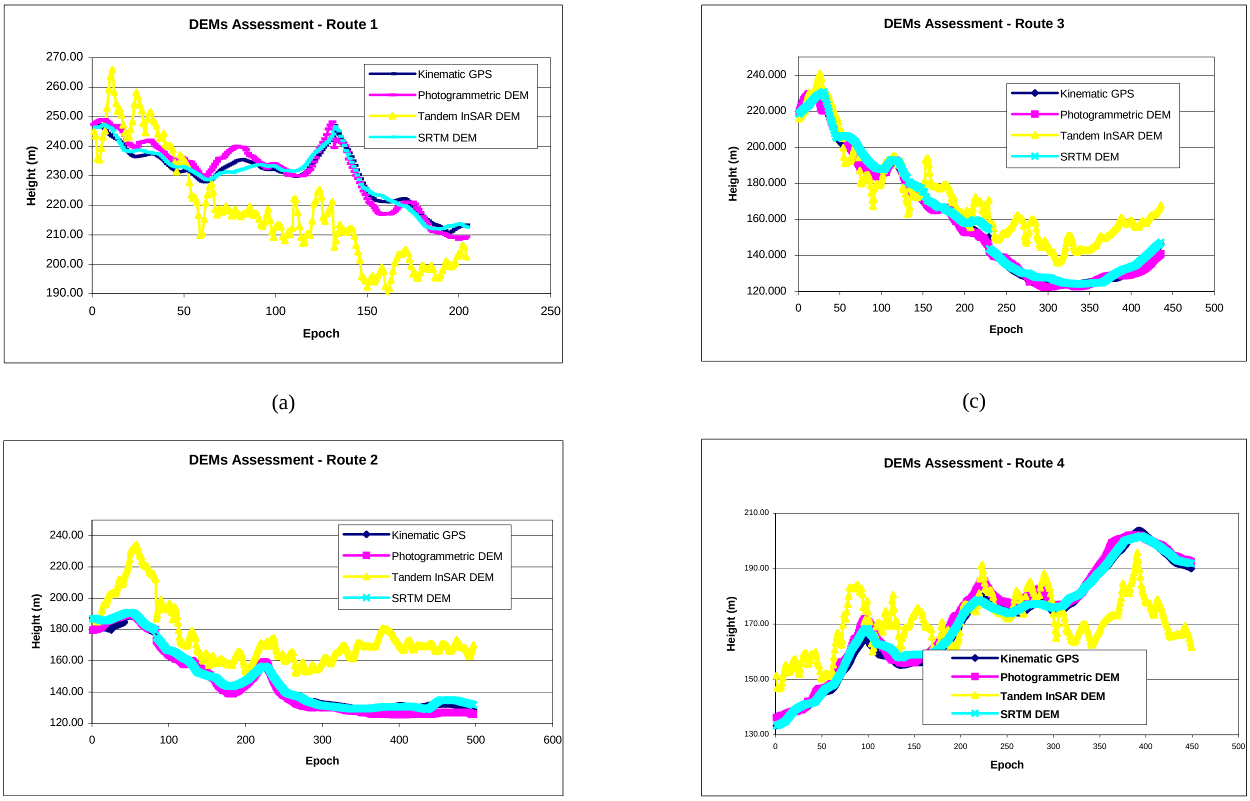
In this paper, 1 arc-second photogrammetric DEM and ERS-1/2 Tandem InSAR DEM as space-borne radar have the pixel size of 30m and 20m, respectively, and SRTM DEM as shuttle-borne radar has the pixel sizes of about 90m. And GPS height profiles were used as ground truth data.

Comparison of three DEMs, i.e. 1 arc-second photogrammetric DEM, SRTM DEM, ERS-1/2 Tandem InSAR DEM against GPS height profiles was used. For this, each height profile of three DEMs was extracted along the same locations where the sampling points in Kinematic GPS positioning were collected. And height profiles of Kinematic GPS positioning were chosen as ground truth data in that the Kinematic GPS positioning had more number of usable sampling points than RTK GPS.

Table 1 indicates the RMSE of height difference between three DEMs and GPS height profiles according to routes.

Tab. 1 RMSE of height difference between three DEMs and GPS height profiles

Sensors / Routes	Photogrammetric DEM	SRTM DEM (Shuttle-borne Radar)	ERS-1/2 Tandem InSAR DEM (Space-borne Radar)
R1	2.95m	1.57m	18.01m
R2	3.26m	2.02m	30.27m
R3	2.16m	3.18m	17.31m
R4	3.24m	1.85m	14.81m



(a) (c)

(b) (d)

Fig. 7 Comparison of height profiles of three DEMs against Kinematic GPS along (a) Route1, (b) Route2, (c) Route 3, and (d) Route 4

The height profiles derived from the photogrammetric DEM and STRM DEM have the mean RMSE of about 2.90m and 2.16m, respectively, while Tandem InSAR DEM has the mean RMSE of about 20.10m against GPS height profiles. In case of Tandem InSAR DEM, this value is likely to be accepted when considering the vertical resolution of ERS images.

Fig. 7 shows that three DEMs have similar trend of heights. Especially, big turbulence of height profile between three DEMs and Tandem InSAR DEM occurred at Route 2. This may be due to satellite inherent errors (e.g., positions and orientations of the satellite), phase unwrapping errors, and atmospheric errors, etc.

Therefore, the detailed information such as satellite orbit information, phase unwrapping algorithm, and especially tropospheric delay to improve the accuracy of InSAR derived DEM is required, and more powerful method like ALS that can validate the accuracy of InSAR derived DEMs should be introduced.

5. Conclusion

This paper dealt with the validation of InSAR derived DEM against GPS height profiles as ground truth data. The results showed that Kinematic GPS positioning had better coverage at the field test, i.e. larger number of usable sampling points than RTK GPS. Therefore, it is expected that Kinematic GPS positioning plays an important role in the validation of InSAR derived DEM because of its cost-effectiveness. But the interference of radio linkage between reference-rover, the tracking satellites and multipath error near and/or under trees are still pending problems to be solved.

Network-Based RTK GPS and the integration SAR with ALS will be an alternative, and what is the most important is that researches related to validation of DEM are further required.

Acknowledgements

This work was supported by the Post-doctoral Fellowship Program of Korea Science & Engineering Foundation (KOSEF).

References

Arora M.K., Patel V. (2002): ***SAR Interferometry for DEM Generation***, (Available in http://www.gisdevelopment.net/technology/rs/techrs0021pf.htm)

Balan P., Mather P.M. (1999): ***Evaluation of the accuracy on InSAR DEM as a function of baseline, wavelength and resolution***, Fringe'99 Workshop, Advancing ERS SAR Interferometry from Applications towards operations, Liège, Belgium, 10~12, November.

Crosetto M., Aragues F.P. (1999): ***Radargrammetry and SAR interferometry for DEM generation: validation and data fusion***, CEOS'99 SAR Workshop, ESA-CNES Toulouse, 26-29, October.

Crosetto M. (2002): ***Calibration and Validation of SAR interferometry for DEM generation***, ISPRS Journal of Photogrammetry & Remote Sensing, 57: 213-227

El-Sheimy N., Valeo C., Habio A., Valeo C. (2005): ***Digital Terrain Modeling: Acquisition, Manipulation And Applications***, Artech House, United Kingdom, 272p.

Hensley S., Riadh M., Paul R. (2001): ***Interferometric Synthetic Aperture Radar (IFSAR)***. In: David F. Maune (Ed.), Digital Elevation Model Technologies and Applications: The DEM Users Manual, ASPRS, Chapter 6, 143-200.

Hofmann-Wellenhof B., Lichtenegger H., Collins J. (1997): ***GPS Theory and Practice (4th Revised Ed.)***, Springer, Wien New York, 389 p.

Kimura H., Yamaguchi Y. (2000): ***Detection of landslide areas using satellite radar interferometer, Photogrammetric Engineering and Remote Sensing***, JOURNAL OF THE AMERICAN SOCIETY FOR PHOTOGRAMMETRY AND REMOTE SENSING, 66(3): 337-344.

Lan H. (1996): ***Development of Real-Time Kinematic System Design, Performance and Results***, .MSc Thesis, Calgary, Alberta, University of Calgary, pp.16.

Manue D.F., Kopp S.M., Crawford C.A., Zervas C.E. (2001): ***Introduction***, In: David F. Maune (Ed.), Digital Elevation Model Technologies and Applications: The DEM Users Manual, ASPRS, Chapter 1, p. 1-31.

Massonnet D., Feigl K.L., Rossi M., Vadon H. (1996): ***Coseismic deformation field of the M 6.7 Northridge, California Earthquake of January 17, 1994, Recorded by Two Radar Satellites using Interferometry,*** Geophysical Research Letters, Vol. 23, No. 9, 969-972.

Molander C.W. (2001): ***Photogrammetry***. In: David F. Maune (Ed.), Digital Elevation Model Technologies and Applications: The DEM Users Manual, ASPRS, Chapter 5, p.121-141.

Peltzer G. (1999): ***Crustal Deformation Studies using SAR Interferometry***, (Available in http://www-radar.jpl.nasa.gov/sect323/InSar4crust/home.html).

Rizos C. (1999): ***GPS enhancements***. In: Chris Rizos (Ed), *Notes on Basics GPS Positioning and Geodetic Concepts*, School of Surveying and Spatial Information, University of New South Wales, pp. 15~16, Electric version (Available in http://www.gmat.unsw.edu.au/snap/gps/gps_notes.htm).

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 88-94

Journal of Global Positioning Systems

Application of Nonlinear Smoothing to Integrated GPS/INS Navigation System

Jaewon Seo, Jang Gyu Lee

School of Electrical Engineering and Computer Science, Seoul National University, Seoul, Korea
e-mail: jwseo1@snu.ac.kr; jgl@snu.ac.kr; Tel: +822-880-7317; Fax: +822-878-8198

Chan Gook Park

School of Mechanical and Aerospace Engineering, Seoul National University, Seoul, Korea
e-mail: chanpark@snu.ac.kr; Tel: +822-880-1675; Fax: +822-888-7958

Hyung Keun Lee

School of Electronis, Telecommunication, and Computer Engineering, Hankuk Aviation University, Gyeonggi-do, Korea
e-mail: hyknlee@hau.ac.kr; Tel: +822-300-0131; Fax: +822-3159-9257

Seong-Baek Kim

Team of Performance & Assessment, Korea Science and Engineering Foundation, Daejeon, Korea
e-mail: kimstar100@kosef.re.kr; Tel: + 8242-769-6321; Fax: + 8242-869-6314

Received: 15 November 2004 / Accepted: 3 October 2005

Abstract. The application of smoothing in the integrated navigation system of 4S-Van is considered. 4S-Van is a mobile mapping system, which provides the position information of various objects on the road. For navigation purpose, it has various sensors such as an inertial measurement unit, a GPS receiver and an odometer. It is also equipped with CCD cameras, laser scanners and video cameras, for mapping purpose. The navigation system of 4S-Van is based on an inertial navigation system, which is aided by GPS and the odometer. Because it can adopt a post-processing method for more accurate and reliable results, the nonlinear smoothing is applied. The nonlinear smoothing, which consists of a forward filter, a backward filter and a smoother, is implemented. For the forward filter, the extended Kalman filter is designed, and for the backward filter, the linearized Kalman filter is constituted. In the smoothing stage, the results of two filters are combined. The algorithm is applied to experimental data and the obtained result demonstrates the effectiveness and good performance of the nonlinear smoothing.

Key words: Inertial Navigation System, GPS, Nonlinear Smoother, Mobile Mapping System

1 Introduction

An Inertial Navigation System (INS) is a system that calculates the position, velocity, and attitude of a vehicle with the output of inertial sensors. The measurements of the inertial sensors contain errors due to physical limitations. These errors are accumulated in the navigation solution of INS, decreasing the accuracy of the solution. Therefore, if the error is not compensated with non-inertial sensors, the information of INS can only be trusted during a short period of time (Siouris, 1993; Titterton, 1997). Nowadays, most INSs are designed with aiding sensors such as GPS.

The mobile mapping system is an information system, which is used for constructing, maintaining, and managing the database of geographical information and facility circumstances. Compared with conventional method mostly depending on surveying, the mobile mapping system is more efficient and easy to operate. 4S-van in Fig. 1 is a mobile mapping system developed by Electronics and Telecommunications Research Institute (ETRI) in Korea. It utilizes CCD cameras, laser scanners, and video cameras for mapping. An inertial measurement unit, GPS receivers, and an odometer are integrated for navigation. With these sensors, the navigation result of 4S-van, such as location, velocity, and attitude can be calculated. For the high accuracy of the database information, the navigation result must be accurate. Therefore, a reliable and accurate navigation system is indispensable for mobile mapping systems. Dissanayake *et al.* (2001) researched the land vehicle navigation system with an inertial measurement unit, a GPS receiver and an odometer. In this paper, for the GPS/INS integrated navigation system of 4S-van, the nonlinear smoother is implemented for efficient estimation of navigation errors and the indirect feedback method is designed for accurate calculation of the navigation solution. The nonlinearity of the navigation system model makes the nonlinear estimators inevitable.

The smoother is a non-real time estimator, which uses all measurements between 0 and T to estimate the state of a system at a certain time t_k, where $0 \le t_k \le T$. In general, smoothing is classified into three categories; fixed-interval, fixed-point, fixed-lag smoothing (Brown and Hwang, 1997). In the fixed-interval smoothing, the time interval of measurements is fixed and optimal estimates at some interior point is searched. Originally the smoother was proposed with three components: a forward filter, a separate backward filter, and a separate smoother (Gelb, 1974). Rauch *et al.* (1965) proposed a scheme with two components: a forward filter and backward smoothing algorithm which incorporates both the backward filter and the smoother. Fraser and Potter (1969) presented a solution to the smoothing problem by interpreting the optimum linear smoother as a combination of two optimum linear filters. For a nonlinear system, Leondes *et al.* (1970) proposed a nonlinear smoother and Yu *et al.* (2002) proposed a nonlinear smoother which is suitable for navigation systems. In this paper, the smoother proposed by Yu *et al.* (2002) is implemented for 4S-van.

The current paper is organized as follows. In section 2, the fixed-interval nonlinear smoother suitable for the navigation of 4S-van is explained. The navigation algorithm for 4S-van is described in section 3 and in section 4 the experimental result is provided. Finally, in section 5, the concluding remarks are presented.

2 Nonlinear smoother

Smoothing is an estimation method using the measurements before and after time t_k for the estimates of the states at time t_k. The following equations show the difference of filtering and smoothing.

Filtering: $\hat{x}(t_k) = E\left\{x(t_k) \middle| Z_k\right\}$ (1)

Smoothing: $\hat{x}(t_k) = E\left\{x(t_k) \middle| Z_m\right\}, m > k$ (2)

where $Z_k = \left\{z_1, z_2, \cdots, z_k\right\}$ and it denotes a measurement set from t_1 to t_k. In equations (1) and (2), it is obvious that the smoothing cannot be used for real time applications.

Smoothers are classified into fixed-point, fixed-lag, and fixed-interval smoother (Brown and Hwang, 1997; Gelb, 1974). In a fixed-interval smoother, the time interval of measurements is fixed and optimal estimates at some interior point is sought. This is the typical problem encountered when processing noisy measurements in non-real time applications. The fixed-interval smoothers have been derived in a few different approaches (Brown and Hwang, 1997; Lewis, 1986). In this paper, however, one kind of the fixed-interval smoothers which is most efficient for the mobile mapping systems will be introduced. The smoother is basically constituted with a forward filter, an independent backward filter and a smoother. Forward and backward filters use measurements independently and the smoother fuses the results of the filters. In the subsequent sections, each filter and smoother will be explained.

The nonlinear system is considered as follows.

$$\dot{x}(t) = f(x(t), t) + G(t)w(t) \tag{3}$$

$$z(t_k) = h(x(t_k), t_k) + v(t_k) \tag{4}$$

where $w(t)$ is white Gaussian noise with zero mean and covariance $Q(t)$; $v(t_i)$ is also white Gaussian noise with zero mean and covariance $R(t)$.

2.1 Forward filter: Extended Kalman filter

As a forward filter of a fixed-interval nonlinear smoother, the extended Kalman filter is used. Assuming that the initial state has a normal distribution with mean $\hat{x}_0$ and covariance P_0, the time propagation of the extended Kalman filter for the system given by equations (3) and (4), is expressed in the following equations:

$$\begin{aligned} &\forall t_{k-1} \le t < t_k \\ &\dot{\hat{x}}(t \mid t_{k-1}) = f(\hat{x}(t \mid t_{k-1}), t),\ \hat{x}(t_{k-1} \mid t_{k-1}) = \hat{x}(t_{k-1}^+) \\ &\dot{P}(t \mid t_{k-1}) = F[t; \hat{x}(t \mid t_{k-1})]P(t \mid t_{k-1}) \\ &+P(t \mid t_{k-1})F^T[t; \hat{x}(t \mid t_{k-1})] + G(t)Q(t)G^T(t), \\ &P(t_{k-1} \mid t_{k-1}) = P(t_{k-1}^+) \end{aligned} \tag{5}$$

where $F[t; \hat{x}(t \mid t_k)] = \left.\dfrac{\partial f(x(t), t)}{\partial x}\right|_{x = \hat{x}(t \mid t_k)}$. The notation $s(t \mid t_k)$ means that the value $s(t)$ is calculated depending on the measurements obtained before and at time t_k.

At the time t_k when a new measurement is available, the filter updates the estimates with the information of the measurement:

$$\begin{aligned}
&\hat{x}(t_k^+) = \hat{x}(t_k^-) + K_k\{z(t_k) - h(\hat{x}(t_k^-), t_k)\} \\
&K_k = P(t_k^-)H^T[t_k; \hat{x}(t_k^-)] \\
&\cdot\{H[t_k; \hat{x}(t_k^-)]P(t_k^-)H^T[t_k; \hat{x}(t_k^-)] + R(t_k)\}^{-1} \\
&P(t_k^+) = \{I - K_k H[t_k; \hat{x}(t_k^-)]\}P(t_k^-) \\
&\cdot\{I - K_k H[t_k; \hat{x}(t_k^-)]\}^T + K_k R(t_k) K_k^T \\
&\hat{x}(t_k^-) = \hat{x}(t_k \mid t_{k-1}), P(t_k^-) = P(t_k \mid t_{k-1})
\end{aligned} \tag{6}$$

where $H[t; \hat{x}(t_k^-)] = \left.\dfrac{\partial h(x(t_k), t_k)}{\partial x}\right|_{x=\hat{x}(t_k^-)}$.

An extended Kalman filter is used for online state estimation and considered as a modification of the linear Kalman filter for nonlinear systems. In the navigation system, the extended Kalman filter estimates the navigation solutions through the perturbed error model.

2.2 Backward filter: Linearized Kalman filter

In nonlinear systems, if the nominal trajectory for linearization is given, the linearized Kalman filter can be used for state estimation. In this paper, the nonlinear smoother adopts the linearized Kalman filter for the backward filter with some modifications (Yu, 2002).

Before explaining the linearized Kalman filter, the nominal trajectory is assumed to be given as follows:

$$\dot{x}_n(t) = f(x_n(t), t), x_n(t_0) = x_0. \tag{7}$$

The time propagation and measurement update of the filter is given as the following equations:

$$\begin{aligned}
&\forall t_{k-1} \le t < t_k \\
&\dot{\hat{x}}(t \mid t_{k-1}) = f(x_n(t), t) \\
&+F[t; x_n(t)]\{\hat{x}(t \mid t_{k-1}) - x_n(t)\} \\
&\dot{P}(t \mid t_{k-1}) = F[t; x_n(t)]P(t \mid t_{k-1}) \\
&+P(t \mid t_{k-1})F^T[t; x_n(t)] + G(t)Q(t)G^T(t) \\
&\hat{x}(t_{k-1} \mid t_{k-1}) = \hat{x}(t_{k-1}^+),\ P(t_{k-1} \mid t_{k-1}) = P(t_{k-1}^+)
\end{aligned} \tag{8}$$

$$\begin{aligned}
&\hat{x}(t_k^+) = \hat{x}(t_k^-) + K_k\{z(t_k) - h(x_n(t_k), t_k) \\
&-H[t_k; x_n(t)]\{\hat{x}(t_k^-) - x_n(t_k)\}\} \\
&K_k = P(t_k^-)H^T[t_k; x_n(t)] \\
&\cdot\{H[t_k; x_n(t)]P(t_k^-)H^T[t_k; x_n(t)] + R(t_k)\}^{-1} \\
&P(t_k^+) = \{I - K_k H[t_k; x_n(t)]\}P(t_k^-) \\
&\cdot\{I - K_k H[t_k; x_n(t)]\}^T + K_k R(t_k) K_k^T \\
&\hat{x}(t_k^-) = x(t_k \mid t_{k-1}),\ P(t_k^-) = P(t_k \mid t_{k-1})
\end{aligned} \tag{9}$$

where $F[t; x_n(t)] = \left.\dfrac{\partial f(x(t), t)}{\partial x}\right|_{x=x_n(t)}$ and $H[t_k; x_n(t_k)] = \left.\dfrac{\partial h(x(t_k), t_k)}{\partial x}\right|_{x=x_n(t_k)}$.

For the backward implementation of the linearized Kalman filter, the variable $\tau = t_f - t$ and equations (10) are newly defined. t_f is the final time in the fixed time interval of the smoothing.

$$\begin{aligned}
&\hat{y}_b(t_f^-) = P_b^{-1}(t_f^-)\hat{x}_b(t_f^-) = 0 \\
&\hat{y}_b(t_k^-) = P_b^{-1}(t_k^-)\hat{x}_b(t_k^-) \\
&\hat{y}_b(t_k^+) = P_b^{-1}(t_k^+)\hat{x}_b(t_k^+)
\end{aligned} \tag{10}$$

where the subscript b denotes that the variables are of the backward filter. Because the backward filter has no information about initial states at time t_f in most cases, $P_b(t_f^-) = \infty$ or $P_b^{-1} = 0$ is naturally assumed. Therefore, it is converted into an information filter with definitions (10) and $\tau = t_f - t$ for the convenience of the implementation. The derived backward filter is given in (11).

$$\begin{aligned}
&\frac{d}{d\tau}P_b^{-1}(\tau) = P_b^{-1}(\tau)F[t_f - \tau; \hat{x}_f(t_f - \tau)] \\
&+F^T[t_f - \tau; \hat{x}_f(t_f - \tau)]P_b^{-1}(\tau) \\
&-P_b^{-1}(\tau)G(t_f - \tau)Q(t_f - \tau)G^T(t_f - \tau)P_b^{-1}(\tau) \\
&\frac{d}{d\tau}\hat{y}_b(\tau) = \{F^T[t_f - \tau; \hat{x}_f(t_f - \tau)] \\
&-P_b^{-1}(\tau)G(t_f - \tau)Q(t_f - \tau)G^T(t_f - \tau)\}\hat{y}_b(\tau) \\
&-P_b^{-1}(\tau)\{f(\hat{x}_f(t_f - \tau), t_f - \tau) \\
&-F[t_f - \tau; \hat{x}_f(t_f - \tau)]\hat{x}_f(t_f - \tau)\} \\
&\hat{y}_b(0) = 0,\ P_b^{-1}(0) = 0 \\
&\hat{y}_b(\tau_k^+) = \hat{y}_b(\tau_k^-) + H^T(\tau_k)R^{-1}(\tau_k)\{z(\tau_k) \\
&-h(\hat{x}_f(\tau_k^-), \tau_k) + H(\tau_k)\hat{x}_f(\tau_k^-)\}
\end{aligned} \tag{11}$$

$$P_b^{-1}(\tau_k^+) = P_b^{-1}(\tau_k^-) + H^T(\tau_k)R^{-1}(\tau_k)H(\tau_k)$$

where $\hat{x}_f(\cdot)$ is the result of the forward filter and in this smoother, it is used for the nominal trajectory of the linearized Kalman filter.

For construction of the indirect feedback navigation system, the backward filter (11) should be perturbed and the following equations can be obtained.

$$\begin{aligned}
&\delta\dot{\hat{y}}_b(\tau) = \{F^T[t_f-\tau;\hat{x}_f(t_f-\tau)] \\
&-P_b^{-1}(\tau)G(t_f-\tau)Q(t_f-\tau)G^T(t_f-\tau)\}\delta\hat{y}_b(\tau) \\
&\dot{P}_b^{-1}(\tau) = P_b^{-1}(\tau)F[t_f-\tau;\hat{x}_f(t_f-\tau)] \\
&+F^T[t_f-\tau;\hat{x}_f(t_f-\tau)]P_b^{-1}(\tau) \\
&-P_b^{-1}(\tau)G(t_f-\tau)Q(t_f-\tau)G^T(t_f-\tau)P_b^{-1}(\tau)
\end{aligned} \tag{12}$$

$$\begin{aligned}
&\delta\hat{y}_b(\tau_k^+) = P_b^{-1}(\tau_k^-)\delta\hat{x}_f(t_f-\tau_k^+) \\
&+\delta\hat{y}_b(\tau_k^-) - H(\tau_k)R^{-1}(\tau_k)\delta z(\tau_k) \\
&P_b^{-1}(\tau_k^+) = P_b^{-1}(\tau_k^-) + H^T(\tau_k)R^{-1}(\tau_k)H(\tau_k)
\end{aligned} \tag{13}$$

where $\delta\hat{y}_b(0)=0$ and $P_b^{-1}(0)=0$ (Yu, 2002).

2.3 Smoother

The estimated states $\hat{x}_s(t_k)$ of the smoother are calculated with the forward filter result, $\hat{x}_f(t_k)$ and the backward filter result, $\delta\hat{y}_b(\tau_k)$.

$$\begin{aligned}
&\hat{x}_s(t_k) = \hat{x}_f(t_k^+) - P_s(t_k)\delta\hat{y}_b(\tau_k^-) \\
&P_s^{-1}(t_k) = P_f^{-1}(t_k^+) + P_b^{-1}(\tau_k^-)
\end{aligned} \tag{14}$$

3 Navigation algorithm

For 4S-van navigation, the indirect feedback method is applied. In the indirect feedback navigation, the navigation errors are estimated with some estimator, and then the navigation solutions are compensated with the estimated errors (Titterton, 1997). In the case of 4S-van, the measurements of GPS and an odometer are used as aiding information for the inertial navigation system.

For application of nonlinear smoothing to the navigation system of 4S-van, the inertial navigation system error model will be necessary. The inertial navigation system error model is described as follows:

$$\begin{aligned}
&\delta\dot{L} = \frac{\rho_E}{R_m+h}\delta h + \frac{1}{R_m+h}\delta v_N \\
&\delta\dot{l} = \rho_N \sec L \tan L\delta L - \frac{\rho_N \sec L}{R_t+h}\delta h + \frac{\sec L}{R_t+h}\delta v_E \\
&\delta\dot{h} = -\delta v_D \\
&\delta\dot{v}^n = [C_b^n f_b]\times\phi - [2\omega_{ie}^n + \omega_{en}^n]\times\delta v^n \\
&+C_b^n\delta f^b + v^n\times(2\delta\omega_{ie}^n + \delta\omega_{en}^n) \\
&\phi = -\omega_{in}^n\times\phi - C_b^n\delta\omega_{ib}^b + \delta\omega_{in}^n
\end{aligned} \tag{15}$$

where $\delta L, \delta l, \delta h, \delta v^n$ and ϕ are the errors of latitude, longitude, height, velocity in the navigation frame, and small tilt angle, respectively. R_m, R_t and C_b^n are the meridian radius, transverse radius, and transformation matrix from the body frame to the navigation frame, respectively. $\rho_{N,E,D}, \delta\omega_{ie}^n$ and $\delta\omega_{en}^n$ are defined as follows:

$$\begin{aligned}
&[\rho_N, \rho_E, \rho_D]^T = [l\cos L, -L, -l\sin L]^T \\
&\delta\omega_{ie}^n = [-\Omega\sin L\delta L, 0, -\Omega cisL\delta L]^T \\
&\delta\omega_{ie}^n \approx \begin{bmatrix} -\frac{\rho_N}{R_t+h}\delta h + \frac{1}{R_t+h}\delta v_E \\ -\frac{\rho_E}{R_m+h}\delta h + \frac{1}{R_m+h}\delta v_N \\ \rho_N\sec^2 L\delta L - \frac{\rho_D}{R_t+h}\delta h + \frac{\rho_D}{v_E}\delta v_E \end{bmatrix} \\
&\omega_{in}^n = \omega_{ie}^n + \omega_{en}^n, \delta\omega_{in}^n = \delta\omega_{ie}^n + \delta\omega_{en}^n
\end{aligned} \tag{16}$$

where Ω is the earth rate. δf^b is the accelerometer noise and $\delta\omega_{ib}^b$ is the gyroscope noise. These noises are modelled as follows:

$$\begin{aligned}
&\delta f^b = \nabla_a + \omega_a \\
&\delta\omega_{ib}^b = \varepsilon_g + \omega_g \\
&\dot{\nabla}_a = 0 \\
&\dot{\varepsilon}_g = 0
\end{aligned} \tag{17}$$

where ∇_a and ε_g are the random bias noises; ω_a and ω_g are the white Gaussian noises. The odometer scale factor error must be compensated. The scale factor error, k_{xo}, is assumed to be a random constant with an initial distribution of $N(0, Q_K)$. The state vectors are given in (18).

$$\begin{aligned}
\delta x^T = \big[&\delta L \quad \delta l \quad \delta h \quad \delta v_N \quad \delta v_E \quad \delta v_D \\
&\phi_N \quad \phi_E \quad \phi_D \quad \nabla_a \quad \varepsilon_g \quad k_{xo}\big]
\end{aligned} \tag{18}$$

The odometer observation model is given as follows:

$$V_m^b \approx (1 - k_{xo}) V_x^b \tag{19}$$

where V_m^b is the output of the odometer, k_{xo} is the scale factor error and V_x^b is the forward direction velocity of the vehicle, respectively. From equation (19), the error model applied to the smoother is derived as follows (Yu, 1999):

$$\delta V_m = \hat{V}^n - V_m^n = \delta v^n - \hat{V}^n \times \begin{bmatrix} \phi_N \\ \phi_E \\ \phi_D \end{bmatrix} - \hat{C}_b^n V_x^b \Delta k + \upsilon_o, \tag{20}$$
$$\Delta k = k_{xo} - \hat{k}_{xo}$$

where $\hat{V}^n$ is the velocity estimates in the navigation frame, V_m^n is the measured velocity by the odometer in the navigation frame, $\hat{C}_b^n$ is the transformation matrix calculated with the estimated values, $\hat{k}_{xo}$ is the estimated scale factor error and υ_o is the white Gaussian noise with zero mean and appropriate covariance.

Position measurement of a GPS receiver is modelled as follows:

$$\delta P = \hat{P} - P_m = \delta P + \upsilon_p \equiv \begin{bmatrix} \delta L \\ \delta l \\ \delta h \end{bmatrix} + \upsilon_p \tag{21}$$

where $\hat{P}$ is the estimated position vector in latitude, longitude and height, P_m is the measured data and υ_p is the white Gaussian noise with zero mean and appropriate covariance.

The covariances are given in the manufacturer's specification table or can be determined by the statistical result of repetitive experiments.

4 Experimental result

4S-van in Fig. 1 is a mobile mapping system which has been developed by the Electronics and Telecommunications Research Institute in Korea. It has CCD cameras, lase scanners, and video cameras for mapping. For constitution of integrated navigation system, the LN-200 developed by Northrop Grumman Co. is used as an inertial measurement unit and two GPS receivers by Trimble Co. are equipped on top of the vehicle. The position information for the aided inertial navigation system is acquired in the DGPS method (Hofmann-Wellenhof, 1994; Kaplan, 1996). Therefore, the provided position measurements are very accurate. The noise characteristics of the sensors are given by the manufacturers and the dominating errors are given in Table 1.

Fig. 1 4S-van

Tab. 1 Specifications of sensors

Sensors			Standard deviation (1σ)
IMU (LN-200)	Accelerometer	Bias repeatability	1mg
		Scale factor stability	300ppm
		Bias variation	50ug with 60 second correlation time
		White noise	50ug
		Bandwidth	100Hz
	Gyro	Bias repeatability	$1^\circ / hr$
		Random walk	$0.07^\circ / \sqrt{hr}$
		Scale factor stability	100ppm
		Bias variation	$0.35^\circ / hr$ with 100 second correlation time
		Bandwidth	>500Hz
DGPS (Position accuracy)			50cm
Odometer (Velocity accuracy)			1.2m/s

The test area in this experiment is an asphalt-paved road in Daejeon, Korea. Total distance of navigation is about 11km and total time is 2150 seconds. The heights above the mean see level change from 50m to 110m in that area.

The navigation results of the experiment are shown in Fig. 2, Fig. 3 and Fig. 4. Fig. 2 shows the navigation result in plane view. Fig. 3 shows the results of north direction, east direction and height. Fig. 4 shows the estimated attitude result. The estimation errors of the result are summarized in Table 2.

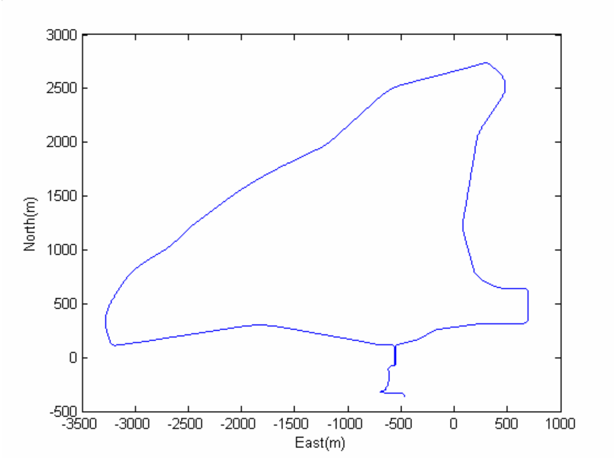


Fig. 2 The estimated position result of 4S-van in plane view

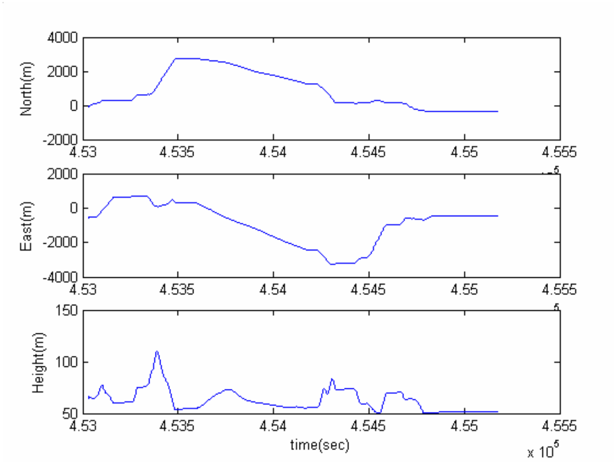


Fig. 3 The estimated position result in north, east and height direction

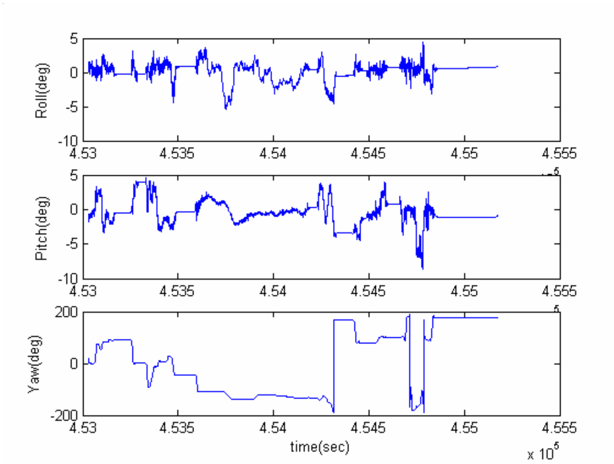


Fig. 4 The attitude estimation result of 4S-van

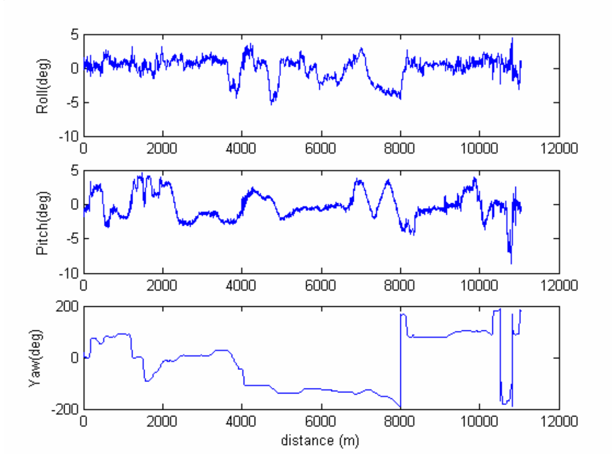


Fig. 4 The attitude estimation result of 4S-van (continued)

Tab. 2 The navigation errors

	Error (RMS)
North	0.299m
East	0.396m
Height	0.726m
Roll	0.248deg
Pitch	0.214deg
Yaw	0.201deg

According to these results, 4S-van can provide its position with errors of 30~40cm in the horizontal plane. The attitude error of 4S-van also causes the position error in the mapping result. For the case of a 30m-apart object from 4S-van, the error of 0.1 deg in the attitude of 4S-van corresponds to about 5cm error in the position of the object. Therefore, the above navigation results cause mapping errors up to about 60cm in horizontal plane which can be calculated with north, east, and yaw errors. With the current algorithm and 4S-van hardware, the map of about 60cm-errors can be made and it is sufficient for the intermediate level of mobile mapping.

5 Conclusion

In this paper, the nonlinear smoother is applied to the GPS/INS integrated navigation system of 4S-van. 4S-van is equipped with an inertial measurement unit, GPS receivers and an odometer for navigation. For accurate navigation solution, the error model of inertial navigation system is studied and the nonlinear smoothing algorithm is applied. The approximate error model is linear and 16th order. With this error model, the fixed-interval nonlinear smoothing is implemented. It consists of three components: Extended Kalman filter, Linearized Kalman filter, smoother. The implemented algorithm is applied to the data obtained by an experiment. The result of the implemented algorithm has sufficient accuracy for mapping applications.

Acknowledgment

The authors would like to acknowledge the support and assistance of ETRI (Electronics and Telecommunications Research Institute), ACRC in SNU, and BK-21 project in Korea.

References

Brown R. G.; Hwang P. Y. C. (1997): ***Introduction to random signals and applied Kalman filtering***, John Wiley & Sons, Inc.

Dissanayake G.; Sukkarieh S.; Nebot E.; Durrant-Whyte H. (2001): The aiding of a low-cost strapdown inertial measurement unit using vehicle model constraints for land vehicle applications, ***Institute of Electrical and Electronics Engineers Transactions on Robotics and Automation***, 17(5):731-747

Fraser D. C.; Potter J. E. (1969): The optimum linear smoother as a combination of two optimum linear filters, ***Institute of Electrical and Electronics Engineers Transactions on Automatic Control***, 7(8):387-390

Gelb A. (1974): ***Applied optimal estimation***, The MIT Press

Hofmann-Wellenhof B.; Lichtenegger H.; Collins J. (1994): ***Global positioning system: theory and practice***, Springer-Verlag Wien New York

Kaplan E. D. (1996): ***Understanding GPS principles and applications***, Artech House, Inc.

Leondes C. T.; Peller J. B.; Stear E. B. (1970): Nonlinear smoothing theory, ***Institute of Electrical and Electronics Engineers Transactions on System Science and Cybernetics***, 6(1):63-71

Lewis F. L. (1986): ***Optimal estimation***, John Wiley & Sons, Inc.

Rauch H. E.; Tung F.; Striebel C. T. (1965): Maximum likelihood estimates of linear dynamic systems, ***American Institute of Aeronautics and Astronautics Journal***, 3(8):1445-1450

Siouris G. M. (1993): ***Aerospace avionics systems***, Academic Press, Inc

Titterton D. H.; Weston J. L. (1997): ***Strapdown inertial navigation technology***, Peter Peregrinus Ltd., :385-386

Yu M.-J.; Lee J. G.; Park H.-W. (1999): Comparison of SDINS in-flight alignment using equivalent error models, ***Institute of Electrical and Electronics Engineers Transactions on Aerospace and Electronic Systems***, 35(3):1046-1054

Yu J.; Lee J. G.; Park C. G.; Han H. S. (2002): Design of nonlinear fixed-interval smoother for off-line navigation, ***Journal of Control, Automation and Systems Engineering***, 8(11):984-990

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 95-105

Journal of Global Positioning Systems
2005

Real-time Doppler/Doppler Rate Derivation for Dynamic Applications

Jason Zhang

School of Mathematical and Geospatial Sciences, RMIT University
e-mail: jzhang@eos-aus.com Tel: + 61-02-6222 7949

Kefei Zhang, Ron Grenfell, Yong Li and Rod Deakin

School of Mathematical and Geospatial Sciences RMIT University

Received: 6 November 2004 / Accepted: 15 October 2005

Abstract. Precise GPS velocity and acceleration determination relies on Doppler and/or Doppler rate observations. There are no direct Doppler rate measurements in GPS. Although every GPS receiver measures Doppler shifts, some receivers output only "raw" Doppler shift measurements and some don't output any at all. In the absence of raw Doppler and Doppler rate measurements, a differentiator is necessary to derive them from other GPS measurements such as the carrier phase observations. For real-time dynamic applications, an ideal differentiator should have a wideband frequency response to cover all the dynamics. It should also have a group delay as short as possible. In addition, a low-order differentiator is more favourable for easy implementation.

This paper provides an overview of methods in differentiator design for applications of GPS velocity and acceleration determination. Low-order Finite Impulse Response (FIR) differentiators proposed by Kavanagh are introduced. A class of first-order Infinite Impulse Response (IIR) differentiators are developed on the basis of Al-Alaoui's novel differentiator. For noise attenuation, it is proposed to selectively use Kavangagh's FIR differentiators, and the first-order IIR filters derived for adaptation to different dynamics.

Key words: GPS velocity determination, GPS acceleration determination, differentiator design, FIR filter, IIR filter, Doppler.

1. INTRODUCTION

Previously proposed methods for GPS velocity and acceleration determination fall in two categories. One is to derive velocity and acceleration directly from GPS determined positions, another is based on the Doppler shift method. The latter has several advantages: it doesn't rely on the precision of the positions from GPS, nor will the accuracy dramatically degrade with an increase in sampling rate (say 10Hz or more). Since there is no direct Doppler rate observation in GPS measurements, as a "virtual" observable, it must be derived in order for the formulae presented by Jekeli and Garcia (1997) to be applied directly in the Doppler shift method.

Every GPS receiver measures Doppler shifts. However, this is primarily an intermediate process to obtain accurate carrier phase measurements. Thus the quality of Doppler shift output varies from receiver to receiver depending on manufacturer. The Trimble 5700™ geodetic receiver, for instance, has a measurement precision of ±1mm/s. The observed Doppler is from a tracking loop that is updated at a very high rate. This also enables the receiver to sense phase accelerations (Harvey 2004). Unfortunately the sensed phase acc-elerations and the Doppler shift on L2 are discarded. Some other GPS receivers, for example the Superstar II™ from NovAtel, have only code and L1 phase outputs (SuperstarII 2004) and the Doppler shifts are masked out of the measurements. For our purposes to obtain accurate velocity and acceleration using these types of receivers, it is necessary to derive the Doppler shifts, i.e. the change rates of the carrier phase from the measured carrier phase measurements.

Differentiators are required to get the Doppler rate "observable" for any type of receiver, or to get the Doppler shift from the carrier phase. In real-time and dynamic applications it is also desirable that the designed differentiator should have a wideband frequency response to cover the system dynamics. It should also have a group delay as short as possible so as to get the Doppler shift or Doppler rate instantaneously. For those receivers that output only "raw" Doppler shifts, the derivation of precise Doppler from the carrier phase plays a key role in precise velocity and acceleration determination. This is because the precision of carrier phase observables can be fully exploited. The objective of this paper is to explore the techniques to derive Doppler rate from GPS measurements, or to derive precise Doppler shift from the carrier phase in real time and in dynamic situations.

Several investigations have been conducted for this purpose in the GPS measurement domain, and the proposed methods can be categorised into:

(1) Curve fitting (Fenton and Townsend, 1994);

(2) Kalman smoother/filtering (Hebert, Keith et al. 1997);

(3) Taylor series approximation (Hebert, Keith et al. 1997; Cannon, Lachapelle et al. 1998; Bruton, Glennie et al. 1999);

(4) Finite Impulse Filter (FIR) by using Fourier series with window techniques (Bruton, Glennie et al. 1999); and

(5) FIR optimal design using the Remez exchange algorithm (ibid).

The FIR filtering technique based on Taylor series approximations was recently adopted to derive phase accelerations by Kennedy (2003).

This paper briefly describes the digital differentiator theory and states the design problems in real-time dynamic GPS applications. It is followed by a comprehensive literature review on each method referred to in the above section. By comparing the various differentiator designs, a series of first-order Infinite Impulse Filters (IIR) are presented which are capable of delivering the derivatives from input signals in real-time dynamic situations. An adaptive scheme is also proposed for noise attenuation.

2. Digital filter and digital differentiator Design

2.1. Digital Filtering

Suppose there is a discrete signal sequence of x_n (n is an integer) with a sampling period of T. A digital filter can be regarded as a linear combination of the discrete samples x_{n-k}, together with the previous output y_{n-k}., which can be defined by the following formula (Hamming 1977, p2):

$$y_n = \sum_{k=-\infty}^{k=-\infty} c_k x_{n-k} + \sum_{m=1}^{M} d_k y_{n-m} \tag{1}$$

where y_n is the output of the filter, c_k and d_k are filter coefficients which are referred to as the impulse response of the filter, which is the filter response for a unit input signal pulse (ibid).

The coefficients of a filter completely define the property of the filter and selectively suppress or enhance particular parts of signals. When the coefficients of the second term on the right hand side of Eq. (1) are nonzero, the filter is referred to as a *recursive* filter since the output of y_{n-k} has been used recursively. The filter coefficients c_k and d_k are usually time-invariant in classical filter designs. Their values are carefully chosen to achieve the desired filtering result. However, their values can be assigned on-line to respond to the change of situations in the so-called adaptive filter design. For practical applications, the length of a realisable digital filter is always finite.

2.1.1. Transfer function

The transfer function of a discrete filter is defined as the Z-transform of the filter output signal over the Z-transform of the input signal, i.e.

$$H(z) \equiv \frac{Y(z)}{X(z)} = \frac{\sum_{k=-\infty}^{k=\infty} c_k \cdot Z(x_k)}{1-\sum_{m=1}^{M} d_m \cdot Z(y_m)} = \frac{\sum_{k=-\infty}^{k=\infty} c_k \cdot z^{-k}}{1-\sum_{n=1}^{M} d_k \cdot z^{-n}} \tag{2}$$

where z is a complex variable, and the Z-transform is a linear transform whereby a discrete-time signal value of x_n is defined as

$$Z(x_{n-k}) = X_{n-k}(z) = z^{-k} x_n \tag{3}$$

and where z^{-1} serves as a unit delay operator. The transfer function is most important in filter design and analysis. With the transfer function having been determined, one can directly write out the impulse response of the filter (filter coefficients), and further analyse the performance of the filter either in the time domain or in the frequency domain.

2.1.2. Frequency and amplitude response

The frequency response of a filter is defined as the discrete Fourier transform of output signals over the discrete Fourier transform of input signals

$$H(\omega)=\frac{Y(\omega)}{X(\omega)}=\frac{\sum_{k=-\infty}^{k=\infty} c_k\cdot e^{-j\omega k}}{1-\sum_{m=1}^{M} d_m\cdot e^{-j\omega m}} \tag{4}$$

The above frequency response function is obtained by simply replacing the variable *z* in the transfer function by the Fourier transform variable e^{jw}. The frequency response function allows us to evaluate the frequency response of a filter on the unit cycle.

Factoring the magnitude of the frequency response into the following form

$$\|H(\omega)\|=G(\omega)\cdot e^{j\Theta(\omega)} \tag{5}$$

gives the amplitude response *G(ω)* which is the gain of the filter. The phase response *Θ(ω)* shows the radian phase shift experienced by each sinusoidal component of the input signal. The phase and group delays of a filter give the time delay in seconds experienced by each sinusoidal component of the input signals:

$$\text{phase delay}\equiv\frac{\Theta(\omega)}{\omega}\qquad \text{group delay}\equiv\frac{d}{d\omega}\Theta(\omega) \tag{6}$$

In the case of a filter that has a linear phase response, the group delay and the phase delay are identical, for example when $\Theta(\omega)=\pi/2\cdot\omega$.

2.1.3. Noise amplification

A digital filter is a linear combination of input signals that are usually contaminated by noise. For simplicity we assume that the noise is Gaussian white, and thus the error propagation law applies. This allows us to estimate the noise amplification of the filter. Assume that the noise of a series of L1 carrier phase measurements $x_n=x_n^0+\varepsilon_n$ is Gaussian white, where x_n^0 stands for the true value of x_n, and then the outcome of the finite non-recursive filter is

$$y_n=\sum_{k=-K}^{k=K} c_k x_{n-k}=\sum_{k=-K}^{k=K} c_k x_{n-k}^0+\sum_{k=-K}^{k=K} c_k\varepsilon_{n-k} \tag{7}$$

and the variance of the filter can be evaluated by (Hamming 1977, p14)

$$E\left\{\sum_{k=-K}^{k=K} c_k\varepsilon_{n-k}\right\}^2=\sum_{k=-K}^{k=K} c_k^2 E[\varepsilon_{n-k}]^2=\sigma_x^2\sum_{k=-K}^{k=K} c_k^2 \tag{8}$$

This shows that the sum of the squares of each coefficient of a filter determines the noise amplification of the filtering process.

Supposing that the variance of a recursive filter is $\sigma_{y_n}^2$, and applying the preceding procedures, we have

$$\sigma_{y_n}^2=\sigma_x^2\sum_{k=-\infty}^{k=\infty} c_k^2+\sum_{m=1}^{M} d_{n-k}^2\sigma_{y_{n-m}}^2 \tag{9}$$

Let us further assume that $\sigma_{y_{n-1}}^2=\sigma_{y_{n-2}}^2\cdots=\sigma_{y_{n-M}}^2=\sigma_{y_n}^2$, and then the variance of the filter can be estimated by

$$\sigma_{y_n}^2\approx\frac{\sum_{k=-\infty}^{k=\infty} c_k^2}{1-\sum_{m=1}^{M} d_{n-m}^2}\cdot\sigma_x^2 \tag{10}$$

This indicates that we can either roughly estimate the variance of the recursive filter or “precisely” calculate the filter variance by computing the initial variance of the recursive filter using Eq. (10), and then estimating the variance of the filtered signals using Eq. (9).

2.2. Statement of Problem of Differentiator Design

Differentiator design has been the subject of extensive investigation in digital signal processing. A main issue is that a differentiator amplifies noise at high-frequencies (Carlsson, Ahlen et al. 1991). As GPS signals are of low

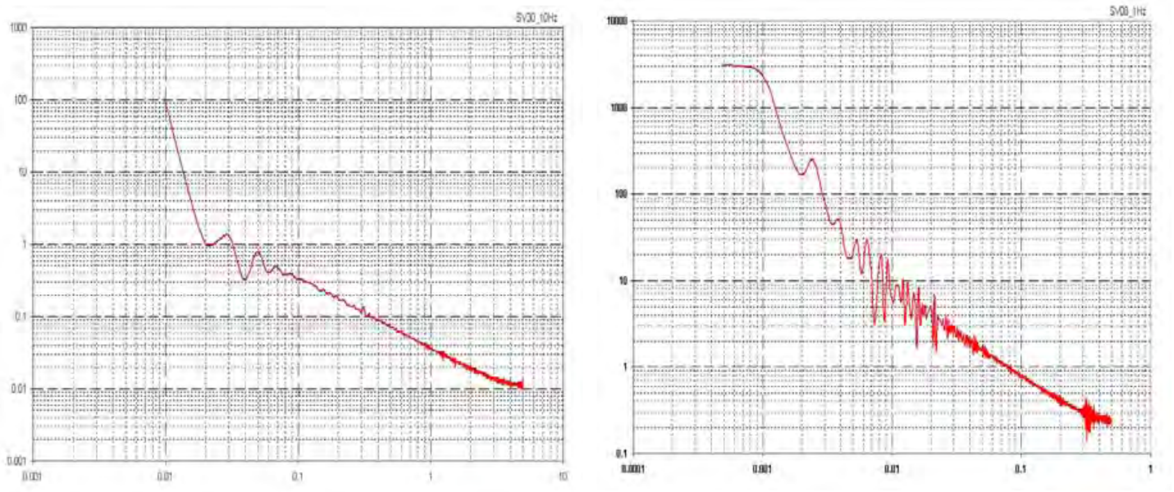


Fig. 1: Power spectral densities for the 1Hz and 10Hz carrier phase signals

frequency character, (see Fig. 1), it is suggested that a low pass filter would be suitable for the design of differentiators. However, the change of dynamics in a system is normally of high frequency. Hence we have to deal with the complicated high frequencies with a broad/full band differentiator. Another complication arises from the signal correlation. It is shown that the GPS carrier signals can be regarded as Gaussian white only when the sampling rate is lower than 1Hz; when the sampling rate goes higher, time correlations must be considered (Bona 2000; Borre and Tiberius 2000). Thirdly, the differentiation may be affected due to lack of information on future signals since the application is real time oriented. Finally there might be aliasing problems due to sampling.

So the problem is to get the derivative from GPS observations where both the signals and noise have random characteristics. In the case of corrupting noise being wideband white and the signal being a Gauss-Markov process (mostlikely for GPS applications), it is apparent that no differentiator is going to be perfect in passing the desired derivative whilst suppressing the noise (Brown and Hwang 1992, p172). This is a typical Wiener filter problem (ibid). The solution is a compromise between good differentiation and low noise sensitivity to achieve a small total error.

The Kalman filter is a space-state solution of the Wiener filter problem (ibid), which is formulated by using the minimum mean-square-error estimation criterion in a two-step recursive procedure. By assuming that both the process driving noise and the measurement noise are Gaussian white and there is no correlation between them, it first predicts the signal state using the system dynamic equation, and then updates the prediction with measurements to get estimates. A successful Kalman filter is subject to proper modelling of system dynamics and the associated stochastic random process. It is suggested that the less than satisfactory performance of the Kalman filter in the case of Heber et al. (1997) is not due to the Kalman filter approach itself, but due to the improper modelling of the system state when it is highly dynamic.

When the sampling rate is high, the theoretical difficulties in Kalman filtering are mainly in the determination of the random process of system driving noise, and the handling of correlations of measurement noise and the cross-correlation between the measurement and signal noises. Another associated practical problem is the heavy computational load in real-time data processing. Finally the outcome of a Kalman filter is a smooth, band-limited solution (Bruton, Glennie et al. 1999). Therefore, it is reasonable to find solutions in the frequency domain rather than in the state space using Kalman filters.

The digital differentiator design oriented in the frequency domain should still consider the variance of the output. Thus the criteria of the differentiator may be summarised as follows:

- the magnitude of frequency response is accurate in low frequencies and is as close to the ideal differentiator $H(\omega) = j\omega$ (Stearns 2003, p127) as possible in a broad band sense depending on the system dynamics;
- the phase response is linear or approximately linear;
- the group delay is acceptably small;
- the sum of the squares of filter coefficients can be minimized; and
- easy to be implemented in real time , i.e. to be causal and low order since there are cycle slips and loss-of-lock of signals.

3. Taylor series approximations

Taylor series approximations have been widely used to derive differentiators. The differentiators used by (Cannon, Lachapelle et al. 1998), Hebert (1997) and Kennedy (2002; 2003) are of low order Taylor series. They are all in the form of central difference approximations such as

$$y_n = \sum_{k=-N}^{N} c_k \cdot x_k \tag{11}$$

where N is the order of Taylor series approximation. Fig. 2 depicts the frequency responses of some low order central difference Taylor series approximations.

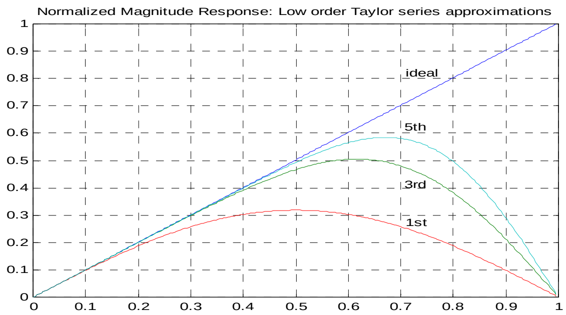


Fig. 2: Frequency response of low order central difference Taylor series approximation

It is apparent that the higher the order, the closer that a Taylor series approximation is to the ideal differentiator. This suggests that broad band differentiators can be designed based on Taylor series, and this can be observed in Khan and Ohba (1999), who gave the explicit coefficients c_k by

$$c_0 = 0$$
$$c_k = c_{-k} = \frac{(-1)^{k+1} N!^2}{k(N-k)!(N+k)!} \tag{12}$$

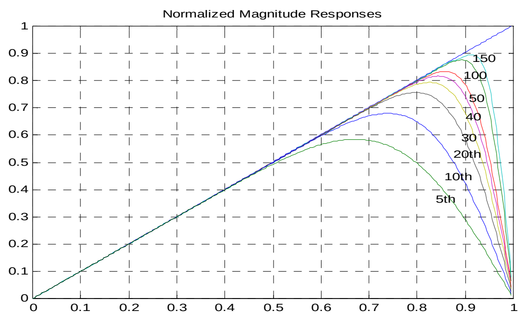


Fig.3: Frequency responses for arbitrary order Taylor series approximations

As can be seen from Fig. 3, this type of differentiator is characterised as having zero amplitude response in both ω=0 and ω =1 (Nyquist frequency). Actually this is the property of type III FIR filters (Chen 2001,p299) which will be discussed later.

4. Curve fitting with window

Jekeli and Garcia (1997) used fifth-order B-splines to derive phase accelerations, and Fenton and Townsend (1994) adopted parabolic functions to obtain the precise Doppler. The referenced curve fitting techniques use sliding windows wherein the data are fitted into polynomials using the least squares approach. The derivative of the central point of a window is obtained by differentiating the polynomials with respect to time accordingly.

Bruton (1999) gave an in-depth review of the curve fitting differentiators. It is concluded that whether a curve fitting uses a polynomial, a parabola, or a cubic spline, the resultant differentiator approaches the ideal only at lower frequencies. Since it is band-limited and lowpass, it is suitable only for low dynamic or static applications. Furthermore, performing the least squares estimation involves intensive computation. Moreover, to obtain the current derivative at t_0, the curve fitting with window requires the input at t_k, which is a signal in the future. Therefore we may conclude that the windowed curve fitting approach is inappropriate for real-time dynamic applications.

5. FIR filters

A Finite Impulse Response (FIR) filter consists of a series of multiplications followed by a summation. The FIR filter operation can be represented by the following equation (Hamming 1977)

$$y_n = \sum_{i=-K}^{K} c_i \cdot x_{n-i} \tag{13}$$

A filter in this form is named FIR because the response to an impulse dies away in a finite number of samples. Note that this form is non-causal and unrealisable. In order to present a causal FIR differentiator, changing the form is required. This leads to

$$y_n = \sum_{i=0}^{N} c_i \cdot x_{n-N} \tag{14}$$

The Fourier series with window are classical in the design of FIR filters where the impulse response is calculated by the inverse discrete Fourier transform of the transfer function, i.e. (Chen 2001)

$$c_d[n] = \frac{1}{2\pi}\int_{\omega=-\pi}^{\omega=\pi} j\omega \cdot e^{-j(n-M)\omega} \cdot d\omega$$
$$= \frac{\cos[(n-M)\pi]}{(n-M)} - \frac{\sin[(n-M)\pi]}{\pi(n-M)^2} \qquad (15)$$

where M=N/2 and the infinite length of Fourier terms is truncated into finite terms. The truncation may cause a discontinuity at the edges of the window and leads to residual oscillations named Gibbs oscillations (ripples in the amplitude response against frequency). Different window methods can be used to smooth the glitches, truncate the filter coefficients, and sharpen the frequency response. Fig.4 gives the comparison between direct truncation and applying the Kaiser window technique.

5.1. Type III FIR Differentiator Design

A FIR filter of type III has an odd length and anti-symmetric impulse response. In this case, the differentiator's coefficients are

$$c_d(n) = \begin{cases} \dfrac{\cos[(n-M)\pi]}{(n-M)} & \text{for} \quad n \neq M \\ 0 & \text{for} \quad n = M \end{cases} \qquad (16)$$

where the *sine* term in Eq. (15) vanishes. To eliminate the Gibbs phenomenon due to the finite truncation, a window function is required. Among many windows that are available, the Kaiser window is most popular. It can be evaluated to any desired degree of accuracy using the rapidly converging series of the zero-order Bessel function of the first kind (Farlex 2004). The ripple of the stopband can also be controlled by an adjustable variable α to meet the optimal criteria given by Kumar and Roy (1988) and Selesnick (2002). With the above procedures, one can also design FIR differentiators with different cut-off frequencies.

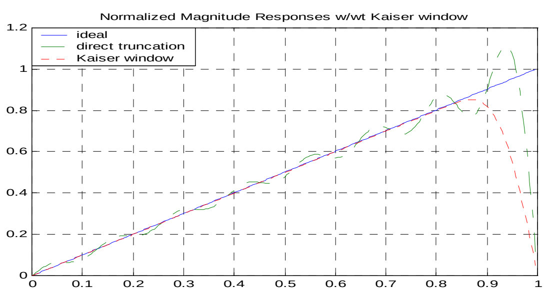


Fig. 4: Magnitude responses of FIR differentiators based on the window technique

Theoretically, FIR filters of type III can be designed to meet requirements at nearly all frequencies, as long as we increase the filter order. However, since the frequency response to the Nyquist frequency is zero, it is impossible to design a full band type III differentiator. Although such filters are causal and are linear in phase, the actual derivative obtained is with respect to time *t-(N/2)T*. This means that the more taps in a FIR filter, the longer the group delay will be. This property of the FIR filter is detrimental to the real-time requirements. However it can be alleviated if the sampling period *T* is small. The difficulty is that increasing the sampling frequency will result in more noisy derivatives. Therefore trade-off and compromise must be made to introduce this type of FIR in real-time applications.

5.2. Type IV FIR Differentiators

Since a FIR filter of type III has the limitation that the amplitude response must go to zero at the Nyquist frequency, it is impossible to get a full band differentiator using a finite number of coefficients. This can be shown in Fig. 3 where transition frequency range of 0.85~1.0 is associated with the 150th order (length of 301) central difference Taylor series differentiator.

A FIR filter of type IV has an even length and anti-symmetric impulse response. The type IV FIR is preferable to a type III as a differentiator in terms of the frequency response. This can be evidenced by the simplest FIR differentiator of $y_n=x_n-x_{n-1,}$ which has a frequency response of

$$H(z) = \frac{Y(z)}{X(z)} = 1 - z^{-1}$$
$$\Rightarrow H(\omega) = 1 - e^{-j\omega} = j \cdot 2\sin\frac{\omega}{2} \cdot e^{-j\frac{\omega}{2}} \qquad (17)$$

The corresponding amplitude response against low order Taylor series approximations is shown in Fig. 5

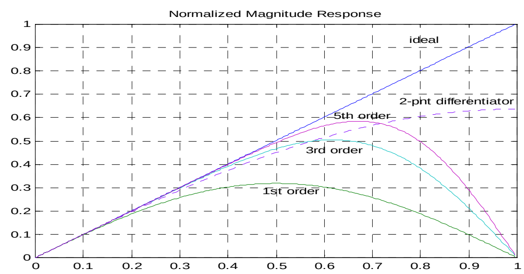


Fig. 5: Frequency response of the simplest IV differentiator against low order Taylor Series FIR filters

It can be seen that even though the differentiator is the simplest form, it is closest to the ideal at low frequencies (<0.2). It has a better amplitude response for the rest of frequency band than its type III counterpart of first-order. It also has a linear frequency response and therefore has a constant group delay at half the sampling period. The type IV FIR differentiators are superior to the type III FIR differentiators in terms of the frequency response, since they have no disadvantageous characteristic of being zero at ω=1.

Details of type IV differentiator design are referred to Chen (2001,p332). An example differentiator of length 8 (7^{th} order) is given with the transfer function of

$$H(z) = -0.0260 + 0.0509z^{-1} - 0.1415z^{-2} + 1.2732z^{-3} - 1.2732z^{-4} + 0.1415z^{-5} - 0.0509z^{-6} + 0.0260z^{-7} \quad (18)$$

which minimizes

$$E = \int_{-\pi}^{\pi} \left| H(e^{j\omega}) - j\omega e^{jN\omega/2} \right|^2 d\omega \quad (19)$$

Therefore it is an optimal differentiator in the sense of least squares with an excellent frequency response at high frequency band. The noise amplification can be calculated from Eq. (10) as $\sigma^2 y_n$=3.2887, which is acceptable so far.

It may be expected that a type IV FIR obtained from the Remez exchange algorithm (Parks and McClellan 1972) would be able to deliver a better performance. This is because the Remez exchange algorithm is a minimax optimal, i.e. minimize {maximum [$H_{ideal}(\omega)$-$H_{disigned}(\omega)$]} for all frequencies, and is more difficult to mathematically compute, but guarantees that the worst case error has been reduced to a quantifiable value. To verify this, the frequency responses have been depicted in Fig. 6 for the 7^{th} and 25^{th}-order filters respectively by the Remez algorithm

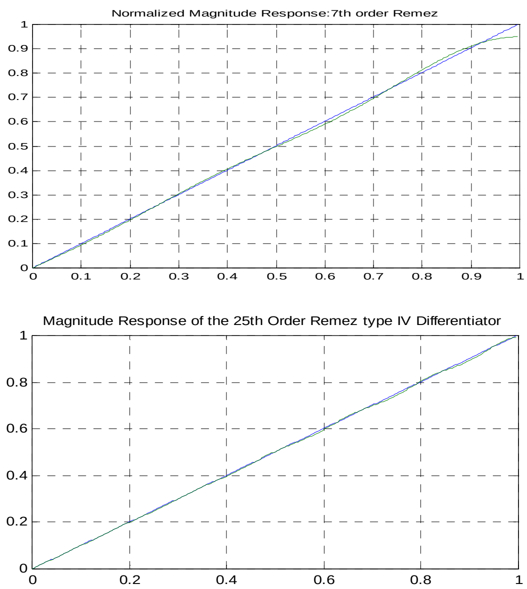


Fig. 6: Frequency response of type IV FIR filters by the Remez exchange algorithm

The FIR filter design by the Remez algorithm is referred to as the equal ripple design. This is because the method can suppress the ripples from the Gibbs phenomenon (Antoniou 1993) to a certain level and turn them into equal ripples in both the passband and stopband.

It seems that type IV FIR differentiators using the Remez exchange algorithm will give us a closing solution. However, the resultant filters provide the first derivative with system biases and higher level of noise.

Type IV FIR differentiators based on Taylor series (Khan, Ohba et al. 2000) have also been tested in this research. It has been found that wideband type IV differentiators are associated with heavy noise amplifications and big biases. Our investigation of type IV FIR filters for differentiator design is still at an early stage and continuing.

5.3. Other FIR Differentiators

In a series of publications, Kumar and Dutta (1988; 1988; 1989; 1989) presented optimal and maximally linear FIR differentiators for low-frequency, mid-frequency, and around specific frequency respectively. They gave the explicit formulae and efficient recursive algorithms to calculate the impulse response of filters. Their contributions are highly appreciated, for example, as the state of art differentiators by Al-Alaoui (1993). In the case of signals that have low frequency components contaminated by wideband noise, FIR differentiators of optimum white-noise attenuation are desired. Kavanagh (2001) investigated the impact of quantization noise on signal from systems with low-frequency rates of change. It is showed that the differentiator proposed by Vainio et al. (1997)

$$h_n = \frac{6(N-1-2n)}{N(N^2-1)} \qquad 0 \le n \le N-1 \quad (20)$$

has an optimum white-noise attenuation and a constant group delay. Kavanagh also proposed a better differentiator for the rate experiencing slow changes

$$h_n = \begin{cases} \dfrac{1}{N-1} & n = 0 \\ 0 & 0 \le n \le N-2 \\ -\dfrac{1}{N-1} & n = N-1 \end{cases} \quad (21)$$

This differentiator has the characteristic of minimising the worst-case error. Clearly when N=2 (type IV), this becomes the simplest two-point differentiator and when N=3 (type III), this turns into the three-point first-order differentiator of a Taylor series approximation.

6. IIR filters

There is another category of filters known as the Infinite Impulse Filter (IIR). A causal IIR filter is represented by

$$y_n = \sum_{k=0}^{N} c_k x_{n-k} + \sum_{m=1}^{M} d_m y_{n-m} \tag{22}$$

where the output signal at a given instant is obtained as the weighted sum of the signal x_{n-k}, and the past outputs of y_{n-m}. As suggested by its name, an impulse input has a response that lasts forever since the output will be recursively used. It is the recursive characteristic that allows IIR filters to be implemented with a lesser order and better performance when compared with FIR filters. Thus IIR filters are attractive for real-time applications.

An IIR filter is unstable if its response to a transient input increases without bound. Poles and zeros are used to analyse the stability of an IIR filter. The poles are the roots of the denominator and the zeros are the roots of the numerator in the transfer function. An IIR filter is stable if and only if, all poles of H(z) are inside the unit circle on the z-plane (Stearns, 2003, p83).

The IIR filter cannot be designed by calculating the impulse response from the known frequency response as is the case in FIR designs. Many IIR filters can be derived from the analogue filter designs and then transformed into the sampled z-plane. Another popular method is the bilinear transform. The IIR differentiator design has been of considerable interest (Rabiner and Steiglitz, 1970). Among various recursive differentiator designs, Al-Alaoui's second order IIR family (1992; 1993; 1994) has been highly acknowledged and widely used, for example (Chen and Lee, 1995).The novel approach of designing digital differentiators by Al-Alaoui is an extension of the method in designing analogue differentiators by using integrators. That is, in the analogue signal processing, differentiators are often obtained by inverting the transfer functions of analogue integrators.

The general procedures to derive the Al-Alaoui family are as follows

- design an integrator that has the same range and accuracy as the desired differentiator;
- invert the obtained transfer function of the integrator;
- reflect the poles that lie outside the unit circle to inside, in order to stabilise the resultant transfer function; and
- compensate the magnitude using the reciprocals of the poles that lie outside the circle.

6.1. Al-Alaoui's First-Order Differentiator

A first-order IIR differentiator was developed by Al-Alaoui (1993) with an effective range 0.78 of the Nyquist frequency based on a non-minimum phase digital integrator. The integrator is a synthesis of the rectangular integrator and the trapezoidal integrator. By assigning weighting factors of ¾ and ¼ to the transfer functions of the integrators respectively, the ideal integrator, which has the following transfer function, is approximated

$$\begin{aligned} H_I(z) &= \frac{3}{4}H_R(z) + \frac{1}{4}H_T(z) \\ &= \frac{3}{4}\cdot\frac{T}{z-1} + \frac{1}{4}\frac{T(z+1)}{2(z-1)} = \frac{T}{8}\cdot\frac{z+7}{z-1} \end{aligned} \tag{23}$$

Reflecting the zero z=-7 with its reciprocal -1/7, and compensating the magnitude by multiplying *r*=7, results in a minimum phase digital integrator with the transfer function

$$H_I(z) = \frac{7\cdot T}{8}\cdot\frac{z+1/7}{z-1} \tag{24}$$

Inverting the above transfer function yields the Al-Alaoui's stabilized IIR differentiator of the first order

$$H_D(z) = \frac{8/7}{T}\cdot\frac{z-1}{z+1/7} \tag{25}$$

The characteristics of this differentiator is shown in Fig. 7. This differentiator is able to approximate the ideal differentiator up to 0.78 of the full band, and has an outstanding "linear phase" response. Al-Alaoui reported that within the effective frequency range, it has a less than 2.0% magnitude error. Since it is of first-order, the delay of the filter is just half of the sample thus it meets every requirement to be used in real-time.

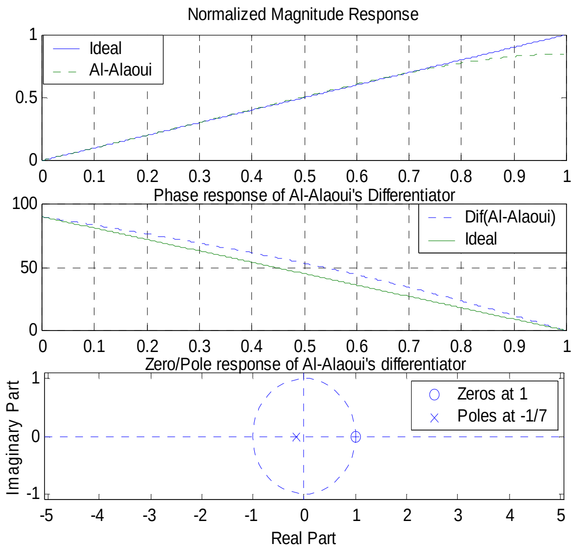


Fig. 7: Characteristics of the first order IIR differentiator

6.2. First-order IIR Differentiator Family

Al-Alaoui contributes the above differentiator as an individual. However, a family of such first order differentiators can be derived following his methodology. That is, while Al-Alaoui designates the weighting factors of ¾ and ¼ empirically, we may get the optimal weights experimentally. To achieve this, a variable α is introduced to adjust the weighting factor in the way of

$$H_I(z) = \alpha H_R(z) + (1-\alpha) H_T(z)$$
$$= \alpha \cdot \frac{T}{z-1} + (1-\alpha)\frac{T(z+1)}{2(z-1)} = \frac{T(1-\alpha)\left[z + \dfrac{1+\alpha}{1-\alpha}\right]}{2(z-1)} \tag{26}$$

where $0<\alpha<1$ serves as a tuner to adjust the integrator so that it better closes to the ideal. α =¾ can be used as a good reference to refine the integrator in the desired range of frequencies. Obviously it has a zero outside the unit circle. Applying Al-Alaoui's procedure to reflect the zero with its reciprocal and to compensate the magnitude, a variable integrator is obtained as

$$H_I(z) = \frac{T(1+\alpha)\left[z + \dfrac{1-\alpha}{1+\alpha}\right]}{2(z-1)} \tag{27}$$

Inverting the transfer function gives a new set of differentiators with transfer functions as

$$H_D(z) = \frac{2(z-1)}{T(1+\alpha)\left[z + \dfrac{1-\alpha}{1+\alpha}\right]} \tag{28}$$

Since $1-\alpha<1+\alpha$, the pole is well located inside the unit circle and the resultant differentiators are, therefore, stable. Setting α=¾ gives the transfer function proposed by Al-Alaoui, and slightly changing α around ¾ results in differentiators which outperform in target bandwidth. The noise amplification of this kind of differentiator can be evaluated using Eq. (10), which is only slightly noisier than the simplest two-point differentiator.

7. Conclusions

The general theory on digital filter design has been introduced. The aim of this research is to find appropriate differentiators that can be used to derive Doppler shifts/Doppler rates from GPS observables in real-time, dynamic applications.

It is concluded that the differentiators obtained from both curve fitting and Kalman filtering require intensive computation and are lowpass. Thus they are not suitable for real-time dynamic applications. Type III FIR differentiators have the inherent nature in frequency response of approaching zero at Nyquist frequency. To extend the performance of type III FIR filters in the higher frequency bands, one has to increase the filter taps. This causes difficulties in managing the data since there are cycle-slips and loss-of-lock signals. It also results in a longer group delay that is detrimental for real-time applications where instant response is desired.

Type IV FIR differentiators using Fourier series have been found to have outstanding frequency response, however, they are noisy and biased. It is found that only the Kavanagh's differentiators of type IV deliver good first derivatives. However, they approximate the ideal differentiator only at low frequencies (lower than 0.2 of Nyquist frequency). Type III FIR filters can be used to derive Doppler/Doppler rate "observables" in post processing mode. Higher order central difference approximations using Taylor series might outperform windowed Fourier series since there is no truncation and the associated Gibbs phenomenon.

It is demonstrated that IIR filters are more favourable for real-time application. Since the outputs of the filter are recursively used, they have much lower orders than the FIR filters. The first-order IIR differentiator from Al-Alaoui is ideal in terms of the frequency response, phase linearity and half sample group delay. The proposed class of first-order IIR differentiators allows us to choose an optimum in the desired frequency range.

It is suggested that Kavanagh's differentiators can be used in static or in constant velocity modes. The proposed first-order IIR differentiators can be adaptively used when systems experience high dynamics.

ACKNOWLEDGEMENTS

The research described in this paper was partially supported by the Cooperative Research Centre (CRC) for microTechnology, Australia endorsed to A/Prof. Kefei Zhang. Partial financial support from the Australia Research Council Linkage Project (LP0455170) endorsed to a research consortium led by A/Prof Kefei Zhang is highly appreciated. Mr. Ravindra Babu at the SNAP, UNSW is highly acknowledged for consultations and discussions.

REFERENCES

Al-Alaoui M.A. (1992): ***Novel Approach to Designing Digital Differentiators***. Electronics Letters 28(15): 1376-1378.

Al-Alaoui M.A. (1993): ***Novel Digital Integrator and Differentiator***. Electronics Letters 29(4): 1376-1378.

Al-Alaoui M.A. (1994): ***Novel IIR Differentiator from the Simpson Integration Rule***. IEEE Transactions on Circuits and System-I Fundamental , Theory and Applications 41(2).

Antoniou A. (1993): ***Digital Filters: Analysis, Design and Appications***. New York, McGraw-Hill 0-07-002121X.

Bona P. (2000): ***Precision, Cross Correlation, and Time Correlation of GPS Phase and Code Observations***. GPS solutions 4(2): 3-13.

Borre K. and Tiberius C. (2000): ***Time Series Analysis of GPS Observables***. proceedings of the ION GPS 2000, Salt Lake City, UT, 1885-1894.

Brown R.G. and Hwang P.Y.C. (1992): ***Introduction to Random Signals and Applied Kalman Filtering***. New York, John Wiley &Sons, INC 0471-52573-1.

Bruton A.M., Glennie C.L. and Schwarz K.P. (1999): ***Differentiation for High-Precision GPS Velocity and Acceleration Determination***. GPS solutions 2(4): 7-21.

Cannon M.E., Lachapelle G., Szarmes M.C., Hebert J.M., Keith J. and Jokerst S. (1998): ***DGPS Kinematic Carrier Phase Signal Simulation Analysis for Precise Velocity and Position Determination***. Journal of the Institute of Navigation 44(2): 231-245.

Carlsson B., Ahlen A. and Sternad M. (1991): ***Optimal Differentiation Based on Stochastic Signal Models***. IEEE Transactions on Signal Processing 39(2): 341-353.

Chen C.-K. and Lee J.-H. (1995): ***Design of High-Order Digital Differentiators using L1 Error Criteria***. IEEE Transactions on Circuits and Systems II: Analogy and Digital Signal Processing 42(4): 287-291.

Chen C.-T. (2001): ***Digital Signal Processing: Spectral Computation and Filter Design***. New York, Oxford, Oxford University Press 0-19-513638-1.

Farlex (2004): ***TheFreeDictionaryCom***: Kaiser Window. http://encyclopedia.thefreedictionary.com/Kaiser%20window.

Fenton P. and Townsend B. (1994): ***NowAtel Communications Ltd,-What's New?*** Proceedings of KIS'94, Banff, AB, CA, 25-29.

Hamming R.W. (1977): ***Digital Filters***. Englewood Cliffs, Prentice-Hall, Inc. 0-13-212571-4.

Harvey N. (2004): ***Doppler and Carrier Phase Measures in Trimble 5700 GPS Receiver***. Christchurch, New Zealand (Personal communication).

Hebert C.J. (1997): ***High Accuracy GPS Velocity Using the Carrier Phase Observable***. Proceedings of KIS'97, Banff, AB, Canada, 265-269.

Hebert C. J., Keith J., Ryan S., Szarmes M., Lachapelle G. and Canon M. E. (1997): ***DGPS Kinematic Carrier Phase Signal Simulation Analysis for Precise Aircraft Velocity Determination***. Proceedings of the ION GPS-97, Albuquerque, NM,

Jekeli C. and Garcia R.. (1997): ***GPS Phase Accelerations for Moving-base Vector Gravimetry***. Journal of Geodesy 71(10): 630-639.

Kavanagh R.C. (2001): ***FIR Differentiators for Quantized Signals***. IEEE Transactions on Signal Processing 49(11): 2713-2720.

Kennedy S. (2002): ***Precise Acceleration Determination from Carrier Phase Measurements***. Proceeding of the ION GPS-2002, Portalnd Oregon, USA, 962-972.

Kennedy S.L. (2003): ***Precise Acceleration Determination from Carrier-Phase Measurements***. Navigation, Journal of the Institute of Navigation 50(1): 9-19.

Khan I.R. and Ohba R. (1999): ***New Design of Full Band Differentiators Based on Taylor Series***. IEEE Proceedings-Vision, Image and Signal Processing 146(4): 185-189.

Khan I. R., Ohba R. and Hozumi N. (2000): ***Mathematical Proof of Explicit Formulas for Tap-coefficients of Full-band FIR Digital Differentiators***. IEE Proceedings-Vision, Image and Signal Processing 147(6): 553-555.

Kumar B. and Dutta S. C. R. (1988): ***Coefficients of Maximally Linear, FIR Digital Differentiators for Low Frequencies***. Electronics Letters 24(9): 563-565.

Kumar B. and Dutta S.C.R. (1988): ***Design of Digital Differentiators for Low Frequencies***. Proceedings of the IEEE 76(3): 287-289.

Kumar B. and Dutta S.C.R. (1989): ***Design of Efficient FIR Digital Differentiators and Hilbert Transforms for Midband Frequency Ranges***. Internal 36(6).

Kumar B. and Dutta S.C.R. (1989): ***Maximally Linear FIR Digital Differentiators for Midband Frequencies***. Int. J. Circuit Theory Appl. 17: 483-488.

Parks T.W. and McClellan J.J. (1972): ***Chebyshev Approximatin for Nonrecursive Digital Filters with Linear Phase.*** IEEE Trans. Circuit Theory 19: 189-194.

Rabiner L. and Steiglitz K. (1970): ***The Design of Wide-band Recursive and Nonrecursive Digital Differentiators***. IEEE Transactions on Audio and Electroacoustics 18(2): 204-209.

Selesnick I.W. (2002): ***Maximally Flat Low-pass Digital Differentiator***. IEEE Transactions on Circuits and Systems II: Analogue and Digital Signal Processing 49(3): 219.

Stearns S.D. (2003): ***Digital Signal Processing with Examples in MATLAB.*** Boca Raton, London, New York, Washington,D.C., CRC press 0-8493-1091-1.

SuperstarII (2004): ***Superstar II Product Brochure***. 2004 http://www.seabed.nl/pdf/superstar.pdf.

Vainio O., Renfors M. and SaramT. (1997): ***Recursive Implementation of FIR Differentiators with Optimum Noise Attenuation***. IEEE Transactions on Instrum. Meas. 46: 1202-1207.

Wolfe J.D., Williamson W.R. and Speyer J.L. (2003): ***Hypothesis Testing for Resolving Integer Ambiguity in GPS. Navigation***, Journal of the Institute of Navigation 50(1): 45-56.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 106-112

2005
© CPGPS
Journal of Global Positioning Systems

Tropospheric Delay Estimation for Pseudolite Positioning

Jianguo Jack Wang, Jinling Wang

School of Surveying and Spatial Information Systems, The University of New South Wales, Sydney, NSW 2052, Australia
e-mail: jianguo.wang@student.unsw.edu.au) Phone: +61-2-9385 4185 Fax: +61-2-9313 7493

David Sinclair, Leo Watts

QASCO Surveys Pty. Limited, 41 Boundary St. South Brisbane, Australia

Hung Kyu Lee

Electronics and Telecommunications Research Institute (ETRI), Rep. of Korea

Received: 15 Nov 2004 / Accepted: 12 July 2005

Abstract. Pseudolites, ground-based GPS signal transmitters, can significantly enhance the GPS satellite geometry or can even be an independent positioning system. However, as pseudolites are very close to the receivers, error effects are different from the traditional GPS and should be considered and modeled in a different way. Tropospheric delay is one of the largest error sources in pseudolite positioning, as pseudolite signal propagates through the lower troposphere which is very difficult to be modeled due to spatial variations in atmosphere. The objective of this research is to analyse pseudolite tropospheric delay modelling methods and to select the optimal tropospheric delay models for different applications.

Several methods to estimate the tropospheric delay for pseudolite positioning are introduced and compared. One approach is to utilize single-differenced GPS tropospheric models. Another one is to compute the tropospheric delay as a function of the local refractivity along the pseudolite signal path. The ratio method used for Electronic Distance Measurement (EDM) can also be applied to estimate tropospheric delay.

Experiments with simulation and real flight test data are conducted in this study to investigate the proposed methods. The advantages and limitations of each method are analysed. The mode defined by RTCA and its modification are suitable for a low elevation and short range application, such as LAAS and local ground based applications. Models derived from single-differenced NMF and Saastamoinen models perform well in long range and high elevation but have a big bias in low elevation. And the model derived from the Hopfield model performs relatively well in all the range and elevation.

Key words: Pseudolite, tropospheric delay, modelling

1 Introduction

GPS has been widely used for precise positioning and navigation applications. However, the accuracy, availability, reliability and integrity of GPS positioning solutions heavily depend on the number and geometric distribution of GPS satellites being tracked. Furthermore, GPS cannot be used without line-of-sight signals between the GPS satellites and a receiver. Pseudolites, which are ground-based GPS-like transmitters, can significantly enhance the satellite geometry or can even construct an independent positioning system.

Compared with GPS satellites, pseudolites are very close to the receivers. Therefore there are many effects that have to be considered and modeled in a different way. Whereas for GPS satellite signals most of the error sources can be eliminated by difference techniques, few of the error sources of the pseudolite signals can be eliminated or mitigated from the analogous approach. The practical solution for pseudolites is to estimate them accurately.

Tropospheric delay can be the largest error source of pseudolite signal as it propagates through the lower troposphere which is very difficult to model due to spatial variations in atmospheric pressure, temperature and humidity. In general, the truth model of tropospheric delay is a function of temperature, atmospheric pressure, relative humidity, elevation angle and range. As atmospheric parameters are normally taken at the reference station, there is always bias when estimating the parameters along the path of signal propagation. This research is to model the tropospheric delay for pseudolite positioning, and to analyze the advantages and limitations of each method by conducting experiments of simulation and real flight data processing.

There are a few tropospheric delay models introduced for pseudolite positioning, such as RTCA model (RTCA, 2000) and Biberger model (Biberger et al., 2003) regarding LAAS (Local Area Augmentation System) for precise aircraft approaching and landing. These two models are suitable for a low altitude measurement, but there are big biases when the altitude is comparable with the scaled heights for the models. The Bouska and Raquet model (2003) and another model derived from the Hopfield model are also investigated in this paper. The above four models are based on a method that computes the tropospheric delay as a function of the local refractivity along the pseudolite signal path.

Another method introduced in this paper is to estimate pseudolite tropospheric delay using single-differenced GPS tropospheric models. As the performance of tropospheric propagation delay prediction models used for GPS typically degrades at very low elevation angles, the models were designed to be used above a limiting elevation angle. Pseudolite tropospheric delay equations derived from single-differenced NMF (Niell Mapping Function) and Saastamoinen models can be applied when the elevation angle is higher than 4 and 10 degrees respectively.

This research reveals the importance of the tropospheric delay modelling for pseudolite positioning, and analyses the performance of the six models proposed for pseudolite tropospheric delay estimation. The advantages and limitations of each model are investigated by analysing simulation results from different aspects. Flight test data were processed to verify the results.

2 Tropospheric Delay Estimation Methods

Common GPS tropospheric delay models are not sufficient for pseudolite positioning. In contrast to the DGPS principle, the signal path from the pseudolite transmitting antenna to the reference station and the signal path from the pseudolite to the rover receiver pass through very different parts of the troposphere. Whereas the distance between pseudolite and reference station is constant, short and the signal runs close to the surface, the distance between pseudolite and rover receiver varies rapidly with time. Even more momentousy, the quantity of tropospheric delay errors is strongly dependent on vertical differences. Therefore, a powerful modeling of the tropospheric effects has to be accurately considered in the pseudolite error model.

Three methods are suggested here to estimate tropospheric delay for pseudolite ranging. One is to compute the tropospheric delay as a function of the local refractivity by integrating local refractivity along the pseudolite signal path. For simplicity, it is called the integration method. Normally, empirical models are employed to represent local refractivity.

Another method utilizes single-differenced GPS tropospheric delay models. The models of GPS tropospheric delay are relatively better developed and can reach very high accuracy. It is reasonable to derive models for pseudolite from them. However, as the performance of GPS tropospheric delay models degrades at very low elevation angles, models derived from these models could have a big bias in a low elevation angle, though they perform well in a long range and high elevation angle. Thus, there is a limited elevation angle of signal slant path for the models derived with single-differenced method.

For short distance measurement at similar altitude, the length ratio method used for Electronic Distance Measurement (EDM) (Rueger, 1996) can be applied to correct pseudolite tropospheric delay. It is assumed that all distance measurements have proportional atmospheric (refractive index) effects. A common scale parameter is assigned to each measurement. No measurements of atmospheric parameters are required. Scales should be adjusted when the receivers are not at the same height.

In this paper the models using the first two methods, integration and single-difference method, are analysed with simulation and flight test data. Four models are applied with the integration method and two models with the single-difference method.

2.1 Models of Integration Method

RTCA has defined the tropospheric delay model for LAAS (RTCA, 2000). The tropospheric correction consists of a dry (hydrostatic) and a wet component.

$$\Delta_{trop} = \Delta_{dry} + \Delta_{wet} \tag{1}$$

The dry and wet components are to be determined separately by Equation (2). The "$*$" in Equation (2) is to be read as the parameter for dry and wet respectively.

$$\Delta_{*} = 10^{-6} \cdot N_{*,0} \cdot R_{rov} \cdot \left(1 - \frac{h_{rov} - h_{PL}}{h_{*,0}}\right) \tag{2}$$

The rover receiver height h_{rov} and the pseudolite height h_{PL} in Equation (2) declare the importance of the vertical distance for tropospheric modeling while R_{rov} is the slope distance between rover receiver and pseudolite. Meteorological data is used to determine the refraction index N_*, which is defined by Equation (3). $h_{*,0}$ is a fixed scaled height for the model which is 42,700m for the hydrostatic component, and 12,000m for the wet component. These heights are arbitrarily defined as the upper boundaries for tropospheric refraction.

$$N_{dry} = 77.6 \cdot \frac{P}{T} \quad \text{and}$$
$$N_{wet} = 22770 \cdot \frac{f}{T^2} \cdot 10^{\frac{7.4475 \cdot (T-273)}{T-38.3}} \tag{3}$$

In Equation (3) T is the temperature, f is the relative humidity and P is the atmospheric pressure sampled on the spot. These parameters then have to be reduced at sea level before they can be used in Equation (2).

Werner (Biberger et al., 2003) proposed some modifications to the RTCA model, which leads Equation (2) to be substituted by Equation (4):

$$\Delta_* = 10^{-6} \cdot N_{*,0} \cdot R_{rov} \cdot \left(1 - 2 \cdot \frac{h_{rov} - h_{PL}}{h_{*,0}} + 2 \cdot \frac{h_{rov}^2 + h_{rov} \cdot h_{PL} + h_{PL}^2}{h_{*,0}}\right) \tag{4}$$

The tropospheric delay model proposed by Bouska and Raquet (2003) is evidently derived from the Hopfield model, with a modified surface height. The N_* in Equation (5) is the refraction index at the height of pseudolite while $N_{*,0}$ in the Equations (2), (4) and (6) is refraction index at sea level.

$$\Delta_* = 2 \cdot 10^{-5} \cdot N_* \cdot R_{rov} \cdot \left(1 - \left(1 - \frac{h_{rov} - h_{PL}}{h_*}\right)^5\right) \cdot \frac{h_*}{h_{rov} - h_{PL}} \tag{5}$$

where $h_* = h_{*,0} - h_{PL}$ and $N_*(h=0) = N_{*,0}$.

Another model directly derived from the Hopfield model is Equation (6), which is an integral from the surface of the sea level (Hofmann-Wellenhof, 2000).

$$\Delta_* = 2 \cdot 10^{-5} \cdot N_{*,0} \cdot R_{rov} \cdot \left(\left(1 - \frac{h_{rov}}{h_{*,0}}\right)^5 - \left(1 - \frac{h_{PL}}{h_{*,0}}\right)^5\right) \cdot \frac{h_{*,0}}{h_{rov} - h_{PL}} \tag{6}$$

It is noticed that Equations (5) and (6) are identical if the Hopfield model is applied for calculating $N_{*,0}$, and different if height-dependent values of pressure, temperature and humidity are used. These four models listed above are based on the integration method that computes the tropospheric delay as a function of the local refractivity along the pseudolite signal path.

2.2 Models of Single-differenced Method

Models derived from the single-differenced GPS tropospheric delay are based on the concept described in Figure 1. The GPS tropospheric delays are calculated from the rover receiver and pseudolite to a GPS satellite in the same line with them using the well-known NMF and Saastamoinen models. The tropospheric delay from the pseudolite to the rover receiver is the difference of the two values.

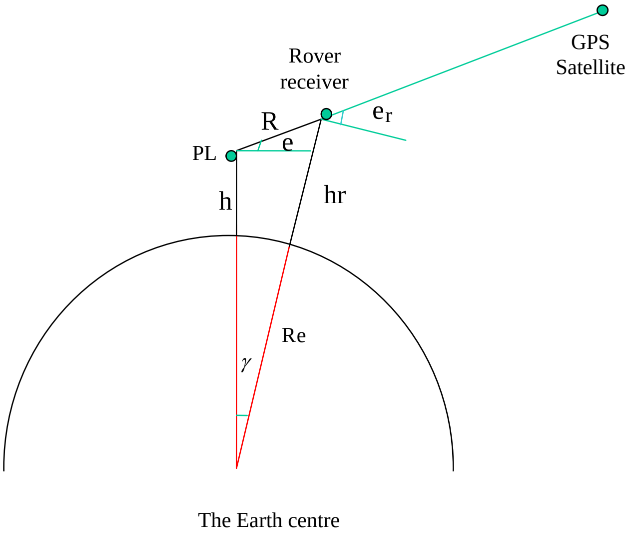


Fig. 1 Concept of single-differenced method

Figure 1 depicts the relation of the positions of pseudolite, the satellite and rover receiver. Equations (7) and (8) are the formulas to calculate the height of the rover receiver and elevation angle from it to the supposed GPS satellite used for NMF or Saastamoinen tropospheric delays models.

$$\begin{aligned} h_r &= \sqrt{(R_e + h)^2 + R^2 - 2 \cdot (R_e + h) \cdot R \cdot \cos(e + \pi/2)} \\ &= \sqrt{(R_e + h)^2 + R^2 + 2 \cdot (R_e + h) \cdot R \cdot \sin e} \end{aligned} \tag{7}$$

and

$$\begin{aligned} e_r &= e + \gamma \\ &= e + \arcsin\left(R \cdot \sin(e + \pi/2) / (R_e + h_r)\right) \\ &= e + \arcsin\left(R \cdot \cos e / (R_e + h_r)\right) \end{aligned} \tag{8}$$

The Saastamoinen model (Remomd, 2000) can precisely estimate GPS tropospheric delay for elevation angle larger than 10 degrees. This model is expressed as:

$$\Delta^{trop} = \frac{0.002277}{\cos Z} \cdot \left[p + \left(\frac{1255}{T} + 0.05 \right) \cdot e - B \cdot \tan^2 Z \right] + \delta R \qquad (9)$$

where ***Z*** denotes the zenith angle and ***p*** is the atmospheric pressure and ***e*** is water vapour pressure in millibar; ***T*** is the temperature in Kelvin. ***B*** and ***δR*** are two correction terms, one being dependent on height of the observing site and the latter on the height and zenith angle. They can be interpolated from the tables provided by a refined model.

The NMF (A.E.Niell, 1996) claims high accuracy estimation for elevation angles larger than 4 degrees. The coefficients in the model depend on the latitude and the height at the observing site and on the day of the year. These two models are applied with the single-differencing method to calculate the tropospheric delay for pseudolites.

3 Simulation Results Analysis

The six models introduced above are programmed under standard atmosphere model (P_0 =1013.25mb, T_0 = 18 C, H_0 = 50%) at sea level. An analysis is conducted by comparing them from different aspects.

3.1 Comparing the Models with Different Elevation Angle

Figures 2 and 3 show the tropospheric delays calculated from the six models in a range of 5km with 0 and 1km reference heights respectively. The elevation angle changes from 0 to 90 degrees. As mentioned above, the values given by the Niell model with elevation angles less than 4 degrees and by Saastamoinen model with elevation angle less than 10 degrees should be ignored. However the figures show that the values given by the Niell model with elevation angle less than 4 degrees do not drift much.

It can be seen from Figures 2 and 3 that the difference of tropospheric delays estimated with different models can reach more than 30cm in a range of 5km. Delay from Hopfield model is larger than that from Bouska model when reference height is higher than zero, and they are identical when the reference height is zero. The delays of Niell, Saastamoinen and Bouska models (and Hopfield model when reference height is zero) are similar when the elevation angle is big, but the estimation of TRCA and modified TRCA models deviate from them at large elevation angles.

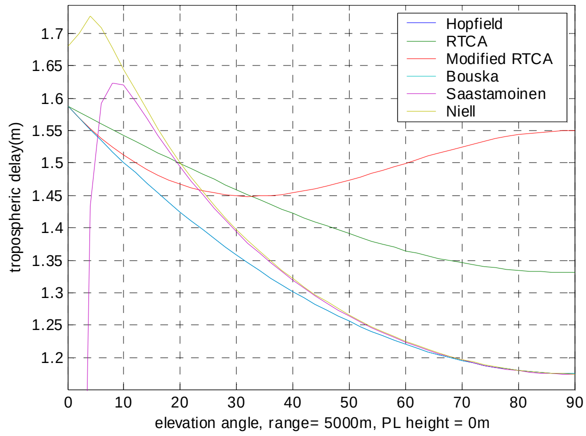


Fig. 2 Changes with elevation angles (h=0m)

The estimations of the Bouska, TRCA and modified TRCA models are the same when the elevation angle is zero. Provided that refractivity at the same height changes little, it is reasonable to believe that the tropospheric delays estimated by models of integration method are accurate at low elevation angle.

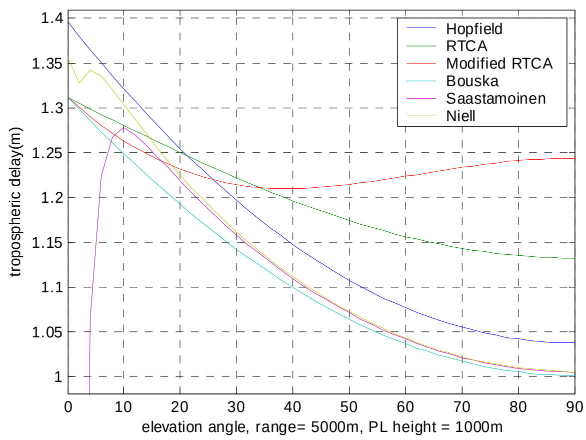


Fig. 3 Changes with elevation angles (h=1000m)

3.2 Comparing the Models with Different Reference Height

Figures 4 and 5 show the tropospheric delays calculated from the six models in a range of 3km with a 10 and 90 degrees elevation angle, respectively. The reference height changes from 0 to 2000 meters.

These figures show that the differences of the estimations between the models become smaller as the reference height increases, especially for a low elevation angle. The estimation of TRCA and modified TRCA model are good with a small elevation angle (Figure 4) but not accurate with a large elevation angle (Figure 5). The estimation of Niell and Saastamoinen models are very similar at different heights and identical at high elevation angles. The line generated by the Hopfield model is less curvy than the others.

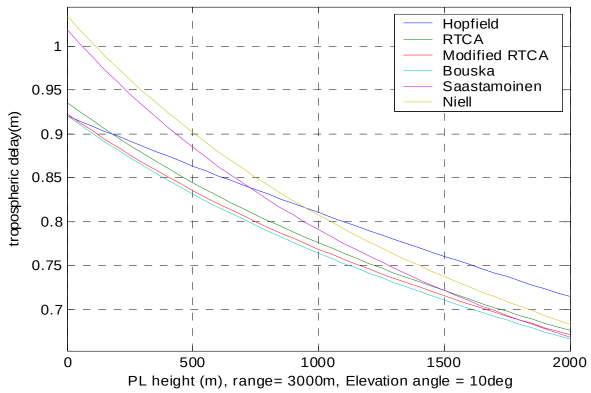


Fig. 4 Changes with reference height (elevation angle is 10 degrees)

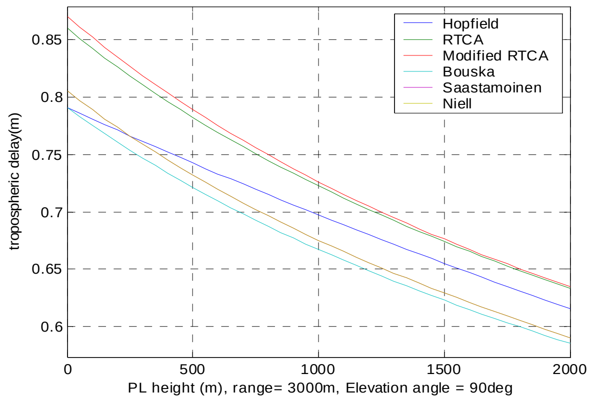


Fig. 5 Changes with reference height (elevation angle is 90 degrees).

In general, it can be concluded from the simulation results that there are observable differences of tropospheric delays estimated with different models. Each model has its strong and weak points. The TRCA and modified TRCA models can be used in the applications with small height difference, such as aircraft landing and land-based applications. However, they are not suitable for large height difference, such as precise airborne geo-referencing. Niell and Saastamoinen models are reliable at high elevation angle but unreliable at very low elevation angle, as shown in Figures 2 and 3. The Hopfield and Bouska models perform relatively stable over the whole range of elevation angle although there is a reference height dependent bias between them. It can be seen from Figures 2 and 3 that the estimation of Bouska model is approaching to that of the RTCA at zero degree and to that of the Niell and Saastamoinen models at 90 degrees. It indicates that this model should have the smallest bias among all the models in the whole range of elevation angle.

4 Flight Test Data Analysis

Flight tests were conducted in April 2003 at the Wedderburn Airfield, Sydney, Australia. Figure 6 shows the sky plot of the GPS satellites and Figure 7 shows the relative height, distance and horizontal trajectory of the aircraft during the period used for data processing.

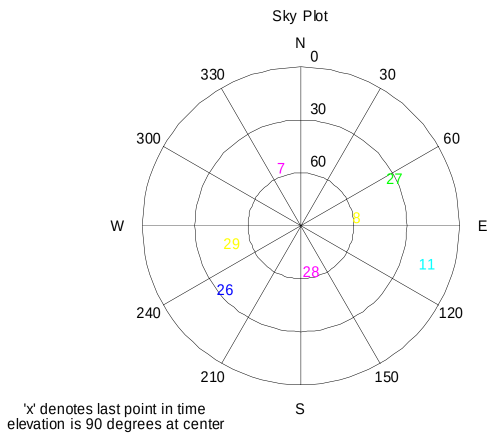


Fig6 Sky plot of GPS satellites

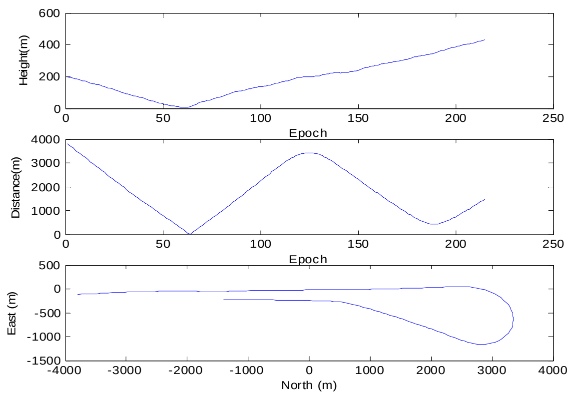


Fig. 7 Flight trajectory used for processing

The pseudolite signal was generated by a Spirent Communications GSS4100P single-channel signal generator pulsing at a 1/11 duty cycle, with a 10MHz oven-controlled crystal oscillator frequency reference. The reference station consisted of a NovAtel Millennium receiver with Leica AT504 choke-ring antenna. Note that use of the choke-ring was for the mitigation of GPS/pseudolite multipath. The airborne system comprised two GPS/pseudolite receivers (NovAtel Millennium) and two antennas. The two GPS antennas were mounted in the aircraft head cone. One was upward-looking and the other downward-looking to track GPS and pseudolite signals. The raw GPS carrier phase measurements from the receivers were processed using an in-house software package; a modified version of the AIMSTM navigation processing software (Lee et al., 2003).

The measurement accuracy cannot be directly evaluated in the kinematic mode without an accurate reference trajectory. Alternatively, a comparison with the independent trajectory obtained by carrier phase dual-frequency DGPS post-processing using the GrafNav/GrafNet software and the double-differenced (DD) residuals computed from GPS/PL-predicted rover GPS antenna positions are used to analyse the performance of different models of pseudolite tropospheric delay.

For a short baseline, the double difference of carrier phase measurement in units of meters can be modelled as

$$\nabla\Delta\Phi = \nabla\Delta\rho + \Delta d_{trop}^{PL} - \Delta d_{trop}^{GPS} + \lambda\nabla\Delta N \qquad (10)$$

The single-differenced pseudolite troposphere delay can be computed based on the double difference of carrier phase measurement (Fukushima et al., 2004)

$$\nabla\Delta\Phi - \nabla\Delta\rho - \Delta d_{trop}^{PL} + \Delta d_{trop}^{GPS} = \lambda\nabla\Delta N \qquad (11)$$

As the GPS troposphere delay is modelled accurately for high elevation angles, Δd_{trop}^{GPS} is treated errorless for the reference GPS satellite. The integer ambiguity is fixed if no cycle slip occurs. If no cycle slip occurred, the left side of Equation (11) should approximately maintain the same value as in the top of Figure 8, which is the carrier phase double difference of two GPS satellites. It is noticed that the double difference values have bias from an integer of cycle, which may be due to multipath effects.

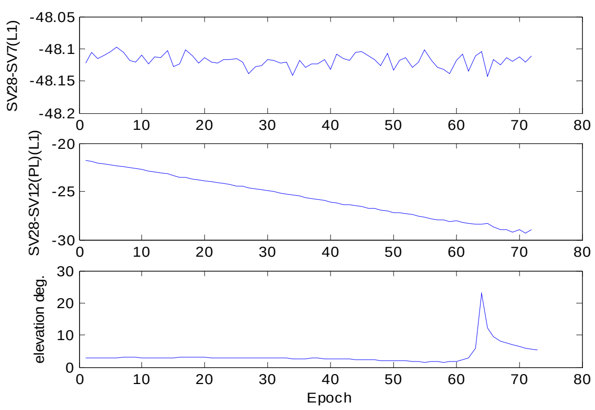


Fig. 8 Carrier phase double difference results and elevation angles

The middle of Figure 8 shows the double differences of the pseudolite and the reference satellite (SV28) without tropspheric delay applied to pseudolite measurements. It has a big drift (about 8 cycles) from the first to the last epoch even though no cycle slip occurred. This indicates that pseudolite tropospheric delay modelling affects the double difference results heavily. The bottom of Figure 8 depicts the elevation angle from the pseudolite to the rover receiver, which is an important parameter for the Saastamoinen and NMF models using the single-differenced method.

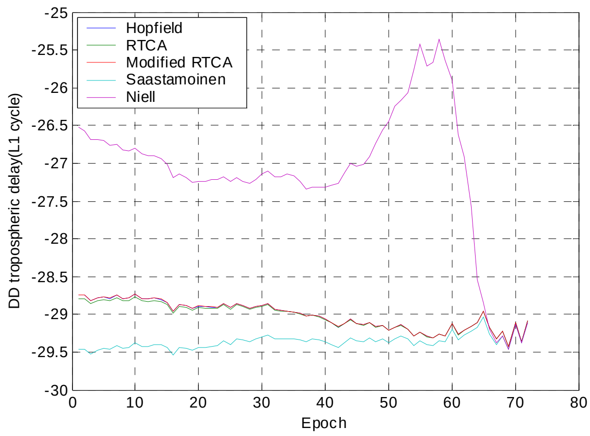


Fig. 9 Carrier phase double difference results of various tropospheric delay models

Figure 9 shows the carrier phase double difference results by applying the pseudolite tropospheric delay models introduced above. The results of the Saastamoinen model vary violently during most epochs. This means that the modelling is inaccurate. As shown in the bottom of Figure 8, the elevation angles from the pseudolite to the rover receiver are less than 5 degrees during most epochs, except the last few ones. The simulation results in Section 3 indicate that the Saastamoinen model cannot estimate correctly below 10 degrees. The results of the Saastamoinen model in the last few epochs become close to other models as the elevation angles increasing.

The results of RTCA, modified RTCA and Hopfield models are almost the same in Figure 9. This agrees with the simulation results in Figure 2, where the estimations of these models are the same when the elevation angle is zero.

The result of the NMF model is the best one among all the models tested as it almost keeps the same value in Figure 9, even if the elevation angle is very small in the flight test. This also agrees with the simulation results in Figures 2 and 3.

Compared to the GPS DD results in the top of Figure 8, the GPS\PL DD results in Figure 9 are more fluctuating. This may be due to the serious multipath effect of pseudolite signals, which is not only sensitive to the attitude of the aircraft as for GPS signals, but also the relative positions of the aircraft and the pseudolite.

5 Closing Remarks

This research reveals the importance of the tropospheric delay modelling for pseudolite positioning, and presents the performances of the six models proposed for pseudolite tropospheric delay estimation. Simulations were conducted for comparing their performances from different aspects and flight test data were processed to verify the simulation results.

It can be concluded from the simulation results that there are some differences between the tropospheric delays estimated with different models. Each model has its strength and weakness. The TRCA and modified TRCA models can be used in the applications with small height differences, such as aircraft landing and land-based applications. However, they are not suitable for these with large height difference, such as precise airborne geo-referencing. Niell and Saastamoinen models are reliable when the elevation is above their limited elevation angles. The Hopfield and Bouska models perform relatively stable for the whole range of elevation angle though there is a reference height dependent bias between them. The Bouska model should have the smallest bias among all the models in the whole range of elevation angle.

The flight test results conform some of the conclusions from the simulation results. The single-differenced method proposed in this paper is effective to estimate the pseudolite tropospheric delay by employing GPS tropospheric models. It is found that the result of the NMF model is the best one among all the models tested even if the elevation angle is very small in the flight test. However, as the flight test data does not cover the range used in the simulation, the other conclusions from the simulation results should be further tested.

Acknowledgements

This study is supported by Australia Research Council Linkage project LP0349297: “Integration of GPS/PL/INS to Geo-reference Airborne Surveying and Mapping Sensors”.

References

Biberger R J.; Teuber A.; Pany T.; Hein G W. (2003): ***Development of an APL Error Model for Precision Approaches and Validation by Flight Experiments***. In: ION (Editor), GPS/GNSS, Portland, OR, pp. 2308-2317.

Bouska C T J.; Raquet J F. (2003): ***Tropospheric Model Error Reduction in Pseudolite Based Positioning Systems***, ION GPS/GNSS 2003, Portland OR, USA, pp. 390-298.

Fukushima S.; Yoshihara T.; Suga S. (2004): ***Evaluation of a Tropospheric Correction Model fpr Airport Pseudolite***, ION GPS/GNSS 2004, Long beach CA, USA.

Hofmann-Wellenhof B.; Lichtenegger H.; Collins J. (2000): ***GPS Theory and Practice***. Fifth revised edition. Springer-Verlag Wien, New York, 382 pp.

Lee H K.; Wang J.; Rizos C.; Park W (2003): ***Carrier phase processing issues for high accuracy integrated GPS/Pseudolite/INS systems***, Proceedings of 11th IAIN World Congress, Berlin, Germany, paper 252.

Niell A E. (1996): ***Global mapping functions for the atmosphere delay at radio wavelengths.*** Journal of Geophysical Research, 101(B2): 3227-3246.

RTCA (2000): ***GNSS Based Precision Approach Local Area Augmentation System (LAAS)-Signal-in-Space Interface Control Document (ICD)***. RTCA/DO-246A, Radio Technical Commission for Aeronautics.

Rueger J M. (1996): ***Electronic Distance Measurement, An Introduction***. Springer, 276 pp.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 113-119

Journal of Global Positioning Systems
2005
© CPGPS

Online Stochastic Modelling for Network-Based GPS Real-Time Kinematic Positioning

Jinling Wang[1), Hung Kyu Lee[2), Young Jin Lee[1), Tajul Musa[1), Chris Rizos[1)

[1)School of Surveying and Spatial Information Systems, The University of New South Wales, Sydney, Australia
e-mail: Jinling.Wang@unsw.edu.au, Tel: 61-2-9385-4203, Fax: 61-2-9313-7493

[2)Telematics Research Division, Electronics and Telecommunications Research Institute, 161 Gajeong-dong, Yuseong-gu, Daejeon 305-350 Korea
e-mail: hkyulee@etri.re.kr, Tel: 82-42-860-1748, Fax: 82-42-860-1611

Received: 16 November 2004 / Accepted: 8 July 2005

Abstract. Baseline length-dependent errors in GPS RTK positioning, such as orbit uncertainty, and atmospheric effects, constrain the applicable baseline length between reference and mobile user receiver to perhaps 10-15km. This constraint has led to the development of network-based RTK techniques to model such distance-dependent errors. Although these errors can be effectively mitigated by network-based techniques, the residual errors, attributed to imperfect network functional models, in practice, affect the positioning performance. Since it is too difficult for the functional model to define and/or handle the residual errors, an alternative approach that can be used is to account for these errors (and observation noise) within the stochastic model. In this study, an online stochastic modelling technique for network-based GPS RTK positioning is introduced to adaptively estimate the stochastic model in real time. The basis of the method is to utilise the residuals of the previous segment results in order to estimate the stochastic model at the current epoch. Experimental test results indicate that the proposed stochastic modelling technique improves the performance of the least squares estimation and ambiguity resolution.

Key words: Stochastic Modelling, GPS, and Network-Based RTK

1. Introduction

Carrier phase-based GPS has become an essential technique for a wide range of precise positioning applications, such as kinematic surveying and vehicle navigation and guidance. The 'relative' positioning concept, in which one GPS receiver's position is derived with the aid of observations made at a stationary reference station, has been used in most implementations of the technique. The current state-of-the-art implementation for high precision GPS positioning is the so-called Real-Time Kinematic (RTK) hardware/software system that can deliver centimetre-level positioning accuracies of a moving (i.e. kinematic) user receiver, *if the integer carrier phase ambiguities are correctly resolved*. However, the impact of baseline length-dependent GPS errors, such as orbit uncertainty, and atmospheric effects, constrains the applicable baseline length between reference and mobile user receiver to perhaps 10-15km. These constraints have led to the development of several network-based RTK techniques, including the Virtual Reference Station (VRS) approach, and the Area Correction Parameter techniques (Wübbena et al., 2001)

The key concept in using multiple reference stations (i.e., a network) to support GPS carrier phase-based positioning is: (a) to generate so-called "correction terms" representing the distance-dependent errors; and (b) to interpolate and apply them to mobile user receiver measurements, so as to significantly diminish the residual error terms, thus making it possible to perform medium or long-range (tens to hundred kilometre receiver separations) positioning. The correction message data are generated based on the pre-determined coordinates of the reference stations. Integer ambiguities among the reference stations must be correctly resolved and station-dependent errors, such as multipath, measurement noise, and antenna phase centre variation, should be mitigated in the correction terms. In addition, the generated correction message data with respect to each reference stations have to be interpolated or modelled for the user's location. Over the past decade, a number of interpolation methods have been proposed. These include *Linear Combination*

Model, *Distance-Based Linear Interpolation Method*, *Linear Interpolation Method*, *Lower-Order Surface Model* and *Least-Square Collocation* (Fotopoulos & Cannon, 2001). However, Dai et al (2001) demonstrated that the performances of all of these methods are similar.

In order to obtain optimal estimates from the least-square solution, both a *mathematical model*, also called a functional model, and a *stochastic model* should be correctly defined. The functional model describes the relationship between measurements and unknown estimates. On the other hand, the stochastic model represents the statistical characteristics of the measurement that is mainly provided by the covariance matrix for the measurements. Whilst the mathematical models for the network-based GPS RTK positioning are sufficiently investigated and well documented (e.g., Han and Rizos, 1996; Raquet, 1997; Wanninger, 1995; Wübbena et al., 1996), stochastic modeling is still an issue under investigation. In the case of network-based GPS RTK, the positioning performance is largely affected by the residual biases due to imperfect network mathematical models (Musa et al., 2004). The residual biases contribute to the noise terms and make it difficult to define a functional model that can deal with them. Hence, they should be taken into account within the stochastic model.

In this paper a stochastic modelling method that can be efficiently implemented for network-based GPS RTK positioning will be introduced. This technique determines the covariance matrix of observations at the current epoch based on the estimation residuals from the previous positioning results. Experimental test results will be presented to demonstrate the application of the stochastic modelling method.

2. Online stochastic modelling

A stochastic model for the network-based GPS RTK has to take into account uncertainties due to: (a) residual vectors at reference stations; (b) residual interpolation; (c) measurements at the mobile receiver. An online stochastic model reflecting all the uncertainties can be derived using the residuals from Kalman Filter solutions (Wang, 2000).

Since the true values of the model errors (measurement noise or process noise) are unknown, stochastic modelling has to be based on the filtering residuals of the measurements and state corrections, which are generated in the process of parameter estimation. A problem here is that the parameter estimation process itself relies on the estimated measurement covariance matrix R and process covariance matrix Q. The process covariance matrix can be fixed to the appropriate values to neglect the impact of the dynamic model on the GPS RTK solutions. Then, Kalman filtering results are in fact identical to the least-squares solutions (see e.g., Wang, 2000). An *adaptive* procedure can be used to estimate the matrix R online.

The basic idea of the adaptive method is that the residuals collected from the previous segment of positioning results are used to estimate the covariance matrices of the measurement noise and process noise for the current epoch. Preset default covariance matrices are needed to seed the adaptive estimation process. Within a segment, the covariance matrices for each epoch are assumed to be the same. Therefore, the formulation of an adaptive stochastic modelling method includes two critical steps, namely: (a) to derive suitable formulae for use in estimating the covariance matrices, and (b) to determine the optimal segment (window width), which is application-dependent (Wang, 2000; Dai, 2002).

Suppose the measurement filtering residuals are:

$$v_{z_k} = z_k - H_k \hat{x}_k \tag{1}$$

where z_k is the measurement vector and H_k is the measurement design matrix. Equation (1), obviously, is the optimal estimator of the measurement noise level because the estimated values $\hat{x}_k$ (not the predicted values $\bar{x}_k$) of the state parameters are used in their computations. In order to obtain the covariance matrix of the measurement filtering residuals, equation (1) is further derived as

$$v_{z_k} = z_k - H_k(\bar{x}_k + G_k d_k) = (E - H_k G_k) d_k \tag{2}$$

where G_k is the gain matrix and d_k is the innovation vector. By applying the error propagation law to equation (2), after extensive computations, one obtains

$$Q_{v_{z_k}} = R_k - H_k Q_{\hat{x}_k} H_k^T \tag{3}$$

In equation (3), if the covariance matrix $Q_{v_{z_k}}$ is computed using the measurement filtering residuals from the previous m epochs, the covariance matrix R_k can be estimated as (Wang, 2000)

$$\begin{aligned} \hat{R}_k &= \hat{Q}_{v_{z_k}} + H_k Q_{\hat{x}_k} H_k^T \\ &= \frac{1}{m} \sum_{i=0}^{m-1} v_{z_{k-i}} v_{z_{k-i}}^T + H_k Q_{\hat{x}_k} H_k^T \end{aligned} \tag{4}$$

which can be used in the computation of epoch $k+1$. In equation (4), m is called *the width of moving windows*. It is noted that the covariance matrix $\hat{R}_k$ estimated with equation (4) is always positive definite because it is the sum of the two positive definite matrices. Equation (4) requires some extra computations for both v_{z_k} and $H_k Q_{\hat{x}_k} H_k^T$, which are not generated by the standard Kalman filtering process. Fortunately, the amount of these additional computations is small and leads no significant time delay in data processing.

The initial covariance matrix is determined by using the previous covariance matrix. Based on the residuals at the previous epochs, the covariance matrix of the observations can be estimated in real-time using Equation (4).

3. Testing results

3.1 The Experiments

Field experiments were carried out on 21st October 2004 at Olympic Park in Sydney. The objective of these experiments was to test performance of the proposed stochastic modelling method. Three reference stations were used in the experiments. Whilst two of them are Continuously Operating Reference Stations (CORS), which belong to *"SydNet"* - a network of GPS reference stations in the Sydney metropolitan area (Rizos et al., 2003), the other one is a permanent GPS station on the roof of the Electrical Engineering Building at The University of New South Wales. Since MGRV was selected as a master reference station, the baseline length between reference and mobile receiver was approximately 32km. The locations of the reference stations and the trajectory of the mobile receiver are illustrated in Figures 1 and 2. During the fifteen minutes of the experiment, seven satellites were tracked. The data interval was one second. In addition, one more GPS station was installed in the experimental area to generate a reference trajectory using short-range data processing.

The acquired data was processed in the post-mission mode using in-house GPS kinematic processing software. However, it should be noted that all the algorithms used are applicable for real-time implementation. The network-based GPS positioning begins with generating network corrections to mitigate baseline distance-dependent errors (see Musa et al., 2004). The generated L1 and L2 carrier phase corrections for the test data are depicted in Figure 3. It can be seen from the figure that the baseline distance-dependent errors are at the few centimetre-level. This is due to the factor that the baseline length in this experiment was about 32km and the test site is geographically located in a middle latitude region where the ionospheric delay effect is relatively small. In the data processing, the initial standard deviation for the pseudo-range and carrier phase measurements was defined as 0.3m and 0.05 cycles, respectively. The width of the moving window for estimating the measurement covariance matrix was set to ten epochs and the initial estimate was calculated using measurement residuals of the least squares from the first ten epochs with ambiguities being fixed.

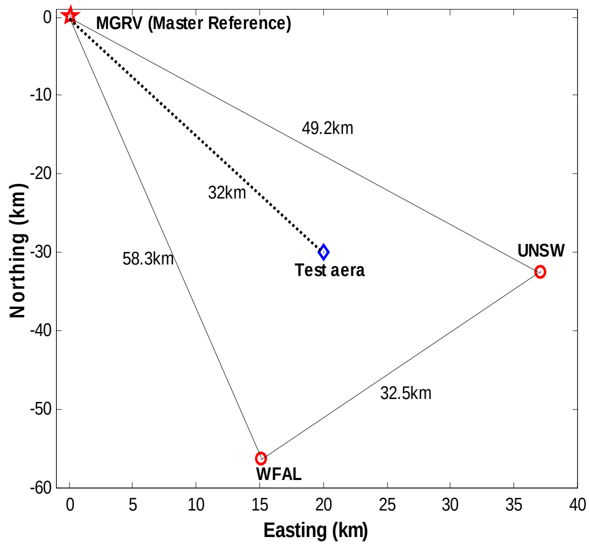


Fig. 1 Configuration of the reference stations and the roving station

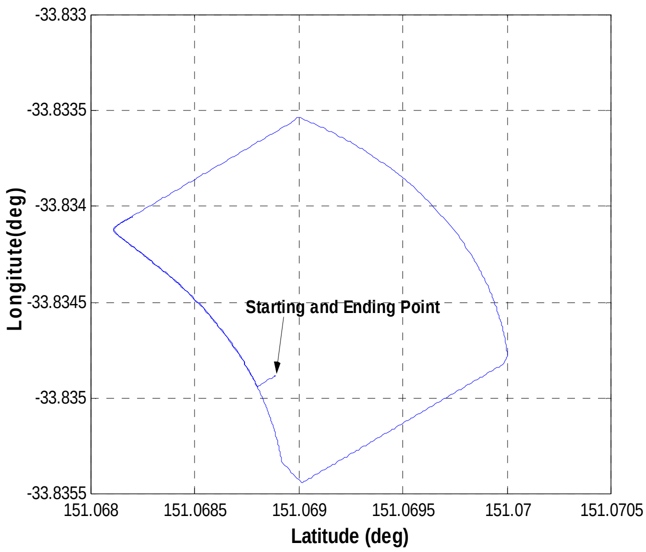


Fig. 2 Trajectory of mobile receiver

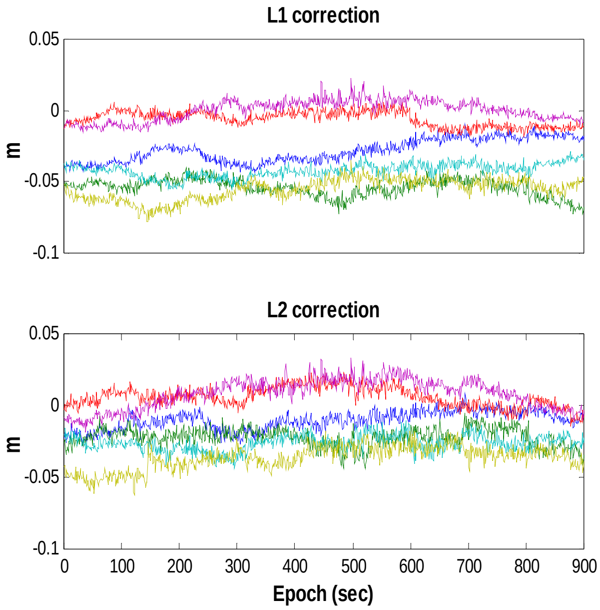


Fig. 3 Network corrections for L1 and L2 carrier phases measurements

3.2 Testing Results

First of all, a comparison between single and multiple reference station solutions was made in order to study potential benefits of using the additional reference stations in the carrier phase-based GPS kinematic positioning. Figures 4 and 5 depict the standard deviations of L1 and L2 carrier phases showing the precision (e.g., noise level) of these measurements and the positioning accuracy.

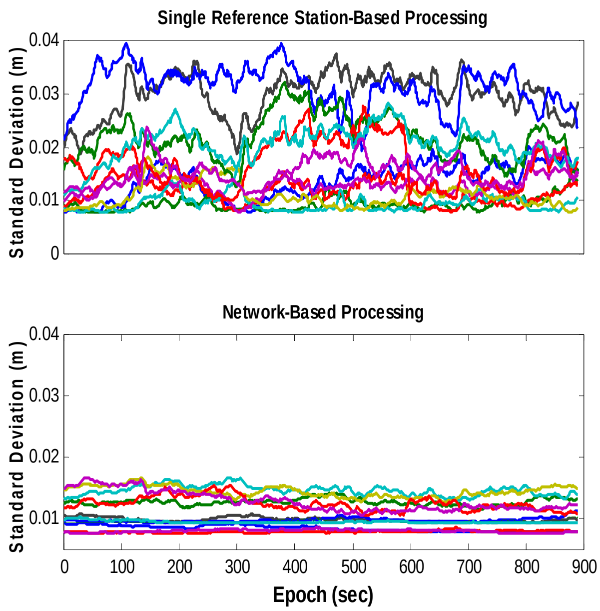


Fig. 4 Standard deviation of L1 and L2 carrier phases measurements

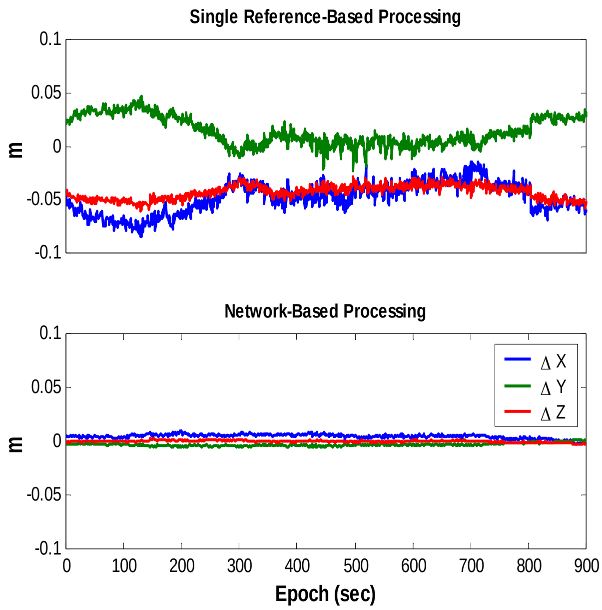


Fig. 5 Accuracy of positioning solution

As already mentioned, the additional reference GPS station installed within the test area with baseline length approximately 100m provided the reference trajectory since its accuracy can be at the few centimetre level with correct integer ambiguity (e.g., short-range kinematic positioning). Hence, it is possible to use the trajectory for evaluating positioning accuracy of the test results. It can be recognised from these results that the application of the network corrections significantly reduces the measurement errors and consequentially improves positioning accuracy, demonstrating the main advantage of the GPS network for medium and long baseline kinematic positioning.

In order to clearly demonstrate effectiveness of the online stochastic modelling method for network-based GPS kinematic positioning, a comparison of the online modelling method with a model based on an apriori assumption of the measurement precision was made. For convenience, they are referred to as ‘Preset’ (based on the error propagation law – see Musa et al., 2004) and ‘Estimated’ (i.e., Equation 4). Figure 6 shows the aposteriori variance value changes obtained from the two different stochastic models. The values should have unity according to the least squares estimation theory (Cross, 1983) if both the functional and stochastic models are correctly defined. In other words, if the variance is significantly different from unity, it is suspected that outliers exist in the measurements, or there is a problem with the fidelity of the stochastic and/or functional models (ibid). The figure shows that the variance values from the preset model are more variable than those from the estimated model. This may be due to the fact that the stochastic model does not realistically reflect the residual biases of the measurements which are mainly caused by the functional model uncertainty. On the other hand, the estimated results show that the variance values are stable and very close to ‘1’, demonstrating that the residuals are appropriately considered by the introduced online modelling method.

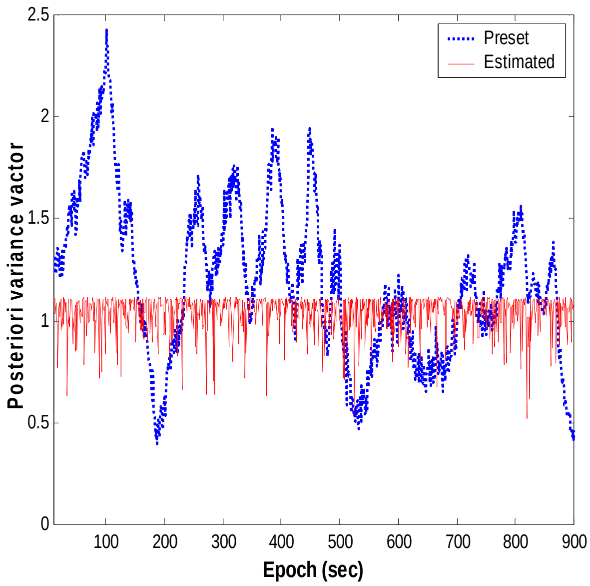


Fig. 6 Posterior variance changes

Figures 7 and 8 illustrate the standard deviations for the DD C1 code and L1 carrier phase observations, respectively. It is evident that big differences exist between the two stochastic models. In reality, the accuracy of the measurements may be influenced by many factors, which must be considered by the appropriate stochastic model. As mentioned early, there are three factors causing the uncertainty (i.e., residual biases) of GPS observations after applying the corrections in the network-based approach. Even though it goes without saying that the uncertainty should be variable due to the satellite and receiver dynamics, the results with the preset stochastic model are almost constant values, which are unrealistic. In contrast, the fluctuations in the estimated results using the proposed on-line stochastic modelling method indicate that such uncertainties are reflected in the stochastic model, making them more realistic.

The determination of the integer ambiguities, commonly referred to as ambiguity resolution (AR), is the most critical data analysis step for high precision GPS based positioning. With fixed integer ambiguities, the carrier phases can be used as unambiguous precise range measurements. A realistic estimation of measurement covariance matrices can provide reliable statistics for ambiguity resolution. To demonstrate this more clearly, both the solutions with the preset and estimated covariance matrices were generated. The ADOP (Ambiguity Dilution of Precision) measure defined by Teunissen and Odijk (1997) is designed to describe the impact of the receiver-satellite geometry on the precision and the correlation of ambiguity parameters. The calculated ADOP values are depicted in Figure 9, indicating significant improvement of the precision and the correlation of the estimated float ambiguities. As a consequence of this it is expected that the ambiguity search volume is reduced and its shape becomes more like a sphere, speeding up the ambiguity searching process (ibid).

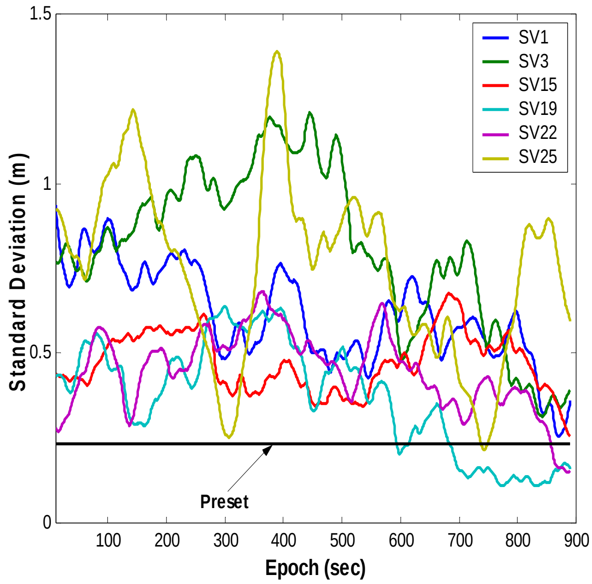


Fig. 7 Standard deviation of C1 pseudo-range measurements

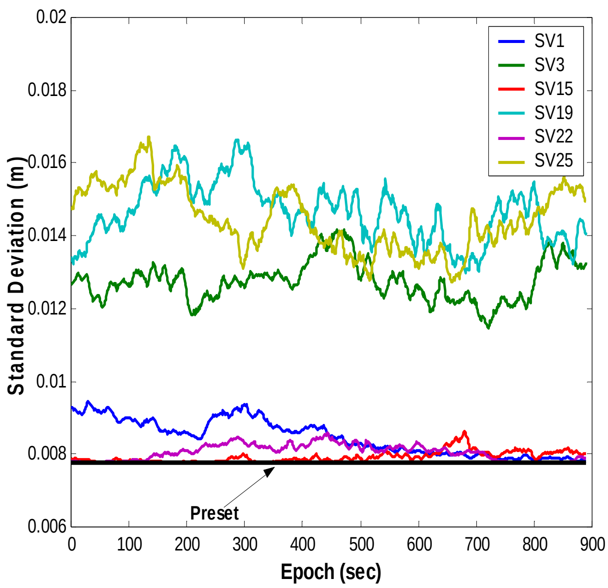


Fig. 8 Standard deviation of L1 carrier phase measurements

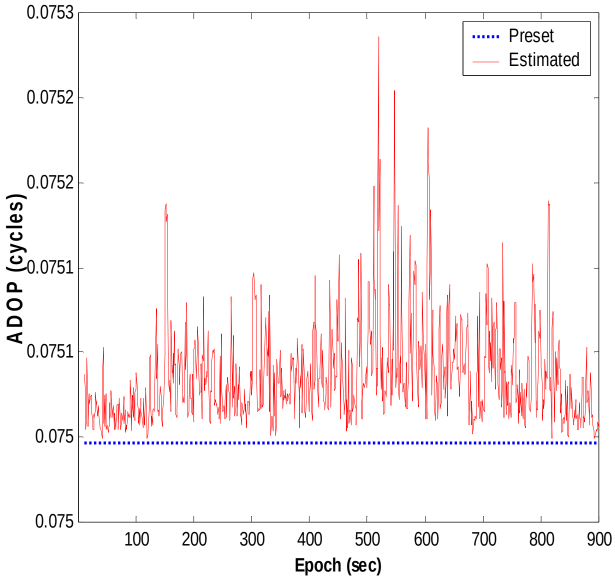


Fig. 9 Ambiguity Dilution of Precision (ADOP)

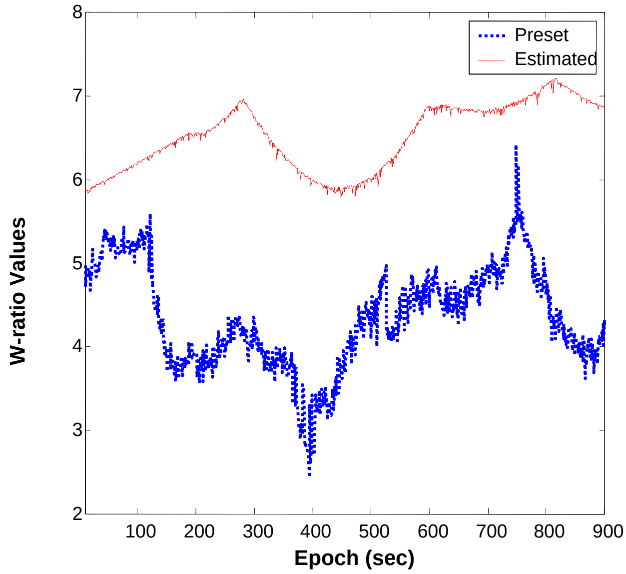


Fig. 10 W-ratio values

It is crucial to ensure that the most likely integer ambiguity combination is statistically better than the second best combination, as defined by the second minimum quadratic form of the least squares residuals, the so-called the *'ambiguity validation test'*. The improved float ambiguity estimates are of great importance for the ambiguity validation test (Wang, 2000; Wang et al., 2003; Lee, 2004). First of all, in order to check the correctness of the best ambiguity combination, the reference ambiguity combination was computed from the reference trajectory obtained from the short-range positioning. A comparison of the best combination with the reference one indicated their ambiguities were exactly the same, hence being able to be considered as the correct ambiguity set. Therefore, it is anticipated that the larger validation test statistics, the higher the probability of validating the correct ambiguity combination.

Figure 10 depicts the validation test statistics using the *W*-ratio proposed by Wang et al. (1998). These results shown in Figure 10 indicate that the ambiguity validation test statistics with the 'Estimated' measurement covariance matrices are much larger than those of the preset measurement covariance matrices. Consequentially, the best ambiguity combination has more probability to be validated by the test, making it possible to reliably resolve correct ambiguities through avoiding type I errors in the statistical hypothesis test.

4. Concluding remarks

Although the impact of baseline length-dependent GPS errors, such as orbit uncertainty, and atmospheric effects, constrains the applicable baseline length between reference and mobile user receiver to perhaps 10-15km, the development of the network-based approaches makes it possible to overcome this constraint. However, the positioning performance is largely affected by the residual biases due to imperfect network mathematical models. These residual biases contribute to the noise terms and make it difficult to define a functional model that can deal with them.

In this paper, an online stochastic modelling method that reflects all the uncertainties of the network-based GPS RTK positioning has been introduced. This stochastic modelling method estimates the covariance matrix of observations at the current epoch based on the estimation residuals from the previous positioning results. In addition, field experiments were carried out to evaluate the performance of the modelling technique. Test results indicate that the proposed technique improves: (a) the covariance matrix of the observations; (b) the model fidelity of the least squares estimation; and (c) the performance and reliability of the ambiguity resolution for network-based GPS kinematic positioning.

References

Dai L, Han S, Wang J, and Rizos C (2001) ***A study on GPS/GLONASS multiple reference station techniques for precise real-time carrier phase positioning***, 14^{th} Int. Tech.

Meeting of the Satellite Division of the U.S. Inst. of Navigation, Salt Lake City, UT, 11-14 September, 392-403

Dai L (2002) ***Augmentation of GPS with GLONASS and pseudolite signals for carrier phase-based kinematic positioning***, PhD Thesis, The University of New South Wales, Australia, pp.175

Fotopoulos G, and Cannon ME (2001) ***An overview of multi-reference station methods for cm-level positioning***, GPS Solutions, 4(3), 1-10

Han S, and Rizos C (1996) ***GPS network design and error mitigation for real-time continuous array monitoring systems***, Proceedings of 9th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation, Kansas City, Missouri, 17-20 September, 1827-1836

Lee HK (2004) ***Integration of GPS/Pseudolite/INS for high precision kinematic positioning and navigation***, PhD Thesis, The University of New South Wales, Australia, pp.221

Musa T, Wang J and Rizos C (2004) ***A stochastic modelling method for network-based GPS positioning***, Proceeding of European GNSS Conference 2004, Rotterdam, The Netherlands, 16-19 May, CD-Rom proc.

Raquet J (1997) ***Multiple user network carrier-phase ambiguity resolution***, Proceedings of Int. Symp. On Kinematic Systems in Geodesy, Geomatics and Navigation (KIS 1997), Banff, Canada, 3-6 June, 45-55.

Teunnissen PJG, Odijk D (1997) ***Ambiguity dilution of precision: definition, properties and application***, Proceedings of 10th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation, Kansas City, Missouri, 16-10 September, 891-900

Wang J, Stewart MP and Tsakiri M (1998) ***A discrimination test procedure for ambiguity resolution on the fly***, Journal of Geodesy, 72, 644-653

Wang J (2000) ***Stochastic modelling for RTK GPS/Glonass positioning***, Navigation, Journal of the US Institute of Navigation, 46(4), 297-305

Wang J, Lee HK, Hewitson S, Rizos C, and Barnes J (2003) ***Sensitivity analysis for GNSS integer carrier phase ambiguity validation test***, Presented in XXIIIth General Assembly of the IUGG, Sapporo, Japan, 30 June – 11 July, CD-Rom proc.

Wanninger L (1995) ***Improved ambiguity resolution by regional differential modelling of the ionosphere***, Proceedings of Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation, San Diego, California, 12-15 September, 55-62.

Wübbena G, Bagge A, Seeber G, Böder V, and Hankemeier P (1995) ***Reducing distance dependent errors for real-time precise DGPS applications by establishing reference station networks***, Proceedings of 9th Inst. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation, Kansas City, Missouri, 17-20 September, 1845-1852.

Wübbena G, Bagge A, and Schmitz M (2001) ***RTK networks based on Geo++ GNSMART – concepts, implementation, results***, Proceedings of 14th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation, Salt Lake City, UT, 11-14 September,368-377

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 120-128

Journal of Global Positioning Systems
2005

6DoF SLAM aided GNSS/INS Navigation in GNSS Denied and Unknown Environments

Jonghyuk Kim and Salah Sukkarieh[1]

Department of Engineering, The Australian National University, ACT200, Australia
e-mail: jonghyuk.kim@anu.edu.au Tel: +61 (0)2 6125 2462; Fax: +61 (0)2 6125 0506,

[1]ARC Centre of Excellence for Autonomous Systems, The University of Sydney, Australia
e-mail: salah@acfr.usyd.edu.au Tel: +61 (0)2 9351 8515; Fax: +61 (0)2 9351 7474

Received: 6 December 2004 / Accepted: 12 July 2005

Abstract. This paper presents the results of augmenting 6DoF Simultaneous Localisation and Mapping (SLAM) with GNSS/INS navigation system. SLAM algorithm is a feature based terrain aided navigation system that has the capability for online map building, and simultaneously utilising the generated map to constrain the errors in the on-board Inertial Navigation System (INS). In this paper, indirect SLAM is developed based on error analysis and then is integrated to GNSS/INS fusion filter. If GNSS information is available, the system performs feature-based mapping using the GNSS/INS solution. If GNSS is not available, the previously and/or newly generated map is now used to estimate the INS errors. Simulation results will be presented which shows that the system can provide reliable and accurate navigation solutions in GNSS denied environments for an extended period of time.

Keywords: 6DoF SLAM, low cost GNSS/INS, vision sensor, features mapping, UAV

1 Introduction

The Global Navigation Satellite System (GNSS) is a space-borne, radio navigation system. In airborne navigation, its complementary characteristics to the Inertial Navigation System (INS) make it an excellent aiding system, resulting in integrated GNSS/INS navigation systems. In UAV application, due to its limited payload capacity and accurate navigation requirement to guide and control the vehicle, the low-cost, light-weighted, and compact-sized GNSS/INS system has been focused significantly.

The main drawback in the cost-effective GNSS/INS system is that the integrated system becomes more dependent on the availability and quality of GNSS information. Unfortunately, GNSS information can be easily blocked or jammed by intentional/unintentional interference. Even a short duration of satellite signal blockage can degrade the navigation solution significantly as shown in Kim 2004.

In this paper, a new concept of terrain-aided navigation, known as Simultaneous Localisation and Mapping (SLAM) is considered to aid INS during GNSS denied situations. SLAM was firstly addressed in the paper by Smith and Cheeseman, 1987 and has evolved from the robotics research community to explore unknown environments, where absolute information is not available (Dissanayake and *et al* 2001, Guivant 2001, Williams and *et al* 2001). Contrary to the exiting terrain aided navigation system, SLAM does not require any pre-surveyed map database. It builds the map incrementally by sensing environment and uses the map to localise the vehicle simultaneously, which results in a truly self-contained autonomous navigation system.

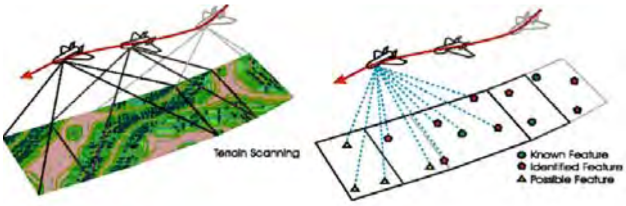


Fig. 1 The overall structure of SLAM is about building a relative map of feature using relative observations, defining a map, and using this map to localise the vehicle simultaneously.

The nonlinear 6DoF SLAM algorithm, incorporating IMU as its core dead-reckoning sensor, was firstly demonstrated in paper by Kim and Sukkarieh, 2004. Its airborne application is described in Figure 1. The vehicle starts navigation from an unknown location and an unknown environment. The vehicle navigates by using its dead-reckoning sensor or vehicle model. As the onboard sensors detect features from the environment, the SLAM estimator augments the feature locations to a map in some global reference frame and begins to estimate the vehicle and map states together with successive observations. The ability to estimate both the vehicle location and the map is due to the statistical correlations that exist within the estimator between the vehicle and features, and between the features themselves. As the vehicle proceeds through the environment and re-observes old features, the map accuracy converges to a lower limit, which is a function of the initial vehicle uncertainty when the first feature was observed (Dissanayake and *et al*, 2001). In addition, the vehicle uncertainty is also constrained simultaneously.

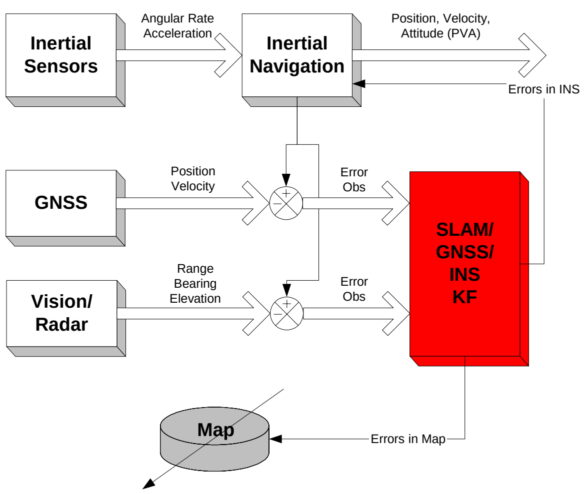


Fig. 2 The architecture of SLAM augmented GNSS/INS system

In this paper the 6DoF SLAM algorithm is augmented to the GNSS/INS navigation system to provide reliable INS aiding in GNSS denied environment. Figure 2 presents the architecture of the SLAM augmented GNSS/INS system. The key feature in this approach is the complementary fusion structure, which has high-speed INS module and low-speed and computationally expensive SLAM/GNSS/INS filter. To achieve this, the nonlinear SLAM algorithm was re-casted into linearised error state form as in the work of Kim, 2004, then it is augmented to fusion filter. In this architecture, the INS and map is maintained outside the SLAM filter and the map is treated as external map database. The fusion filter works as either feature-tracking filter or SLAM filter depending on the availability of GNSS observation. If GNSS provides reliable observations, then the on-board terrain observations are used to build the feature map and SLAM/GNSS/INS filter estimates the errors in INS and map, which results in a feature (or target)-tracking system. However, in GNSS-denied situation, the terrain observations are solely used to estimate the errors in INS and map, which results in SLAM mode operation. Although there are no global observations from GNSS, the constant re-observation and revisit processes can improve the quality of map and navigation performance.

Section 2 will present the external INS loop and map and Section 3 will formulate the error model of SLAM/GNSS/INS algorithm and Kalman filter structure. In Section 4, simulation results are provided based on our Brumby UAV, then Section 5 will provide conclusions with future work.

2 External INS loop and map

In the complementary SLAM/GNSS/INS structure, the SLAM filter aids the external INS loop in a complementary fashion. The inertial navigation algorithm is to predict the high-dynamic vehicle states from the Inertial Measurement Unit (IMU) measurements. In this implementation a quaternion-based strapdown INS algorithm formulated in earth-fixed tangent frame is used (Kim, 2004):

$$\begin{bmatrix} \mathbf{p}^n(k) \\ \mathbf{v}^n(k) \\ \mathbf{q}^n(k) \end{bmatrix} = \begin{bmatrix} \mathbf{p}^n(k-1) + \mathbf{v}^n(k-1)\Delta t \\ \mathbf{v}^n(k-1) + [(\mathbf{q}^n(k-1) \otimes \mathbf{f}^b(k)) \otimes (\mathbf{q}^n)^*(k-1) + \mathbf{g}^n]\Delta t \\ \mathbf{q}^n(k-1) \otimes \Delta\mathbf{q}^n(k-1) \end{bmatrix} \tag{1}$$

where $\mathbf{p}^n(k), \mathbf{v}^n(k), \mathbf{q}^n(k)$ represent position, velocity, and quaternion respectively at discrete time k, Δt is the time for the position and velocity update interval, $(\mathbf{q}^n)^*(k)$ is a quaternion conjugate for the vector transformation, $\otimes$ represents a quaternion multiplication, and $\Delta\mathbf{q}^n(k)$ is a delta quaternion computed from gyroscope readings during the attitude update interval.

3 Complementary SLAM/GNSS/INS Algorithms

The mathematical framework of the SLAM algorithm is based on an estimation process which, when given a kinematic/dynamic model of the vehicle and relative observations between the vehicle and features, estimates the structure of the map and the vehicle's position, velocity and orientation within that map. In this work, the Kalman Filter (KF) is used as the state estimator.

3.1 Augmented Error State

In complementary SLAM, the state is now defined as the error state of vehicle and map:

$$\boldsymbol{\delta}\mathbf{x}(k) = [\boldsymbol{\delta}\mathbf{x}_v(k), \boldsymbol{\delta}\mathbf{x}_m(k)]^T \tag{2}$$

The error state of the vehicle $\boldsymbol{\delta}\mathbf{x}_v(k)$ comprises the errors in the INS indicated position, velocity and attitude expressed in the navigation frame:

$$\boldsymbol{\delta}\mathbf{x}_v(k) = [\boldsymbol{\delta}\mathbf{p}^n(k), \boldsymbol{\delta}\mathbf{v}^n(k), \boldsymbol{\delta}\boldsymbol{\psi}^n(k)]^T \tag{3}$$

The error state of the map $\boldsymbol{\delta}\mathbf{x}_m(k)$ also comprises the errors in 3D feature positions in the navigation frame. The size of state is also dynamically augmented with the new feature error after the initialisation process,

$$\boldsymbol{\delta}\mathbf{x}_m(k) = [\boldsymbol{\delta}\mathbf{m}_1^n(k), \boldsymbol{\delta}\mathbf{m}_2^n(k), \cdots, \boldsymbol{\delta}\mathbf{m}_N^n(k)]^T, \tag{4}$$

where N is the current number of registered features in the filter and each state consists of a 3D position error.

3.2 SLAM Error Model

The linearised SLAM system in discrete time can be written as

$$\boldsymbol{\delta}\mathbf{x}(k+1) = \mathbf{F}(k)\boldsymbol{\delta}\mathbf{x}(k) + \mathbf{G}(k)\mathbf{w}(k) \tag{5}$$

where $\boldsymbol{\delta}\mathbf{x}(k)$ is the error state vector, $\mathbf{F}(k)$ is the system transition matrix, $\mathbf{G}(k)$ is the system noise input matrix and $\mathbf{w}(k)$ is the system noise vector which represents the instrument noise with any un-modelled errors with noise strength $\mathbf{Q}(k)$.

The continuous time SLAM/Inertial error model is based on misalignment angle dynamics and stationary feature model which is a random constant (Kim, 2004):

$$\begin{bmatrix} \boldsymbol{\delta}\dot{\mathbf{p}}^n \\ \boldsymbol{\delta}\dot{\mathbf{v}}^n \\ \boldsymbol{\delta}\dot{\boldsymbol{\psi}}^n \\ \hline \boldsymbol{\delta}\dot{\mathbf{m}}_m^n \end{bmatrix} = \begin{bmatrix} \boldsymbol{\delta}\mathbf{v}^n \\ \mathbf{C}_b^n\mathbf{f}^b \times \boldsymbol{\delta}\boldsymbol{\psi}^n + \mathbf{C}_b^n\boldsymbol{\delta}\mathbf{f}^b \\ -\mathbf{C}_b^n\boldsymbol{\delta}\boldsymbol{\omega}^b \\ \hline \mathbf{0}_m \end{bmatrix}, \tag{6}$$

where $\mathbf{f}^b$ and $\boldsymbol{\omega}^b$ are acceleration and rotation rates measured from IMU, $\boldsymbol{\delta}\mathbf{f}^b$ and $\boldsymbol{\delta}\boldsymbol{\omega}^b$ are the associated errors in IMU measurement, $\mathbf{C}_b^n$ is the direction cosine matrix formed from the quaternion. The discrete-time model can now be obtained by expanding the exponential state transition function and approximating it to the first-order term, and integrating the noise input during discrete sample time (Δt), which result in,

$$\mathbf{F}(k) = \left[\begin{array}{ccc|c} \mathbf{I} & \Delta t\mathbf{f}^n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \Delta t\mathbf{f}^n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{m\times m} \end{array}\right] \tag{7}$$

$$\mathbf{G}(k) = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \sqrt{\Delta t}\mathbf{C}_b^n(k) & \mathbf{0} \\ \mathbf{0} & -\sqrt{\Delta t}\mathbf{C}_b^n(k) \\ \hline \mathbf{0}_m & \mathbf{0}_m \end{bmatrix} \tag{8}$$

$$\mathbf{Q}(k) = \begin{bmatrix} \boldsymbol{\sigma}_{\delta f}^2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma}_{\delta\omega}^2 \end{bmatrix} \tag{9}$$

with $\boldsymbol{\sigma}_{\delta f}$ and $\boldsymbol{\sigma}_{\delta\omega}$ representing noise strengths of acceleration and rotation rate respectively.

3.3 Observation model

The linearised observation model can be obtained in terms of the observation residual, or measurement differences, $\boldsymbol{\delta}\mathbf{z}(k)$ and the error states, $\boldsymbol{\delta}\mathbf{x}(k)$,

$$\boldsymbol{\delta}z(k) = \mathbf{H}(k)\boldsymbol{\delta}\mathbf{x}(k) + \mathbf{v}(k) \tag{10}$$

with $\mathbf{H}(k)$ being the linearised observation Jacobian and $\mathbf{v}(k)$ being the observation noise with noise strength matrix $\mathbf{R}(k)$. The error observations are generated by subtracting the measured quantity, $\tilde{\mathbf{z}}(k)$, from the INS predicted quantity $\hat{\mathbf{z}}(k)$,

$$\boldsymbol{\delta}z(k) = \boldsymbol{\delta}\hat{z}(k) - \boldsymbol{\delta}\tilde{z}(k). \tag{11}$$

As there are two different types of observation in this system, that is a range/bearing/elevation observation and a GNSS position/velocity observation, they should be formulated separately.

3.3.1 Range/Bearing/Elevation observation

In range/bearing/elevation observation, the onboard sensor provides relative observations between vehicle and features. The non-linear observation equation relates these observations to the state as

$$\mathbf{z}(k) = \mathbf{h}(\mathbf{x}(k), \mathbf{v}(k)), \tag{12}$$

where $\mathbf{h}(\cdot)$ is the non-linear observation model at time k, and $\mathbf{v}(k)$ is the observation noise vector. Since the observation model predicts the range, bearing, and elevation for the i-th feature, it is only a function of the i-th feature and the vehicle state. Therefore Equation 8 can be further expressed as

$$\mathbf{z}_i(k) = \mathbf{h}(\mathbf{x}_v(k), \mathbf{x}_{mi}(k), \mathbf{v}_i(k)), \tag{13}$$

with $\mathbf{z}_i(k)$ and $\mathbf{v}_i(k)$ being the i-th observation and its associated additive noise in range, bearing and elevation with zero mean and variance of $\mathbf{R}(k)$. The feature position in the navigation frame is initialised from the sensor observation in the sensor frame and vehicle state as shown in Figure 3.

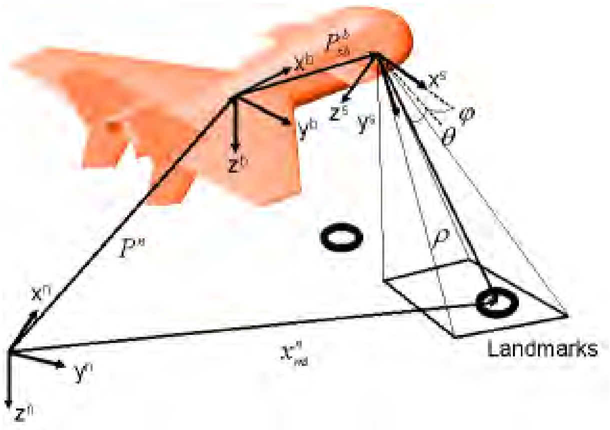


Fig. 3 The range, bearing and elevation observations from the onboard sensor can be related to the location of the feature in the navigation frame through the platform's position and attitude.

The initial feature position in the navigation frame is then computed

$$\mathbf{m}_i^n(k) = \mathbf{p}^n(k) + \mathbf{C}_b^n(k)\mathbf{p}_{bs}^b + \mathbf{C}_b^n(k)\mathbf{C}_s^b\mathbf{p}_{sm}^s(k) \tag{14}$$

where $\mathbf{p}_{bs}^b(k)$ is the lever-arm offset of the sensor from the vehicle's centre of gravity in the body frame, $\mathbf{C}_s^b$ is a direction cosine matrix which transforms the vector in the sensor frame (such as camera instalment axes) to the body frame, and $\mathbf{p}_{sm}^s(k)$ is the relative position of the feature from the sensor expressed in the sensor frame which is computed from the observation:

$$\mathbf{p}_{sm}^s(k) = \begin{bmatrix} \rho\cos(\varphi)\cos(\vartheta) \\ \rho\sin(\varphi)\cos(\vartheta) \\ \rho\sin(\vartheta) \end{bmatrix}, \tag{15}$$

with ρ, φ and ϑ being the range, bearing and elevation angle respectively, measured from the onboard sensor. Hence the predicted range, bearing and elevation between the vehicle and the i-th feature in Equation 8 can now be obtained by rearranging Equation 10,

$$\mathbf{z}_i(k) = \begin{bmatrix} \rho \\ \varphi \\ \vartheta \end{bmatrix} = \begin{bmatrix} \sqrt{x^2+y^2+z^2} \\ \tan^{-1}(y/x) \\ \tan^{-1}(z/\sqrt{x^2+y^2}) \end{bmatrix}, \tag{16}$$

where,

$$\begin{aligned} [x \quad y \quad z]^T &= \mathbf{p}_{sm}^s(k) \\ &= \mathbf{C}_b^s\mathbf{C}_n^b(k)[\mathbf{m}_i^n(k) - \mathbf{p}^n(k) - \mathbf{C}_b^n(k)\mathbf{p}_{bs}^b] \end{aligned} \tag{17}$$

The observation model is non-linear and has two composite functions; a coordinate transformation from the navigation frame to sensor frame, and transformation from Cartesian coordinates to polar coordinates. By calculating Jacobian of this equation, linearised discrete model is obtained:

$$\mathbf{H}(k) = \begin{bmatrix} \frac{\partial\rho}{\partial\mathbf{p}^n} & \frac{\partial\rho}{\partial\mathbf{v}^n} & \frac{\partial\rho}{\partial\boldsymbol{\psi}^n} \\ \frac{\partial\varphi}{\partial\mathbf{p}^n} & \frac{\partial\varphi}{\partial\mathbf{v}^n} & \frac{\partial\varphi}{\partial\boldsymbol{\psi}^n} \\ \frac{\partial\vartheta}{\partial\mathbf{p}^n} & \frac{\partial\vartheta}{\partial\mathbf{v}^n} & \frac{\partial\vartheta}{\partial\boldsymbol{\psi}^n} \end{bmatrix}, \mathbf{R}(k) = \begin{bmatrix} \sigma_\rho^2 & 0 & 0 \\ 0 & \sigma_\varphi^2 & 0 \\ 0 & 0 & \sigma_\vartheta^2 \end{bmatrix} \tag{18}$$

If vision or radar information is available, $\boldsymbol{\delta}\mathbf{z}(k)$ is formed by subtracting the range, bearing and elevation of the sensor from the INS indicated range, bearing and elevation, then it is fed to the integrated fusion filter to estimate the errors in vehicle and map.

3.3.2 GNSS observation

GNSS can provide several observables such as position/velocity, pseudorange/pseudorange-rate, or integrated carrier phase. If the position/velocity observation is used the observation model simply becomes a linear form with,

$$\mathbf{H}(k) = \left[\begin{array}{ccc|c} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0}_m \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0}_m \end{array}\right], \mathbf{R}(k) = \begin{bmatrix} \boldsymbol{\sigma}_p^2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma}_v^2 \end{bmatrix}. \tag{19}$$

If GNSS information is available, $\boldsymbol{\delta}\mathbf{z}(k)$ is formed by subtracting the position and velocity of the GNSS from the INS indicated position and velocity, then they are fed to the fusion filter to estimate the errors in vehicle and map.

3.4 K/F Prediction

With the state transition and observation models defined in Equations 6 and 9, the estimation procedure can proceed. The state and its covariance are predicted using the process noise input. The state covariance is propagated using the state transition model and process noise matrix by,

$$\boldsymbol{\delta}\mathbf{x}(k \mid k-1) = \mathbf{F}(k)\delta\mathbf{x}(k-1 \mid k-1) = \mathbf{0} \quad (20)$$

$$\mathbf{P}(k \mid k-1) = \mathbf{F}(k)\mathbf{P}(k-1 \mid k-1)\mathbf{F}^T(k) + \mathbf{G}(k)\mathbf{Q}(k)\mathbf{G}^T(k)$$

Not only is the linear prediction much simpler and computationally more efficient than in the direct SLAM approach, but furthermore the predicted error estimate, $\boldsymbol{\delta}\mathbf{x}(k \mid k-1)$, is zero. This is because if one assumes that the only error in the vehicle and map states is zero mean Gaussian noise, then there is no error to propagate in the state prediction cycle, and the uncertainty in this assumption is provided in the covariance matrix propagation.

3.5 K/F Estimation

When an observation occurs, the state vector and its covariance are updated according to

$$\boldsymbol{\delta}\mathbf{x}(k \mid k) = \boldsymbol{\delta}\mathbf{x}(k \mid k-1) + \mathbf{W}(k)\mathbf{v}(k) = \mathbf{W}(k)\mathbf{v}(k) \quad (21)$$

$$\mathbf{P}(k \mid k) = \mathbf{P}(k \mid k-1) - \mathbf{W}(k)\mathbf{S}(k)\mathbf{W}^T(k)\,. \quad (22)$$

where the innovation vector, Kalman weight, and innovation covariance are computed as,

$$\mathbf{v}(k) = \mathbf{z}(k) - \mathbf{H}(k)\delta\mathbf{x}(k \mid k-1) = \mathbf{z}(k) \quad (23)$$

$$\mathbf{W}(k) = \mathbf{P}(k \mid k-1)\mathbf{H}^T(k)\mathbf{S}^{-1}(k) \quad (24)$$

$$\mathbf{S}(k) = \mathbf{H}(k)\mathbf{P}(k \mid k-1)\mathbf{H}^T(k) + \mathbf{R}(k)\,, \quad (25)$$

where, for the same reason as in the prediction cycle, $\mathbf{H}(k)\boldsymbol{\delta}\mathbf{x}(k \mid k-1)$ is zero and hence is not explicitly computed.

3.6 Feedback Error Correction

Once the observation estimation has been processed successfully, the estimated errors are now fed to the external INS loop and the map for correction. The corrected position, $\mathbf{p}_c^n(k)$, and velocity, $\mathbf{v}_c^n(k)$, are obtained by subtracting the estimated errors, and The corrected attitude quaternion, $\mathbf{q}_c^n(k)$, is obtained by pre-multiplying the error quaternion to the current quaternion:

$$\mathbf{p}_c^n(k) = \mathbf{p}^n(k) - \boldsymbol{\delta}\mathbf{p}^n(k \mid k) \quad (26)$$

$$\mathbf{v}_c^n(k) = \mathbf{v}^n(k) - \boldsymbol{\delta}\mathbf{v}^n(k \mid k) \quad (27)$$

$$\mathbf{q}_c^n(k) = \boldsymbol{\delta}\mathbf{q}^n(k) \otimes \mathbf{q}^n(k)\,, \quad (28)$$

where the error quaternion $\boldsymbol{\delta}\mathbf{q}_c^n(k)$ is computed from the estimated misalignment angle:

$$\boldsymbol{\delta}\mathbf{q}^n(k) \cong \begin{bmatrix} 1 & \delta\psi_x/2 & \delta\psi_y/2 & \delta\psi_z/2 \end{bmatrix}^T\,. \quad (29)$$

The corrected map positions are directly obtained by subtracting the estimated map position errors from the current positions:

$$(\mathbf{m}_N^n)_c(k) = \mathbf{m}_N^n(k) - \boldsymbol{\delta}\mathbf{m}_N^n(k \mid k)\,. \quad (30)$$

Using these equations the complementary SLAM/GNSS/INS Kalman filter can recursively fulfil its cycle of prediction and estimation with the external INS loop and the map.

3.6. Data Association and Feature Augmentation

Data association is a process that finds out the correspondence between observations at time k and features registered. Correct correspondence of the sensed feature observations to mapped features is essential for consistent map construction, and a single false match can invalidate the entire SLAM estimation process. Association validation is performed in observation space. As a statistical validation gate, the Normalised Innovation Square (NIS) is used to associate observations. The NIS (γ) is computed by

$$\gamma = \mathbf{v}^T(k)\mathbf{S}^{-1}(k)\mathbf{v}(k)\,. \quad (31)$$

Given an innovation and its covariance with the assumption of Gaussian distribution, γ forms a χ^2 (chi-square) distribution. If γ is less than a predefined threshold, then the observation and the feature that were used to form the innovation are associated. The threshold value is obtained from the standard χ^2 tables and is chosen based on the confidence level required. Thus for example, a 99.5% confidence level, and for a state vector which includes three states of range, bearing, and elevation, then the threshold is 12.8. The associated innovation is now used to update the state and covariance. If the feature is re-observed then the estimation cycle proceeds, otherwise it is a new feature and must be augmented into both the external map and the covariance matrix (Kim, 2004).

4. Results

A simulation analysis is performed to verify the proposed algorithm for the Brumby UAV, developed in University of Sydney, under GNSS enabled and disabled scenarios.

4.1 Simulation Environment

A low-cost IMU is simulated with a vision as the range, bearing, and elevation sensor. The vision sensor used in the real system provides range information based on knowledge of target size; hence its range is simulated with large uncertainty. The simulation parameters obtained from the implemented actual sensor specifications are listed in Table I.

Table 1. **The parameters used in simulation**

Sensor	Specification	Parameter
IMU	Sampling rate (Hz)	50
	Accel noise ($m/s^2/\sqrt{Hz}$)	0.5
	Gyro noise ($°/s/\sqrt{Hz}$)	0.5
Vision	Frame rate (Hz)	25
	Field-Of-View (°)	±15
	Estimated range error (m)	≥5
	Bearing noise (°)	0.16
	Elevation noise (°)	0.12
GNSS	Position noise (m)	2.0
	Velocity noise (m/s)	0.5

The flight vehicle undergoes three race-horse trajectories approximately 100m above the ground. The flight time is 460 seconds and the average flight speed is 40m/s. There are 80 features placed on the ground. The vision observation is expressed in a camera frame, which is transformed to navigation frame to be processed in the SLAM node. The biases of the IMU are calibrated precisely using onboard inclinometers in the real implementation thus the biases are not explicitly modelled and studied in the simulation analysis and only white noise is modelled as in Table 1.

4.2 GNSS Active Region: Map Building

Figure 4 shows the SLAM/GNSS/INS estimated vehicle trajectory with the map built during the flight. The map uncertainty ellipsoids are also plotted with 10σ boundaries for the clarification. The vehicle takes off and flies over circuit-1 two rounds. It then transits to circuit-2 and circuit-3. To simulate GNSS denied scenario, GNSS signal is disabled between 130 to 420 seconds from the start. After the vehicle taking off, GNSS signal is available until 130 seconds. The system behaves as a feature-tracking/mapping system in this mode. The error covariance of features around circuit-1 has relatively small value than features around other circuits. Figure 9 presents the evolution of uncertainties of the vehicle and map. It can be clearly observed that the vehicle uncertainty was maintained within one metre until 130 seconds, and the uncertainties of observed features are monotonically decreased. INS error is dominantly estimated from GNSS information, and the GNSS/INS blended navigation solution is used to track features. Contrary to the conventional airborne mapping systems, SLAM/GNSS/INS system maintains the cross-correlation information between the INS and map which can enhance the INS performance, and it is essential for the SLAM operation in GNSS denied conditions.

4.3 GNSS Denied Region: SLAM/INS navigation

In this condition, the SLAM/GNSS/INS system now behaves as a SLAM/INS system. INS and map errors are solely estimated from the feature observation. The pre-registered feature during GNSS active period can be effectively used to estimate the INS and map errors as shown in Figures 5 to 8. After GNSS is disabled, INS uncertainties begin to increase which in turn increases the registered map uncertainties. The accumulated INS error is effectively removed from the closing-the-loop effect, which, in turn, eliminates the INS error embodied in the map. This can be observed in INS and map covariance plot as in Figure 13. Figures 10 and 11 show the evolution of uncertainty of INS velocity and attitude which are constrained effectively for extended period of time without GNSS aiding. This is due to the correlation structure between vehicle and the map in SLAM. The map uncertainty decreases monotonically and whenever the vehicle observes the feature, the vehicle error can be constrained effectively until GNSS signal is available again. Figures 12 shows the final map uncertainties built in GNSS enabled and disabled conditions. When GNSS signal is re-activated, the INS position and velocity errors are directly observed from the GNSS measurements, which, in-turn, improves the map accuracy via the vehicle-map correlation structure within SLAM filter. From these plots, it is obvious that the SLAM/GNSS/INS system can perform navigation and mapping for extended periods of time under GNSS denied conditions.

5 Conclusions

A new concept for UAV navigation is presented based on 6DoF Simultaneous Localisation and Mapping (SLAM) algorithm, and augmenting it to GNSS/INS system. The simulation analysis illustrates that the SLAM system with a range, bearing, and elevation sensor can constraint the INS errors effectively, performing on-line map building in GNSS denied and unknown terrain environments for extended periods of time. It can be applied to various navigation areas such as battlefield situations, urban canyons, indoor, or underwater. The real-time implementation on UAV platform using low-cost sensor are being tackled at the moment.

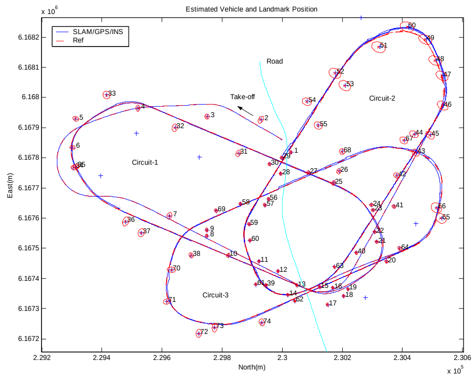


Figure 3. 2D position result of the SLAM augmented GNSS/INS navigation. UAV takes off at (0,0) and flies three different race-horse tracks (circuit-1,2,3) in counter clock-wise. GNSS signal is disabled in circuit-1 and re-enabled at the end of in circuit-3. The vision is abaliable during whole flight time which is used for feature-mapping and SLAM.

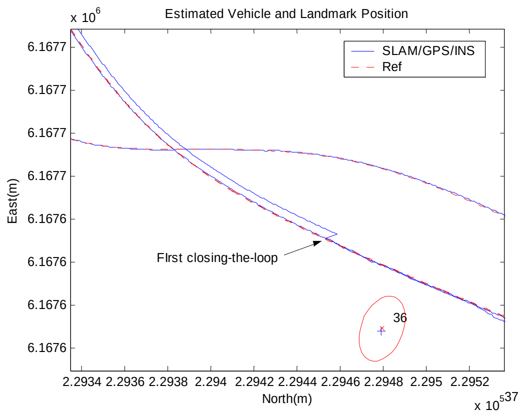


Figure 4. Enhanced view of INS correction by the closing-the-loop effect of SLAM during GNSS disabled condition.

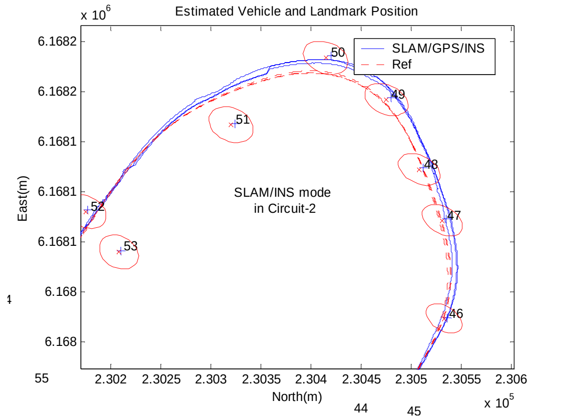


Figure 5. SLAM inclemently builds new map during GNSS disabled condition.

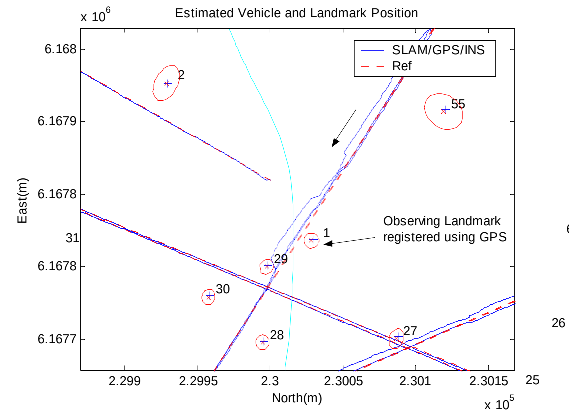


Figure 6. Enhanced view of INS correction by re-observing previously built map by GNSS.

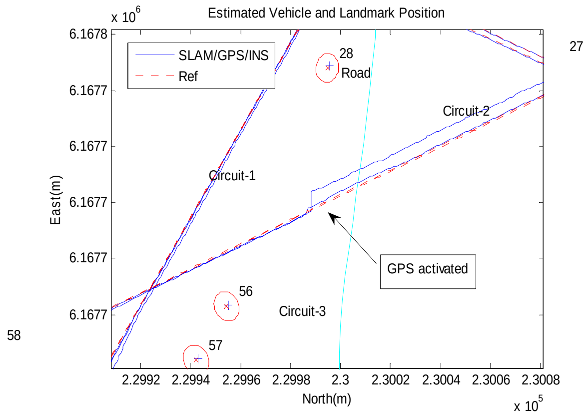


Figure 7. GNSS is re-enabled and it corrects both INS and map error simultaneously.

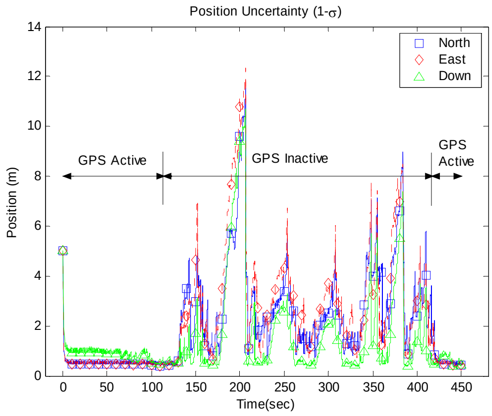


Figure 8. Evolution of INS position uncertainty during flight. It can be observed that SLAM can bound the error growth during GNSS denied condition for extended period of time. GNSS is disabled at 130 second and re-enabled at 420 second

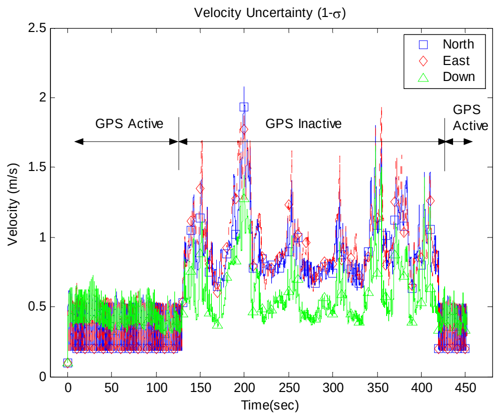


Figure 9. Evolution of INS velocity uncertainty during flight.

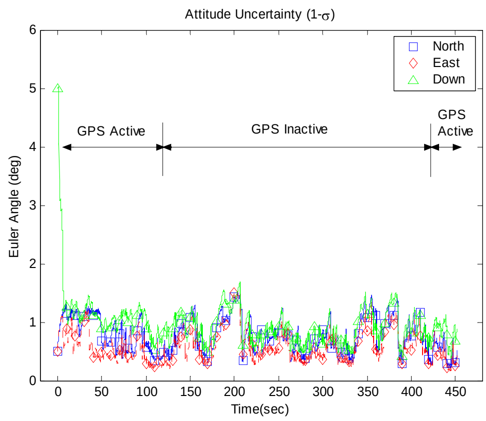


Figure 10. Evolution of INS attitude uncertainty during flight.

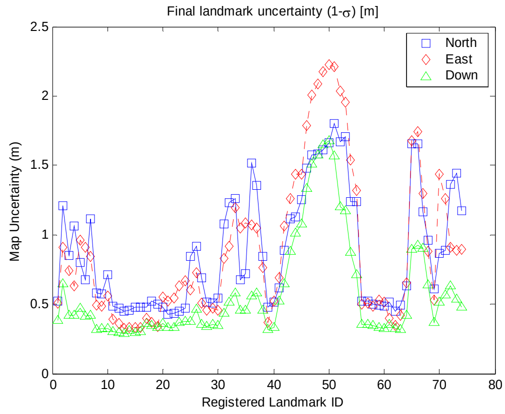


Figure 11. Final map uncertainty. The features registered during GNSS denied condition show larger uncertainties due to the accumulated INS error.

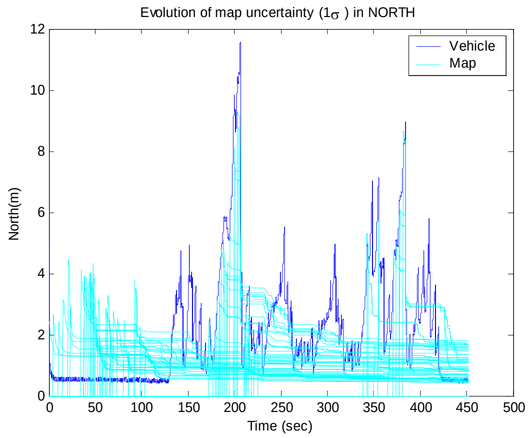


Figure 12. Evolution of INS position and map uncertainties in north direction. The map uncertainty converges to the lower limit monotonically.

Acknowledgements

This work is supported in part by the ARC Centre of Excellence programme, funded by the Australian Research Council (ARC) and the New South Wales State Government.

References

Dissanayake M.W.M.G.; Newman P.; Durrant-Whyte H.; Clark S.; Csorba M. (2001): ***A solution to the simultaneous localization and map building problem,*** IEEE Transactions on Robotics and Automation, June, 17(3):229–241.

Guivant J.; Nebot E. (2001): ***Optimisations of the simultaneous localization and map building algorithm for real-time***

implementation, IEEE Transactions on Robotics and Automation, 17(3):242–257.

Kim J.; Sukkarieh S. (2004): ***Autonomous Airborne Navigation in Unknown Terrain Environments,*** IEEE Transactions on Aerospace and Electronic Systems, 40(3):1031-1045, July

Kim J. (2004): ***Autonomous Navigation for Airborne Applications,*** PhD thesis, Australian Centre for Field Robotics, The University of Sydney.

Smith R.; Cheeseman P. (1987) ***On the Representation of Spatial Uncertainty,*** In International Journal of Robotics Research, 5(4):56–68.

Williams S.B.; Dissanayake M.W.M.G.; Durrant-Whyte H. (2001): ***Towards terrain-aided navigation for underwater robotics,*** In Advanced Robotics, 15(5):533–550.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 129-138

Journal of Global Positioning Systems
2005
© CPGPS

Deformation monitoring and analysis using Victorian regional CORS data

Kefei Zhang[1], Youjian Hu[1,2], Gangjun Liu[1], Falin Wu[1] and Rodney Deakin[1]

[1]School of Mathematical and Geospatial Sciences, RMIT University, Melbourne, Australia
Tel: +61-3-99253272, Fax: +61-3-96632517, Email: Kefei.zhang@rmit.edu.au
[2]Department of Surveying Engineering, China University of Geosciences, Wuhan, China
Tel: +86-27-67884788, Fax: +86-27-67883507, Email: huyoujian@cug.edu.cn

Received: 27 November 2004 / Accepted: 12 July 2005

Abstract. This paper investigates the feasibility using continuously operating reference stations (CORS) in Victoria (termed GPSnet) for deformation monitoring and analysis. A number of critical issues associated with the suitability, geological stability, data quality of the GPS networks system, the precision and reliability of the GPSnet solution are investigated using geological information. Appropriate strategies for GPS data processing and deformation analysis are investigated. The absolute and relative displacement of selected GPSnet stations are analysed using chronological GPS CORS data and dedicated high precision scientific GPS data processing software packages. The latest International Terrestrial Reference Frame is used for deformation analyses. Detailed data-processing strategies and results of deformation analyses are presented and some useful conclusions are drawn.

Results show that the methodology of deformation analysis and data processing based on the regional CORS network data is feasible and effective. It is concluded that high-precision continuous tracking data from GPSnet is a very valuable asset and can provide a technically-advanced and cost-effective geoscientific infrastructure for deformation monitoring analysis. By mining the data from the GPSnet, not only reliable and high precision deformation information can be potentially obtained, but also high expenditure required for establishing dedicated deformation monitoring networks in this area can also be spared.

Key words: CORS, ITRF, precise ephemeris, deformation monitoring and analysis

1 Introduction

Natural disasters are of a problem of global concern and may cause significant social, environmental and human losses and sometimes, threaten geopolitical stability. Natural hazards that impact on Australian communities include earthquakes, landslides, floods, storm surges, severe winds, bushfires, and tsunamis. In Australia, natural hazards are estimated at an average annual cost of $1.25 billion (Geoscience Australia, 2004a). Victoria is one of the regions where earthquake epicentres are relatively concentrated. There is potential risk of earthquakes in this region. For example, Yallourn area within Victoria is a geologically active part and experiences earthquake from time to time (Brown, 2002). Landslide is another considerable geological hazard in Australia and south-eastern Victoria is a very active landslide area (see Figure 9). The Earth's surface deformation due to human activities such as mineral mining may also cause hazards.

For many applications, such as site selection of important engineering projects and constructions and their protection against hazards, the ability to analyse and predict natural and non-natural hazards is of great importance. Such ability depends heavily on precise and reliable deformation information which in turn can be acquired by using advanced technologies through the monitoring and analysis of the Earth's surface displacement, the movement of faults, landslide and some other deformations.

Due to its high precision, 24 hours availability, operability under all weather conditions and automation, GPS technique has been widely used in monitoring ground movement, deformation and subsidence (He et al., 2004; King et al., 1995; Kogan et al., 2000). In Victoria, GPSnet with its high-precision observation data provides a technically-advanced and cost-effective geoscientific infrastructure for deformation monitoring analysis. By mining the data from the GPSnet, not only reliable and high precision deformation information can be extracted, lots of expenditure required for establishing dedicated deformation monitoring networks in this area can also be saved.

Methodologies for GPS data processing and deformation analysis are investigated. The absolute and relative displacement of selected GPSnet stations subnet are calculated using chronological GPS data and AUSPOS (Dawson et al., 2004) on-line scientific data-processing engine. The feasibility and effectiveness of the methodologies put forward are discussed and some useful conlusions are given.

2 Victorian GPSnet

In 1994, Land Victoria, the Department of Sustainability and Environment (DSE) of the State of Victoria/Australia foresaw the rapid developments of global navigation satellite technology and initiated an ambitious project to establish a set of 20 permanent and continuously operating GPS Base Stations (GPSnet) across the State. The primary purpose of the GPSnet is to provide a range of users with a means of obtaining accurate and homogenous positioning within Victoria using the space-borne technology. As an integral part of the new geodetic strategy for Victoria, GPSnet is being established in partnership with industry and academia. GPSnet currently consists of 19 operating base stations and will contain approximately 24 primary stations upon completion. Table 1 lists the chronological developments of the GPSnet stations.

The nominal design spacing of GPSnet stations is approximately 50km in the Melbourne metropolitan region and 100km in rural Victoria, but in remote areas the separations can range up to 200km (see Figure 1). The Melbourne observatory base station has been connected to International GPS Service (IGS) network. The network records, distributes and archives GPS data for accurate position determination with post-processing techniques. Seven sites also transmit local real-time kinematic (RTK) correction signals via radio. The GPSnet system provides a mechanism for centimetre level positioning relative to the Australian National Spatial Reference Systems. Real-time transmission of networked GPSnet data to enable near instantaneous network RTK positioning services using a single GPS receiver is currently under consideration by DSE (Hale, 2004).

GPSnet uses a variety of receivers including Trimble 4000SSE/SSI, 4700 and Leica SR9500 dual-frequency receivers. The receivers use dual-frequency geodetic antennas with ground-planes and record C/A code, L_1/L_2 carrier phase and Doppler data in the RINEX format at all sites. All antennas are permanently mounted to provide an uninterrupted view of the surrounding sky. GPS antennas are usually sited on rooftops of buildings and at the most stable locations free of multipath. Data

Tab. 1 Chronological developments of the Victorian GPSnet stations

GPSnet stations (date of operation)	Year of operation	No of Stations (Total)
Ballarat (01/12)	1995	1 (1)
Epsom (01/07) (relocated in 2002) Melbourne RMIT (01/08)	1996	2 (3)
Geelong (03/09)	1998	1 (4)
Benalla (13/07) Irymple (relocated in 2003) (26/01)	1999	2 (6)
Colac (30/10) Mt Buller (19/12)	2000	2 (8)
Swan hill (05/03) Hamilton (19/03) Shepparton (06/04) Walpeup (14/05) Horsham (02/06) Yalllourn (21/06) (relocated in 2003)	2001	6 (14)
Cann River (01/09) Melb obs (IGS station) (18/11)	2002	2 (16)
Clayton (12/02) Bairnsdale (31/10)	2003	2 (18)
Albury (11/02)	2004	1 (19)

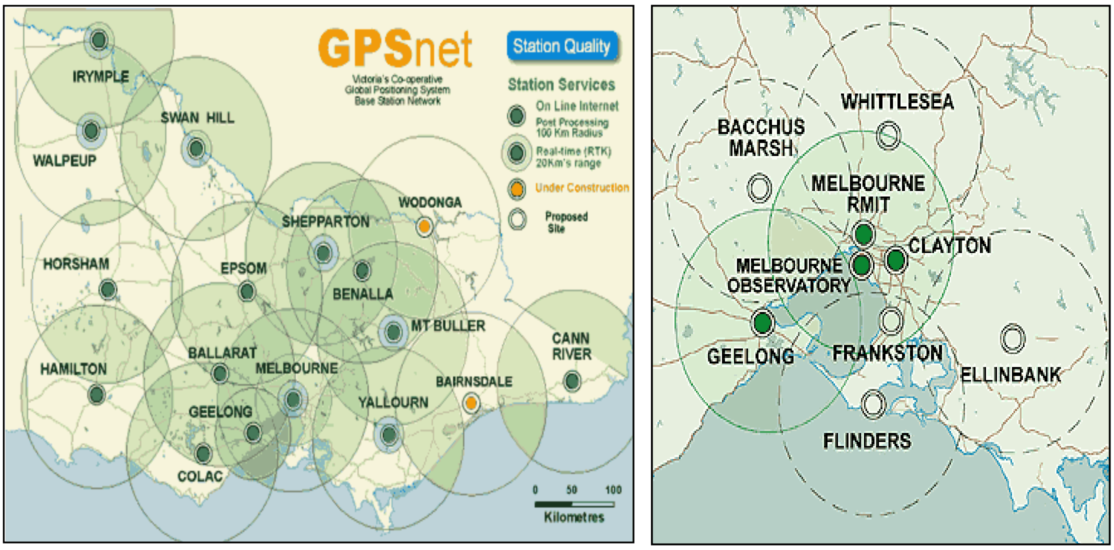


Fig. 1 Victorian rural/regional (left) and Melbourne (right) GPSnet base station network locations and their development status (Land Victoria, 2003)

processing is performed in the International Terrestrial Reference Frame (ITRF) 97 and then transformed to geocentric datum of Australian (GDA) 1994. Preliminary results indicate that RMS of daily solutions is in the order of 2-4mm in easting and northing and 3-8mm in height using IGS final orbits products (Brown, 2002). Apart from high-precision geodetic applications, the GPSnet has been widely used since its inception, including but not limited to navigation, mapping, GIS, surface deformation monitoring (eg open pit coal mining), agriculture and surveying applications (Zhang and Roberts, 2003).

Land Victoria has also developed a mechanism for data quality check of the GPSnet measurements. This is measured through visual indications of cycle slips in carrier phases, multipath effects and data completeness respectively. Figures 2-4 show cycle slips occurred, multipath effects and data completeness for Ballarat station from 7 October to 6 November 2004 (Land Victoria, 2004).

This information provides a rough idea on the quality of the data measured in a particular station and this is very valuable for GPSnet users.

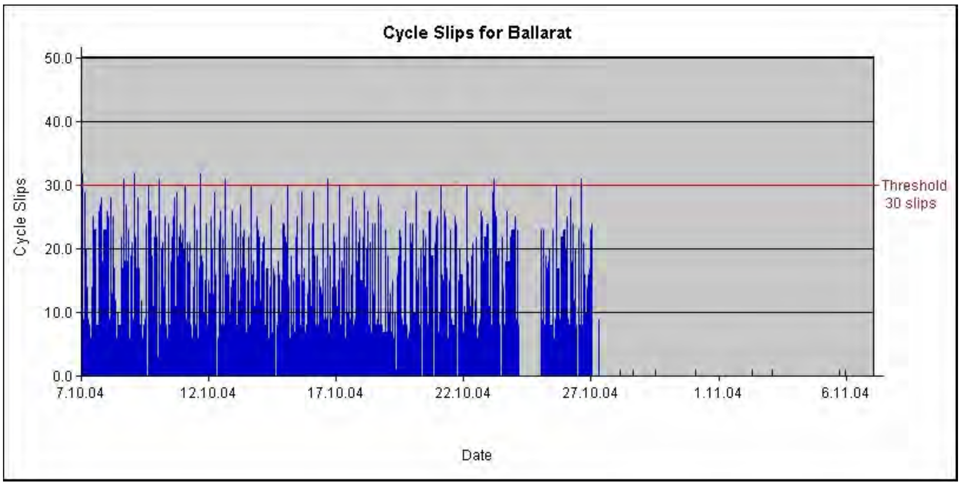


Fig. 2 Cycle slips detected for Ballarat station during October 2004

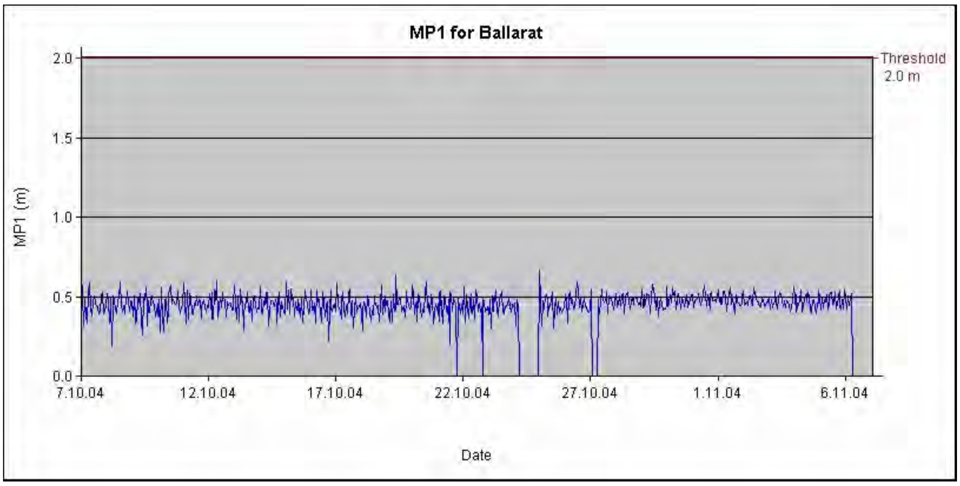


Fig. 3 Multipath effects detected for Ballarat station during October 2004

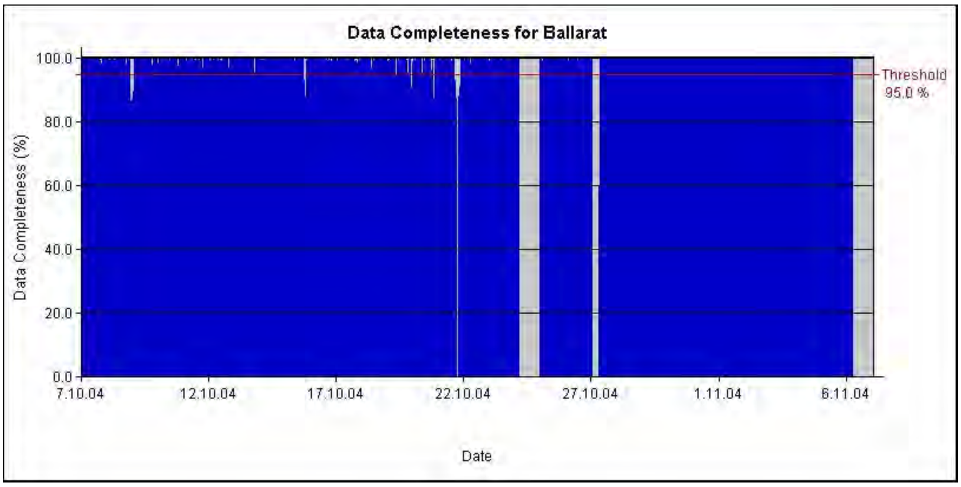


Fig. 4 Data completeness results for Ballarat station during October 2004

3. Method of deformation analysis

Victorian GPSnet is of high precision (mm level in horizontal position), and most antennas are of good quality and high stability. The GPSnet is, therefore, capable of providing reliable and high-precision deformation data, such as the determination of both velocity and direction of Earth's surface displacement, relative movement of large geological faults, the relation between the Earth's surface displacement and tectonic motion, landslide deformation and the crustal deformation caused by mineral mining.

Figure 5 outlines a detailed process of this investigation using GPSnet measurements for deformation analysis. Major steps for data processing, deformation analysis and some important contributing factors are presented. The technical requirements and procedures of data-processing and deformation analysis are usually different for different types of deformation analyses, and the reference datum of deformation analysis and the GPS base stations in the GPSnet should be properly chosen to form an optimal deformation analysis subnet.

3.1 Selection of deformation analysis datum

A number of ITRFs (i.e. ITRF 93, 94, 96, 97 and 2000) are involved in Victorian GPSnet data due to historical evolution. To obtain reliable results of deformation analyses, coordinate reference frames of GPSnet stations must be identical. Obviously, the latest and most accurate reference frame of ITRF2000 should be used as the unique coordinate reference frame so that both the position change of entire GPSnet caused by tectonic motion of the Australian continent and the relative position changes of GPSnet reference stations caused by other factors (fault movement， landslide， mineral mining, etc.) can be precisely estimated.

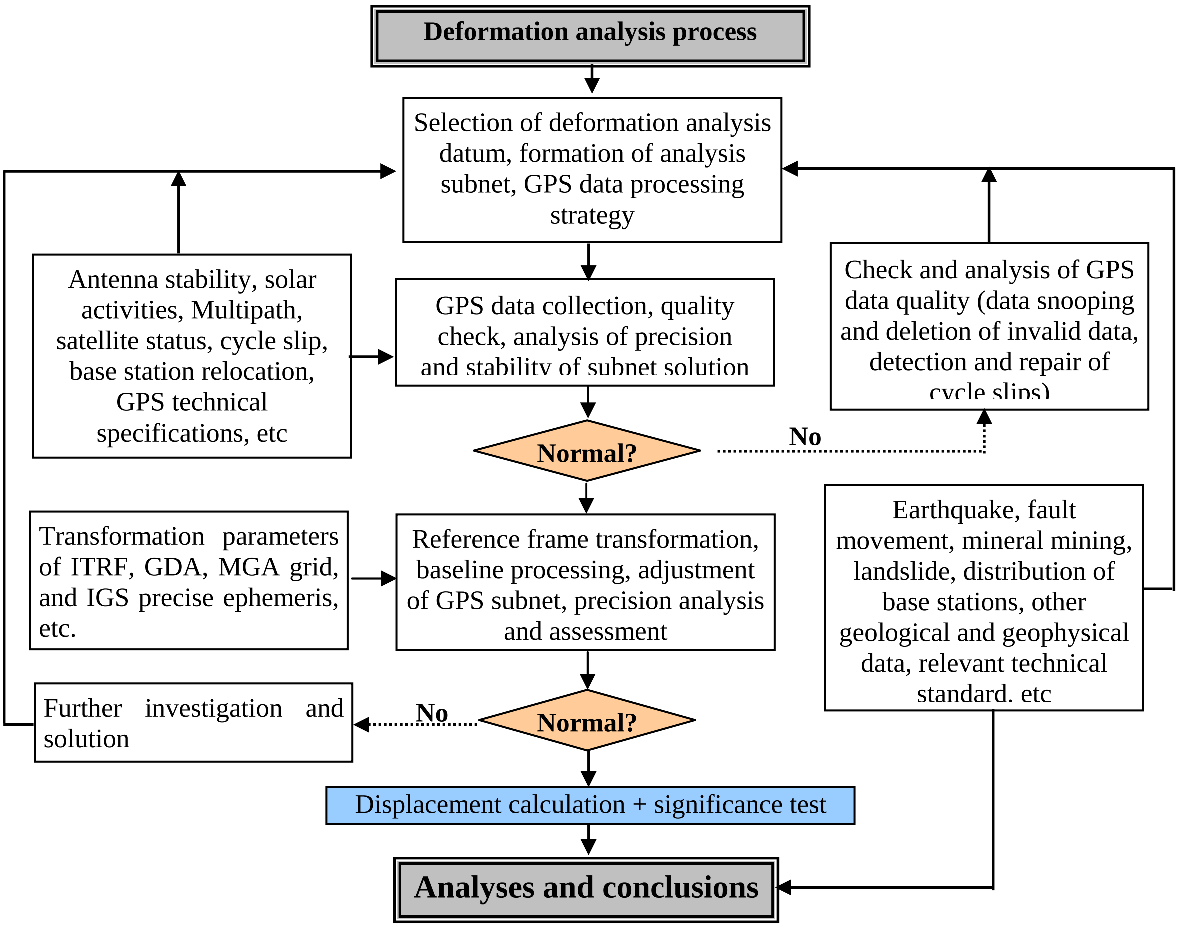


Fig. 5 Flowchart of GPSnet data processing and stability analysis

Coordinate transformation of an ITRF system to ITRF2000 can be performed using the transformation parameters provided by the International GPS Services (IGS, 2000). The deformation analysis can also be conducted in GDA94 or Map Grid of Australia (MGA) as long as the coordinates of GPSnet reference stations in ITRF2000 are transformed to GDA94 or MGA Grid using the transformation parameters between ITRF2000 and GDA94/MGA Grid (Dawson, 2002). The displacement of GPSnet reference stations derived from the coordinate differences in ITRF2000 from two different epochs reflects the resultant effects of all contributing factors on the stability of GPSnet stations. If the effects from Australian continent motion are subtracted from the "absolute" displacement, then the relative displacement of GPSnet reference stations can be obtained.

To obtain precise relative displacement, it is more desirable that one relative stable station in GPSnet is used as the datum of deformation analysis and the GPS network for displacement analysis is adjusted using a non-constrained free network adjustment method. In GPSnet, the "Melbourne Observatory" station in IGS network should be ideally used as a relatively stable datum because it is built directly on bedrocks and of high stability. However, the station was established in 2002 and became operational since November 2002. Before then, no station in the GPSnet can be regarded of high stability since all of the GPSnet station antennas are mounted on rooftops of buildings. Therefore, currently, to compute and analyse relative displacement of the GPSnet, relatively stable and precise IGS/ARGN reference stations close to the GPSnet have to be selected and subsequently used as a stable datum for relative displacement analysis of the GPSnet stations, ie the GPS network for displacement analysis is adjusted using the free network adjustment method with no fixed datum.

3.2 Formation of deformation analysis subnet

There are a number of different deformation analyses required, for example, displacement analysis of the entire GPSnet, local deformation analysis, comprehensive deformation analysis of the effects of multiple factors and individual analysis of the effects of a single factor. For different deformation analysis, GPSnet stations should be chosen to form an optimal deformation analysis network - "subnet" for a particular corresponding purpose.

The subnet used for a certain deformation analysis purpose should use the same network shape, same deformation analysis datum and compatible precision whenever the subnet data is processed and adjusted. By doing so, potential systematic errors caused by adopting different deformation analysis datums and minimised.

3.3 Data processing strategy of subnet

Given the fact that the velocity of the Earth's surface displacement is usually within a few centimetres per year and the current relative baseline precision of GPS measurement is in the order of 10^{-6}~10^{-8}, it is, in general, not necessary to process the GPSnet data continuously or in a short time interval. Instead, the GPSnet data processing should be carried out at one time per year or one time per season scenario (so that the surface movement/displacement is large enough to be reliably detected). However, when the Earth's surface is active due to some reasons, the interval of the data processing sessions should be properly increased or the session interval can even be as high as possible in order to extract real time and kinematic displacement information.

To achieve reliable deformation analysis results, the solution of the deformation analysis subnet needs to be precise enough and stable. The precision and stability of the network solutions are strongly related to the amount of GPS measurements used to generate the solution, which is usually measured in the length of observation time (for a given sampling rate). Research on the amount of data required has been conducted and solutions from a minimal of six hours data are usually considered precise enough and stable for a high precision deformation monitoring and analysis (Dawson et al., 2004). However, there are a number of important factors contributing to the stability of a GPS network solution, such as the length of observation time, the amount of valid data collected, baseline length, quality of GPS signal recorded, the station environment (e.g. multipath, solar activities, satellite status), and cycle slip, etc. Among these, many factors vary with time. Therefore, it is necessary to investigate numerically the proper amount of data required to generate a reliable and precise solution from the deformation analysis subnet.

In GPS data processing, analysis of precision and stability of GPS network solution can be conducted using precise GPS data processing software, such as GAMIT (Gamit, 2004), BERNESE (Bernese, 2004) or AUSPOS (Dawson et al., 2004). A number of trials are carried out to test the "best" software package for this research and it is found that all the three packages give very similar baseline solution. AUSPOS is chosen due to its automation and access to the solutions in different reference frames (details see below). Note that for same deformation analysis network, the same data processing software should be used whenever the GPS network data is processed and adjusted so that any potential errors caused by different computational models and algorithms can be minimised.

4. Precision and stability of GPSnet solution

The longest baseline (Cann River-Irymple, 723km) in Victorian GPSnet with a fixed datum derived from three IGS stations （HOB2, STR1, TIDB） is selected to form an experimental network for precision and stability analyses of the GPSnet (see Figure 6). GPS data pre-processing software "TEQC" (TEQC, 2004) is used for editing and quality check of the GPS data. The precise GPS data processing software AUSPOS is used to generate the solution of the experimental network.

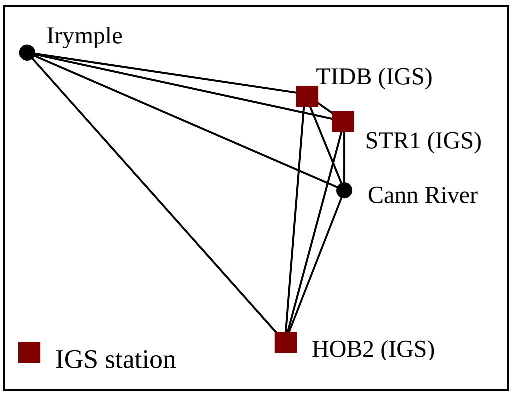


Fig. 6 An experimental network for precision analysis of GPSnet solution (not to scale)

AUSPOS allows users to submit their data via the Internet. The RINEX data needs to be static and geodetic quality (i.e. dual frequency) and the turn-around time of the processing results is very short. The quality of the coordinates with 6 hours of data is: horizontal precision is better than 10 mm and vertical precision is better 20 mm (Dawson, 2002). AUSPOS processing report provides coordinates in ITRF, GDA94 and MGA Grid, precision of coordinates, RMS of observations, percentage of observations removed etc. This information is very useful for analysing the precision and stability of the experimental network solution. AUSPOS processing engine uses IGS precise ephemeris products, Earth orientation and station coordinate and velocity parameters and differential technique to several IGS stations. The data processing is undertaken in accordance with the International Earth Rotation Service (IERS) computation standards.

The experimental network data recorded on 14 April 2004 is used for precision and stability analysis of solutions. Figure 7 shows the relation between precision (s_X, s_Y, s_Z) of coordinates computed (in ITRF2000) and the amount of data used. Figure 8 shows the coordinate differences (dx, dy, dz) between the coordinate derived from different session lengths (amount of data) and the "ground truth" values that are derived from 24-hour data.

From Figures 7 and 8, we can conclude that:

(1) Overall the accuracy of the coordinates derived from different lengths of observations varies and their differences can be upto one decimetre level. The accuracy can be improved when more data is used and the solution is pretty stable when more than 20 hours of data is used. The differences of coordinates decrease when the length of the session increases. This means that solutions converge (to the "ground truth") when the length of data sessions increases.

(2) When the session length is less than six hours, the RMS error of coordinates can be more than 15mm, and the coordinate differences can be more than 20mm, which cannot meet the requirement for a high precision deformation monitoring. In addition, the solution is not stable enough, particularly when the session length is less than 2 hours. Figures 7 (b) and 8 (b) show that some coordinate differences and coordinate errors derived from 2 hours data are obviously more than those derived from 1-hour data, which are, theoretically, not normal. This is most likely due to the fact that the 2-hour data is too noisy and there exists a large random error (see Figures 2-4 for example).

(3) When the length of a session is 12 hours, the coordinate error is about 5mm, and the coordinate differences can be less than 10mm, which means that the solution is relatively stable. When the session length is close to 24 hours, such as more than 20 hours, the precision of coordinates is 3~5mm, and the coordinate differences can be less than 5mm, which mean that the solution becomes quite stable.

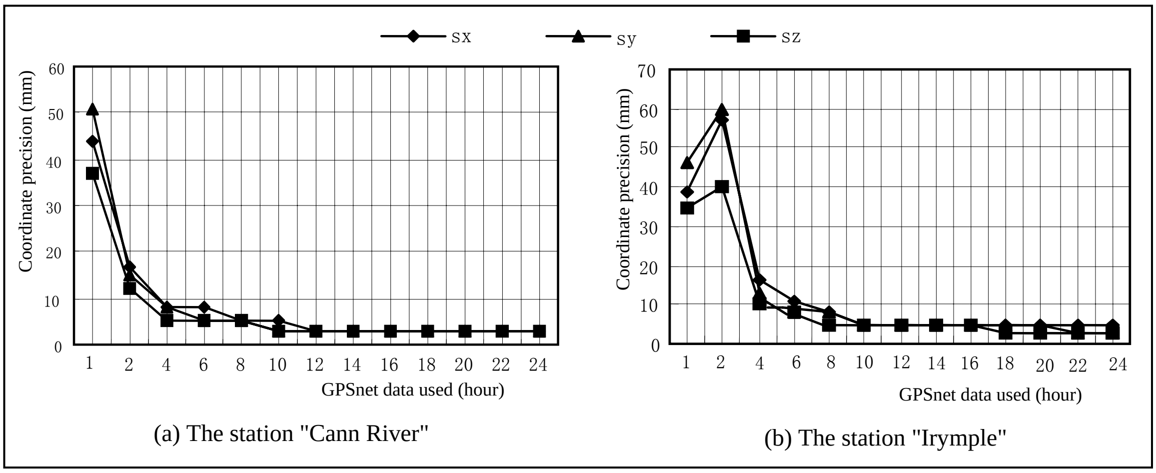


Fig. 7 Precision evaluation of GPSnet solution in ITRF2000 using different session lengths of data at stations (a) "Cann River" and (b) "Irymple" respectively

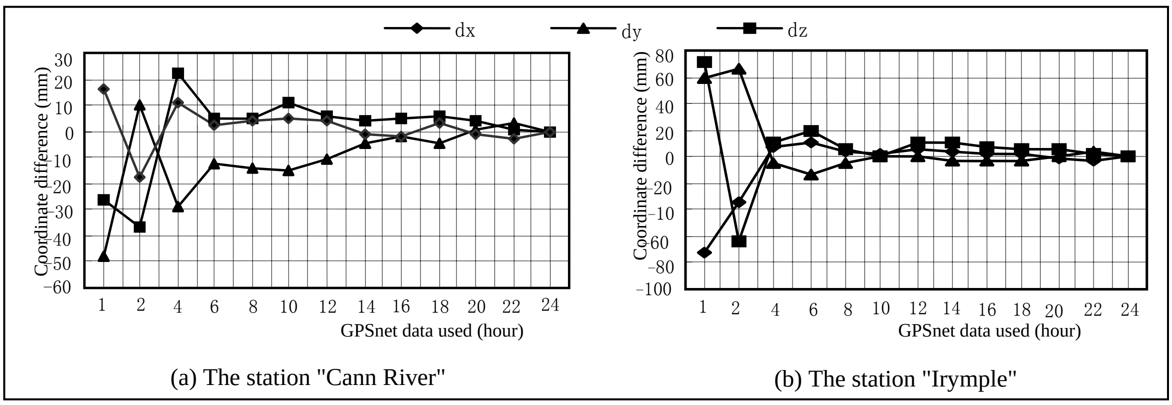


Fig. 8 Differences of the coordinate solutions in ITRF2000 using different session lengths of GPSnet data in comparison with the solution from 24 hours data

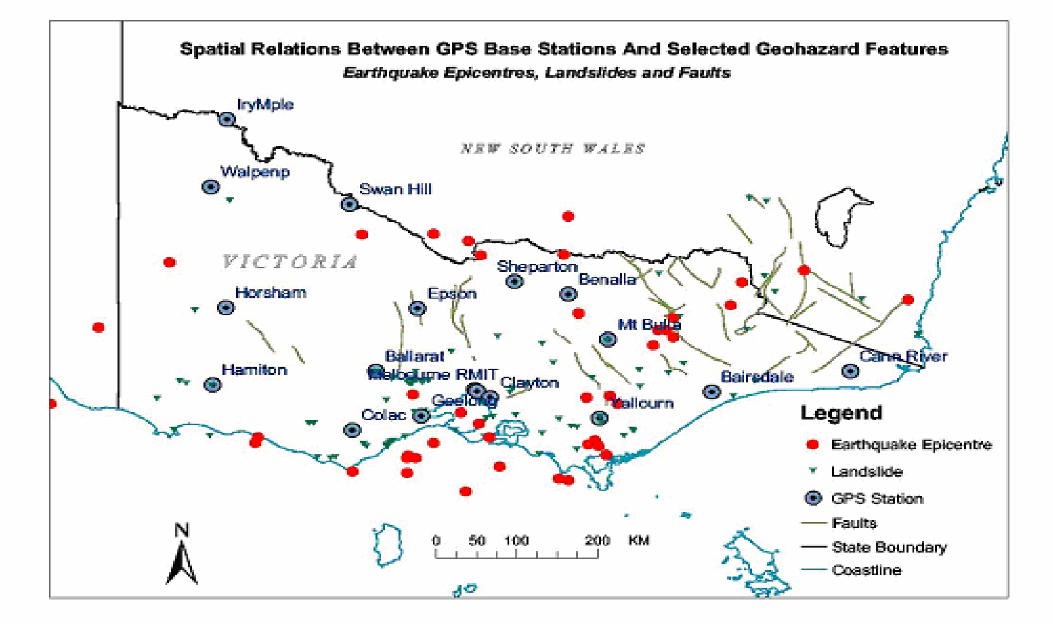


Fig. 9 Schematic figure of the displacement vector at selected stations and spatial relations between GPSnet stations and geological features

Thus it can be seen that daily GPSnet solution (24 hours of data) is of high precision and sufficient stability. It is, therefore, possible to use this data for high precise regional deformation monitoring and analysis.

5. Calculation and analysis of deformation

Figure 9 shows the distribution of some geological features (earthquake epicentres, faults and landslides) in Victoria and the spatial position relations between Victorian GPSnet stations and these geological features. There are more than 10 relatively large faults within Victoria and some stations are close to faults and/landslide sites (eg Epson). Victoria, in particular south-eastern Victoria, is one of the regions where both earthquake epicentres and landslide sites are relatively concentrated. There are potential risks of earthquake and landslide in this area. In addition, human activities such as mineral mining can also cause deformation of the Earth's surface. Therefore, according to the result of displacement analysis of GPSnet stations, the regional deformation of the Earth's surface and the stability of some faults and landslide sites can be inferred.

5.1 Calculation of GPSnet Station Displacement

A number of factors are taken into consideration when choosing experimental network, length of sessions and epochs of comparisons. These factors include data file losing, improper data format and relocation of some stations. GPSnet data from 14 April 2002 to 14 April 2004 and seven base stations (see Figure 10) are used in this paper for local deformation analysis.

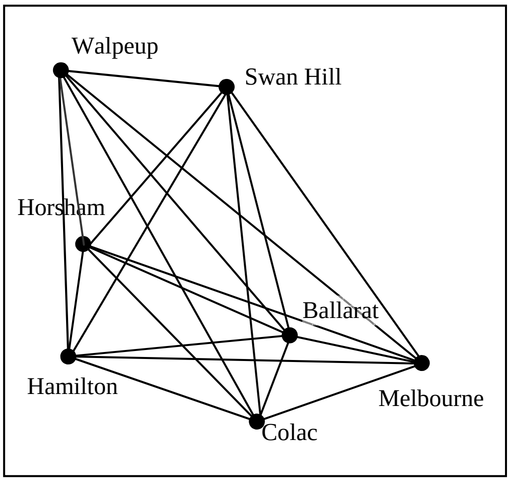


Fig. 10 A seven-station GPSnet subnet selected for displacement analysis (not to scale)

There are 21 simultaneous observation baselines in the subnet. The longest baseline length (Walpeup-Melbourne) is 399km and the shortest baseline length (Colac-Ballarat) is 90km. The subnet is adjusted using the free network adjustment method (with no fixed datum). The absolute displacements in horizontal directions (ΔE=Easting, ΔN=Northing) and vertical direction (ΔU=up direction) of the subnet stations (derived from the transformation of ITRF2000 to Australian Map Grid) are shown in Table 2.

Tab. 2 Absolute and relative displacements of the GPSnet subnet stations

station	Absolute displacement (mm) and velocity (mm/yr)						relative horizontal displacement (mm)		
	ΔE	ΔN	ΔU	V	V/2	significance test	ΔE_r	ΔN_r	significance test
Melbourne	21	124	32	130	65	✓	-6	-5	×
Ballarat	24	121	25	125	62	✓	-3	-8	×
Colac	32	125	18	128	64	✓	5	-4	×
Hamilton	35	128	16	134	67	✓	8	-1	×
Horsham	20	135	34	139	70	✓	-7	6	×
Walpeup	33	138	47	148	74	✓	7	9	×
Swan Hill	27	131	45	142	71	✓	0	2	×

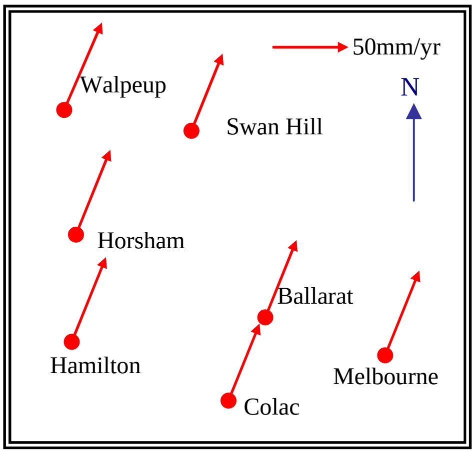


Fig. 11 Amplitude and direction of the absolute

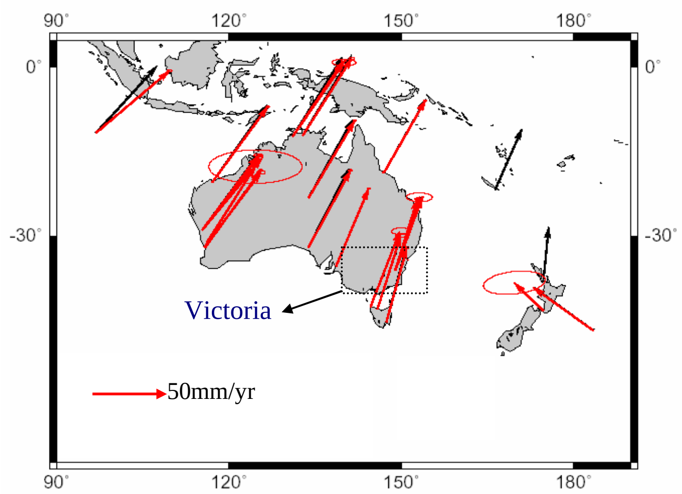


Fig. 12 Australian tectonic motion vector from IGS

The total displacement magnitude "*V*" is calculated by the following formula: $V = \sqrt{(\Delta E)^2 + (\Delta N)^2 + (\Delta U)^2}$. "*V/2*" is the mean annual velocity of the displacement. "ΔE_r" and "ΔN_r" are relative horizontal displacements and are free from the systematic horizontal displacement of the whole subnet.

5.2 Deformation Analysis

The significance of both absolute and relative displacements in Table 2 are tested. The displacement significance of the whole subnet is tested using F-Test and the displacement significance of single station is tested using T-test. The results of significance test are listed in Table 2. The symbols "✓" and "×" indicate significant and insignificant respectively. The results of significance test show that the absolute displacements of all the subnet points are significant. The absolute displacement directions of all the subnet points are shown in Figure 11. The average displacement velocity of the subnet points is 6.8 cm/year. Both the magnitude and direction of the absolute displacement of all the base stations in the subnet agree well with the velocity of approximately 7cm/year and direction of current Australia tectonic motion (see Figure 12) derived from other IGS measurements (Geoscience Australia, 2004b).

Since the precision of vertical coordinate (height) is about 2-3 times lower than that of horizontal coordinates, the relative vertical displacements of the subnet points are not precise enough and reliable for high precise deformation analysis. Therefore, the relative vertical displacement of the subnet is not analysed in this paper. The results of significance test show that the relative horizontal displacements of all the subnet points are not significant. Thus it can be seen that the relative horizontal positions of the subnet points are not notably affected from local geological features. According to this, it can be inferred that currently, the faults and/or landslide body near these base stations are relatively stable. Of course, the stability of the faults and landslide body still needs to be continuously analysed in the future.

5. Conclusive remarks

The solution of Victorian regional CORS network is not precise enough and stable for high precise deformation monitoring and analysis if the amount of data used to generate the solution is less than 12 hours. The precision of 3D coordinates derived from daily GPSnet solution (24-hour data) is 3~5mm and the solution is quite stable. This can meet the requirements of high precision deformation analysis. Therefore, continuous tracking data from GPSnet is a very valuable asset and can provide a technically-advanced and cost-effective geoscientific infrastructure for regional deformation monitoring and analysis.

The average velocity of the displacement at subnet points is 6.8 cm/year. Both the magnitude and direction of the whole subnet displacement agree well with the velocity of ~7cm/year and direction of current Australian continent derived independently. The relative horizontal positions of the subnet points are not notably affected from local geological features. It can be inferred that the faults and/or landslide body near these base stations are relatively stable.

Preliminary results indicate that the methodology of data processing and deformation analysis based on CORS is feasible and effective. However, further investigation is required when more GPSnet data covering a larger chronological span and more GPSnet base stations can be used for deformation analysis. It is recommended that geological information needs to be taken into account when any new CORS stations are established. The improvement of data quality, stability of antenna, precision and reliability of the GPSnet solution will be of great help in the analysis of both absolute and relative displacement of the GPSnet stations. It is, therefore, anticipated that the GPSnet will play an important role in the regional deformation monitoring and analysis.

Acknowledgements:

The authors would like to thank P. Ramm, M. Hale, P. Oates, J. Millner and E. Retimana from the Department of Sustainability and Environment, Victoria for their consistent support and access to the GPSnet data. Financial support ("211" academic exchange fund program) awarded to Prof Youjian Hu from China University of Geosciences is gratefully acknowledged. Partial financial support from the Australia Research Council Linkage project (LP0455170) endorsed to a research consortium led by A/Prof Kefei Zhang is highly appreciated.

References

Bernese (2004): ***Bernese GPS Data Processing Software Package***, Version 4.2, http://www.bernese.unibe.ch/.

Brown N. (2002): ***Quality Control and Integrity Monitoring of the Victorian GPS Reference Station Network,*** FIG XXII International Congress, Washington DC, April, 2002. 19-26.

Dawson J.; Steed J. (2002): ***International Terrestrial Reference Frame to GDA94 Coordinate Transformations***, http://www.ga.gov.au/image_cache/GA3795.pdf.

Dawson J.; Govind R.; Manning, J. (2004): ***The AUSLIG Online GPS Processing System (AUSPOS)***, http://www.ga.gov.au/nmd/geodesy/sgc/wwwgps/pdf/auspos.pdf.

GAMIT (2004): ***GAMIT-GLOBK GPS Data Processing Software Package***, http://www-gpsg.mit.edu/~simon/gtgk/index.htm.

Geoscience Australia (2004a): ***National Datasets online GIS***. http://www.agso.gov.au/map/national/

Geoscience Australia (2004b): ***Australia Space Geodesy analysis Center***, http://www.ga.gov.au/nmd/geodesy/sgc/gps/index.htm.

Geoscience Australia (2004c): ***Differences between coordinate systems,*** http://www.ga.gov.au/nmd/geodesy/datums/cosys.jsp.

Geoscience Australia (2004d): ***AUSPOS-Online GPS Processing Service,*** http://www.ga.gov.au/nmd/geodesy/sgc/wwwgps/

Hale, M. (2004): ***Personal Communication***.

He XF.; Guang Y.; Ding XL.; Chen YQ (2004): ***Application and evaluation of a GPS-multi-antenna system for dam deformation monitoring***. *Earth Planets and Space* 56 (11): 1035-1039.

IGS (2000): ***International GPS Services (IGS) - International Terrestrial Reference Frame (ITRF),*** http://itrf.ensg.ign.fr/.

King, N. E.; Svarc J. L.; Fogleman E. B.; Gross W. K.; Clark K. W.; Hamilton G. D.; Stiffler C. H.;Sutton J. M. (1995): ***Continuous GPS observations across the Hayward fault, California,*** 1991-1994, *J. Geophys. Res., 100,* 20,271-20,283, 1995.

Kogan M. G., Steblov G. M., King R. W., Herring T. A., Frolov D. I., Egorov S. G., Levin V. Y., Lerner-Lam A., Jones A. (2000): ***Geodetic constraints on the rigidity and relative motion of Eurasia and North America,*** *Geophys. Res. Lett., 27,* 2041-2044, 2000.

Land Victoria (2003): ***GPSnet GPS Base Station Logon and Site Details,*** http://www.land.vic.gov.au/.

TEQC (2004): ***TEQC: The Toolkit for GPS/GLONASS/SBAS Data,*** http://www.unavco.org/facility/software/teqc/teqc.html.

Zhang K.; Roberts C. (2003): ***Recent Development of the Victorian High Precision Permanent GPS Tracking Network with Real-time Positioning Capacity***, Proceedings of the 5th International Symposium on Satellite Navigation Technology and Application, Melbourne, Australia, 22-25 July.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 139-143

Journal of Global Positioning Systems
2005
© CPGPS

Using GPS to enhance digital radio telemetry

K J Parkinson

School of Surveying and Spatial Information Systems, University of New South Wales, Sydney NSW 2052 Australia
Tel: +6494157255 Email: k.parkinson@student.unsw.edu.au

Received: 27 November 2004 / Accepted: 12 July 2005

Abstract. The precise time available from the atomic clocks orbiting the earth in GPS satellites is used in many systems where time synchronization is important. The satellite clocks are monitored and adjusted by ground based control telemetry to within one microsecond of Universal Time. A number of commercial GPS receivers have the ability to provide a time synchronised output, typically one pulse per second, that is locked to this precise time base. This easily accessible timing source is often the justification for including a GPS receiver as an integral component of a complex system. There are additional benefits to be gained from integrating a GPS receiver as an embedded component of a mobile radio telemetry system, where GPS information can also be used to enhance the overall performance. This paper examines some research into combining some transmission techniques with time synchronisation from GPS receivers located in the mobile and in the base equipment to improve a digital radio channel. Using this combined approach, a reverse data channel can be eliminated where a single direction data stream is the predominant requirement.

Key words: 1PPS, FEC, SFH, TDMA, FDMA, Synchronisation.

1 Introduction

Radio communications systems for digital telemetry have undergone many enhancements over the past few decades. The ability to reliably transmit high speed digital data streams is a common requirement. As an example, the IEEE 802.11 Radio Frequency (RF) Local Area Network (LAN) has been widely used, not only in the office environment, but for many applications because of its relatively low cost. As with any RF data channel, some real challenges arise when the remote stations become mobile, and data rates are pushed higher. They suffer from well known propagation difficulties including multi-path and deep signal fade (Lee, 1993), all of which introduce data errors. While there are a number of methods used to correct these errors on the fly, the impact is often unacceptable transmission delays. Particularly damaging to throughput are the methods of successively reducing the data speed and transmission retries. For many static applications such as PC networks, error correction and data integrity are more important than variable delays. The delays experienced during error recovery must be tolerated by the user. For real time systems, this situation is unacceptable because it lengthens the transmission delay making the system response time unpredictable.

This project investigated the application of GPS within a digital telemetry system to bring advantages of robustness, simplification and to help solve some of these problems. This allowed the combining of some conventional methods in a system architecture that may not otherwise be considered viable.

2 802.11 RF LAN Tests

2.1 Initial equipment setup

The aim of this first step was to build and verify the equipment at each end of the radio link before introducing new radio channel equipment using GPS. Commercial 802.11b RF LAN equipment was used as the initial “proof of concept” transport medium connected as shown in Figure 1. 802.11 systems typically use Direct Sequence Spread Spectrum (DSSS) radio system in the licence-free Industrial Scientific and Medical (ISM) 2.4GHz band for the physical layer with a TCP/IP interface to the host equipment.

During development, an opportunity was taken to measure the real time performance of the RF LAN equipment under mobile conditions for later comparison. The accurate time information from a variable number of SuperStar II GPS receivers was transmitted back to a central computer. An embedded microprocessor in an FPGA was used to format a 150 byte message packet and to synchronise the transmission start time to the 1PPS signal from the GPS receiver. The message transmission was started from each remote station simultaneously at 100ms intervals. Because the transmission start time was known, this was able to be compared with the arrival time as measured at the central GPS receiver, in order to reveal the transmission delay.

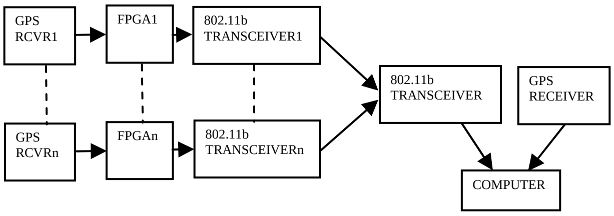


Figure 1. Test system data flow

2.2 Test results and analysis

Although the link was reliable under all Line Of Sight (LOS) conditions, the transmission delay was found to be quite unacceptable. The delay increased with the number of remote stations added, even under ideal conditions, as shown in Figure 2. The small drop in delay using 5 stations suggests that the data packet being sent was close to the optimum size for the internal RF protocol block. Delays of this type were expected because the conventional approach when un-correctable errors are detected in an RF LAN is to use an Automatic Repeat reQuest (ARQ) technique. Naturally, this reduces throughput as it takes extra time to re-send the data. Different message lengths and repetition rates were not tested but it is suspected that the data was broken into smaller blocks to help Forward Error Correction (FEC) and reduce the number of ARQ requests.

Changing the minimal tuning parameters for message length and wait time produced almost no improvement.

This was not intended to be an exhaustive test but it did show that the equipment is not ideal for real time deterministic transmission between more than 3 to 5 remote mobile stations sharing one base station.

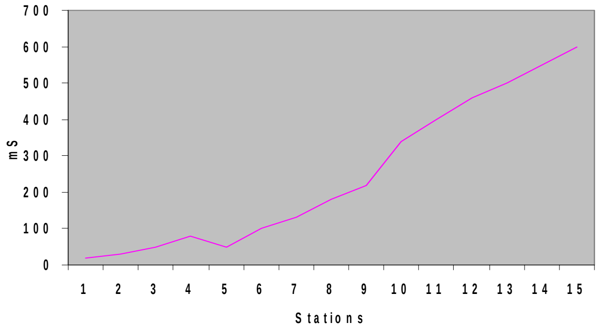


Figure 2. 802.11 average transmission delay

3 A GPS Synchronous solution

A new system was needed to deliver continuous bursts of real time digital data over a variable LOS distance of up to 5km from up to 30 remote mobile transmitters, at 100ms intervals. The remote transmitter needed to be light-weight, battery-operated and small enough to be man- portable. Consistent transmission delay for real time performance and data integrity were important.

The time base from the GPS network was employed as an integral part of the system. The One Pulse Per Second (1PPS), available from many GPS receivers (Mumford, 2003), was provided in this case by Novatel Superstar II receivers as a synchronous time base for both transmit and receive. Testing confirmed that the 1PPS signal between two Superstar II receivers was on average no more than 150ns apart.

This allowed the system to be fully synchronised at both ends using a combination of Slow Frequency Hopping (SFH), Time Division Multiple Access (TDMA) and Frequency Division Multiple Access (FDMA) to provide robust performance.

The ISM band transmitter output was set at a maximum of 400mW which allowed a run time of approximately 2 hours at full power using a small 950mAH cell phone size Lithium Ion battery. Receiver sensitivity was -100dB.

The data speed selected was 288Kbps giving a bit time of 3.47µs. This was fast enough to do the job but low enough to reduce excessive exposure to propagation-induced data errors.

The modulation scheme used was Gaussian Minimum Shift Keying (GMSK) for RF power amplifier efficiency and lower battery drain (Eberspacher and Vogel, 1999).

An FPGA was used in both the transmitter and receiver to perform signal processing, error recovery, frequency hop and synthesiser control functions.

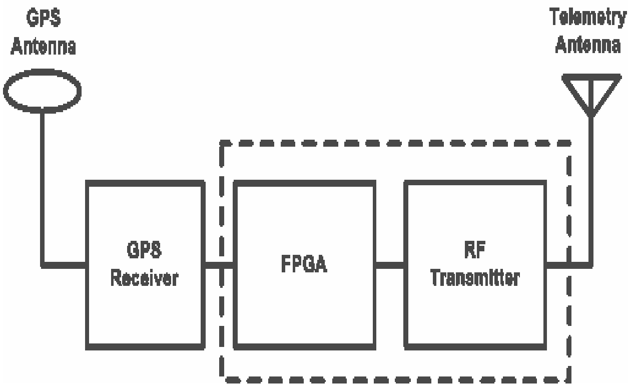


Figure 3. Remote transmitter architecture

3.1 Frequency Hopping Spread Spectrum (FHSS)

A lack of useable spectrum in the ISM band means that there is a risk of interference from other users. Slow Frequency Hopping (SFH) using 75 channels, spaced at 5MHz intervals between 2.405GHz and 2.470GHz, was chosen for this system. Although it is recognised that this approach is vulnerable to partial-band interference, to overcome this, the technique of Dual Frequency Diversity (DFD) (Proakis and Saheli, 2000) was used. The same information is transmitted on two successive frequency hops within the 100ms epoch and the received data from each hop is combined to improve interference rejection and anti-jam capability

Because 1PPS and GPS time were used to synchronise the system, the start time and duration of each hop is known at each transmitter-receiver pair. This simplified the design considerably because the usual adaptive timing recovery circuits and tracking procedures were not required.

3.2 Time Division Multiple Access (TDMA)

TDMA is a common technique used to increase capacity in a communication channel. A successful radio example of this is the Global System for Mobile communications (GSM) mobile telephone network. In order to maintain synchronisation of mobiles, the GSM base station transmits signals on a dedicated channel (Eberspacher and Vogel, 1999). The mobiles must use these signals to synchronise both time and operating frequency.

When using TDMA in a GPS synchronous system, the need for complicated time slot synchronising is eliminated. Each time slot is determined relative to the 1PPS signal. In this case the time slots chosen were 8.5ms long which allowed 5 different transmissions of 2448 bits at 288Kbps in each half (50ms) of the 100ms epoch, as shown in Figure 4.

3.3 Combining SFH, TDMA and FDMA

Using some custom-designed logic in an FPGA at each end of the radio link it was possible to combine both techniques described above with FDMA using the 1PPS signal.

With a pre-allocated orthogonal frequency hopping plan that was known to all transmitters and receivers, it was possible to dedicate a transmitter-receiver pair to a given channel in a given time slot. This effectively added FDMA to the system. Furthermore, it was possible to have a group of transmitter-receiver pairs operate in parallel, knowing that the frequency in use was exclusive to each member of the group. By distributing the channel occupancy of the receivers across the 100ms epoch, it was possible to use less receivers than remote transmitters. The receivers operated in every time slot while the transmitters only operated in two time slots per epoch to achieve DFD.

In this case there were 6 parallel operating receivers using 5 time slots per half epoch (50ms) receiving data from 30 remote transmitters.

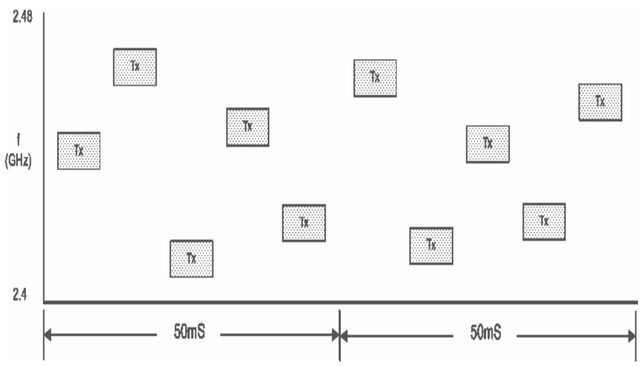


Figure 4. Time slots for one receiver

The maximised use of the spectrum and the radio equipment in this way requires that the hopping table entries are random to satisfy the channel occupancy time and avoid adjacent channel interference. The hopping table contained 300 entries which allowed sequential channel use over a 30 second period before repeating the sequence.

3.4 Error handling approaches

Error handling in transmission often employs a combination of procedures to detect and correct errors after receiving.

Because the data stream contained GPS measurements, there was an opportunity to interpolate some missing samples in downstream processing. Any samples that were completely unrecoverable due to failure to correct errors were marked as unusable. Based on this it was decided to eliminate the ARQ function from the architecture to reduce complexity and power consumption.

This is somewhat unique to the application, as it removes the need for a guaranteed delivery mechanism. In addition to using DFD, this did mean that stronger embedded data stream FEC measures were required to recover errors where they could have been recovered by ARQ.

A combination of approaches was used, as explained below.

3.4.1 Forward Error Correction (FEC)

The FEC technique used relies on the transmission of enough redundant data so that multiple bit errors can be corrected. In this case convolutional encoding was performed in the transmitter while the Viterbi algorithm (Viterbi and Omura 1979) was used for decoding in the receiver.

The encoder processed the message bits with k=1 and v=2 doubling the number of bits to give 100% data redundancy. The constraint length (K) was set to 9 to gain increased robustness. This was built into the transmitter FPGA using an additive shift register of K-1 stages to encode the data using polynomials (1) for the first bit and (2) for the second:

$$g_0(x) = 1+x+x^2+x^3+x^5+x^7+x^8 \quad (1)$$

$$g_1(x) = 1+x+x^2+x^3+x^4+x^8 \quad (2)$$

As the K value increases to strengthen the error correction capability, the downside is the increased processing load in the Viterbi decoder. Although the selection of K=9 was large, this posed no problem at the receiver end because an FPGA is ideally suited to this task.

3.4.2 Interleaving

Interleaving is used to reduce the susceptibility to fading by spreading the data so that all adjacent bits are separated. The level of protection against fade duration is impossible to set for all likely conditions of the radio channel because of the dynamically changing environment. Signal fading characteristics have been modelled (Lee, 1993) to show that fade rate increases as speed increases while fade depth is inversely proportional to speed, and can fluctuate over a large dynamic range from 10 dB to 50 dB.

In this case the message was broken into four blocks of 612 bits giving 612/288000 = 2.125ms of fade duration protection. The buffer was arranged as a rectangular matrix so that data was written by columns and extracted by rows, as shown in Figure 6. This is done inside a RAM buffer in the FPGA before transmit, and reconstructed using the reverse procedure in the receiver.

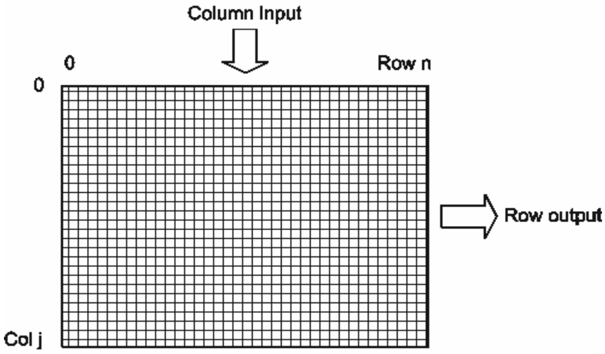


Figure 6. Transmit interleaver

3.4.3 Error checking

The ability to check for errors after receiving is essential because not all errors can be corrected. In this case two measures were used to confirm the reliability of the data.

The first was a 32 bit Cyclic Redundancy Check (CRC) generated from the buffer in the transmitter FPGA and appended to the message before transmission. This gave the ability to detect all error bursts of 32 bits or less at the receiver. Bursts greater than this were also detected but with only slightly less reliability.

The second integrity check was to verify that the 64 bit predicted GPS time value from the Superstar II receiver Measurement Record 23 was actually incrementing in 100ms steps. Any variation of this in the data from the telemetry receiver gave an indication of an error in the data.

4. Performance

The transmission delay was as expected from a synchronous system, and is shown overlayed with the RF LAN results in Figure 7. The transmission delay was 110ms, consistently measured regardless of the number of transmitters operating.

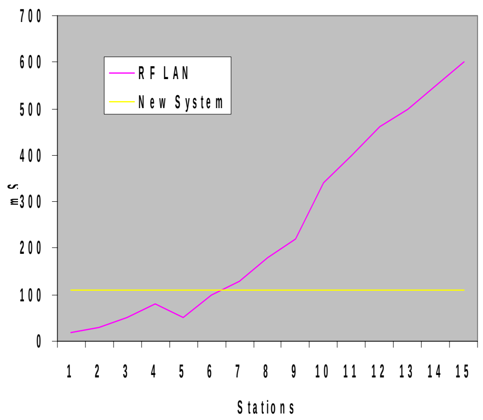


Figure 7. Transmission delay comparison

Table 1 gives a comparison of Bit Error Rates (BER) between different communication paths, including the new system described here.

Table 1. BER comparison

Communication path	Nominal BER
RF, No error correction	10-1 to 10-3
RF LAN	10-5
GSM	10-5 to 10-6
New telemetry system	10-7

5. Concluding remarks

GPS provides advantages when designed into a radio telemetry system because of the ability to make use of synchronisation. While this system has the benefit of multiple receivers, effectively providing a dedicated channel during each frequency hop, there is the added option to use TDMA. The concept allows design flexibility and scalability to meet a number of requirements limited only by the selected frequencies and the processing speed of the electronics.

One of the key system features is that when the number of remote transmitters is scaled up, there is no penalty in transmission delay. The limit may only be a regulatory issue with channel occupancy and dwell time.

Acknowledgements:

The guidance of Prof. Chris Rizos, Dr. P. Wakeman and Dr. Joel Barnes are gratefully acknowledged in supporting this work.

References

Eberspacher J.; Vogel H J. (1999): ***GSM Switching, Services and Protocols,*** J. Wiley & Sons, Chichister, England

Lee W C Y. (1993): ***Mobile Communication Design Fundamentals,*** 2nd ed, J. Wiley & Sons, New York

Mumford P J. (2003): ***Relative timing characteristics of the one pulse per second (1PPS) output pulse of three GPS receivers,*** Satnav 2003, Melbourne, Australia

Proakis J G.; Salehi M. (2000): ***Contemporary communication systems using MATLAB,*** Brookes/Cole, California, 416-421

Viterbi A.; Omura J K. (1979): ***Principles of Digital Communication and Coding,*** McGraw-Hill, New York

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 144-150

Journal of Global Positioning Systems
2005

Development of Navigation Algorithm to Improve Position Accuracy by Using Multi-DGPS Reference Stations' PRC Information

Kyung Ryoon Oh

KARI, 45 Eoeun-dong Yuseong-gu Daejeon Rep. Of Korea
e-mail: bigoh@kari.re.kr Tel: +82-42-860-2825; Fax: +82-42-860-2009

Jong Chul Kim

KARI, 45 Eoeun-dong Yuseong-gu Daejeon Rep. Of Korea
e-mail: jckim@kari.re.kr Tel: +82-42-860-2340; Fax: +82-42-860-2009

Gi Wook Nam

KARI, 45 Eoeun-dong Yuseong-gu Daejeon Rep. Of Korea
e-mail: gwnam@kari.re.kr Tel: +82-42-860-2365; Fax: +82-42-860-2009

Received: 6 December 2004 / Accepted: 20 October 2005

Abstract. In this paper, the linearly interpolated PRC (Pseudorange Correction) regenerating algorithm was applied to improve the DGPS positioning accuracy at user's position by using the various PRC information obtained from multi-DGPS reference stations. The unknown user's position can be calculated from the regenerated PRC which can be expressed as the linear combination of multi-DGPS reference station's known position and PRC values of common satellite from multi-DGPS reference stations. Two sets of 3 DGPS reference stations were selected to compare the performance of the linearly interpolated PRC regenerating algorithm. To test the performance, linearly interpolated PRC regenerating algorithm adopted multi-channel DGPS receiver was developed. The results show that the DGPS positioning accuracy is improved by about 40% and with the modification of the navigation solution software of GBAS receiver, GBAS positioning accuracy improvement is expected without any modification of GBAS reference station's equipment.

Key words: Multi-DGPS coverage, weighting coefficients, Multi-PRC values, linearly interpolated PRC generating algorithm

1 Introduction

Since 1994, Korea Aerospace Research Institute has been conducting the research on the GBAS (Ground Based Augmentation System) for precision approach and landing of aircraft based upon the concept of CNS/ATM (Communication Navigation Surveillance/Air Traffic Management). As the results of this research, Korea Government has developed a plan to install GBAS at each domestic airport for the safety of civil airlines. If the Government's plan is implemented, around the metropolitan area, especially around Seoul, multi-GBAS environment will be established considering the minimum GBAS service coverage as 23NM (Nautical Mile).

The PRC information of each GPS satellite is not varying rapidly; it is possible to assume that the variation of PRC information of each GPS satellite is linear. So the linearly interpolated PRC regenerating algorithm can be applied to improve the DGPS positioning accuracy at user's position by using the various PRC information obtained from multi-DGPS reference stations.

The user's position can be calculated from the regenerated PRC which can be expressed as the linear combination of multi-DGPS reference station's known position and PRC values of common satellite from multi-DGPS reference stations.

To test the performance of the linearly interpolated PRC regenerating algorithm, maritime DGPS reference stations' PRC data were used in RTCM format. 11 maritime DGPS reference stations and 1 inland DGPS reference station are in service since 1999. Two sets of 3 DGPS reference stations are selected to compare the performance of the linearly interpolated PRC regenerating algorithm. The DGPS positioning accuracy was dramatically improved by about 40%.

Even though common PRC was extracted from the RTCM format, the suggested PRC regenerating algorithm in this paper can be applied to improve the DGPS positioning accuracy in GBAS for civil aviation.

With the change of the navigation solution software of the GBAS receiver, GBAS positioning accuracy improvement is expected without any modification of GBAS reference station's equipment.

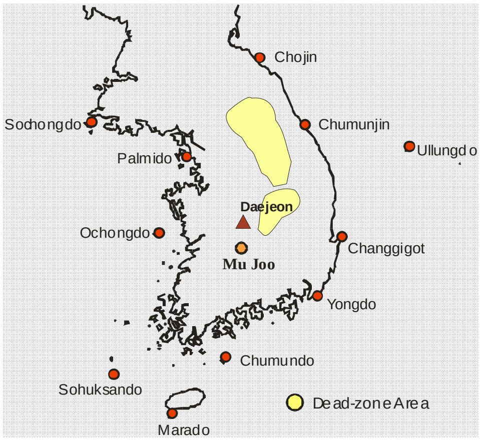


Fig. 1 MOMAF DGPS reference stations in Korea (Jun. 2004)

2 Maritime DGPS reference stations in Korea

Korean Government, Ministry of Maritime Affairs and Fisheries (MOMAF), has started DGPS service from 1999 by the IMO (International Maritime Organization) recommendation of using GNSS in maritime navigation. MOMAF will extend its DGPS infra structure into the inland to establish nationwide DGPS (NDGPS) system by 2006. (Jong Chul Kim, 2002)

In 2004, the first inland DGPS reference station of MOMAF start to provide DGPS service (See the right of Figure 2). 5 more inland DGPS reference stations will be constructed by 2006 to get rid of the dead zone area as shown in Figure 1. So the multiple coverage area will be increased.

3 Navigation solution using multiDGPS information

Due to the characteristics of the GBAS, somewhat extensive network of DGPS reference stations need to be established. In the GBAS coverage, it is possible to receive valid corrections from a number of stations. Within a multiple DGPS reference station solution, all the pseudo-range corrections received from pre-selected reference stations are used to position the mobile station. There are a number of different approaches to providing such a solution.

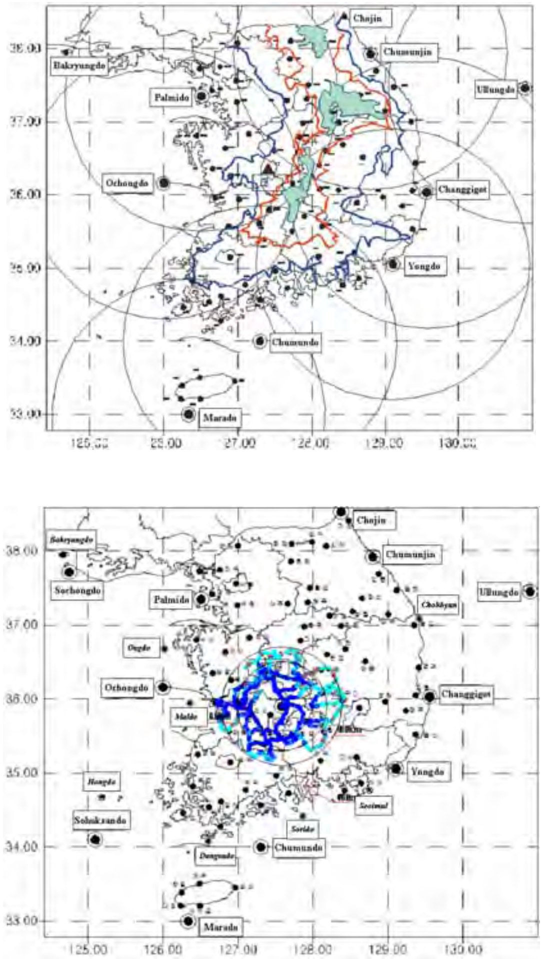


Fig. 2 MOMAF DGPS service coverage (left) and Muju DGPS service coverage (right)

- Position domain approach (See the left of Figure 3): This is the simplest approach which computes an independent position using each reference station from which corrections are received. The resultant positions are later combined by taking a weighted average.
- Centroid approach (See the middle of Figure 3): The pseudo-range corrections from all reference stations are combined to form one correction for each satellite in view. This correction should fit the centroid of the area defined by the reference stations that are used. Additional directional corrections can also be developed by examining the correlation between the composite centroid corrections and those at particular reference stations. The pseudo-range corrections for the centroid can be generated

either at a land-based hub or at the mobile station itself. The advantage of the former is that the mobile station needs only to receive one set of pseudo-range corrections.

- All-in-view approach (See the right of Figure 3): All the pseudo-range corrections received from the reference stations are incorporated into one positioning solution with no pre-processing (except for validity checks). For instance, the correction for satellite PRN 12 may be received from 4 different reference stations and will be used separately to correct the pseudo-range observed at the mobile station from PRN 12 - thus adding 4 observations to the system.

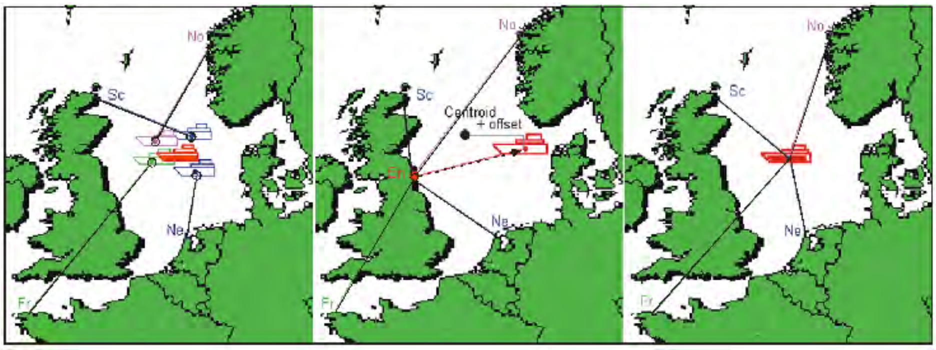


Fig. 3 The positioning methods using multi-DGPS references

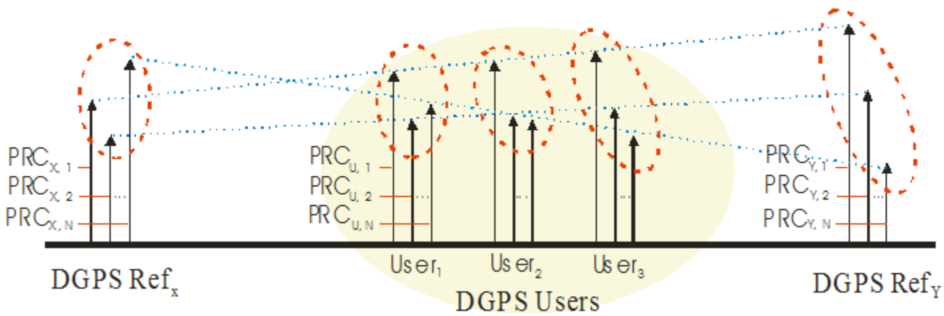


Fig. 4 Linear interpolation of the PRC from two DGPS Reference stations

3.1 Developing the linearly interpolated PRC regenerating algorithm

In developing the linearly interpolated PRC regenerating algorithm, there are some basic assumptions:

a. The user only uses the common in view satellites to calculate hhe positions for both sides of the user and reference stations.

b. At least, 4 common satellites exist between the user and reference stations.

c. The variation of the correction data of a satellite is small enough to assume that the characteristic of the PRC variation for each satellite is linear.

■ **PRC linear interpolating Algorithm:**

In Figure 4, the user will be at user1 or user2, 3 location between DGPS reference station x and station y. The DGPS correction(PRCx,i , PRCy,i, i=1,2,..n) value of common satellite is not the same, so there is a gradient of the DGPS correction value for the common satellite between the DGPS reference stations. If the user can use this gradient information, more precise position information is achievable. (Loomis et al., 1995)

The unknown user's position (longitude, latitude) can be calculated from the regenerated PRC which can be expressed as the linear combination of multi-DGPS reference station's known positions and the PRC values of the common satellite from the multi-DGPS reference stations. (Hong, 1990)

The unknown user's position can be expressed by using the relative geometry information of the stations. (van Essen et al., 1997)

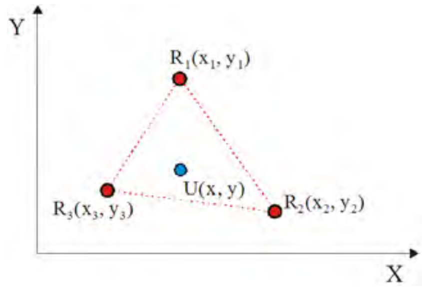


Fig. 5 Geometry relation between the user and DGPS reference station's locations

$$x = \sum_{i=1}^{n} a_i x_i = a_1 x_1 + a_2 x_2 + a_3 x_3 + \ldots + a_n x_n \quad (1)$$

$$y = \sum_{i=1}^{n} a_i y_i = a_1 y_1 + a_2 y_2 + a_3 y_3 + \ldots + a_n y_n \quad (2)$$

$$1 = \sum_{i=1}^{n} a_i = a_1 + a_2 + a_3 + \ldots + a_n \quad (3)$$

Let's assume that the number of reference stations is r, marks as n_r, and the number of satellites in line of sight is s, marks as n^s. And each reference station observes the same GNSS satellites, but the PRC values of specific satellites differ from the DGPS reference stations. Then the linearly interpolated PRC(∇^i_j ,i=1,2,...n^s, j=1,2,..n_r) at the user's spot can be expressed as:

$$\nabla^i_j = \nabla^i_1 + a^i_1 (x_j - x) + a^i_2 (y_j - y) \quad (4)$$

In the above equation, x_i is the latitude y_i is longitude respectively in WGS-84. The parameters a^i_1 and a^i_2 are the coefficients of a plane which contains all the DGPS reference station coordinates.

For the case of using 3 DGPS reference stations, Equation 4 can be written as the following matrix format:

$$\begin{bmatrix} \nabla^i_2 - \nabla^i_1 \\ \nabla^i_3 - \nabla^i_1 \end{bmatrix} = \begin{bmatrix} \Delta x_2 & \Delta y_2 \\ \Delta x_3 & \Delta y_3 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \quad (5)$$

Or,

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \Delta x_2 & \Delta y_2 \\ \Delta x_3 & \Delta y_3 \end{bmatrix}^{-1} \begin{bmatrix} \nabla^i_2 - \nabla^i_1 \\ \nabla^i_3 - \nabla^i_1 \end{bmatrix} \quad (6)$$

Where $\Delta x_j = x_j - x_1$, $\Delta y_j = y_j - y_1$

For the case of using more than 4 DGPS reference stations, the above equations can be written as in general format (van Essen et al., 1997):

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = (G^T G)^{-1} \begin{bmatrix} \nabla^i_2 - \nabla^i_1 \\ \nabla^i_3 - \nabla^i_1 \\ \nabla^i_4 - \nabla^i_1 \\ \vdots \\ \nabla^i_r - \nabla^i_1 \end{bmatrix} \quad (7)$$

Matrix G is a set of known coordinate information of the DGPS reference stations. On the right side of Equation 7, [$\nabla^i_j - \nabla^i_1$] (i=1,2,...n^s, j=1,2,..n_r) term value can be determined using the measurement of PRC information from each DGPS reference station.

With the values for a_1 and a_2, the linearly interpolated PRC (∇^i_j) can be determined using Equation 4.

■ Generating linearly interpolated PRC :

In Figure 6, GPS time in GPS raw data and Modified Z count in DGPS information are compared to check if the data is time synchronized with each other or not. If data is time synchronized, the common satellite number in the data from the Reference Stations is checked. If the number of common satellites is less than 4, the data will be discarded and the next epoch data will be used.

If more than 4 common PRC data exist, the procedure moves to next step. To get the linearly interpolated PRC information, input the user's position into the linearly interpolated PRC regenerating algorithm. Then PRC linear interpolating algorithm will regenerate the new PRC value.

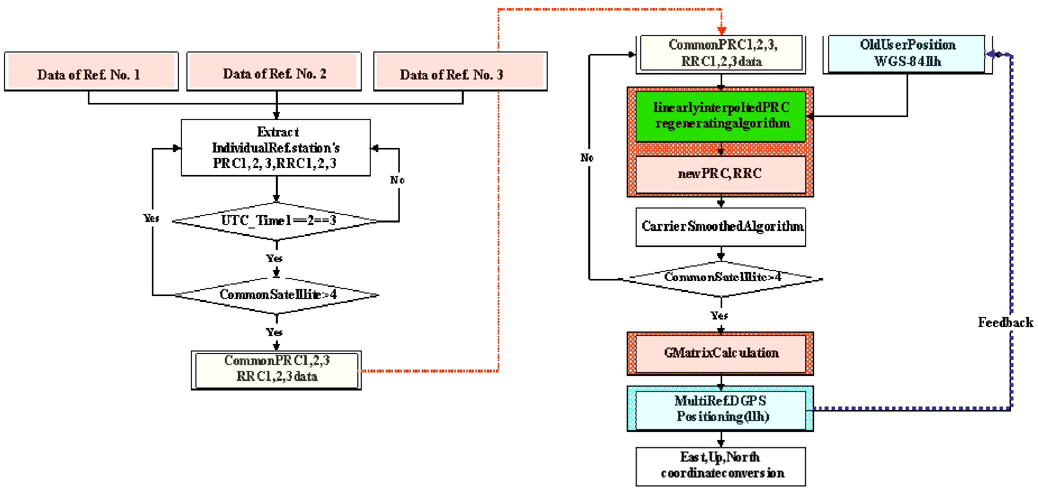


Fig. 6 Diagram of extracting common PRC and generating linearly interpolated PRC

The next procedure to get the DGPS position is explained in Figure 7. In this procedure, the number of common satellites is critical. If the common satellite number is more than 4, the regenerated PRC will be input into the DGPS navigation solution algorithm based on the carrier smoothed algorithm (Park et al, 2003).

3.2 Analysis the effect of the linearly interpolated PRC

■ Phase I :

In phase I analysis, the PRC information of Multi DGPS station gathered through the landline. Each maritime DGPS reference station stored the broadcasted DGPS information every 5 sec. So the analysis was carried out as a post processing.

To analyse the effect of the linearly interpolated PRC algorithm, three sets of DGPS reference stations combination were used. There are 4 DGPS reference stations in the first set, 3 in the second set, 2 in the third set.

As a result of phase I analysis, the second set shows the best results. Comparing the position accuracy with the stand alone DGPS reference station, an average of 33% improvement was achieved. Table.1 shows the results of the analysis of the second set.

In the case of using 2 DGPS reference stations' PRC information, the DGPS position accuracy was 1.8m. Other case of using 3 DGPS reference stations' PRC information, the DGPS positioning was 0.788m and 1.164m depending on which combination of DGPS reference stations used. The last case of using 4 DGPS reference stations' PRC information, the worst result achieved. DGPS position accuracy was 2.449m.

■ Phase II :

In phase II analysis, one 3 channel DGPS receiver was built to field test the performance of the linearly interpolated PRC regenerating algorithm. The built receiver (See Figure 10) can have up to 6 channels.

For the field test, a river side area was selected rather than inland. By the rule of thumb, the medium wave signal propagation characteristic around the river side is better than that of inland.

The PRC values were analysed in real situation and the results shows (See the Table 2) that there was an averaged 4.2% difference between the PRC values of each GPS satellite. The PRC value changes of inbound and outbound GPS satellites are shown in Figure 11.

Tab. 1 The position accuracy using 3 DGPS reference stations

	Multi-Ref.	Changgigot	Ochongdo	Sochongdo	Chumunjin
Distance(Km)		202	127	279	214
Position Error (m)	1.164	1.959	1.607	1.223	
	0.788		1.607	1.223	1.239
Improvement (%)	24.3	40.6	27.6	4.8	
	41.0		51.0	35.6	36.4

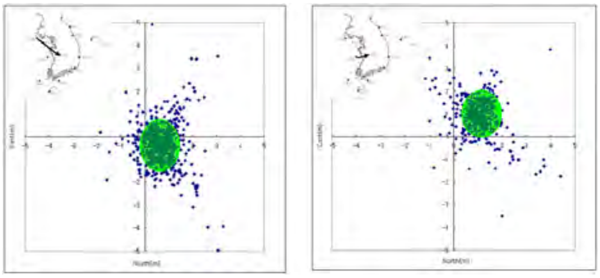


Fig. 7 Scatter plot of DGPS positioning of Sochongdo (left) and Ochongdo (right)

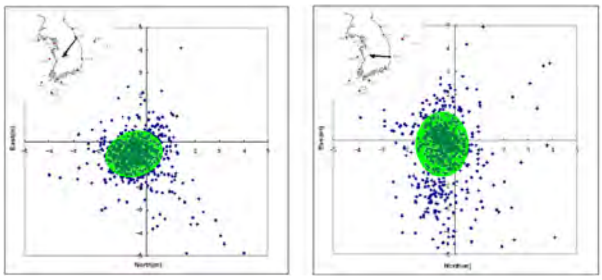


Fig. 8 Scatter plot of DGPS positioning of Chumunjin (left) and Changgigot (right)

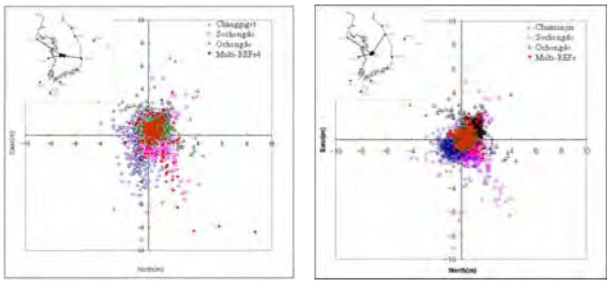


Fig. 9 Multi-Ref.s DGPS positioning accuracy; 1.164 m (left) and 0.788m (right)

Tab. 2 The maximum gap of PRC values

SV / Ref. S.	SV1	SV6	SV14	SV16	SV20	SV25
Muju	-7.1951	-11.521	-11.191	-12.100	-18.858	-10.069
Ochongdo	-6.9579	-12.078	-11.328	-11.218	-18.387	-10.034
Palmido	-7.3114	-11.718	-11.691	-11.914	-18.574	-10.181
PRC gap	4.84(%)	4.61(%)	4.28(%)	7.29(%)	2.50(%)	1.45(%)

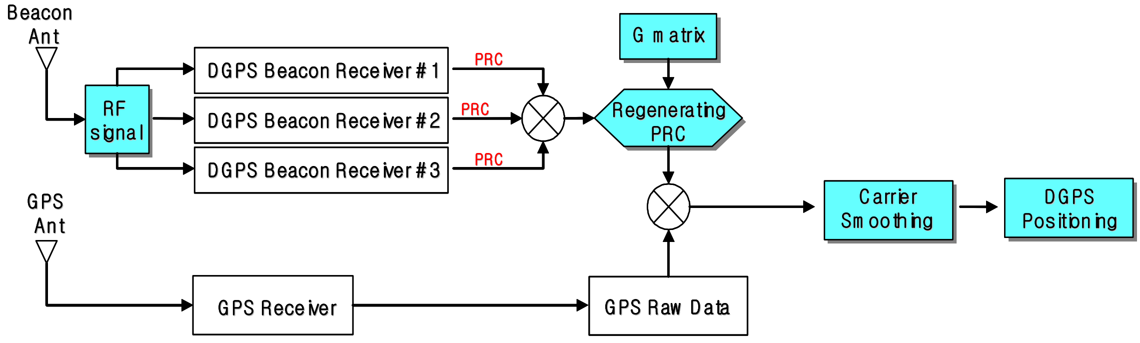


Fig. 10 The configuration of built receiver for field test

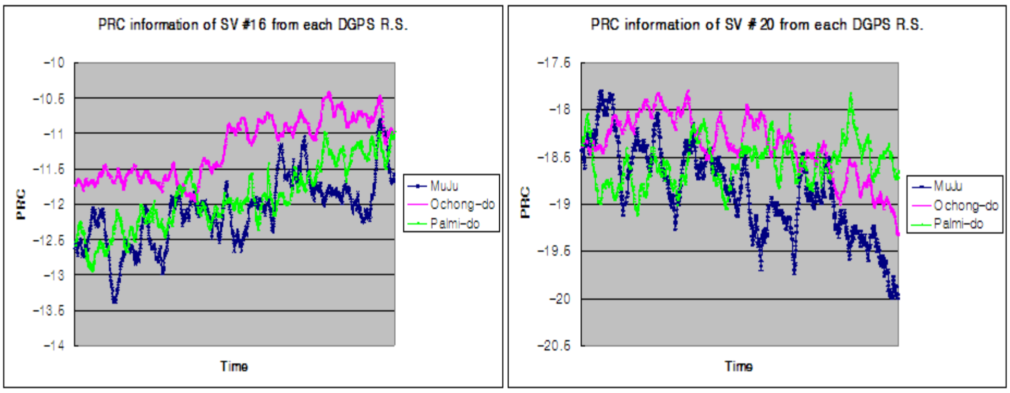


Fig. 11 The tendency of PRC value changes

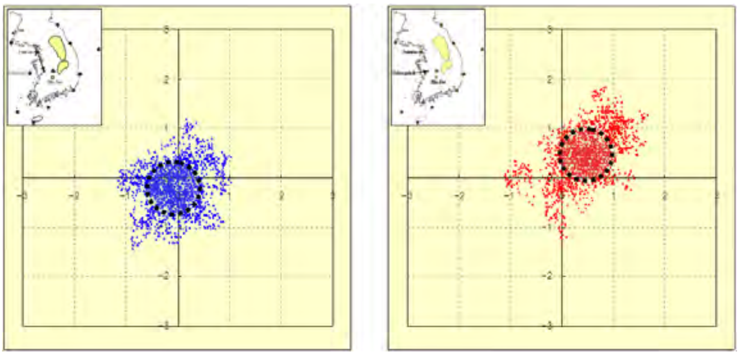


Fig. 12 Field test results of DGPS positioning of Palmido (left) and Ochongdo (right)

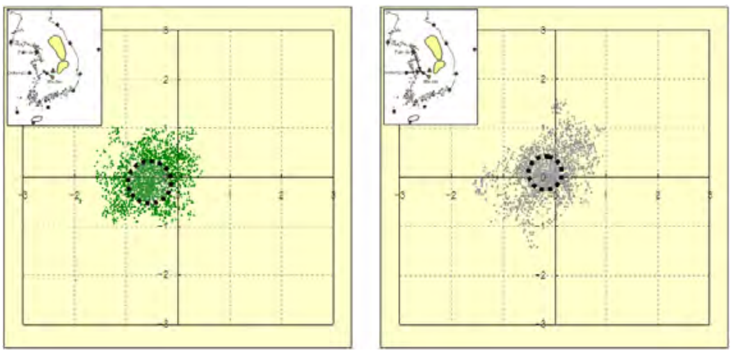


Fig. 13 Field test results of DGPS positioning of Muju (left) and Multi-DGPS R.S (right)

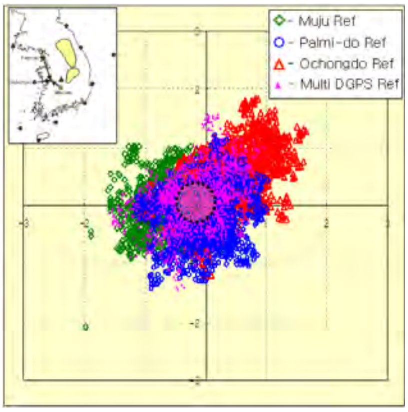


Fig. 14 Scatter plot of the field test results of DGPS positioning

4 Conclusions

The linearly interpolated PRC regenerating algorithm has been proposed in this paper, which can improve the DGPS position accuracy by about 40% without any changes in DGPS reference station's system. In the phase I study, off-lined PRC data was used. DGPS RF signals directly from the DGPS stations are available through a multi channel DGPS beacon receiver developed in phase II study. The results of phase II study shows that the PRC regenerating algorithm works well.

Acknowledgement s

This study was supported by the 2004 National R&D Program. The authors greatly appreciate the support of Korea Research Council of Public Science & Technology (KORP), Korea.

References

Hong G.D., (1990): ***Linear controllable systems.*** Nature, Vol. 135, 18-27

Kim J.C., Oh K.R. (2002): ***Development of the strategies for NDGPS in Korea.*** Korea Aerospace Research Institute, Final Report, Aug.

Loomis P, Sheynblatt L, Muller T(1995): ***Differential GPS Network Design*.** Proceedings on ION-GPS '95, The Institute of Navigation, Alexandria, VA, U.S.A., 511-520

van Essen R.F., Offermans G.W.A., Helwig A.W.S., van Wiiligen D. (1997): ***Regional Area Augmentation Concept for Eurofix-Reducing Spatial Decorrelation Effects through Multi-station DGPS.*** 26th Annual Technical Symposium of the International Loran Association, Ottawa, Canada, 6-9 Oct.

Park Y.H., Oh K.R., Kim J.C. (2003): ***Development of Linearly Interpolated PRC Generating Algorithm to Improve Navigation Solution Using Multi-DGPS Reference Stations.*** 10th GNSS workshop, phoenixpark, Rep. of Korea, 21-22 Nov., 360-363

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 151-159

Journal of Global Positioning Systems
2005
© CPGPS

Quality Monitoring for Multipath Affected GPS Signals

Maurizio Fantino, Fabio Dovis

Politecnico di Torino, Electronics Department , c.so Duca degli Abruzzi 24 10129 Torino Italy
e-mail: name.surname@polito.it Tel: + 390112276416; Fax: +390115644099

Jinling Wang

School of Surveying & Spatial Information Systems, University of New South Wales Sydney Australia
e-mail: jinling.wang @unsw.edu.au Tel: + 61293854203; Fax: +61293137493

Received: 8 December 2004 / Accepted: 12 July 2005

Abstract. The ability to monitor and detect any disturbances on the PRN code signals transmitted from the navigation satellite constellation is of primary importance. It is known that the tracking performance of a navigation receiver stems from the correlation property of the PRN code signals transmitted. These anomalies can be detected in several different ways, either observing the outputs of navigation user receivers, or processing the received signal within the receiver. Quality control is the process that defines how well the solution of a problem is known and in the context of navigation, it consists of assuring an agreed level of accuracy, reliability and robustness for the measurements. In this work a modified version of the conventional tracking scheme will be proposed with the aim of monitoring the quality of the measurements at the signal processing level. The proposed tracking scheme is able to give a measure of the distortion of the correlation function and consequently, of the reliability of the signal tracked. In particular the problem of multipath distortion is considered The amplitude and multipath delay can be estimated with an extension of the linear Kalman Filter which can be implemented inside the traditional DLL architecture. Simulations show that due to its prediction capability, Kalman Filter enhances the robustness of the system when weak signals are present or there is loss of lock on the signals, trading off the performance improvement with an increase in complexity of the new architecture. The recognition of a multipath corrupted signal estimating the amplitude and delay of the reflection can be used to select the more reliable pseudo-range measurements for the evaluation of the positioning equations. Mitigation of the multipath effects may be performed where the number of tracked signals is not sufficient.

Key words: GPS, Multipath, Kalman Filter, Quality Control, Monitoring.

---

1 Introduction

Quality control is the process that defines how well the solution of a problem is known and it consists of assuring an agreed level of accuracy, reliability and robustness for the estimated position. Techniques for evaluating the quality of the estimated position solution can be based on the observation of several different parameters and they can be assessed at different navigation system levels. A well established way to evaluate the quality of positioning is based on the processing of several Position, Velocity and Time (PVT) solutions, at the output of the receivers. For instance, Receiver Autonomous Integrity Monitoring (RAIM) techniques have been developed and refined over the past 10 years to ensure that a given solution is within tolerable constraints. Hewitson et al (2004) provide a good overview of the literature relating to the significant developments and studies in RAIM techniques. Other possible strategies can be based on Dilution of Precision (DOP) indicators such as BDOPs for baseline relative positioning and ADOPs specifically for ambiguity resolution, see Leick (2004).

A different approach to the quality monitoring can be based on the processing of the raw data received, examples such as, raw pseudoranges, or carrier frequency/phase, priori to the solution computation. For example, there has been a large increase in the use of quality control techniques in the context of GPS surveying where they are applied to ensure accurate and reliable survey results see Wang et al (1999).

Although they have become an integral part of the GPS Surveying process, the implementation of Quality Control principles inside the tracking loop at the signal processing level, prior to the measurements stage, are not so common. Significant efforts over the past four years have been made to develop and analyse Monitoring Techniques and interference detection strategies based on the analysis and shape of the PRN autocorrelation function. Shloss et al. (2002), amongst others, discusses the threats, detection requirements, and detector design approach to mitigate the failures in the WAAS LNAV/VNAV system. Macabieu et al. (2000) analyses the latest proposed ground Signal Quality Monitoring (SQM) techniques against several types of failures and Evil Waveforms on the GPS signal. A multicorrelator scheme for interference monitoring and a metric test based approach for signal validation is presented in Mitelman et al. (2000) More recently Mark L. et al. (2002) and Jee et al. (2002) propose to use an Extended Kalman Filter based tracking loop for weak and multipath affected GPS signals.

The conventional Delay Locked Loop (DLL) uses discriminator functions constructed from the combination of early, prompt, and late correlators; for example, early-minus-late to detect code tracking error. It is well known that this architecture suffers from performance degradation due to error sources like multipath, loss of signal, and weak signals. A possible extension of this architecture is to integrate the quality process in tracking measurements which consists of a loop with multiple correlators, an opportune Kalman filter and a loop filter. The Kalman filter estimates the code tracking errors from the corrupted input signals by averaging multiple samples of the PRN code autocorrelation function. In the instance of an opportune stochastic model, the Kalman filter can be used to evaluate the multipath components, and mitigate the loss gain in the discrimination function and predict the system evolution of the incoming signal even during a momentary loss of the signal condition. The inclusion of Kalman filters into the tracking block can also be used to estimate the reliability of the incoming signal by comparing the measurement results with an opportune cost function and, where possible, mitigate the influence of temporary interferences and system lags using, for example, maximum likelihood techniques.

This paper will discuss a new way to track GNSS signals using quality control techniques in order to improve the system performance in accuracy and robustness, and also mitigate the effects due to the distortion of the autocorrelation function caused by a single multipath ray. Section 2 introduces the interferences which may affect the GPS signal; the multipath affected channel model is presented in Section 3. Section 4 introduces the GPS signal model while Section 5 deals with the Kalman based architecture. Finally, Section 6 presents and analyses some simulation results for a single multipath corrupted signal.

2 Potential Faults and Interference in GPS

Potentially hazardous signal faults may occur due to unintentional or jamming signals in user, ground or space segments of GPS, leading to corrupted C/A code spectrum and distortions in the correlation function. The most significant GPS interferences and faults are listed below, see Phelts *et al*., (2000):

Wide Band & Narrow Band Interferences

Wide band interference (i.e. white Gaussian noise) is a signal with a constant energy spectrum over all frequencies, whereas Narrow band interference has a limited bandwidth, usually less than a few MHz.

Evil Waveforms

Under the name of "Evil Waveforms" are classified, all the signal failures resulting in a malfunction of the signal generator on board the GPS space vehicles. These anomalies may cause severe distortion in the autocorrelation shape and peak, but fortunately they occur rarely. However, in local area differential systems, undetected Evil Waveforms may result in large pseudorange errors.

Multipath

The signal distortion caused due to reflections is a well known phenomenon, which will be discussed extensively in Section 3.

This paper focuses on detection and monitoring scheme for the distortions caused by multipath in the PRN autocorrelation function.

3 Multipath Channel Model

Multipath is caused by the reflection of the satellite signals from the environment around the receiver such as the ground, buildings or other obstacles. The received signal can be modeled as the sum of the line of sight satellite signal and the reflections with different amplitudes and delays, see Braasch (2001):

$$s_R(t) = s(t) + \sum_{k=1}^{N} m_k s(t - \tau_k) \tag{1}$$

where $s(t)$ is the nominal C/A code, m_k, τ_k are the relative amplitude and delay of the kth echo, respectively. This expression may, in general, be used to model ground-based multipath as well as anomalies originating on board the space vehicles, as is the case with misterminated transmission lines which can be represented with the equation above. This type of failure is not uncommon in radio frequency applications involving transmission lines. In this case, the echoes decrease geometrically in amplitude while the delays are multiples of the round trip times.

The correlation peak for the case of a single reflection is shown in Fig. 1. The thin solid line represents the nominal correlation due to the line of sight signal; the dashed line is the echo; and the heavy solid line is the composite peak which is what a receiver actually processes.

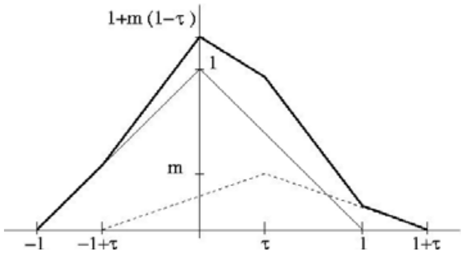


Fig. 1: Single reflection contribution

4 Signal Model

The received GPS C/A code signal can be modelled as, Kaplan (1996)

$$s(t) = \sqrt{2P_s} D(t-\tau) PN(t-\tau) \cdot \sin\left[2\pi f_c t + \theta_D(t)\right] + n_T(t) \tag{2}$$

where P_s is the transmitted signal power, $D(t)$ is the data modulation at 50 bit/s, and f_c the L1 carrier frequency of 1575.42 MHz. $PN(t)$ is the pseudorandom code modulation defined by

$$PN(t) = \sum_{m=-\infty}^{\infty} \sum_{k=0}^{L_{ca}-1} c_k P_{T_c}(t - kT_c - mT_{ca}) \tag{3}$$

where $L_{ca} = 1023$ chips, $T_{ca} = 1\,ms$, $T_c = T_{ca} / L_{ca}$ the chip duration. P_{T_c} is the pulse function, c_k is the C/A code sequence, and $n_T(t)$ is the white Gaussian noise.

Ignoring noise, a sampled model of the received signal is given as

$$s(k) = \sqrt{2P_s} D(k) \cdot PN(k) \cdot \sin\left[2\pi(f_c + \Delta f_D) t_k + \phi_0\right] \tag{4}$$

The quadrature signals of the channel, *I (in-phase) and Q (quadrature)*, can be obtained by multiplying the received signal with the locally generated code and carrier estimates

$$I_{\delta_j} = \frac{\sqrt{2P_s}}{2} R_{PN}(\tau + \delta_j) \sin\left[2\pi \Delta f_D T_n + \Delta\phi\right] \cdot \frac{\sin(\pi \Delta f_D T_n)}{\pi \Delta f_D T_n} + n_I \tag{5}$$

$$Q_{\delta_j} = \frac{\sqrt{2P_s}}{2} R_{PN}(\tau + \delta_j) \cos\left[2\pi \Delta f_D T_n + \Delta\phi\right] \cdot \frac{\sin(\pi \Delta f_D T_n)}{\pi \Delta f_D T_n} + n_Q \tag{6}$$

where T_n is the average time of the accumulator output, and $R_{PN}(\tau)$ is the code autocorrelation function of PN sequence which can be expressed by:

$$R_{PN}(\tau) = -\frac{2P_s}{L_{ca}} + \frac{L_{ca}+1}{L_{ca}} \cdot R(\tau) \otimes \sum_{m=-\infty}^{\infty} \delta(\tau - mL_{ca}T_c) \tag{7}$$

with

$$R(\tau) = \begin{cases} 2P_s\left(1 - \frac{|\tau|}{T_c}\right) & |\tau| \le T_c \\ 0 & \text{elsewhere} \end{cases} \tag{8}$$

The received signal is considered to be affected by a single multipath component. According to (1) the received GPS signal represented by (10) can therefore be written as:

$$\begin{aligned} s(k) &= A_{LoS} D(k - \tau_{LoS}) \cdot PN(k - \tau_{LoS}) \\ &\cdot \sin\left[2\pi(f_c + \Delta f_D) t_k + \phi_{LoS}\right] \\ &+ A_M D(k - \tau_{LoS} - \tau_M) \cdot PN(k - \tau_{LoS} - \tau_M) \\ &\cdot \sin\left[2\pi(f_c + \Delta f_D) t_k + \phi_M\right] + n_k . \end{aligned} \tag{9}$$

where A_{LoS} and τ_{LoS} are the Line of Sight (LoS) amplitude and signal delay, respectively and A_M and τ_M are the amplitude and signal delay of the multipath ray.

5 Extended Kalman Filter based tracking architecture

5.1 Tracking Loop Architecture

In a navigation receiver the code tracking block tries to maximize the cross-correlation between the local generated code and the received signal, on the basis of the autocorrelation function of the PRN codes. In fact, when the codes are perfectly aligned the auto-correlation assumes the maximum value. Lock of the signal might be maintained by feeding back a proper control signal which regulates the local code phase.

When multipath is present, it changes the cross-correlation function used for the alignment in the tracking stage. There are several techniques to mitigate the multipath effect, such as the Narrow correlator, Edge and Strobe correlators. None of these methods give information about how much the cross-correlation function departs from the triangular shape as a fault consequence, Jee (2002).

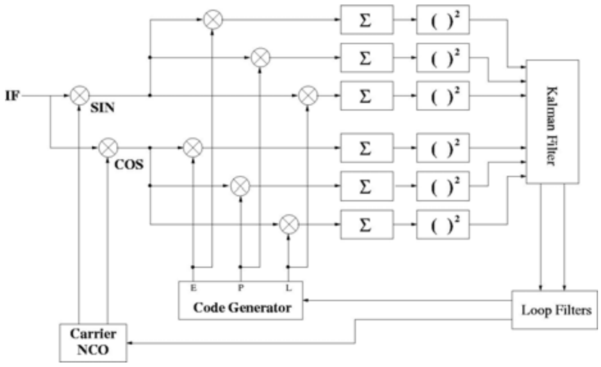


Fig. 2: Modified Tracking Loop Architecture

Extended Kalman Filter (EKF) can be included in the tracking loop with an aim of detecting anomalies on the cross-correlation function and monitoring the reliabilities of the tracked signals. The modified tracking architecture is presented in Fig. 2. In the case of the single component multipath model, where the amplitude and delay of the reflection are measured it is also possible to evaluate the error on the measurement and then estimate the current tracked signal quality.

5.2 Extended Kalman Filter (EKF)

Kalman Filter is a set of mathematical equations that provide an efficient computational means to estimate the state of a process, in a way that minimizes the mean of the squared error. The filter is a very powerful tool and it can estimate past and future states even when the precise nature of the modeled system is unknown.

The EKF uses the system model and measurement model. The system model projects the state ahead in time in the presence of noise v_k :

$$x_{k+1} = f(x_k) + v_k \, , \tag{10}$$

where $f()$ is the non-linear transition function and $v_k \approx N(0, Q(k))$. The noise is assumed to be white, zero-mean Gaussian with variance $Q(k)$. The measurement model relates the observations z_k to the state of the system, and has the following form:

$$z_k = h(x_k) + w_k \, , \tag{11}$$

where $h()$ is, usually, a non linear function and $w_k \approx N(0, R(k))$. It is assumed that the measurements are corrupted by additive, white, zero-mean Gaussian noise with variance $R(k)$. The estimation algorithm produces an estimate of the state $\hat{x}_{k+1,k}$ of the system at the step $k+1$ based on the previously updated estimate of the state $\hat{x}_{k,k}$ and the observations z_{k+1}. The basic steps of the algorithm are, Ronald (1999):

State time propagation:

$$\hat{x}_{k+1,k} = f(\hat{x}_{k,k}) \tag{12}$$

Covariance time propagation:

$$P(k+1,k) = f(k)P(k,k)F^T(k) + Q(k) \tag{13}$$

where $F(k)$ is the Jacobian of f obtained by linearizing about the updated state estimate $\hat{x}_{k,k}$:

$$F(k) = \nabla f(\hat{x}_{k,k}). \tag{14}$$

Kalman gain calculation

$$K(k+1) = P(k+1,k)H^T(K+1) \cdot \\ \left[H(k+1)P(k+1,k)H^T(k+1) + R(k+1) \right] \tag{15}$$

where $H(k+1) = \nabla h(\hat{x}_{k+1,k})$. (16)

State measurement update:

$$\hat{x}_{k+1,k+1} = \hat{x}_{k+1,k} + K(k+1)\left[z_{k+1} - h(\hat{x}_{k+1,k}) \right] \tag{17}$$

Covariance measurement update:

$$P(k+1,k+1) = \left[I - K(k+1)H(k+1) \right] P(k+1,k) \tag{18}$$

5.3 First-order Divided Difference (DD1)

Until now the EKF has undoubtedly been the dominant estimation technique. The EKF is based on the first-order Taylor approximations of state transition and observation equations about the estimated state trajectory. Application of the filter is therefore based upon the assumption that the required derivates exist and can be obtained with a reasonable effort. The Taylor linearization provides an insufficiently accurate representation in many cases. Significant bias, or even convergence problems, are also encountered due to the overly crude approximation.

This study shows that a new non linear extension of the celebrated Kalman filter can have a better performance in the tracking loop problem discussed here. The first-order divided difference filter proposed by Magnus et *al.*

(2000) is based on polynomial approximations of the nonlinear transformation obtained with particular multidimensional extension of Stirling's interpolation formula. Let the operator f'_{DD} perform the following operation (h denotes a selected *interval length*)

$$f'_{DD}(\bar{x}) = \frac{f(\bar{x}+h) - f(\bar{x}-h)}{2h} \tag{19}$$

With $x = \bar{x}$ the point around the interpolation is made, the first-order Stirling's interpolation formula can be expressed as:

$$f(x) \approx f(\bar{x}) + f'_{DD}(\bar{x})(x-\bar{x}) \tag{20}$$

In contrast to the Taylor approximation no derivatives are needed in the interpolation formula; only function evaluations. This accommodates an easy implementation of the filter, and it enables state estimation even when there are singular points in which the derivatives are undefined.

5.4 Filter Design

To design a non linear estimator for the problem, several assumptions must be made. The signal parameters could be modeled as a random walk sequence, with noise as a zero mean Gaussian process. The received signal is considered to be affected by a single multipath component. Under these assumptions, according to (9) the j^{th} branch correlator output can be written as:

$$\begin{aligned} I_{\delta_j} &= A_{LoS} R_{PN}(\tau_{LoS} + \delta_j) \\ &\sin\left[2\pi\Delta f_D T_n + \Delta\phi_{LoS}\right] \cdot \frac{\sin(\pi\Delta f_D T_n)}{\pi\Delta f_D T_n} \\ &+ A_M R_{PN}(\tau_{LoS} + \tau_M + \delta_j) \\ &\sin\left[2\pi\Delta f_D T_n + \Delta\phi_M\right] \cdot \frac{\sin(\pi\Delta f_D T_n)}{\pi\Delta f_D T_n} + n_I \end{aligned} \tag{21}$$

$$\begin{aligned} Q_{\delta_j} &= A_{LoS} R_{PN}(\tau_{LoS} + \delta_j) \cdot \\ &\cos\left[2\pi\Delta f_D T_n + \Delta\phi_{LoS}\right] \frac{\sin(\pi\Delta f_D T_n)}{\pi\Delta f_D T_n} \\ &+ A_M R_{PN}(\tau_{LoS} + \tau_M + \delta_j) \cdot \\ &\cos\left[2\pi\Delta f_D T_n + \Delta\phi_M\right] \frac{\sin(\pi\Delta f_D T_n)}{\pi\Delta f_D T_n} + n_Q \end{aligned} \tag{22}$$

Defining the state variable vector as

$$\bar{x} \equiv \left[A_{LoS}, \tau_{LoS}, A_M, \tau_M, \Delta f_D, \phi_{LoS}, \phi_M\right]^T \tag{23}$$

the corresponding system dynamic matrix model under the previous assumptions is given as:

$$\bar{F} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2\pi T_n & 1 & 0 \\ 0 & 0 & 0 & 0 & 2\pi T_n & 0 & 1 \end{bmatrix} \tag{24}$$

while the vector measurements for the estimators are the I and Q samples from the correlator branches.

$$\bar{z} \equiv \left[I_{\delta_1}, I_{\delta_2}, \cdots, I_{\delta_i},\ Q_{\delta_1}, Q_{\delta_2}, \cdots, Q_{\delta_i}\right]^T \tag{25}$$

In cases where the EKF is used inside the loop, the corresponding linearized measurement matrix H is given by evaluating the following derivative, Jee (2002):

$$\begin{aligned} &\frac{\partial\left\lfloor I_{\delta_j}, Q_{\delta_j}\right\rfloor}{\partial A_{LoS}}, \frac{\partial\left\lfloor I_{\delta_j}, Q_{\delta_j}\right\rfloor}{\partial \tau_{LoS}}, \frac{\partial\left\lfloor I_{\delta_j}, Q_{\delta_j}\right\rfloor}{\partial A_M}, \frac{\partial\left\lfloor I_{\delta_j}, Q_{\delta_j}\right\rfloor}{\partial \tau_M}, \\ &\frac{\partial\left\lfloor I_{\delta_j}, Q_{\delta_j}\right\rfloor}{\partial \Delta f_D}, \frac{\partial\left\lfloor I_{\delta_j}, Q_{\delta_j}\right\rfloor}{\partial \phi_{LoS}}, \frac{\partial\left\lfloor I_{\delta_j}, Q_{\delta_j}\right\rfloor}{\partial \phi_M} \end{aligned} \tag{26}$$

No derivatives must be computed in the case of the DD1 Filter.

6 Simulation Results

The architecture presented in Fig. 2 has been tested in simulation with both the EKF and DD1 filter. For the sake of simplicity, a single multipath ray channel as described in Section 2.4 is considered. The processing gain is adjusted to be high so that the autocorrelation function $R_{PN}(\tau)$ can be well approximated as in (8). To analyse a signal under the real GPS conditions a Signal to Noise Ratio of 40 dBHz is considered. A 10 MHz Intermediate Frequency (IF) filter bandwidth has been used to reduce the channel noise on the simulated GPS signal.

The simulation results presented in Fig. 3 to Fig. 6 shows the capability of the EKF to estimate a single multipath component in terms of amplitude and relative delay, when the ray with ¼ power of the direct component is 0.3 chip distant from the direct component LOS. The same analysis is performed by employing the DD1 filter, and the results are shown in Fig. 7 to Fig. 10. The performances of the filters have been evaluated by comparing the actual observations, along with three times their (estimated) standard deviations (confidence intervals – dashed lines).

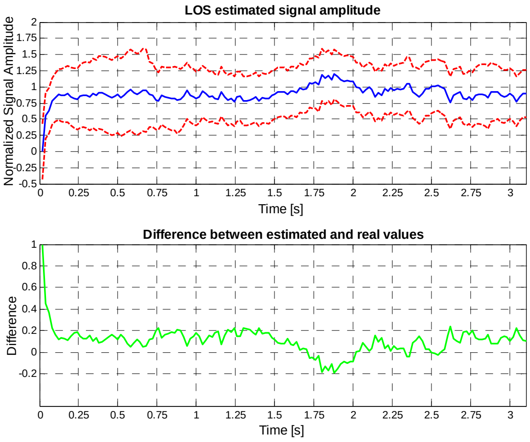


Fig. 3 LOS Estimated Amplitude (EKF)

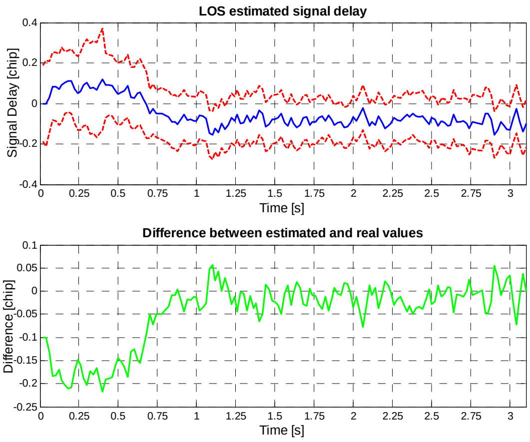


Fig. 4 LOS Estimated Delay (EKF)

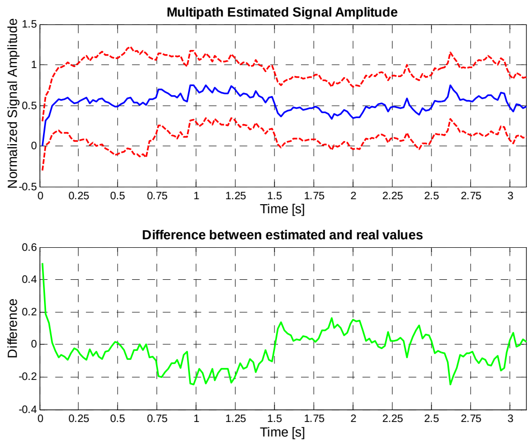


Fig. 5 Estimated Multipath Amplitude (EKF)

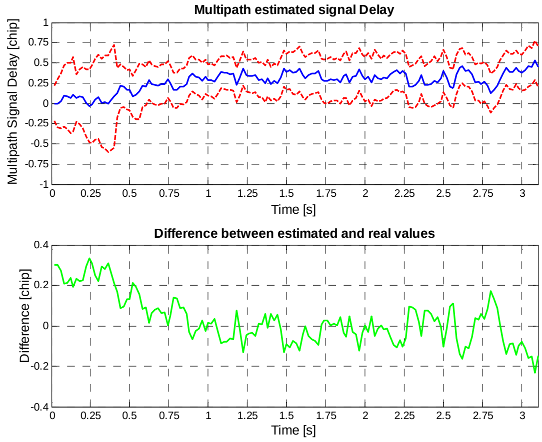


Fig. 6 Estimated Multipath Delay (EKF)

In Tab. 1 and Tab. 2 the major statistics on the results have been reported.

	Mean Value	Average error	Error Variance	Upper confidence limit	Lower confidence limit
Normalized LOS Amplitude	0.9531	0.0469	9.9E-03	1.2624	0.5346
LOS Delay [chip]	-0.088	-0.012	6.2E-04	-0.2164	0.0187
Normalized Multipath Amplitude	0.4964	0.0036	7.2E-03	0.8525	0.1121
Multipath Delay [chip]	0.3278	-0.0278	6.6E-03	0.6974	0.195

Tab. 1 Statistics for a single multipath ray affected signal when EKF is used

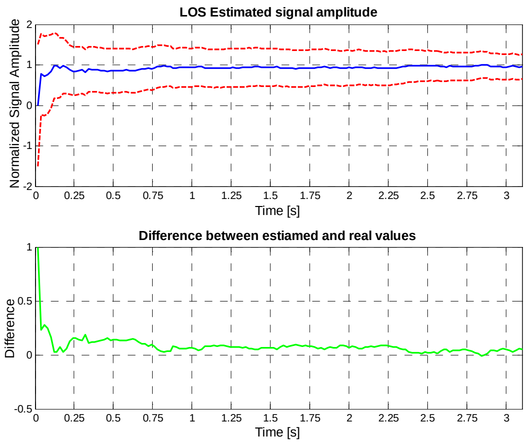


Fig. 7 LOS Estimated Amplitude (DD1)

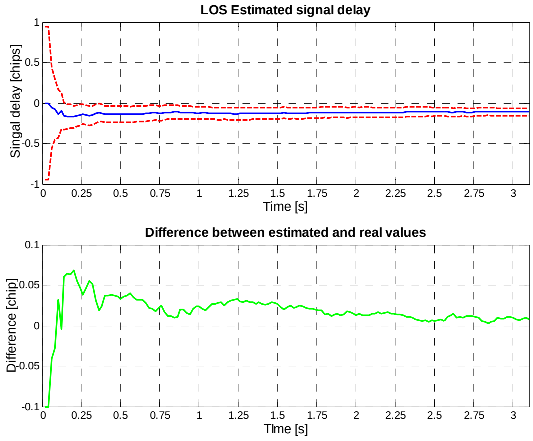


Fig. 8 LOS Estimated Delay (DD1)

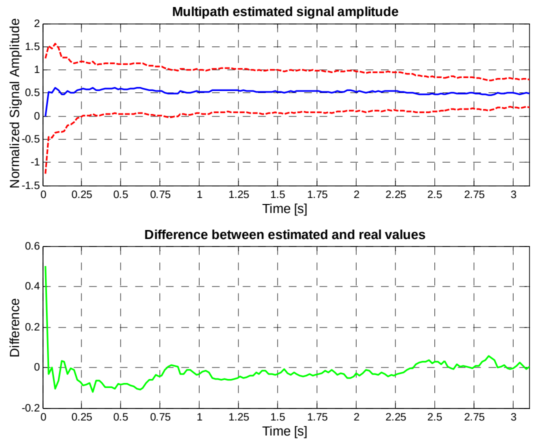


Fig. 9 Estimated Multipath Amplitude (DD1)

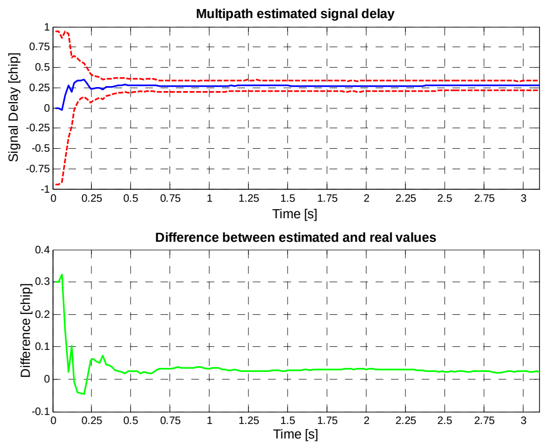


Fig. 10 Estimated Multipath Delay (DD1)

	Mean Value	Average error	Error Variance	Upper confidence limit	Lower confidence limit
Normalized LOS Amplitude	0.9467	0.0533	6.3E-04	1.2602	0.6468
LOS Delay [chip]	-0.1121	0.0121	2.5E-05	-0.1525	-0.0633
Normalized Multipath Amplitude	0.5085	-0.0085	7.1E-04	0.7968	0.1914
Multipath Delay [chip]	0.273	0.027	1.3E-05	0.3352	0.22

Tab. 2 Statistics for a single multipath ray affected signal when DD1 is used

By comparing the results it is possible to deduce that the DD1 filter has a shorter convergence time than the EKF, and the states observations have smaller confidence intervals which mean a better accuracy and estimation closer to the real values.

Moreover, due to the nature of the problem, the EKF suffers from stability problems and tuning sensitivity of the Q and P matrices. This stems from the shape of the autocorrelation functions and theirs derivatives. These functions, besides being dependent on the IF filter bandwidth as reported in Fig. 11 and 12 can only give a raw approximation of the non linear model. The DD1 filter which is based on an interpolation technique is less sensitive and consequently more stable. Furthermore the DD1 filter can be successfully used in the monitoring of multipath component with a relative delay of 0.1 chip or less.

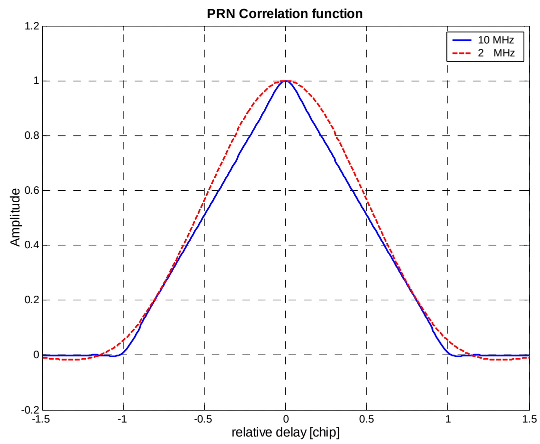


Fig. 11 Correlation function shaping for different IF filter bandwidths

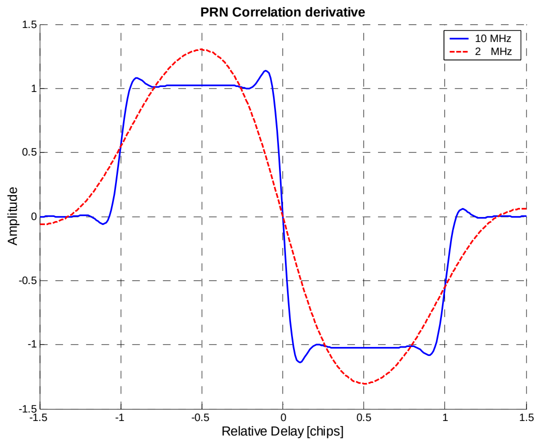


Fig. 12 Correlation function derivative for different IF filter bandwidths

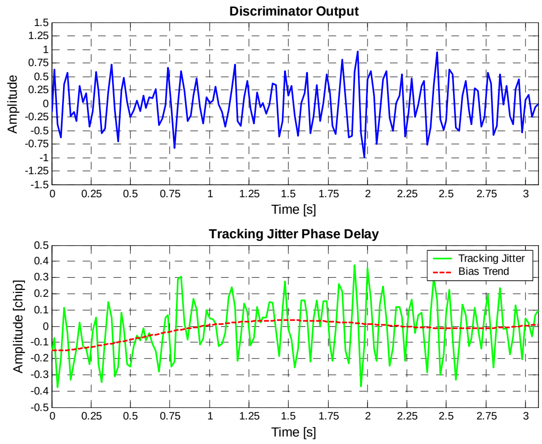


Fig. 13 Discriminator Output and Tracking Jitter with Bias Trend

Tab. 3 shows the results of complete monitoring for different multipath delays in the case of DD1 filter. In the single component multipath model, by estimating the amplitude and delay of the reflection it is possible to evaluate the error on the measurements and then estimate the current tracked signal quality. The recognition of a multipath corrupted signal associated with the amplitude and delay of the reflection may be used to select the pseudorange measurements more reliably in the system positioning, or the multipath fault may be mitigated where the number of tracked signals is not sufficient.

Relative Delay [chips]	LOS Amplitude [Normalized]	LOS Delay [chips]	MP Amplitude [Normalized]	MP Delay [chips]	Bias Error [m]
0.1	1.1733	-0.116	0.326	0.1006	2.52
0.2	1.0061	-0.111	0.4939	0.1822	1.88
0.3	0.9467	-0.112	0.5085	0.2724	1.81
0.4	1.0012	-0.106	0.5085	0.3857	1.43
0.5	1.004	-0.104	0.4878	0.4955	1.23
0.6	1.0122	-0.105	0.4863	0.6055	0.80
0.7	1.0085	-0.106	0.4739	0.6974	1.53
0.8	0.9976	-0.104	0.5072	0.7942	0.35
0.9	0.994	-0.103	0.497	0.8726	0.18
1.0	0.9742	-0.105	0.4526	0.9385	-2.78

Tab. 3 Monitoring of multipath component for several different delays

In Fig. 13 the discriminator output and behavior of the tracking jitter have been depicted. It can be seen how knowledge of the multipath parameters may be used to remove the bias present on the autocorrelation function. Fig. 14 shows the comparison between the rejection capability of a 0.2 chip spacing narrow correlator and the Kalman based tracking architecture. In such a case the error on the pseudo range measurements is within 2.5 meters and -2.5 meters.

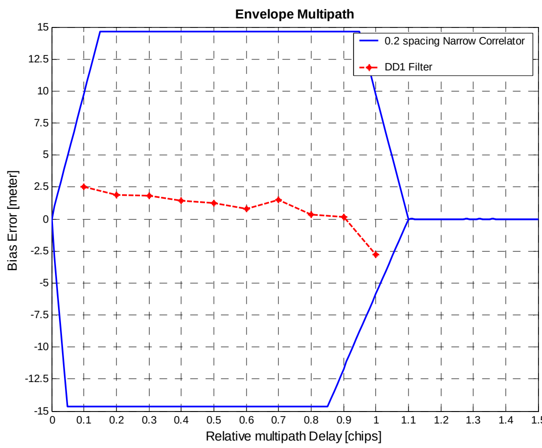


Fig. 14 Multipath Figure Envelop comparisons

7 Conclusion

A tracking scheme capable of monitoring the quality of the autocorrelation function has been analysed. Two different extensions of the traditional Kalman filter have been compared. The results reveal how accurate monitoring of a single ray multipath component can be performed by using the modified tracking architecture. Due to the prediction capability, the Kalman filter enhances the robustness of the tracking loop even in the presence of weak signals. The possibility of checking the quality of the autocorrelation function from several types of anomalies using EKF or DD1 filters, in real time, can justify the increase in system complexity of the new scheme.

ACKNOWLEDGEMENTS

This work has been performed during the primary author's visit at the UNSW. Maurizio Fantino would like to thank Alenia Spazio- Laben that is sponsoring his Ph.D grant.

References

Braasch M. (2001): *Performance Comparison of Multipath Mitigation Receiver Architectures*, IEEE Proceedings of Aerospace conference, 3, 1309-1315.

Enge P. (1999): *Local Area Augmentation of GPS for the precision approach of aircraft*, Proc. IEEE, 87, 111-132.

Hewitson S. (2003): *GNSS Receiver Autonomous Integrity Monitoring: A separability Analysis*, 16th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation, Portland, OR, 9-12 September, 1502-1509,

Hewitson S.; Hung K.; Wang J. (2004): *Localizability Analysis for GPS/Galileo Receiver Autonomous Integrity Monitoring*, The journal of Navigation, 57, 245-259.

Jee G.; Kim H.; Lee Y.; Park C. (2002): *A GPS C/A Code Tracking Loop Based on Extended Kalman Filter with Multipath Mitigation*, 15th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation, Portland, OR, 24-27 September, 446-452.

Kaplan ED. (1996): *Understanding GPS: Principles and Aplications*, Artech, Boston Massachusetts.

Leick A. (2004): *GPS Satellite Surveying* – Third Edition, John Wiley & Sons, Inc.

Macabiau C. (2000): *Impact of Evil Waveforms on GBAS Performance*, IEEE Position Location and Navigation Symposium, San Diego, 22-29.

Mitelman P.; Phelts P.; Akos D.; Pullen S.; Enge P. (2000): *A Real-Time Signal Quality Monitor For GPS Augmentation Systems*, Institute of Navigation's GPS conference, Salt Lake City, UT, 177-84.

Norgaard M.; Poulsen NK.; Ravn O. (2000): *Advances in Derivative-Free State Estimation for Nonlinear Systems*, Technical University of Denmark, Technical report: IMM-REP-1998-15.

Phelts P.; Akos D.; Enge P. (2000): *Robust Signal Quality Monitoring and Detection of Evil Waverforms*. 13th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation, Salt Lake City, Utah, 19-22 September, 1180-1190.

Psiaki ML.; Hee J. (2002): *Extended Kalman Filter Methods for Tracking Weak GPS Signals*, 15th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation, Portland, OR, 24-27 September, 2359-2553.

Ronald A. (1990): *Joint Estimation of PN Code Delay and Multipath Using the Extended Kalman Filter*, IEEE Trans. Commun, 38, 1677-1685.

Shloss P.; Phelts RE.; Walter T.; Enge P. (2002): *A Simple Method of Signal Quality Monitoring for WAAS LNAV/VNAV*, 15th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation, Portland, OR, 24-27 September, 800-809.

Wang J.; Chen Y. (1999): *Outlier Detection and Reliability Measures for Singular Adjustment Models*, Geomatic Research Australasia, No. 71, 57-72.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 160-165

Journal of Global Positioning Systems
2005
© CPGPS

A Novel Antenna Array for GPS/INS/PL Integration

Mingquan Lu[1], Jinling Wang[2], Ravindra Babu[2], Dan Li[1] and Zhenming Feng[1]

[1]Department of Electronic Engineering, Tsinghua University, Beijing, P. R. China
Tel: +86-10-62792380, Fax: +86-10-62770317, Email: lumq@tsinghua.edu.cn

[2]School of Surveying & SIS, University of New South Wales, Sydney, Australia
Tel: +61 2 9385 4203, Fax: +61 2 9313 7493, Email: jinling.wang@unsw.edu.au

Received: 16 November 2004 / Accepted: 18 October 2005

Abstract. In order to improve signal reception performance in GPS/INS/PL (Pseudolite) integration applications, a semi-sphere antenna array is proposed in this paper. It inherits the wide coverage characteristic of conventional spherical arrays and utilizes only about half the number of elements compared to a planar array to cover the upper-semi-sphere space above earth plane. It can process signals from both overhead and horizontal directions at the same time. Thus, unlike common planar arrays, this novel antenna array with a special geometry can receive satellite and Pseudolite (PL) signals from all directions, even from the horizon. Both Capon's and constraint methods have been used in the simulations of Direction of Arrival (DOA) estimation and beam-forming. These simulation results have demonstrated the advantages of the new array.

Key words: Antenna array, DOA, beamforming, semi-sphere, GPS/INS/PL.

1 Introduction

Due to its lower cost and good performance, GPS is considered as the primary source of navigation for many applications. However, GPS has its own limitations. A stand-alone GPS receiver can work well only if it can receive signals at least from 4 satellites with a reasonably good geometry. In addition, the GPS signal can also be jammed very easily by either intentional or unintentional interferences (Kaplan, 1996).

In order to get high positioning performance (e.g. accuracy and robustness), many multiple sensor integration schemes have been proposed, including GPS/GLONASS, GPS/INS, and GPS/INS/PL (Wang et al., 2001). Besides these integration methods, employing antenna arrays and adaptive filters can also substantially improve the GPS based positioning performance. An antenna array with a set of multi-channel receivers can significantly improve the position and location performance due to its anti-jamming and multipath mitigation property. As a result, the development of these spatial and temporal processing GPS receivers has become a hot topic in the research and industrial community (Amin et al 2003; Enge, 1999; Brown & Gerein, 2001).

Recently, several techniques based on antenna-array reception have been proposed (Malmström, 2003; Amin et al 2003). Among them, conventional planar antenna array is the most widely used in the researches (Brown & Gerein, 2001). However, the coverage of a general planar antenna array is limited in space. For example, these arrays cannot provide good reception performance if signals are received at low elevations. In fact, it is required to receive the satellite or Pseudolite signals at very low elevations in some environments. Obviously, employing a wide coverage antenna array with a GPS/INS/PL integrated system will be a good solution.

In order to overcome the limitations of the conventional planar antenna arrays, and improve signal reception performance in GPS, GPS/PL and GPS/INS/PL integrated applications, a semi-sphere antenna array is proposed in this paper. This array inherits the wide coverage characteristic of conventional spherical arrays and utilizes only about half the number of elements compared to a planar array to cover the upper-semi-spherical space above the earth plane. This property enhances the signal reception from both overhead and horizontal directions simultaneously. Thus, unlike common planar arrays, this novel antenna array with special geometry can receive satellite and Pseudolite signals from all directions, even from the horizon. Both Capon's and constraint methods have been used in the simulations of Direction of Arrival (DOA) estimation and beam-forming. The simulation results have demonstrated the advantages of the new array.

The paper is organized as follows. Section II presents the array geometry and derives the received signal expressions.Section III gives two simulation examples to demonstrate the advantages of the array: one for DOA estimation, and the other for beam-forming. Section IV concludes the paper.

2 ANTENNA ARRAY ILLUSTRATION

2.1 Array Geometry

The geometry of the semi-sphere array is given in Figure 1. The array is assumed to have one element on the top with M circles where N elements are uniformly placed in each circle. Therefore, a total of $MN+1$ elements are distributed on the semi-spherical surface. The coordinates of each element in the (r,θ,ϕ) polar coordinate system can be expressed as:

$$\overrightarrow{r_{mn}} = (r_{mn},\theta_{mn},\phi_{mn}) = \begin{cases} (r,\theta_{mn},\phi_{mn}), & 1\le m\le M, 1\le n\le N \\ (r,0,\#), & m=0, n=0 \end{cases} \quad (1)$$

where r represents the radius of the sphere, the symbol '#' represents any real number, and $(r,0,\#)$ is the coordinate of the element on the top of the semi-globe.

A far field source is assumed to be located at $\vec{r_s} = (r_s,\theta_s,\phi_s)$. For convenience, the centre of the bottom circle is selected as the phase reference point. Hence, every source impinging on the array will have a phase advance on each element relative to the reference point. The phase advance on $(m,n)_{th}$ element can be calculated with the following formulae:

$$\psi_{mns} = \frac{2\pi d_{mns}}{\lambda} \quad (2)$$

$$d_{mns} = r\frac{\overrightarrow{r_{mn}} \bullet \overrightarrow{r_s}}{r\cdot r_s} \quad (3)$$

where λ denotes the wavelength corresponding to the array working frequency, symbol '$\bullet$'means inner product, and $\frac{\overrightarrow{r_{mn}} \bullet \overrightarrow{r_s}}{r\cdot r_s}$ is the cosine of the angle between the two vectors. In fact, $\vec{r} = (r,\theta,\phi)$ can also be denoted by Cartesian co-ordinates (x,y,z). A mapping between these two frames can be defined as

$$x = r\sin\theta\cos\phi,\ y = r\sin\theta\sin\phi,\ z = r\cos\theta \quad (4)$$

Thus, d_{mns} in equation (3) can be recalculated as

$$\begin{aligned} d_{mns} &= r(\sin\theta_{mn}\cos\phi_{mn}\sin\theta_s\cos\phi_s) \\ &+ (\sin\theta_{mn}\sin\phi_{mn}\sin\theta_s\sin\phi_s) \\ &+ (\cos\theta_{mn}\cos\theta_s) \end{aligned} \quad (5)$$

Specifically, $d_{00s} = r\cos\theta_s$. For almost all practical applications, equation (5) is normally used to calculate the phase advance.

2.2 Received Signals

With equations in section 2.1, the received signal on $(m,n)_{th}$ element can be expressed as

$$s_{mns} = f_{mn}(\theta_s,\phi_s)a_s e^{j\psi_s} e^{j\frac{2\pi}{\lambda}d_{mns}} \quad (6)$$

where $f_{mn}(\theta_s,\phi_s)$ is the pattern of $(m,n)_{th}$ element, $a_s e^{j\psi_s}$ represents the amplitude and phase characteristics of the impinging signal, and $f_{mn}(\theta_s,\phi_s)e^{j\frac{2\pi}{\lambda}d_{mns}}$ is the element of the so called steering vector, which can be written as

$$\begin{aligned} S(\theta_s,\phi_s) = [\, & f_{00}(\theta_s,\phi_s)e^{j\frac{2\pi}{\lambda}d_{00s}},\ f_{11}(\theta_s,\phi_s)e^{j\frac{2\pi}{\lambda}d_{11s}}, \\ & \ldots,\ f_{mn}(\theta_s,\phi_s)e^{j\frac{2\pi}{\lambda}d_{mns}},\ldots,\ f_{MN}(\theta_s,\phi_s)e^{j\frac{2\pi}{\lambda}d_{MNs}}]^T \end{aligned} \quad (7)$$

Generally, D far field uncorrelated signals are assumed to come from D different directions $\{(\theta_i,\phi_i), i=1,\cdots,D\}$ with received power p_i. Noise on each element is uncorrelated, and white with zero mean and variance σ_n^2. The received signals on $(m,n)_{th}$ element is formulated as

$$x_{mn} = \sum_{i=1}^{D} s_{mni} + n_{mn} \quad (8)$$

where n_{mn} is the noise on $(m,n)_{th}$ element. The array received signal can be expressed as

$X = [x_{00}, x_{11},\cdots,x_{mn},\cdots,x_{MN}]^T$, which is a vector of length $MN+1$.

Now the covariance of the received signals at the entire array can be written as

$$R_{XX} = E[XX^H] \quad (9)$$

With some algebra manipulation, equation (9) can be reformulated as

$$R_{XX} = \sum_{i=1}^{D} p_i s_i s_i^H + \sigma_n^2 I \quad (10)$$

where $s_i = [e^{j\psi_{00i}}, e^{j\psi_{11i}}, \cdots, e^{j\psi_{mni}}, \cdots, e^{j\psi_{MNi}}]^T$. Figure 1 is the illustration of the proposed semi-sphere antenna array geometry to receive a far field signal.

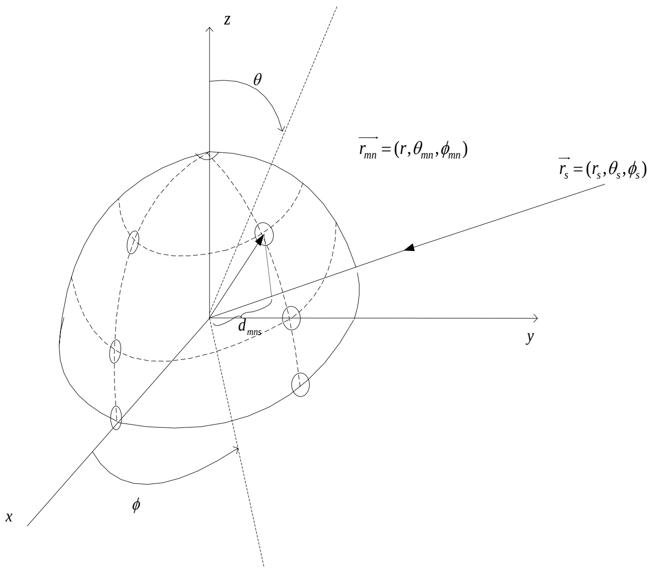


Figure 1. The proposed semi-sphere array

3 SIMULATION RESULTS

In this section, a semi-sphere array with 19 elements (circles and elements per circle) is simulated. The projection of the element circles on the plane forms a concentric circle. The radius of the sphere is the wavelength of the L1 frequency (1575.42MHz) in free space. The difference between adjacent circles is the difference between adjacent elements in the same circle.

All the elements are assumed to have some pattern in their faced directions, where denotes the angle difference between the field and element position vectors. The single element pattern is shown in Figure 2, and the element patterns of the array are illustrated in Figure 3. For simulations, all signals and interferences are BPSK modulated, and assumed to be narrowband and uncorrelated with each other. Also, the space is defined in co-ordinates with units in degrees.

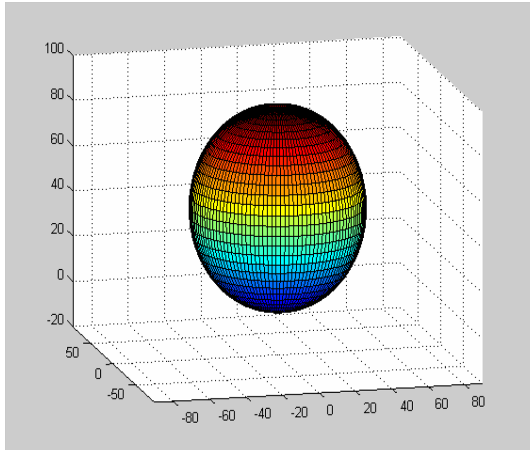


Figure 2. Single element pattern

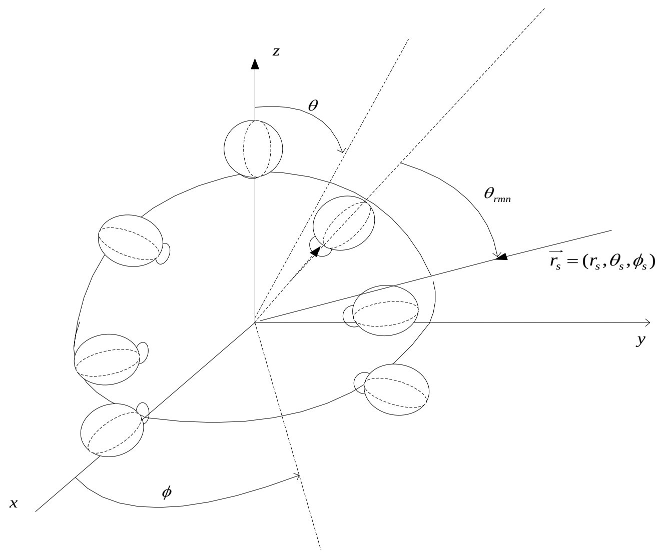


Figure 3. Element patterns of the array

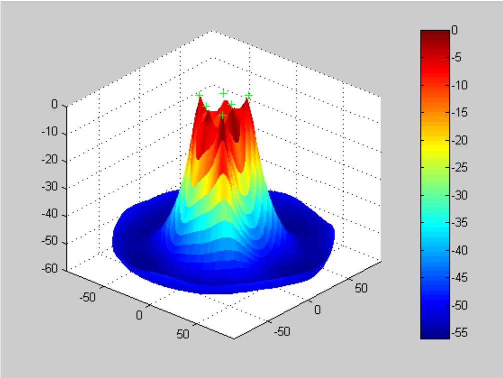
3.1 DOA Estimation

In DOA estimation, 200 snapshots were taken and Capon's method (Capon, 1969) is employed to generate the spatial spectrum. The weights for calculating the spatial spectrum is obtained by

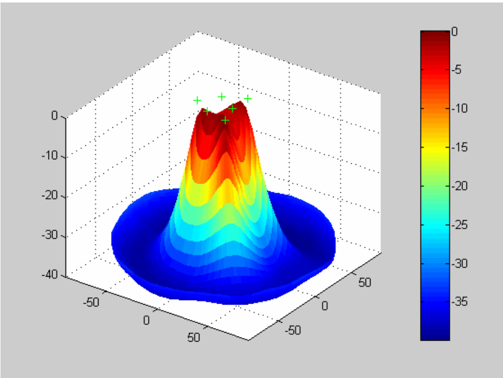
$$W_{\theta,\phi} = (R_{xx}{}^{-1} S_{\theta,\phi}) / (S_{\theta,\phi}^H R_{xx}{}^{-1} S_{\theta,\phi}) \quad (11)$$

Various simulation experiments have been carried out to demonstrate the array's utilities. In Figure 4, 6 BPSK modulated signals clustered around vertex act as GPS signals with angles (5, 10), (15, 70), (10,130), (15,190), (10,250), (15,310), and a DOA 3D spectrum is plotted. SNRs are 20dB and 10dB for (a) and (b) respectively. The peaks representing the directions of the desired signals can be easily located when SNR is 20dB, while the peaks are hard to distinguish when SNR is 10dB.

From this figure, we also observe that the array can simultaneously locate 6 signals from the overhead.



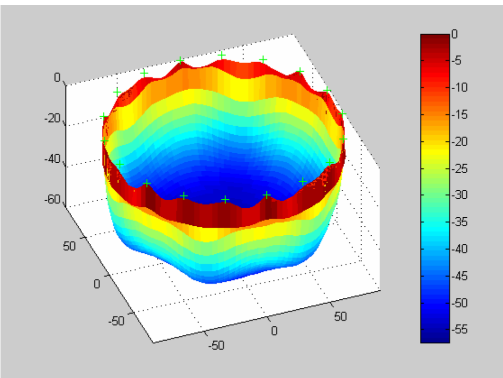
(a) SNR=INR=20dB



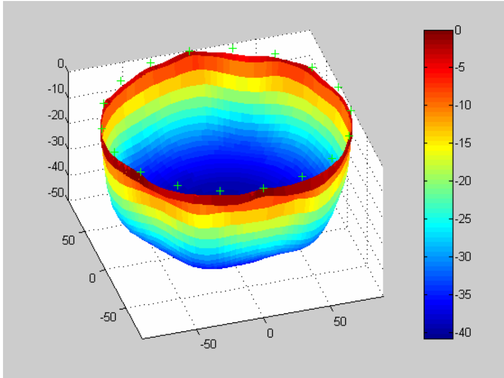
(b) SNR=INR=10dB

Figure 4. DOA spectrum for GPS signals with angles (5,10), (15,70), (10,130), (15,190), (10,250), (15,310).

In Figure 5, 18 BPSK modulated signals which simulate the Pseudolites are placed at the horizon at an angle (85,10), is uniformly distributed around the circle with 200 intervals, and a DOA 3D spectrum is plotted. In Figures 5 (a) and (b), SNR's are 20dB and 10dB respectively. The results show that 18 signals can be simultaneously located in horizon when the SNR's are high.

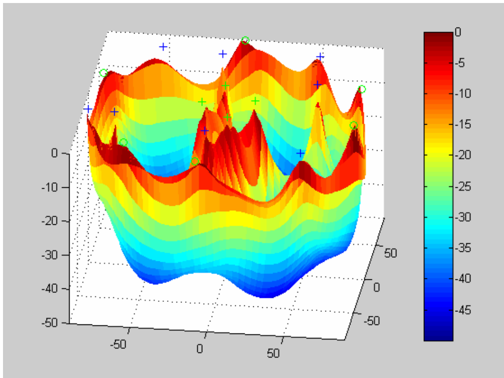


(a) SNR=INR=20dB

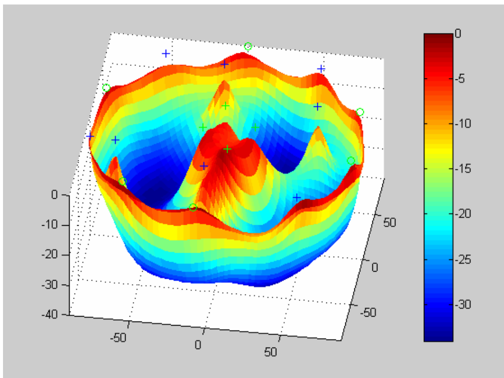


(b) SNR=INR=10dB

Figure 5. DOA spectrum for Pseudolite signals with directions (85,10), (85,30), (85,350).



(a) SNR=INR=20dB

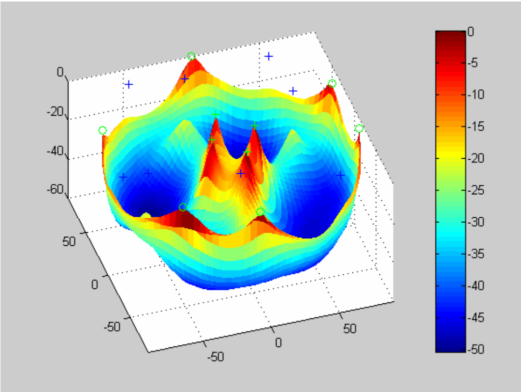


(b) SNR=INR=10dB

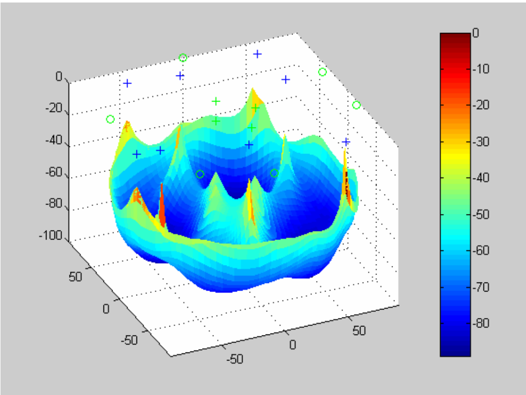
Figure 6. DOA spectrum for GPS, Pseudolite signals and interferences with angles (12, 10), (15,100), (10,190), (15,280), (82, 20), (87, 90), (84,160), (85,230), (87,265), (83,345), (30,260), (45, 30), (50,100), (55,200), (82, 55), (85,125), (87,195), (86,310).

In Figure 6, 4 BPSK modulated signals clustered around vertex act as GPS signals at angles (12,10), (15,100), (10,190), (15,280); 6 BPSK modulated signals from the horizon take the role of Pseudolite signals with angles (82,20), (87,90), (84,160), (85,230), (87,265), (83,345);8 BPSK modulated signals at angles (30,260), (45, 30), (50,100), (55,200), (82, 55), (85,125), (87,195), (86,310) are considered as interferences. DOA 3D spectrum for SNR=20dB and SNR=10dB are respectively plotted in Figures 6 (a) and (b). From this figure, it can be observed that the array can simultaneously locate 18 signals.

The effect of the INR vibration on the DOA spectrum is shown in Figure 7. DOA's are the same as those in Figure 6; INR's are 10dB and 60dB respectively in Figures 7 (a) and (b), and SNR is 20dB in both subfigures.



(a) SNR=20dB, INR=10dB



(b) SNR=20dB, INR=60dB

Figure 7. DOA spectrum for GPS, Pseudolite signals and interferences with angles (12, 10), (15,100), (10,190), (15,280), (82, 20), (87, 90), (84,160), (85,230), (87,265), (83,345), (30,260), (45, 30), (50,100), (55,200), (82, 55), (85,125), (87,195), (86,310)

3.2 Beam-forming

In beam-forming, DOA's are assumed to be known. A multiple constraint method can be employed to generate multiple beams for the desired signals and nulls at interferences by placing a unit response in the desired directions and zero response in undesired directions. Thus, the array weights can be obtained using the equation

$$W = C[C^H C]^{-1} f \tag{12}$$

where C, f are the signal steering vectors and corresponding constraints respectively; the equation satisfies $C^H W = f$. For validations, the beam patterns of symmetrical signal distributions are simulated.

In Figure 8, 6 BPSK modulated signals clustered around vertex act as GPS/Pseudolite signals with angles (15, 10), (15, 70), (15,130), (15,190), (15,250), (15,310); 6 BPSK modulated signals from the horizon take the role of Pseudolite signals with angles (85, 10), (85, 70), (85,130), (85,190), (85,250), (85,310); 6 BPSK modulated signals at angles (45, 10), (45, 70), (45,130), (45,190), (45,250), (45,310) are considered as interferences; and a DOA 3D spectrum is plotted. From this figure, it can be concluded that the array is not very effective in receiving the signals from horizon.

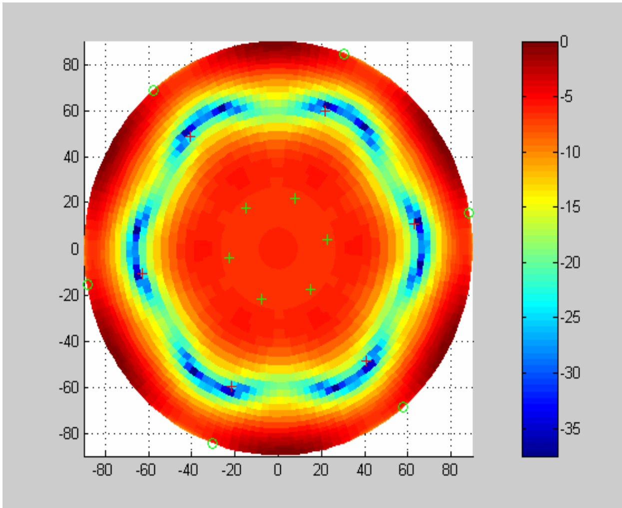


Figure 8. Beam pattern overview for GPS signals (15, 10), (15, 70), (15,130), (15,190), (15,250), (15,310). Pseudolite signals (85, 10), (85, 70), (85,130), (85,190), (85,250), (85,310) and interference (45, 10), (45, 70), (45,130), (45,190), (45,250), (45,310).

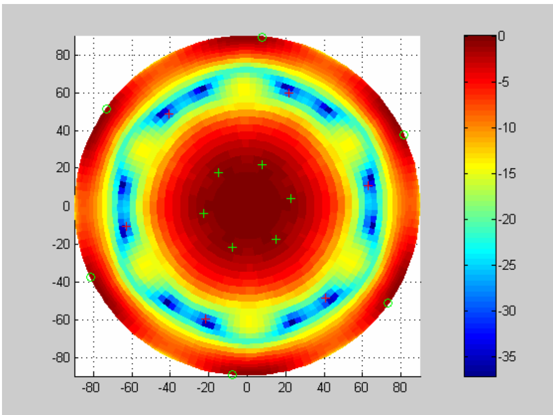


Figure 9. Beam pattern overview for GPS signals (15,10), (15,70), (15,130), (15,190), (15,250), (15,310), Pseudolite signals (85,25), (85,85), (85,145), (85,205), (85,265), (85,325) and interference (45,10), (45,70), (45,130), (45,190), (45,250), (45,310).

With GPS and interference signals direction fixed, the direction of Pseudolite signals are increased by 150, and the result is shown in Figure 9. It can be seen that both signal reception and interference rejection have been improved.

With the GPS and Pseudolite signal directions fixed, the directions of interferences are decreased to 200, and the result is plotted in Figure 10. It can be seen that the horizontal signal reception becomes weaker.

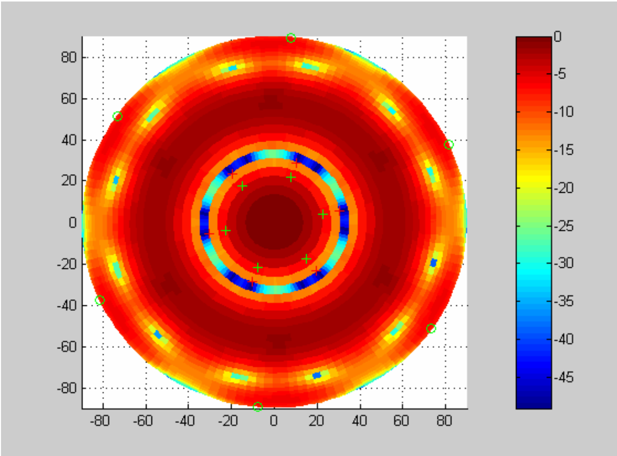


Figure 10. Beam pattern overview for GPS signals (15,10), (15, 70), (15,130), (15,190), (15,250), (15,310), Pseudolite signals (85,25), (85,85), (85,145), (85,205), (85,265), (85,325) and interference (20,10), (20,70), (20,130), (20,190), (20,250), (20,310).

4 Conclusions and outlook

In this paper, a semi-sphere array for GPS/INS/PL integration is proposed. The basic expressions are derived for this proposed new antenna array. The proposed semi-sphere antenna array can receive satellite and Pseudolite signals from all directions. Combined with adaptive processing, these semi-sphere antenna arrays based GPS, or GPS/PL, GPS/INS/PL integration receivers will possess strong anti-jamming capability. Therefore, this antenna array can be widely used in various applications to mitigate the multipath and interference signals, and also to receive low elevation Pseudolite signals. Simulation results on both the DOA estimation and beam-forming demonstrated that the array can simultaneously process multiple signals from arbitrary directions in the upper semi-sphere. The power effect and spatial distribution effect of signals to the DOA spectrum and beam pattern were also analysed. This property of the semi-sphere antenna makes it very attractive in GPS based applications, including GPS/INS/PL integration.

Acknowledgements

This research is partly supported by an ARC (Australian Research Council) – Discovery Research Project on ‘Robust Positioning Based on Ultra-tight Integration of GPS, Pseudolites and Inertial Sensors’.

References

Amin M.; Sun W.; and Lindsey A. (2003): ***Adaptive arrays for GPS receivers***, in Ibnkahla M.Ed., Signal Processing for Mobile Communications Handbook, CRC Press.

Capon J. (1969): ***High-resolution frequency-wave number spectrum analysis***, Proc. IEEE, 57:1408-1418.

Enge P. (1999): ***Local area augmentation of GPS for the precision approach of aircraft***, Proc. IEEE, 87(1):111-132.

Kaplan ED. (1996): ***Understanding GPS: Principles and Applications***, Artech House, Inc.

Wang J.; Dai L.; Tsuiji T.; Rizos C.; Grejner-Brzezinska D.; Toth C. (2001): ***GPS/INS/Pseudolite Integration: Concepts, Simulation and Testing***. 14th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation, Salt Lake City, Ohio, 11-14 September, 2708-2715.

Malmström J. (2003): ***Robust Navigation with GPS/INS and Adaptive Beamforming***, Scientific Report, FOI-R—0848—SE, Swedish Defence Research Agency, Sweden.

Brown A.; Gerein N. (2001): ***Test Results of a Digital Beamforming GPS Receiver in a Jamming Environment***, ION GPS 2001, 11-14 September 2001, Salt Lake City, UT, 894-903.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 166-175

Journal of Global Positioning Systems
2005
© CPGPS

A Comparison of Single Reference Station, Correction-Based Multiple Reference Station, and Tightly Coupled Methods using Stochastic Ionospheric Modelling

P. Alves and G. Lachapelle

Position, Location And Navigation Group, Department of Geomatics Engineering, University of Calgary, Canada
Tel: (1-403) 210 9796, Fax: (1-403) 2841980, Email: lachapelle@geomatics.ucalgary.ca

Received: 6 December 2004 / Accepted: 12 July 2005

Abstract. The multiple reference station approach to carrier phase-based positioning uses a network of GPS reference stations to model the correlated errors in a geographic region. This paper compares two methods for multiple reference station positioning under a low and a high level of ionosphere. The first method tested is the conventional method for multiple reference station positioning, which is usually a three-step process, namely (1) estimation of the carrier phase ambiguities in the network, (2) prediction of the measured network errors at the location of the rover, and (3) application of the corrections in a practical format. The second method is called the tightly coupled or in-receiver approach, which uses the data from the rover and integrates it with the network solution to better model the effect of the ionosphere. In this approach there are no explicit corrections. These two methods are compared with the single reference station approach for data from two days collected from the Southern Alberta Network in Canada, a medium scale network with inter-stations distance of 34 to 59 km.

Key words: Multiple reference station GPS RTK, VRS, single reference station GPS, least squares collocation

1 Introduction

Network RTK (Real Time Kinematic) implementation consists of three main steps (Lachapelle and Alves, 2002). In the first step, the errors at the reference stations are estimated using carrier phase observations. The second step interpolates these errors to the rover receiver location and finally the corrections are transmitted to the rover in the third step. This process is usually carried out through the generation of virtual reference station (VRS) data that the rover can accept in a single reference station data format. In this way, standard single reference station RTK software can be used at the receiver.

Real-time kinematic network positioning is limited by many factors, one of which is the communication network used between the network control centre and the rovers. Due to bandwidth limitations with multiple rovers and an attempt to allow for user privacy, network RTK positioning methods have attempted to operate a broadcast-only system (one-way communication), whereby the network corrections are broadcast to all rovers and there is no information communicated from the rover back to the network.

If alternate communication methods are used then not only can the network stations assist the rover but the rover can also assist the network with additional information. In this case the rover actually becomes part of the network and the reduced inter-receiver distances and additional ambiguity constraints provided by the rover improve the overall ambiguity resolution process very significantly using the now established receiver multiplicity; concept initially proposed by Lachapelle et al. (1993) and further tested by Luo and Lachapelle (2003). This enhanced procedure is also ideal for post-mission applications such as, verification of hydrographic surveys, airborne surveys, and land surveying.

Network RTK systems use reference stations to precisely measure the correlated errors affecting the region. These errors can only be measured when all other parameters are precisely determined, namely the station position and carrier phase ambiguities. With this in mind, the better a station's position and ambiguities are known, the more accurately one can separate measurement errors and systematic biases. Reference stations are an obvious choice because their positions are known, but any receiver can be used to estimate measurement errors. For example, a static or kinematic rover can be treated as a reference station. In terms of error modelling, multiple rovers in an area can each give an indication of the local environmental error conditions, e.g. the state of the ionosphere that is a major error source. Combining all of this information into a coherent model allows for new network rovers, with less defined position and velocity estimates, to benefit from decreased measurement errors.

The assistance of the rover to the network can be seen in the baseline configurations for the network. Ambiguity resolution performance is a function of the inter-antenna distance separation because the correlated errors increase in magnitude as the inter-antenna distance increases. In a broadcast-only Network RTK system, baselines are formed between the various reference stations. Rovers within the network will, by definition, be between two or more reference stations. Therefore connecting baselines to the rovers as well as the reference stations will shorten the overall network inter-antenna separations within the network, thus giving a higher likelihood of resolving the carrier phase ambiguities.

Instead of applying a simple weighted average (prediction) approach, the rover's data and estimated states are added to the network filter. The network filter is used solely to estimate and resolve the network ambiguities in the real-time approach. The addition of the rover's information into the network filter maintains all the information used in the correction-based approach and adds the rover's measurements. The difference is that the network not only assists the rover but the rover also assists the network.

In the loosely coupled (correction-based) approach, network ambiguities (and other nuisance parameters, such as the ionosphere) are estimated using Bayes filtering. The ambiguity estimates are then searched, and if validated, resolved. The resulting ambiguities and measurement residuals are then used to predict the errors at the rover locations.

The integrated approach does not require the error prediction phase of the loosely coupled approach because the error estimates are reduced from the rover's estimates when the rover's position is estimated. This is accomplished by the signal covariance function, which is used to evaluate the contribution of each of the reference station's observations on the rover. The covariance function is a statistical measure of the correlated errors, namely orbital, tropospheric, and ionospheric errors, between measurements. If two measurements are highly correlated, then when they are differenced, the variance of the resulting differenced observation is reduced. Consider the case where the rover is connected to every baseline in the network. The reference station observations that are highly correlated with the rover are assigned a low variance and as a result, are given more weight in the adjustment than an observation whose errors are different than those of the rover. This method of weighting produces an error model using all of the surrounding reference station data.

Pugliano et al. (2003, 2004), Alves (2004), and Alves et al. (2004), show significant improvements when using the tightly coupled approach on a variety of rover and network configurations.

2 Tightly Coupled Implementation

The design matrix of the tightly coupled (in-receiver) approach has the form

$$A = \begin{bmatrix} \frac{\partial \Delta\nabla\Phi_1}{\partial x} & \frac{\partial \Delta\nabla\Phi_1}{\partial y} & \frac{\partial \Delta\nabla\Phi_1}{\partial z} & \lambda & 0 & & 0 & 0 & \\ \frac{\partial \Delta\nabla\Phi_2}{\partial x} & \frac{\partial \Delta\nabla\Phi_2}{\partial y} & \frac{\partial \Delta\nabla\Phi_2}{\partial z} & 0 & \lambda & & 0 & 0 & \\ & & & & & \ddots & & & \\ 0 & 0 & 0 & 0 & 0 & & \lambda & 0 & \\ 0 & 0 & 0 & 0 & 0 & & 0 & \lambda & \\ & & & & & & & & \ddots \end{bmatrix} \tag{1}$$

where the first n rows correspond to the double difference observations between the rover and one of the reference stations and the second set of m rows correspond to double difference observations between the fixed reference stations with known coordinates. n is the number of double difference observations between the rover and the reference station(s) and m is the number of double difference observations between reference stations. The first three columns correspond to the rover's position estimates and the following n+m columns correspond to the ambiguities of all the double difference observations. Φ represents the carrier phase measurements in unit of length (m) and λ is the measurement wavelength in unit of length (m). The partial derivatives of the coordinates with respect to the reference stations are zero because the reference station coordinates are known and held fixed.

The design matrix can be extended to accept any number of reference stations and rovers. The processing shown includes the code and carrier phase observations processed in a single Bayes filter. This model can be expanded to incorporate any observation (system) model (estimating atmospheric errors and/or the rover's velocities, for example). The selection of the double difference observables is based on the shortest inter-receiver separations, with the conditions of linear independence and connectivity being preserved. Thus a rover may be connected to one or several reference stations, depending on the reference station-rover receiver configuration. Short distances are selected to minimize the magnitude of the differential errors. As an example, in the case of four reference stations and one rover, the double differences over the shortest four linearly independent receiver separations would be used. The rover may be involved in one to four sets of double differences.

In order to maintain the information from the correction-based approach, mathematical and stochastic information must be preserved in the integrated approach. The mathematical correlation is due to inter-receiver separations that share a common reference station (or rover) that uses the same observations in the double difference measurements. This is represented in the filter by the double difference measurement matrix, B. This matrix is not block diagonal because the observations from one station may be used in multiple baselines. The measurement matrix for the scenario where there are four stations and each station is used in a maximum of two baselines is

$$B = \begin{bmatrix} B_{sd} & -B_{sd} & 0 & 0 \\ 0 & B_{sd} & -B_{sd} & 0 \\ 0 & 0 & B_{sd} & -B_{sd} \end{bmatrix} \tag{2}$$

where B_{sd} is the single difference measurement matrix for each of the stations, assuming that each station has the same satellites in view. This correlation is often neglected in multiple baseline processing.

3 Correction-Based Method

The tightly coupled approach described above is compared herein to the typical correction-based multiple reference station approach. The correction-based model uses the network ambiguities to measure the residual errors. These residuals are interpolated to the location of the rover station using an interpolating function. In this case least-squares collocation is used to interpolate (or predict) the reference station residuals to the location of the rover. The corrections are calculated as

$$\begin{aligned} \delta\hat{\mathbf{l}} &= C_l B^T (B C_l B^T)^{-1} (B\Phi - \Delta\nabla\mathbf{N}) \\ \delta\hat{\mathbf{l}}_r &= C_{l_r,l} B^T (B C_l B^T)^{-1} (B\Phi - \Delta\nabla\mathbf{N}) \end{aligned} \tag{3}$$

where $\delta\hat{\mathbf{l}}$ is the correction vector, C is a covariance matrix, B is the observation double difference matrix, Φ is a vector of the carrier phase ambiguities minus range, and $\Delta\nabla\mathbf{N}$ is a vector containing the double difference carrier phase ambiguities in unit of length. The subscript "r" refers to the measurements of the reference station used in the single difference processing. This method was introduced by Raquet (1998).

4 Covariance Function

Stochastic correlation is defined by the signal covariance function. The covariance function states the likelihood of two values being the same based on a physical process. For example, it is known that the ionosphere is a spatially correlated error, therefore two stations close to each other are likely to have a similar ionospheric error. Stochastic correlation is represented in the Bayes filter in the variance-covariance matrix of the observations. This is a fully populated matrix because each of the measurement pairs should be somewhat correlated. The following covariance function form is used

$$\sigma_{a,b} = e^{-\frac{d}{\beta_d}} e^{-\frac{\alpha}{\beta_\alpha}} \sigma_T^2 + e^{-\frac{d_I}{\beta_{d_I}}} \sigma_I^2 \tag{4}$$

where the covariance between observations a and b ($\sigma_{a,b}$) is a function of the distance, d, between the two reference stations, the great circle angle, α, between the measurements and the ionosphere pierce point distance, d_I. σ_T^2 and σ_T^2 are the variances of the troposphere and ionosphere components respectively. The parameters of the covariance function (β_d, β_α, σ_T^2, β_{d_I}, and σ_I^2) are estimated in real-time so that they can adaptively respond to changing atmospheric conditions, as proposed by Alves (2004).

5 Ionosphere Modelling

Stochastic ionospheric modelling is used to reduce the effect of the ionosphere on the estimated position and ambiguities. This model estimates the dual frequency slant ionospheric delays using the ionosphere-free model (Odijk, 1999). In addition to the rover's position, velocity, and ambiguity states, an ionospheric parameter for each dual frequency satellite pair is estimated. The corresponding rows and columns are added to the design matrix in Equation 1.

The following description is in terms of the tightly coupled approach, however the single reference station approach uses the same methodology except that only one reference station is used. The design matrix is partitioned into sub-matrices as

$$A = \begin{bmatrix} A_{(1,1)} & A_{(1,2)} \\ A_{(2,1)} & A_{(2,2)} \end{bmatrix} \tag{5}$$

where the first row of two matrices refers to the measurements of baselines that include the rover as one of the stations. The second row refers to measurements of baselines with only network stations. As there are no common estimated parameters between the network and rover,

$$A = A_{(1,2)} = A_{(2,1)} = 0 \tag{6}$$

$$A_{(1,1)} = \begin{bmatrix} A_{pos} & \lambda_{L1} I & 0 & I \\ A_{pos} & 0 & \lambda_{L2} I & \frac{f_{L1}^2}{f_{L2}^2} I \end{bmatrix} \tag{7}$$

where the two rows of the above sub matrix represent the two measurement types used, namely L1 and L2 phase measurements. A_{pos} is a 3-column sub-matrix that includes the partial derivatives of the double difference measurements with respect to the three position components of the rover. If the velocity is estimated, then A_{pos} contains six columns where the last three columns are zero. f is the measurement frequency in Hertz and I is the identity matrix.

The sub matrix for the network observations has the same form as the sub matrix for the rover observations. The difference is that the first row is removed as shown.

$$A_{(2,2)} = \begin{bmatrix} \lambda_{L1} I & 0 & I \\ 0 & \lambda_{L2} I & \frac{f_{L1}^2}{f_{L2}^2} I \end{bmatrix} \tag{8}$$

The first column is removed because no position states are measured by the network observations. Each row of this sub matrix represents the same observation types as the rover's observations.

6 Test Methodology

The performance evaluation of the three methods, correction-based multiple reference station approach, tightly coupled approach, and the single reference station approach is broken down into two parts, namely convergence and steady state. To test the convergence performance of each of the methods, the processing filters are reset each hour. The rover position is initialized with an error of 1 m when the filters are reset. This process gives 24 convergence trials over a period of 24 hours. Each epoch during convergence is averaged across the trials. The root mean squared (RMS) position error after convergence is presented as an indication of the steady state position accuracy. A data rate of 30 seconds is used in all of the tests.

The use of stochastic ionospheric modelling is also evaluated in this study. To simulate realistic real-time operational performance, the network correction computation always implements stochastic ionospheric modelling. Although the rovers used in these tests are in static mode, they have been processed in kinematic mode, i.e. no batch solution. The positions and velocities are estimated as a first order Gauss-Markov process.

MultiRef™ and MultiRefPM™ were used for processing (e.g., Alves et al 2004). These software packages were developed at the University of Calgary. The MultiRef™ software was used to calculate the network corrections for the correction-based approach. All of the rover positions are calculated using MultiRefPM™. To ensure that the same processing methods are used for all of the compared approaches, the tightly coupled algorithm is used for processing in all cases, however only one baseline is processed in the single and correction-based scenarios.

MultiRef™ and MultiRefPM™ estimate both L1 and L2 ambiguities with the option of implementing stochastic ionosphere modelling. The best available ambiguities (be it fixed for float) are used for the final positioning solution and correction computation.

7 Southern Alberta Network

The Southern Alberta Network (SAN) was used to conduct the performance analysis. This network of dual-frequency NovAtel Modulated Precision Clock (MPC) receivers is operated and maintained by the PLAN Group of the University of Calgary. The network of 14 reference stations covers approximately 150 km north-south and 200 km east-west around Calgary. It is used in a variety of GPS research projects at the University of Calgary including Network RTK and GPS meteorology.

The sub-network shown in Figure 1 was used herein to evaluate the improvement due to the in-receiver multiple reference station approach whereby UOFC is considered the roving receiver. The baselines used for the in-receiver approach are the shortest set of independent baselines. The baseline lengths of the network range from 24 to 49 km. The closest reference station is used to evaluate the performance of the single reference station approach. This creates a single reference station inter-antenna distance of approximately 24 km.

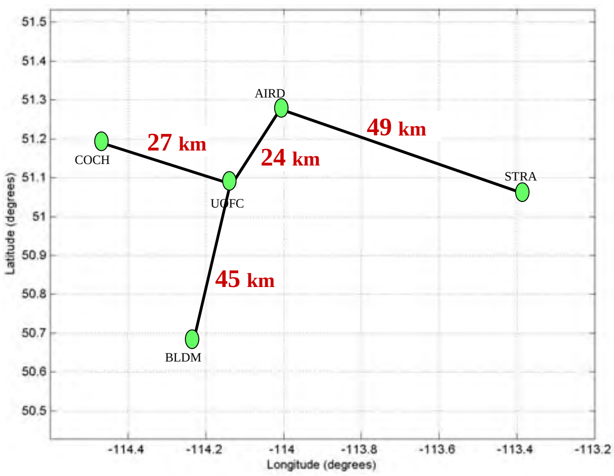


Fig. 1 Network configuration for the in-receiver multiple reference station approach with UOFC as the rover

The network configuration shown in Figure 2 is used for the correction-based multiple reference station approach. None of the network baselines connect to the rover in this approach, which makes the baseline lengths slightly longer. The baseline lengths range from 34 to 59 km for the correction-based approach.

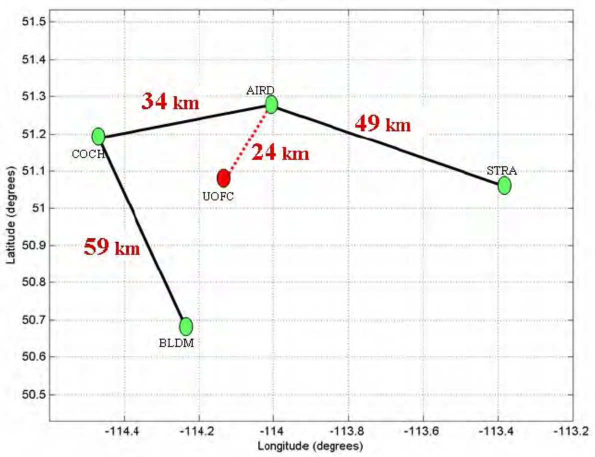


Fig. 2 Network configuration for the correction-based multiple reference station approach with UOFC as the rover

Two days of data were used for the performance evaluation, namely May 24 and April 6, 2004. These were selected because they represent a relatively low and high level of ionospheric error respectively. Figure 3 shows the estimated double difference slant ionospheric effect for a 60 km baseline for the two days. The ionosphere is more variable and higher in magnitude for April 6 (bottom) then for May 24 (top). On April 6, the double-difference ionospheric effect reaches about 5 ppm.

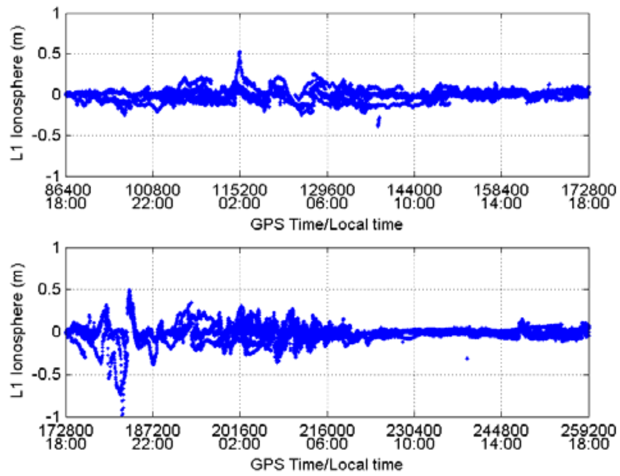


Fig. 3 Estimated double difference slant ionosphere for May 24 (top) and April 6 (bottom) for a 60 km baseline

8 May 24 (Low Ionosphere)

Figure 4 shows the convergence performance for May 24 without stochastic ionospheric modelling. The tightly coupled method converges faster than the other methods, followed by the correction-based approach. The single reference station approach converges very slowly relative to the multiple reference station methods. The individual convergence periods, representative of the 24 periods used to derive Figure 4, are shown in Figure A1 of the Appendix.

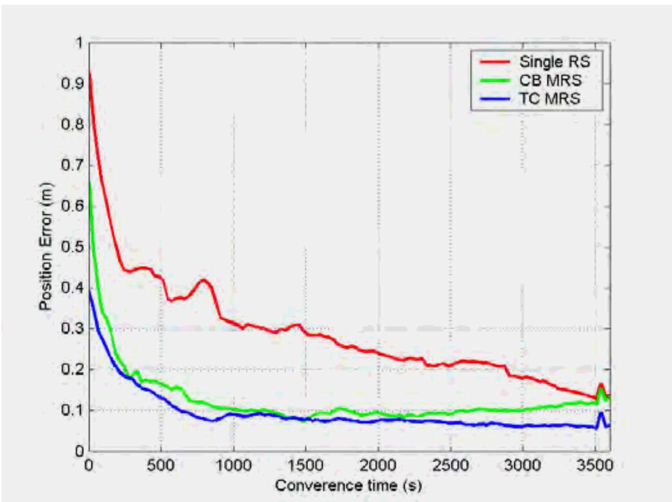


Fig. 4 Convergence of the single reference station, correction-based multiple reference station and tightly coupled approaches for May 24 (low ionosphere) without stochastic ionosphere modelling

Table 1 shows the RMS position errors for the three methods *after* convergence. The multiple reference station methods perform much better than the single reference station approach. This difference is due to network modelling of the ionosphere in the correction-based approach and ionospheric error averaging in the tightly coupled approach. The results improve drastically when stochastic ionospheric modelling is used.

Table 1 RMS position errors for the single reference station, correction-based multiple reference station, and tightly coupled approaches, *after* convergence for May 24 (low ionosphere) without stochastic ionospheric modelling

	Single (cm)	CB MRS (cm)	CB MRS Improvement	TC MRS (cm)	TC MRS Improvement
North	4.5	2.0	56 %	1.8	60 %
East	5.2	1.2	77 %	1.1	79 %
Height	9.3	5.3	43 %	5.5	41 %
3D	11.5	5.8	50 %	5.9	49 %

Figure 5 shows the convergence performance for May 24 with stochastic ionospheric modelling. The individual convergence periods, representative of the 24 periods used to derive Figure 5, are shown in Figure A2 of the Appendix. The convergence in general is very good. The tightly coupled method converges faster than the other methods, followed by the correction-base approach. The tightly coupled approach convergences in approximately 1200 seconds, while the correction-based and single reference station approaches converge after 2000 and 3400 seconds, respectively.

The single reference station and tightly coupled methods perform better with ionospheric modelling. The correction-based approach performs very similarly in both cases. This is because the stochastic ionospheric model decreases the degree of freedom of the position estimation adjustment. For example, the single epoch degree of freedom of a single baseline without stochastic ionospheric modelling is (s – 1) * 2 – (6 + (s – 1) * 2) where s is the number of satellites. The first term {(s – 1) * 2} is for the L1 and L2 carrier phase observations together. The number 6 refers to the 3 position and 3 velocity states and (s – 1) * 2 is for the L1 and L2 ambiguities for each satellite pair. The degree of freedom changes to (s – 1) * 2 – (6 + (s – 1) * 3) when the ionosphere is estimated. The degree of freedom is negative which shows that more than one epoch are required to observe all of the estimated parameters. The increase in model noise due to the decrease in the degree of freedom is usually less than the magnitude of the ionosphere biases that are estimated and removed. The network corrections are effective in removing the ionospheric errors from the measurements, which increases the model noise without decreasing the measurement biases.

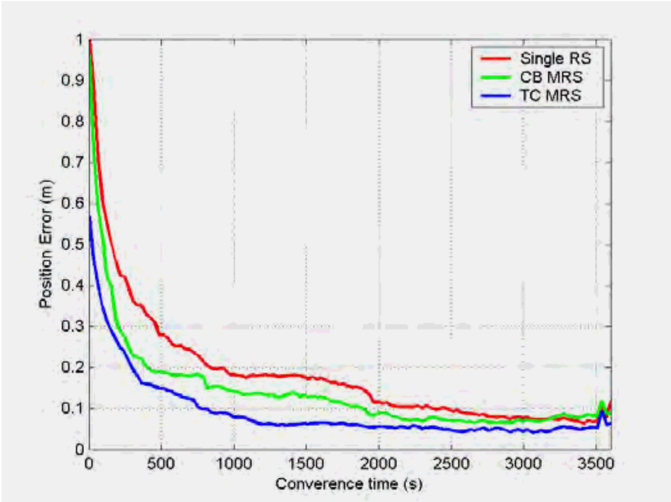


Fig. 5 Convergence of the single reference station, correction-based multiple reference station, and tightly coupled approaches for May 24 (low ionosphere) with stochastic ionosphere modelling.

Table 2 shows the RMS position error for the three methods after convergence. The difference in performance with a low level of ionosphere and stochastic ionospheric modelling is negligible. This is due to the effectiveness of ionospheric modelling.

Table 2 RMS position errors for the single reference station, correction-based multiple reference station, and tightly coupled approaches *after* convergence for May 24 (low ionosphere) with stochastic ionosphere modelling.

	Single (cm)	CB MRS (cm)	CB MRS Improvement	TC MRS (cm)	TC MRS Improvement
North	1.4	1.3	7 %	1.2	14 %
East	0.8	0.8	0 %	0.8	0 %
Height	2.7	3.0	-11 %	3.1	-14 %
3D	3.2	3.3	-3 %	3.4	-6 %

A comparison of the results from Tables 1 and 2 reveals that the stochastic ionospheric modelling decreases the position RMS by slightly less than half for the multiple reference station methods and more than three times for the single reference station approach. This further shows the effectiveness and importance of ionospheric modelling, even under fairly benign ionospheric conditions.

9 April 6 (High Ionosphere)

Figure 6 shows the convergence of the three methods for April 6, which has a high level of ionospheric effect. No ionosphere modelling is applied. The convergence is slow in all cases although the multiple reference station methods perform much better than the single reference station approach. The convergence periods used to derive Figure 4, are shown in Figure A3 of the Appendix. Figure A3 shows that the convergence is representative of the ionosphere levels throughout the day.

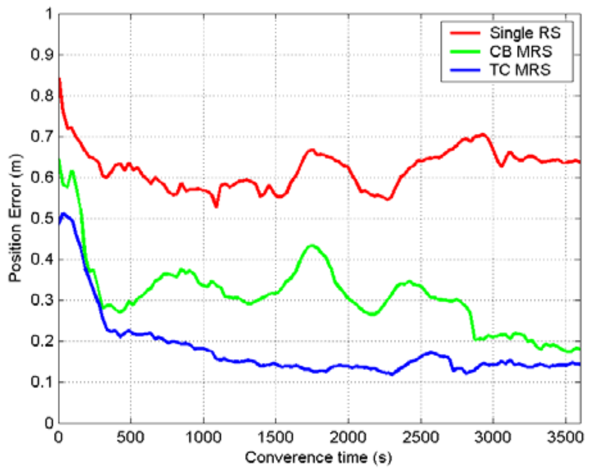


Fig. 6 Convergence of the single reference station, correction-based multiple reference station, and tightly coupled approaches for Apr. 6 (high ionosphere) without stochastic ionosphere modelling.

Figure 7 shows the 3D position errors for the three methods over the entire 24-hour period. The accuracies of the methods are correlated to the magnitude of the differential ionospheric effect shown in Figure 3. The error at the beginning of the data set is up to 2 m for the single reference station approach. This is due to the large variability of the ionosphere shown in Figure 3. This decreases to less than 10 cm when the ionospheric effect is lower near the end of the day. The correction-based approach performs best during this test, which shows that the corrections are effective in reducing the effect of the ionosphere. The tightly coupled approach performs better than the single reference station approach but is still affected by the ionosphere.

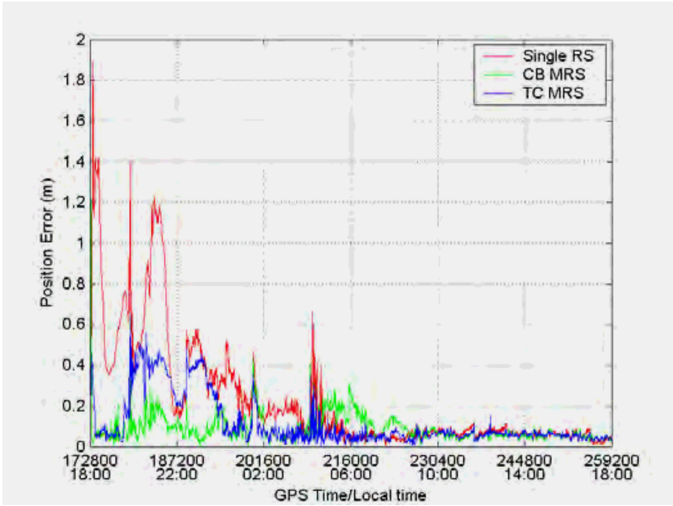


Fig. 7 3D position error over time for the single reference station, correction-based multiple reference station and tightly coupled approaches, for Apr. 6 (high ionosphere) without stochastic ionosphere modelling.

The performance is improved when stochastic ionospheric modelling is applied. Figure 8 shows the convergence of the methods in this case. The individual convergence periods shown in Figure A3 when applying the ionospheric modelling are shown in Figure A4. For this data set the correction-based approach performs slightly better than the tightly coupled approach, which performs significantly better than the single reference station approach. Comparing Figures 6 and 8 shows the improvement due to stochastic ionospheric modelling. Without stochastic ionospheric modelling, none of the methods can achieve an accuracy better than 10 cm mean position error within the one hour convergence test. When stochastic ionospheric modelling is used, all the methods converge to an accuracy better than 10 cm in less than 2500 seconds.

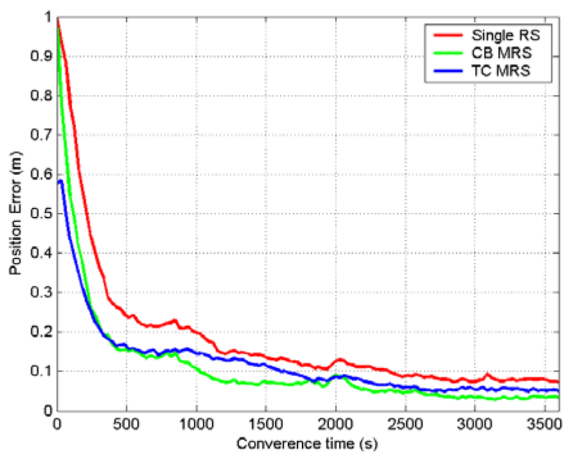


Fig. 8 Convergence of the single reference station, correction-based multiple reference station, and tightly coupled approaches, for Apr. 6 (high ionosphere) with stochastic ionosphere modelling.

Figure 9 shows the 3D position errors over time with stochastic ionospheric modelling. A comparison with Figure 7 shows a major improvement. Table 3 shows the RMS position errors for the methods after convergence. The position accuracy for all of the methods is exceptional (better than 4 cm 3D position RMS). Although the multiple reference station methods provide a noticeable reduction of the convergence time there is little if any improvement in the converged position accuracy for this data set.

A comparison of the results of Table 2 and 3 reveals the same level of accuracy for both a low and high level of ionospheric effect when stochastic ionosphere modelling is applied. This somewhat surprising result further confirms the effectiveness of the ionospheric modelling approach used in the software.

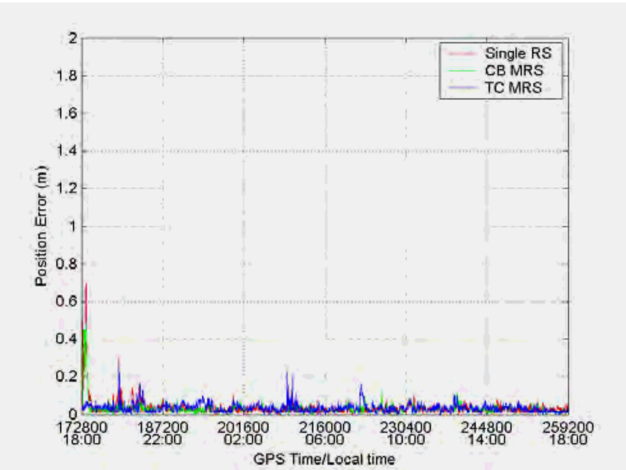


Fig. 9 3D position error over time for the single reference station, correction-based multiple reference station and tightly coupled approaches, for Apr. 6 (high ionosphere) with stochastic ionosphere modelling

	Single (cm)	CB MRS (cm)	CB MRS Improvement	TC MRS (cm)	TC MRS Improvement
North	1.6	1.3	19 %	1.3	19 %
East	0.8	0.8	0 %	0.9	-13 %
Height	3.0	2.9	3 %	3.7	-23 %
3D	3.5	3.3	6 %	4.0	-14 %

Table 3 RMS position errors for the single reference station, correction-based multiple reference station, and tightly coupled approaches *after* convergence, for Apr 6 (high ionosphere) with stochastic ionosphere modelling

10 Conclusions

This paper compares three different carrier phase based DGPS RTK methods, namely the traditional single reference station approach, a collocation-based correction-based multiple reference station approach, and a tightly coupled multiple reference station approach. Two days of data from a medium scale network are used to assess the methods. These days represent a relatively high level and low level of ionosphere activity.

The multiple reference station approaches are effective in reducing convergence time by more than half in some cases. The tightly coupled approach converged slightly faster than the correction-based approach although they both performed well. When using stochastic ionospheric modelling, there was little position accuracy difference between the approaches after convergence. However without stochastic ionospheric modelling, the multiple reference station approaches perform significantly better than the single reference station approach, with a 3D position error reduction of nearly 50 percent.

All the results show a high level of improvement due to stochastic ionospheric modelling in terms of both convergence time and position accuracy during and after convergence. In most of the tests, the use of stochastic ionospheric modelling reduced the position errors by half and sometimes as much at three times. Even with stochastic ionospheric modelling, the multiple reference station approaches provide significant improvements in terms of convergence time relative to the single reference station approach.

References

Alves P (2004) ***Development of two novel carrier phase-based methods for multiple Reference station positioning.*** PhD Thesis, published as UCGE Report No. 20203, Department of Geomatics Engineering, The University of Calgary.

Alves P, Lachapelle G, and Cannon ME (2004) ***In-Receiver Multiple Reference Station RTK Solution***, Proceedings of the Institute of Navigation Saetllite Division Technical Meeting, GPS 04 (September 2004, Long Beach, USA).

Dao D, Alves P and Lachapelle G (2004) ***Performance Evaluation of Multiple Reference Station GPS RTK for a Medium Scaled Network.*** Proceedings of the International Symposium on GPS/GNSS, Sydney, Australia (6-8 December).

Lachapelle G, and Alves P (2002) ***Multiple Reference Station Approach: Overview and Current Research***, Invited Contribution, Expert Forum. Journal of Global Positioning Systems, 1(2), 133 – 136.

Lachapelle G, Liu C, and Lu G (1993) ***Quadruple Single Frequency Receiver System for Ambiguity Resolution on the Fly***. Proceedings of the International Technical Meeting of the Satellite Division of the Institute of Navigation, ION GPS 03 (September 1993, Salt Lake City, USA), 1167 – 1172.

Luo N, and Lachapelle G (2003) ***Relative Positioning of Multiple Moving Platforms using GPS.*** IEEE

Transactions on Aerospace and Electronic Systems, 39(3), 936 – 948.

Odijk D (1999) ***Sochastic Modelling of the Ionosphere for Fast GPS Ambiguity Resolution.*** Proceedings of General Assembly of the International Association of Geodesy (July 1999, Birmingham, England), volume 121, 387 – 392. International Association of Geodesy, Springer-Verlag.

Odijk D (2000) ***Weighting Ionospheric Correction to Improve Fast GPS Positioning Over Medium Distances.*** Proceedings of the National Technical Meeting of the Satellite Division of the Institute of Navigation, ION GPS 2000 (September 2000, Salt Lake, USA), 1113 – 1124.

Pugliano G, Alves P, Cannon ME, and Lachapelle G (2003) ***Performance Analysis of a Post-Mission Multi-Reference RTK DGPS Positioning Approach.*** Proceedings of the International Association of Institutes of Navigation World Congress (October 2003, Berlin, Germany)

Puliano G, Obrizza F, Pingue F, Sepe V, Alves P, and Lachapelle G (2004) ***Monitoring the Neapolitan Volcanic Area Using an Advanced Multiple Reference Station RTK DGPS Technique***, Proceedings of the Institute of Navigation Satellite Division Technical Meeting, GPS 04 (September 2004, Long Beach, USA)

Raquet J (1998) ***Development of a Method for Kinematic GPS Carrier-Phase Ambiguity Resolution Using Multiple Reference Receivers***. PhD Thesis, published as UCGE Report No. 20116, Department of Geomatics Engineering, The University of Calgary.

APPENDIX

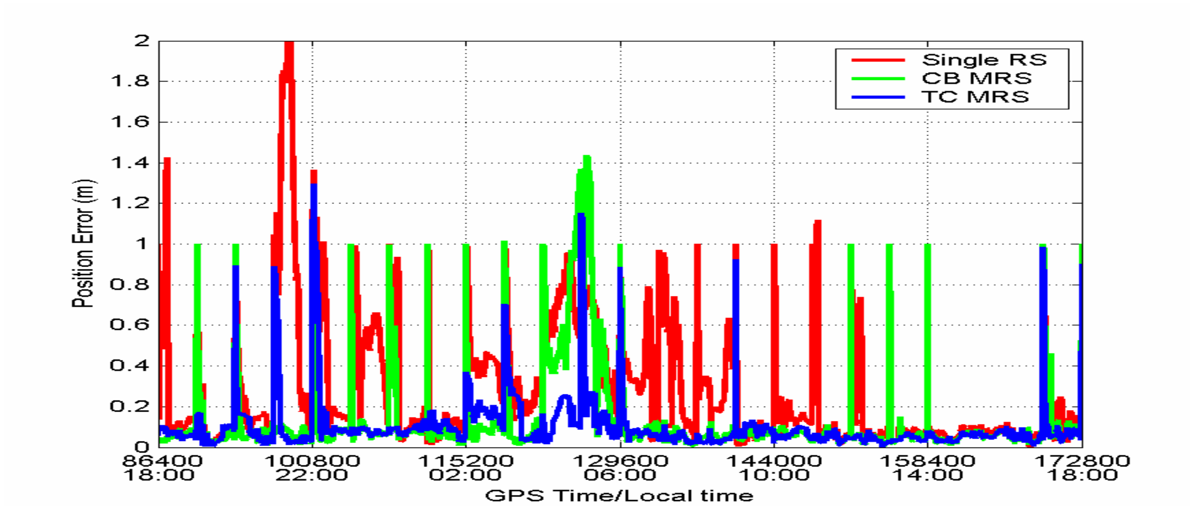


Fig. A1 All convergence periods of the single reference station, correction-based multiple reference station and tightly coupled approaches for May 24 (low ionosphere) without stochastic ionosphere modelling.

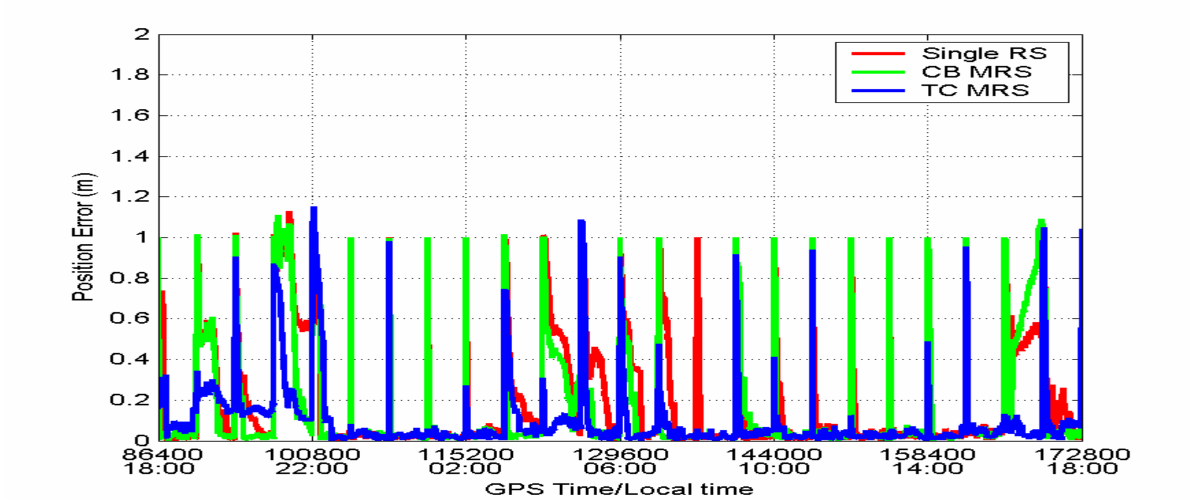


Fig. A2 All convergence periods of the single reference station, correction-based multiple reference station and tightly coupled approaches for May 24 (low ionosphere) with stochastic ionosphere modelling.

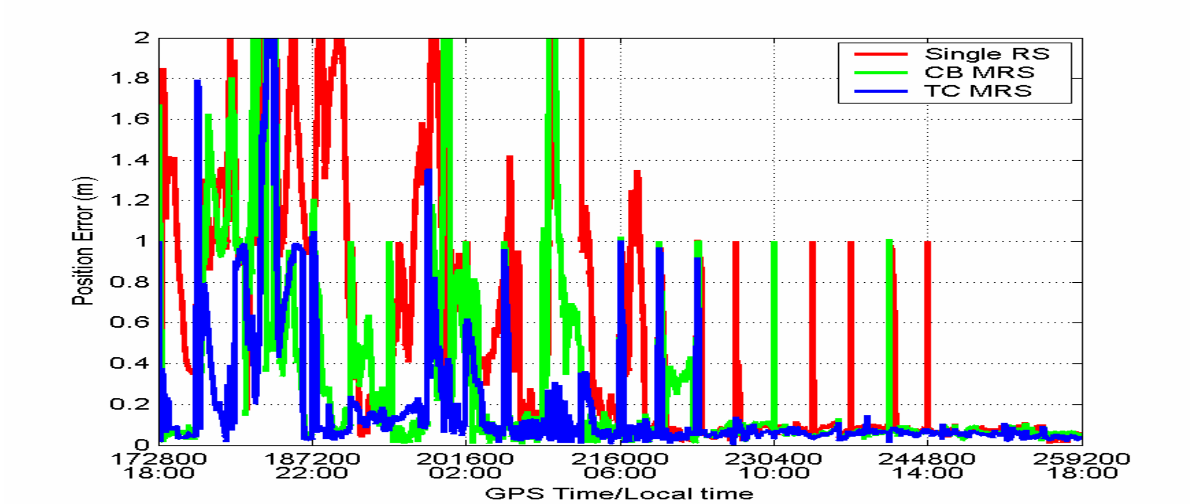


Fig. A3 All convergence periods of the single reference station, correction-based multiple reference station and tightly coupled approaches for April 6 (high ionosphere) without stochastic ionosphere modelling.

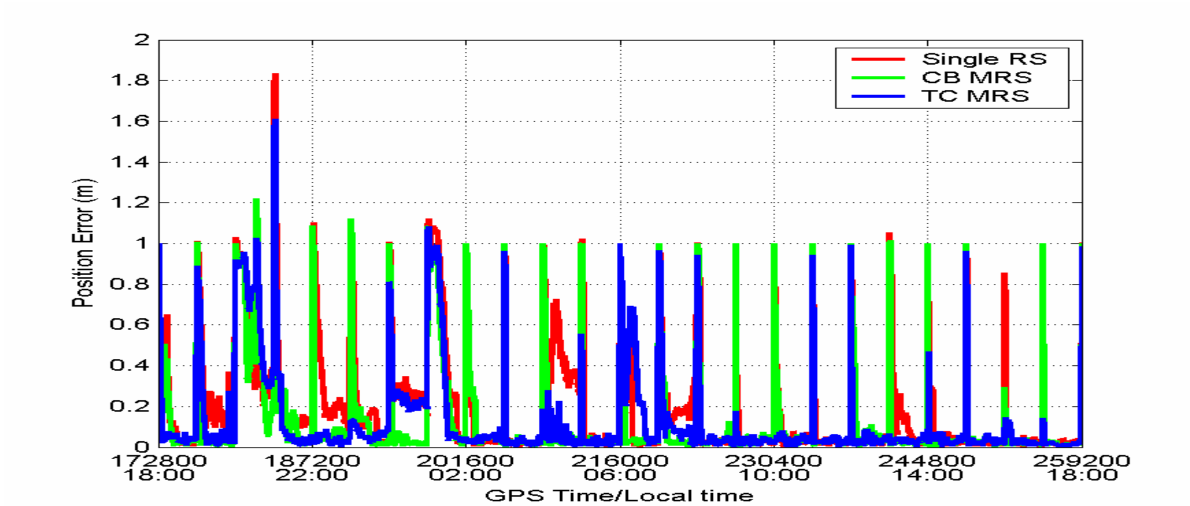


Fig. A4 All convergence periods (same as Figure A2) of the single reference station, correction-based multiple reference station and tightly coupled approaches for April 6 (high ionosphere) with stochastic ionosphere modelling.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 176-183

Galileo Receiver Core Technologies

Pavel Kovář, František Vejražka, Libor Seidl, Petr Kačmařík

Czech Technical University in Prague, Faculty of Electrical Engineering, Technická 2, 166 27 Prague 6, Czech Republic
Tel: +420 2 2435 2244 Fax: Email: kovar@fel.cvut.cz

Received: 30 November 2004 / Accepted: 12 July 2005

Abstract. The modern satellite navigation system Galileo is developed by European Union. Galileo is a completely civil system that offers various levels of services especially for civil users including service with safety guarantee. Galileo system employs modern signal structure and modern BOC (Binary Offset Carrier) modulation. The Galileo Receiver is investigated in the frame of the GARDA project solved by consortium under leadership of Alenia Spacio – LABEN. The aim of Galileo Receiver Core Technologies subtask is to investigate the key problems of the Galileo receiver development. The Galileo code and carrier tracking subtask of the Galileo Receiver Core Technologies is carried out at the Czech Technical University. The problem was analysed and split to the particular tasks. The aim of this paper is focused on BOC correlator architecture. The correlation function of the BOC modulation is more complex with a plenty of correlation peaks. The delay discriminator characteristic of such signal has several stable nodes, which cause stability problem. The standard solutions of this problem like BOC non-coherent processing, very early – very late correlator and deconvolution correlator are analysed. The new correlator architecture for BOC modulation processing has been developed. The developed correlator has two outputs, one for fine tracking and the second one for correct node detection. The second output is based on comparison of the correlation function envelopes. The simplified method of correlation function envelope calculation is described in this paper. The correlator is planned to be tested in the GRANADA software simulator including a sophisticated method of correlator output combination.

Key words: Galileo, Galileo core technologies, Galileo receiver, code tracking, carrier tracking.

1. INTRODUCTION

1.1 Galileo

The European GNSS system Galileo (that is currently under development) operates on the same ranging principle as the existing GPS and GLONASS systems do. The big benefit of this system is that it is a completely civil system, which offers to the user various types of services, which are adjusted to the civil user requirements. Besides Open Service, which is free of charge, the system offers services with guarantee of the service performance by the system provider, customer driven local element services and Public Regulated Service for governmental needs.

Galileo shares the same basic operating principle with the GPS, but the system architecture and service model are based on the latest knowledge.

1.2 GARDA project

The basic architecture of the Galileo user receiver is similar to the GPS one, yet some approaches to the receiver design are more complex. Galileo receiver development is investigated within the GARDA (GAlileo Receiver Development Activity) project, performed by a consortium established under the leadership of Alenia Spacio – LABEN. GARDA is funded by the GJU (Galileo Joint Undertaking) in the frame of the Galileo R&D activities under the EC 6th Framework Program. The project consists of three tasks, which cover Galileo user receiver development including development plan consolidation, software Galileo receiver development, receiver prototyping, and last, but not least, core technology task.

1.3 GRANADA

GRANADA (Galileo Receiver ANalysis And Design Application) is the software simulator of the Galileo developed in the frame of GARDA project by Deimos Space company. Software simulator consists of Bit-True GNSS SW Receiver Simulator and GNSS Environment and Navigation Simulator.

Mono-channel Bit-True GNSS SW Receiver Simulator serves for detail analyses of the Galileo signal processing, signal propagation, multipath propagation, interference, and other related problems. Bit-true simulator is based on detail modelling of the signal processing inside the Galileo receiver.

On the other hand, the Environment and Navigation Simulator is determined for analyses of the position determination algorithms, satellites constellation etc. The macro model of the receiver behaviour, propagation channel, noise, etc. are employed in this simulator.

The only basic most common features and algorithms of the Galileo receiver are implemented to the simulator. Some marginal problems of the Galileo are simplified or not implemented.

1.4 Galileo Core Technologies

The aim of the core technologies subtask is to investigate the critical principles and technologies of the Galileo system. The technologies are to be tested with the GRANADA software simulator. The other goal of the core technology task is to implement the new features to the GRANADA software.

The Galileo receiver core technology task was launched in January, 2005, thus the current state of the task is the preliminary phase and the problem is being analysed. The analysis and preliminary experiments results of the Galileo receiver core technology are concentrated in this paper.

Two main Galileo core technologies have been assigned to the Czech Technical University:

- Galileo code tracking,
- Galileo phase tracking.

The present simulation results with GRANADA tool have mainly verification purpose. The fundamental problems like performance of tracking loops in presence of additive white Gaussian noise were analysed. The performance parameter (variance of tracking error in this case) was compared with theoretical assumptions with good agreement. This simulation also showed some weakens and inconveniences of GRANADA mainly belong to impossibility to perform a multi frequency signal tracking.

2. PROBLEM ANALYSIS

The code and carrier tracking are very complex problems, which very closely relate to each other. The main function of the Galileo receiver is an estimation of the code delay and the carrier phase of the receiving signal. The estimation is usually realized by use of correlation reception principle, where the replica of the Galileo signal is synchronized with the received signal. The feedback tracking circuits are commonly used. The tracking loops can be classified to the following main categories:

1. Single frequency scalar tracking loops – individual tracking loops for each satellite signal
2. Multicarrier scalar tracking loops – complex tracking loops for all signals of individual satellite
3. Multicarrier vector tracking loop (VDLL) – one complex tracking loop for all signal components of all satellites

The other classification criterion of the signal tracking methods is according to interaction of the code and carrier tracking:

1. Independent code tracking and carrier tracking
2. Independent carrier tracking and code tracking with carrier aided
3. Integrated (joined) code and carrier tracking

The last classification approach to the code and phase tracking is according to the design principle of the loop low pass feedback filter:

a) Deterministic approach (classical control filter),

b) Stochastic approach (Wiener or Kalman filter).

The problem can be analyzed according to many other criteria. Basically code and carrier tracking is very similar to the GNSS signal tracking, but several problems arise in consequence with higher Galileo signal complexity. This problem has been identified and some of them will be solved in the frame of core technology project. The identified particular problems are listed below:

1. BOC and AltBOC discriminator
    a. Delay discriminator
    b. Phase/Frequency discriminator
    c. Detection of the correct peak of the correlation function

   d. Sensitivity of the discriminator to multipath
2. Cycle slip detection technique
3. Ambiguity resolution
4. Tracking strategy
   a. Independent code tracking and carrier tracking
   b. Independent carrier tracking and code tracking with carrier aided
   c. Integrated (joined) code and currier tracking
5. Tracking loops
   a. Tracking loop development method
   b. Dynamic performances of the tracking loops
   c. Loop stability
6. Tracking strategy in environment with shadowing

In this early phase of Galileo development, the research is focused on the basic solution of most critical problems. The one of the key problem of the Galileo receiver is the processing of the ranging signal with BOC (Binary Offset Carrier) modulation. This problem is analyzed in the rest of this paper.

3. STANDARD GNSS CORRELATOR

The essential navigation receiver block for an estimation of the pseudorange is called correlator. The standard GNSS correlator is designed for BPSK modulated ranging signal. The adoption of the standard GNSS correlator for BOC modulated ranging signals is discussed in this paragraph.

The architecture of adopted delay correlator is very similar to the BPSK one, see Figure 1. The ranging code $c\left(\lfloor Nf_0 t \rfloor\right)$ and digital carrier $\operatorname{sgn}\left(\sin\left(2\pi M f_0 t\right)\right)$ can be multiplied and the resulting code $c_{M,N}(t)$ can be used for the despreading of the received signal.

$$c_{M,N}(t) = c\left(\lfloor Nf_0 t \rfloor\right) \cdot \operatorname{sgn}\left(\sin\left(2\pi M f_0 t\right)\right) \tag{1}$$

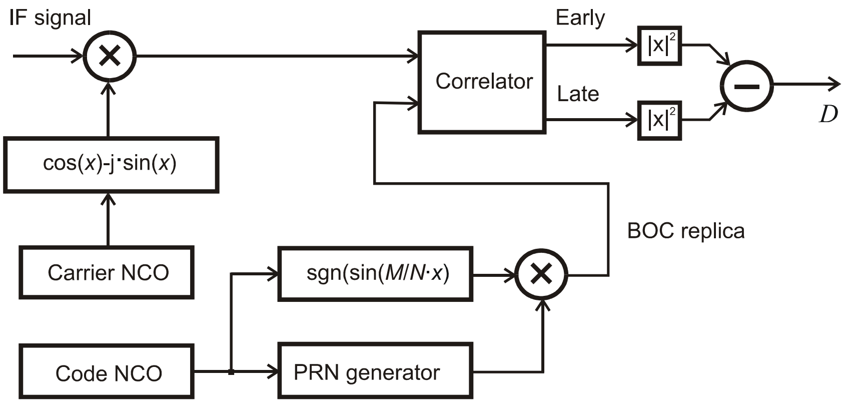


Figure 1. BOC correlator

The BOC delay discriminator characteristic of Early minus Late amplitude discriminator and Early minus Late power discriminator for BOC(1,1) modulation are displayed on the Figure 2.

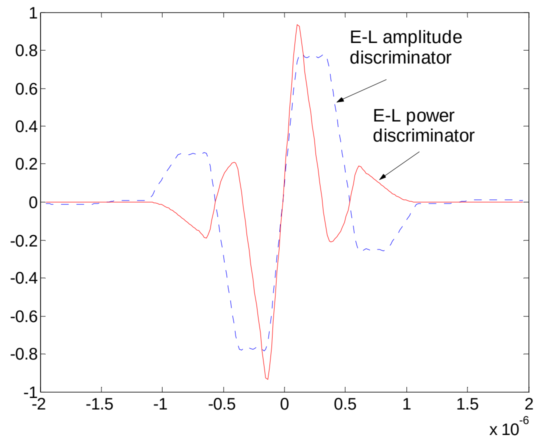


Figure 2. BOC(1,1) delay discriminator characteristic

The problem of the BOC correlator is in existence of more than one stable node on the discriminator characteristic, see Figure 3. The problem with multiple stable nodes is even more complicated for higher order BOC modulation, where a plenty of these nodes occur.

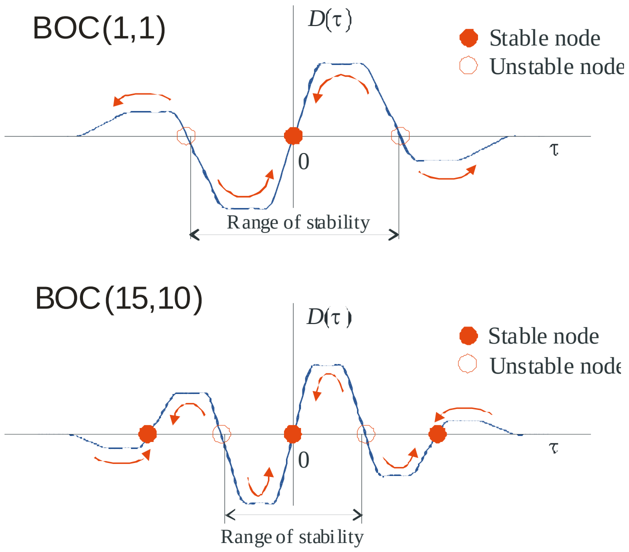


Figure 3. Stable and Unstable nodes of the BOC discriminator characteristics

The number of false stable nodes in coherent delay discriminator characteristic for modulation BOC(N, M) is given by

$$S = 2\left\lfloor \frac{2N-1}{2M} \right\rfloor. \tag{2}$$

This problem causes significant reduction of the range of the delay lock loop (DLL) stability. The DLL can potentially track false stable node without any indication.

Discussed problem is demonstrated by the following simulation, see Figure 4. The several experiments of the DLL hang-up stage are displayed on this figure. The initial delay error of each experiment is set to zero value. The DLL mostly tracks the correct node. Some of the experiments diverge to the false node or totally diverge due to the noise in loop.

False node tracking of BOC modulated signal is a very serious problem, which must be solved.

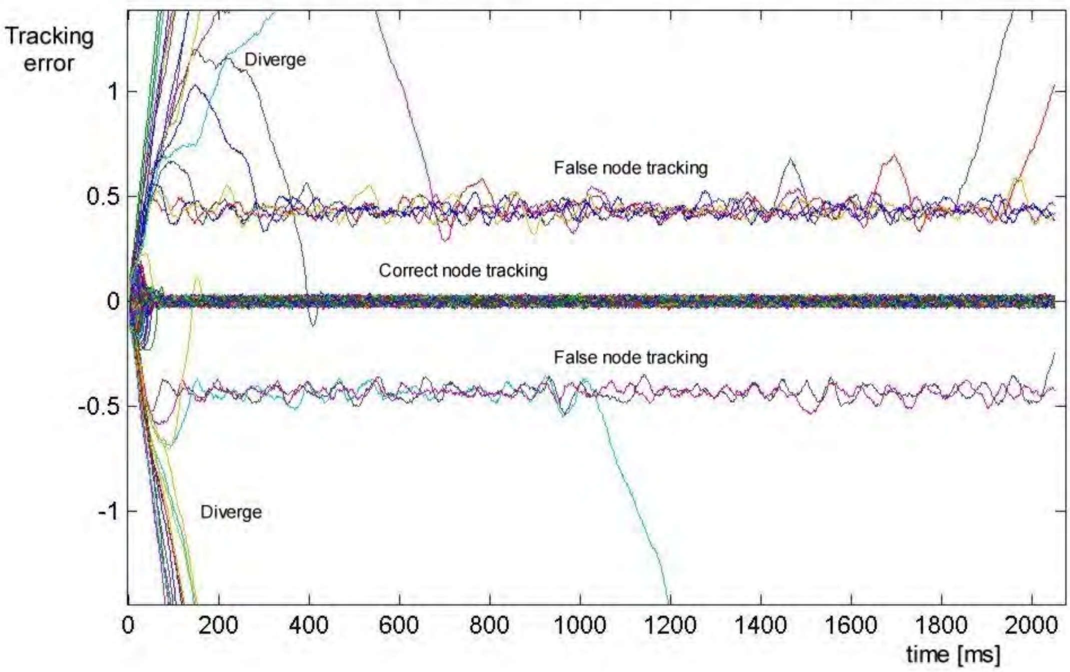


Figure 4. Simulation results of the tracking errors of BOC(1,1) signal by Early minus Late power correlator

4. EXISTING BOC CORRELATORS

4.1 Non-coherent BOC processing

Since the both sidebands of BOC modulation contain the same information the particular sideband can be processed separately and result can be non-coherently combined, see Figure 5. Of course, this method is non optimal and does not use BOC modulation benefits. On the other hand, the particular sidebands can be easily processed in classical BPSK manner. The separate sideband processing can also be useful when one of the two sidebands is corrupted with interference.

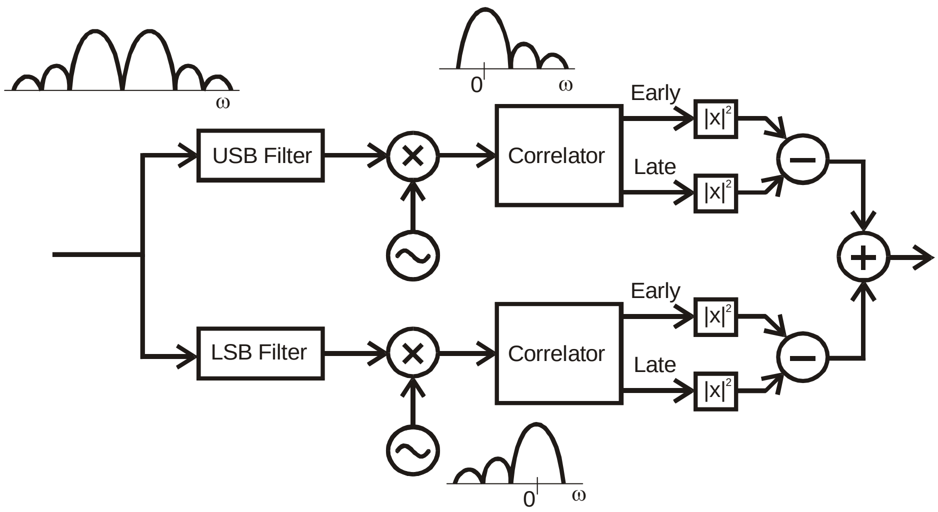


Figure 5. BOC non-coherent processing.

4.2 Very early – very late correlator

The most obvious way to handle the problem with tracking of correct peak of BOC modulation correlation function is the technique denoted as very early – very late (VEVL) correlator, also known as "bump-jump" method, see Fine and Wilson (1999), Barker et al.(2002). In comparison to classical early – late correlator structure, VEVL has a further couple of early and late taps, see Figure 6. This extra couple of taps are adjusted to track the side-peaks of correlation function.

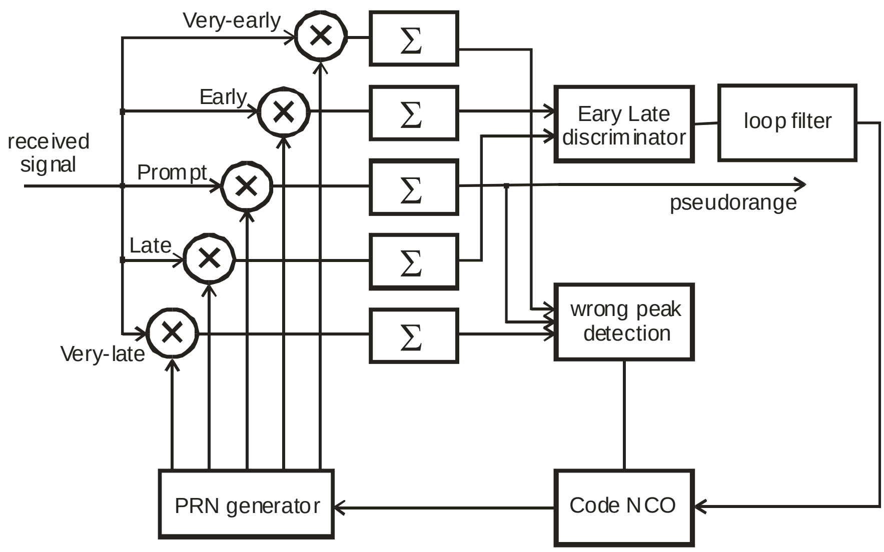


Figure 6. Structure of Very Early Very Late correlator.

The early and late taps together with prompt tap are intended for tracking the correct (centre) peak of the correlation function like in the classical early – late correlator. The spacing (a correlator width) is adjusted to enable tracking the narrow peak of particular type of BOC correlation function. The additional very early – very late taps are set to watch the side-peaks of the correlation function. When the correlator tracks the correct correlation function peak, the prompt tap output is greater than from very-early and very-late ones. In case of repetitively greater output from very-early or very-late taps, the wrong peak tracking is declared. Then the phase of a local signal replica is adjusted to restore the tracking of the correct peak.

4.3 Deconvolution correlator

This method is based on the linearization of discriminator characteristic (S-curve function) with using of multiple taps in the correlator structure, see Fante (2003). The discriminator characteristic of the classical no-coherent two taps early and late correlator (NCEL) is given by

$$S(\tau)=\left|R(\tau+D/2)\right|^2-\left|R(\tau-D/2)\right|^2, \tag{3}$$

where $R(\tau)$ is cross-correlation function, τ is tracking error and D is the spacing between the early and late taps. The two taps discriminator characteristic for BOC modulation has multiple wrong stabile nodes (Figure 3). To obtain the linear monotonic discriminator characteristic in the entire range of tracking error τ, the number of taps are incorporated into correlator structure. The outputs of particular taps are then scaled by $a(m)$ coefficients to meet this demand. The discriminator characteristic is then given by

$$S(\tau)=\sum_{m=1}^{2N}a(m)\left|R(\tau+(m-N+0.5)D)\right|^2, \tag{4}$$

where the N is the number of couples of taps. In comparison to early late structure, this correlator has worse sensitivity.

5. PROPOSED BOC CORRELATOR

The aim of the development of the new correlator is to find such a correlator that fully utilize the BOC modulation benefits and is not sensitive to the false node tracking. The developed correlator should have two outputs; first output should be equal to the tracking error of coherent processing of BOC modulated signal and the second one should compare envelopes of correlation or similar product which has only one stable tracking node.

The first section of the correlator is comprised of the BOC delay correlator (Figure 1). The second section is a sum of the both side-band early minus late discriminators $D_U(\tau)$ and $D_L(\tau)$ which is derived from side-band correlators outputs $R_U(\tau)$ and $R_L(\tau)$ (Figure 7).

The upper-side-band correlator $R_U(\tau)$ gives correlation between received BOC modulated signal $c_{N,M}(t)$ and spectrally shifted PRN code $x_{N,M}(t)$,

$$x_{N,M}(t)=c\left(\lfloor Nf_0t\rfloor\right)\cdot e^{\mathrm{j}2\pi Mf_0t}. \tag{5}$$

The BOC modulated signal can be decomposed to Fourier series as follows

$$\begin{aligned}&c_{M,N}(t)=\\&=\frac{2}{\pi}c\left(\lfloor Nf_0t\rfloor\right)\cdot\sum_{n=-\infty}^{\infty}\frac{-\mathrm{j}\cdot\mathrm{sgn}(2n+1)}{2n+1}e^{\mathrm{j}2\pi(2n+1)Mf_0t}=\\&=\frac{2}{\pi}\sum_{n=-\infty}^{\infty}\frac{-\mathrm{j}\cdot\mathrm{sgn}(2n+1)}{2n+1}x_{N,(2n+1)\cdot M}(t).\end{aligned} \tag{6}$$

We can resolve this situation in frequency domain

$$\begin{aligned}&\mathrm{F}\left[c_{M,N}(t)\right]=\\&=\frac{2}{\pi}\sum_{n=-\infty}^{\infty}\frac{-\mathrm{j}\cdot\mathrm{sgn}(2n+1)}{2n+1}X_N\left(2\pi Mf_0(2n+1)\right)\end{aligned} \tag{7}$$

where $X_N(\omega)$ is spectrum of the PRN code with chip-rate Nf_0.

The signal $\frac{-\mathrm{j}\cdot\mathrm{sgn}(2n+1)}{2n+1}x_{N,(2n+1)M}$ is one of the PRN components of the BOC modulated signal. Due to the limited (however non-zero) cross-correlation between $x_{N,(2i+1)M}$ and $x_{N,(2j+1)M}$, $i\neq j$, the proposed upper sideband correlator $R_U(\tau)$ estimates cross-correlation between spectrally shifted PRN code $x_{N,M}(t)$ and related component $\frac{2}{\pi}x_{N,M}$ of the received signal. The correlator output $R_U(\tau)$ is given by

$$R_U(\tau)=\frac{2}{\pi}R_N(\tau)+\varepsilon(\tau), \tag{8}$$

where the $R_U(\tau)$ is an autocorrelation function of PRN code with chip-rate Nf_0 and component $\varepsilon(\tau)$ covers the cross-correlation remainder of other signal components $w_{N,M}(t)$

$$\begin{aligned}w_{N,M}(t)&=\frac{2}{\pi}\sum_{\substack{n=-\infty\\n\neq0}}^{\infty}\frac{-\mathrm{j}\cdot\mathrm{sgn}(2n+1)}{2n+1}x_{N,(2n+1)\cdot M}(t)=\\&=c_{M,N}(t)-\frac{2}{\pi}x_{N,M},\end{aligned} \tag{9}$$

$$\varepsilon(\tau)=\int_{T_1} c_{M,N}(t+\tau)\,w^*_{N,M}(t)\mathrm{d}t\ . \qquad (10)$$

Analogically, the lower-size-band correlation is given by

$$R_L(\tau)=\frac{2}{\pi}R_N(\tau)+\varepsilon^*(\tau)\,. \qquad (11)$$

The output of discriminator second section $D_2(\tau)$ summarizes the sideband outputs $D_U(\tau)$ and $D_L(\tau)$. Suitability of discriminator characteristic is conditioned by monotony of the $R_U(\tau)$ and $R_L(\tau)$ sides. It depends mainly on the relationship of the wanted correlation $\frac{2}{\pi}R_N(\tau)$ and the parasitic correlation $\varepsilon(\tau)$. The situation is much better for higher order BOC modulation ($M >> N$).

Thus, this correlator (Figure 7) has been designed and simulated. The calculated discriminator characteristic of the correlator for low order modulation BOC(1,1) is shown on the Figure 8. The discriminator characteristic of the proposed correlator has only one stable node, which is convenient.

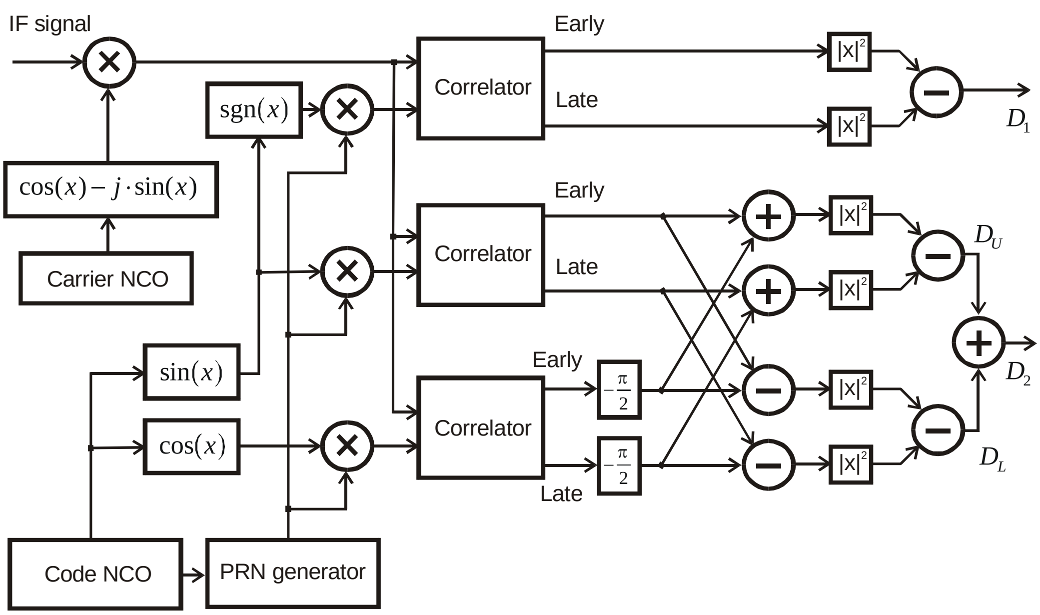


Figure 7. Proposed BOC correlator.

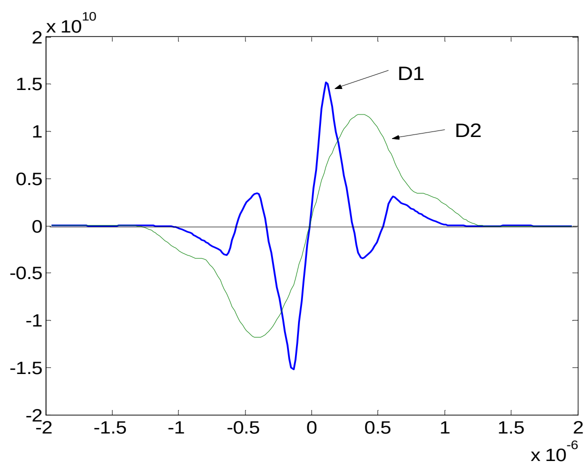


Figure 8. Discriminator characteristic for BOC(1,1) modulation

In the frame of GARDA project described BOC correlator is planned to be investigated and tested in GRANADA Galileo system simulator. For example, the sophisticated method of combining information from both correlator outputs should be developed and tested.

6. CONCLUSIONS

The Galileo receiver development is carried out in the frame of GARDA project. The project is financed by the GJU (Galileo Joint Undertaking) in the frame of the Galileo R&D activities under the EC 6th Framework Program. The key technologies concerning Galileo receiver and Galileo correlators are developed. The Czech Technical University is GARDA project consortium member with responsibility for the Galileo code and carrier tracking problems.

The Galileo system uses some modern sophisticated modulation schemes based on the BOC modulation. The correlation function of the BOC modulated signal has several correlation peaks, which cause the problem of detection of the correct one. In the frame of the project the new correlator for processing the BOC modulated signal has been developed. The developed correlator has two delay discriminator outputs: the first for fine tracking and the second based on comparison of the correlation function envelope power. The discriminator characteristic has only one stable node and serves for the detection of incorrect tracking node. The correlator is planned to be tested with the GRANADA tool.

REFERENCES

Barker B.; Betz J.; Clark J.; Correia J.; Gillis J.; Lazar S.; Rehborn K.; Straton J. (2002): ***Overview of the GPS M Code Signa***l [online], MITRE Technical Papers Archive, [cit. 2004-11-04] http://www.mitre.org/work/tech_papers/tech_papers_00/betz_overview/betz_overview.pdf

Fante R. (2003): ***Unambiguous Tracker for GPS Binary-Offset-Carrier Signals*** [online], MITRE Technical Papers Archive, [cit. 2004-11-04] http://www.mitre.org/work/tech_papers/tech_papers_03/fante_tracker/fante.pdf

Fine P.; Wilson W. (1999): ***Tracking Algorithm for GPS Offset Carrier Signals*** Proceedings of ION 1999 National Technical Meeting, Institute of Navigation, January 1990. 671–676.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 184-191

Journal of Global Positioning Systems
2005
© CPGPS

A Demonstrative GPS-aided Automatic Landslide Monitoring System in Sichuan Province

Pinggen Zhou, Baishen Zhou

China Institute of Geo-environmental Monitoring, Beijing, 100081,China
e-mail: zhoupg@mail.cigem.gov.cn; Tel:86-10-62179926; Fax: 86-10-62173426

Jingjun Guo, Donghang Li, Zhigang Ding

Tsinghua University, Beijing, 100084, China
e-mail: lidonghang@tsinghua.org.cn; Tel:86-10-62702528; Fax: 86-10-62702529

Yanming Feng

Cooperative Research Centre for Satellite Systems, Queensland University of Technology, GPO 2434,Q4001
e-mail: y.feng@qut.edu.au; Tel:0061-7-38641363; Fax: 0061-7-38641517

Received: 15 November 2004 / Accepted: 12 July 2005

Abstract. In China, the geological disasters of landslide and mud-rock flows cause losses of over 1000 lives and total economic losses of over 10 billions of RMB each year. There have been about 90,000 identified landslide sites, mostly distributed in several southern and north-western provinces of China. In the reservoir area of the Three-Gorge project only, over 1000 landslide sides have been identified. A joint research was launched involving authors from a number of scientific institutions to explore technologies and methodologies for landslide monitoring with focus on the characteristics of the geological disasters at the up and middle reaches of Yangtze River. This paper studies the combined technologies for landslides monitoring and presents a demonstrative automatic landslide monitoring system in a chosen ancient landslide site, where the creeping movement process continues since its latest large sliding on August 25, 1981. The landslide test-bed is 500 m long and 300 m wide located in the Ya'an-Xiakou area in Sichuan province. To study the mechanism of the sliding process, 15 permanent GPS monuments were built in the area for regular observations. Automatic ombrometers, digital thermometers, underground water-level-meter and ground fissure-displacement-meter were set-up as well. The data from these automatic sensors are collected and automatically sent to the data process centre in Beijing via the Beidou-1 communication satellite. The paper also compares the landslide results from three GPS observing campaigns, demonstrates the feasibility to identify the displacements at the accuracy level of 2 mm using the dual-frequency GPS receivers. The results are encouraging and further analyses will be conducted, considering influences of non-GPS measurements.

Key words: Augmented Reality, GPS, Inertial Sensors, Kalman Filter, Integrated Systems

1 Introduction

China is a country frequently suffered from various geological hazards. With the boom of economy and a great deal of ongoing engineering projects such as transportation, water conservancy and resource development, as well as the influence of environmental changes, damages caused by landslide are sharply increasing. According to Tang (2004), in recent years, rock falls, landslide and debris flow have caused an economic loss of about 10 billion RMB and over 1000 lives per annum. In China, about 700 Counties (local governments) have suffered from geologic hazards for long and tens of millions of residents living in these areas face threatens of serious geologic hazards, but lack of a sense of safety. This sometimes causes serious social problems. Therefore, it is particularly important to develop cost-effective reliable landslide monitoring systems and technologies.

Landslide monitoring is a complex of technological means. Deformation monitoring is one of the most important parts, providing an important basis for identifying landslide. Conventional geodetic surveying methods for deformation monitoring include transit traverse survey, triangulation method, levelling survey, total station methods. These methods provide reasonable accuracy, but requiring skilled professionals to conduct the work in situ, resulting in heavy workload, high personnel risk and low efficiency. Monitoring and timely alarms in case of hazard cannot be realized at night or in continuous rain. Use of GPS can improve the situations considerably (Xu et al, 2003, Xu et al.,2003, Gili and Corominas, 2000). First, GPS can operate in all-weather conditions and inter-station visibility is not required; second, GPS static relative positioning can achieve millimetre accuracy to meet the requirements (Guo et al, 2004 and Zhao et al, 2001). Third, if the satellite communication is available to transfer the observation data from the remote site lacking of landline communication to an urban data processing centre, thus realizing automatic monitoring is more feasible.

Addressing the characteristics of geologic hazards in the middle and upper reach of Yangtze River in China, China Institute of Geo-Environment Monitoring and Tsinghua University jointly constructed GPS-based landslide monitoring demonstrative network at the Ya'an-Xiakou landslide site in Sichuan Province. Regular monitoring results since 2003 have demonstrated that GPS can provide required accuracy for landslide monitoring in the demonstrative case, which can lead to completely replacing conventional geodetic surveying instruments, showing also advantages in terms of speed and efficiency (Guo et al, 2004). The test-bed is also equipped other monitoring devices such as automatic pluviometer, digital thermometer, underground water gauge and geofracture displacement meter, to form a complete set of automatic landslide monitoring system. All the monitoring data are transmitted to Beijing Monitoring Analysis Centre through the China Beidou No 1 navigation/satellite communication channel. The system is 24/7 automatic unmanned monitoring system.

2 GPS landslide monitoring experiments and results

2.1 Description of the experimental landslide area

As shown in Figure 1, the Ya'an- Xiakou, also called Wujia Mountain, landslideis situated in the east bank of the gorge in the middle reaches of Longxi River in Longxi Village, north of Ya'an City, Sichuan Province. It is the largest landslide in the local area. The whole landslide area consists of an ancient rock fall and landslide deposit of about 10 million M2. (unknown time), a reactivation landslide of 2.60 million M2 (1981) as well as a deformable body of 0.75 million M2 (the part deformed in late 1990s after sliding in 1981). Overall sliding is direction I is from East to West. The sliding mass is the ancient rock fall deposit, gravel mixed with mauve clay, and the stone's diameter is 0.2to 2m. The sliding mass features favourable conditions for influent seepage and shallow layer of underground water. The region enjoys warm and wet climate, and that is why Ya'an is also called Rain City. The mean annual temperature there is 16ºc, and the annual rainfall is 1,800mm, mostly in the season June to September.

On the night of August 19th, 1981, a storm broke out, and a large amount of rain flowed into the slope, triggering off a large-scale slide and resulting in collapse of houses and ruin of roads and channels. Different degrees of deformation occurred in the creep deformation body of the reactivation landslide following the rain season in 1995, and formed a potential sliding mass(Zhao). Experimental area is located at the reactivation body, 500m long and 300m wide, running from east to west slightly southwards about 25.50º.

Fig. 1 Sliding mass at the Xiakou landslide 2.1 Reference Frames

2.2 Ya'an-Xiakou GPS landslide monitoring network

Reference points. Two points at the stable rock mass outside the sliding mass were chosen as reference sites. One is in the east and the other is in the northern ridge, numbered as ya15 and ya17, respectively

Monitoring points. We set up monitoring points according to the features of the local sliding mass. These points should reflect the general deformation direction of the sliding mass, magnitude of the deformation, as well as range of the sliding mass and deformation speed. Reception of signals from the satellite was considered at each point. There should be no large-area screening objects over the monitoring point (Liu *et al.*, 1996). There are in total fifteen monitoring points placed in the test area, of which there are five points placed from top to bottom crossing the section of landslide, each being close to the inclinometer and crack displacement measuring device. Three monitoring points are set up next to the pluviometer and the underground water level measuring device. The remaining seven points are scattered around the sliding mass. The average distance between the points is 120 m, with the maximum of 229 m and the minimum of 17 m. At each monitoring point, the GPS monument was cast with reinforced concrete with 1.5 m underneath, and 1.5 m above the ground. A forced centring device is also placed on the top of the observation pillar. Figure 2 illustrates the layout of the monitoring points of the GPS network.

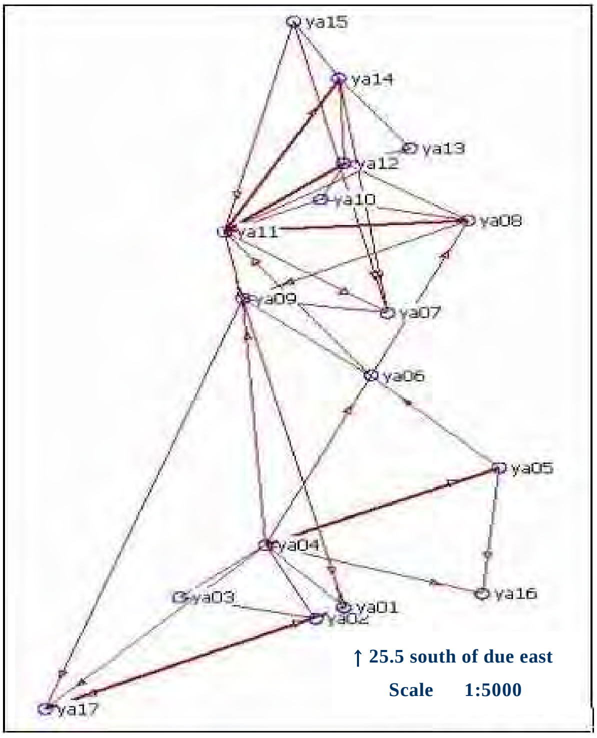


Fig. 2 Layout of GPS observation points

2.3 Collection and processing of GPS data

Four Novatel PROPAK-II-TR2 dual-frequency GPS receivers were used for GPS data collection and observation conducted in March and October of 2003 in the area. The data collection with a PDA-based device developed by the researchers from Tsinghua University GPS data were recorded at 5 seconds interval with 1 hour session length. The elevation cut-off angle of 10º was set. Meanwhile, Activesync software was used to enable PDA and computer communication. The observation data were converted to the standard RINEX format for processing. The baseline positioning was completed using Trimble Geomatics Office software. The results show the maximum RMS values of 8 mm, the minimum value of 2 mm and the average value of 3.4 mm. 2.2.1 Kalman filter functional model for the GPS observations

2.4 Results and Analysis

Baseline consistency check

To ensure accuracy of the observation results, in each round of experiments, several baselines were observed twice, allowing to conduct internal consistence checks between the solutions. Table 1 gives the results from the test conducted in March 2003. As seen, the maximum difference between the two baseline solutions is 4 mm, while the RMS is 3 mm for all the baselines.

Tab. 1 Baseline consistency Check

From	To	Baseline length (m)	Computation method	Ratio	RMS (m)
ya04	ya05	142.184	L1 fixed	4.5	.003
ya04	ya05	142.188	L1 fixed	5.7	.003
ya08	ya11	136.728	L1 fixed	23.5	.003
ya08	ya11	136.732	L1 fixed	53.6	.003
ya11	ya12	84.006	L1 fixed	22.2	.003
ya11	ya12	84.003	L1 fixed	31.8	.003
ya11	ya14	133.899	L1 fixed	11.4	.003
ya11	ya14	133.899	L1 fixed	33.2	.002

Close- loop consistence checks

The zero theoretical value for sum of the coordinate differences in a loop formed by three or more baseline vectors can be used to check the results as well. Referring to Figure 2, each triangle in the network can form a close loop. There are totally 40 close loops, of which 39 passed the test. The failed one related to the monitoring point ya06, which is close to a house and same threes. The quality of GPS signals might be affected.

GPS network adjustment

The baseline solutions from the test in March 2003 was adjusted using rank defect free network adjustment (Cui et al., 2000). Table 2 shows the positional RMS accuracy of the network adjustment results.

Gauss projection

To minimize the deformation of the coordinates in the landslide monitoring, Gauss map projection is made at the arbitrary zone closest to the survey area. That is, the longitude 103 º 01’ E is taken as the central meridian and projection is made at north latitude 30º as latitude of pedal with the projection height of 820 m (geodetic height), thus minimizing the projection error and ensuring that the projection error will not affect the accuracy of monitoring. To ensure that Y coordinate is the positive one, we moved Y coordinate eastwards to the longitude 103 ºE. Table 3 gives the coordinate projection results.

Tab. 2 Positional accuracy after network adjustment process

Points	Error in vertical axis（m）	Error in horizontal axis（m）
Ya01	0.009	0.004
Ya02	0.003	0.002
Ya03	0.003	0.003
ya04	0.002	0.002
ya05	0.003	0.002
ya06	0.003	0.003
ya07	0.002	0.002
ya08	0.002	0.002
ya09	0.002	0.001
ya10	0.003	0.002
ya11	0.002	0.002
ya12	0.002	0.002
ya13	0.002	0.003
ya14	0.002	0.002
ya15	0.003	0.003
ya16	0.003	0.003
ya17	0.002	0.002
Maximum	0.009	0.004
Minimum	0.002	0.001
Average	0.003	0.002

Tab. 3 Results of Gauss Map Projection

Points	North coordinate（m）	Error in vertical axis（m）	East coordinate（m）	Error in horizontal axis（m）
ya01	7840.813	0.009	1463.268	0.004
ya02	7858.419	0.003	1462.534	0.002
ya03	7920.010	0.003	1509.314	0.003
ya04	7860.236	0.002	1523.291	0.002
ya05	7718.432	0.003	1519.071	0.002
ya06	7753.409	0.003	1610.950	0.003
ya07	7725.563	0.002	1648.389	0.002
ya08	7656.025	0.002	1690.707	0.002
ya09	7793.487	0.002	1692.598	0.001
ya10	7723.173	0.003	1739.654	0.002
ya11	7781.867	0.002	1741.487	0.002
ya12	7700.098	0.002	1758.348	0.002
ya13	7662.237	0.002	1752.438	0.003
ya14	7675.655	0.002	1815.950	0.002
ya15	7680.316	0.003	1864.527	0.003
ya16	7767.043	0.003	1439.250	0.003
ya17	8022.828	0.002	1467.256	0.002

Conversion of coordinates

To facilitate analysis on the sliding mass, we convert the gauss projection coordinates to rectangular coordinate system at the sliding mass direction. The main sliding mass direction ya01-to-ya12 is the negative direction for X- axis, and Y- axis is normal to X- axis to form the right hand rectangular coordinate system. Table gives the coordinates of each point after the coordinate conversion.

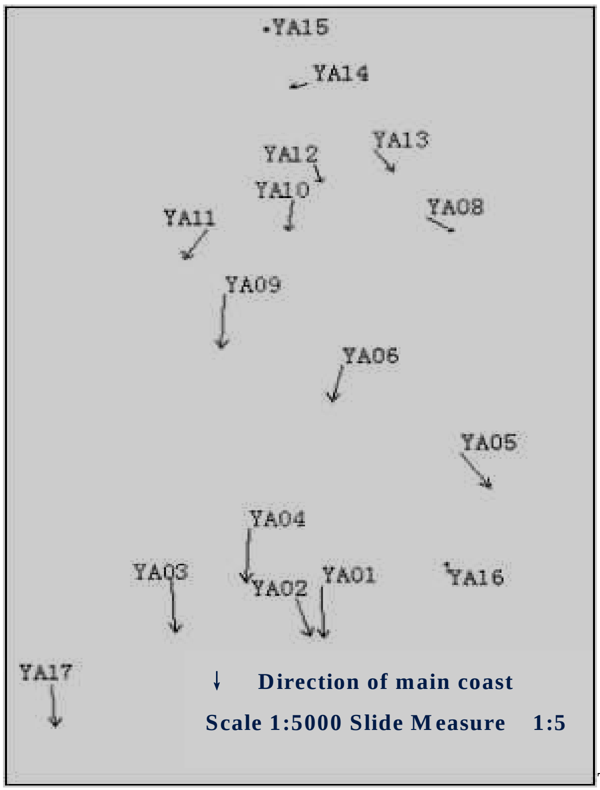
Tab. 4 Results of Coordinate Conversion

Points	X after conversion (m)	Y after conversion (m)
ya01	917.109	846.902
ya02	908.869	831.326
ya03	924.583	755.597
ya04	962.927	803.534
ya05	1020.155	933.346
ya06	1088.032	862.227
ya07	1133.811	871.246
ya08	1201.940	915.798
ya09	1144.478	790.908
ya10	1217.218	834.120
ya11	1193.608	780.353
ya12	1244.024	846.902
ya13	1254.986	883.620
ya14	1306.538	844.170
ya16	927.183	923.826
ya17	842.364	680.894

Table 4 gives the results from the experiment conducted in March 2003. The second experiment took place in October. The same observation methods and data processing strategies were taken as in March 2003, but the coordinates of the reference point ya15 known from the pervious time were used in the network adjustment. In addition, no observation was taken for Point ya07, due to the complete blockage of signals by crops.

Tab. 5 Displacement of points

	dx(m)	dy(m)
YA01	-0.040	-0.001
YA02	-0.031	0.008
YA03	-0.042	0.000
YA04	-0.042	-0.003
YA05	-0.028	0.016
YA06	-0.027	-0.008
YA08	-0.012	0.015
YA09	-0.042	-0.003
YA10	-0.025	-0.005
YA11	-0.021	-0.014
YA12	-0.015	0.004
YA13	-0.017	0.010
YA14	-0.003	-0.012
YA15	0.000	0.000
YA16	0.008	-0.002
YA17	-0.032	0.001



Ta

Fig. 3 Map of points displacement

In Table 5, Dx and Dy are the coordinates differences between two experiments for x and y components, respectively.

As seen from Table 5 and Figure 3, sliding mass not only moves along the main sliding direction of the landslide, but in the direction normal to the main sliding direction. This is because the sliding mass projects in the middle, and sinks at the two sides. Results from GPS landslide monitoring accurately reflect sliding tendency and displacement of the sliding.

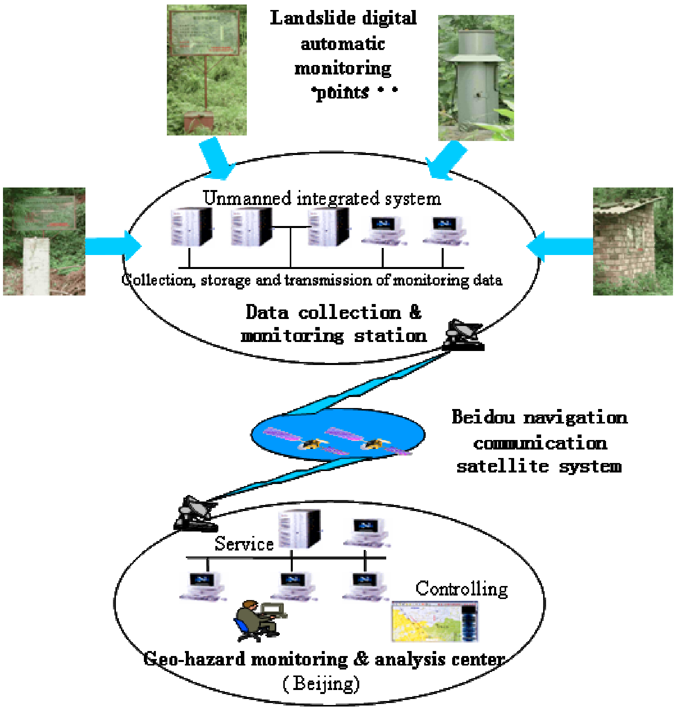


Fig. 4 Composition of landslide remote monitoring system

3 An automatic landslide monitoring system remotely controlled though Beidou-1 communication satellite

3.1 Components of the system

As shown in Figure 4, the remotely controllable automatic landslide monitoring system, built in the Ya'an landslide area, has four major components: in-field data collection & monitoring station, digital automatic landslide monitoring point, Beidou-1 navigation satellite communication system and "geologic hazard monitoring & analysis center". Detailed descriptions are given as follows.

(1) Field data collection & monitoring station

The field data collection & monitoring station is a set of unmanned integrated data collection system. It includes an information monitoring and collection platform for various geological measurements, such as absolute displacement monitoring for landslide ground, monitoring for displacement of the deeper parts of the landslide, dynamic monitoring for landslide underground water, landslide relative displacement monitoring, and monitoring for inducing factors of landslide. These elements are consolidated and integrated into one unit to collect, store, compress and transfer various data. Figure 5 is the integrated rack.

Fig. 5 Field data collection & monitoring station

Fig. 6 Underground water level monitoring point

(2) Digital automatic landslide monitoring points

These digital automatic landslide monitoring points are scattered everywhere in the landslide area in order to use various digital monitoring methods to collect data concerning the geological environmental parameters, such as underground water level, water temperature, rainfall, displacement of the earth surface, deformation of deeper parts and so on. The monitoring devices include borehole tiltmeter, geofracture displacement meter, digital automatic pluviometer, automatic water level meter and so on. Figure 6 indicate one of the monitoring points-underground water level monitoring point.

(3) Beidou-1 navigation/ communication satellite system

"Beidou-1 navigation communication satellite system"is a regional navigation system in China. It can provide all-weather, around-clock satellite navigation information, time service and two-way communication service (Ha, 2004). Communication signal of this system is stable, and high intensively encrypted measures are designed to ensure security and reliability, so it is very suitable for monitoring landslide hazard in harsh and complicated field conditions. This system is composed of space segment, a ground controlling center and Beidou-1 user terminal. Figure 7 illustrates the concept of the Beidou-1 system and services.

The space segment of Beidou-1 navigation communication satellite system includes two earth synchronous orbit satellites (GEO), namely BDSTAR-1 and BDSTAR-2. The signal transmission device on the satellite is available to complete the relay task for two–way wireless signals between ground controlling central station and the client terminal.

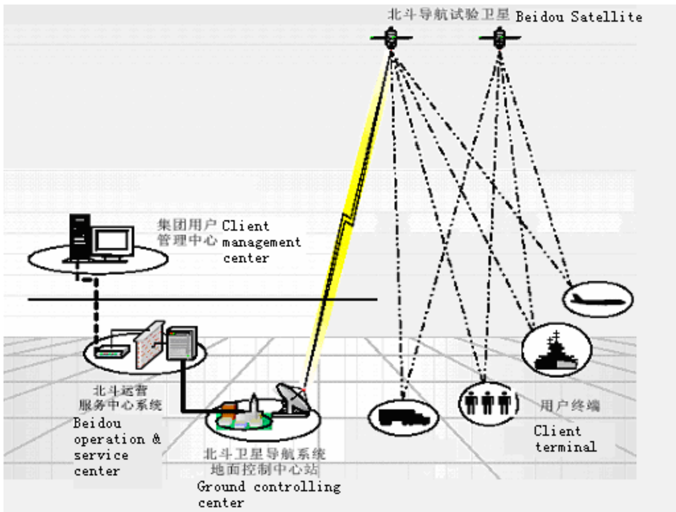


Fig. 7 Diagram for comprehensive information service structure of Beidou-1 navigation communication satellite system

(4) Geologic hazard monitoring & analysis center

The geologic hazard monitoring & analysis center comprises a Beidou-1 communication user receiver, a database server and an analysis unit. In terms of its functions, the system includes the controlling software for landslide remote monitoring system operation a database and monitoring data analysis software. Through Beidou-1 communication client receiver, the center uses satellite signal channel to receive data collected and sent from the field, and process and store the data and results in the database server. The system controlling software can automatically feed back information, and meanwhile telecontrol field data collection & monitoring station according to users' requirements for field data collection. The data analysis software can read data from the database any time, analyse and process the data collected, obtain the changing and alarm information of geologic feature, and thus providing valuable reference to the researchers and identifying landslide hazard, and make accurate early-warning. Figure 8 shows the components of the experimental system for monitoring and analysis center.

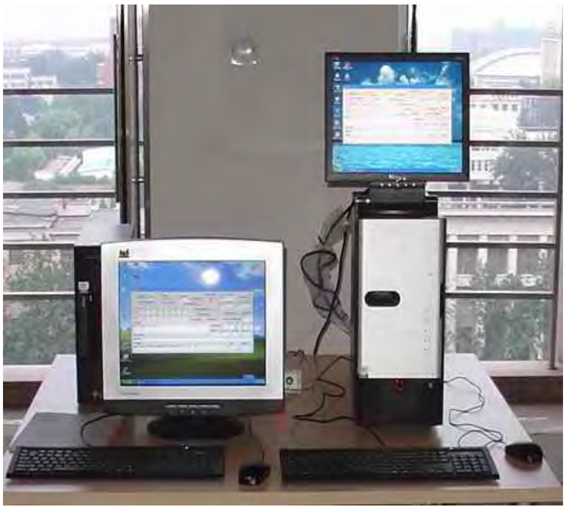


Figure 8. Geo-hazard monitoring & analysis system (experimental center)

3.2 Operation and monitoring function of the system

The remote automatic landslide monitoring system, based on Beidou-1satellite communication, operates as follows:

(1) Field data collection, storage and processing of the geological environmental features, including underground water level, water temperature, rainfall, the earth's surface displacement, deformation of the deeper parts collected by the various digital automatic landslide monitoring points scattering in the landslide monitoring region;

(2) Regular data packaging and compression, or via telecontrol according to the needs, using the transmission and communication function of Beidou-1 satellite navigation system to directly send the data to the geologic hazard monitoring & analysis center, where the data are analyzed and processed. Meanwhile it can send feedback information and control instructions to the field system through Beidou-1 system as well.

The system enables the center facility to complete the remote monitoring and control over the landslide monitoring sensor through Beidou-1 satellite communication link; while the landslide monitoring sensor regularly transmits the data processed primarily and locally stored by the field data collection & monitoring station to the central facility. The system is also equipped with satellite communication receiver for cross checking and mutual checking. It can also display the sensor status, check the sensor data; query the current record file in the monitoring station, request for sending the data on the specified date and time, request for sending all the data of designated file, telecontrol running of the monitoring station from the starting point, display in real time the sensor data; time mark, the received data; display the particular running status of the program the instructions ID length of received information; telecontrol the normal exit (save the setting of the current state) and cancellation of the program (exit of the program without change of the original state). Meanwhile the central facility has the capacity of automatically processing remote monitoring data and publicizing the early-warning information promptly.

4 Conclusions

GPS monitoring experiment results from the Ya'an-Xiakou testbed have demonstrated that GPS can provide sufficient accuracy to meet landslide displacement monitoring requirements, leading to the replacement of the conventional geodetic surveying methods. Use of GPS-based methods can significantly improve the efficiency of landslide monitoring and reduce the work intensity of labour. In addition, with means of the Beidou-1 satellite communication system, remote automatic monitoring of the landslide becomes feasible. The demonstrative GPS-aided automatic landslide monitoring system developed has performed well, providing an excellent technological platform for the academics and experts engaged in the research of geological and environmental hazards to timely, accurately and conveniently obtain the real-time monitoring data at each dangerous landslide region. As a result, it saves costs for labour and data acquisition in filed and avoid potential hazard to personnel in the field. Meanwhile, the comprehensive information volume obtained through the system is far greater than the effective data obtained through the conventional means. Moreover, a database with rich and detailed resources can be set up, which would be of great significance to the landslide hazard investigation analysis, forecast, early-warning and hazard mechanism research in the future.

ACKNOWLEDGEMENTS

This work was carried out with financial support from the Ministry of Science and Technology (MOST, P. R. China and partially from Australia-China Special Funds for Scientific and Technological Cooperation bilaterally offered by MOST and Department of Education, Sciences and Technology (DEST), Australia.

References

Yin Y. (2004): ***Initial study on the hazard-relief strategy of geological hazard in China,*** The Chinese Journal of Geological Hazard and Control, 15 (1): 1-8

Zhou P. (2004): ***Indicators and techniques of landslide monitoring***, Journal of Geomechanics (in Chinese), 10(1):19-26

Xu, S, W. Cheng, X. Huang, D. Xu (2003): ***The investigation of landslide monitoring in the Three Gorges Reservoir region by applying GPS***, Journal of Hydraulic Engineering (in Chinese). (1):114-118

Gili J.A.; Corominas J. (2000): ***Using Global Position System techniques in landslide monitoring***, Engineering Geology (special issue), 55:167-192

Guo J J.; Yang Z.; Ding P.; Zhou X Zhu. (2004): ***Application of GPS on landslide monitoring-A case study of Xiakou landslide***, Journal of Geomechanics (in Chinese). 10(1): 63-70

Zhao Y. (2001): ***Mechanical analysis of sliding process of Ya'an Xiakou landslide and prediction of its hazard model***, Journal of Engineering Geology(in Chinese). 9(2): 188-193

Liu, D Y Shi, J Guo (1996): ***The Elements and Data Processing of Global Position System***, Tongji University Press, Shanghai

X Cui, Z Yu, B. Tao (2000): ***The Generalized Surveying Adjustment***, Survey and Mapping Press, Beijing

Ha B. (2004) ***Beidou Navigation and Position Satellites System***, China Metrology,(3):22-23

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 192-200

Journal of Global Positioning Systems
2005

Comparative study of interpolation techniques for ultra-tight integration of GPS/INS/PL sensors

S.Ravindra Babu and Jinling Wang

School of Surveying and Spatial Information Systems, University of New South Wales, Sydney - 2052, Australia
Email: s.ravi@ unsw.edu.au, Tel: 61-2-9385 4206 Fax: 61-2-9313 7493

Received: 26 November 2004 / Accepted: 12 July 2005

Abstract: Ultra-tight architecture plays a key role in improving the robustness of the integrated GPS/INS/PL (Pseudolite) system by aiding GPS receiver's carrier tracking loops with the Doppler information derived from INS (Inertial Navigation System) velocity measurements. This results in a lower carrier tracking loop bandwidth and subsequent improvement in measurement accuracy. Some other benefits using this architecture include: robust cycle-slip detection and correction, improved anti-jam performance, and weak signal detection.

Typically the integration/navigation filter run at a rate of 1 to 100 Hz, which is insufficient to aid the carrier tracking loop as such loops normally run at about 1000 Hz. Two approaches were envisioned to solve this problem. One approach is to run the navigation Kalman filter at a higher rate, and the other is to run the filter at a lower rate and interpolate the measurements to the required rate. Although the first approach seems to be straightforward, it is computationally very intensive and requires a huge amount of processing power, adding to the cost and complexity of the system. The second method interpolates the low rate Doppler measurements from the navigation filter using multirate signal processing algorithms. Due to its efficiency and simpler architectures the interpolation method is adopted here.

Filtering is the key issue when designing interpolators as they remove the images caused in the upsampling process. Although direct form of filtering can be adopted, they increase the computations. To reduce the computational burden, two efficient ways of implementing the interpolators are proposed in this paper: Polyphase and CIC (Cascaded Integrator Comb). The paper summarizes the design and analysis of these two techniques, and our initial results suggest that CIC is relatively better in terms of performance and computational requirements.

Keywords: Tracking loops, Doppler, INS, interpolators, Polyphase, CIC.

1 Introduction

A conventional GPS receiver can track the signal if the received power and the vehicle dynamics are within its operational limits. But, the demands of the proliferating applications are much more. The receiver is expected to operate in reduced signal strength, multipath, interference, intentional and unintentional jamming environments. Moreover, automotive applications involve dynamics such as acceleration and jerk. Unfortunately, optimizing a single receiver to meet all these requirements is almost an impossible task; the design is usually optimized to cater to a particular environment. However, adding additional sensors not only increases its operational areas but also its reliability and robustness, and in fact it is this philosophy that drives the growth of integrated GPS/INS systems. Of many possible sensors, inertial navigation system (INS) is found to be optimal due to its immunity to electromagnetic signals and also its ability to provide navigation data at higher data rates.

Increasingly, Pseudolites (ground based GPS transmitters) are also seen as attractive aiding sensors primarily due to their capability to improve the geometrical strength, and also providing signals at places where GPS signals cannot be received (Wang et al., 2001). In the loose, tight and ultra-tight integration architectures, the time dependent systematic errors of the inertial sensor are calibrated using precise GPS/Pseudolite positioning solutions. During loss of GPS signals, Pseudolites can continue to calibrate the inertial sensor errors thereby improving the robustness of the integrated architecture. Therefore, for applications such as indoors, foliage etc., integrating Pseudolites with GPS and INS systems will certainly improve the performance especially in terms of robustness and reliability.

The integrated GPS/INS/Pseudolite systems are not new in the field of navigation, and are being developed for nearly two decades; however, the architectures in which these two systems can be integrated have changed over the period of time. The principal idea behind these architectures is if GPS or a Pseudolite can calibrate inertial sensor errors during normal operation, then the calibrated INS can provide navigation during GPS outages. Traditionally, both these systems were integrated in the so called loosely coupled architecture, where the navigation solutions from both the systems were combined in an external Kalman filter to provide an optimal solution. Though the implementation of this system looks simple, nevertheless there are limitations in this type of architecture (Farrell, 2000). To overcome some of these shortcomings, tightly coupled architecture was developed where a GPS/Pseudolite receiver is not considered as a navigation system but as a sensor that provides pseudo-ranges (PR) and pseudo range-rates (PRR) which can be integrated with INS variables. Some of the advantages of this system are: it can provide navigation even with one satellite though with a degradation, and lesser correlation of the integration variables (PR, PRR) reduces the complexity of the integration Kalman filter.

The recent development in this series is the Ultra-tight integration, i.e. integration of I (in-phase) and Q (quadrature) variables from the receiver's tracking loops with INS. The inherent property of this system is the integration of INS derived Doppler feedback to the carrier tracking loops. This forms an important advantage of this system, as the INS Doppler aiding removes the vehicle Doppler from the GPS/Pseudolite signal, it facilitates a significant reduction in the carrier tracking loop bandwidth (Babu & Wang, 2004); on a comparative scale the dynamics on the pseudorandom noise code is very less due to its low frequency nature. The bandwidth reduction improves the anti-jamming performance of the receiver, and also increases the post correlated signal strength. In addition, due to lower bandwidths, the accuracy of the raw measurements is also increased.

But, the INS aiding of the receiver tracking loops require higher Doppler update rates from INS. As the update rate of the tracking loops is normally about 1 KHz, the derived Doppler rates should be generated at the same rate for the aiding to be efficient. One possible method is to run the Kalman filter at a high rate, i.e. 1000 Hz; however, this requires an extensive processing power. The second method is to generate Doppler at lower rates and then interpolate to the required rate (Beser et al., 2000; Gardner, 1993). This is the method adopted in this paper. The Kalman filter from which the Doppler is generated typically runs at 1 or 100 Hz, and the Doppler measurement is then interpolated by a factor of 10 or 100 for aiding.

An increase in sampling rate can be accomplished by using interpolators which can efficiently be designed using multi-rate signal processing techniques (Mitra, 1999; Crochiere & Rabiner, 1983). A lowpass FIR (finite impulse response) filter is used in the interpolators to remove the images caused in the upsampling process. The filter transfer function is efficiently realized using Polyphase and CIC (Cascaded integrator comb) techniques (Hentschel, T., & Fettweis, 1990). While the polyphase method involves decomposing the filter transfer function into parallel stages, CIC implements the interpolator transfer function without using multipliers. This paper discusses on the design issues of both these techniques with their advantages and disadvantages.

2 Doppler estimates from INS

The GPS/Pseudolite receiver computes its velocity by measuring the Doppler offsets on the GPS and Pseudolite signals. Therefore, measuring the Doppler signal accurately becomes imperative. After down converting the L1 signals to IF (intermediate frequency), the acquisition loops coarsely measures the carrier frequency and code offsets, and then pass these coarse measurements to the tracking loops for fine tracking. Due to their low loop bandwidths (typically about 12 to 18 Hz), the tracking loops are sensitive to the Doppler changes, whereas acquisition loops with a Doppler bin size of about 500 Hz are almost insensitive except in circumstances of very high dynamics. This places severe constraints on the tracking loops. For tracking high dynamics (acceleration and jerk), the bandwidth should be greater than 18 Hz with the order of the loop increased to 3Hz (Kaplan, 1996); however, this affects the quality of measurements and stability of the loop.

The received Doppler from satellites and Pseudolites are given as

$$f_{rx\_gps} = f_{tx}\,(1 - \frac{v_{rel}\,\vec{a}}{c}) \tag{1}$$

where f_{tx} is the transmitted GPS/Pseudolite L1 frequency (1575.42 MHz), $v_{rel} = v_t - v_r$ is the relative velocity between satellite and receiver, $\vec{a}$ is the line of sight vector, and c is the velocity of light. The total Doppler on the received signal is due to the satellite and receiver motion, and satellite and receiver clock biases as shown in equation (2).

$$f_{rx\_gps} = f_{rx\_motion} + f_{sat\_motion} + f_{clk\_bias} + f_{sat\_clk} + bias \quad (2)$$

The average rate of change of Doppler due to satellite motion is about 0.5Hz/s (Tsui, 2000), and the satellite clock bias is transmitted in the navigation data. Therefore, ignoring these two terms equation (2) can be simplified as

$$f_{rx\_gps} = f_{rx\_motion} + f_{clk\_bias} \quad (3)$$

The tracking loop bandwidth is determined by the receiver motion and clock bias as shown in equation (3). However, with oscillators better than 1 ppm, the Doppler is primarily dictated by the motion, i.e. $f_{rx\_gps} = f_{rx\_motion}$. The order and bandwidth of the carrier tracking loop is determined based on the expected dynamics. If there is only velocity in the receiver motion, a stable second order tracking loop can be used, but if the receiver experiences acceleration and jerk, to minimize the dynamic stress error a 3rd order loop is used. The design of a 3rd order loop is quite complex and also causes stability issues (Ward, 1998).

Ultra-tight tracking loop, as shown in Figure 1, overcomes this by integrating the INS derived Doppler with the tracking loops. This derived Doppler closely reflects the Doppler on the GPS and Pseudolite signals caused due to receiver motion, and if integrated, removes the Doppler from the base band signal; i.e. the Doppler due to receiver clock oscillator and any residual bias from the Kalman filter will only remain. This residual Doppler is usually small, and therefore, the tracking loop bandwidth can be reduced to about 3 to 5 Hz.

The Doppler derived from INS is given as

$$f_{rx\_ins} = f_{rx\_motion} + f_{res\_bias} \quad (4)$$

where f_{res_bias} is the Doppler caused by the residual bias in the complementary Kalman filter. Integrating this Doppler signal with the tracking loops gives

$$\begin{aligned} f_{res\_dopp} &= f_{rx\_gps} - f_{rx\_ins} \\ &= f_{clk\_bias} - f_{res\_bias} \end{aligned} \quad (5)$$

Therefore, in the ultra-tightly integrated system, the bandwidth is determined by the receiver clock bias and any residual bias in the Kalman filter facilitating a reduction in the carrier tracking bandwidth. However, to leverage the benefits from this system the Doppler from INS should have the same update rate as that required by the tracking loops. Normally, the update rate of the integration Kalman filter is about 1 to 100 Hz, but the tracking loops are updated at about 1 KHz rate. Interpolators are therefore used to increase the sampling frequency of Doppler. The subsequent sections discuss the design and efficient realization of the interpolators.

2.1 Interpolators design for Doppler re-sampling

To convert the low frequency INS derived Doppler to the high frequency rate required by the tracking loops, an interpolator is required. The design of the interpolator is critical as any signal distortion will have a direct impact on the loop bandwidth. In general, the interpolator has two blocks as shown in Figure 2: an upsampler which inserts $L-1$ zero samples between two successive input samples where L is the interpolation factor, and a low-pass filter to remove the images caused in the upsampling process.

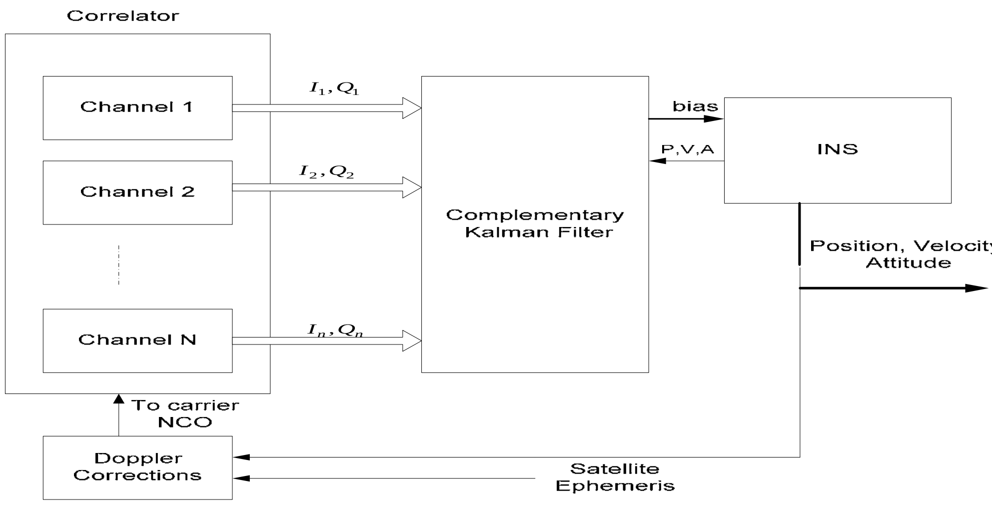


Fig. 1 Ultra-tight architecture

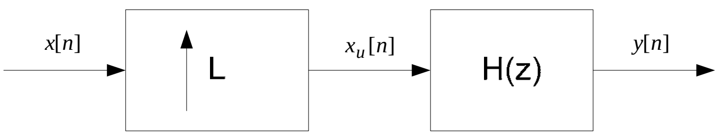


Fig. 2 Interpolator

The transfer function for the upsampler is given as

$$x_u[n] = \begin{cases} x[n/L], & n = 0, \pm L, \pm 2L, \ldots\ldots, \\ 0, & \textit{otherwise.} \end{cases} \tag{6}$$

where $x[n]$ is the input sequence, $x_u[n]$ is the output sequence. From equation (6), it can be clearly observed that the sampling rate of $x_u[n]$ is L times larger than the input sequence. However, the process of adding zeros in the upsampler results in a signal whose spectrum is an L-fold repetition of the input signal spectrum as given by (Mitra, 1999)

$$X(z) = \sum_{n=-\infty}^{\infty} x[n]\, z^{-n}$$

$$X_u(z) = \sum_{n=-\infty}^{\infty} x_u[n]\, z^{-n} \quad = X(z^L) \tag{7}$$

As a result, these $L-1$ additional images of the input spectrum distort the original spectrum. Therefore, a low pass filter $H(z)$ is used after the up-sampler removes these additional images and also fills the zero samples with non-zero values.

In this design, the Kalman filter is updated at every 100Hz and the tracking loops are updated at every 1 KHz. Therefore, an interpolation factor of 10 is required to convert the Doppler rate to 1000 Hz. As a first step, the upsampler inserts nine zeros between two successive input Doppler samples to increase the sampling rate, and then a Remez lowpass filter is used to remove the images caused by the upsampler. The input and output of upsampler is shown in Figure 3.

The distorted output is due to the insertion of zero samples. To remove this distortion and to smooth the output spectrum, an FIR Remez filter with a length of 80 samples was designed. The transfer function of the filter is given as

$$\left|H(e^{jw})\right| = \begin{cases} L, & |w| \leq w_c / L, \\ 0, & \pi / L \leq |w| \leq \pi \end{cases} \tag{8}$$

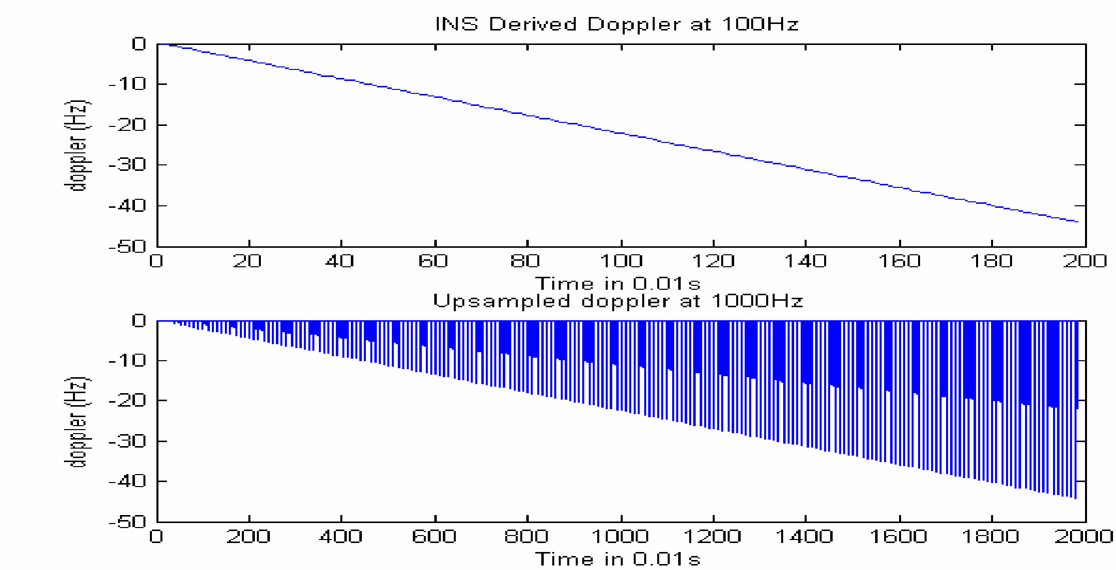


Fig. 3 Input and Output of Up Sampler

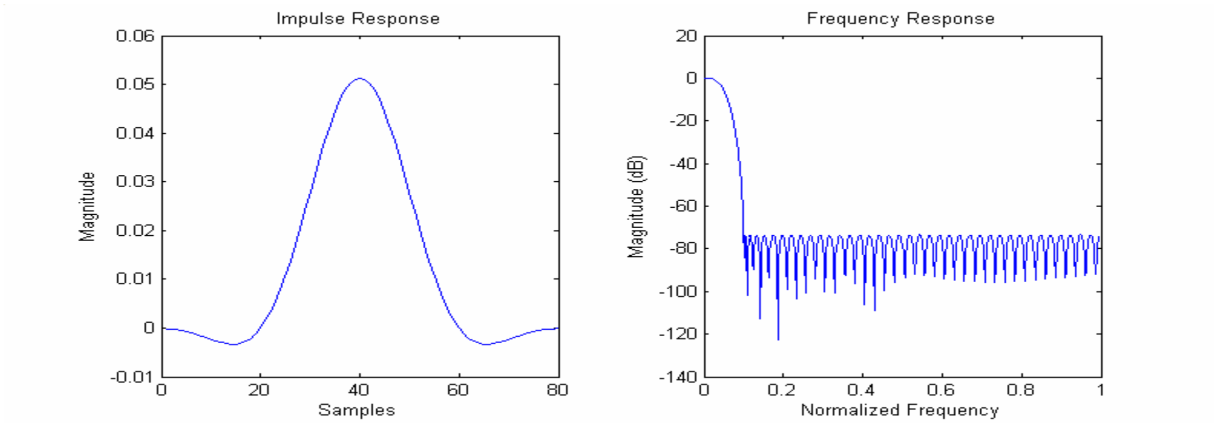


Fig. 4 Impulse and Frequency response of Remez filter

To preserve the signal shape, the pass-band edge should be at $w_p = w_c / L$, where w_c is the highest frequency in the input spectrum. The impulse and the frequency response of the Remez filter is shown in Figure 4.

The up sampled Doppler is then filtered to remove the images. The output of the filter shown in Figure 5 has almost the same shape as the original Doppler but with an increase in number of samples.

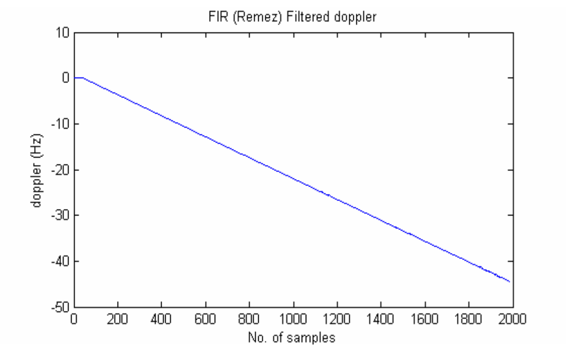


Fig. 5 Interpolated Doppler

Although the FIR filter has a linear phase, it is computationally intensive. Therefore, efficient structures such as Polyphase and CIC based techniques can be adopted to realize the low pass transfer function $H(z)$ which is the focus of the subsequent sections.

2.2 Polyphase Decomposition

Efficient realization of the interpolation filter $H(z)$ in equation (8) can be obtained using polyphase decomposition technique (Vaidyanathan, 1990). It is a method by which the original transfer function can be divided into L different branches given by

$$H(z) = \sum_{k=0}^{L-1} z^{-k} H_k(z^L) \tag{9}$$

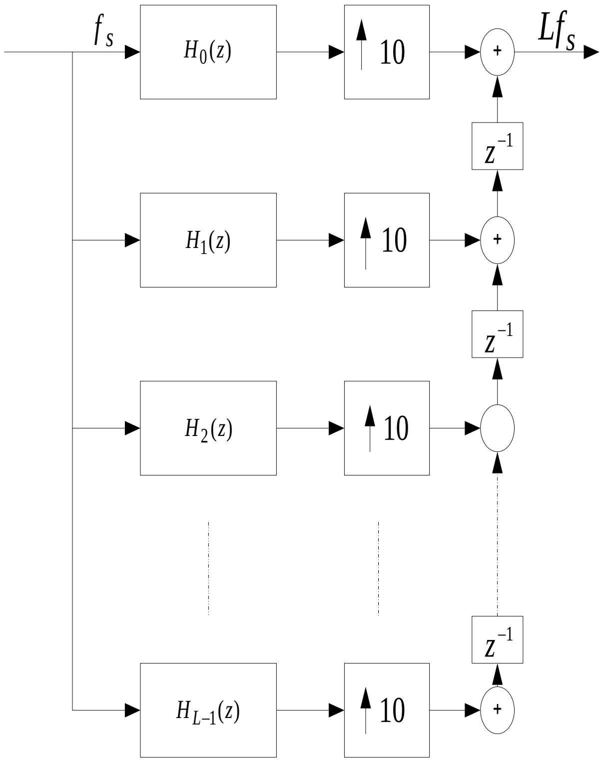
where

$$H_k(z) = \sum_{n=-\infty}^{\infty} x_k[n] z^{-n} = \sum_{n=-\infty}^{\infty} x[Ln+k] z^{-n} \qquad k = 0,1,..........L-1. \tag{10}$$

The subsequences $x_k[n]$ are called the polyphase components of the parent sequence $x[n]$, and the functions $H_k(z)$, given by the z-transform of $\{x_k[n]\}$, are called the polyphase components of $H(z)$. The transfer function given in equation (10) can be realized using Type II Polyphase decomposition as shown in Figure 6.

Note that in Figure 6, the input of the polyphase filters run at the low sampling rate f_s, while the output sampling frequency is $L\, f_s$, the increase is due to the generation of L samples from the parallel stages; i.e. for a single input sample there are ten output samples. The original impulse response of length 80 is split into 10 stages with each stage having 8 samples as shown in Figure 7. The relationship between the original FIR filter $h[n]$ with a length M and the polyphase filters $H_k[n]$ is given as

$$H_k[n] = h(k + nL) \qquad \begin{array}{l} k = 0,1,2,................L-1 \\ n = 0,1,2,...............K-1 \end{array} \tag{11}$$



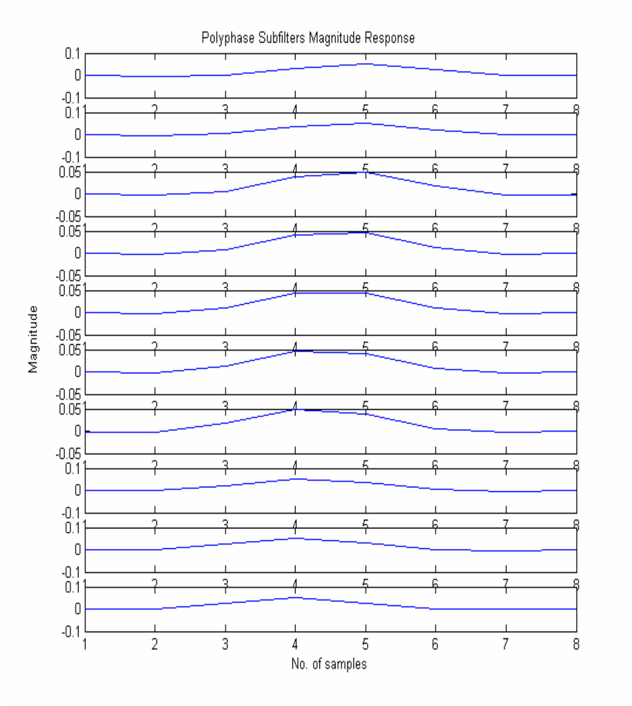


Fig. 7 Impulse responses of Polyphase filters $H_k(z)$

Fig. 6 Type II Polyphase decomposition with L=10

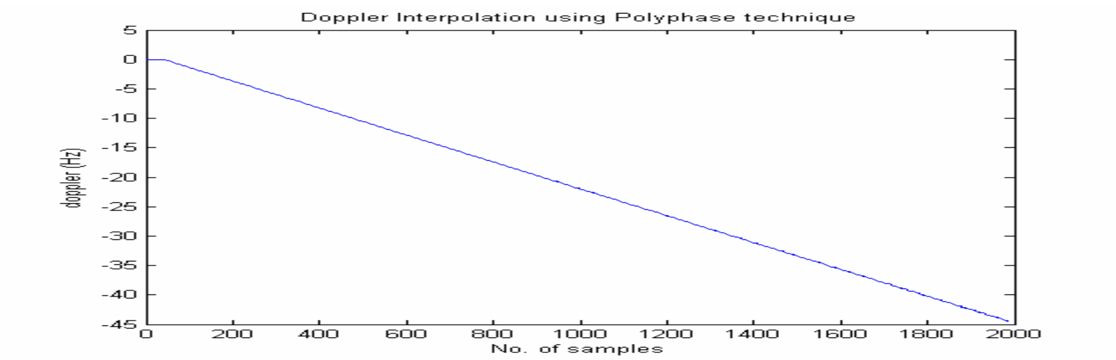


Fig. 8 Interpolated Doppler using polyphase techniques

where $K = M / L$ is an integer. However, all the subfilters may not have the same symmetrical impulse response property like the original filter; in the present experiment only the subfilter $H_k[5]$ has a symmetrical response, however, the other subfilters have a relationship that can be exploited to reduce the number of computations, i.e. $H_k[4]$, $H_k[3]$, $H_k[2]$, $H_k[1]$ are mirror images of $H_k[6]$, $H_k[7]$, $H_k[8]$, $H_k[9]$. These relations can be effectively utilized in developing an efficient architecture using only 36 multipliers and 79 two input adders. This is a significant reduction in computation when compared with the original FIR filter which uses 80 multipliers and 79 two input adders.

The Doppler samples at a rate of 100 Hz are fed to the polyphase subfilters $\{H_0(z)........H_{L-1}(z)\}$. Following the procedure mentioned above, ten Doppler samples are collected from the ten stages for each input Doppler sample. This increases the sampling rate to 1000 Hz as required by the tracking loops. Figure 8 shows the interpolated Doppler using polyphase decomposition technique.

2.3 CIC based interpolation

Cascaded integrator-comb (CIC), also called Hogenauer filters, are multi-rate filters that are used for sampling rate conversions. The main advantage of this filter is that it does not use multipliers; it only uses simple arithmetic operations like addition and subtraction to realize the sampling rate change (Hogenauer, 1981). The two fundamental blocks in a CIC filter are the comb filter and an integrator. Comb filters are linear phase FIR filters characterized by the transfer function (Crochiere & Rabiner, 1983)

$$h(n) = \begin{cases} 1, & 0 \le n \le N-1 \\ 0, & otherwise \end{cases} \tag{12}$$

where N is the number of taps in the filter. For a rate change of R (same as L in the polyphase filter), the comb filter can be described by $y[n] = x[n] - x[n-RM]$, where M is the differential delay; the value for M is usually limited to 1 or 2. The corresponding transfer function is given by $H_c(z) = 1 - z^{-RM}$. An integrator is a single-pole IIR filter with a unity feedback coefficient given by the transfer function $y[n] = y[n-1] + x[n]$. The frequency response is given by $H_I(z) = (1 - z^{-1})^{-1}$. By cascading the N integrator sections with N comb sections the CIC architecture is realized. One of the distinguishing factors of CIC is, the sampling rate of comb filters is different from the sampling rate of integrator, i.e. the comb runs at a lower sampling frequency f_s / R, whereas the integrator runs at f_s, which makes it easily programmable. The transfer function of the CIC at f_s is given by (Xilinx, 2003)

$$H(z) = H_c(z)\, H_I(z) = \frac{\left(1 - z^{-RM}\right)^N}{\left(1 - z^{-1}\right)^N} \tag{13}$$

The magnitude response at the output of CIC is

$$|H(f)| \approx \left| RM \frac{\sin(\pi M f)}{\pi M f} \right|^N \quad for\ 0 \le f < \frac{1}{M} \tag{14}$$

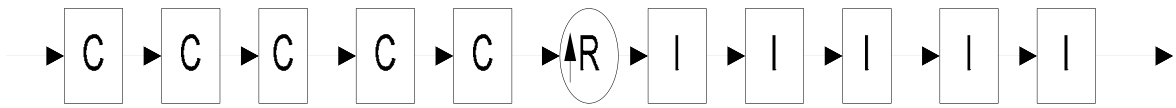


Fig. 9 CIC Interpolator with N = 5, M=1, R=10

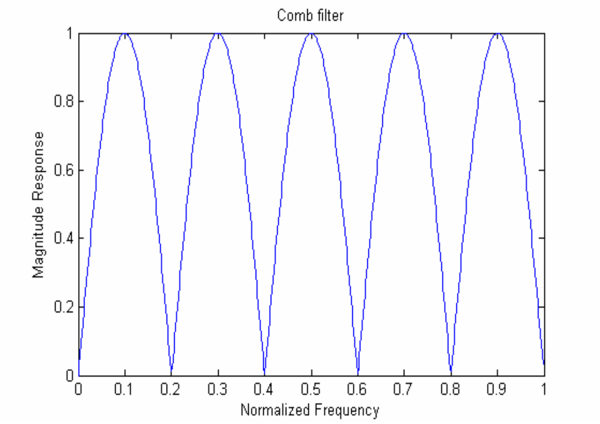


Fig. 10 Comb filter response for R=10

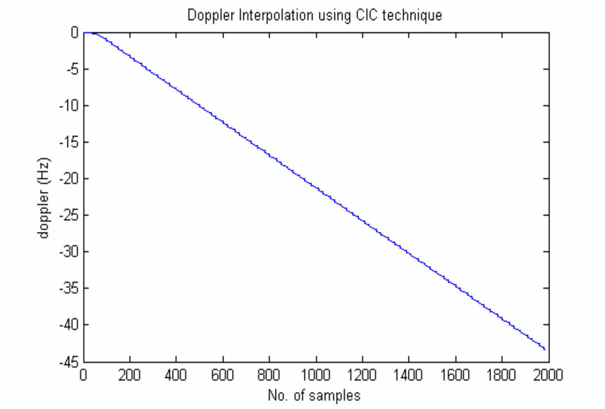


Fig. 11 Interpolation using CIC

Note from equation (14) that the output spectrum has nulls at $f = 1/M$. The filter is designed such that the images that result from the rate change conversion are placed at these nulls. The factors R, M and N are adjusted to optimize the filter for passband attenuation, stopband rejection, and passband droop. To increase the sampling frequency of INS derived Doppler to 1000Hz, a rate change factor R = 10 with 5 stages of comb and integrator are chosen. The block diagram of the CIC interpolator is shown in Figure 9.

The magnitude response of the comb filter and the CIC interpolated Doppler are shown in Figures 10 and 11 respectively. The nulls in Figure 10 represent the frequencies where the images are created by the insertion of $R-1$ zeros at the output of the comb filter. The output Doppler shows that the images are effectively removed, and the input shape is maintained.

3 Comparison between Polyphase and CIC

The Doppler from the navigation Kalman filter is interpolated using both Polyphase decomposition and CIC techniques. To compare the effectiveness of both, one out of 10 samples is taken from the outputs of both the interpolators and compared with the low rate input Doppler, and the results are plotted in Figure 12. The results show that the interpolated Doppler has a constant bias of about 0.5Hz, whereas the Doppler output from CIC closely matches the input Doppler. In addition, the CIC is computationally very intensive as there are no multipliers. Therefore, our preliminary analysis suggests that CIC has relatively superior performance than polyphase techniques.

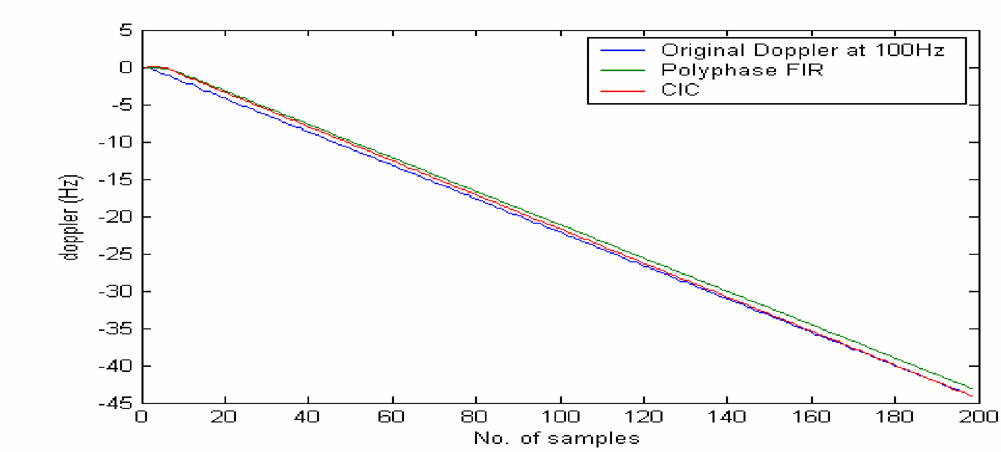


Fig. 12 Polyphase and CIC Interpolators performance

4 Conclusion

Aiding of the GPS/Pseudolite receiver carrier tracking loop with the INS derived Doppler is an inherent property in Ultra-tightly integrated systems. However, the derived Doppler cannot be directly used for aiding due to its low sampling rate. This paper has proposed an interpolation based technique by which the sampling rate can be increased. Although a direct form filtering method can be adopted, it is computationally intensive. Two algorithms are proposed to reduce the computational burden: Polyphase decomposition and CIC. While Polyphase technique is based on decomposing the original transfer function to L parallel stages, CIC increases the sampling frequency without any multipliers. A Doppler signal at 100 Hz is interpolated to a sampling frequency of 1000 Hz. The results from both methods are compared. The preliminary analysis suggests that the CIC is relatively more effective than the Polyphase decomposition technique.

Acknowledgements

This research is supported by an ARC (Australian Research Council) – Discovery Research Project on ‘Robust Positioning Based on Ultra-tight Integration of GPS, Pseudolites and Inertial Sensors’.

References

Babu R.; Wang J. (2004): ***Improving the Quality of IMU-Derived Doppler Estimates for Ultra-Tight GPS/INS Integration.*** GNSS 2004, Rotterdam, The Netherlands, 16-19 May.

Beser J.; Alexander S.; Crane R.; Rounds S.; Wyman J.; Baeder B. (2002): ***Trunav™: A Low-Cost Guidance/Navigation Unit Integrating a SAASM-based GPS and MEMS IMU in a Deeply Coupled Mechanization.*** 15th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation, Portland, OR, 24-27 September, 545-555.

Crochiere RE.; Rabiner LR. (1983): ***Multirate Digital Signal Processing***. Englewood Cliffs, NJ: Prentice Hall.

Farrell, JA.; Barth M. (1999): ***The Global Positioning System and Inertial Navigation***. McGraw-Hill.

Gardner FM. (1993): ***Interpolation in Digital Modems – Part I: Fundamentals***. IEEE Transactions on Communications, COM-41(3), 501-507.

Hentschel T.; Fettweis G. (2000): ***Sample Rate Conversion for Software Radio.*** IEEE Communications Magazine, August, 2-10.

Hogenauer EB. (1981): ***An Economical Class of Digital Filters for Decimation and Interpolation***. IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-29, 155-162.

Kaplan ED. (1996): ***Understanding GPS: Principles and Applications***. Artech House, MA.

Mitra SK. (1999): ***Digital Signal Processing – A Computer Based Approach.*** Tata McGraw-Hill Edition, India, ISBN 0-07-463723-1.

Tsui JBY. (2000): ***Fundamentals of Global Positioning Receivers – A Software Approach***. John Wiley & Sons, Inc.

Vaidyanathan PP. (1990): ***Multirate Digital Filter, Filter Banks, Polyphase Networks, and Applications: A Tutorial***. Proc. IEEE, 78, 56-93.

Vesma J. (1999): ***Optimizations and Applications of Polynomial-Based Interpolation Filters***. PhD Thesis, Tempere University of Technology, Finland, ISBN 952-15-0206-1.

Wang J.; Dai L.; Tsuiji T.; Rizos C.; Grejner-Brzezinska D.; Toth C. (2001): ***GPS/INS/Pseudolite Integration: Concepts, Simulation and Testing***. 14th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation, Salt Lake City, Ohio, 11-14 September, 2708-2715.

Ward P. (1998): ***Performance Comparisons Between FLL, PLL and a Novel FLL-Assisted PLL Carrier Tracking Loop Under RF Interference Conditions***. 11th Int. Tech. Meeting of the Satellite Division of the U.S. Inst. of Navigation, Nashville, Tennessee, 15-18 September, 783-795.

Xilinx. (2003): ***Digital Up Converter (DUC) DS276 v1.0 Product Specification***. (http://www.xilinx.com/ipcenter/catalog/logicore/docs/duc.pdf).

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 201-206

Journal of Global Positioning Systems
2005

Performance Analysis of GPS Integer Ambiguity Resolution Using External Aiding Information

Sebum Chun, Chulbum Kwon, Eunsung Lee, Young Jae Lee, Teasam Kang

Department of Aerospace Engineering, Konkuk University, Seoul, Korea
e-mail: koreagnss@koreagnss.com Tel: +82-2-458-0164 Fax: +82-2-444-6670

Gyu-In Jee

Department of Electronics Engineering, Konkuk University, Seoul, Korea
e-mail: gijee@konkuk.ac.kr Tel: +82-2-450-3070 Fax: +82-2-3437-5235

Received: 27 November 2004 / Accepted: 26 October 2005

Abstract. The integer ambiguity should be resolved at the beginning stage of GPS carrier phase positioning. In this procedure, some additional information can be used to improve the ambiguity resolution performance, for example, the positioning information of INS/GPS integrated system and baseline constraint of two adjacent antennas. This improvement is well known by experiment or simulation. But the quantitative characteristic of the improvement is not known yet. In this paper, we analyse this improvement quantitatively using the success rate.

Key words: GPS, RTK, Success rate, Baseline Constraint, INS/GPS

1 Introduction

GPS carrier phase is applicable for real time navigation, attitude determination and geodetic surveys which require sub-centimeter positioning accuracy. But it needs to resolve GPS carrier phase integer ambiguities. The most popular way to determine integer ambiguities is a searching method. Many searching type algorithms to resolve the integer ambiguities have been proposed. But, the way to estimate the initial real ambiguities is mainly based on GPS C/A code information. Therefore the reliability is very low because GPS code information has a positioning error of 10~15m.

Bad precision of initial position increases the calculation time because it enlarges the searching space and memory usage. The size of searching space is eight times larger when the variance of initial position is doubled (Hofmann-Wellenhof et al., 1997).

To overcome this problem, many external aiding methods to improve initial position precision have been proposed. Aiding information gives more precise position. So the method which uses external aiding has better performance in resolving GPS carrier phase integer ambiguities. But previous researches show the performance experimentally, not quantitatively.

In this paper, we analyse the performance improvement of integer ambiguity resolution with external aiding using the success-rate. This result can be used in the design stage of integrated systems as the basic information for reliability analysis.

2 Resolution of integer ambiguities with external aiding

We can classify the externally aided methods to resolve integer ambiguity into three groups (Lu, 1994).

2.1 Searching space fixing

It is impossible to resolve integer ambiguities using the general analytic methods. Therefore we need to use a searching method. In the searching sequence, the first step is to fix searching space. The size of the searching space depends on the confidence interval of float ambiguities. So the searching space can be reduced by some external aiding information that makes the confidence interval of float ambiguities small.

Computation time and memory consumption can also be reduced by a smaller searching space size.

2.2 Object function

External aiding information can be used in the object function. It has the same effect as increasing satellite visibility. Furthermore, it has better quality than original GPS measurements in some conditions. So we can get more advantages than the increasing satellite visibility.

2.3 Validation

As mentioned above, the external aiding information has better quality, so it can be used as a standard to confirm if the resolved integer ambiguity is true or not.

In this paper, we concentrate on the use of baseline constraint between adjacent antennas and positioning information of INS/GPS integrated system as external aiding information. And we analyze the changes of the success rate with these aiding information.

3 Success rate

The integer ambiguity of GPS carrier phase can be resolved in the discrete integer domain. Its distribution has a shape of probability mass function (p.m.f) and the p.m.f. of integer ambiguity can be described with equation (1).

$$P(\breve{a} = z),\ \mathrm{z} \in \mathrm{Z}^n \tag{1}$$

'a' is the true integer ambiguity vector. Equation (2) is the probability distribution of float ambiguity.

$$p_{\hat{a}}(x) = (2\pi)^{-\frac{n}{2}} \sqrt{\left|Q_{\hat{a}}^{-1}\right|} \exp\left(-\frac{1}{2}\|x-a\|_{Q_{\hat{a}}}^2\right) \tag{2}$$

'$Q_{\hat{a}}$' is covariance matrix of float ambiguity. Therefore equation (3) is the probability to fix true integer ambiguity and it is defined as the integer ambiguity success rate (Teunissen, 1998a)

$$P(\breve{a} = z) = \int_{s_z} p_{\hat{a}}(x)dx,\ \forall_z \in Z^n \tag{3}$$

Success rate can be defined by the probability that the float ambiguity vector is in the pull-in region. Pull-in region is the region which makes the float ambiguity which is inside it true integer ambiguity. Figure 1 shows the distribution of float ambiguity and figure 2 is the pull-in region.

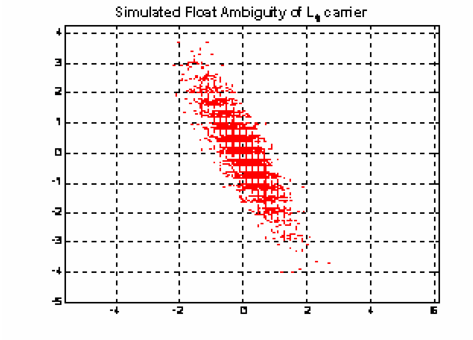


Fig. 1 Distribution of float ambiguity

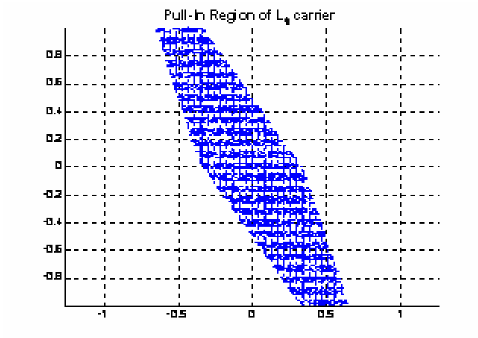


Fig. 2 Pull-in region

Success rate is the probability to fix the true integer ambiguity with a given float ambiguity. Estimation of the success rate doesn't need any real measurement, so it can be utilized as a designing tool for navigation system in the early stage of development procedure.

4 External aiding information

Improved navigation information can be provided by integrated navigation system which uses each sensor optimally. Followings are the method using baseline information and INS/GPS integrated system information. Each system has similarities that they decrease the variance of float ambiguity.

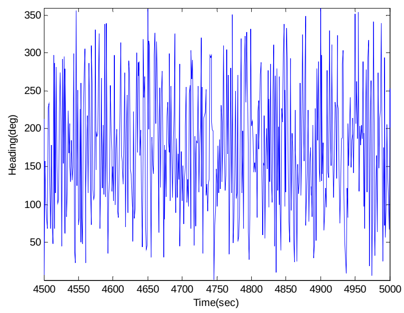
4.1 Baseline information between antennas

It is possible to measure the baseline length between two antennas in most cases to determine the attitude of a body. We apply baseline length constraints to improve initial position precision and it makes float ambiguity more precise

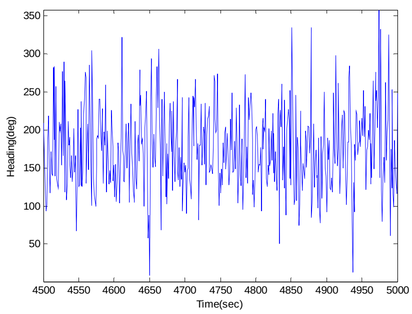
But the baseline observation model is nonlinear and the nonlinearity makes the final estimation results worse when general estimation methods, such as the least squares method or extended Kalman filter are applied. Particularly, it is remarkable when the baseline length is very short and the precision of initial position is poor. Equation (4) is nonlinear baseline observation model.

$$l = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2} \tag{4}$$

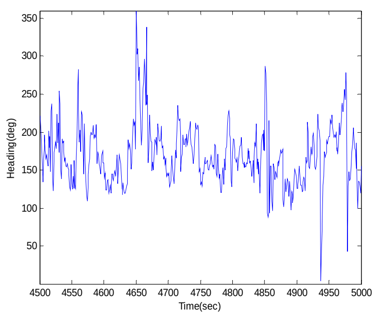
The unscented Kalman filter is proposed to reduce this linearization error. It utilizes unscented transform which is adapted for nonlinear transformation. The unscented transform uses several sigma points. Sigma points are generated from original state vector and its covariance matrix (Julier, 2004).



(a)



(b)



(c)

Fig. 3 Attitude determination result using C/A code and baseline constraint

(a:C/A code only, b:extended Kalman filter, c:unscented Kalman filter)

Figure 3 is the estimation result of initial attitude. Figure 3-(a) is the result of C/A code only, (b) is the result of the extended Kalman filter, which uses the linearized baseline observation model and (c) is the result of the unscented Kalman filter which uses a nonlinear model. We can find that the extended Kalman filter has a larger error than the unscented Kalman filter. Table 1 is the comparison of the estimation result with various estimation methods. The least squares case which uses the baseline constraint has the worse result than the C/A code only case (without the baseline constraint). It is the linearization error that makes the least squares method have larger errors when the baseline constraint is applied.

Tab. 1 Comparison of estimation result (unit: degree)

Least squares method (without BL)	Least squares method (with BL)	EKF (with BL)	UKF (with BL)
85.6758	91.1626	51.3863	41.5691

Figure 4 is the case that the initial position has a large error. This result shows that the extended Kalman filter has larger initial position sensitivity than the unscented Kalman filter. Therefore the extended Kalman filter takes a long time to converge. It looks like that the unscented Kalman filter is immune to the initial position error.

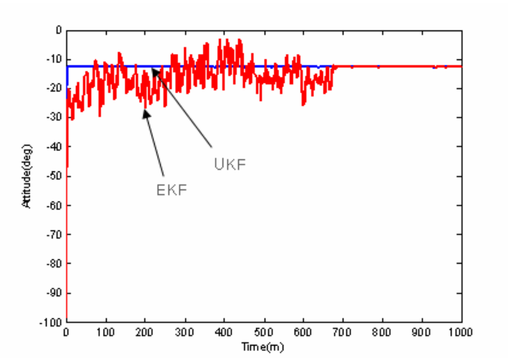


Fig. 4 Large initial position error

4.2 Positioning information of integrated INS/GPS system

An integrated INS/GPS system has higher availability than GPS carrier positioning information because we can assume that the navigation information of INS is given before GPS carrier phase information is given and we can utilize it in the estimation of float ambiguity. Figure 5 shows the trajectory of a vehicle and its variance of positioning result.

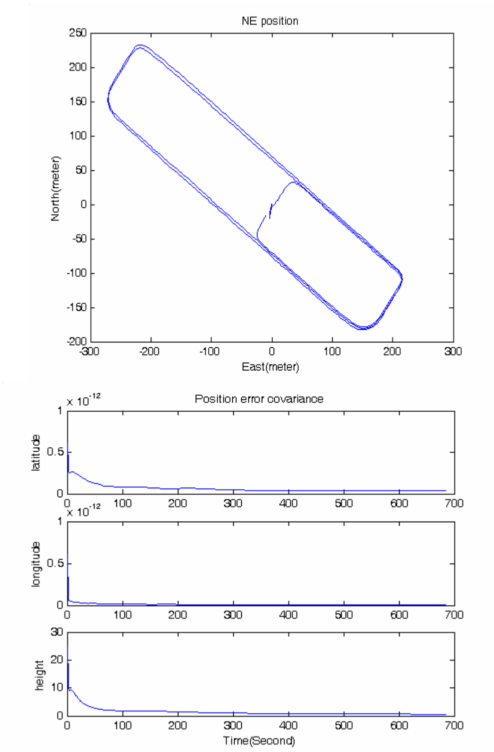


Fig. 5 Vehicle trajectory and the covariance of positioning result

5 Estimation of success rate

As mentioned above, the success rate cannot be estimated by analytic methods. Therefore we applied the Monte Carlo simulation method to estimate it (Teunissen, 1998b).

5.1 Success rate simulation

The covariance of float ambiguity must be estimated to analyse the success rate when the external aiding information is applied. The covariance of float ambiguity is determined by the covariance of initial position and carrier phase measurements. Equation (6) shows the relationship between the variance of float ambiguity and others.

$$Q_{\hat{a}} = Q_{\Phi} + \frac{1}{\lambda^2} H Q_{dx} H^T \tag{6}$$

Q_{Φ} is the covariance of carrier phase measurements and Q_{dx} is the covariance of initial position.

We can generate the adequate number of float ambiguity vector whose variance is derived by equation (6). And the success rate estimation can be performed by the Monte Carlo simulation with the generated float ambiguity vectors.

5.2 LAMBDA

We can estimate the success rate using the generated float ambiguity and Monte Carlo simulation. But large dimension of float ambiguity vector makes it impossible because of its huge computation time and memory consumption. LAMBDA (Least squares AMBiguity Decorrelation Algorithm) is the answer of this problem. LAMBDA uses decorrelation and sequential bootstrapping methods. These methods are effective in reduction of the computation time. Figure 6 depicts a full procedure of success rate estimation (Jong de P. J., 1996)

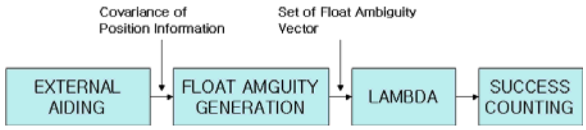


Fig. 6 The full procedure of success rate simulation

6 Result of success rate analysis

We applied Monte Carlo simulation method to estimate the success rate for the ambiguity resolution methods aided by the baseline constraint and INS/GPS integrated navigation system.

6.1 Baseline length constraint applied

The initial position information which is aided by the baseline constraint between adjacent antennas has smaller variance than the GPS C/A code only case. This information can make the precision of float ambiguity better. Figure 7 shows the change of observable satellites.

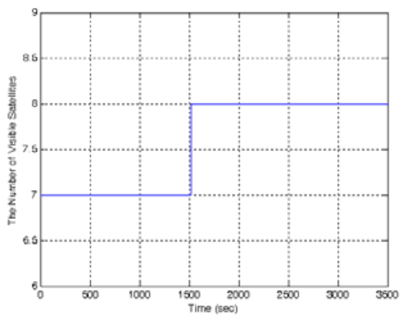
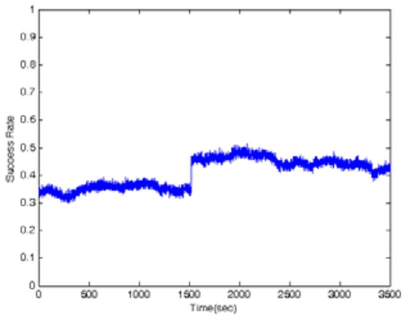
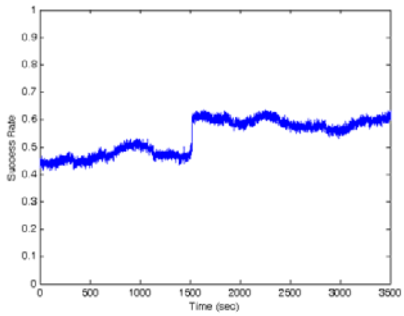


Fig. 7 Change of Observable Satellites

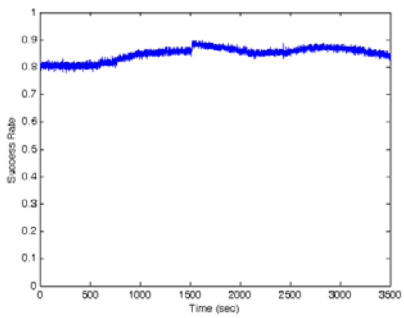
3500 float ambiguity vectors were generated and the success rate was estimated epoch by epoch. Figure 8-(a) is the estimation results of the success rate which uses C/A code information only. (b) is the result of the extended Kalman filter and (c) is the result of the unscented Kalman filter. From the figures, we can find the difference between these cases.



(a)



(b)



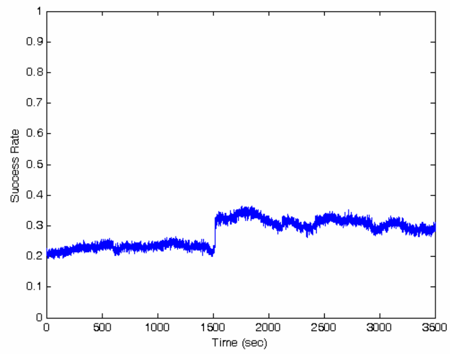
(c)

Fig. 8 Comparison of the Estimated Success Rates

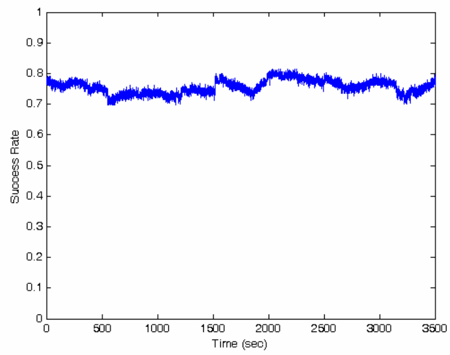
(a:C/A Code Only, b:EKF + Baseline Constraint, c:UKF+Baseline Constraint)

6.2 Positioning information of GPS/INS integrated system applied

It is possible to provide better positioning information when the INS/GPS integrated system gives the aiding information. Like a baseline constraint case, the INS/GPS integrated system can also increase the integer ambiguity success rate. To confirm it, the same simulation was repeated for the INS/GPS aided case.



(a)



(b)

Fig 9 Comparison of the Estimated Success Rates

(a:C/A Code Only, b:UKF+INS/GPS)

Figure 9-(a) is the result for the C/A code information only case, and (b) the result of INS/GPS integrated system. We can see that the success rate was significantly improved when we used INS/GPS position information.

7 Conclusion

The improvement of the success rate with the external aiding information has been analysed in this paper. We concentrate on the use of baseline constraint and positioning information of INS/GPS integrated navigation system as the aiding information.

We can see the improvement of success rate in both cases. Especially, an unscented Kalman filter has a higher ambiguity resolution success rate because the baseline observation model has large nonlinearities. It makes large error when we use a linearized model like an extended Kalman filter. INS/GPS aided case has a similar tendency with the baseline constraint case, but it is barely influenced by the change of observable satellites.

This result can be used as basic information for designing of integrated systems with various navigation sensors, because it does not need any real measurements, and only needs simulated measurements.

Acknowledgements

This paper was performed for the Aerospace Technology Research and Development Program funded by the Ministry of Commerce, Industry and Energy of Korea.

References

Hofmann-Wellenhof B.; Lichtenegger H.; Collins J. (1997): ***GPS Theory and Practice (4th Ed.).*** Springer, Wien New York.

Lu G. (1994): ***Development of a GPS Multi-Antenna System for Attitude Determination***, Dissertation, December 1994, Calgary, Alberta.

Teunissen P. J. G.; Kleusberg A. (1998a): ***GPS for Geodesy (2nd Ed.)***. Springer, Berlin Heidelberg.

Julier S. J. (2004): ***Unscented filtering and nonlinear estimation***, Proceeding of IEEE, Vol. 92, Issue 3, 401-422.

Teunissen P. J. G. (1998b): ***Success probability of integer GPS ambiguity rounding and bootstrapping***, Journal of Geodesy, Vol 72, 606-612.

Jong de P.J., C.C.J.M. Tiberius (1996): ***The LAMBDA method for integer ambiguity estimation: implementation aspects***, Delft Geodetic Computing Centre LGR Series No. 12.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 207-214

Journal of Global Positioning Systems
2005
© CPGPS

Ambiguity Resolution in GPS-based, Low-cost Attitude Determination

Shenglin Fan[1,2]

[1]School of Mathematical and Geospatial Sciences, RMIT University, Victoria, Australia
Email: shenglin.fan@rmit.edu.au, Tel: +61-3-99253278; Fax: +61-3-96632517
[2]Nanjing university of Aeronautics & Astronautics, Nanjing, Jiangsu Province, China
Email: slfan@nuaa.edu.cn, Tel: 61 +86-25-84892304; Fax: +86-25-84892184

Kefei Zhang, Falin Wu

School of Mathematical and Geospatial Sciences, RMIT University, Victoria, Australia
Email: kefei.zhang@rmit.edu.au, Tel: +61-3-99253272; Fax: +61-3-96632517

Received: 22 April 2005 / Accepted: 12 July 2005

Abstract. Reliable determination of integer ambiguities is a critical issue in high-precision global positioning system (GPS) applications such as kinematic positioning, fast control surveying and attitude determination. This paper discusses the integer ambiguity resolution procedures in attitude determination using single frequency carrier phase measurements. An optimised ambiguity search algorithm is proposed. This method can not only improve the computation efficiency and reduce the time for resolving ambiguities, but also improve the reliability of the ambiguity solution. The ambiguity search space is determined using float solutions and their variance and covariance matrices estimated by applying Kalman filter algorithm. The integer Gaussian transformation is then used to reduce the size of the search space and Cholesky factorisation algorithm is used to improve the efficiency of the integer ambiguity searching process. Finally, an ambiguity validation method by using the known baseline length and the relationship between the primary and secondary ambiguity groups is presented. The algorithms have been implemented within two low-cost Allstar GPS OEM boards. A number of field experiments have been conducted and the results show that a valid integer ambiguity solution in cold start mode can be identified within 3 minutes.

Key words: Ambiguity resolution, attitude determination, carrier phase, GPS

1 Introduction

GPS can usually provide two types of direct measurements. One is the pseudorange and the other is the carrier phase. The carrier phase measurements can be used for various high precision applications including kinematic positioning, static survey and attitude determination due to their low noise level. However the carrier phase measurements are ambiguous by an unknown integer number of cycles. This is well-known integer ambiguity resolution problem which requires a time-consuming initialisation process for both attitude determination and high-precision kinematic positioning.

There are several methods for resolving the integer ambiguities. In general they can be divided into two categories: search-based ambiguity resolution (Quinn, 1993; Sutton, 2002) and motion-based ambiguity resolution (Crassidis et al. 1999). Motion-based methods need to collect data for a period of time during which obvious changes of the visible GPS constellation or an apparent Rotation of the platform have occurred. The search-based methods use only single epoch measurements to identify the most likely ambiguity combination although sometimes it may be not the correct ambiguity due to the level of noise.

This paper mainly focuses on discussing the search-based method. There are three steps in this integer ambiguity resolution algorithm. First, a search space is determined in the ambiguity space (Juang, 2003; Xu, 2002) or solution space (Sutton, 1997; Sutton, 2002). The most likely integer ambiguity is, then, identified from the search space using some deterministic techniques such as least squares. Finally, the resolved ambiguities are validated and confirmed. This step is considered as a vital additional step for the GPS-based attitude system.

This paper will first outline the principle of attitude determination by using GPS carrier phase measurements. Then a fast and reliable algorithm for integer ambiguity resolution by using the known baseline length and the relationship between the primary and secondary ambiguity groups will be presented in detail. A number of experiments have been conducted to test the performance of the method. Finally, some useful results and conclusions are given.

2 GPS-based attitude determination using two antennae

The basic idea of attitude determination using GPS carrier phase measurements is similar with the principle used in interferometry. It is assumed that for a short baseline the unit vectors from both receivers to a given satellite are the same. This is based on the fact that the baseline length is negligibly small compared to the distance between GPS satellites and the user (approximately 22, 000km). This is shown in Fig. 1.

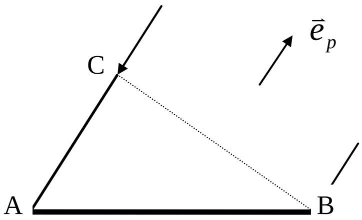


Fig.1 Interferometry principle

The difference between the true ranges from satellite "p" (p-th satellite) to antennae A and B can be expressed as:

$$|AC| = R_n^T \vec{e}_p = \rho_A^p - \rho_B^p \tag{1}$$

where, R_n is the baseline vector determined by A and B, $\vec{e}_p$ is a unit directional vector from antenna A or B to satellite "p", ρ_A^p and ρ_B^p are the distances between the antennae A and B and satellite "p" respectively.

The double-differenced observation equation can be expressed as:

$$\begin{aligned}\lambda DD_{AB}^{pq} &= R_n^T \vec{e}_p - R_n^T \vec{e}_q + \lambda (N_0)_{AB}^{pq} + \lambda \Delta v_{AB}^{pq} \\ &= (\vec{e}_p - \vec{e}_q)^T R + \lambda (N_0)_{AB}^{pq} + \lambda \Delta v_{AB}^{pq}\end{aligned} \tag{2}$$

where, DD_{AB}^{pq} is the phase double-differenced measurement, λ is the wavelength of the GPS signal (~20cm), $(N_0)_{AB}^{pq}$ is the integer double-differenced carrier phase ambiguity, and Δv_{AB}^{pq} is the measurement noise.

If there are M Satellites in view, all the measurements can be written in the following matrix form:

$$DD = HR_n + \lambda N + V \tag{3}$$

$$DD = \lambda \begin{bmatrix} DD_{AB}^{21} & DD_{AB}^{31} & \cdots & DD_{AB}^{M1} \end{bmatrix}^T$$

$$N = \begin{bmatrix} (N_0)_{AB}^{21} & (N_0)_{AB}^{31} & \cdots & (N_0)_{AB}^{M1} \end{bmatrix}^T$$

Where, R_n is a baseline vector in the local level system (North-East-Down coordinate system in this paper), H is the vector matrix of line of sight from antennae to GPS satellites in the local level system, V is the carrier phase difference-doubled measurement noise vector.

In equation (3), if the ambiguities have been fixed to integers, there will be only 3 unknowns (three components of the baseline vector R_n (x_n , y_n , z_n)). Therefore, if there are 4 satellites in view, there will be 3 independent double-differenced observations, the baseline vector (R_n) can be estimated by using a weighted least-squares method as follows:

$$\hat{R}_n = \left(H^T W H\right)^{-1} H^T W (DD - \lambda N) \tag{4}$$

where, $W = [\text{cov}(V)]^{-1}$.

If the relative position of two antennae can be determined with a sub-centimetre accuracy using the carrier phase observables, two of the three attitude parameters, usually heading and pitch angles of the platform can be estimated.

Suppose that the baseline is mounted along longitudinal direction, then the baseline vector in body frame is: $R_b = \begin{bmatrix} b & 0 & 0 \end{bmatrix}^T$, where b is the length of the baseline. The estimated baseline vector in the local level system is: $R_n = \begin{bmatrix} x_n & y_n & z_n \end{bmatrix}^T$

Then the heading and pitch angles can be calculated using:

$$\psi = \tan^{-1}(y_n / x_n) \tag{5}$$

$$\theta = -\tan^{-1}(z_n / \sqrt{x_n{}^2 + y_n{}^2}) \tag{6}$$

3 Carrier phase integer ambiguity resolution algorithm

It can be seen from Section 2 that the fast resolution of integer ambiguities is crucial to GPS-based attitude determination algorithm. Only when the integer ambiguities are resolved, the attitude angles can be calculated. In this section an ambiguity resolution algorithm for single frequency GPS receiver is proposed. In this algorithm, before the integer ambiguities can be determined, the ambiguities are treated as real values and estimated along with baseline vector. The real-value solution is often called the float solution. In this paper Kalman filter method is used to estimate the float solution of ambiguities, and the ambiguity search space is determined by using the float solution and its variance-covariance.

3.1 Float ambiguity estimation

The dynamic model of GPS-based attitude determination system can be described by the following equation:

$$\dot{X} = \begin{bmatrix} 0_{3\times3} & I_{3\times3} & 0_{3\times m} \\ 0_{3\times3} & (1/\tau)_{3\times3} & 0_{3\times m} \\ 0_{m\times3} & 0_{m\times3} & 0_{m\times m} \end{bmatrix} X + U \tag{7}$$

where, $X = [x \quad y \quad z \quad \dot{x} \quad \dot{y} \quad \dot{z} \quad N_1 \quad N_2 \quad \cdots \quad N_m]$ is the state vector, τ is a time constant relative to dynamic condition of the platform, U is the measurement noise vector, x, y, z are the components of baseline vector (R_n), $\dot{x}$, $\dot{y}$, $\dot{z}$ are the variance of components of the baseline vector R_n, $I_{3\times3}$ is a 3×3 unit matrix.

By re-writing equation (3), the measurement model can be expressed as:

$$Z = H'X + V \tag{8}$$

Where, $Z = DD$, $H' = [H \quad 0_{3\times3} \quad \lambda I_{m\times m}]$.

GPS-based attitude determination equations can be written in a discrete form according to equations (7) and (8) as follows:

$$X_k = \phi_{k/k-1} X_{k-1} + \Gamma_{k/k-1} U_{k-1} \qquad U_k \sim N(0, Q_k) \tag{9}$$

$$Z_k = H_k' X_k + V_k \qquad V_k \sim N(0, R_k) \tag{10}$$

where N is a normal distribution operator, the two variables in the bracket are mean value vector and variance-covariance matrix, $\phi_{k/k-1}$ is the state transition matrix, $\Gamma_{k/k-1}$ is the system disturbance matrix, H_k is the observation matrix, X_k is the state vector at epoch k, Z_k is the observation vector at epoch k.

Kalman filtering estimation can be expressed as:

$$\hat{X}_{k/k-1} = \Phi_{k/k-1} \hat{X}_{k-1/k-1} \tag{11}$$

$$\hat{X}_{k/k} = \hat{X}_{k/k-1} + K_k \left[Z_k - H'_k \hat{X}_{k/k-1} \right] \tag{12}$$

$$K_k = P_{k/k-1} H'^T_k \left[H'_k P_{k/k-1} H'^T_k + R_k \right]^{-1} \tag{13}$$

$$P_{k/k-1} = \Phi_{k/k-1} P_{k-1/k-1} \Phi^T_{k/k-1} + \Gamma_{k/k-1} Q_{k-1} \Gamma_{k/k-1} \tag{14}$$

$$P_{k/k} = \left[I - K_k H'_k \right] P_{k/k-1} \tag{15}$$

Where $\hat{X}_{k/k-1}$ is the predicted state vector, $\hat{X}_{k/k}$ is the estimation of filtering, K_k is the system gain matrix, $P_{k/k-1}$ is the variance matrix for $\hat{X}_{k/k-1}$, $P_{k/k}$ is the variance matrix for $\hat{X}_{k/k}$.

In theory, if a precise float solution can be obtained and rounded to the nearest integers, this should in most cases lead to the correct integer ambiguity set. Unfortunately, the float estimate obtained by this method is not precise enough, especially for a short observation period (Mohamed et al, 1998). Therefore, the correct ambiguity set usually need to be identified by a dedicate search method.

3.2 Determination of the search space

In order to further reduce the number of possible ambiguity candidates, all GPS satellites in view are divided into two groups: primary and secondary groups. The primary group which contains 5 satellites is used to determine the search space. The secondary group which includes remaining satellites is used to validate the correctness of the identified ambiguity set. So the integer ambiguity N, the float solution of integer ambiguity $\hat{N}$ and its variance-covariance P_N estimated by using Kalman filter technique can be re-written respectively as:

$$N = [N_1 \quad N_2] \qquad \hat{N} = [\hat{N}_1 \quad \hat{N}_2] \qquad P_N = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$$

where N_1 and N_2 are the integer ambiguities of the primary group and secondary group respectively, $\hat{N}_1$ and $\hat{N}_2$ are the float solutions of the ambiguities of the primary group and secondary group respectively, and P_{11} and P_{22} are ambiguity variance matrices of the primary group and secondary group respectively.

The integer ambiguities can be obtained by minimising the following cost function:

$$J = (\hat{N}_1 - N_1)^T P_{11}^{-1} (\hat{N}_1 - N_1) \qquad N_1 \in Z^m \tag{16}$$

Due to the integer constraints on the ambiguities and the fact that the ambiguity variance matrix is non-diagonal, the solution of Eq. (16) must be obtained by means of statistical criteria and a search method. The search space is described by the following inequality:

$$\left(N_1 - \hat{N}_1\right)^T P_{11}^{-1}\left(N_1 - \hat{N}_1\right) < \chi^2 \qquad (17)$$

This ellipsoidal region is centred at $\hat{N}_1$, its shape and orientation is governed by P_{11}, and its size can be controlled through the selection of positive constant χ^2. The size is assumed to be set such that the sought integer ambiguities are indeed in the search space. The solution is then obtained by searching through the entire search space. The efficiency of the search is poor, however, when the search space is highly elongated and the principal axes do not coincide with the grid axes. In order to reduce the correlation between ambiguities and make searching process more efficiently, the integer Gaussian transformation (Mohamed et al., 1998) is applied. The Gaussian transformation procedure of P_{11} is as follows:

[1] Factorise the matrix P_{11} using upper triangular factorisation: $P_{11} = U_1 D_{U1} U_1^T$, where U_1 is an upper triangular matrix, and D_{U1} is a diagonal matrix;

[2] Invert U_1 and round all elements of U_1^{-1} to their nearest integers, get Z_{U1};

[3] Transform the matrix P_{11}, and get: $P_{ZU1} = Z_{U1} P_{11} Z_{U1}{}^T$;

[4] Factorise the matrix P_{ZU1} using lower triangular factorisation: $P_{ZU1} = L_1 D_{L1} L_1^T$, where L_1 is a lower triangular matrix, and D_{L1} is a diagonal matrix;

[5] Invert L_1 and round all elements of L_1^{-1} to their nearest integers, get Z_{L1};

[6] Transform the matrix P_{ZU1}, and get: $P_{ZL1} = Z_{L1} P_{ZU1} Z_{L1}{}^T$.

Repeat above steps until Z_{L1} becomes an identity, and then the integer Gaussian transformation matrix is calculated by:

$$Z = \prod_{i=k}^{1} Z_{Li} Z_{Ui} \qquad (18)$$

where k is the total iteration steps. The transformed ambiguity vector is:

$$\hat{Z}_N = Z\hat{N}_1 \qquad (19)$$

and the transformed variance of $\hat{N}$ is:

$$P_{ZN} = Z P_{11} Z^T \qquad (20)$$

By substituting the equations (19) and (20) into inequality (17), the transformed search space can be formulated as

$$\left(Z_N - \hat{Z}_N\right)^T P_{ZN}^{-1}\left(Z_N - \hat{Z}_N\right) < \chi^2 \qquad (21)$$

and the cost function becomes:

$$J = (\hat{Z}_N - Z_N)^T P_{ZN}{}^{-1}(\hat{Z}_N - Z_N) \qquad Z_N \in Z^m \qquad (22)$$

Because the Gaussian transformation can decorrelate the ambiguities as much as possible, the float estimates become more precise and its variance becomes more diagonal-like. The transformed search space becomes more spherical.

Note that inequality (21) is a quadratic constraint, it is difficult to perform the searching process directly. Therefore the confidence interval of every ambiguity is used to replace the constraint of inequality (21).

$$z_i \in \left[round(\hat{z}_i - k\sqrt{Q_{ii}}) \quad round(\hat{z}_i + k\sqrt{Q_{ii}})\right] \qquad (23)$$

where Q_{ii} is the i-th row and i-th column element of P_{ZN}, z_i is i-th element of Z_N, $\hat{z}_i$ is i-th element of $\hat{Z}_N$, k is the confidence coefficient. By using the Gaussian transformation, the number in the search space defined by inequality (23) can be reduced dramatically.

The integer ambiguity search space is determined by equation (23) and all of the integer combinations in the search space are candidate ambiguity combinations.

3.3 Ambiguity searching process

The aim of ambiguity searching is to find the ambiguity combination which can minimise the cost function J in the search space. It is usually a time-consuming process. In order to improve the efficiency of the searching process, Cholesky factorisation is applied in this paper. By using Cholesky factorisation, P_{ZN}^{-1} can be expressed as:

$$P_{ZN}^{-1} = CC^T \qquad (24)$$

where, $C = \begin{bmatrix} c_{11} & 0 & 0 & 0 \\ c_{21} & c_{22} & 0 & 0 \\ c_{31} & c_{32} & c_{33} & 0 \\ c_{41} & c_{42} & c_{43} & c_{44} \end{bmatrix}$

Substituting equation (24) into (22), we have:

$$\begin{aligned} J &= \left(\hat{Z}_N - Z_N\right)^T P_{ZN}^{-1}\left(\hat{Z}_N - Z_N\right) \\ &= \left(\hat{Z}_N - Z_N\right)^T CC^T\left(\hat{Z}_N - Z_N\right) \\ &= f^T f = f_1^2 + f_2^2 + f_3^2 + f_4^2 \end{aligned} \qquad (25)$$

where, $f = C^T\left(\hat{Z}_N - Z_N\right) = \left[f_1 \quad f_2 \quad f_3 \quad f_4\right]$

$$
\begin{aligned}
f_4 &= (\hat{z}_4 - z_4)C_{44} \\
f_3 &= (\hat{z}_4 - z_4)C_{43} + (\hat{z}_3 - z_3)C_{33} \\
f_2 &= (\hat{z}_4 - z_4)C_{42} + (\hat{z}_3 - z_3)C_{32} + (\hat{z}_2 - z_2)C_{22} \\
f_1 &= (\hat{z}_4 - z_4)C_{41} + (\hat{z}_3 - z_3)C_{31} + (\hat{z}_2 - z_2)C_{21} + (\hat{z}_1 - z_1)C_{11}
\end{aligned} \quad (26)
$$

Therefore, J in equation (25) can be calculated by the following iterative procedure:

$$J(k+1) = J(k) + f_k^{\ 2} \quad J(0) = 0 \quad J(4) = J \quad k = 1,2,3,4 \quad (27)$$

Because $f_i^{\ 2} \geq 0$ in equation (26), $J(k)$ keeps increasing along with the increase of index number k. Therefore a fast cutting-off search method is used to reduce the calculation load. This method can be described as:

[1] Give an ambiguity combination from the ambiguity search space;
[2] Calculate equation (27) step by step;
[3] If in the k-th step, $J(k)$ is larger than the threshold, this ambiguity combination can be rejected and the iteration calculation will stop and jump to step (1).
[4] If $J(4)$ is smaller than the threshold, then replacing the threshold using $J(4)$.

Because the calculation load of f_k decreases along with the increase of k, this method can reduce the calculation load of J dramatically.

Once the ambiguity combination $\hat{z}$ that minimises J in equation (25) is found, the initial ambiguity combination can be estimated using the inverse Gaussian transformation:

$$\hat{N}_1 = Z^{-1}\hat{Z}_N \quad (28)$$

3.4 Ambiguity validation

The ambiguity combination which produces the minimum sum of the squared residuals does not necessarily indicate that correct ambiguities are identified. There are a number of factors contributing to this, such as poor system geometry and high measurement noise. Thus, ambiguity validation and evaluation procedure has to be applied to further validate and confirm its correctness. Although integer ambiguities should be determined as early as possible, the reliability of the ambiguity resolution is of paramount importance for GPS-based attitude determination system.

Traditionally, ambiguity validation test procedures have been based on the so-called F-ratio test (Erickson, 1992). When the ratio of the second minimum and the minimum of J in equation (25) is larger than a threshold, the best ambiguity combination which produces the minimum of J is considered as the correct ambiguity set. But in practical application, sometimes the difference between the second minimum and the minimum of J may not be larger enough because of the high noise, therefore it will need a long time to confirm the correct ambiguity although the best ambiguity combination already has been the correct ambiguity set.

In this paper a new validation method by using baseline length and the relationship between primary and secondary ambiguity groups is proposed to validate the estimated ambiguity combination $\hat{N}_1$. By using this method, the correct ambiguity can be quickly and reliably identified. This method includes two steps which can be described as follows:

[1] **Baseline length test:** Substituting the estimated ambiguities into equation (3) and calculating the baseline length. For the correct ambiguity combination, the estimated and the given known baseline length should be consistent with a certain tolerance.

$$\left| L_{estimate} - L_{true} \right| < \delta \quad (29)$$

The numerical value of the baseline length tolerance δ in equation (29) depends on the carrier phase measurements noise, multipath and geometry distribution of satellites in view. In this paper, δ is empirically chosen as 0.02m.

[2] **Primary and secondary ambiguity groups test:** calculating a coarse baseline vector using the estimated primary ambiguity combination; calculating all the float ambiguities by means of substituting the coarse baseline vector into equation (3); calculating the secondary integer ambiguities by means of rounding the float ambiguities to the nearest integer values. For the correct ambiguity combination, the difference of integer and float ambiguities in the secondary should be less than a threshold value “β”, which can be empirically set to 0.02m (about tenth of L1 wavelength).

Only when the estimated ambiguity combination passes the above tests, the correct ambiguity combination is considered identified.

4 Experimental results

A low-cost GPS-based attitude determination system is developed. The hardware of the system mainly includes: PC/104 computer, two Allstar GPS OEM boards, two antennae, battery and other auxiliary accessories. The GPS OEM board can output raw measurements including L1 carrier phase, pseudorange and navigation information at a rate of 1HZ. The algorithms proposed in this paper are implemented in its software which is designed using C++ language. All the available satellites with elevation angle more than 10 degrees are used in the data processing. Considering that usually the ephemeris data can be received whin 30 seconds in cold start mode, the Kalman filtering procedure begins from 30-th second in the software. The integer ambiguity search procedure begins from 80-th second. After this epoch, the ambiguity combination which minimizes cost function (16) can be found at every epoch, but only when it passes the validation test, correct ambiguity combination is considered reliably identified. Once the ambiguities are fixed, the attitude parameters can be estimated by using the equations (5) and (6).

To evaluate the performance of the system, a number of experiments have been conducted. There always more than 7 satellites being tracked during these experiments. The results of ambiguity resolution time test are shown in Tab. 1. It is indicated from Tab. 1 that ambiguity can be fixed whin 3 minutes in a cold start mode.

The attitude solution for a 3m baseline is shown on Fig. 2, 3 and 4, and the attitude solution for a 6m baseline is shown on Fig. 5, 6 and 7. The average and standard deviation of the attitude solution are listed in Tab. 2. From Tab. 2, it can be concluded that the accuracy of the GPS-based attitude determination system becomes higher with the increase of the baseline length.

Tab. 1 List of ambiguity resolution time with different baseline lengths

NO.	GPS date (d/m/y)	GPS time (seconds)	Number of Satellites	Ambiguity Resolution Time (seconds)	Length of baseline (meters)
1	19/04/2002	529971	7	89	2.40
2	19/04/2002	530348	7	137	2.40
3	22/04/2002	203356	8	103	3.00
4	22/04/2002	203706	8	138	3.00
5	22/04/2002	204106	8	95	3.00
6	28/04/2002	128804	7	103	6.00
7	28/04/2002	129104	7	109	6.00
8	28/04/2002	129404	7	97	6.00
9	28/04/2002	129704	7	119	6.00

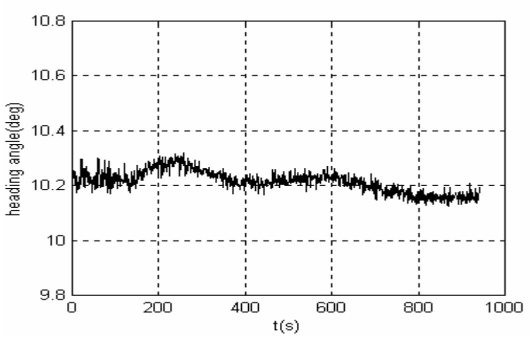


Fig. 2 Heading angle (3m baseline)

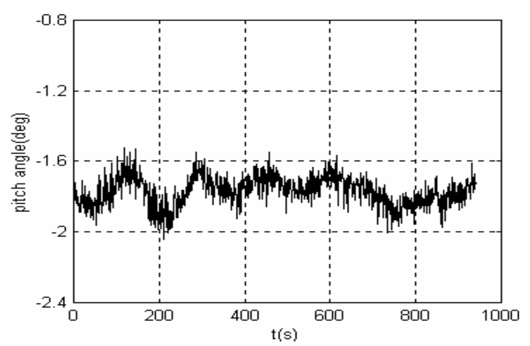


Fig. 3 Pitch angle (3m baseline)

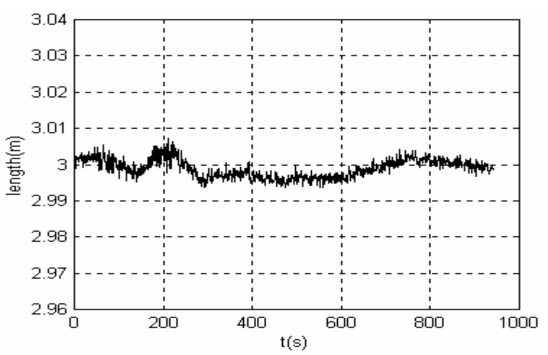


Fig. 4 Baseline length (3m baseline)

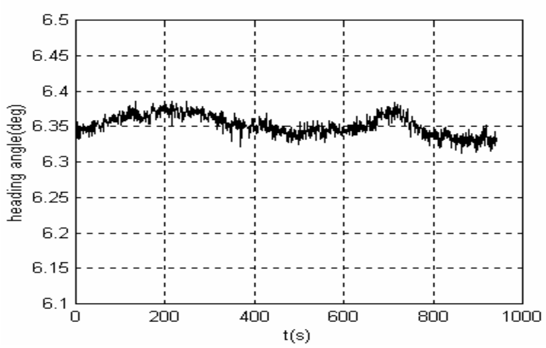


Fig. 5 Heading angle(6m baseline)

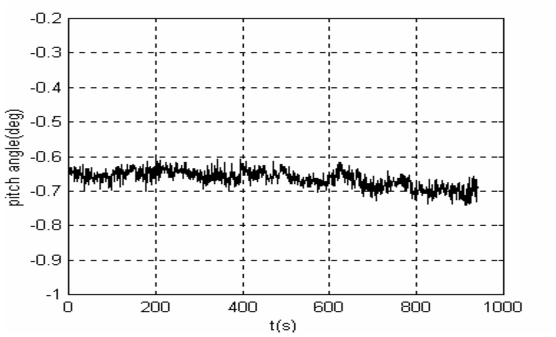


Fig. 6 Pitch angle (6m baseline)

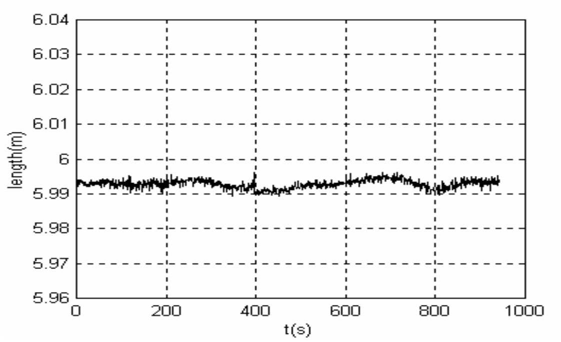


Fig. 7 Baseline length (6m baseline)

Tab. 2 Average and standard deviation of attitude solutions with different baseline lengths

	3m		6m	
	Average	Standard deviation	Average	Standard deviation
Baseline (meters)	2.9989	0.0025	5.9935	0.0027
Heading angle (degrees)	10.2114	0.045	6.3519	0.0168
Pitch angle (degrees)	-1.7684	0.0833	-0.6550	0.0314

5 Conclusions

This paper presents an algorithm for fixing the integer ambiguity in attitude determination. The method can not only improve the computation efficiency and reduce the time for ambiguity resolution, but also improve the reliability of the ambiguity solution.

A number of experiments are carried out to evaluate the effectiveness of the proposed algorithm, the results show that correct ambiguities in cold start mode can be identified within 3 minutes and there is no evident relationship between the ambiguity resolution time and the length of baseline. The results also show that heading angle of the GPS-based attitude determination system can achieve an accuracy of 0.045deg (RMS) for a 3m baseline and a higher accuracy for longer baselines.

Acknowledgements

This research is partially supported by the Cooperative Research Centre for microTechnology, Australia. A/Prof Kefei Zhang is supported through an Australia Research Council Project (LP0455170).

References

Crassidis JL, Markley FL, Lightsey EG (1999): ***Global Positioning System Integer ambiguity Resolution without attitude knowledge,*** J. Guidance Control and Dynamics 22(2): 212-218.

Gao Y., Mclellau J F. and Schleppe J B (1996): ***An optimised GPS Carrier Phase Ambiguity Search Method Focusing on Speed and Reliability,*** IEEE Aerospace and Electronic System Magazine. 11(12): 22-26**.**

Erickson C. (1992): ***An analysis of Ambiguity Resolution Techniques for Rapid Static GPS Surveys using Single Frequency Data,*** Proceeding of ION GPS-92, 453-462.

Juang J. (2003): ***GPS Integer Ambiguity Resolution Based on Eigen-Decomposition,*** Proceedings of ION GPS-2003, 775-782, Portland.

Kee C., Kim D, and Jang J. (2003): ***A New Approach for Fast and Reliable Ambiguity Search Technique,*** Proceedings of ION GPS-2003, 2112-2119, Portland.

Mohamed A H, Schwarz K P (1998): ***A Simple and Economical Algorithm for GPS Ambiguity Resolution on the Fly Using a Whitening Filter***, Navigation, 45(3): 221-231.

Quinn, P. G. (1993): ***Instantaneous GPS Attitude Determination,*** Proceedings of ION GPS-93, 603-615, Alexandria, VA: Institute of Navigation.

Sutton E. (1997): ***Optimal Search Space Identification for Instantaneous Integer Cycle Ambiguity Resolution,*** Proceedings of ION GPS-97, 313-322.

Sutton E. (2002): ***Integer Cycle Ambiguity Resolution under Conditions of Low Satellites Visibility,*** ION GPS-2002, 91-98.

Xu G. (2002): ***A General Criterion of Integer Ambiguity search,*** Journal of Global Positioning System, 1(2): 122-1.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 215-222

Journal of Global Positioning Systems
2005
© CPGPS

Evaluation of the pseudorange performance by using software GPS receiver

Shun-Ichiro Kondo, Nobuaki Kubo and Akio Yasuda

2-1-6 Etchujima Koto-ku Tokyo Japan
e-mail: shun@denshi.e.kaiyodai.ac.jp Tel: +81-3-5245-7376,Fax:+81-3-5245-7376

Received: 15 November 2004 / Accepted: 12 July 2005

Abstract. An algorithm is developed to process IF signal data from a GPS RF front-end module, which consists of a down converter and an ADC. The down converter converts the signal from RF to IF, the ADC samples the IF signal. All the other processing including signal acquisition, tracking, data decoding and solving position are implemented in software using base-band signal processing techniques. The local C/A codes and carrier replica signal are pre-generated, stored in memory, and used respectively during signal acquisition and tracking. In order to evaluate the algorithm, this paper demonstrates standalone positioning using the measured pseudoranges of which accuracy depends on DLL parameter of the correlator of early/late spacing. This paper presents the explanation and evaluation of the algorithm.

Key words: Software GPS receiver, signal acquisition, signal tracking, accuracy of pseudorange

1 Introduction

In a conventional GPS receiver, a GPS signal acquisition and tracking channel is made up of both hardware and software (Misra and Enge, 2001). Typical hardware components include mixers, correlators, NCO, and code generator for a code tracking loop or sine/cosine tables for a carrier lock loop. Software equips discriminators, tracking loop filters, navigation data bit/frame sync and ephemeris demodulator, and implements various reading from and writing to uniquely addressed registers. The software GPS receiver has advantage of flexibility over such conventional receivers.

In the software GPS receiver, down-converter converts the signal from RF to IF, ADC samples the IF signal, and all the other processing are implemented in software. Thus the accuracy and function are controllable according to the algorithms and parameters**.** The signal quality and/or multipath effect can be estimated from the correlation values of the programmable code tracking Loop (Pany, Eissfeller and Winkel, 2003).

The algorithm is developed on MATLAB to investigate the effect of GPS signal processing parameters on positioning in the present paper. The algorithm itself is also evaluated.

2 Software GPS receiver

The feature of the software GPS receiver is to use an analog-to-digital converter (ADC) which converts the input signal into digital data at the earliest possible stage in the receiver (Tsui, 2000). Fig.1 illustrates a general structure of software GPS receiver (Pany, Moon, Irsigler, Eissfeller, and Furlinger, 2003). The signals transmitted from the GPS satellites are received at the antenna. Through a down converter, the input signal is amplified to a proper amplitude and the frequency is converted to a desired output frequency. An ADC is used to digitize the input signal. Acquisition means to find the signal of a certain satellite. The tracking program tracks the code phase and the carrier phase, and finds the phase transition of the navigation data. The code phase is used to obtain the pseudorange. The navigation data can be obtained from the phase transition of the data. Ephemeris data can be obtained from the navigation and are used to obtain the satellite position. Finally, the user position, velocity, and time can be deduced.

The prototype software receiver developed in this paper is comprised of RF front-end module and signal processing program on MATLAB.

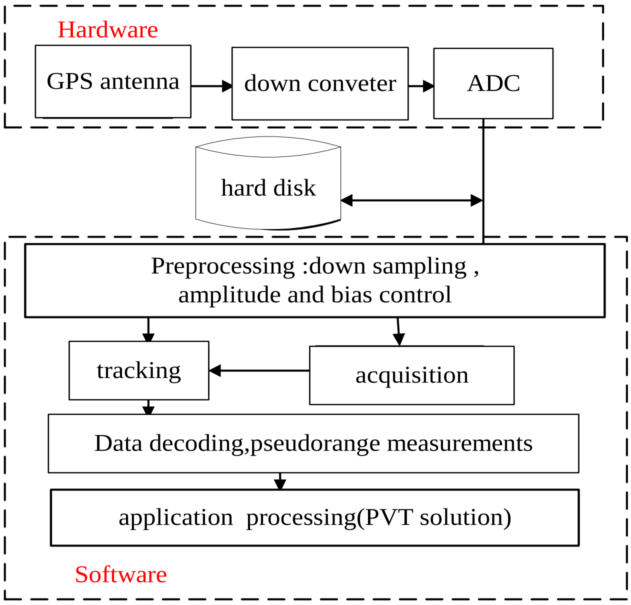


Fig.1 General structure of software GPS receiver

3 GPS RF front-end module

GPS Signal Tap (manufactured by Accord Software and System Private Limited) is used as a GPS L1 RF front-end. Table.1 summarizes the characteristics of the GPS Signal Tap. The L1 GPS signal is down converted to the intermediate frequency of 15.42MHz in two stages of a down converter. The IF is sampled by the ADC at a frequency of user's choice, which can be selected from 2MHz to 20MHz in the GUI application. The ADC output is stored once in the on-board SDRAM of the Signal Buffer and transferred to host PC via USB for post processing. Since the capacity for the Signal Buffer is limited to 64MB, the data duration depends on the sampling frequency. Fig. 2 shows the picture of the GPS Signal tap.

Table.1 Characteristics of the GPS Signal Tap

Frequency	GPS L1 at 1.57542GHz
GPS code	C/A
IF	15.42MHz
Sampling rate	From 2MHz to 20MHz in an increment of 1kHz
Signal gain (Second IF stage)	120dB
Band Width (Second IF stage)	8MHz

Fig.2 Picture of the GPS Signal Tap

4 Acquisition

Acquisition is comprised of a two-dimensional search process that determines a coarse Doppler offset of the carrier and the beginning point of the C/A code. Conventional GPS receivers generally use a serial search method for acquisition (Kaplan, 1996). The present method searches correlation peak by multiplying the code with delay within 1023chips with frequency shifted carrier signal within ± 10 kHz Doppler band in coarse increments of delay and shift. It requires a large processing time to search the correct Doppler and code shift. To expedite the process, parallel search methods have been suggested. Here, algorithms based on fast Fourier transform (FFT) and inverted FFT (IFFT) techniques are used to search a maximum correlation power of signals between the locally generated code/carrier signal and the input IF signal. A circular convolution method is utilized for efficient implementation. In this section, two different acquisition methods, conventional approach and FFT approach, will be described.

The conventional approach performs signal acquisition in time domain. A non-coherent correlator in time domain, shown in Fig. 3, is used since the phase of received signal is random (Johanson, Mollaei, Thor and Uusitalo, 1998). The correlation is approximated by the discrete sum.

$$R^2[m] = \sum_{j=0}^{K-1}\left(\left[\sum_{n=jNL}^{(j+1)NL-1} x[n]\cdot CA[n+m]\cdot\cos[\Omega n]\right]^2\right) + \sum_{j=0}^{K-1}\left(\left[\sum_{n=jNL}^{(j+1)NL} x[n]\cdot CA[n+m]\cdot\sin[\Omega n]\right]^2\right), \quad (1)$$

where Ω is the radian frequency of the IF signal. The digital IF, $x[n]$ is multiplied first by the replicated C/A code, $CA[n+m]$. Here n represents the n^{th} sample and m represents the number of phase shifted samples of the replicated C/A code. L is the sample length of one C/A code period. n and m are the numbers of 0 to $L-1$. K is the non-coherent integration time. After the code removal, the in-phase(I) and quadrature-phase(Q) components are generated. The I and Q components are accumulated for one or more code periods, N . The accumulated sum is squared. Next, K correlations are accumulated to produce an averaged correlation point. If the correlation peak is lager than a certain threshold, it is assumed that the satellite is acquired.

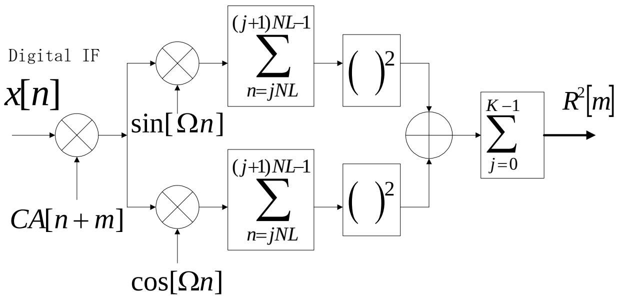


Fig.3 Non-coherent correlator in time domains

The conventional approach can also be preformed in circular convolution. The received data are correlated with the replica code by a circularly shifting the replica code. This may be expressed as (Johanson, Mollaei, Thor, and Uusitalo, 1998)

$$R[m]=\sum_{n=0}^{L-1} x[n]\otimes CA[(n+m)_L] \tag{2}$$

The FFT approach performs circular convolution in the frequency domain. The discrete Fourier transform (DFT) and its inverse is used to calculate $R[m]$:

$$R[m]=\sum_{n=0}^{L-1} x[n]\otimes CA[m+n]$$
$$=F^{-1}\left(F\left(x[n]\right)\cdot F\left(CA[n]\right)^{*}\right),$$
(3)

where F and F^{-1} denote FFT and inverse FFT. A non-coherent correlator in frequency domain can be adapted to acquisition of GPS signals as shown in Fig.4 (Johanson, Mollaei, Thor and Uusitalo, 1998). Here the input signal is mixed to base band and the I and Q components are used as the real and imaginary inputs when calculating the DFT. The result is multiplied by the complex conjugate of DFT of the C/A code. The circular convolution is obtained by taking the magnitude of inverse DFT. The fast Fourier transform algorithm is used to implement the DFT and inverse DFT; hence this acquisition method may be called FFT approach.

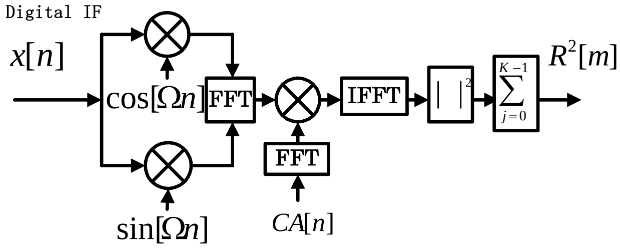


Fig.4 Non-coherent correlator in frequency domain

Fig. 5, 6, and 7 show the results of acquisition using FFT approach (IF: 4.5MHz, sampling frequency: 20MHz, non-coherent integration time: 10ms). Fig.5 shows the correlation matrix for SV3. After the circular correlation, the beginning of the C/A code of SV3 is found to be 18236 in Fig. 6. The correlation power at n = 18236, for 21 frequency components of SV 3 every 1 kHz, centered at 4.58MHz, is shown in Fig. 7. The highest component at 4.575MHz corresponds to real frequency of the IF signal.

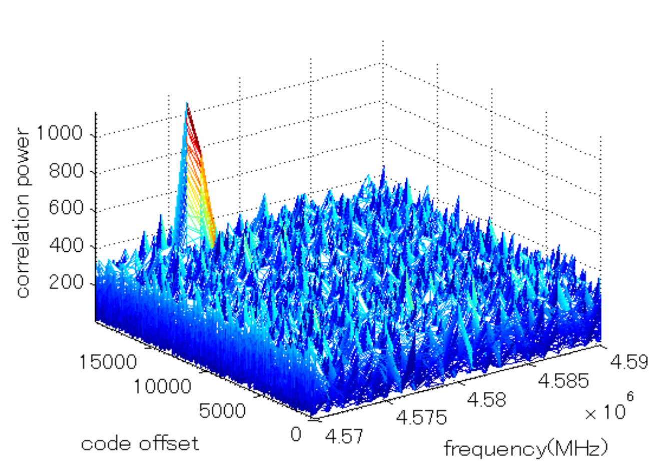


Fig.5 Correlation matrix for SV3

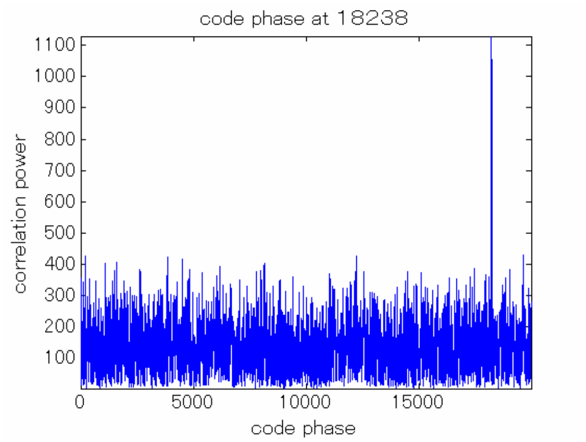


Fig.6 Beginning of C/A code of SV3

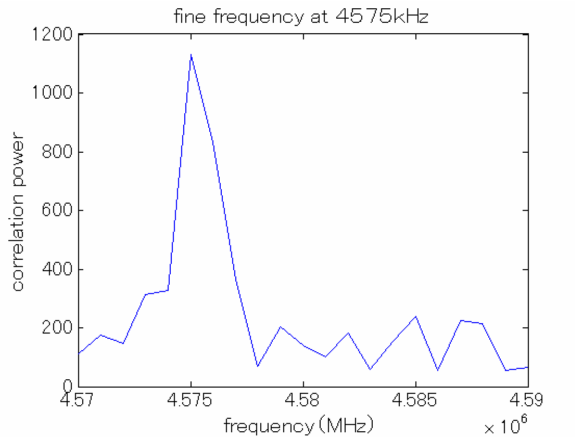


Fig.7 Frequency component of SV3

5 Tracking

Once the signal is acquired, it must be tracked to obtain the navigation data. The tracking program uses two parameters obtained from the acquisition process: the beginning point of C/A code and Doppler offset of carrier. Two loops are needed to track one GPS satellite signal. One loop is often referred to as code loop, which tracks the C/A code. The other one is the carrier loop, which tracks the carrier frequency of the down-converted input signal. These two loops must be coupled together as shown in Fig. 8 (Tsui, 2000).

The code loop uses three locally generated C/A code to track the C/A code of the input signal. The three locally generated codes are usually used: a prompt, an early and a late replica. The early and the late replica codes are shifted a few samples in early and delay directions, respectively. The prompt code is applied to the digitized input signal and strips the C/A code from the input signal. The output will be a continuous wave (cw) signal with phase transition caused only by the navigation data. The regenerated carrier signal in the carrier loop is applied to the digitized input signal to strip the carrier from the signal. The output is a signal with only a C/A code and no carrier frequency, which is introduced to the code loop.

5.1 Code tracking

The code loop is known as a delay lock loop (DLL) (Misra and Enge, 2001). The prompt code is to be matched to the beginning of the C/A code in the input signal. The correlation outputs from three codes can be used to determine accurately the beginning of the C/A code in the input signal. This information is used to adjust the initial phase of locally generated prompt code to match the code phase of input signal better. Fig.9 shows the correlation between input C/A code and local replica C/A one. Fig.10 shows the example of the averaged correlation of the SV25 signal for 200ms.

The discriminator algorithm is used to calculate the code phase error. The discriminator outputs signal r is given as,

$$r = \frac{y_l}{y_e} \tag{4}$$

where y_l and y_e are the correlation powers associated with the late code and early code, respectively. The time distance x from the peak of the correlation powers can be written as

$$x = \frac{(1-r)(1-d)}{(1+r)} \tag{5}$$

where d is the time distance from prompt code to early or late code. If the correlator spacing is 1 chip (~20 samples at the sampling frequency of 20.0MHz), then d is given by 0.51151(10$\times$50/977.5ns). The ratio r is corresponds to the degree to which the beginning point of the C/A code is shifted, the sampling interval is 50ns (1/20.0MHz) and the one chip duration is 977.5ns (1/1.023MHz). If the maximum misalignment occurs within a sampling interval, then x has a value within the range of $\pm$0.026chip (=25/977.5). If x is greater than 0.026 chips, then the local codes should be shifted to the left. Or, if x is smaller than –0.026 chips, then the local codes should be shifted to the right. The logic for the code shift is based on the value of x, and can be expressed as follows:

If x < -0.026chip

The local codes is shifted to the right

Else if x > 0.026chip

The local codes is shifted to the left

Else

The local codes isn't shifted

End

Once x is calculated using Eq. (5), the code phase deduced from x is used to determine the pseudorange. The accuracy of the code phase has a direct effect on positioning accuracy. The measurement noise can be significantly reduced by a simple averaging. In this paper, r is averaged over 10 ms for the noise reduction.

5.2 Carrier tracking

After code tracking loop determines code phase, the carrier tracking loop is used to demodulate navigation data from the IF input signal. The loop filter of the carrier loop can be designed in the form of Phase Lock Loop (PLL) or Frequency Lock Loop (FLL). PLL tracks the phase error between two carrier signals and FLL tracks frequency errors. Generally, PLL provides a higher tracking performance than FLL, but is more sensitive to noise and dynamics. On the other hand, FLL is less sensitive to dynamics but results in lower accuracy in general. In this paper, PLL based algorithm is implemented for fine carrier tracking.

The loop filter (Tsui, 2000, Tsui, Stockmaster and Akos, 1997) is given by

$$F(z) = \frac{(C_1 + C_2) - C_1 z^{-1}}{1 - z^{-1}} \tag{6}$$

where

$$C_1 = \frac{8\varsigma\varpi_n t_s}{K \cdot [4 + 4\varsigma\varpi_n t_s + (\varsigma\varpi_n t_s)^2]} \tag{7}$$

$$C_2 = \frac{4(\varsigma\varpi_n t_s)^2}{K \cdot [4 + 4\varsigma\varpi_n t_s + (\varsigma\varpi_n t_s)^2]} \tag{8}$$

K :loop gain (400π)

ς :damping factor(0.707)

ϖ_n :natural frequency(37.7143)

t_s :sampling interval(1^{-3})

By substituting the values for the parameters, the implemented discrete loop filter is given by

$$F(z) = \frac{0.000209 - 0.000184 z^{-1}}{1 - z^{-1}} \tag{9}$$

In this paper, atan2() function is used for the discriminator function. Fig.11 shows tracking outputs. The amplitude changes temporally according to navigation data bits. After carrier tracking is completed, the navigation data are extracted from the in-phase signals. It is expressed in phase difference π, which is transformed into the navigation data as ± 1 values.

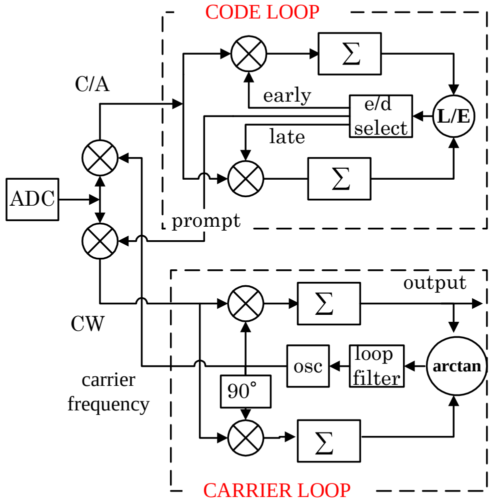


Fig.8 Code and carrier tracking loops

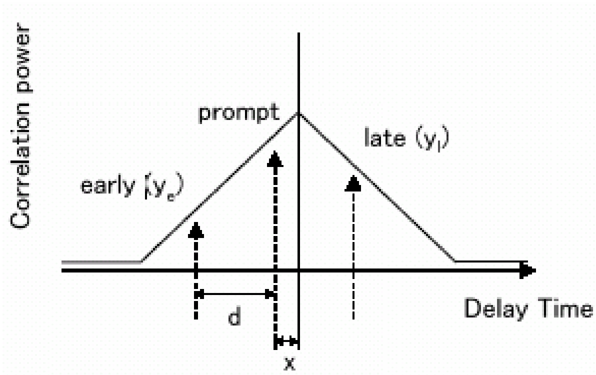


Fig.9 Correlation between the input and the replica C/A signals

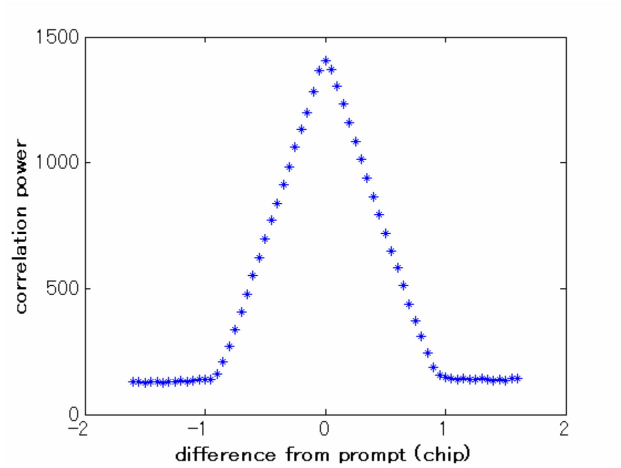


Fig.10 Correlation of SV25 averaging for 200 ms

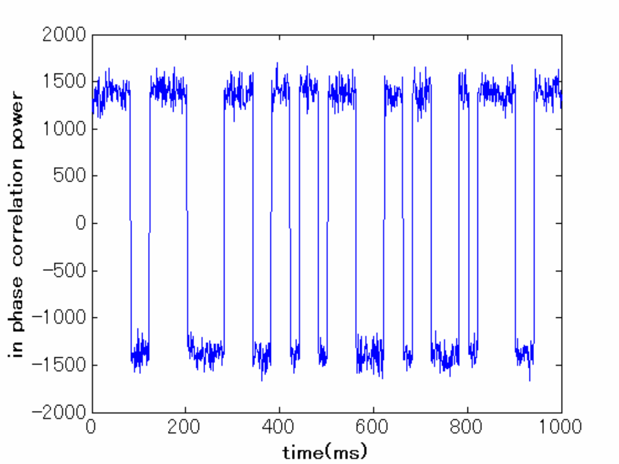


Fig.11 Tracking outputs of SV25

6. Sub frame matching and pseudorange measurements

After the tracking results are converted into navigation data, the next step is to find the subframes (Tsui, 2000) in these data. A subframe will start with the preamble of a pattern (10001011) in the first word. The second word is HOW (the hand of word); bits 1-17 are TOW (time of week) and bits 20-22 are the subframe ID. However, the polarities of the words in a frame may change. Therefore, one should perform correlation on only one word at a time. The code to match the preamble can be written as (1 –1 –1 –1 1 –1 1 1). Since the polarity of the word is not known, the matched results can be ± 8. One can repeat this method to find, 300data points (1 subframe) later there should be another preamble match. If a match is not found, the first match is not a preamble. One can be repeat this method to find the beginning of several subframes.

The pseudorange measurements are performed after obtaining the beginning of the subframes. However, in collecting the digitized data, there is no absolute time reference and the only time reference is the sampling frequency. In other word, the time of the reception can't be obtained. As a result, the pseudorange can be measured only in a relative way (Tsui, 2000) as shown in Fig. 12. In this figure, the points represent individual input digitized data (beginning of the C/A code obtained from tracking program). The relative pseudorange is the distance (or time) between two reference points. The beginning point of subframe is used as a reference point. All the beginning points of subframe from different satellites are transmitted at the same time expect for the clock correction terms of each satellite. Since the beginnings of subframe from different satellites are received at different times, this difference time represents the time (or distance) difference from the satellite to the receiver. Therefore, it represents the relative pseudorange.

Pseudorange measurements (Tsui, 2000) are performed in the following steps.

1. The beginning points of subframe in terms of the actual digitized input data points are found.
2. The average point of all the beginning points is calculated. Time differences from average point to each beginning point are calculated respectively.
3. The transit time of the beginning points of subrame are calculated by assuming that transit time of the average point is 73ms. And pseudoranges of each satellite are measured.
4. The pseudoranges of each satellite change with time by adding the C/A code shift for one second.

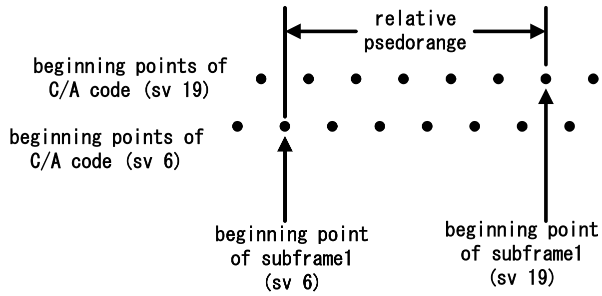


Fig.12 Relative pseudorange

7. The results

In our test, the signal processing was repeated with changing the correlator spacing into 0.1, 0.2 and 1chip. Table 2 shows the parameters of collecting the data. In the case of 20MHz in sampling frequency, maximum collection time is limited to 24 seconds due to the GPS Signal Tap's capacity. Thus the standalone positioning was performed using ephemeris data collected by another GPS receiver. Fig.13 shows the horizontal errors for the cases of 0.1, 0.2 and 1 chip of early/late spacing of the correlator in standalone positioning under light multipath environment. Compared with the horizontal errors of 1 chip, the horizontal errors of 0.1 and 0.2 chip are reduced. It is considered that the correlator could track the C/A code accurately in the case of 0.1 and 0.2 chip.

Fig.14 shows two horizontal errors (0.2 and 1 chip) under heavy multipath environment. Compared with the results under light multipath environment, the horizontal errors are larger and the positions deviate more. The correlator couldn't track enough number of satellites for the positioning in the case of 0.1 chip spacing. The reason is that the correlation powers were very weak and the shape around correlation peaks of the C/A code was contaminated by mulithpath effect and gentle (Misra and Enge, 2001). Thus narrow correlator (0.1 chip) couldn't acquire the peak well.

The example of the peak of the correlation averaged for 200 ms (801~1000ms) under heavy mulitipath condition is shown in Fig.15. As shown in this figure, a triangle shape of the correlation power is very gentle, and the 0.1 chip correlator can't acquire the peak. Therefore, it can't be converted into the navigation data shown in Fig.16.

Table.2 Prameters of collecting the data

IF	15.42MHz
Sampling Frequency	20MHz
Collection Time	24seconds
Band Width	8MHz

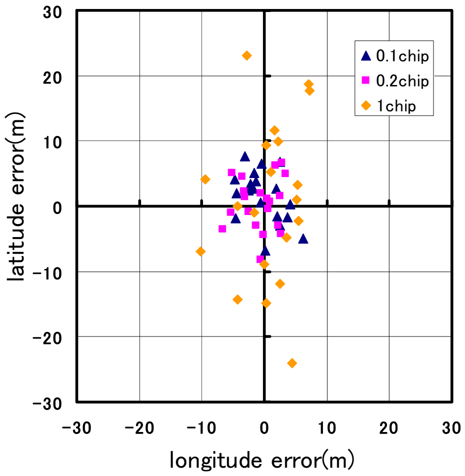


Fig.13 Horizontal errors (light multipath)

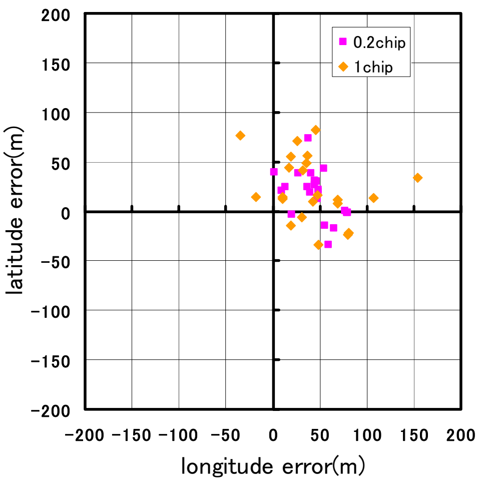


Fig .14 Horizontal errors (heavy mulitpath)

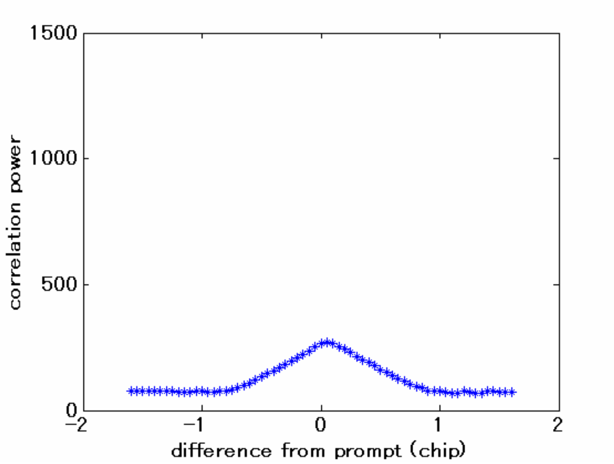


Fig.15 Correlation of SV11

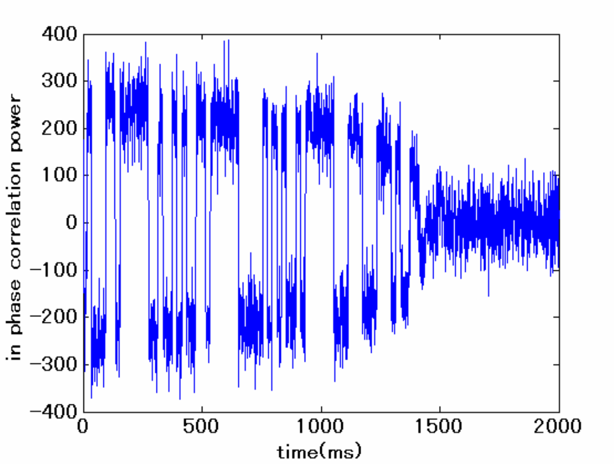


Fig.16 Tracking outputs of SV11 (0.1chip correlator)

8 Conclusions

A prototype PC based and IF sampling software GPS receiver has been developed successfully. This paper presented the performance of our software GPS from a point of accuracy. It also evaluated the relationship between the signal quality and the positioning errors in our post processing. It is found that much more information in the signal processing can be obtained by using software GPS receiver compared with the raw outputs of conventional GPS receiver. As shown in these results, our software GPS doesn't work as well as conventional GPS receivers. One of the reasons is that the carrier aided code loop can't be used in this algorithm. The future work is firstly to improve the accuracy up to the level of the conventional GPS receivers. The second future work is to reveal the balance between the tracking threshold and the accuracy of the pseudorange under heavy multipath condition.

References

Wu F.; Kubo N.; Yasuda A.; HOJO Harumasa. (2003): ***Development of a Prototyoing Platform for Software GPS Receiver.*** Journal 111th of Japan Institute of Navigation

Ko S.; Won J; and Lee S. (2003): ***FFT and PLL Based GPS Signal Processing for Software GPS Receiver.*** Wuhan University Journal of Natural Science

Pany T.; Eissfeller B.; Winkel J. (2003): ***Tracking of High Bandwidth GPS/Galilieo Signals with a Low Sample Rate Software Receiver***. Proc. ENC-GNSS2003, Graz.

Tsui J. B.Y. (2000): ***Fundamentals of Global Positioning System Receiver – A software Approach.*** A John Wiley & Sons, Inc. Publication

Pany T.; Moon T. S. W.; Irsigler M.; Eissfeller B.; Furlinger K. (2003): ***Performance Assement of an Under Sampling SWC Receiver for Simulated High-bandwidth GPS/Galileo Signals and Real signals.*** proceeding of 16th International Technical meeting of the Satellite Division of The Institute of Navigation, Portland, Oregon, USA, pp.103-106

Kaplan E D. (1996): ***Understanding GPS: Princcples and application,.*** Boston:Arech House,1996.

Misra P.; Engse P. (2001): ***GRPBAL POSITIONING SYSTEM, Signal Measurements, and performance.***. Ganga-Jamuna Press

Tsui J. B.Y.; Stockmaster.; Akos D M. (1997): ***Block Adjustment of synchronizing Signal (BASS) for Global Positioning System (GPS) Receiver Signal Processing,*** Proceeding of ION GPS 97.

Johanson F.; Mollaei R.; Thor J.; Uusitalo J. (1998): ***GPS Satellite Signal Acquisition and Tracking***, Division of Signal Processing, Luleca University of Technology, S-971 87 Luleca, weden, Undergraduate Projects 1998

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 223-229

Journal of Global Positioning Systems

Real Time Quality Assessment for CORS Networks

Simon Fuller, Philip Collier, Allison Kealy

Department of Geomatics, The University of Melbourne, Melbourne, Australia
e-mail: s.fuller2@pgrad.unimelb.edu.au Tel: + 61 3 8344 4509

Received: 15 November 2004 / Accepted: 13 July 2005

Abstract. The growing use of real time high accuracy Global Positioning System (GPS) techniques has resulted in an increase in the number of critical decisions made on the basis of a GPS derived position. When making these decisions mobile users require assurance that the GPS position quality meets their requirements. Providers of Continually Operating Reference Stations (CORS), whom mobile users are generally reliant upon, must also be able to assure users that their data meets agreed quality standards. Unfortunately, the realistic and reliable description of position and data quality is an area in which GPS has traditionally been weak. Research being undertaken as part of the Cooperative Research Centre for Spatial Information (CRC-SI) is attempting to address this problem by assessing and reporting on the quality of raw GPS observations in real time. This paper examines a number of existing approaches to assessing the quality of raw GPS observations and presents a conceptual architecture for the development of a real time quality control system.

Key words: GPS, Quality, Real Time, Stochastic Modelling, CORS

1 Introduction

The increasing usage of high accuracy real time GPS positioning in a wide range of applications has resulted in a proportional rise in the number of critical decisions made on the basis of GPS positions. These decisions may be critical from a safety-of-life, financial, or environmental perspective. In making these decisions GPS users must be capable of determining if the quality of the position meets their requirements. Furthermore, they must be confident that the indicators of position quality that their decision is based on are realistic and reliable in all conditions, at any time.

To obtain high accuracy real time GPS positions, mobile users rely heavily on information from external sources, be it from a local GPS basestation, a regional CORS network, or a global correction service. Thus the quality of the mobile user's position is intrinsically linked to the quality of the external data. Mobile users must be assured that the information provided to them is of sufficient quality to meet their requirements. It follows that suppliers of GPS data products, e.g. CORS providers, must be able to deliver quality information to the mobile user in real time.

Research being undertaken as part of the positioning program of the CRC-SI is attempting to address many of the issues associated with the real time assessment of CORS and mobile user positioning quality. The aim of this research is to develop real time procedures for CORS networks and mobile users that will improve the reliability of the mobile user's position and provide a realistic assessment of the position quality. To accomplish this an understanding of existing approaches to quality control and the ability of these approaches to be adapted to real time operation is required. This paper presents a review of the current methods for assessing the quality of CORS and mobile user data and positions, in conjunction with an analysis of the potential of these methods to operate in real time. Finally a conceptual architecture for the real time quality control of CORS networks and mobile users is proposed.

2 Quality control for CORS networks and mobile users

The positioning accuracy and quality achievable by GPS is dependent on the raw data quality and the processing algorithm chosen. The quality control of GPS observations falls into two parallel categories – the validation and description of the raw data quality, independent of its future application, and the quality control undertaken as part of the processing algorithm (Brown et al., 2003). Given the wide range of processing algorithms available the quality control processes employed by these algorithms is not of particular concern at this stage. Suffice it to say that the quality control and end results of the chosen processing algorithm will be dependent on the provision of the high quality observation data and an accurate stochastic model, both of which are a direct outcome of quality control of raw observation data.

The methods and procedures for the validation and description of raw data quality are generally independent of the processing algorithm chosen. The aspects of raw data quality control considered here include data completeness, the detection and repair of cycle slips and receiver clock jumps, and the description of raw data quality in the form of stochastic models.

2.1 Data completeness

The most basic form of raw data quality control consists of statistics that describe the amount and completeness of the data collected by a GPS receiver. The consequences of ignoring data completeness in a quality control process can be severe, leading to difficulty in detecting outliers and cycle slips, increased time to resolve ambiguities, the introduction of multiple ambiguities, a weaker solution due to limited data availability, and in the worst case an inability to compute a solution (Brown et al., 2003). Three aspects of data completeness are generally considered; data gaps - being epochs with incomplete or no observations; missing epochs - whereby observations are not recorded for a satellite that is visible; and the availability of sufficient ephemeris information for a satellite. Statistics on data completeness, when analysed over extended time periods, can be useful for determining problems with receiver hardware and software (Brown et al., 2003), site-specific problems (Brown et al., 2003, Jonkman and de Jong, 2000a), and abnormalities in the satellite constellation or ephemeris information (Jonkman and de Jong, 2000a). Current quality control software packages such as GQC (Brown et al., 2003) and TEQC (Estey and Meertens, 1999) operate in a post-processing mode and are well suited to this sort of task.

From a real time quality control (RT-QC) perspective data gaps, missing epochs, satellite constellation problems and so forth need to be closely monitored and appropriate action taken to notify users of any problems that may impact on the quality of their position solution. Additionally, data gaps and missing epochs are likely to have a detrimental impact on the ability of any real time algorithms for the detection of cycle slips, or the generation of stochastic models, to carry out their assigned tasks..

2.2 Systematic biases in observation data

High accuracy GPS positioning is dependent upon the identification and removal of the main error sources that impact upon the observation quality. In relation to the quality control of raw data, receiver clock jumps, cycle slips, and quasi-random (e.g. multipath, diffraction, ionospheric scintillation etc.) effects are the main error sources that can degrade observation quality. The impact of quasi-random errors are not considered here but are dealt with briefly in the section describing stochastic modelling. However, the influence of quasi-random errors does hamper cycle slip detection, mainly due to the fact that their influence on the phase observations is not limited to an integer number of cycles (Kim and Langley, 2001). The treatment of true systematic errors in the RT-QC context is discussed in the following sections.

2.2.1 Receiver clock jumps

GPS receivers align themselves with GPS time using a variety of techniques. Some receivers constantly synchronise their clock with GPS time (so called "Clock Steering") whilst others allow their clock to drift and periodically introduce corrections of approximately 1 millisecond to keep the clock close to GPS time (Fig. 1). Other receivers allow the clock to drift unchecked and simply keep track of the bias and bias rate of change (Rizos, 1999, Gurtner, 1999, Fraser, 2004).

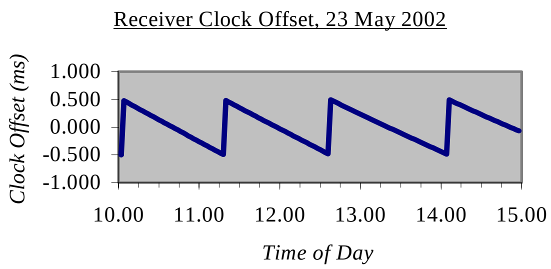


Fig 1: Receiver Clock Jumps

Of concern from a RT-QC perspective is the second technique (illustrated in Fig. 1), whereby clock jumps are introduced into the raw observations. These jumps produce a systematic bias in the undifferenced code and phase observations, as shown in the following equation:

$$\Phi(t+\Delta) = \Phi(t) + \dot{\Phi}\cdot\Delta = \Phi(t) + \dot{\rho}\cdot\Delta - c\cdot\Delta \qquad (1)$$

where Φ represents the carrier phase (or pseudorange) observation and $\dot{\Phi}$ its rate of change with respect to time; Δ represents the clock jump; $\dot{\rho}$ is the satellite dependent geometric range rate; and c is the speed of light in a vacuum. The clock jumps themselves are quite small (less than or equal to 1 millisecond) but they have two distinct effects on the code and phase observables. The term $c \cdot \Delta$ represents a constant receiver dependent effect on the geometric range whilst the term $\dot{\rho} \cdot \Delta$ represents the contribution of the satellite dependent geometric range rate at the time of the clock jump (Kim and Langley, 2001). The first of these terms ($c \cdot \Delta$) is removed during subsequent single or double difference processing. The latter term ($\dot{\rho} \cdot \Delta$) does not cancel during differencing, as it is dependent on a particular satellite-receiver combination.

Thus the term $\dot{\rho} \cdot \Delta$ introduces a systematic bias into the geometric range rate. The size of the bias is dependent on the particular geometric range rate. Assuming a maximum possible rate of 900m/s, a one-millisecond jump could potentially introduce 0.9m of error into the geometric range. From a RT-QC perspective it is crucial that these effects are estimated and removed in real time. Without correcting for such an effect it may be difficult to detect and repair cycle slips, estimate an accurate stochastic model, and undertake subsequent quality control (e.g. during the processing algorithm).

2.2.2 Cycle slip detection and repair

Cycle slips are discontinuities of an integer number of cycles in the carrier phase observations caused by a loss of lock in the receiver's carrier tracking loops. Hofmann-Wellenhof et al. (1992) describe three potential causes for cycle slips. Firstly, the most likely cause of cycle slips are physical obstructions to the satellite signal due to natural or man-made features (e.g. buildings, trees, bridges etc.). Secondly, low signal to noise ratios (SNR) due to ionospheric conditions, multipath, rapid changes in receiver position, or low satellite elevation can produce cycle slips. Finally, failures in the receiver software or malfunctioning satellite oscillators may cause cycle slips, however such incidents are rare.

To take advantage of the superior measurement precision of the phase observables, cycle slips must be removed from the phase data before further processing can occur. This process involves detecting the location of the cycle slip (in time), determining the number of L1 and/or L2 cycles that comprise the slip, and then correcting all phase observations of the affected satellite subsequent to the time of the cycle slip (Kim and Langley, 2001, Hofmann-Wellenhof et al., 1992).

The focus on Real Time Kinematic (RTK) positioning in recent times has moved the detection and repair of cycle slips, traditionally a post-processed activity, into the real-time domain. RTK positioning is dependent on the resolution of the integer ambiguities, a process greatly aided by the presence of clean, cycle slip free data. The push for instantaneous ambiguity resolution has lead to the development of real-time algorithms for the detection and repair of cycle slips.

One such algorithm is the instantaneous cycle slip correction technique proposed by Kim and Langley (2001). This algorithm utilises the triple difference (TD) observables of the carrier phases in conjunction with Doppler and code observables. TD observations are generally free of the majority of GPS biases, such as receiver and satellite clock offsets, integer ambiguities, atmospheric effects, multipath, and satellite orbits. Thus, the size of the remaining biases and noise should be less than a few centimetres, provided that the observation interval is relatively short. Cycle slips would then be evident in the TD observations as large spikes, several orders of magnitude larger than the mean bias and noise. These assumptions may not hold in all cases, for example severe ionospheric disturbances, very long baselines, or rapid variations in the receiver position may lead to the triple difference biases and noise exceeding the L1 and L2 wavelengths, without cycle slips being present. In such situations the observation interval can be reduced to a level such that the biases and noise exhibited by the TDs are once again at the centimetre level and therefore, useful in detecting and repairing cycle slips (Kim and Langley, 2001).

Cycle slip candidates are obtained by examining the mean and variance of the predicted TD residuals (being the difference between the observed TDs and the computed TDs). If dual frequency carrier phase observations are available the number of candidates can be reduced through the use of TDs formed from the geometry free linear combination observations. Following identification of the cycle slip candidates a least squares estimation is carried out to determine the two candidates (best and second best) that minimise the least squares residuals. The statistical likelihood of these two candidates is assessed and if they are considered significantly different then the best candidate is accepted and the slip is repaired. In a final step a reliability test on the cleaned data is carried out to determine if further, unspecified, errors remain in the observations.

Another example of an algorithm capable of real-time cycle slip detection and repair has been proposed by de Jong (1998) and was implemented in the Dutch Permanent GPS Network and during the International GLONASS Experiment (Jonkman and de Jong, 2000b). The algorithm is based on the use of a Kalman filter in conjunction with the recursive Detection, Identification and Adaptation (DIA) procedure developed by Teunissen (1990). The DIA procedure consists of an overall model test to detect any unspecified errors in the observation or dynamic models (Detection). If an unspecified error is encountered a number of alternative models, incorporating different bias parameters, are tested. The model producing the highest test statistic is considered the most likely to represent the “correct” observation model (Identification). Finally the original observation model is modified to reflect the identified bias (Adaptation).

The DIA algorithm was developed to be independent of the positioning application the data was intended for. Thus no external information such as receiver-satellite geometry, clock offsets or atmospherics should be required. This is accomplished through the use of the geometry free linear combination for both the observation and dynamic models (Jonkman and de Jong, 2000b). Of particular note is that this method is applied on a satellite-by-satellite basis for a single receiver, thus no observation differencing is required. This is advantageous in the sense that data from other receivers is not required to detect cycle slips. Further studies by de Jong (1998) showed that the DIA geometry free approach, on a satellite-by-satellite basis, was theoretically capable of detecting slips of a single cycle in magnitude, provided the observation interval is relatively short.

2.3 Stochastic Modelling

GPS data processing involves the determination of various unknown parameters (e.g. station coordinates, tropospheric estimates, integer ambiguities etc.) from a set of observations. Generally these observations consist of different types of measurements on different frequencies (e.g. code and phase measurements on L1 and L2) and there are usually large numbers of them when compared to the unknown parameters. In the positioning community the accepted methodology for determining the parameters is least squares (LSQ) estimation. LSQ estimation relates the observed quantities to the unknown parameters through a set of mathematical equations known as a functional model. The noise or precision of the observed quantities is represented using a stochastic model.

A great deal of work has been put into the development of functional models for GPS data processing. The stochastic model has received less attention from researchers until relatively recently. As a result simple stochastic models are frequently used in LSQ based GPS data processing algorithms (Tiberius et al., 1999).

Stochastic models are used in three phases of GPS processing, quality control of the raw observations, ambiguity resolution, and computation of the unknown parameters (Kim and Langley, 2001). The statistical quantities used in cycle slip detection and repair algorithms are derived from the chosen stochastic model. Thus the use of incorrect or oversimplified stochastic models may result in faulty slip detection, thereby introducing biases into the ambiguity resolution and parameter estimation processes. Similarly, the performance of instantaneous, real-time ambiguity resolution strategies is greatly improved when using accurate stochastic models. Accurate stochastic models reduce the ambiguity search space and ensure that the fixed ambiguities are correct. An incorrect stochastic model could potentially result in faulty ambiguity resolution, with unsatisfactory consequences for the accuracy of the positioning application. Finally, the estimated quality of the unknown parameters (obtained from the LSQ estimation) are implicitly dependent on the a priori stochastic model. An incorrect a priori model may lead to overly optimistic estimates of the derived position quality, leading users to believe they have met quality requirements when, in fact they have not (Tiberius et al., 1999).

A number of methods have been proposed to provide more realistic stochastic models for the various GPS observables. Four approaches will be considered here - the elevation dependent method (Euler and Goad, 1991), the SNR or C/No approach (Brunner et al., 1999, Richter and Euler, 2001), a rigorous least squares estimation approach, and a method based on time differencing (Kim and Langley, 2001).

2.3.1 Elevation dependent modelling

The dependence of observation noise on satellite elevation has been known for some time and can mainly be attributed to the receiver antenna’s gain pattern, with additional contributions from atmospheric attenuation and multipath (Kim and Langley, 2001, Tiberius et al., 1999). Modelling the observation noise with respect to satellite elevation can be carried out using functions tailored to individual receivers (Euler and Goad, 1991) or using general functions that can be applied regardless of receiver type (Hugentobler et al., 2004). One drawback of the elevation dependent approach is that it only considers the variance of the individual observations. Cross correlations between observations types (e.g. C1 and P2) are neglected, as are spatial and temporal correlations. Thus a fully populated variance covariance matrix is not available when using this method.

2.3.2 C/N_0 Based modelling

GPS signal power is expressed in the form of carrier-to-noise power density ratios (C/N_0), also known as signal to noise ratios (SNR). The C/N_0 measurements generated by GPS receivers are an indication of how well the receiver hardware is tracking the incoming GPS signals. As such they provide a direct indication of the quality of the phase observations (Richter and Euler, 2001, Kim and Langley, 2001, Brunner et al., 1999). The C/N_0 approach to stochastic modelling seeks to take advantage of this information to provide a more realistic assessment of the observation noise.

C/N_0 values are highly correlated with satellite elevation, due in the most part to the antenna gain pattern, but also influenced by atmospheric refraction and multipath. Initial work focussed on this link to produce stochastic models that were in effect, elevation dependent (Hartinger and Brunner, 1999). Further work by (Brunner et al., 1999) extended the simple C/N_0 models to account for the fact that C/N_0 is also influenced by signal diffraction. C/N_0 values observed in "clean" environments can be treated as a "known" template for C/N_0 values observed in other environments. Deviations of the observed values from the template are considered to be the result of diffraction and down weighting (or removal) of the observations occurs as a result. The practical difficulties of providing templates for the various receiver-antenna combinations has been discussed in Richter and Euler (2001).

Problems with this method include the dependence on C/N_0 values, which may not be available from all receivers, and the fact that cross, spatial, and temporal correlations are not considered.

2.3.3 Least squares estimation

The least squares estimation approach offers a rigorous solution to the problem of estimating a priori stochastic models. Results in Barnes et al. (1998) indicate that using the optimal stochastic model, estimated from the LSQ residuals, significantly effects positioning results, when compared to alternative modelling approaches (e.g. C/N_0 approach). The basis of this approach is the direct estimation of every element in the a priori variance covariance matrix from the a posteriori observation residuals. Due to the recursive nature of this process it can be incorporated into a Kalman filter or sequential least squares adjustment (Kim and Langley, 2001).

One technique to carry out the estimation of the variance covariance elements is Minimum Norm Quadratic Unbiased Estimation (MINQUE) developed by Rao (1971) and utilised for static baseline processing by Wang (1998). Unfortunately, MINQUE and similar techniques are computationally intensive and not suited to real-time processing. The optimality of the least squares estimation approach is not guaranteed, as the estimation technique may make assumptions about the correlations that do not hold in all cases (e.g. temporal correlations may be ignored). Furthermore, a certain level of observation redundancy is required to produce reasonable estimates, a situation that may not exist in all positioning scenarios (Kim and Langley, 2001).

2.3.4 Differencing in the time domain

Differencing in the time domain was proposed by Kim and Langley (2001) to overcome the three main problems in the existing modelling approaches - the lack of a fully populated variance covariance matrix, no temporal correlations, and no observation redundancy in long baseline solutions. This method takes the view that high order differencing in time (differencing TDs to produce quadruple differences (QDs), then differencing QDs to produce quintuple differences (dQDs)) will remove all systematic biases and correlations, leaving only white noise.

The assumption that systematic biases and correlations are removed is justified on the basis that the differencing process is in effect the application of consecutive subtractive filters. These filters remove biases (e.g. receiver and satellite clock offsets), damp low frequency effects (e.g. atmospherics, multipath), and amplify high frequency effects (e.g. noise, ionospheric scintillation). For short baselines the effects of the correlated biases are assumed to be ignorable, thereby implying the temporal correlations are also ignorable. However, temporal correlations may still exist, particularly in high multipath environments, thus high order differencing is still required. For long baselines the correlated biases are not ignorable and consequently time correlations will exist (Kim and Langley, 2001). Assuming the dQDs are free of systematic biases and correlations they represent white noise at the dQD level. The variance covariance matrix of the dQDs can then be formed from a set of arbitrary dQD samples. Using the mathematical relationship between the various differencing levels, variance covariance matrices for any difference (i.e. zero, single, double) can be derived.

Of concern here is the generation of the dQD variance covariance matrix. One solution is the estimation of covariance functions. However, this is a computationally intensive process not really suited to real-time use. If a simpler technique is utilised one must question its effectiveness in correctly modelling the cross, spatial, and temporal correlations, particularly when extrapolating back from the dQDs. Furthermore, this method is

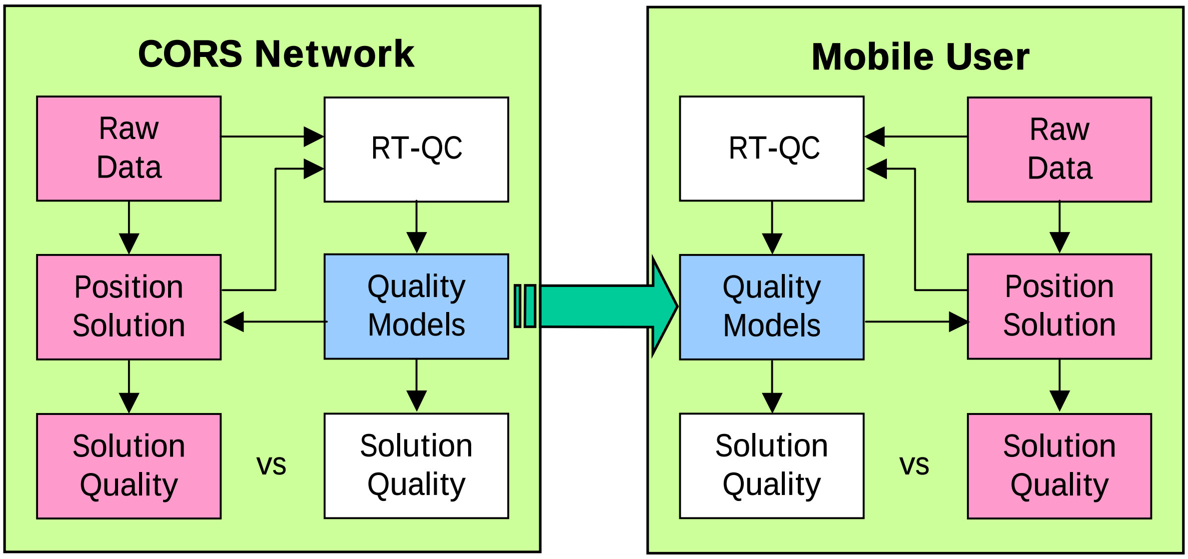


Fig. 2: Proposed RT-QC Architecture.

dependent on the selection of an appropriate time interval for the differencing. The assumption that the dQD observable represents white noise requires the high frequency biases and correlations (which are amplified by the use of subtractive filters) to be insignificant. This may not always be the case (e.g. in unstable ionospheric conditions) and it may be necessary to adjust the time interval in response to changes in the behaviour of the high frequency biases.

3 RT-QC Architecture

The aim of the research being undertaken is the development of real time procedures for CORS networks and mobile users that will improve the reliability of the mobile user's position and provide a realistic assessment of the position quality. Through an examination of the existing approaches to assessing raw data quality an understanding of the various aspects and limitations of raw data quality assessment has been developed. To proceed further, a conceptual architecture for a proposed RT-QC system has been developed and is shown in Fig 2.

The RT-QC architecture is built around the idea that the assessment of raw data quality (RT-QC box) should be carried out independent of the processing algorithm (Position Solution box). However, in the initial stages of the project information from the position solution will be considered during the quality control process. The red boxes indicate the current approach to assessing the quality of CORS and mobile users position and raw data. As Fig. 2 shows, this research is attempting to develop procedures whereby quality models of the CORS network data can be transmitted to a mobile user, thereby improving the quality of the mobile user's position and the estimates of position quality.

4 Conclusions

The number of critical decisions made on the basis of GPS positions has increased proportionally with the use of GPS within the community. When faced with a decision that may have severe consequences GPS users must be confident that their position has been determined to a sufficient level of quality to justify the decision and that the indicators of quality their decisions is based are realistic and reliable. The quality of a GPS position is a direct result of the raw data quality and the processing algorithm chosen. This paper has presented a review of some existing methods for the assessment and reporting of raw GPS data quality and the potential of these methods to be adapted for use in a real time environment. A conceptual architecture of a RT-QC system has been presented as a way forward for future research in this area.

Acknowledgements

This work has been undertaken as part of Project 1.2 in the CRC-SI. The authors would like to acknowledge the support of the industry partners in this project.

References

Barnes, J. B., N. Ackroyd, et al. (1998). ***Stochastic modelling for very high precision real-time kinematic GPS in an engineering environment.*** FIG XXI International

Congress. Commission 6, Engineering Surveys, Brighton, UK.

Brown, N., A. Kealy, et al. (2003). ***Quality control concepts for regional GPS reference station networks***. SatNav 2003, The 6th International Symposium on Satellite Navigation Technology, Melbourne, Australia; 1-13.

Brunner, F. K., H. Hartinger, et al. (1999). ***GPS signal diffraction modelling: the stochastic SIGMA-D model***. Journal of Geodesy 73: 259-267.

de Jong, C. D. (1998). ***A unified approach to real-time integrity monitoring of single- and dual-frequency GPS and GLONASS observations***. ActaGeodaetica et Geophysica Hungarica 33(2-4): 247-257.

Estey LH, Meertens CM (1999). ***TEQC: The Multi-Purpose Toolkit for GPS/GLONASS Data***. GPS Solutions, 3(1):42-49.

Euler, H. J. and C. C. Goad (1991). ***On optimal filtering of GPS dual frequency observations without using orbit information***. Bulletin Geodesique 65(2): 130-143.

Fraser, R. (2004). ***SIRF Output Message ID 28***. Personnel Communication.

Gurtner, W. (1999). ***Ashtech clock jumps in RINEX files***. http://listserv.unb.ca/bin/wa?A2=ind9909&L=canspace&T=0&F=&S=&P=4271 .

Accessed 3/11/04

Hartinger, H. and F. K. Brunner (1999). ***Variances of GPS phase observations: The SIGMA-E Model***. GPS Solutions 2(4): 35-43.

Hofmann-Wellenhof, B., H. Lichtenegger, et al. (1992). ***GPS Theory and Practice***. New York, Springer-Verlag Wien.

Hugentobler, U., R. Dach, et al. (2004). ***Bernese GPS Software, Version 5.0 DRAFT***. Bern.

Jonkman, N. F. and C. D. de Jong (2000). ***Integrity Monitoring of IGEX-98 Data. Part I: Availability.*** GPS Solutions 3(4): 10-23.

Jonkman, N. F. and C. D. de Jong (2000). ***Integrity Monitoring of IGEX-98 Data. Part II: Cycle Slip and Outlier Detection.*** GPS Solutions 3(4): 24-34.

Kim, D. and R. B. Langley (2001). ***Quality Control Techniques and Issue in GPS Applications: Stochastic Modelling and Reliability Test***. International Symposium on GPS/GNSS, Jeju Island, Korea.

Land Victoria (2004). Victorian GPSnet Website. http://www.land.vic.gov.au .

Accessed 11/11/04

Rao, C. R. (1971). ***Estimation of Variance and Covariance components - MINQUE Theory***. Journal of Multivariate Analysis 1: 257-275.

Richter, B. and H. J. Euler (2001). ***Study of improved observation modelling for surveying type applications in multipath environment***. http://www.leicaatl.com/support/gps/Technical_papers/727422en.pdf .

Accessed 8/11/04

Rizos, C. (1999). ***Simultaneity considerations for GPS phase reductions***. http://www.gmat.unsw.edu.au/snap/gps/gps_survey/chap6/638.htm .

Accessed 2/9/04

Teunissen, P. J. G. (1990). ***An integrity and quality control procedure for use in multi sensor integration***. ION GPS 90, Colorado Springs; 513-522.

Tiberius, C. C. J. M., N. F. Jonkman, et al. (1999). ***The Stochastics of GPS Observables***. GPS World 10(2): 49-54.

Wang, J. (1998). ***Stochastic modeling for static GPS baseline data processing***. Journal of Surveying Engineering 124(4): 171-181.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 230-239

Journal of Global Positioning Systems
2005
© CPGPS

Troposphere Modeling in a Regional GPS Network

S. Skone and V. Hoyle

Department of Geomatics Engineering, University of Calgary, 2500 University Dr. N.W., T2N 1N4 Calgary, Canada
e-mail: sskone@geomatics.ucalgary.ca Tel: + 01-403-220-7589; Fax: +01-403-284-1980

Received: 16 November 2004 / Accepted: 12 July 2005

Abstract. By using a regional network of Global Positioning System (GPS) reference stations, it is possible to recover estimates of the slant wet delay (SWD) to all GPS satellites in view. SWD observations can then be used to model the vertical and horizontal structure of water vapor over a local area, using a tomographic approach. The University of Calgary currently operates a regional GPS real-time network of 14 sites in southern Alberta. This network provides an excellent opportunity to study severe weather conditions (e.g. thunderstorms, hail, and tornados) which develop in the foothills of the Rockies near Calgary. In this paper, a 4-D tomographic water vapor model is tested using the regional GPS network. A field campaign was conducted during July 2003 to derive an extensive set of truth data from radiosonde soundings. Accuracies of tomographic water vapor retrieval techniques are evaluated for 1) using only ground-based GPS input, and 2) using a ground-based GPS solution augmented with vertical wet refractivity profiles derived from radiosondes released within the GPS network. Zenith wet delays (ZWD) are computed for both cases, by integrating through the 4-D tomography predictions, and these values are compared with truth ZWD derived from independent radiosonde measurements. Results indicate that ZWD may be modeled with accuracies at the sub-centimeter level using a ground-based GPS network augmented with vertical profile information. This represents an improvement over the GPS-only approach.

Key words: troposphere, water vapor, tomography, GPS, positioning, atmospheric errors Ionosphere, WADGPS, WAAS

1 Introduction

GPS range observations are derived under the assumption that GPS signals travel at the speed of light (or, equivalently, the index of refraction is equal to one) along the satellite-receiver signal path. For GPS orbits of approximately 20,000 km altitude, the signal must travel through the Earth's ionosphere and neutral atmosphere. In these regions indices of refraction can differ significantly from assumptions, such that range errors arise from signal propagation through the Earth's atmosphere. The range errors induced by the ionosphere are dispersive and 99% of the ionospheric effect may be removed using dual frequency GPS observations (Brunner and Gu, 1991).

Range errors associated with propagation through the neutral atmosphere can be classified as a hydrostatic component and a wet delay. The total delay Δs is related to the neutral refractivity N as follows:

$$\Delta s = \int_{path} \frac{N}{10^6} ds \qquad (1)$$

where N may be expressed as (cf. Ware et al., 1997)

$$N = \underbrace{77.6\frac{P}{T}}_{\text{hydrostatic}} + \underbrace{3.73\times10^5 \frac{e}{T^2}}_{\text{wet}} = N_h + N_w \qquad (2)$$

The variable P represents air pressure in millibars, T is the temperature in degrees Kelvin and e is the partial pressure of water vapor in millibars. The variables N_h and N_W represent the hydrostatic and wet refractivities, respectively. The total tropospheric range delay (Equation 1) can therefore be expressed as the sum of both wet and hydrostatic components:

$$\Delta s = SHD + SWD \qquad (3)$$

where SHD and SWD refer to the slant hydrostatic and slant wet delays, respectively, along the signal path.

Range delays arising from the hydrostatic component (SHD) can be computed with accuracies of a few millimeters using existing models, provided that surface barometric or meteorological data are available (Bevis et al., 1992). By using carrier phase-based differential GPS techniques and removing the hydrostatic component, it is possible to recover estimates of the slant wet delay (SWD) for all satellites in view. Previous research has demonstrated that double difference slant water vapor may be determined with millimeter-level accuracy (translating into better than 1 cm accuracy for SWD), for satellite elevation angles greater than 20 degrees (Ware et al., 1997).

Extensive measurements of SWD may be derived from dense geodetic networks of continuously operating GPS reference stations. The time-varying vertical and horizontal structure of wet refractivity may then be modeled by using the SWD as input observables in a tomographic approach. Flores et al. (2000) have developed a 4-D modeling technique in which the wet refractivity (or functions describing the wet refractivity) is estimated for discrete voxels. Horizontal and vertical smoothing constraints are applied to compensate for undetermined voxels. Perturbations of 3.5 mm/km in vertical profiles are resolved for altitudes below 4 km. Gradinarsky and Jarlemark (2002) have proposed a slightly modified approach, in which wet refractivity values in individual voxels are related via cross-correlation (covariance) information, as opposed to applying smoothing constraints. Results of such studies are promising, and suggest that water vapor fields may be derived with sufficient accuracy for meteorology and precise positioning applications.

2 Background

2.1 Tomography

2.1.1 Measurement model

In the derivations presented here, the following properties are assumed for the wet refractivity N_w:

1) Horizontal variations of N_w can be described as a low-order expansion in latitude and longitude.

2) Vertical variations of N_w can be described as constant values in discrete layers.

This approach is similar to the voxel algorithms, in that the troposphere is considered to consist of discrete vertical layers. Wet refractivity values for each vertical layer are related in the filtering approach via covariance information. Horizontal variations are estimated using a functional approach, which is essentially equivalent to the smoothing constraints applied in voxel models.

The slant wet delay is related to the wet refractivity through Equation 1. This expression may be re-written for the slant wet delay component as follows:

$$SWD = 10^{-6}\int_{path} N_w(\phi,\lambda,h)ds \tag{4}$$

where Nw is a function of latitude (ϕ), longitude (λ) and height (h). In assuming that Nw is constant in a given vertical layer, Equation 4 can be approximated as a summation:

$$SWD = \sum_{j=1}^{n} N_{wj}(\phi_j,\lambda_j,h_j)ds_j \tag{5}$$

where the troposphere consists of n vertical layers and Nwj represents the wet refractivity at the mid-point (ϕ_j, λ_j, h_j) of the ray with length ds_j in layer j. This concept is illustrated in Figure 1. Equation 5 can be further re-written to include the functional relationship describing horizontal variations in N_w:

$$SWD = \sum_{j=1}^{n}(a_{0j} + a_{1j}\Delta\phi_j + a_{2j}\Delta\lambda_j + a_{3j}\Delta\phi_j^2 + a_{4j}\Delta\lambda_j^2 + a_{5j}\Delta\phi_j\Delta\lambda_j)ds_j \tag{6}$$

where

$a_{0j},\ldots,a_{5j}$ are the expansion coefficients for layer j at height hj

$\Delta\phi_j = \phi_j - \phi_0$

$\Delta\lambda_j = \lambda_j - \lambda_0$

(ϕ_0,λ_0) is the expansion point (generally chosen as the centroid of the GPS network)

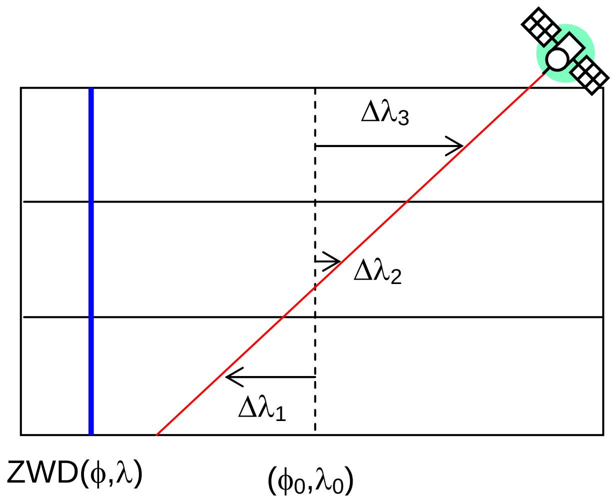


Fig. 1 Sample geometry of wet refractivity estimation in three discrete vertical layers

For the purposes of the testing conducted here, it is assumed that the troposphere consists of eight discrete vertical layers at approximately 750 m intervals. There are a total of 48 unknowns in the adjustment.

2.1.2 System model

The model unknowns (a_{ij} where i=0,1,…,5 and j=1,…,n) are approximated as stochastic processes in time. A first order Gauss-Markov process is assumed for temporal correlations in wet refractivity, and the following system model is employed to describe temporal variations in the model coefficients:

$$a_{ij}(t_{k+1}) = e^{-\beta(\Delta t)} a_{ij}(t_k) + w \qquad (7)$$

where $1/\beta$ is the *correlation time* and $\Delta t = t_{k+1} - t_k$.

Equation 7 provides a statistical description of how model coefficients vary over time. The coefficients at a given time are only partially correlated with those at later epochs, with the normalized autocorrelation function being given as $e^{-\beta(\Delta t)}$. The uncorrelated part of the predicted coefficient $a_{ij}(t_{k+1})$ is described by a white noise sequence w with variance q(t):

$$q(t) = \sigma^2 [1 - e^{-2\beta(\Delta t)}] \qquad (8)$$

where q(t) is the process noise. For the model implemented here, a correlation time Δt of 1800 s is assumed, while the values of σ^2 are set as follows:

a_0: $\sigma^2 = 10\ (\text{mm/km})^2$

a_1, a_2: $\sigma^2 = 2\ (\text{mm/km})^2/\text{deg}^2$

a_3, a_4, a_5: $\sigma^2 = 0.5\ (\text{mm/km})^2/\text{deg}^4$

2.1.3 Prediction and update equations

The standard discrete Kalman filter equations are given as follows (after Gelb (1974)), where the superscripts - and + denote prediction and update, respectively.

1) Prediction (from time t_k to t_{k+1})

$$\mathbf{x}^-(t_{k+1}) = \mathbf{\Phi}(t_k, t_{k+1})\mathbf{x}^+(t_k) + \mathbf{w} \qquad (9)$$

$$\mathbf{P}^-(t_{k+1}) = \mathbf{\Phi}(t_k, t_{k+1})\mathbf{P}^+(t_k)\mathbf{\Phi}(t_k, t_{k+1}) + \mathbf{Q}(t_k) \qquad (10)$$

2) Update (at time t_{k+1})

$$\mathbf{x}^+(t_{k+1}) = \mathbf{x}^-(t_{k+1}) + \mathbf{K}[\mathbf{z}(t_{k+1}) - \mathbf{H}(t_{k+1})\mathbf{x}^-(t_{k+1})] \qquad (11)$$

$$\mathbf{P}^+(t_{k+1}) = [\mathbf{I} - \mathbf{K}\mathbf{H}(t_{k+1})]\mathbf{P}^-(t_{k+1}) \qquad (12)$$

where $\mathbf{K}$ is the gain matrix:

$$\mathbf{K} = \mathbf{P}^-(t_{k+1})\mathbf{H}^T(t_{k+1})[\mathbf{H}(t_{k+1})\mathbf{P}^-(t_{k+1})\mathbf{H}^T(t_{k+1}) + \mathbf{R}(t_{k+1})]^{-1} \qquad (13)$$

The vector **x** represents the unknown coefficients (a_{0j},…, a_{5j} for all vertical layers j), $\mathbf{\Phi}$ is the transition matrix, and **H** is the design matrix. The matrices **R** and **P** are covariance matrices for the observations **z** and estimates of the unknowns **x**, respectively. Variances for the observations are estimated as follows:

$$\sigma^2 = (1.6\text{cm}^2)/\sin E \qquad (14)$$

where E is the satellite elevation angle. The observation variances are based on processing conducted at the University of Calgary, where Bernese software was used to derive SWD observations over a period of several weeks. The SWD estimates were compared with pointed water vapor radiometer observations (truth data) and errors computed for various ranges of elevation angles (Skone and Shrestha, 2003).

The **P** matrix is fully populated, where cross-covariances are used to model the correlations between parameters in different vertical layers. The cross-correlation is derived as a function of distance between the given layers. Covariances also depend on height, where lower correlations are assumed for the lower troposphere layers – where inversion events and irregular variations in the vertical wet refractivity profile may occur.

3 Southern Alberta Network and A-GAME

The Southern Alberta Network (SAN) consists of 14 GPS receivers across southern Alberta, deployed in 2003 by the Geomatics Engineering Department at the University of Calgary (Figure 2). The spacing between SAN stations was designed to be approximately 50 km in order to give optimal results for mesoscale numerical weather prediction, and at the same time allow for precise positioning applications. In general, equipment at each SAN station consists of a NovAtel 600 antenna, NovAtel MPC receiver and Paroscientific MET3A meteorological sensor, although some sites do not have a MET3A instruments due to cost limitations.

During July 14-28, 2003 the A-GAME (Alberta – GPS Atmospheric Monitoring Experiment) data collection campaign took place within this network. This campaign was a collaborative effort between the Geomatics Engineering Department at the University of Calgary, the Meteorological Service of Canada (MSC), and Weather Modification Inc. (a private company employed in detection and mitigation of severe weather). Data were collected from the SAN, and radiosondes were released at a number of locations within the network at regular intervals as well as during storm periods.

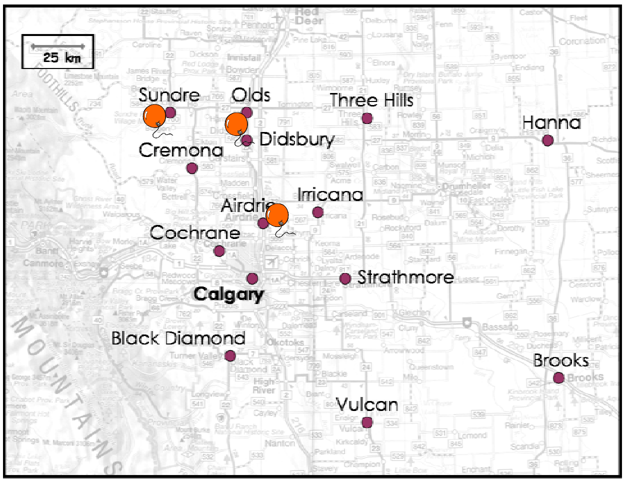


Fig. 2 The Southern Alberta Network during A-GAME 2003. GPS stations are shown as purple dots, and locations of radiosonde launches are shown as orange balloons

The radiosondes were launched at Airdrie approximately three times per day (by personnel from the MSC), and at both Sundre (by personnel from University of Alberta) and Olds/Didsbury airport (by Weather Modification Inc.) once per day - at noon local time. The Sundre radiosonde observations were of questionable quality since the instruments had been stored for some time previously and were tracked visually; these observations were not used to derive results presented in this paper. The Airdrie and Olds/Didsbury instruments were manufactured by Vaisala (2004). In the processing conducted here, radiosonde observations are used as both vertical constraint information (Airdrie) and truth data for assessment of model accuracies (Olds/Didsbury). An example of a single sounding from Airdrie is shown in Figure 3. Weather Modification Inc. also collected radar images within the network (with their TITAN instrument), which allowed correlation of storm evolution with GPS modeling results.

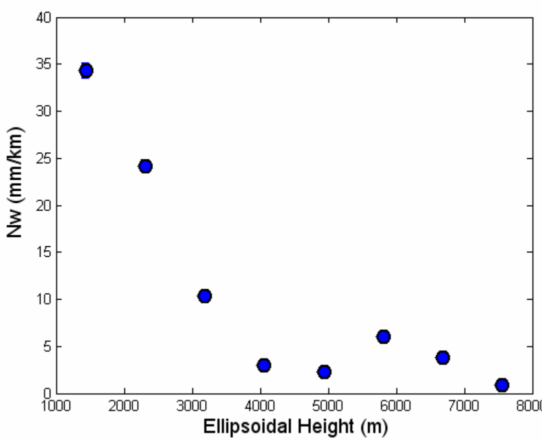


Fig. 3 Sample profile of wet refractivity derived from radiosonde observations at Airdrie

4 Simulation results

A flat network geometry may lead to inaccuracies in vertical profiles of N_w derived using a tomographic approach with only ground-based GPS input. Accuracies of integrated ZWD predictions are compromised to some extent through inability to resolve vertical features. In order to assess such limitations for the SAN, simulations were conducted to evaluate vertical resolution as a function of network geometry. The simulations are based on a suite of MATLAB programs in the Satellite Navigation Toolbox 2.0™ developed by GPSoft. These programs simulate the GPS constellation and range observations for given site coordinates. Slant wet delay observables are generated for the given satellite constellation at various locations in the simulated regional GPS network (network in Figure 2). The tomographic model is then employed (Section 2) to derive refractivity profiles and assess accuracies of model ZWD predictions.

4.1 Method

The approach described in Section 2 is implemented using simulated SWD observations generated every 30 seconds at all reference sites. An elevation cutoff angle of five degrees is assumed, in order to be consistent with further testing conducted in Section 5. Accuracies of the 4-D model were assessed for different tropospheric conditions.

Accuracies of wet refractivity were assessed for two simulated atmospheric profiles:

1) Standard profile where N_w decreases smoothly with altitude.

2) Inversion event where N_w increases with altitude in the lower troposphere, and decreases with altitude at heights above 2 km.

The simulated SWD values are derived through integration of theoretical N_w along each satellite-receiver line-of-sight (e.g. Equation 1). The focus of these tests is to assess the model capabilities in resolving vertical atmospheric structure. The wet refractivity is therefore assumed to have negligible horizontal variations. The vertical distribution of N_w is simulated using the second term in Equation 2 and the following expressions for water vapor (e) and temperature (T) as a function of height (H):

$$T = T_0 - 6.5H \tag{15}$$

$$e = \frac{U}{100}\exp(-37.2465 + 0.213166T - 0.000256908T^2) \tag{16}$$

where H is in kilometers, T is in Kelvin and and e is in millibars. The variable U represents humidity (in percent) and T_0 is the temperature at sea level. For the simulations presented here, U is assumed to be 50 percent and T_0 is 293 °K. The simulated SWD observations have additional random errors imposed as a function of elevation angle, with magnitudes determined from Equation 14. The inversion event is simulated by using Equations 15 and 16 for heights above 2 km, but imposing a positive gradient (as a function of height) in the altitude range 0-2 km.

4.2 Results

Figure 4 shows the wet refractivity estimates generated by the model (after 30 minutes of processing) for the standard profile. The truth data (the N_w profiles used to generate the initial SWD observations) are also plotted for comparison purposes. The N_w values predicted by the model average through the truth profile – representing a smoothed approximation of the vertical atmospheric features. The N_w values are particularly poor at the lower heights, where only one GPS site is located at an altitude sufficiently low enough to observe the bottom atmospheric layer.

Figure 5 shows the wet refractivity estimates for the inversion event (after 30 minutes of processing) versus the truth profile. The irregular inversion profile at lower altitudes is not resolved in the tomography model, with accuracies as poor as 10 mm/km at lower altitudes. Similar to Figure 4, the model values represent a smoothed average of the vertical atmospheric features present above the GPS network. The ground-based GPS observations alone would not allow resolution of inversion profiles.

Results in Figures 4 and 5 demonstrate the impact of network geometry - in particular, vertical station separation within the network - in deriving wet refractivity profiles using ground-based regional networks. Vertical resolution is limited for a flat network such as the SAN, with deficiencies in resolving irregular profiles. For the case of an inversion event, the vertical N_w values generated with a flat network represent only the low-order variations – with an overall smoothing of the true vertical profile.

Results in this section demonstrate that it is difficult to resolve vertical N_w profiles for a flat network geometry, using ground-based GPS data alone. Potential exists, however, to exploit existing sources of vertical information (such as radiosondes or climate models) to constrain the vertical profiles in a tomography approach. By achieving improved vertical resolution through assimilation of such external data sources, it is anticipated that improved ZWD predictions may be derived for GPS users within the GPS network. This type of approach is explored in the next section.

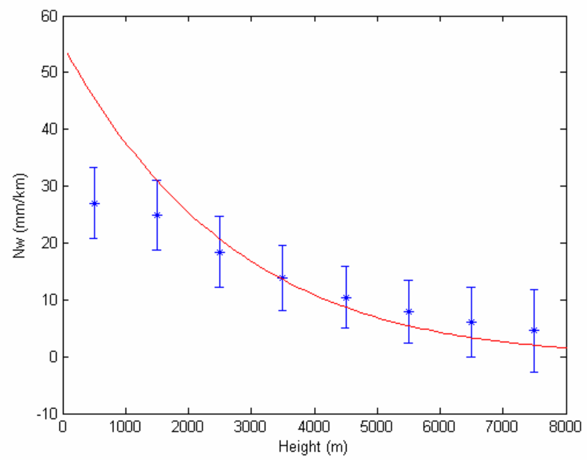


Fig. 4 N_w estimates (blue stars) versus truth (red curve) – standard N_w profile

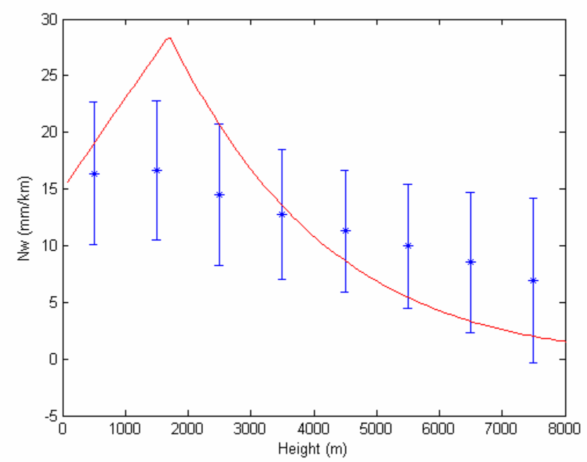


Fig. 5 N_w estimates (blue stars) versus truth (red curve) – inversion event

5 Model results: A-GAME

This section shows model results derived using SAN data, augmented with radiosonde observations, for the A-GAME 2003 campaign. Results are derived for a number of days representing various weather conditions. A brief description of the processing approach and specific data sets follows.

5.1 Estimation of SWD

Hourly estimates of total zenith delays were derived at each receiver in the SAN, with the exception of Olds and Didsbury (Figure 2), using Bernese version 4.2 (Hugentobler et al., 2001), with an ionosphere-free fixed approach using 30-second observations and an elevation mask of five degrees. The hydrostatic component of the total zenith delay was removed using the Saastamoinen model for hydrostatic delay (cf. Bar-Server and Kroger, 1998):

$$D_H = \frac{0.22765P}{(1-0.00266\cos 2\phi - 0.00028h)} \tag{17}$$

where D_H is the hydrostatic delay in centimeters, P is the pressure at the station in millibars, φ is the station latitude in degrees and h is the station height in kilometers. The remaining zenith wet delay was then mapped to the appropriate elevation angle using the Niell wet mapping function (Niell, 1996). In this way, SWD observations were derived for all satellites in view at each available station within the SAN (Figure 2). These observations are used as input observables in the tomography model.

5.2 Radiosonde vertical N_W constraints and truth data

As described in Section 3, radiosonde observations were available at a number of SAN sites during the A-GAME 2003 campaign. These measurements can be used as additional input observations for the tomographic model - serving essentially as vertical profile constraints. The addition of such high-resolution vertical information allows improved 4-D modeling using the tomographic approach, when combined with the SWD estimates from GPS reference sites. For the tests conducted here, two sets of radiosonde observations are used:

- Airdrie: vertical profiles of N_W are derived and assimilated into the tomography model.
- Olds/Disbury airport: vertical profiles of N_W are derived but excluded from the tomography adjustment and instead used as independent truth data – to assess model prediction accuracies.

Note that Airdrie and Olds/Didsbury airport sites are ~50 km apart. Neither Olds nor Didsbury GPS observations were used in the tomography processing, in order to independently assess model predictions in this region when compared with the local (Olds/Didsbury Airport) radiosonde truth values.

In order to assimilate the Airdrie radiosonde observations into the tomography model, values of wet refractivity were estimated for each sounding. Single observations of N_W were derived from radiosonde measurements for each layer defined in the tomography model (e.g. the eight vertical layers of thickness 750 m), by averaging all N_W point measurements made in the given layer as the balloon ascended. The N_W values were estimated using the following equations from the ICS (2004):

$$\log_{10}(e_s) = \frac{-2937.4}{T} - 4.9283\log_{10} T + 23.5470 \tag{18}$$

$$e = \frac{RH(\%)}{100} e_s \tag{19}$$

where

e_s is the saturation pressure of water vapor in hectoPascal or millibars

T is the temperature in Kelvin

e is the water vapor pressure in hectoPascal or millibars

As a final step N_W was calculated from Equation 2 as

$$N_W = 3.73\times10^5\left(\frac{e}{T^2}\right) \tag{20}$$

Observation variances were derived from the laws of error propagation with temperature and relative humidity having uncertainties as given by Vaisala (2004). Wet refractivity profiles and associated error bars were derived for the Airdrie radiosondes in this manner, for direct assimilation into the tomography model. These radiosonde profiles are generally assumed to be valid for a one-hour period, and the analyses presented here focus on periods just after radiosonde launch.

In order to adequately assess the 4-D wet refractivity predictions versus truth, it is important that the Airdrie radiosonde constraint information and the Olds/Didsbury airport radiosonde truth data be available at approximately the same times. Unfortunately, the radiosonde observations at Olds/Didsbury did not always occur at the same time as the radiosonde launches from Airdrie. On the days used for processing, the time differences for these launches were

- July 19, 2003 – same time
- July 20 and 25, 2003 – one hour apart
- July 26, 2003 – two hours apart

5.3 Results and analysis

5.3.1 Data set and processing

Two days were processed as "quiet" days since no meteorological events of interest happened in the network during these times – July 19 and 25, 2003. On July 20 and 26, 2003 large storms passed through the SAN and these times are presented as storm days. As stated earlier, Olds and Didsbury GPS data were excluded from the tomography adjustment since this is where the truth comparison (model predictions compared with radiosonde truth) takes place.

GPS results shown here are processed using as many stations in the SAN as had surface pressure measurements and GPS data on days of interest. Some data drop-outs were encountered at sites during the A-GAME 2003 campaign, and thus the numbers of stations used for processing were as follows:

- July 19 & 25, 2003 (7 and 8 stations)
- July 20 & 26, 2003 (6 and 5 stations)

In order to retrieve absolute (and not relative) troposphere measurements, three IGS stations were included in the Bernese software processing to derive SWD values: ALGO (Algonquin Park in Ontario, Canada), DRAO (Dominion Radio Astrophysical Observatory in B.C., Canada) and NLIB (North Liberty, U.S.A.) which are approximately 2680 km, 430 km and 1890 km from the network, respectively.

Two types of processing are conducted:

- Ground-based GPS stations alone. In this case, the tomography model uses only SWD input from available GPS stations. This approach is herein referred to as "GPS".
- The GPS approach is augmented by including radiosonde observations from Airdrie as observational input to the tomography model. This approach is herein referred to as "GPS + RS".

Wet refractivity and zenith wet delay values are shown in the following sections for times when Olds/Didbury and Airdrie radiosonde launches took place within enough time of each other for valid model versus truth comparisons to be conducted (two hours or less apart). The ZWD values are derived from the model predictions (for the two different test cases) by integrating upwards through the N_W field predicted by the tomography model at the location of interest (the Olds/Didsbury truth site). Similarly, the Olds/Didsbury N_W truth values are integrated vertically to derive truth ZWD estimates – for comparison with model predictions.

5.3.2 Quiet days

Figures 6 and 7 show results for the first quiet day: July 19, 2003. The results for GPS + RS best match truth for both vertical N_W profiles and integrated ZWD plots. ZWD accuracies of 0.3 cm are achieved for model predictions when radiosonde observations are assimilated into the tomography model, versus accuracies of 2 cm for using ground-based GPS observations alone. The N_W profile obtained from GPS observations has negative values at the lower altitudes, which is clearly in error.

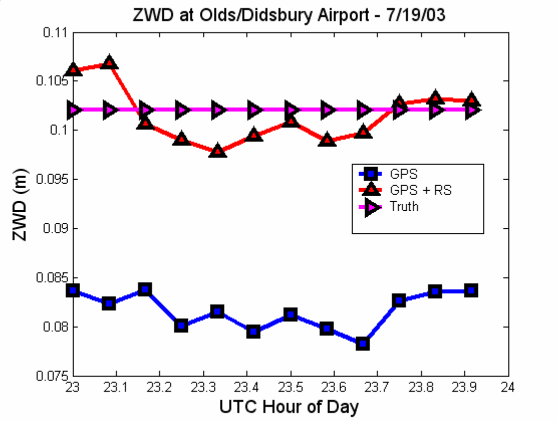


Fig. 6 Integrated ZWD solutions at Olds/Didsbury airport for July 19, 2003

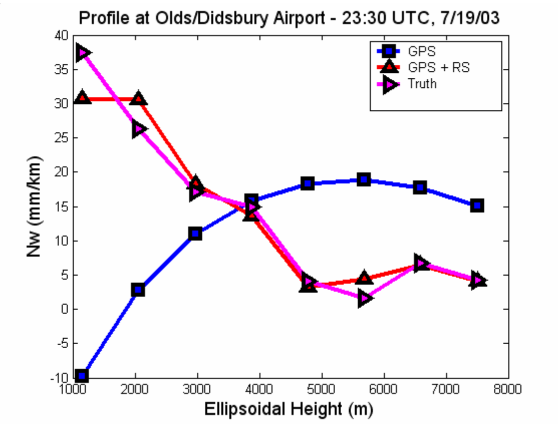


Fig. 7 Vertical N_W profile at Olds/Didsbury airport July 19, 2003 at 23:30 UTC

Integrated ZWD results for July 25 are shown in Figure 8. The GPS + RS solution has an overall accuracy of approximately 1 cm, while the GPS solution has errors of 3 cm. The GPS N_W profile for this day exhibits the correct trend (higher values at lower altitudes) when compared with the July 19 results, but it deviates significantly from the truth values (Figure 9). Overall, results in this section demonstrate the improved modeling of tropospheric wet delay that may be achieved by assimilating vertical profile observations into the tomographic model.

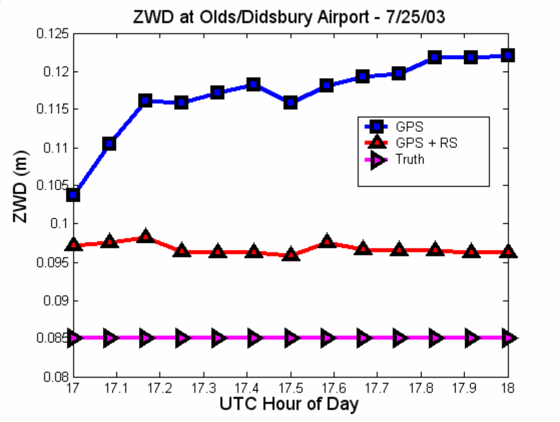


Fig. 8 Integrated ZWD solutions at Olds/Didsbury airport for July 25, 2003

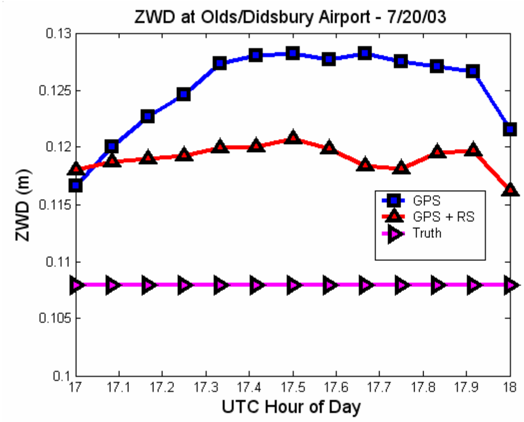


Fig. 10 Integrated ZWD solutions at Olds/Didsbury airport for July 20, 2003

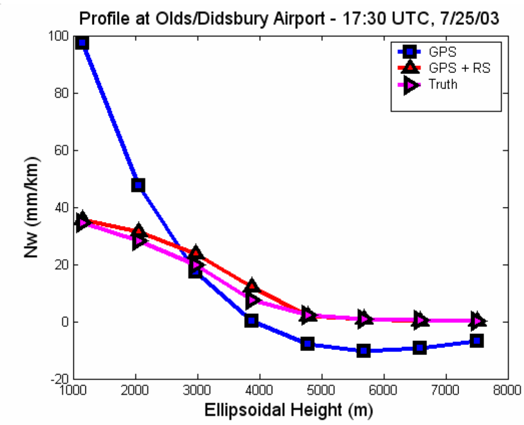


Fig. 9 Vertical N_W profile at Olds/Didsbury airport for July 25, 2003 at 17:30 UTC

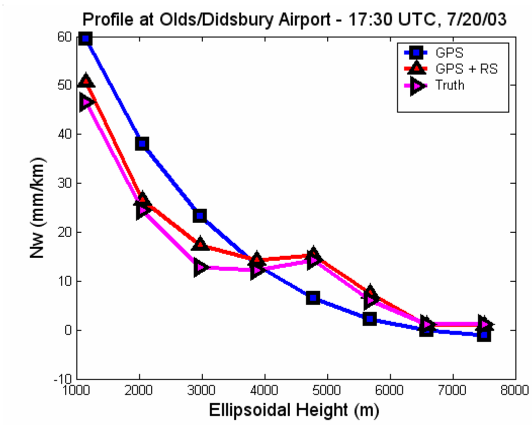


Fig. 11 Vertical N_W profile at Olds/Didsbury airport for July 20, 2003 at 17:30 UTC

5.3.3 Storm days

Figures 10 and 11 show the integrated ZWD values and N_W vertical profiles, respectively, for July 20, 2003. Integrated ZWD values for the GPS + RS solution have improved accuracies (approximately 1 cm) versus the GPS solution. This is consistent with results in Section 5.3.2 for the quiet days. Profiles for July 20 show small-scale variation in the GPS + RS and truth vertical profiles. These features, which appear to be real, are smoothed through in the GPS and GPS solutions.

Figures 12 and 13 show the integrated ZWD values and N_W vertical profiles, respectively, for July 26, 2003. The GPS + RS solution has overall accuracies of approximately 0.5 cm with respect to the truth solution, while the GPS ZWD solutions are approximately 2 cm higher than truth. Again, ZWD results are improved by including radiosonde information in the tomography solution. The GPS N_W profile appears to (incorrectly) represent an inversion structure, while the GPS + RS N_W profile follows the truth profile more closely.

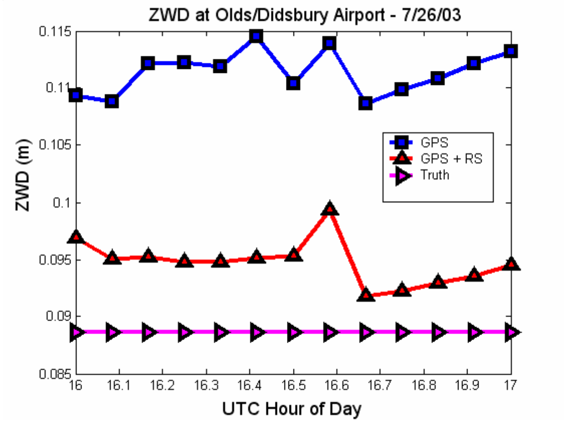


Fig. 12 Integrated ZWD solutions at Olds/Didsbury airport for July 26, 2003

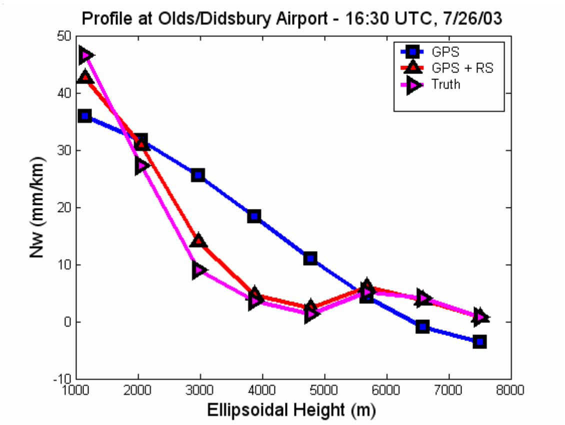


Fig. 13 Vertical N_W profile at Olds/Didsbury airport for July 26, 2003 at 16:30 UTC

5.3.4 Summary

Table 1 summarizes the integrated ZWD results for all cases presented in Sections 5.3.2 and 5.3.3. Accuracies on the quiet day July 25 are the most significantly improved by assimilating radiosonde observations into the tomography model. Overall, the results show promising potential for exploiting ground-based GPS networks and available radiosonde data to model ZWD with cm-level accuracy.

Note that the model ZWD accuracies are perhaps better than expected for the GPS (without radiosonde) solutions, given the poor vertical resolution of the model (e.g. the GPS N_W profile in Figure 7). The tomography N_W solutions are non-unique, however – such that identical integrated quantities may be derived from significantly different N_W profiles. It is possible to derive accurate integrated ZWD estimates from apparently non-realistic vertical N_W profiles. The addition of vertical profile constraints does, however, improve both the model N_W profiles and ZWD predictions.

Tab. 1 Zenith wet delay accuracies from tomography model during times where radiosonde observations are available (storm days in italics)

	RMS (CM)	
Date/Time	GPS	GPS+RS
July 19	2.0	0.3
July 20	1.8	1.1
July 25	3.2	1.2
July 26	2.3	0.6

6 Conclusions

Resolving vertical structures of water vapor using data from a flat GPS network (e.g. the SAN) alone, using a tomography approach, results in poor vertical resolution of wet refractivity, although integrated ZWD quantities are accurate to approximately 1-3 cm. By exploiting other sources of vertical profile information, improvements may be made in tomographic modeling of wet refractivity. Potential sources of vertical profile information include radiosonde data, climate models, microwave profilers, and radio occultation estimates. The addition of radiosonde point measurements from a location within the GPS network (GPS + RS) to ground-based GPS tomography improves the integrated ZWD solution by at least 0.7 cm when compared to the GPS-only tomographic solution, and improves the vertical wet refractivity profiles derived from the tomography model. Absolute ZWD accuracies, when compared to truth values, are in the range 0.3-1.2 cm for both quiet and storm conditions, for this augmented approach.

Water vapor profiles can also be derived from radio occultations using low Earth orbiting (LEO) satellites. Currently several LEO satellites exist with GPS payloads (e.g. CHAllenging Minisatellite Payload – CHAMP, Satelite de Aplicanciones Cientificas C – SAC-C) and there are plans in place for a six-satellite system in the near future (Constellation Observing System for Meteorology, Ionosphere and Climate – COSMIC). First results from the CHAMP mission have indicated that vertical profiles of humidity agree well with European Centre for Medium-Range Weather Forecast (ECMWF) and National Centers for Environmental Prediction (NCEP) specific humidity data (Wickert et al., 2001) to about 1.5 kilometers above the surface of the Earth, where atmospheric water vapor and multipath degrade the solution (Gregorius and Blewitt, 1998). Since occultation data is likely to become more readily accessible and timely in the future, these measurements could be assimilated into the tomographic estimation routine described in this paper. Future plans for follow-on work in fact include assimilation of N_W profiles derived from radio occultations into the tomography model, and to determine their benefit for flat GPS network wet refractivity tomography.

Acknowledgements

The authors wish to acknowledge the International GPS Service, for orbit products and raw station data used in the Bernese processing. We acknowledge our collaborators: Craig Smith for formatting of radiosonde observations; and Geoff Strong and Terry Krauss for consultation on meteorological phenomena. Rinske van Gosliga is acknowledged for Bernese processing of A-GAME 2003 data. This work has been funded in part by the Canadian Foundation for Climate and Atmospheric Sciences (CFCAS). Portions of this paper have also been submitted for publication in Proceedings of the ION GNSS 2004 conference.

References

Bar-Sever Y E.; Kroger P M. (1998): ***Estimating horizontal gradients of tropospheric path delay with a single GPS receiver***, Journal of Geophysical Research, 103(B3):5019-5035

Bevis M S.; Businger S.; Herring T A.; Rocken C.; Anthes R A.; Ware R H. (1992): ***GPS meteorology: remote sensing of atmospheric water vapor using the Global Positioning System***, Journal of Geophysical Research, 97(D14):15787-15801

Braun J.; Rocken C. (2003): ***Water vapor tomography within the planetary boundary layer using GPS***, International Workshop on GPS Meteorology, Tsukuba, Japan, 14-17 January, http://dbx.cr.chiba-u.jp/Gps_Met/gpsmet

Brunner F K.; Gu M. (1991): ***An improved model for the dual frequency ionospheric correction of GPS observations***, Manuscripta Geodaetica, 16:205-214

Flores A.; Ruffini G.; Ruis A. (2000): ***4D tropospheric tomography using GPS slant wet delays***, Annales Geophysicae, 18:223-234

Gelb A. (1974): ***Applied Optimal Estimation***, Cambridge, Mass.: MIT Press

Gradinarsky L.; Jarlemark P. (2002): ***GPS tomography using the permanent network in Goteborg: simulations,*** Proceedings of IEEE Positioning, Location and Navigation Symposium, Palm Springs, CA, 15-18 April, 128-133

Gregorius T.; Blewitt G. (1998): ***The effect of weather fronts on GPS measurements***, GPS World, May:52-60

Hugentobler U.; Schaer S.; Fridez P. (2001): ***Bernese GPS Software Version 4.2 Manual***, Astronomical Institute University of Berne, February

Institute of Information and Computing Science (ICS) Web Site (2004), ***Temp, humidity and dew-point ONA***, http://www.cs.uu.nl/wais/html/na-dir/meteorology/temp-dewpoint.html, Accessed: August 10

Niell A E. (1996): ***Global mapping functions for the atmosphere delay at radio wavelengths***, Journal of Geophysical Research, 101(B2):3227-3246

Skone S.; Shrestha S. (2003): ***4-D modeling of water vapor using a regional GPS network***, Proceedings of the ION National Technical Meeting, Anaheim, CA, 22-24 January

Vaisala (2004): ***RS80 Spec Sheet***, http://www.vaisala.com/DynaGen_Attachments/Att2743/2743.pdf, Accessed August 13

Ware R.; Alber C.; Rocken C.; Solheim F. (1997): ***Sensing integrated water vapour along GPS ray paths***, Geophysical Research Letters, 24:417-420

Wickert J.; Reigber C.; Beyerle G.; König R.; Marquardt C.; Schmidt T.; Grunwaldt L.; Galas R.; Meehan T.; Melbourne W.; Hocke K (2001): ***Atmosphere sounding by GPS radio occultation: first results from CHAMP***, Geophysical Research Letters, 28:3263-3266

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 240-244

Journal of Global Positioning Systems

Benefits of Telecommunications Technology to GPS Users

Thomas Yan

School of Surveying & SIS, University of New South Wales, Sydney NSW 2052, Australia
e-mail: thomas.yan@unsw.edu.au Tel: +61 2 9385 4189; Fax: +61 2 9313 7493

Received: 27 November 2004 / Accepted: 17 October 2005

Abstract. For many years, telecommunications technology has assisted GPS users in accomplishing their tasks. Dial-up system over copper phone line enables users to download data from base station at remote locations. Radio modem provides wireless communications link between a base station and a rover to enable surveyors to carry out RTK-surveys. While these techniques are still very much in use, developments in telecommunications technology over the last decade or so has brought more services providing easier use, faster speed and wider coverage. Fast spread of Internet has made TCP/IP protocols ubiquitous resulting in more devices being IP-enabled and Internet-connected. Wireless technology such as GPRS and 3G make better use of bandwidth providing faster speed and better coverage to mobile users. This paper looks at these new emerging technologies and how they could have impact on GPS users. It also discusses recent GPS-related protocols such as Ntrip and RTCM 3.0 which were designed in response to these new developments. Examples will be presented based on local trends, settings and conditions in Australia.

Key words: telecommunications, network, wireless, mobile, protocols, GPS

1 Introduction

While Global Positioning System (GPS) is mainly used for navigation and surveying purposes, many aspects of the system rely on telecommunications technology. All three segments that make up the system, the space segment (a constellation of satellites), the control segment (ground base stations) and the user segment (the signal receiver) are essentially common building blocks to a telecommunications system. Both the space and control segments of GPS are controlled solely by the U.S. Department of Defence. The user segment, the receivers, are designed and produced by various manufacturers and used widely by public. Naturally, this renders the user segment to much innovative development and fast adoption of new technology.

The simplest use of GPS receiver is by using it autonomously, independent of external feedback. This technique typically only gives a positioning accuracy of approximately 15 m. This figure can significantly be improved by using differential measurement techniques, making it more useful for many applications. These techniques require user's receiver to communicate to one or more other receivers to produce measurement with higher accuracy. Traditionally, users would need to setup their own reference receiver and communication link to the rover receiver. In some areas, the reference receivers might be operated as a service by government agencies or commercial companies. For example, in some countries beacon transmitters have been established along the coasts to assist marine crafts in navigation. All these require dedicated communication link to be established separately. Within the last decade however, Internet and mobile networks have grown very rapidly. This provides users with servicing technology with which they can employ differential techniques widely.

2 Protocols

A **protocol** defines the format and the order of messages exchanged between two or more communicating entities, as well as the actions taken on the transmission and/or receipt of a message or other event (Kurose and Ross, 2003). Usage of common protocols means compatibility and interoperability. This section looks into several current and new protocols that the author believed will have significant role to GPS users.

2.1 RTCM 3.0

Several different protocols exist for exchange of GPS data but two protocols have become standard, NMEA 0813 and RTCM. As the names suggest, these protocols were produced by the National Marine Electronics Association (NMEA) and Radio Technical Commission for Maritime Services (RTCM). In February 2004, RTCM released the third version of their recommended standards for differential GNSS service commonly referred to as RTCM 3.0.

RTCM 3.0 has been developed as a more efficient alternative to previous versions. It was developed based on requests from service providers and vendors for a new standard that would be more efficient, easy to use and more easily adaptable to new situations. The main complaint was that the parity scheme of Version 2 was wasteful of bandwidth. Another complaint was that the parity was not independent from word to word. Furthermore, even with so many bits devoted to parity, the actual integrity of the message was not as high as it should be. RTCM 3.0 is intended to correct these weaknesses (RTCM, 2004).

RTCM 3.0 consists primarily of messages designed to support real-time kinematic (RTK) operations. The reason for this emphasis is that RTK operation involves broadcasting a lot of information, and thus benefits the most from an efficient data format. RTCM 3.0 provides messages that support GPS and GLONASS RTK operations, including code and carrier phase observables, antenna parameters and ancillary system parameters. However, the format is specifically designed to make it straightforward to accommodate modifications to these systems (e.g., new L2C and L5 signals) and to new systems that are under development (e.g. Galileo).

RTCM 3.0 has been designed using a layered approach adapted from the Open System Interconnection (OSI) standard reference model. A diagram of the OSI standard reference model is shown here.

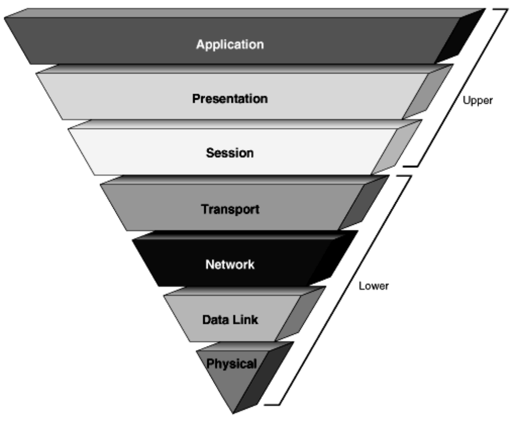


Fig. 1 Seven layers of OSI model (Doyle & Zecker, 1996)

The protocol defines message format on Application, Presentation and Transport layers. The bulk of the document is on the Presentation Layer and describes the message, data elements and data definitions. Implementation on Data Link and Physical layers are left to service providers to determine as they see appropriate to the application.

The higher efficiency of RTCM 3.0 will make it possible to support RTK services with significantly reduced bandwidths. This is especially relevant in wireless and mobile networks where the bandwidth available is much less than that of wired network. The expected performance of this protocol will open ways to more stringent and unique applications of high-accuracy positioning technique. Bock *et al*. (2003) presented a network-based RTK technique in which raw data from several reference stations were aggregated and delivered to users via wireless channel. Such application would benefit tremendously from reduced use of bandwidth.

For wireless link users, who are charged by the amount of bandwidth used, reduced bandwidth means reduced operating cost. For CORS network operators with leased data lines from telecommunications service provider, the reduced bandwidth allows them to provision link with lower data rate and naturally, lower charge from the service provider.

Major GPS manufacturers such as Trimble, Leica and NovAtel have expressed support for RTCM 3.0 by providing firmware upgrade to their products and integrating RTCM 3.0 capability into their current products.

2.2 TCP/IP

The Internet protocols are the world's most popular open-system protocol suite because they can be used to communicate across any set of interconnected networks and are equally well-suited for Local Area Network (LAN) and Wide Area Network (WAN) communications. The Internet protocols consist of a suite of communications protocols, of which the two best known are the Transmission Control Protocol (TCP) and the Internet Protocol (IP).

TCP/IP was first developed in the mid-1970s and has since become the foundation on which the Internet is based. IP occupies the Network Layer on OSI reference model while TCP occupies the Transport Layer. TCP/IP is the foundation on top of which many other Application level protocols such as HTTP and FTP are built.

2.3 Ntrip

Ntrip stands for "Networked Transport of RTCM via Internet Protocol". It is an Application layer level protocol which is used to stream Global Navigation Satellite System (GNSS) data over the Internet. Ntrip is a generic, stateless protocol based on the Hypertext Transfer Protocol (HTTP). The HTTP objects are enhanced to GNSS data streams. Ntrip was built on top of the TCP/IP foundation. It was developed by the Federal Agency for Cartography and Geodesy (known as BKG), Germany.

Ntrip is designed for disseminating differential correction data (e.g. in the RTCM-104 format) or other kinds of GNSS streaming data to stationary or mobile users over the Internet, allowing simultaneous PC, Laptop, PDA or receiver connections to a broadcasting host. Ntrip supports wireless Internet access through Mobile IP networks such as GSM, GPRS, EDGE or UMTS. Recently, Ntrip has been adopted by RTCM as their recommended standard.

The Ntrip system consists of three software components: NtripClient, NtripServer and NtripCaster. The NtripCaster is the actual HTTP server program while NtripClient and NtripServer act as HTTP clients. In the diagram below, NtripServer receives data from a source (typically a GPS reference receiver) and forward it to the NtripCaster. The NtripCaster acts as a 'switchboard' which connects NtripClients to their required streams.

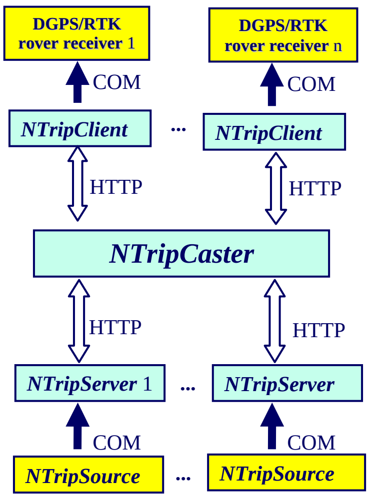


Fig. 2 Connection in an Ntrip system (Weber, 2004)

2.4 Ethernet

The term *Ethernet* refers to the family of Local Area Network (LAN) products covered by the IEEE 802.3 standard that defines what is commonly known as the CSMA/CD (Carrier Sense Multiple Access Collision Detect) protocol. Three data rates are defined in the standard with 10 Mbps (10Base-T Ethernet) and 100 Mbps (Fast Ethernet) being the most common rates at the moment (Cisco Systems, 2003).

Other technologies and protocols have been touted as likely replacements but the market has spoken. Ethernet is currently used for approximately 85 percent of the world's LAN-connected PCs and workstations. Ethernet has survived as the major LAN technology because it is easy to understand, implement, manage and maintain. In provides extensive topological flexibility for network installation. Being a *de facto* standard also implies guaranteed interconnection and operation with other products regardless of manufacturer. Most networked devices have Ethernet port as their standard network connection, from common PC and laptop to the more controversial Internet-enabled fridge and air-conditioner.

3 Applications

3.1 Network Appliance GPS Receivers

A GPS reference receiver typically has three to four serial (RS-232) ports over which it communicates with other devices. In a reference station setup, this receiver is usually connected to a computer which logs measurement data from the receiver which is then distributed to users. This scheme works well for a single base station but has its limitation. As the number of base stations increased, it becomes desirable to control the data centrally. For post-processing use, it is more manageable to have a central data repository compared to multiple computers storing its own set of data. Issues such as user access, fault management, backup, archive and data distribution are all easier to handle with a centralised system. Also, different levels of user needs may mean different type of streams which is currently limited to the physical number of serial ports on the receiver.

For real-time use, network-based solutions such as Network-RTK has proven to be more reliable and offer better performance compared to a single station solution. Network-based solutions require real-time data streams from multiple stations to be aggregated into a central processing system.

A network-enabled GPS receiver provides a good solution to these issues and an elegant way to distribute data and manage the unit. Integration of Ethernet and IP protocols into a GPS receiver brings about the concept of *network appliance* to GPS receivers. With Internet Protocol (IP) as the primary communications method, public domain tools such as web browser and FTP client can be used to configure receiver and access logged data files.

As a network appliance, GPS receiver can provide services to all users attached to the receiver through the network. Different streaming services may be configured on different TCP or UDP ports, for example, with differing data rates or smoothing configurations. To obtain a service, the client has only to connect to a specific port. This allows multiple users to access different streaming services simultaneously.

A network-enabled GPS receiver also provides better remote access to operator. With a web browser, operator can access configurations of the receiver via a webpage without having to connect directly to it. This is especially critical for operators of CORS network where the reference stations are spread over a large area. Previously, to change settings on a GPS reference receiver, operators need to physically connect the computer running the control software to the receiver via serial ports. Obviously, it is not ideal if an operator has to travel hundreds of kilometres only to modify observations rate or other minor settings.

In a CORS network, this concept allows the central server to connect to multiple GPS receivers. With RS-232, the number of data streams is usually limited to the number of physical serial ports available on the computer which is around two to four ports. With Ethernet and IP protocols, it is possible for the server to connect to tens or hundreds of data streams. RS-232 is also much slower compared to Ethernet. RS-232 typically has maximum speed around 115,200 bps whereas Ethernet's speed is typically 10 Mbps – a hundred times faster than RS-232 – with 100 Mbps connection becoming more and more common.

Using Ethernet protocol allows for repeatability, multiple connections and compatibility with higher layer protocols such as TCP/IP or UDP/IP. Being the most popular link layer protocol, Ethernet enables for easy connectivity with other protocols. For example, in SydNET – a real-time permanent GPS network in Sydney, Australia – it allows GPS data to be aggregated to a central server via fibre-optic network which runs on ATM protocol.

As of the time of writing, the author is aware of two network-enabled GPS receivers in the market which are equipped with built-in Ethernet and IP capability. Trimble produces a model called the *NetRS* while Thales Navigation also produces a model called the *iCGRS*.

3.2 Built-in Ntrip

Most parts of Ntrip implementation from BKG have been released under GNU General Public License which means it is open source. This makes it possible for service providers and vendors to incorporate an Ntrip implementation into their products. As of September 2004, implementations of NtripClient are available for PC, Pocket PC PDA and Symbian mobile phone. Some GPS receiver manufacturers such as Trimble and Leica have also added NtripClient and NtripServer implementations into their receiver software.

A list of hardware and software supporting Ntrip is available from BKG's website (http://igs.ifag.de/ntrip_down.htm).

3.3 Using Mobile IP Networks & Wireless Broadband

While RTCM 3.0 addresses bandwidth issue by reducing the message size, mobile networks have also developed in terms of coverage and bandwidth. For example, GSM technology – which gained popularity in Australia and Europe – originally has data capability of 9600 bps. This is only about one-fifth of a 56k dial-up modem. Currently, all GSM networks in Australia have been upgraded with GPRS technology which provides higher data rate and IP network connectivity.

General Packet Radio Service (GPRS) is an upgrade to GSM, providing packet-switching technology and bridging the mobile network to IP network. Unlike GSM, GPRS is an always-on connection where user is charged not by the length of connection but by the amount of data traffic. GPRS provides substantial improvement to data speed with a theoretical limit of up to 170 kbps. GPRS also allows mobile users to access the Internet directly. All GSM networks in Australia, which include Telstra, Optus and Vodafone, support and offer GPRS.

CDMA2000 1X is another technology offered in Australia by Telstra. It offers data rate of around 144 kbps and connectivity to Internet. As with GPRS, users have a choice of PCMCIA access card or a PDA with built-in CDMA capability.

3G technology is another evolution from CDMA with even higher data rate. It claims to provide up to 384 kbps of data rate under ideal condition. So far, only PCMCIA-type card is available to access the service. 3G network is offered in Australia by Hutchison 3G Australia Pty Ltd.

While GPRS, CDMA2000 1X and 3G have all evolved from mobile voice networks, *wireless broadband* started as a service providing broadband connection via radio. Unlike voice networks, this technology is pure IP and is dedicated to data service. Wireless broadband has superior data rate compared to the three technologies mentioned previously. Its maximum speed is at around 1 Mbps which is almost three times that of 3G.

The main disadvantage of wireless broadband network is because they're relatively new their network coverage is nowhere that of voice networks such as GSM and CDMA. The iBurst network for example, is only available in metropolitan area of New South Wales, Queensland and Victoria. Another network, *Unwired*, so far is only available in Sydney. However, it is expected that as this technology gains popularity, their network coverage will expand and cover more areas.

All the technologies mentioned here offer a new and better ways to GPS users to obtain correction service in the field. GSM or CDMA networks, being more ubiquitous in terms of coverage, maybe the choice to many users while those who demand even higher data rate can choose to use personal broadband technology such as *iBurst* or *Unwired* where available.

4 Conclusions

This paper introduced several new technologies which may have significant impact and use to GPS users in the very near future. A new and improved protocol from RTCM allows for reduced bandwidth which means users with slow or limited data link could now exchange data in a standard protocol. Additionally, users now may be able to provision several data streams using the same data link due to the reduced bandwidth. From business point of view, reduced bandwidth means reduced operating cost.

Widespread of Internet and the protocols behind it means a common set of protocols are being adopted. Ethernet and IP protocols have gained wide popularity with many devices adopting it as their standard communication protocols. The concept of GPS receiver as network appliance offer many benefits over the current serial device implementation.

Finally, these protocols allow GPS users to utilise telecommunications networks such as mobile networks to disseminate correction service without having to build new infrastructure. The advent of wireless broadband service with its substantially higher data rate could enable new techniques not possible previously to be used by GPS users.

References

Bock Y.; de Jonge P.; Honcik D.; Fayman J. (2003): ***Wireless Instantaneous Network RTK: Positioning and Navigation***. Proceedings of ION GPS 2003, Portland Oregon, 9-12 Sept 2003, 1397-1405.

Chen R.; Li X. (2004): ***Test Results of An Internet RTK System Based on The Ntrip Protocol.*** European Navigation Conference GNSS 2004, Rotterdam, 16-19 May 2004.

Cisco Systems, Inc. (2003): ***Internetworking Technology Handbook***. http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito_doc/

Doyle P.; Zacker C. (1996): ***Upgrading and Repairing Networks.*** Macmillan Computer Publishing, Indianapolis.

iBurst Homepage, http://www.iburst.com.au.

Kurose J.F.; Ross K.W. (2004): ***Computer Networking A Top-Down Approach Featuring The Internet***. Pearson Addison-Wesley, Boston.

Ntrip Homepage, http://igs.ifag.de/index_trip.htm.

Radio Technical Commission for Maritime Services (2003): ***Networked Transport of RTCM via Internet Protocol, Version 1.0.*** RTCM Paper 167-203/SC104-315, June 2003.

Radio Technical Commission for Maritime Services (2004): ***Recommended Standards for Differential GNSS Service, Version 3.0***. RTCM Paper 86-2004/SC104, February 2004.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 245-257

Journal of Global Positioning Systems
2005
© CPGPS

Inverse Diffraction Parabolic Wave Equation Localisation System (IDPELS)

Troy A. Spencer [1], Rodney A. Walker [1], Richard M. Hawkes [2]

[1] Cooperative Resarch Centre for Satellite Systems (CRCSS), QUT University, Brisbane, Queensland, Australia
e-mail: ta.spencer@qut.edu.au; Tel: + 61 7 38641772; Fax: 61 7 3864 1517

[2] Navigation Warfare, Electronic Warfare and Radar Division, Defence Science Technology Organisation (DSTO), Edinburgh, South Australia, Australia

Received: 22 April 2005 / Accepted: 13 July 2005

Abstract. While GPS is a relatively mature technology, its susceptibility to radio frequency interference (RFI) is substantial. Various investigations, including the Volpe Report (Volpe, 2001) which was the result of a US Presidential Decision Directive (PDD-63) assigned to the Department of Transportation (DOT), have recommended that methods should be developed to monitor, report and locate interference sources for applications where loss of GPS is not tolerable. With GPS becoming an integral utility for developed society, the significance of research projects that enhance and expand the capabilities of GPS RFI localisation is highly important.

In response to this requirement for GPS interference localisation, a novel technique called "Inverse Diffraction Parabolic Equation Localisation System" (IDPELS) has been developed. This technique exploits detailed knowledge of the local terrain and an inverse diffraction propagation model based on the Parabolic Equation method (PEM). In wave-propagation theory, an inverse problem may involve the determination of characteristics concerning the source, from field values measured at a certain point or certain regions in space. PEM is an electromagnetic propagation modelling tool that has been extensively used for many applications. This paper will present simulation and field trial results of IDPELS. Simulation results show that this technique has good promise to be useful in locating GPS jamming sources in highly-complex environments, based on networks of GPS sensing antennas. Results also show that the method is capable of locating multiple interference sources. Trials concerning the practical application of IDPELS are also provided. With measured lateral field profiles recorded with a single moving sensor platform in a van, results indicate IDPELS to be a pragmatic localisation technique.

Key words: Parabolic, Inverse Diffraction, Propagation, Localisation

1 Introduction

An area that has received considerable research and development in recent times is the mechanism that discovers the spatial relationship between objects. This process is referred to as *localisation* and has been extensively applied. Areas where localisation can applied include autonomous mobile robot navigation (Adams, 1999), local neural networks (Weaver et al., 1996), E-911 (Biacs et al., 2002) and airborne electronic warfare (EW) systems (Stimson, 1998).

After World War II, there was extensive research and development of radar-directed weapon systems and secure communication systems. This saw EW-based receiving systems, particularly Electronic Support Measures (ESM) evolve to provide the location of an enemies signal's source (Sherman, 2000). The development of such EW intelligence functions occurred for several reasons. One reason for ESM localisation is because the correlation of source location with the electronic order of battle (EOB) can aid in identifying the signal being analysed. Another reason for ESM localisation concerns the ability to assist in real-time threat assessment. Real-time threat assessment provides an increase in situation awareness (SA) for both people and various on-board electronic systems (Vaccaro, 1993). With real-time situational awareness, effective prompt responses can be performed to avoid or minimise the impact of an opponents attack. By having localisation and real-time information, both of these factors will assist the mitigation of hostile electromagnetic aggression against GPS.

GPS which is used as a supplemental means of navigation by avionics systems also has a wide range of other important uses (Spencer et al., 2003). These diverse uses range from network time synchronisation, criminal investigations or even for archaeological discoveries. With the unabated penetration of GPS into civil infrastructure, the possible loss of GPS service could have damaging consequence to users (Baker, 2001). Since the inception of GPS as a supplemental means of navigation, the FAA became an organisation with a primary interest in ensuring the integrity and availability of GPS signals. While early FAA programs focussed on RFI prevention, it latter became clear the GPS would require a significantly greater real-time RFI source localisation capability (Geyer et al., 1999). Other RFI localisation research has shown that a network of sensors can provide an instantaneous estimate in relation to the direction-of-arrival (DOA) of RFI signals (Jan et al., 2001). The continuous real-time estimation ability of a network is a significant benefit compared to a single moving sensor platform. This capability is a substantial factor that should be considered in the design of RFI localisation systems. Consequently a network-centric framework should be chosen for the implementation of this novel localisation technique. For a network to provide highly accurate and real-time position estimation, the configuration of the network will be a significant factor to consider. Investigations concerning the orientation of networks that can be self-configurable or adaptive have also been undertaken (Bulusu et al., 2004).

The importance of GPS availability and it susceptibility to intentional interference has provided the motivation for this interference localisation research. The IDPELS research program has been developed with the objective of ensuring the integrity and availability of GPS signals required by aviation navigation systems and users.

2 Localisation

In performing localisation, there are various approaches that can be undertaken, each of which uses various parameters. Trilateration and Triangulation are two basic geolocation approaches that can be performed with networks. The respective parameters that are required for each of these geolocation approaches are *range* and *direction-of-arrival* (DOA). The estimation of these parameters is found to be a classical process with radar, sonar and geophysical exploration (Vanderveen, 1997).

With trilateration, possible techniques for estimating range can be based on signal strength or the transit time of the signal. Prevalently range based on transit time has been employed due to the greater accuracy that is available. By finding the intersection of three range measurements, the location of the source is able to be unambiguously estimated. Localisation being performed with trilateration and triangulation is graphically presented in Figure 1.

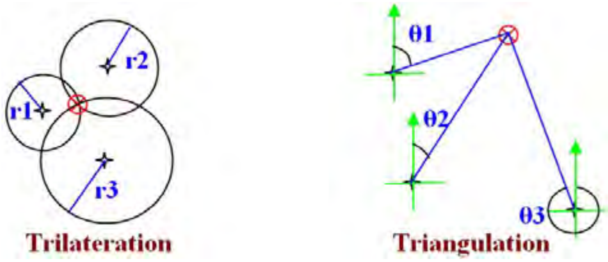


Figure.1 Geolocation using Range and DOA

In a hostile environment, localisation should be performed passively. This is to ensure the enemy can't apply retrospective electronic counter measures (ECM). This means there would only be one-way transmission, i.e. from the interference source. While the signals time of arrival (TOA) is simple to measure, there is no way of knowing when the signal was transmitted. This makes finding the transmission time infeasible. Only cooperative systems such as GPS are able to perform trilateration with one-way transmission.

Triangulation requires the DOA parameter to be used. A network consisting of two sensors is able to estimate the location of the source. Localisation based on DOA has been extensively applied in EW. This is because a hostile emitter can not easily alter the DOA parameter. As a result of the reduced susceptibility to ECM, DOA has become an invariant sorting parameter in the deinterleaving of radar signals for ESM (NAWCWD, 2003). This provides a strong foundation for IDPELS, which can determine the DOA parameters.

The inability of trilateration to resolve the transmission time in a hostile scene can be overcome by using Time Difference of Arrival (TDOA) localisation. The TDOA method requires the difference in a signals arrival time between baseline sensors to be measured. With this measurement, a *line-of-positions* (LOP) indicating where the source can be found is provided. This LOP is known as an Isochrone. The isochrone is an infinite hyperbolic line containing all possible locations where the emitter may be found (Boetcher et al, 2002). Various isochrones, corresponding to different TDOA are displayed in Figure 2. For localisation to be performed with TDOA, multiple baselines are required. The sources location will be at the intersection of the isochrones.

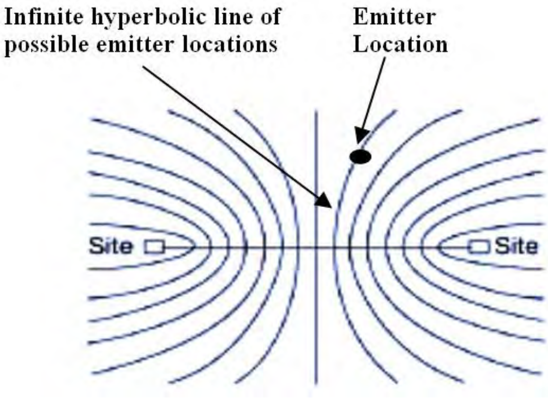


Figure.2 Hyperbolic Isochrone LOP

Another precise localisation techniques based on LOP intersection is the Frequency Difference of Arrival (FDOA) technique (Adamy, 2003). A desirable property of FDOA concerns the dynamics of network sensors. Here a similar difference with TDOA between baseline sensors is found, but with frequency and not time. The result of FDOA is a three dimensional surface defining all possible transmitter locations. By taking a planar cross-section, the curve is called an Isofreq. A set of isofreq curves for various frequency differences is shown in Figure 3, where the baseline sensors are moving in the same direction. As with TDOA, multiple baselines are required for the emitter location to be determined with FDOA.

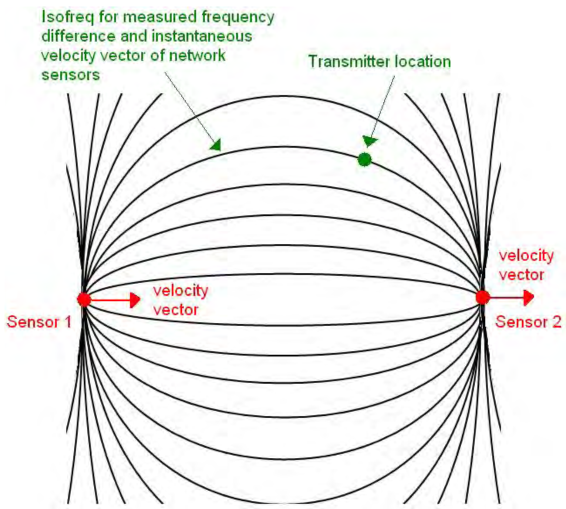


Figure.3 FDOA Isofreq LOP

While FDOA can be based on moving localisation sensors, the computational load associated with moving interference sources will most often be too large. FDOA is therefore generally used only on stationary or slowly moving targets. In practise, localisation systems will typically use multiple platforms. This allows multiple solutions to be considered. A system that combines TDOA and FDOA measurements can determine the precise localisation of an emitter location with a single baseline, which is displayed Figure 4. The multiplicity of combined TDOA and FDOA solutions produces more accurate results over a wider range of operational conditions. IDPELS could enhance current localisation systems by providing multiple solutions.

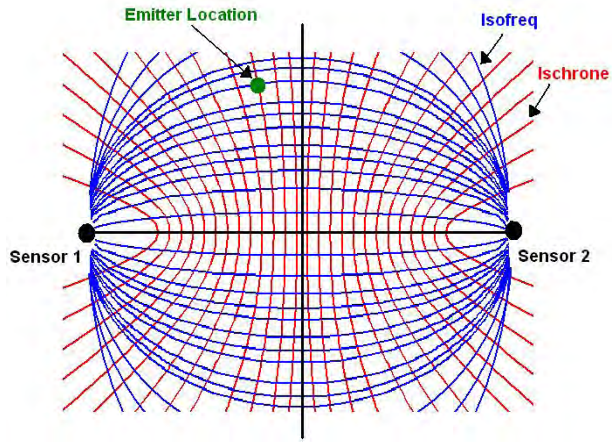


Figure.4 Single Baseline Localisation

There are also localisation techniques that are intended for an urban environment. Intended for picocell and microcell multipath scenarios, the database correlation method (Wolfe et al., 2002) compares a signals path-loss with a look-up table. Depending on the urban layout, the workload for adequate resolution in the look-up tables could be considerable. While any technique that contributes to interference signal localisation in an urban environment should be considered valuable, this database correlation method is not functional in hostile scenarios. In urban EW, there is no method to determine the hostile interference transmission power level. As a result, no path-loss calculations can be made. This renders the database method unsuitable for RFI localisation in an urban EW scene.

With IDPELS localisation being based on the DOA parameter, several common DF techniques will be briefly reviewed. The simplest DF method uses amplitude comparison and a mechanically rotated narrow-beam antenna. While highly accurate DF can be yielded, the probability of signal interception is relatively low (Tsui, 1986). To overcome this low probability of interception (LPI), an array can be configured to provide 360° coverage. This coverage is displayed with a four-quadrant amplitude DF system in Figure 5.

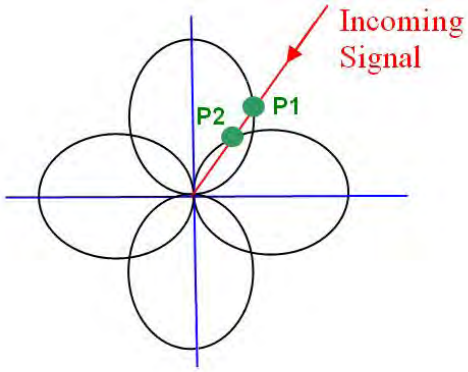


Figure.5 Monopulse DF System

By identifying the greatest (P1) and second greatest (P2) received power levels, the DOA can be determined. While amplitude comparison systems are frequency independent and able to cover wide bandwidths, the DOA estimation has a high probability of being contaminated by multiple signals simultaneously received. These systems also require calibration with signals that have known DOA information. Another common DF technique employed in EW is Phase interferometry shown in Figure 6.

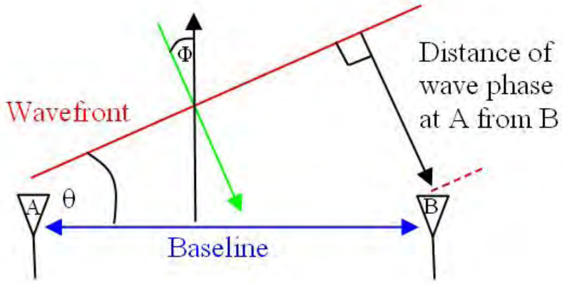


Figure.6 Phase Interferometry

Application of interferometry is however restricted to narrow-band signals. By measuring the phase difference between baseline sensors, the DOA can be determined via trigonometry. In most interferometric systems, the baseline is between 0.1 and 0.5 λ. A baseline less than 0.1 λ does not provide enough accuracy and if over 0.5 λ ambiguous results are provided. There are many other DF finding techniques that could have been analysed, however a tutorial of many existing DOA estimation methods is provided by Godara (1996). A special class of these DF techniques that has high-resolution capabilities will conclude this discussion on localisation techniques.

The high-resolution DF methods are the subspace class of spectral estimation techniques that determine a signals DOA by computing the spatial spectrum and finding the local maximas of the spectrum. Subspace techniques require the noise and signal subspace to be extracted from the covariance matrix of signal observations. Eigen-analysis can be used on symmetrical matrices, or Singular value decomposition (SVD) can be applied with asymmetric matrices. Both of these techniques will elliptically fit the observed covariance matrix (Therrien, 1992). Subspace based DF methods have the ability to surpass the limiting behaviour of classical Fourier-based DF methods. They are also referred to as super-resolution algorithms. The first subspace method was developed by Pisarenko (1973), which is referred to as Pisarenko Harmonic Decomposition (PHD). It should be noted that PHD does not directly estimate DOA, instead it determines frequency and power of real sinusoids in additive white noise. PHD is based on Caratheodory`s theorem which is an indication of the required data-set size for dynamics of desired parameters to be captured (Sidiropouls, 2001). The extension of PHD to DOA estimation was made by Schmidt (1982) with Multiple Signal Classification (MUSIC). One of the limitations associated with MUSIC is that the number of sensors must be greater than the number of signals present. The Joint Angle and Delay Estimation (JADE) method developed by Vanderveen (1997) can overcome this limitation, provided that signal fading is constant. JADE is based on multiple channel estimates and is best suited for TDMA systems where training signals are available. While blind estimation is possible with JADE, it is considered to have an undesirable intensive computational load. The simplicity of IDPELS in comparison with JADE in performing localisation is significant.

3 Research Objectives

From the localisation methods previously discussed, there are different limitations associated with each. These limitations range from the jammer/sensor dynamics to intensive computational loads. As noted with the combined TDOA/ FDOA method, multiple solutions and simplicity is the ultimate goal of localisation methods. The primary objectives of the IDPELS development were twofold;

- To investigate if an inverse EM propagation model can be used to provide an accurate localisation solution.
- To determine if an improved localisation can be made if detailed knowledge of the local terrain is known.

The new solution can be combined with other methods to provide multiple localisation solutions. Where networks already exist, the integration of IDPELS is also intended to be relatively simple, provided the receiving sensors are available. All that is required for the extra solution is the software for inverse diffraction propagation. The application of IDPELS with a moving single sensor is also a possible configuration option if a network configuration in not possible.

4 Methodology

The principle methodology of IDPELS is based on applying inverse diffraction to the Parabolic wave equation propagation model (PEM). Fundamentally IDPELS involves measuring a received signal profile using a large antenna array or from a moving vehicle. This received field profile is then inverse propagated over the terrain profile. The source location is then identified by determining the field convergences. This classifies IDPELS as an inverse problem. With the development of IDPELS being based on the PEM and inverse theory, these two areas will be briefly reviewed.

The classification of inverse problems was defined by Keller (1976). Keller defines two problems as *inverses* of one another if the formulation of each involves all or part of the solution of the other. One of the problems has been extensively studied (forward problem), while the other is not so well understood (inverse problem). From a mathematical perspective, the decision of what is direct or inverse can be arbitrary. However in reference to physics (i.e. astronomy, mechanics, geophysics, wave propagation, etc) the decision of which problem is forward or inverse is not as arbitrary. Turchin et al. (1971) define forward problems in the physics domain as a process that is oriented along a cause – effect sequence. A corresponding inverse problem is associated with the reversed, effect – cause sequence. This means that a forward problem involves determining what observation will be made, given various parameters of the systems. An inverse problem will determine the unknown parameters of the system, from the observation made with respect to the system. Another important link that should be considered with forward and inverse problems is the model identification problem (Aster et al., 2004). The combination of these three factors of *Inverse Theory* is shown in Figure 7.

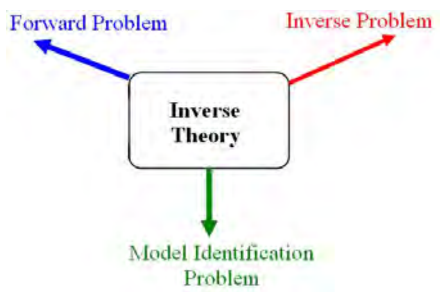


Figure.7 Inverse Theory

Models concerning the physical properties or processes of the systems are generally already known. Over history there have been many mathematicians and physicists who discovered and identified models for many different systems. Various examples of such people include Gauss, Faraday, Maxwell and Einstein. The model used for IDPELS is the numerically efficient PEM which has been extensively researched and developed (Lee et al., 2000). With wave-propagation theory, a possible forward problem could be the computation of a field radiated by the source. A corresponding inverse problem could involve the determination of the source position from the knowledge of the radiated field. The solution of this inverse problem is the intended function of IDPELS, where the applied model is PEM.

The use of inverse theory has been extensively applied in imaging. The X-ray computer tomography (CT) technique developed in 1971 (Hounsfield, 1973) is the first case of medical images obtained as a process involving inverse problems. Other topics that have applied inverse theory include atmospheric sounding, particle scattering or seismology. One inverse problem that is similar to IDPELS is sonar-based and was studied by Zhu (2001). This sonar research was concerned with image reconstruction by back propagating the PEM model with a focus-marching procedure.

PEM being the model employed by IDPELS, was originally proposed by Mikhail Aleksandrovich Leontovich (1944) for long range radio propagation. In 1946, Leontovich together with Fock (1946) were able to provide a planar and spherical PEM solution. PEM involves approximating the elliptic Helmholtz wave equation with a parabolic partial differential equation to reduce the difficulties experienced in obtaining a Helmholtz solution. After the original development of PEM, application of PEM remained significantly restricted till the 1970`s when computer technology had advanced to allow numerical solution to be developed. With advances in computer technology, Frederick D. Tappert and R. H Hardin (1973) introduced the parabolic approximation to oceanic acoustic propagation with the efficient Split-Step method that can propagate a signal with the Fast-Fourier-Transform (FFT). Claerbout (1976) latter developed a finite difference PEM version for geophysical applications. Eventually PEM returned to radio propagation where propagation over a littoral environment (i.e. sea or flat terrain) was initially considered. With the development of faster algorithms, Kuttler and Dockery (1991) were able to adapt the split-step method (developed by Tappert) for radio propagation. Further application of PEM was made possible with researchers such as Barrios (1994), who tested the Tappert approach on a variety of irregular terrain profiles. Walker (1996) extended PEM for use in GPS propagation studies. Because radio domains are generally large with respect to the wavelength, approximation of Maxwell's solution must be made. With the efficiency and accuracy provided by PEM, it has largely superseded geometric optics and mode theory in achieving the approximations. Current PEM applications are wide ranging.

Having provided a brief historical background to PEM, the process of applying inverse diffraction within PEM will now be provided. While finite difference or the Fourier split-step technique (FSS) can be used to solve the PEM, discussion will be restricted to FSS as this was the method employed in the code. One of efficiencies offered by PEM over other propagation models is that it is an open boundary problem. The numerical solution for an elliptical wave equation requires all boundary condition to be specified. This situation is not required by PEM. With FSS, the forward propagation provided by PEM involves marching an input field profile as shown in Figure 8.

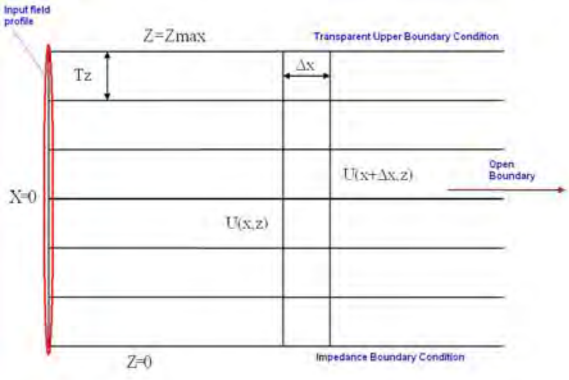


Figure.8 Open Boundary PEM Marching

The initial field profile is transformed into the angular spectrum via the fast Fourier transform (FFT). This angular spectrum is also referred to as the vertical spatial frequency spectrum as it involves the vertical component of the wavenumber. In the angular spectrum a propagator is multiplied with the transformed field profile, which effectively propagates the field to the next marching step. By inverse transforming the angular spectrum that has been propagated, the field at x+Δx is able to be determined. The propagator term is also referred to as the Diffraction function, which is multiplied with the angular spectrum. The actual equation employed in PEM may slightly vary depending of factors that are considered in the model. One example could be to account for the atmospheric refractive index. A propagator that has been employed by Hawkes (2003) is shown in Eq. (1) and is based of the Fourier imaging domain method suggested by Eibert (2002).

$$D(p) = \exp\left\{ j\Delta x\left(\sqrt{k^2 - p^2}\right)\right\} \qquad (1)$$

where,

k is the spatial frequency spectrum

p is the vertical spatial frequency spectrum

Δx is the distance covered in a propagation step

With the previously discussed forward propagation problem, the diffraction term must be multiplied with the angular spectrum. A high level equation representing this forward propagation is provided by Eq. (2).

$$u(x + \Delta x) = T^{-1}(U \times D) \qquad (2)$$

where,

u is the envelope function of the signal

U is the angular spectrum of the signal

(i.e. FFT of u)

As IDPELS intends to apply inverse diffraction with back propagation in order to resolve the location of the source, it will divide the diffraction term with the angular spectrum. A high level equation representing inverse propagation is provided in Eq. (3).

$$u(x - \Delta x) = T^{-1}(U \div D) \qquad (3)$$

Simulation results of inverse propagation with IDPELS are provided in the following section. The final factor that will be discussed with respect to PEM and IDPELS is associated with their upper boundary condition. To ensure there is no reflection of signal from the upper boundary, a windowing function must be applied. With IDPELS, the propagation domain height was doubled for application of the window. Figure 9 provides a display of the gradual signal attenuation in the window domain.

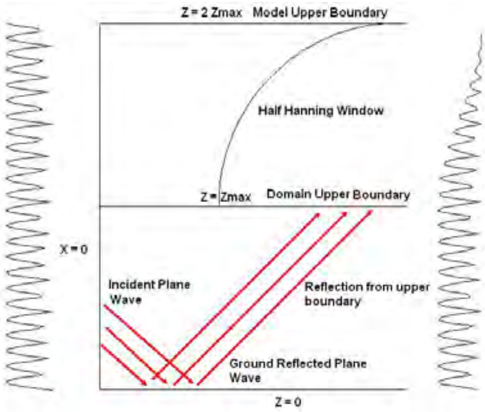


Figure.9 Hanning Window

The chosen window for PEM and IDPELS was the Hanning window. This note is important as it must be considered when viewing the IDPELS display of Figure 13. Further information concerning radio propagation with PEM, is provided by Levy (2000).

5 Simulation Results

Simulation results will be presented to demonstrate the theoretical feasibility of IDPELS to perform geolocation. When analysing IDPELS under simulation, the generation of a forward propagation field with PEM is a prerequisite. The first example is a simple demonstration of IDPELS where the terrain profile is a single block with a height and width of 20m as shown in Figure 10. The transmission source is chosen to be on the far left-hand side of the block, and 20m above the block. A range of 100m was chosen for field analysis.

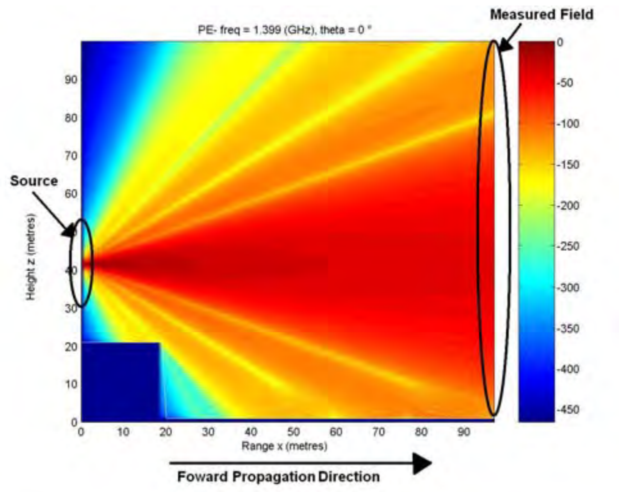


Figure.10 PEM – Block

The corresponding IDPELS result is shown in Figure 11.

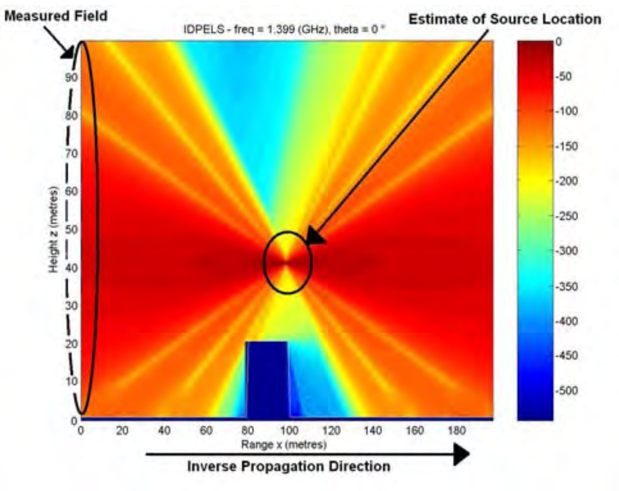


Figure.11 IDPELS – Block

The geolocation capability of IDPELS is clearly demonstrated in this simple scenario with unobstructed line-of-sight (LOS) paths. Where the inverse propagated field acutely converges, this is a highly accurate estimation of the sources locations. It should be noted that the inverse propagation range has been extended an extra 100m. It's also important to recognise that Figure 11 is a reversed view of the forward propagated field. This means the input field profile for IDPELS being on the left-hand side of Figure 11, is the same field profile located on the far right-hand side of Figure 10. This reversed view is present in all other simulated IDPELS displays. The next demonstration of IDPELS is with respect to a wedge. In this evaluation of IDPELS, the source was chosen to be located 20m above the floor of the domain. A display of the forward propagated field is provided in Figure 12.

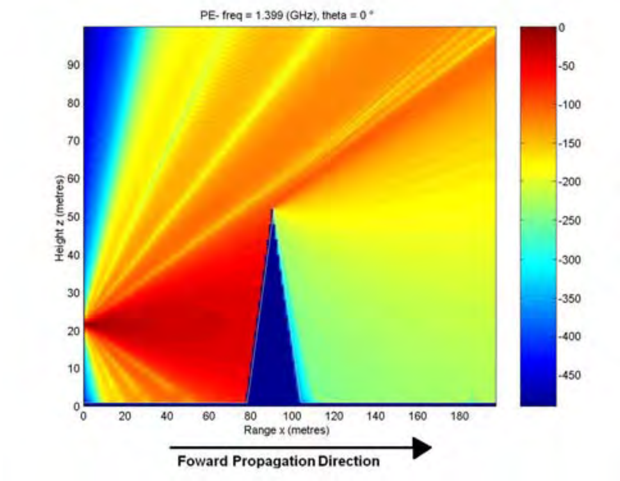


Figure.12 PEM – Wedge

The corresponding IDPELS result is shown in Figure 13.

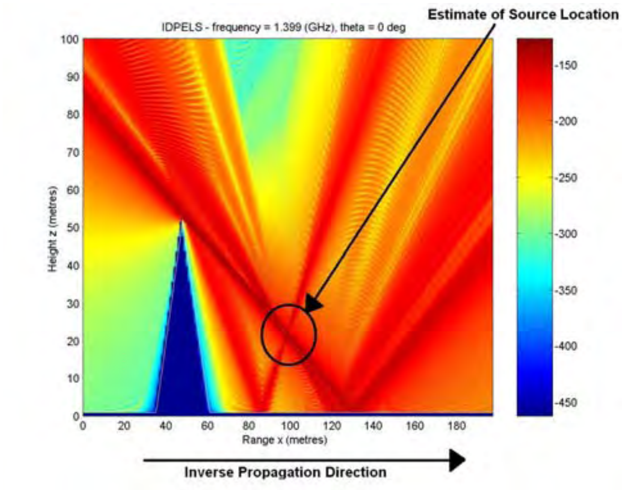


Figure.13 IDPELS – Wedge

This scenario was investigated to consider the feasibility of IDPELS when a *non-line-of-sight* (NLOS) will exist with inverse diffraction propagation. It should be noted that the position of the source at 20m height still allows LOS paths above 88m on the right-hand side of Figure 12. Measuring antenna elements will be required to be positioned at heights greater than 88m in this scenario. With the input IDPELS field profile corresponding to the right-hand of Figure 12, accurate localisation is again provided by IDPELS as indicated by the intersection of the ground reflected beam with the downward directed beam originating from the left-hand side of Figure 13. Please note that visual interpretation must be currently made to determine the location of the source.

With Figure 13 using an input field profile analogous to principles associated with Synthetic-Aperture-Radar (SAR), a further investigation was made with a reduced set of antenna measurements to examine how a network configuration will affect localisation results. A 9-element uniform-linear-array (ULA) configuration is applied to the measure field profile from the right-hand side of Figure 12. The corresponding IDPELS field is shown in Figure 14.

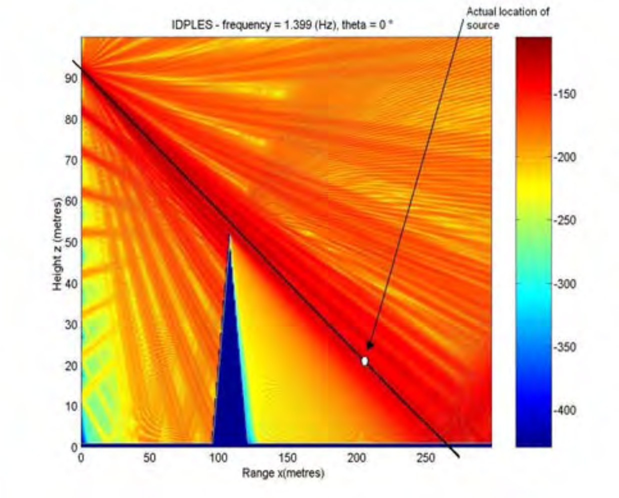


Figure.14 Network Configuration

With the array configuration, there is no definite indication of the interference source location. Only a LOP is provided by the sensor that experienced a LOS to the source. It should also be noted that the sensors with a NLOS to the source did converge to the apex of the wedge that shadows the source. This indication can provide assistance for localisation being conducted in an urban environment.

The next evaluation of IDPELS was to consider its effectiveness against multiple interference sources. A display of three sources simultaneously transmitting interference signals is provided in Figure 15.

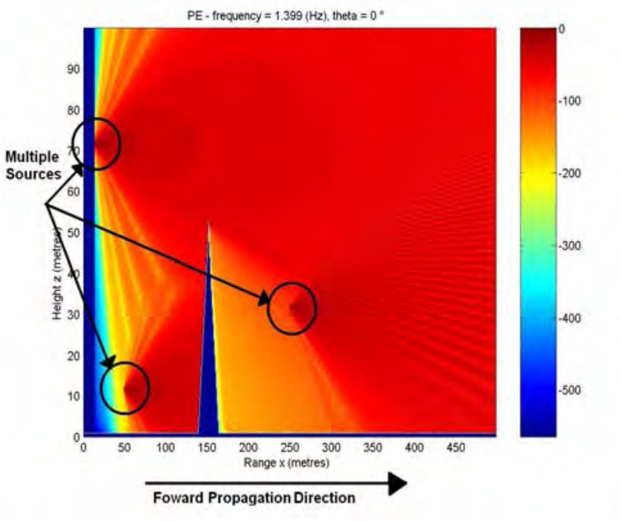


Figure.15 PEM – Multiple Sources

With multiple sources, one source was positioned to be completely obstructed by the wedge. The IDPELS input has no account of this source. Figure 16 shows the IDPELS field generated for the multiple sources in Figure 15.

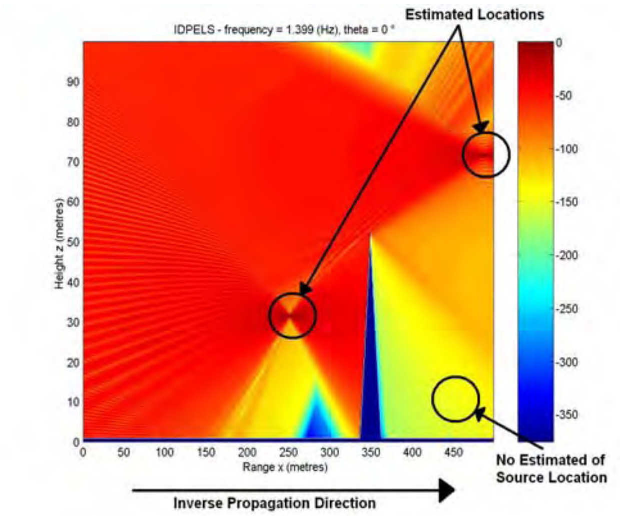


Figure.16 IDPELS – Multiple Sources

All sources with a LOS where able to be localised. This is indicated by the field convergence at their relative positions in Figure 16. The source that did not have a LOS was not able to be localised.

While the geolocation feasibility of a network based IDPELS was not demonstrated in Figure 14, this was due to the NLOS orientation of the source. A demonstration of IDPELS functionality with an array configuration of two antenna array elements is provided in Figure 17.

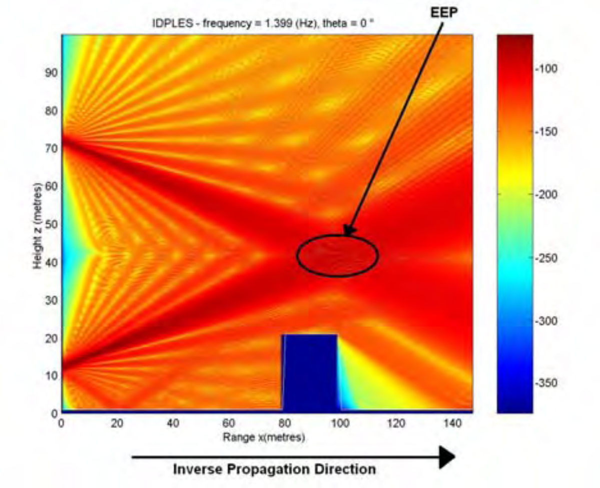


Figure.17 Two-element array configuration

The accuracy of the estimated source location in Figure 17 is subject to a large elliptical-error of probability (EEP) compared with Figure 11. This localisation error can however be reduced according to the array configuration. Factors that govern the localisation error are,

- number of sensors used
- sensor aperture

The localisation error is reduced by increasing either of these two factors. Figure 18 shows this affect where there is an increase in the number of field sensors used, all of which have a relatively larger aperture.

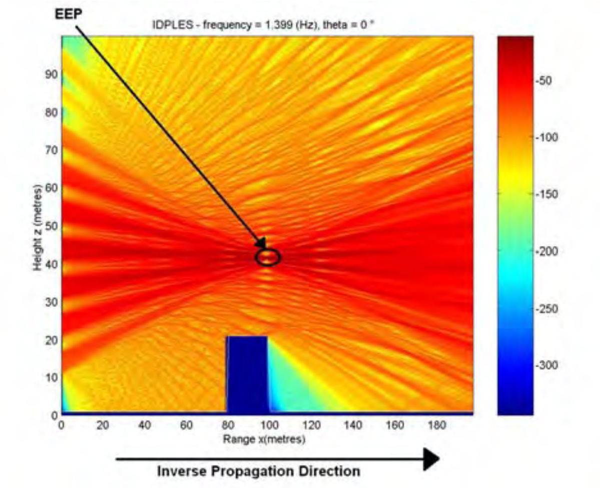


Figure.18 Reduced Localisation Error

6 IDPELS Field Trials

To test the practical application of IDPELS, field trials were conducted in collaboration with the Navigation Warfare, Electronic Warfare and Radar Division of DSTO, Edinburgh, South Australia. The transmission source was a 1.399GHz tone signal being transmitted from a helix antenna as shown in Figure 19.

Figure.19 Helix Transmission Antenna

Two sets of data were collected. One data set concerns the transmission source approximately 13km east of Truro, SA (34°25’2.85” S, 139°14’10” E) at the base of the Mt Lofty Ranges (Figure 20). The other data set has the transmission source positioned at DSTO Radio Research Station (34°43’26.2” S, 138°32’15.6” E) at St Kilda, SA (Figure 21).

Figure.20 Mt Lofty Base Trials

Figure.21 St Kilda Trials

The input field profiles for IDPELS were recorded based on the SAR analogy. An overview of the signal recording process is shown in Figure 22.

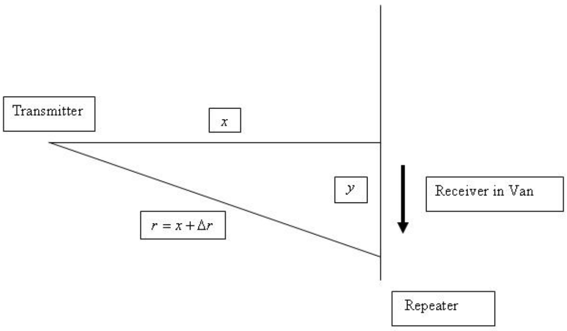


Figure.22 Signal recording process

With Truro data sets, Figure 23 is a display of the IDPELS field where the signal was recorded in a moving van approximately 4.8kms from the transmission site on Baldon road. Figure 24 corresponds to data recorded on Woolshed road, approximately 5.9kms from Baldon rd.

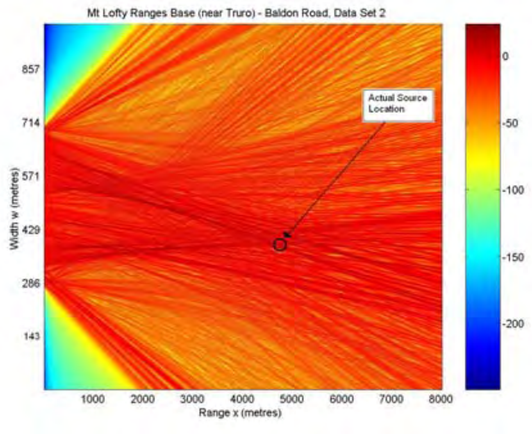


Figure.23 Signal Recorded on Pine Creek Track

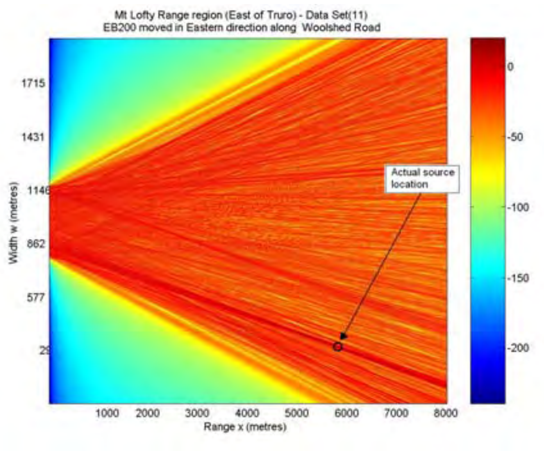


Figure .24 Signal recorded on Woolshed Road

These IDPELS results have not provided a solution as accurate in comparison with simulation results. While data recorded on Pine Creek Track has shown a clear convergence region, Woolshed road data only provided a LOP. Various causes for the solution degradation include noise, clutter, multipath factors and a non-linear phase shift in the recorded signal. A factor that will have contributed to a non-linear phase shift is the road section not being perfectly straight. While scattering and reflection will also have had some impact on the recorded signal, modelling of obstacles was not incorporated into the prototype IDPELS code. This is because the selected region was considered to approximate a littoral environment. A photo of the general terrain profile at the base of the Mt Lofty ranges is shown in Figure 25.

Figure.25 Littoral Mt Lofty Base Region

A photo of the McEvoy road section used to record the test signal is shown in Figure 26. The displayed repeater was used to account for Doppler shift generated by the movement of the van.

Figure .26 McEoy Road

The displayed IDPELS result concerning McEvoy road shown in Figure 27 has similar visible field convergence to Pine Creek Track. The range of McEvoy rd from the St Kilda Radio Research Station is approximately 3.9kms.

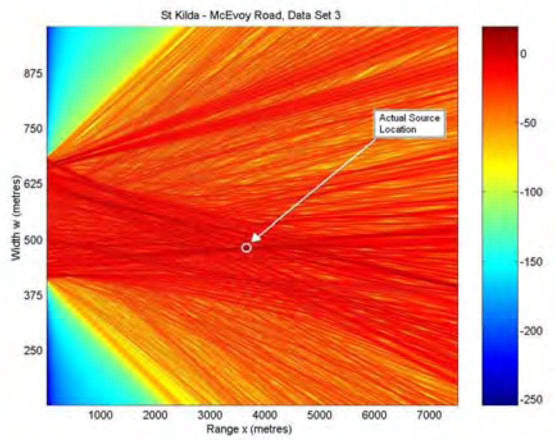


Figure.27 McEvoy Road

The IDPELS field profile corresponding to data recorded on Pt Gawler road is shown in Figure 28. The range to Pt Gawler approximated 10.8kms.

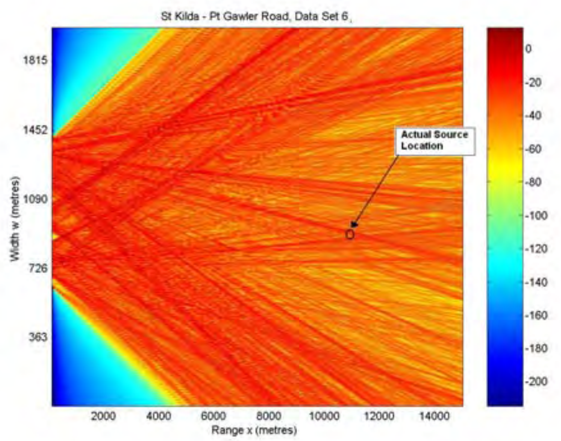


Figure.28 Pt Gawler Road

The localisation result for Pt Gawler road has many field convergent regions. This demonstrates Rayleigh fading where there is no dominant wave component. The terrain profile for this region was also considered to be semi-urban. Van speeds were also greater compared with all data sets. The general driving pattern was initiated with a steady acceleration and maintain at a constant speed. Near the end of the recording session, a steady de-acceleration was applied to being completely stationary. Speeds reached on Pt Gawler road varied between 80 – 100 km/hr, while all other data sets varied between 10 – 30 km/hr.

Conclusion

The simulation results of IDPELS has demonstrated that inverse diffraction propagation is capable in providing a geolocation estimate to multiple sources that have a direct LOS to network sensors. While IDPELS was unable to geolocate a source that only has a NLOS, it could indicate the direction to objects that shadow the source. This could be beneficial in an urban environment. A network configured IDPELS was also shown to improve accuracy with the number of sensors, and sensor aperture.

Field trial results demonstrating the practical feasibility of inverse diffraction propagation were based on a SAR analogy for generation of the input field profile. Trials conducted in regions that approximated a littoral environment indicated the method to be feasible. The trials however also showed that great care must be taken to ensure the phase-shift in the recorded signal profile is linear. Other factors such as multipath propagation and noise also degraded localisation accuracy. It should also be noted that given the experimental nature of the trials, experienced conditions and measurements conducted were not ideal.

For localisation to be performed with novel *inverse diffraction* propagation methods, further research and development is required for an efficient localisation method to be readily available and operational.

References

Adams M. D. (1999): ***Sensor Modelling, Design and Data Processing for Autonomous Navigation***. World Scientific Publishing. Singapore.

Adamy D. (2003): ***Precision Emitter Location by FDOA***. Journal of Electronic Defense – EW 101, January.

Aster R.C.; Borchers B.; Thurber C. (2004) ***Classification of Inverse Problems***. in: Parameter Estimation and Inverse Problems, Academic Press, New Mexico Technology, 1-4.

Baker C. (2001): ***GPS Vulnerability to Interference***. SatNav 2001, Canberra, ACT, Australia.

Barrios A. E. (1994): ***A terrain parabolic equation model for propagation in the troposphere***. Antennas and Propagation, IEEE Transactions on, 42(1): 90-98.

Biacs Z.; Marshall G.; Moeglein M.; Riley W. (2002): ***The Qualcomm/SnapTrack Wireless-Assisted GPS Hybrid Positioning System and Results from Initial Commercial Deployments***. ION GPS 2002, Oregon Convention Center, Portland, Oregon.

Boettcher P.; Shaw G.; Sherman J. (2002): ***A Distributed Time-Difference of Arrival Algorithm for Acoustic Bearing Estimation***. SensIT PI Meeting, Santa Fe, New Mexico.

Bulusu N.; Heidemann J.; Estrin D.; Tran T. (2004): ***Self-Configuring Localization Systems: Design and***

Experimental Evaluation. *ACM Transactions on Embedded Computing Systems (ACM TECS),* 3(1): 24-60.

Claerbout J. F. (1976): ***Fundamentals of Geophysical Data Processing with Application to Petroleum Prospect***. McGraw-Hill. New York.

Eibert T. F. (2002): ***Irregular terrain wave propagation by a Fourier split-step wide-angle parabolic wave equation technique for linearly bridged knife-edges***. Radio Science, 37(1).

Enge P.; Jan S-S. (2001): ***Finding Source of Electromagnetic Interference (EMI) to GPS Using a Network Sensors***. ION National Technical Meeting Proceedings, Westin Hotel, Long Beach, California.

Geyer M.; Frazier R. (1999): ***FAA GPS RFI Mitigation Program***. ION GPS-99, 12th International Technical Meeting of the Satellite Division of the Institute of Navigation, Nashville Convention Center, Nashville, Tennessee.

Godara L. C. (1996) ***Limitations and capabilities of direction-of-arrival estimation techniques using an array of antennas: A mobile communication perspective***. IEEE International Symposium on Phased Array Systems and Technology, Boston, Massachusetts, USA, 327 – 333.

Hardin R. H.; Tappert F. D. (1973): ***Applications of the split-step Fourier method to the numerical solution of nonlinear and variable coefficient wave equations***. Society of Industrial and Applied Mathematics (SIAM), 15(423).

Hawkes R. M.; Baker C. P. (2003): ***Tropospheric Propagation Model for Land Warfare***. Land Warfare Conference 2003, Adelaide Convention Centre, Adelaide, Australia.

Hounsfield G. N. (1973): ***Computerized transverse axial scanning (tomography). Part I: Description of system. Part II: Clinical applications'***. British Journal of Radiology, 46(1016-1022).

Keller J. B. (1976): ***Inverse Problems***. American Mathematical Monthly, 83(107 – 118).

Kuttler J. R.; Dockery G. D. (1991): ***Theoretical description of the parabolic approximation/Fourier split-step method of representing electromagnetic propagation in the troposphere.*** *Radio Science*, 26(2): 381-393.

Lee D.; Pierce A. D.; Shang E-C. (2000): ***Parabolic Equation Development in the Twentieth Century***. Journal of Computational Acoustics, 8(4): 527-637.

Leontovich M. A. (1944): ***A method of solution of problems of electromagnetic wave propagation along the earth`s surface***. Bulletin de l` Academie des Sciences de l`URSS, ser. phys., 8(1): 16.

Leontovich M. A.; Fock V. A. (1946): ***Solution of the Problem of Propagation of Electromagnetic Waves along the Earth's Surface by the Method of Parabolic Equations***. Journal of Physics of the USSR (In English), 10(1): 13-24.

Levy M. (2000): ***Parabolic Equation Methods for Electromagnetic Wave Propagation***. London, Institution of Electrical Engineers. 352.

NAWCWD (2003): ***Signal Sorting Methods and Direction Finding***. in: Electronic Warfare and Radar Systems Engineering Handbook, Avionics Department of the Naval Air Warfare Center Weapons Division, Point Mugu, California, 5-8.1 - 5-8.9

Pisarenko V. F. (1973): ***The Retrieval of Harmonics from a Covariance Function***. Geophysical Journal of the Radio Astronomical Society, 33(347-366).

Schmidt R. O. (1982): ***A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation***. Ph.D. dissertation. Stanford University, Stanford, California.

Sidiropoulos N. D. (2001): ***Generalizing Caratheodory`s Uniqueness of Harmonic Parameterization to N Dimensions***, IEEE Transaction on Information Theory, 47(4): 1687 – 1690.

Spencer T. A.; Walker R. A. (2003): ***A Case Study of GPS Susceptibility to Multipath and Spoofing Interference***. Australian International Aerospace Congress incorporating the 14th National Space Engineering Symposium 2003, Brisbane, Queensland, Australia.

Stimson G. W. (1998): ***Electronic Counter Counter Measures (ECCM)***. in: *Introduction to Airborne Radar*. SciTech Publishing, Inc, Mendham, New Jersey, 457 – 468.

Therrien C. W. (1992): ***Discrete Random Signals and Statistical Signal Processing***. Englewood Cliffs. New Jersey, Prentice Hall Signal Processing Series.

Tsui JB-Y. (1986): ***Parameters Measured by EW Receivers***. in: Microwave Receivers with Electronic Warfare Applications, John Wiley & Sons, New York, 81 – 112.

Turchin V. F.; Kozlov V. P.; Malkevich M. S. (1971): ***The use of mathematical-statistics methods in the solution of incorrectly posed problems.*** Soviet Physics Uspekhi, 13(681-840).

Vaccaro D. D. (1993): ***Passive Direction Finding and Geolocation***. in: Electronic Warfare Receiving Systems, Artech House, Boston - London, 221- 259.

Vanderveen M. C. (1997): ***Estimation of Parametric Channel Models in Wireless Communication Networks***. Ph.D. dissertation. Scientific Computing and Computational Mathematics, Stanford University, Stanford.

Volpe National Transport Systems (2001): ***Vulnerability Assessment of the Transportation Infrastructure Relying on the Global Positioning System***. Office of the Assistant Secretary for Transportation Policy, U.S. Department of Transport, 29 August.

Walker R. A. (1996): ***The Operation and Simulation of GPS Positioning in Harsh Environments***. Electrical and Electronic Systems Engineering, Queensland University of Technology, Brisbane.

Weaver S. E.; Baird L. C.; Polycarpou M. M. (1996): ***An Analytical Framework for Local Feedforward Networks***. International Symposium of Intelligent Control, Dearborn, Michigan.

Wolfle G.; Hoppe R.; Zimmeramnn D.; Landstorfer F. M. (2002): ***Enhanced Localization Technique within Urban***

and Indoor Environments based on Accurate and Fast Propagation Models. European Wireless, Florence, Italy.

Zhu D. (2001): ***Application of a Three-Dimensional Two-way Parabolic Equation Model for Reconstructing Images of Underwater Targets***. Journal of Computational Acoustics, 9(3): 1067-1078.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 258-267

Journal of Global Positioning Systems
2005

Null-steering LMS Dual-Polarised Adaptive Antenna Arrays for GPS

W C Cheuk, M Trinkle & D A Gray

CRC for Sensor Signal and Information Processing, Dept of Electrical and Electronic Engineering, University of Adelaide, Adelaide SA
e-mail: dgray@eleceng.adelaide.edu.au Tel: + 61 83036425; Fax: +61 83934369

Received: 27 November 2004 / Accepted: 12 July 2005

Abstract. The implementation of a null-steering antenna array using dual polarised patch antennas is considered. Several optimality criterion for adjusting the array weights are discussed. The most effective criteria minimises the output power of the array subject to maintaining a right hand circular polarisation (RHCP) response on the reference antenna. An unconstrained form of this criteria is derived, in which the reference channel is the RHCP output of the reference antenna and the LHCP output of the reference antenna is included as one of the auxiliary channels. An FPGA implementation of the LMS algorithm is then described. To prevent weight vector drift a variant of the circular leakage LMS algorithm was used. The implementation details of a simplified circular leakage algorithm more suited to an FPGA implementation are presented. This simplified leakage algorithm was shown to have a similar steady state weight vector as the full algorithm.

Key words: GPS, Polarisation, Interference Mitigation, Adaptive filters.

1 Introduction

A GPS receiver is relatively susceptible to interference and a number of antenna and signal processing techniques have been investigated to overcome this deficiency. These include :

- Fixed antenna enhancements: Manz (2000), Kunysz (2000).
- Single Channel Adaptive Filters: Dimos et al (1995), Trinkle and Gray (2001).
- Adaptive Beamformers: Trinkle and Gray (2001), Zoltowski and Gecan (1995), Jian et al (1998), Fante and Vaccaro (2000).
- Polarisation Cancellers: Brassch et al (1998), Trinkle and Cheuk (2003).
- Adaptive Beamformers with Polarisation Diversity: Nagai et al (2002), Fante and Vaccaro (2002), Compton (1981).
- Modifying the tracking loop of the GPS receivers: Manz et al(2000),Legrand et al (2000).
- Integrating GPS and INS sensors: Soloviev and van Grass (2004).

In this paper combined spatial and polarisation null steering are considered.

A single GPS antenna with an adaptive polarisation response can be used to reject interferences with polarisations other than the GPS signal. The GPS signal uses RHCP (right hand circular polarisation), i.e., it contains both horizontal and vertical polarizations 90 degrees out of phase; thus it is possible to remove any linearly polarised signal with only a 3 dB loss in GPS signal power see e.g. Brassch et al (1998). The key idea of polarisation cancellers is thus to reject the interference by adaptively mismatching the polarisation response of the antenna to the polarisation of the interfering signal. Adaptive polarisation antennas can be implemented by adaptively combining the two outputs of a dual polarised patch antenna see e.g. Brassch et al (1998), Compton (1981). Clearly this technique becomes ineffective if the interference has the same polarisation as the GPS signal (i.e., RHCP).

Single antenna polarisation cancellers operate purely in the polarisation domain, while adaptive beamformers operate only in the spatial (angle domain). It is possible to combine these two domains in a single algorithm by applying beamforming techniques to an antenna array with dual polarised antenna elements. Using a such an array, will approximately double the degrees of freedom available for interference cancellation compared with using a conventional antenna array see e.g. Jain (1998), Trinkle and Cheuk (2003), Nagai et al (2002), and it has recently been shown in Jian et al (1998) and Fante and Vaccaro (2002) that GPS polarimetric antenna arrays can cancel more interferences than a standard antenna array with the same number of antenna elements. Thus polarisation diversity can be used to significantly reduce the size of GPS antenna arrays, which is important in many applications.

Field Programmable Gate Array, FPGA, technology has lately become an attractive alternative for the implementation of a wide range of DSP applications because of its flexibility and speed. An FPGA allows a large number of multipliers and accumulators to be configured and inter-connected in such a way as to suit a particular algorithm. This fine-grain parallelism allows most adaptive beamforming techniques to be implemented much more efficiently than on a standard DSP. Algorithms implemented on the largest of the current generation FPGAs can achieve several hundred Giga Operations per Seconds (GOPs)

The FPGA implementation of an adaptively null-steering dual-polarised antenna array is proposed. The novelty of this approach lies in

- The choice of the reference and auxiliary channels in the power minimisation algorithm.
- The use of a modified version of the complex LMS algorithm to minimise computations.
- The use of Hilbert transforms to generate the analytic signal.
- The use of the circular leakage LMS algorithms rather than the standard leaky LMS to prevent weight vector drift.
- A simplification of the circular leakage LMS algorithm more suited to FPGA implementation.

2 Digital Null-Steering Structure

System Architecture

A polarimetric adaptive antenna array is placed in front of the GPS receiver. Each element of the array is a patch antenna with two outputs corresponding to the horizontal and vertical polarizations of the received signal. See Figure 1 for a four channel digital null-steering system using two dual-polarized antenna elements.

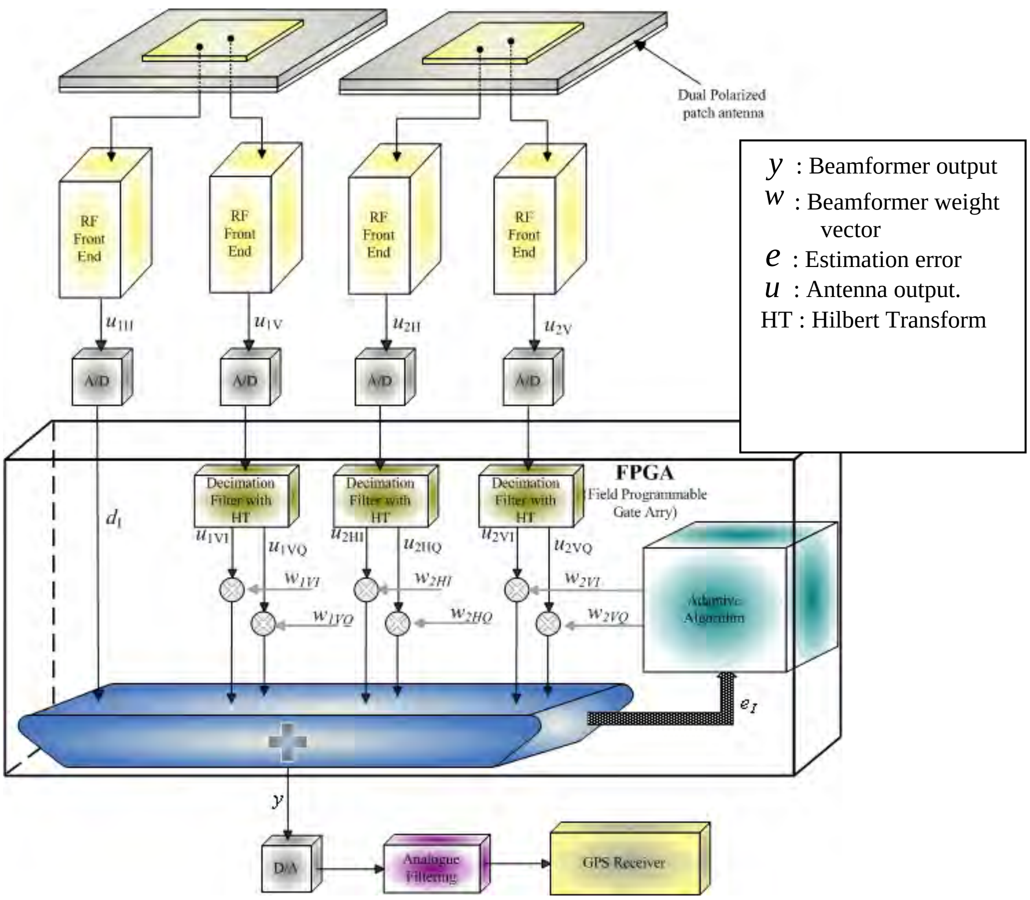


Figure 1. Basic structure of digital null-steerer using a polarimetric antenna array

In this example the horizontal component of the first antenna is used as a reference channel as originally proposed in Fante and Vaccaro (2002). The subscript $_H$ denotes the horizontal polarization, $_V$ denotes the vertical polarization, $_I$ denotes the real part of the complex signal, and $_Q$ denotes the imaginary part of the complex signal.

The output of each polarization channel is converted into a digital signal using an analog to digital (A/D) converter. Since the L1 band GPS signal is transmitted at 1.575GHz and most A/D converters do not have an input bandwidth large enough to sample the GPS signal, a down-converter is used to convert the incoming GPS signal to a lower intermediate frequency. The digital signal from the A/D converter is then fed into an FPGA, see Figure 1, which implements further filtering and the complete adaptive antenna array processing. The single channel output, *y*, of the adaptive antenna array is then converted back to the analogue domain by a D/A converter and after further analogue filtering and mixing the signal can be played into a standard GPS receiver.

Adaptive Algorithm

Optimum null-steering algorithms are usually applied to antenna arrays by simply minimising the output power of a weighted sum of receiver outputs subject to one of the multiplier weights (the reference channel) being fixed at unity, see e.g. Trinkle and Gray (2001). This leads to an optimum weight vector given by

$$\underline{w} = \frac{R^{-1}\underline{c}}{\underline{c}^H R^{-1}\underline{c}}$$

where

$$\underline{c} = [1,0,0,\ldots,0]^T$$

and R is the covariance matrix of the receiver outputs defined by

$$R = E\left\{\underline{u}(k)\underline{u}^H(k)\right\}.$$

The vector $\underline{u}(k)$ contains the antenna outputs at time kt_s where t_s is the sampling interval. With reference to Figure 1, it is assumed that each channel is Hilbert transformed allowing the use of the analytic (complex) signal representation.

For polarimetric antenna arrays, better performance may be achieved by choosing as the reference channel the right hand circularly polarized component of the reference antenna, thus avoiding a 3dB loss in SNR of desired GPS signals in the absence of interferences see e.g. Trinkle and Cheuk (2003). To handle the dual channels of each polarised antenna the vector of complex receiver outputs is written as:

$$\underline{u}(k) = \left(u_{1H}[k], u_{1V}[k], u_{2H}[k], u_{2V}[k], \ldots u_{NH}[k], u_{NV}[k]\right)^T \tag{2.1}$$

where

$$u_{nH}[k] = u_{nHI}[k] + ju_{nHQ}[k]$$

and

$$u_{nV}[k] = u_{nVI}[k] + ju_{nVQ}[k].$$

For right hand circular polarisation the horizontal and vertical components of the electric field are –90 degrees out of phase with each other. Thus, for an ideal narrowband RHCP signal the complex vector of outputs from the first reference channel, is given by

$$\underline{u}_1(k) = \begin{bmatrix} u_{1H}[k] \\ u_{1V}[k] \end{bmatrix} = s[kt_s]\begin{bmatrix} 1 \\ -j \end{bmatrix}$$

Combining, with complex weights, w_{1H} and w_{1V}, the outputs of the horizontal and vertical channels, such that the response due to a circularly polarised signal of unit amplitude is fixed at unity implies that the complex weights are constrained to satisfy

$$w^*_{1H} - jw^*_{1V} = 1.$$

Generalising for the full array, the constraint becomes

$$\underline{w}^H \underline{c} = 1$$

where

$$\underline{c} = [1, -j, 0, \ldots, 0]^T.$$

From Trinkle and Cheuk (2003), the optimum weight vector becomes

$$\underline{w} = \frac{R^{-1}\underline{c}}{\underline{c}^H R^{-1}\underline{c}}. \tag{2.2}$$

The advantage of this constraint, is that it allows, for an *N* channel receiver, *N*-1 nulls to be formed as opposed to the previous approaches which restricted the number of nulls to be *N*-2 see e.g. Trinkle and Cheuk (2003). In order impose this constraint with an unconstrained LMS adaptive algorithm, an orthogonal projection matrix is used. This transformation takes the horizontally and the vertically polarized complex signals from the dual polarized antenna and converts them to a RHCP signal and another signal that is orthogonal to the RHCP signal (LHCP) as illustrated in Figure 2 below.

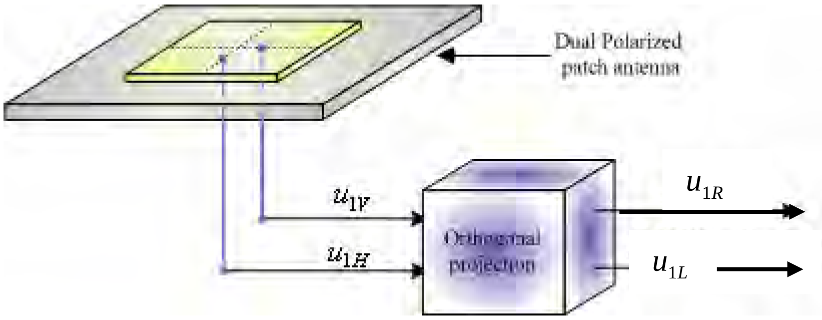


Figure 2. Orthogonal projection matrix.

The orthogonal projection matrix transformation is given by

$$\begin{bmatrix} u_{1R}[k] \\ u_{1L}[k] \end{bmatrix} = \begin{bmatrix} 1 & -j \\ 1 & j \end{bmatrix} \begin{bmatrix} u_{1V}[k] \\ u_{1H}[k] \end{bmatrix}$$

The outputs of the orthogonal projection matrix are then fed to the adaptive algorithm. The RHCP channel, $u_{1R}[k]$, is the reference channel, as it has the same polarization as the GPS signal. The orthogonal projection of the RHCP channel, i.e., $u_{1L}[k]$ is the first auxiliary channel, which is multiplied by a weight determined by the adaptive algorithm. The other auxiliary channels are the vertical and horizontally polarised outputs of the second antenna. Note that there is no need to transform these outputs into right and left circularly polarised components as the signals are assumed narrow-band so the required phase shift can be incorporated into the adaptive algorithm.

To estimate the weights we could first estimate *R* from the input data and then substitute into Eq. 2.2 to give the desired weights. This approach was not taken as it would require a matrix inversion which is not simple to implement in an FPGA. To avoid direct matrix inversion, gradient descent algorithms are used to iteratively minimise the mean square error.

As the output of the beamformer needs to be converted back to a real signal, a simplified version, of the standard complex LMS algorithm was applied. The trade-off between this algorithm and the full complex LMS algorithm has been considered in Horowitz and Senne (1981).

3 Implementation of the LMS Algorithm

The use of finite precision arithmetic in the LMS algorithm can cause drift in the weight vectors, Sethares et al (1986), particularly in the presence of strong interferences. To overcome this a leakage factor can be incorporated into the LMS algorithm. An investigation of a number of leaky LMS algorithms is carried out below.

Leaky LMS Algorithm

The leaky LMS algorithm prevents weight vector drift in finite precision implementations by inserting a leakage factor, α_L, into the weight vector update loop. This leakage factor avoids weight vector drift by containing the energy in the impulse response of the LMS adaptive filter see e.g. Haykin (2002). The leaky LMS algorithm is given by :

$$\underline{w}(k+1) = (1-\mu\alpha_L)\underline{w}(k) + \mu\{\underline{u}(k)e[k]\} \tag{3.1}$$

where $e[k]$, the error signal, is given by

$$e[k] = d[k] - \underline{w}^H(k)\underline{u}'(k) = u_{1R}(k) - \underline{w}^H(k)\underline{u}'(k)$$

and

$$\underline{u}'(k) = (u_{1L}[k], u_{2H}[k], u_{2V}[k], \ldots\ldots\ldots\ldots u_{NH}[k], u_{NV}[k])^T$$

The difference between the leaky LMS and the conventional LMS algorithm is the leakage factor $(1-\mu\alpha_L)$, the first term on the right hand side of the Eq. 3.1. To stabilize the algorithm (i.e., to avoid overflows), the leakage factor, α_L, must satisfy the condition

$$0 \le \alpha_L < 1/\mu \tag{3.2}$$

The leakage factor can also prevent stalling in the LMS algorithm. Stalling occurs when the correction term (i.e. $\mu[\underline{u}(k)e(k)]$) in the weight update equation is smaller than the least significant bit (LSB), thus the LMS algorithm will stop adapting. Stalling can also be prevented by dithering the error term.

A disadvantage of the leaky LMS algorithm is that it requires more computational and hardware resources. It also biases the weight vector away from the optimal solution, which degrades the performance of the filter. Since the average weight vector noise in the leaky LMS will never converge to the zero, the MSE (Minimum Square Error) is normally larger than for the conventional LMS. Decreasing the value of α_L will reduce this problem, but if it is too small then stalling will occur, as the estimation error is not noisy enough.

Circular Leakage LMS Algorithm

The circular leakage LMS described in Nascimento and Sayed (1999) is another leakage-based algorithm to prevent the drift problem. It has an advantage over the leaky LMS algorithm as it does not introduce a bias in the weight estimates. The circular leakage LMS algorithm is a modified version of the $switching-\sigma$ algorithm described in Ioannou and Tsakalis (1986), but uses less hardware resources while maintaining the same stabilization performance in the adaptive filter. The weight update equation for the circular leakage LMS can be written as

$$\underline{w}(k+1) = (1-\mu\alpha_C(k))\underline{w}(k) + \mu[\underline{u}(k)e(k)] \quad (3.3)$$

where $\alpha_C(k)$ is a nonlinear and time-varying circular leakage term. It is defined as follows

In the above equation the α_0 is a pre-defined constant, $0 < C_1 < C_2$ are pre-specified levels for the magnitudes of the components of the estimated weight vectors and $D = 0.5\{C_2 - C_1\}$. Their choice is not discussed here.

As indicted above the circular leakage term has four different regions which determine the value of the time-varying circular leakage term to be applied to the weight vector recursion It is also only applied to one channel at each iteration (i.e. in a multi-channel antenna array each channel has a single tap weight $w_n(k)$, where n is the channel number). At the first iteration, the magnitude of the weight vector component corresponding to the first channel, i.e., $|w_1(k)|$ is checked to determine into which of the ranges of Eq 3.4 it falls. The circular leakage term appropriate to that range is then applied to the weight vector update equation. In the next iteration, the magnitude of the weight vector component corresponding to the second channel, i.e., $|w_2(k+1)|$ is examined to determine which circular leakage term to apply to the update equation, etc. The procedure continues until the weights of all channels have been examined and then, in the next iteration, the process returns to the first channel and cycles around again.

A typical circular leakage function plot is illustrated in Figure 3 where the positive constants chosen are : α_0 = 0.3, C_1 = 0.6, C_2 = 0.9. These values can be calculated using the equations given in Nascimento and Sayed (1999).

$$\alpha_C(k) = \left\{ \begin{array}{ll} \alpha_0 & \text{if} \quad |w_n(k)| \ge C_2 \\ \alpha_0 - \dfrac{\alpha_0}{2}\left(\dfrac{C_2 - |w_n(k)|}{D}\right)^2 & \text{if} \quad 0.5\{C_1 + C_2\} \le |w_n(k)| < C_2 \\ \dfrac{\alpha_0}{2}\left(\dfrac{|w_n(k)| - C_1}{D}\right)^2 & \text{if} \quad C_1 < |w_n(k)| < 0.5\{C_1 + C_2\} \\ 0, & \text{otherwise} \end{array} \right\} \quad (3.4)$$

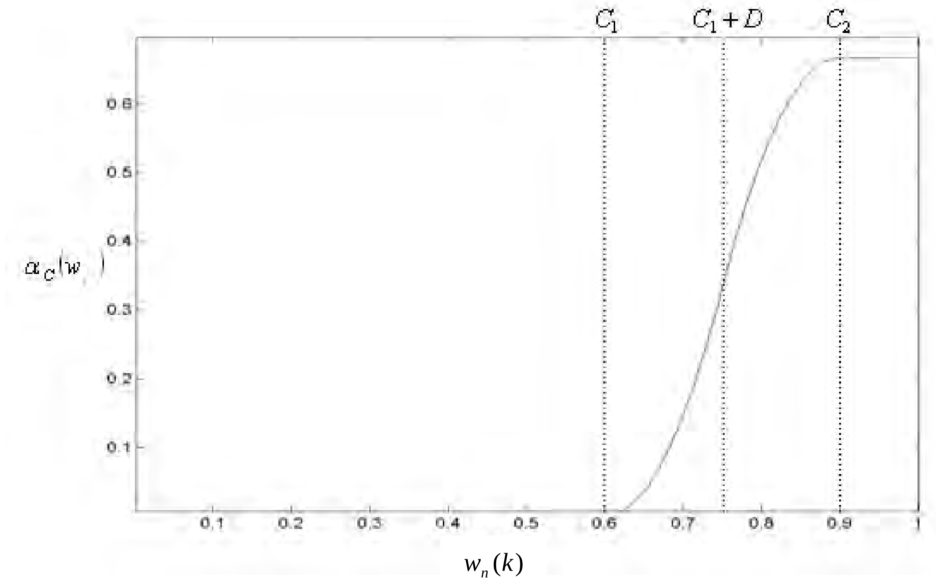


Figure 3. Circular leakage function.

When implementing the circular leakage in real time on the FPGA, one of the main factors that needs to be consider is latency. The standard LMS weight update loop has a delay of one clock cycle. From Eq. 3.3 and Eq. 3.4 the circular leakage algorithm requires six multiplications, two subtractions and two divisions per weight vector update. This introduces significant latency in the weight estimation loop, requiring it to take more than one clock cycle to complete all the calculations. Such a delay would cause the LMS algorithm to behave like the DLMS (Delayed Least Mean Square algorithm), resulting in increased convergence times and an overall loss in performance.

Modified Circular Leakage LMS Algorithm

As the adaptive null-steerer is to operate in real time, the above circular leakage algorithm was replaced by a simplified linear algorithm that required less arithmetic operations. The linear equations for determining the leakage factor α_C are given by

Eq. 3.5 has the same four regions as Eq. 3.3, but since the quadratic terms are replaced by linear ones the modified function uses less multiplications and no division and so its computational complexity is significantly reduced. As a result, the modified function can complete the weight update calculation in one clock cycle, making it more suitable for a real time implementation.

Figure 4 shows a plot of α_c for the standard circular-leakage (blue), and the modified circular-leakage (red). As shown in the plot, the original function rises quickly when the weight estimation reaches the C_1 (0.88) bound, it then goes up almost linearly as w increases, and then rolls off smoothly shortly before reaching the C_2 (0.93) bound.

$$\alpha_C = \begin{cases} \alpha_0 & \text{if} \quad |w_n(k)| \geq C_2 \\ 5.2(w - 0.9133) + 0.58 & \text{if} \quad 0.5\{C_1 + C_2\} \leq |w_n(k)| < C_2 \\ 17.4(w - C_1) & \text{if} \quad C_1 < |w_n(k)| < 0.5\{C_1 + C_2\} \\ 0, & \text{otherwise} \end{cases} \tag{3.5}$$

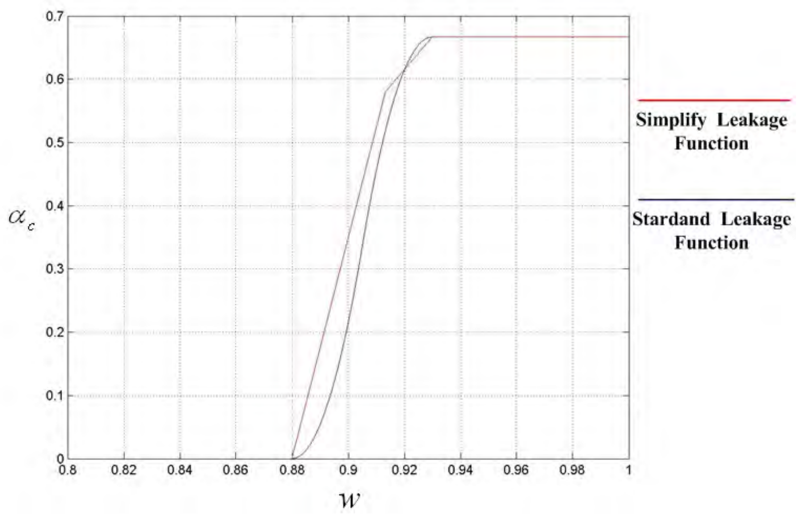


Figure 4. Simplified/standard circular leakage function with $C_1 = 0.88, C_2 = 0.93$.

The linear approximation has a slightly higher numerical value of leakage factor (α_c) at the same weight value (w) compared with the standard one. This ensures stability of the algorithm. Figure 5 shows simulation results indicating that the modified circular leakage algorithm remains stable and behaves almost the same as the original one albeit at what appears to be a higher level of misadjustment noise.

4 FPGA Implementation

This section considers the FPGA implementation of the adaptive null-steerer. The FPGA has four main processing modules including, digital decimation filtering, Hilbert transform filtering and the implementation of the LMS adaptive algorithm using the simplified circular leakage algorithm. The signal flow diagram is illustrated in Figure 6.

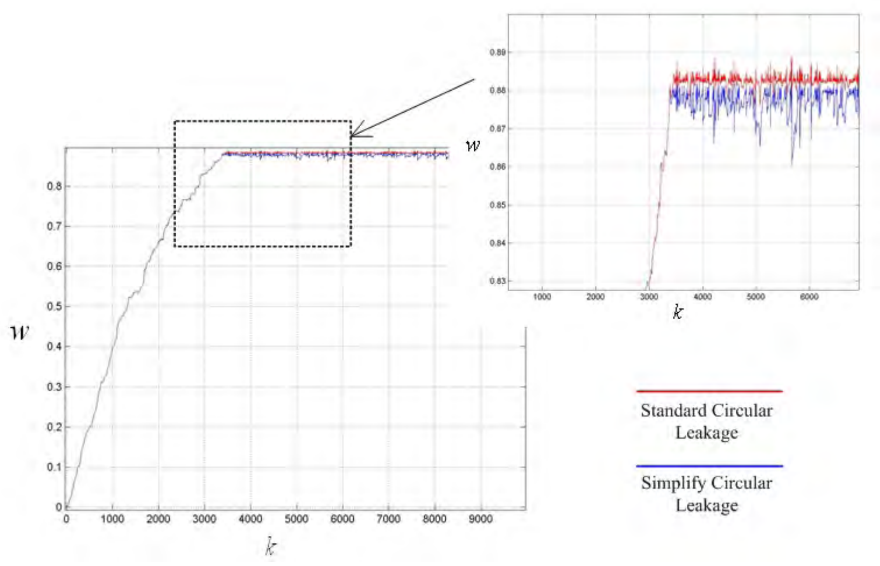


Figure 5 w for standard/simplified circular leakage $\mu\alpha_o = 0.1667$, $C_1 = 0.88, C_2 = 0.93$.

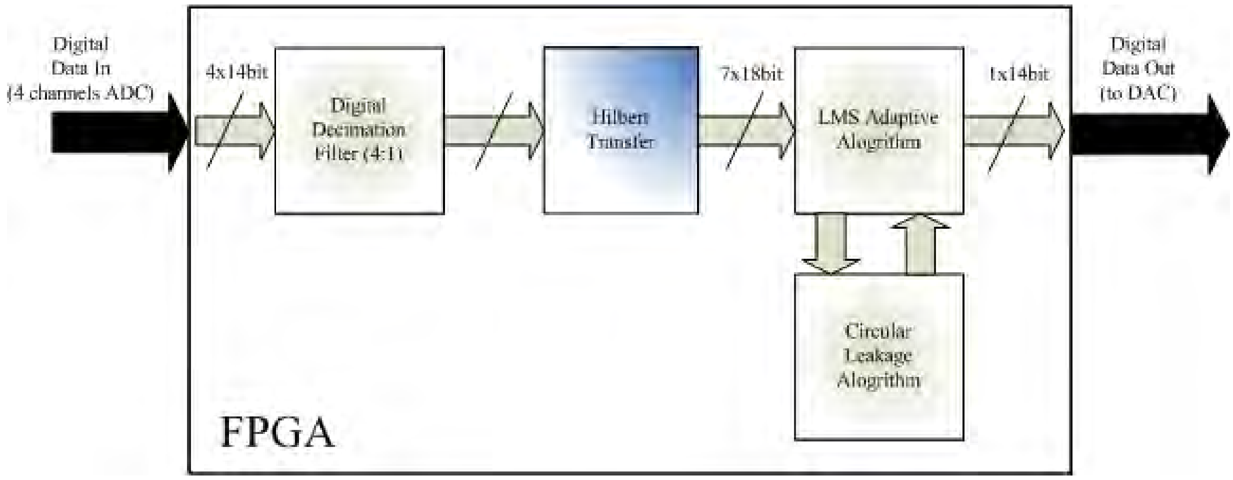


Figure 6 FPGA signal flow diagram

In this section, the FPGA implementation of the circular leakage algorithm is discussed.

Simplified Circular Leakage LMS Algorithm FPGA Implementation

The Xilinx System Generator was used to implement the code for the adaptive null-steerer. A schematic block diagram which represents the structure of the code is shown in Figure 7.

In Figure 7, there are 8 inputs on the left hand side (these are from top to bottom: u1HI, u1HQ, u1VI, u1VQ, u2HI, u2HQ, u2VI and u2VQ) and one output on the right hand side. The top four inputs on the left hand side are fed to the orthogonal projection block, this block converts the horizontal polarization and vertical polarization to a RHCP and LHCP output signal. The RHCP channel is chosen to be the reference channel (i.e. u_RHCP_I), which is treated as the desired signal $d_I(k)$. Then the next pair is the real part and imaginary part of the signal from the orthogonal projection of RHCP channel (i.e. u_RHCP(1+j)_I and u_RHCP(1+j)_Q), then the next pair is u2HI and u2HQ etc. So there are a total of one reference and three pairs of auxiliary channels.

Each of the real / imaginary signals are fed to the LMS based adaptive algorithm, shown in Figure 8. Then all the outputs of each LMS block are fed into the adder and subtracted from the desired signal to give the error signal which is then is fed back to the LMS block to calculate the updated weights. The LMS loop operates continuously, at the input sample rate, and the optimum weights are continuously estimated.

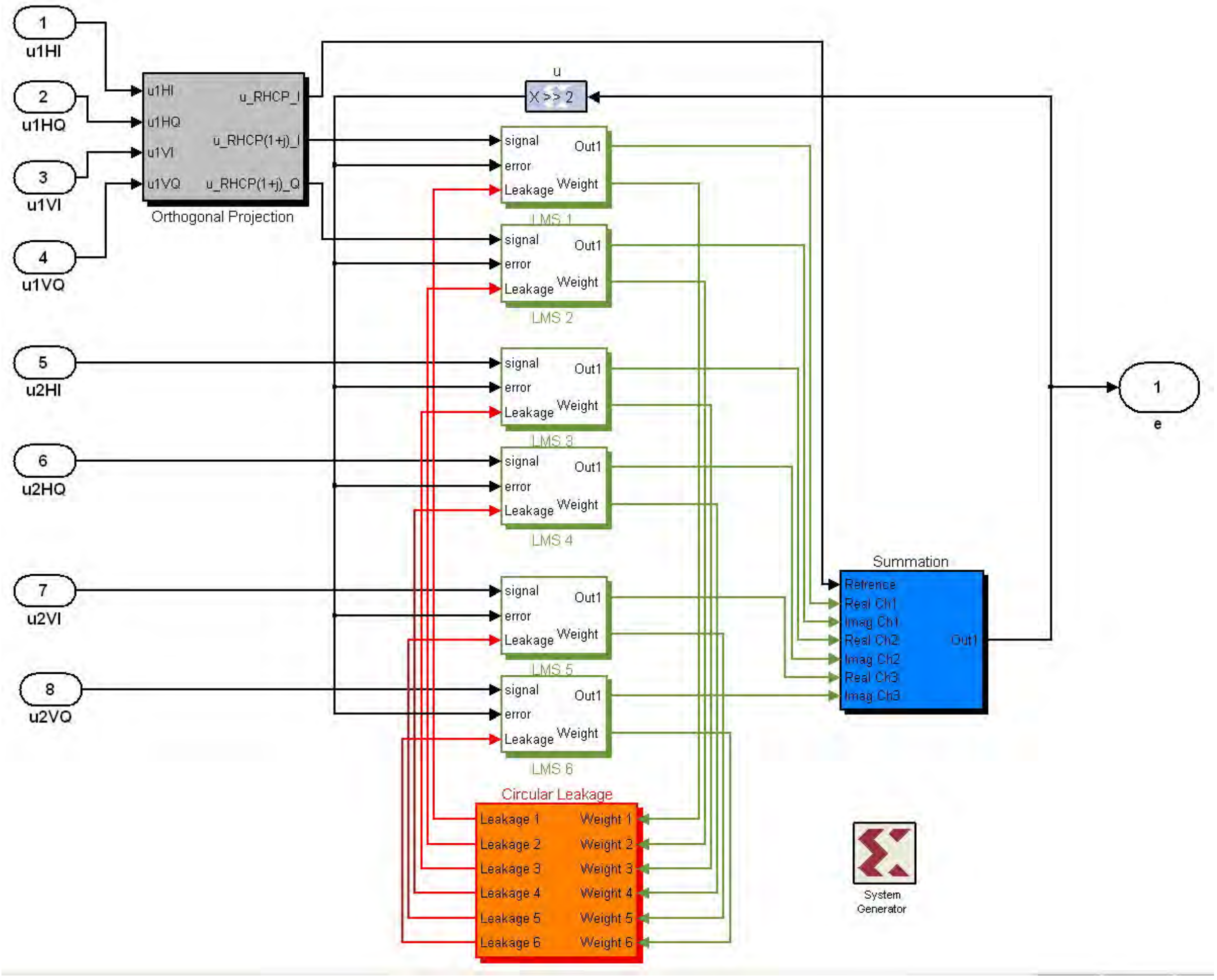


Figure 7 Four channel digital null-steerer - FPGA schematic.

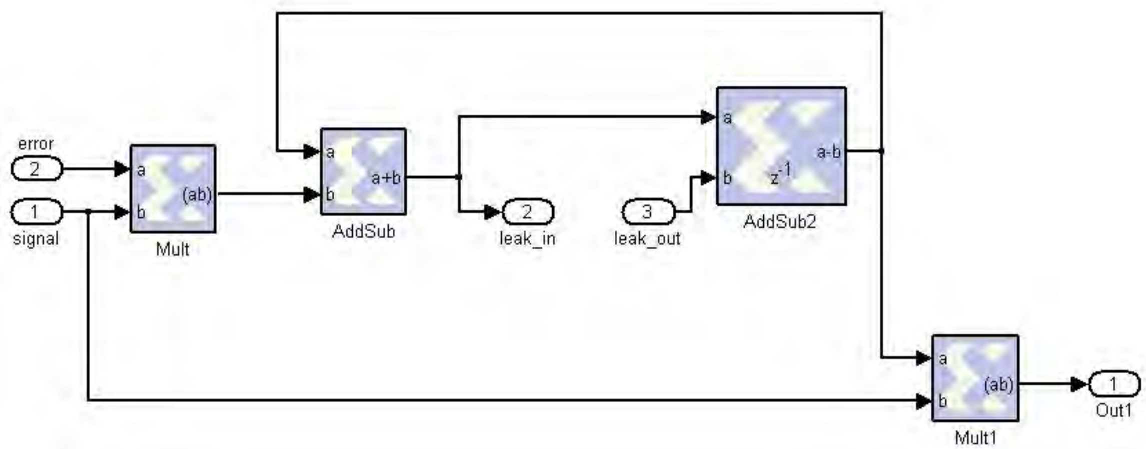


Figure 8 LMS base algorithm block - FPGA schematic.

The circular leakage algorithm is implemented in VHDL, as shown in Figure 7 in the block called circular leakage. It has six inputs on the right hand side and same number of outputs in left hand side. It implements the if/then statement of Eq. 3.5. In each iteration, it measures the weight estimate for one channel, and outputs the leakage value according to the Eq. 3.5. This value is then fed back to the leak_out (Figure 8), and subtracted from the current weight vector component. Then in the next iteration, the circular leakage block will check the w of the next channel.

Conclusions

The FPGA implementation of an adaptive LMS algorithm suitable for dual polarised GPS antenna arrays has been presented. The LMS algorithm minimises the output power of the array while maintaining unity gain for RHCP signals on the reference antenna. To prevent weight vector drift in the LMS algorithm a circular leakage algorithm was implemented. A simplified version of the leakage algorithm was derived which introduces minimal additional bias and allows the LMS updates to occur at the input data rate. The complete algorithm was implemented on an FPGA using the Xilinx System Generator.

References

Manz A.; Shallerg K.; Shloss P.; (2000): ***Improving WAAS Receiver Radio Frequency Interference Rejection.*** ION GPS 200, September 2000, pp471-479

Kunysz W. (2000): ***A Novel GPS Survey Antenna.*** ION NTM 2000, January 2000

Kunysz W. (2000): ***High Performance GPS Pinwheel Antenna***. ION NTM 2000, September 2000

Dimos G.; Upadhyay T.; Jenkins T. (1995): *Low-Cost Solution to Narrowband GPS Interference Problem* Proc IEEE 1995 Naecon Meeting, vol I pp145-153 May 1995.

Trinkle M.; Gray D. (2001): ***GPS Interference Mitigation; Overview and Experimental Results.*** SatNav 2001, July, 2001.

Braasch M. S.; Rosen M. W. (1998): ***Low-Cost GPS Interference Mitigation Using Single Aperture Cancellation Techniques.*** Navigation 2000, 1998 ION National Technical Meeting.

Zoltowski M. D.; Gecan A. S. (1995): ***Advanced Adaptive Null Steering Concepts for GPS.*** MILCOM 95 Vol. 3, pp. 1214-1218.

Jian K.; Guosheng R.; Xizhang K. (1998): ***Power Inversion Adaptive Algorithm for GPS Receiver Antijamming.*** International Radar Symposium, September 1998. pp.1395-1401.

Fante R. L.; Vaccaro J. J. (2000): ***Wideband Cancellation of Interference in a GPS Receive Array.*** IEEE Transaction on Aerospace and Electronic Systems Vol. 36, No. 2 April 2000.

Trinkle M.; Cheuk W. (2003): ***Null-Steering GPS Dual-Polarized Antenna Arrays.*** SatNav July 2003.

Nagai E. C.; Blejer D. J.; Phuong T.; Herd J. (2002): ***Anti-jam Performance of Small GPS Polarimetric Array.*** IEEE Antennas and Propagation Society Int. Symp. 2002 Digest, Vol. 2, 16-21 June 2002, pp. 128-131.

Fante R. L.; Vaccaro J. J. (2002): ***Evaluation of Adaptive Space-Time-Polarization Cancellation of Broadband Interference.*** Proceedings of the 2002 IEEE Precision Location & Navigation Symposium pp.1-3.

Compton R. T. JR. (1981): ***On the Performance of a Polarization Sensitive Adaptive Array.*** IEEE Transaction on Antennas and Propagation, Vol. AP-29, No. 5, pp. 718-725, September 1981.

Legrand F.; Macaiau C.; Issler J. L.; Lestarquit L.; Mehlen C. (2000): ***Improvement of Pseudo Range Measurements Accuracy By Using Fast Adaptive Bandwidth Lock Loops.*** ION GPS 200, September 2000, pp. 2346-2356

Soloviev A.; van Grass F. (2004): ***Combining Low-Cost Inertial System with GPS Applications for General Aviation***. GPS World, March 2004 pp. 38-45.

Horowitz L.; Senne D. (1981): ***Performance Advantage of Complex LMS for Controlling Narrow-Band Adaptive Arrays.*** IEEE Transactions on acoustics, and signal processing, Vol. ASSP-29, NO.3 June 1981.

Sethares W. A.; Dale A.; Lawrence C.; Johnson R. JR.; Bitmead R. R. (1986): ***Parameter Drift in LMS Adaptive Filters.*** IEEE Transactions on Acoustic, Speech and Signal Processing Vol. ASSP-34, No 4, August 1986.

Nascimento V. H.; Sayed A. H. (1999): ***Unbiased and Stable Leakage-Based Adaptive Filter.*** IEEE Transactions on Signal Processing, Vol. 47, No. 12 December 1999.

Ioannou P.; Tsakalis K. ***Robust discrete-time adaptive control*** in Narendra K. S, Ed.,(1986): ***Adaptive and Learning Systems: Theory and Applications.*** New York: Plenum, 1986, pp. 73-87.

Haykin S. (2002): ***Adaptive Filter Theory 4th Edition***. Prentice Hall 2002.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 268-276

Journal of Global Positioning Systems
2005

The Application of GPS Technique in Determining the Earth's Potential Field

WenBin Shen and Jinsheng Ning

Dept. Of Geophysics/ Key Laboratory of Geospace Environment and Geodesy, School of Geodesy and Geomatics, Wuhan University, 129 Luoyu Road, Wuhan 430079, China
e-mail: wbshen@sgg.whu.edu.cn; Tel: 0086 27 87851394; Fax: 0086 27 68778825

Received: 15 November 2004 / Accepted: 20 October 2005

Abstract. Two approaches to determining the Earth's external potential field by using GPS technique are proposed. The first one is that the relation between the geopotential difference and the light signal's frequency shift, between two separated points, is applied. The second one is that the spherical harmonic expansion series and a new technique dealing with the "downward continuation" problem are applied. Given the boundary value provided by GPS "geopotential frequency shift" on the Earth's surface, the Earth's external field could be determined based on the "fictitious compress recovery" method. Given the boundary value derived by on-board GPS technique on the satellite surface, the Earth's external field could be determined by using a new technique for solving the "downward continuation" problem, which is also based on the "fictitious compress recovery" method. The main idea of the "fictitious compress recovery" is that an iterative procedure of "compress " and "recovery" between the given boundary (the Earth's surface or the satellite surface) and the surface of Bjerhammar sphere is executed and a fictitious field is created, which coincides with the real field in the domain outside the Earth. Simulation tests support the new approach.

Key words: GPS observation, fictitious compress recovery, geopotential frequency shift, potential field determination

1 Introduction

The GPS technique plays an important role in geoscience and has very broad applications in various fields, especially in determining the coordinates of the interested points. One possible application of the GPS technique, which might not be taken good attention in geodesy, is that the geopotential difference between two (even far away) separated points on the Earth's surface might be directly determined by using GPS signals (Shen et al, 1993; Brumberg and Groten, 2002). Conventionally, the potential difference between two points are determined by gravimetry and levelling, the drawback of which is that it is almost impossible to connect two points which are located on two continents, because it is well known that the potential surface of the mean sea level is not an equi-potential surface. Hence, if GPS "geopotential frequency shift" approach could be applied in practice in the future, a unified world datum might be established. Meanwhile, once the geopotential on the Earth's surface is given, the Earth's external geopotential field could be strictly determined by using the "fictitious compress recovery" method (Shen, 2004), which can be also applied for determining the Earth's external geopotential field (and consequently solving the "downward continuation" problem), if the gravitational potential on the satellite surface (which is defined by the flying satellite) is given, for which the GPS technique provides a good opportunity.

Satellite on-board GPS receiver provides the position $x_i(t)$ of the satellite (e.g., CHAMP or GRACE satellite), and consequently its velocity $v_i(t)$ by using differential approach. Suppose an accelerometer is equipped with the flying satellite, one could determine the Earth's gravitational potential along the satellite orbit based on the energy conservation law, e.g., using the well known energy integral approach (e.g., Gerlach et al, 2003; Visser et al, 2003). Then, one can determine the potential field based on the truncated spherical harmonic expansion (in this case there are finite harmonic coefficients to be determined) and by using, e.g., the least squares adjustment (e.g., Rummel et al, 1993), which is referred to as the conventional approach for convenience.

However, the problem is that so determined field might not be valid in the domain between the Earth's surface and the surface of Brilloun sphere, the smallest sphere that encloses the whole Earth, because it could not be guaranteed that the spherical harmonic expansion series is convergent in that domain (e.g., Moritz, 1978; Sjöberg, 1980; Rummel et al, 1993). To solve this problem, two approaches could be applied. If the provided (discrete) potential values are uniformly distributed on the satellite surface, one could determine the fictitious field that coincides with the real field in the whole domain outside the Earth based on the "fictitious compress recovery" method, naturally solving the "downward continuation" problem, which is greatly interested by geodesists. Otherwise, using the conventional approach one first determines the potential field, which is valid in the domain outside Brillouin sphere; then, choose a simply closed surface (it is recommended that an ellipsoidal or spherical surface is chosen, Cf. Remark 2 in Sec.6) that encloses Brillouin sphere, and applying the "fictitious compress recovery" method one could determine the real field between the Earth's surface and the surface of Brillouin sphere, also solving the "downward continuation" problem.

Since the "fictitious compress recovery" method (Shen, 2004) is essential, it will be summarized in Sec.2. In Sec.3, the GPS "geopotential frequency shift" approach is presented, and in Sec.4, the "downward continuation" approach is presented. In Sec.5, preliminary simulation results are provided, and finally, the conclusions and discussions end this paper.

2 The "fictitious compress recovery" method

Choose an inner sphere or Bjerhammar sphere K (Bjerhammar, 1964), which is an open set (excluding the boundary ∂K of the sphere K) and entirely located inside the Earth, with its centre coinciding with the Earth's mass centre (Shen, 2004). In the domain $\overline{K}$, which denotes the domain outside the sphere K , including the boundary ∂K , there exists such a potential field $V^*(P)$ satisfying

$$\begin{aligned} &\Delta V^*(P) = 0, \quad P \in \overline{K} \\ &V^*(P)\,|_{\partial K} = V^*{}_{\partial K} \\ &\lim_{P\to\infty} V^*(P) = 0 \end{aligned} \qquad (1)$$

and on the boundary of the Earth, $V^*(P)$ has the same value as the Earth's real potential field $V(P)$ has. Then, the fictitious potential field $V^*(P)$ coincides exactly with the Earth's real potential field $V(P)$ in the domain $\overline{\Omega}$ (Shen, 2004), where Ω denotes the domain (open set) occupied by the Earth (excluding the Earth's boundary), $\overline{\Omega}$ denotes the domain outside the Earth (including the boundary of the Earth), and let $\partial\Omega$ denote the boundary of the Earth. To realize this idea, the "compress" and "recovery" technique (Shen, 2004) could be used.

Set the observed gravitational potential value $V(P)\,|_{\partial\Omega} \equiv V_{\partial\Omega}$ (note that $V_{\partial\Omega}$ is obtained by the observed geopotential W subtracting the centrifugal potential Q) on the surface of K along the radial direction. Then, based on Poisson integral a regular harmonic solution $V^{*(1)}(P)$ could be determined in the domain $\overline{K}$:

$$V^{*(1)}(P) = \frac{r^2 - R^2}{4\pi R} \int_{\partial K} \frac{V_{\partial\Omega}}{l^3} d\sigma, \quad P \in \overline{K} \qquad (2)$$

which can be taken as the first approximation of the Earth's real potential field $V(P)$ in the domain outside the Earth. The first-order residual field $\delta^{(1)}(P)$ is defined as follows (Shen, 2004):

$$\delta^{(1)}(P) = \delta^{(0)}(P) - V^{*(1)}(P) \equiv V(P) - V^{*(1)}(P), \quad P \in \overline{\Omega} \qquad (3)$$

where $\delta^{(0)}(P) \equiv V(P)$ is the Earth's real potential field. It should be noted that $\delta^{(1)}(P)$ is defined only in the domain $\overline{\Omega}$. With $V^{*(1)}(P)$ the boundary value $V^{*(1)}\big|_{\partial\Omega} \equiv V^{*(1)}_{\partial\Omega}$ can be calculated. Based on Eq.(3), set the first-order residual boundary value

$$\delta^{(1)}\big|_{\partial\Omega} = V_{\partial\Omega} - V^{*(1)}_{\partial\Omega} \qquad (4)$$

again on the surface of Bjerhammar sphere K (note that the boundary value $\delta^{(1)}\big|_{\partial\Omega}$ is identically compressed on ∂K along the radial direction), a second-order regular harmonic solution $V^{*(2)}(P)$ can be determined in $\bar{K}$:

$$V^{*(2)}(P) = \frac{r^2 - R^2}{4\pi R} \int_{\partial\Omega} \frac{\delta^{(1)}\big|_{\partial\Omega}}{l^3} d\sigma, \quad P \in \bar{K} \qquad (5)$$

$V_1^* + V_2^*$ can be taken as the second approximation of the Earth's real potential field $V(P)$ in the domain outside the Earth.

Similarly, the second-order residual field $\delta^{(2)}(P)$ is defined as follows:

$$\delta^{(2)}(P) = \delta^{(1)}(P) - V^{*(2)}(P), \quad P \in \bar{\Omega} \qquad (6)$$

and set the second-order residual boundary value

$$\delta^{(2)}\,|_{\partial\Omega} \equiv \delta^{(1)}\,|_{\partial\Omega} - V^{*(2)}\,|_{\partial\Omega} \qquad (7)$$

again on the surface of the sphere K (along the radial direction), a third-order regular harmonic solution $V^{*(3)}(P)$ in $\bar{K}$ is determined. This procedure can be repeated until a series solution $V^*(P)$ is provided in the domain outside the sphere K (Shen, 2004):

$$V^*(P) = \sum_{n=1}^{\infty} V^{*(n)}(P), \quad P \in \bar{K} \tag{8}$$

which is a regular harmonic function in $\bar{K}$, and coincides exactly with the Earth's real potential field $V(P)$ in the domain $\bar{\Omega}$, the domain outside the Earth:

$$V^*(P) = \sum_{n=1}^{\infty} V^{*(n)}(P) \equiv V(P), \quad P \in \bar{\Omega} \tag{9}$$

Hence, once the geopotential or gravitational potential on the Earth's physical surface $\partial\Omega$ is given, the Earth's external field can be exactly determined. It should be pointed out that the "fictitious compress recovery" method has wide applications in geophysics (Shen *et al.*, 2004).

3 The "geopotential frequency shift" approach

The geopotential frequency shift approach by using GPS signals was briefly proposed in Shen *et al*. (1993), and the technical details could be found in Shen (1998).

Suppose a light signal with frequency f is emitted from point P by an emitter, and the signal is received at point Q by a receiver. Because of the geopotential difference between these two points, the frequency of the received light signal is not f but f'. Using f_P and f_Q to denote f and f' respectively, the following equation holds (Pound and Snider, 1965; Shen *et al*., 1993):

$$\Delta f \equiv f_Q - f_P = -\frac{f}{c^2}\Delta W_{PQ} \equiv -\frac{f}{c^2}(W_Q - W_P) \tag{10}$$

where c is the velocity of light in vacuum, W_P and W_Q are the geopotentials at points P and Q, respectively. Expression (10) is called in literature the gravity frequency shift equation (Pound and Snider, 1965), or properly called the geopotential frequency shift equation (due to the fact that the frequency shift is caused by the geopotential difference). Katila and Riski (1981) confirmed Eq.(10) with the accuracy level 10^{-2}. Vessot *et al*. (1980) proved that Eq.(10) is correct to the accuracy of 10^{-4}. Scientists believe that Eq.(10) is correct, because it is a result derived from the theory of general relativity. In fact, Eq.(10) can be also derived out based on quantum theory and energy conservation law (Shen, 1998).

Suppose the geopotential at point P is given, then, from Eq.(10) the geopotential at an arbitrary point Q can be determined by measuring the geopotential frequency shift between P and Q:

$$W_Q = W_P - \frac{c^2\Delta f}{f} \tag{11}$$

If the point P is chosen on the geoid, it holds that

$$W_Q = C_0 - \frac{c^2\Delta f}{f} \tag{12}$$

where C_0 is the geoid geopotential constant. It should be noted that C_0 might not be correct because of a constant shift δW, which will give rise to W_Q the same shift at an arbitrary point Q. This is a systematic error, and it could be filtered out by using the"fictitious compress recovery" method (the determination of C_0 or δW in details is beyond the scope of present paper and will be explored in a separated paper). Then, the geopotential at an arbitrary point Q on the Earth's physical surface $\partial\Omega$ can be determined based on the geopotential frequency shift equation. Then, the main problem is how to measure the frequency shift between two points. The basic principle of measuring the frequency shift can be stated as follows.

Set at point P an emitter which emits a light signal with frequency f and a receiver at point Q, which receives the light signal emitted by the emitter at point P. Suppose the received signal's frequency is f'. Then, it could be compared the frequency f' of the received light signal with it-self's standard frequency f (this is not only the emitting frequency at point P but also the standard innate frequency of the receiver at point Q), and the frequency shift $\Delta f = f' - f$ can be determined. Consequently, according to Eq.(11) the geopotential difference ΔW_{PQ} between P and Q can be determined. Applying the same principle it will be found the geopotential difference $\Delta W_{OP} = W_P - W_O$ between the geoid and the point P, where $W_O = C_0$ is the geopotential at point O located on the geoid. If C_0 is a known constant, W_P as well as W_Q can be found.

Now, suppose the light signal emitter E is located in a satellite. Two GPS receivers at P and Q receive the light signals coming from E corresponding to an emitting time t (seeing Fig.1), and suppose the received signals' frequencies corresponding to time t are recorded by receivers at P and Q in some way, respectively, i.e., f_P and f_Q corresponding to time t are recorded by receivers at P and Q, respectively. Note that the time at which the signal is received by P is generally different from that by Q. By comparing the received frequencies f_P and f_Q it could be determined the geopotential difference $\Delta W_{PQ} = W_Q - W_P$ (Shen *et al*., 1993), which is just given by Eq.(10).

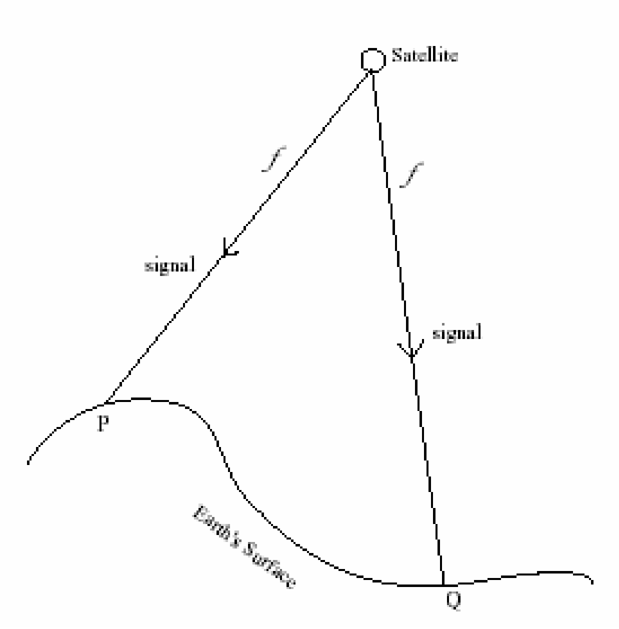


Fig. 1 Two receivers at points *P* and *Q* receive simultaneously the satellite-emitted light signal with frequency f

By this way, theoretically, the geopotential on the Earth's whole surface could be determined based on the geopotential frequency shift approach by using GPS technique. Then, the "fictitious compress recovery" method could be applied for determining the Earth's external potential field $V(P)$. Note that by GPS technique the coordinates x^i at any point on the Earth's physical surface can be determined, where x^i denote $x^1 \equiv x,\ x^2 \equiv y,\ x^3 \equiv z$. Consequently, once the geopotential W on the Earth's surface is determined, the gravitational potential V on the Earth's physical surface is determined.

4 The "downward continuation" approach

Suppose there are quite a few satellites flying around the Earth and many observation stations distributed at various points on the Earth's physical surface. Generally, the Earth's gravitational potential V could be expanded into spherical harmonic series (Heiskanen and Moritz,1967):

$$V(p) = GM\sum_{n=0}^{\infty}\sum_{m=0}^{n}\frac{a^n}{r^{n+1}}(a_{nm}\cos m\lambda + b_{nm}\sin m\lambda)P_{nm}(\cos\theta) \quad (13)$$

where G is the gravitational constant, M the Earth's mass, a the Earth's semi-major axis, r the distance between the coordinate origin o and the field point $P(x^i)$, a_{nm} and b_{nm} are constant coefficients to be determined based on various observations (GPS observations in our case), $P_{nm}(\cos\theta)$ are the associated Legendre functions, λ and θ are longitude and co-latitude, respectively. The expression (13) is at least correct in the domain outside a satellite surface ∂S. In the domain near the Earth's surface, the series (13) might be divergent (Moritz, 1978; Sjöberg, 1980; Rummel et al, 1993; Shen, 1995).

Now, suppose Eq.(13) holds in $\overline{S}$, the domain outside the satellite surface ∂S. To determine the field $V(P)$, the truncation technique should be applied (otherwise the infinite harmonic coefficients a_{nm} and b_{nm} can't be determined), i.e., only the first terms until degree N are left:

$$V_1(P) = GM\sum_{n=0}^{N}\sum_{m=0}^{n}\frac{a^n}{r^{n+1}}(a_{nm}\cos m\lambda + b_{nm}\sin m\lambda)P_{nm}(\cos\theta) \quad (14)$$

Hence, only the finite harmonic coefficients a_{nm} and b_{nm} are left as the unknown parameters. Then, by establishing the relation between the potential (or the gravitation $\partial_i V$ either the gravitational gradient $\partial_i\partial_j V$) and the GPS observations $x^i(t)$ (note that once $x^i(t)$ are determined as GPS observation values, $dx^i(t)/dt$ and $d^2x^i(t)/dt^2$ are also determined as GPS "observation values"), the finite harmonic coefficients a_{nm} and b_{nm} could be determined by using the least-squares method. That means the potential field $V_1(P)$ outside the satellite surface ∂S is determined. However, as mentioned before, it can't be guaranteed that Eq.(14) holds also in the domain near the Earth's surface due to the divergence problem caused by Eq.(13). Hence, it occurs the "downward continuation" problem, which has attracted many geodesists' interests and attention. The "downward continuation" problem was not solved satisfactorily by using conventional methods due to the well-known "ill-posed" problem. Recently, this problem was solved satisfactorily (Shen and Ning, 2004; Shen and Wang *et al*., 2005; Cf. Remark 1) by using the "fictitious compress recovery" method (Shen, 2004).

By GPS observations, suppose it has been established a model of the Earth's potential field, noted as EGM1, which is correct in the domain $\overline{S}$ (the domain outside the satellite surface ∂S). Then, on the boundary ∂S it has been known the boundary value $V_{\partial S} = V_1\big|_{\partial S}$, which can be assumed very accurate without loss of generality. Choose Bjerhammar sphere K, using the boundary value $V_{\partial S}$ (which is calculated by EGM1) and by applying the "fictitious compress recovery" method it could be determined a regular harmonic function $V^*(P)(P \in \bar{K})$, which coincides exactly with the real field $V(P)$ in the domain $\bar{S}$ under the assumption that the boundary value $V_{\partial S}$ is error-free (Shen, 2004). Furthermore, it has been proved that the determined fictitious field $V^*(P)(P \in \bar{K})$ coincides also with the real field $V(P)$ in the domain $\bar{\Omega} - \bar{S}$, the domain between the satellite surface and the Earth's physical surface (Shen and Ning, 2004; Shen and Wang *et al*., 2005; Cf. Remark 1). For convenience, the determined field constrained in $\bar{\Omega}$ based on the "fictitious compress recovery" method is referred to as FEGM (or FEGM1).

Remark 1: Suppose the boundary value $V_{\partial S}$ on the satellite surface ∂S (it can be also a spherical surface ∂K_Γ) is given. It will be briefly proved that the real field can be determined based on the "fictitious compress recovery" method, referred to Shen and Ning (2004). Set

$$\phi(P) = V^*(P) - V(P), \quad P \in \bar{\Omega} \tag{15}$$

where $V^*(P)$ is a fictitious regular harmonic solution (based on $V_{\partial S}$ and the "fictitious compress recovery" method) in the domain $\bar{K}$, the domain outside Bjerhammar sphere K. Hence, $\phi(P)$ is a regular harmonic function in the domain $\bar{\Omega}$, and satisfies the following equation:

$$\phi(P) = V^*(P) - V(P) = 0, \quad P \in \bar{S} \tag{16}$$

Based on Eq.(15), on the boundary $\partial\Omega$ it holds:

$$\phi(P)|_{\partial\Omega} = V^*(P)|_{\partial\Omega} - V(P)|_{\partial\Omega} \tag{17}$$

Applying the "fictitious compress recovery" method (Shen, 2004), it can be determined a regular harmonic function $\phi^*(P)$ in the domain $\bar{K}$, and it holds

$$\phi^*(P) = \phi(P), \quad P \in \bar{\Omega} \tag{18}$$

Since $\phi^*(P)$ is regular and harmonic in $\bar{K}$, it can be expanded into a (uniformly convergent) spherical harmonic series (the mathematical expression form looks like Eq.(13)). Using Eqs.(16) and (18) it must hold that $\phi^*(P) \equiv 0$ $(P \in \bar{K})$. Then, from Eqs.(15) and (18) it holds that $V^*(P) = V(P)$ $(P \in \bar{\Omega})$. The proof is completed.

Hence, after applying the new "downward continuation" approach, referred to as the "fictitious downward continuation" (Shen and Wang et al, 2005; Shen and Yan et al, 2005), the Earth's external potential field can be determined based the established model EGM1 (the Earth's potential field in the domain $\bar{S}$). If more precise EGM1 (in the domain $\bar{S}$) is established by various satellite observations (e.g. GPS, CHAMP, GRACE, GOCE, etc), a more precise field $V(P)$ in $\bar{\Omega}$ could be determined.

If given the gravitational potential values distributed uniformly on the satellite surface, a fictitious potential field in the domain outside Bjerhammar sphere can be directly determined based on the "fictitious compress recovery" method, and the determined fictitious field coincides with the real field in the whole domain outside the Earth. In this way, FEGM is directly established.

5 Preliminary simulation results

In this section a simulation test is provided, which is referred to (Li, 2005; Shen and Li, 2005). Choose a spherical coordinate system (r, θ, λ) and a gravitational potential model, a 4-sphere anomaly model: three small spheres O_i ($i = 1,2,3$) are located inside a large sphere O_0, with the parameters listed in Tab.1.

Tab.1 Parameters of the four spheres

Sphere	O_0	O_1	O_2	O_3
Centre [km, deg,deg]	(0, 0, 0)	(4000, 0, 0)	(2000, 90, 120)	(3000, 120, 240)
Radius [km]	R_0 6371	R_1 300	R_2 500	R_3 600
Potential at surface [m^2s^{-2}]	V_0 1000	V_1 -100	V_2 400	V_3 200

Based on the above model, the real potential field $V(P)$ outside the large sphere (which is assumed as the "Earth") is known, expressed as

$$V(P) = \frac{R}{r}V_0 + \sum_{i=1}^{3}\frac{R_i}{r_i}V_i, \quad P \in \bar{\Omega} \tag{19}$$

where r_i ($i = 1,2,3$) is the distance from O_i to the field point P.

Now, it is supposed that only the boundary value $V_{\partial S}$ on a satellite surface is known, calculated from Eq.(19), and the aim is to determine the real field $V(P)$ in the domain outside the "Earth" (i.e., in the domain outside the large sphere), based only on the given boundary value $V_{\partial S}$. The boundary value $V_{\partial S}$ on the satellite surface is supposed to be obtained based on a polar satellite (equipped with a GPS receiver and an accelerometer), using the well known energy integral approach (e.g., Gerlach et al, 2003; Visser et al, 2003). The satellite surface is supposed to be a rotation-ellipsoidal surface, with its geometric centre coinciding with the coordinate origin, the major-axis $a = 6371 + 250$ km, and the eccentricity $e = 0.01$. The radius of the inner sphere (i.e., Bjerhammar sphere) is taken as 6000 km. The simulation calculation (especially the calculation of Poisson integral) is executed based on grid approach, with $1^o \times 1^o$ grid defined by parallel latitude line and longitude line on the surface of the inner sphere. The grids on the satellite surface are 1-1 corresponding to the grids on the spherical surface. Consequently there are 64800 discrete values, which are uniformly distributed on the satellite surface.

Now, with the given (discrete) boundary values $V_{\partial S}$, the fictitious distribution on the surface of the inner sphere is determined based on the "fictitious compress recovery" method (Cf. Sec.2), and consequently the fictitious field $V^*(P)$ in the domain outside the inner sphere is determined. In theory, the fictitious field $V^*(P)$ coincides with the real field in the domain outside the "Earth". Hence, we need only to compare the calculated values with the real values on the surface of the "Earth", because, if two regular harmonic fields (in the domain outside the "Earth") coincide on the boundary of the "Earth", they must coincide in the whole domain outside the "Earth". The calculated results are summarized in Tab.2, where $n = 8$ and $n = 15$ express the iterative procedure times, respectively, $\partial\Omega$ and ∂S express the satellite surface and the surface of the "Earth", respectively, and $\Delta V = V - V^*$ expresses the residual potential value between the real value and the calculated (fictitious) value. Fig.2 shows the residuals between the real values and the calculated corresponding values on the "Earth's surface".

Tab.2 Results based on the boundary values on the satellite surface

n	$\max(\lvert\Delta V\rvert)$ [m²s⁻²]		$mean(\Delta V)$ [m²s⁻²]		RMS [m²s⁻²]	
	On ∂S	**On $\partial\Omega$**	On ∂S	**On $\partial\Omega$**	On ∂S	**On $\partial\Omega$**
8	0.006	**0.01**	-1.7× e-4	**-1.7× e-4**	1.0× e-3	**0.002**
15	0.002	**0.008**	2.0× e-5	**9.2× e-5**	2.4× e-4	**0.001**

From Tab.2, we can draw the following conclusions: suppose the given boundary value $V_{\partial S}$ is error-free, then, based on the boundary value $V_{\partial S}$ and the "fictitious compress recovery" method, after 15-times iterative procedures, one gets a fictitious (regular harmonic) field $V^*(P)(P \in \overline{K})$, which coincides with the real field $V(P)$ on the "Earth's surface" (seeing the black numbers in Tab.2) under the accuracy (RMS) level 0.1 mm (note that $0.001 m^2 s^{-2}$ corresponds to the height 0.1 mm), and based on the extreme value principle (e.g., Kellogg, 1929) we can conclude that the fictitious field $V^*(P)(P \in \overline{K})$ coincides with the real field $V(P)$ in the whole domain outside the "Earth" at least under the accuracy (RMS) level 0.1 mm, which is confirmed by further experiments, seeing Tabs.3 and 4 (note that in theory, the experimental tests summarized in Tabs.3 and 4 are not necessary).

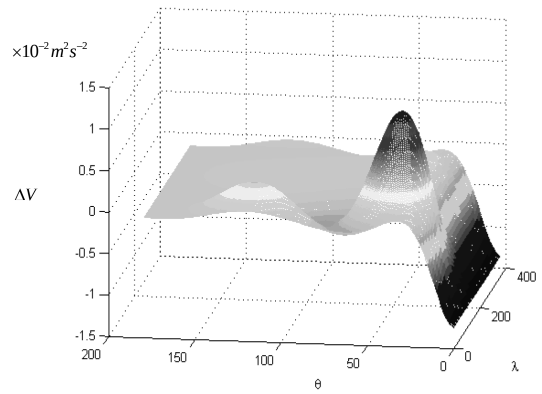


Fig. 2 Residual potential values on the surface of the "Earth" (From Shen and Li, 2005)

Quite arbitrarily, 12 test field points, which are located in the domain between the satellite surface and the surface of the "Earth", are chosen, and then, the residual values (i.e., the differences between the calculated fictitious values and the corresponding real values) on those points are calculated. The results are listed in Tab.3, from which it can be seen that the fictitious field coincides with the real field (at least at the chosen points) under a high accuracy (RMS) level, around 0.08 mm.

Tab. 3 Results at test points in the domain between two boundaries ($\Delta V = V - V^*$; unit: m²s⁻²)

(r, θ, λ) [km, deg, deg]	ΔV	(r, θ, λ) [km, deg, deg]	ΔV
(6380, 0, 0)	-0.0026	(6380, 90, 120)	-8.15e-7
(6600, 0, 0)	-0.0014	(6420, 100, 250)	-9.65e-8
(6400, 15, 0)	-2.62e-4	(6610, 120, 60)	5.01e-7
(6500, 10, 60)	-1.07e-4	(6500, 125, 200)	-2.72e-10
(6550,2 5, 180)	-2.55e-5	(6480, 135, 60)	6.81e-7
(6600, 20, 270)	-5.18e-5	(6570, 150, 210)	-9.68e-8
$mean(\Delta V)$	-3.70e-4		
RMS	7.67e-4		

Further, 5 test field points, which are located in the domain outside the satellite surface, are chosen, and then the residual values on those points are calculated. The results are listed in Tab.4, from which it can be seen that the fictitious field coincides with the real field (at least at the chosen points) under the accuracy (RMS) level around 0.03 mm.

Tab.4 Results at test points in the domain out side the satellite surface ($\Delta V = V - V^*$)

(r, θ, λ) [km, °,°]	(6700, 0, 0)	(6750, 30, 90)	(6800, 60, 120)	(6720, 120, 240)	(6850, 180, 0)
ΔV [m^2s^{-2}]	-6.8× e-4	1.6× e-5	-2.2× e-6	-4.8× e-11	-5.4× e-4
$mean(\Delta V)$	-2.4e-4 m^2s^{-2}				
RMS	3.0e-4 m^2s^{-2}				

In summary, the "fictitious compress recovery" is valid and reliable, based on which the "downward continuation" problem is satisfactorily solved.

Previous to the above mentioned simulation test, Shen and Wang et al (2005) completed an experimental test, which is summarized as follows. Two spherical surfaces ∂K_1 and ∂K_2 with radii $R_1 = 6378$ km and $R_2 = 6680$ km were chosen (Cf. Fig.3), where ∂K_1 and ∂K_2 simulate the Earth's surface (exactly saying the surface of Brillouin sphere) and the satellite surface, respectively.

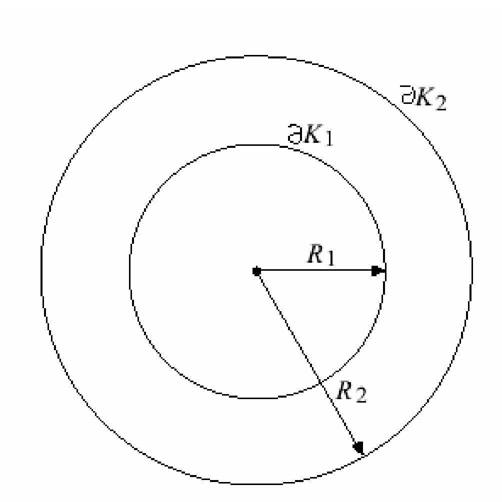


Fig.3 ∂K_1 and ∂K_2 simulate the surface of the Earth or Brillouin sphere and the satellite surface, respectively

With EGM96 model the potential value $V|_{\partial K_1}$ on ∂K_1 is known, and using Poisson integral the potential value $V|_{\partial K_2}$ on the "satellite" surface was calculated, which was assumed to be the "observations", that means the potential values $V|_{\partial K_2}$ on the satellite surface were taken as the initial boundary values. Then, with $10^o \times 10^o$ grid and based on the "fictitious compress recovery" method the fictitious field $V^*(P)$ was calculated, which is compared with the "real" value $V|_{\partial K_1}$ on ∂K_1. The largest difference between the real value $V_1|_{\partial K_1}$ and the calculated fictitious value $V^*|_{\partial K_1}$ is 0.04 m^2s^{-2}, which corresponds to a height variation 0.4 cm. Hence, the experimental test (Shen and Wang et al, 2005) supports the "fictitious compress recovery" method and the "fictitious downward continuation".

It is noted that the satellite surface ∂S can be also replaced by a spherical surface, which completely encloses the Earth (Cf. Remark 2). The simulation test in details as well as various other simulation tests will be provided in a separated paper.

6 Conclusions and discussions

If the geopotential on the Earth's boundary $\partial\Omega$ is determined, the Earth's external potential field $V(P)$ can be determined based on the "fictitious compress recovery" method. To realize this, the GPS "geopotential frequency shift" approach (Shen et al, 1993) is proposed. However, it is most likely that in the very near future it is very difficult to realize this approach in practice, due to the fact that the accuracy of the determined potential difference by using the GPS "geopotential frequency shift" approach is too low, which depends on the frequency stability of the signal receiver. At present, the frequency stability is around $10^{-15} - 10^{-16}$ (HMC Project, 2005), which corresponds to the height variation about 1m.

If the gravitational potential $V(P)$ on the satellite surface ∂S or the surface ∂K_Γ of a sphere K_Γ (which completely encloses the whole Earth) is determined, e.g., using the energy integral approach (Cf. Gerlach et al, 2003; Visser et al, 2003), the Earth's external potential field $V(P)$ can be also determined based on the "fictitious compress recovery" method (Cf. Remark 2). This is a new approach for solving the "downward continuation" problem, referred to as the "fictitious downward continuation" (Cf. Sec.4). To realize this, it can be first determined the field $V_1(P)$ outside the satellite surface (or outside the sphere K_Γ) based on the spherical harmonic expansion (14) and by using GPS observations, and then the whole field outside the Earth could be determined, or the potential field could be directly determined if the determined boundary values are uniformly distributed on the satellite surface. Hence, in any case, the real field $V(P)$ in $\overline{\Omega}$ can be exactly determined based on the boundary value $V_{\partial S}$ (or $V_{\partial K_\Gamma}$) and by using the "fictitious compress recovery" method.

At present, the determined position deviation (from the real position) by on-board GPS receiver is around 5 cm, which gives rise to the velocity deviation about 10 cm/s, if the time keeping error is neglected. Consequently, the deviation of the "observed gravitational potential" on the satellite surface due to the position deviation is around 0.01 m^2/s^2 (of course this is not the real case, because the accuracy of the accelerometer is relatively low). Hence, neglecting other error sources, the determined potential field has the deviation around 0.01 m^2/s^2.

Generally, suppose the accuracy of the given value on the satellite surface (several hundred kilometres above the Earth's surface) is $\sigma_{\partial S}$, then, the accuracy $\sigma(r)$ of the determined fictitious field (based on the "fictitious compress recovery" method) is on the same accuracy level as σ in the domain between the Earth's surface and the satellite surface, expressed by the following relation (Shen and Tao, 2004):

$$\sigma(r) = \frac{r}{R}\sigma_{\partial S} \equiv (1+\frac{h}{R})\sigma_{\partial S} \tag{19}$$

where R is the average radius of the Earth, h is the height of the field point above the Earth's surface.

Remark 2: The satellite surface ∂S can be replaced by the surface ∂K_Γ of a sphere K_Γ that encloses the whole Earth. In this case, Eqs.(13) and (14) hold also in the domain $\overline{K}_\Gamma$, the domain outside the sphere K_Γ (there does not exist divergence problem any more in the domain $\overline{K}_\Gamma$). Based on Eq.(14) and satellite observations, the field $V_1(P)$ in the domain $\overline{K}_\Gamma$ could be determined. Then, after the "fictitious compress recovery" method is applied, the real field $V(P)$ in the whole domain $\overline{\Omega}$ outside the Earth could be determined, under the assumption that $V_1(P)$ coincides with the real field $V(P)$ in the domain $\overline{K}_\Gamma$

Acknowledgements

The authors would like to express their sincere thanks to the two anonymous reviewers, who provided valuable comments and suggestions to the original manuscript. This study is funded in parts by Natural Science Foundation China (No.40374004), and the Foundation of the Key Laboratory of Geospace Environment and Geodesy, Ministry of Education China (the Surveying and Mapping Foundation China 03-04-13).

References

Bjerhammar A (1964): ***A new theory of geodetic gravity.*** Royal Institute of Technology, Geodesy Division, Stockholm.

Brumberg V.A.; Groten E. (2002): ***On determination of heights by using terrestrial clocks and GPS signals.*** J. Geod. 76: 49-54.

Gerlach Ch.; Földvary L.; Svehla D.; Gruber Th.; Wermuth M.; Sneeuw N.; Frommknecht B.; Oberndorfer H.; Peters Th.; Rothacher M.; Rummel R.; Steigenberger P. (2003): ***A CHAMP-only gravity field model from kinematic orbits using the energy integral.*** Geophysical Research Letters 30(20), 2037.

Heiskanen W.A.; Moritz H. (1967): ***Physical Geodesy.*** Freeman and Company, San Francisco.

HMC Project (2005): ***http://cfa-www.harvard.edu/hmc/***

Katila T.; Riski Q. (1981): ***Measurement of the interaction between electromagnetic radiation and gravitational field using ^{67}Zn Mössbauer spectroscopy.*** Physics Letters, 83A: 51-54.

Kellogg O.D. (1929): ***Foundation of potential theory.*** Frederick Ungar Publishing Company, New York.

Li Jin (2005): ***Simulation tests on the fictitious compress recovery method for gravitational potential.*** Bachelor Degree Thesis, Wuhan University, Wuhan, China.

Moritz H. (1978): ***On the convergence of the spherical harmonic expansion of the geopotential at the Earth's surface.*** Bollettino di Geodesia e Scienze Affini, 37: 363-381.

Pound R.V.; Snider J.L. (1965): ***Effect of gravity on gamma radiation.*** Physical Review, 140(3B): 788-803.

Rummel R.; Sansó F.; van Gelderen M.; Brovelli M.; Koop R.; Miggliaccio F.; Schrama E.; Sacerdote F. (1993): ***Spherical harmonic analysis of satellite gradiometry.*** Netherlands Geodetic Commission, Publications on Geodesy, New Series No. 39

Shen W.B. (1995): ***A general expression of the Earth's gravitational potential's series expansion.*** Bollettino di Geodesia e Scienze Affini, 54: 361-372.

Shen W.B. (1998): ***Relativistic physical geodesy. Habilitation*** Thesis, Graz Technical University, Graz, Austria.

Shen W.B. (2004): ***The fictitious compress recovery method for gravitational potential.*** Geomatics and Information Science of Wuhan University 29(8): 720-724

Shen W.B.; Chao D.B.; Jin B.R. (1993): ***On the relativistic geoid.*** Bollettino di Geodesia e Scienze Affini, 52: 207-216.

Shen W.B.; Li Jin (2005): ***Simulation tests on the fictitious compress recovery method and the fictitious downward continuation.*** Chinese Journal of Geophysics, (submitted).

Shen W.B.; Ning J.Sh. (2004): ***A proposal on the determination of the Earth's potential field.*** Presented at the 18th International Meeting of CEOS, November 16-18, 2004, Beijing, China.

Shen W.B.; Tao B.Z. (2004): ***The accuracy analysis of the gravity model based on the fictitious compress recovery.*** Presented at the 2nd AOGS General Assembly, July 6-9, 2004, Singapore

Shen W.B.; Wang Zh.T.; Chao D.B. (2005): ***An approach for determining the Earth's external gravity field by using satellite gravity data and a simulation experiment test.***

Geomatics and Information Science of Wuhan University, (accepted).

Shen W.B.; Yan J.G.; Chao D.B. (2005): ***The application of the fictitious compress recovery approach in "downward continuation" and simulation experiment tests.*** Acta Geodaetica et Cartographica Sinica, (submitted).

Shen W.B.; Zhong Q.J.; Han M.T.; Ning J.Sh. (2004): ***The fictitious compress recovery principle and its applications in geophysics.*** Annual of the Chinese Geophysical Society 2004. Xi'an Map Press, Xi'an, China, pp. 20-21.

Sjöberg L. (1980): ***On the convergence problem for the spherical harmonic expansion of the geopo- tential at the surface of the Earth.*** Bollettino di Geodesia e Scienze Affini, 39: 261-271.

Vessot R.F.C.; Levine M.W.; Mattison E.M.; Blomberg E.L.; Hoffman T.E.; Nystrom G.U.; Farrel B.F.; Decher R.; Eby P.B.; Baugher C.R.; Watts J.W.; Teuber D.L.; Wills F.D. (1980): ***Test of relativistic gravitation with a space-borne hydrogen maser.*** Physical Review Letters, 45: 2081-2084.

Visser P.N.A.M.; Sneeuw N.; Gerlach C. (2003): ***Energy integral method for gravity field determination from satellite orbit coordinates.*** J. of Geod. 77: 207-216.

Weinberg S. (1972): ***Gravitation and cosmology.*** John Wiley & Sons, New York.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 277-283

Journal of Global Positioning Systems
2005
© CPGPS

Joint Australian Engineering (Micro) Satellite (JAESat) - A GNSS Technology Demonstration Mission

Werner Enderle

Cooperative Research Centre for Satellite Systems at the Queensland University of Technology, GPO Box 2434, Brisbane QLD 4001, Australia
e-mail: w.enderle@qut.edu.au , Tel: +61 (0)7 3864 1436, Fax: +61 (0)7 3864 1517

Cameron Boyd

Australian Space Research Institute Ltd (ASRI), PO Box 3890, Manuka ACT 2603, Australia
e-mail: cameron.boyd@concepts.aero, Tel: +61 (0)7 3366 0498

Jan A. King

Southern Cross Space and Communications PTY. LTD, Lake Weyba QLD, Australia
e-mail: jking@eclipticenterprises.com , Tel: +61 (0)7 5471 0657

Received: 27 November 2004 / Accepted: 12 July 2005

Abstract. JAESat is a joint micro-satellite project between Queensland University of Technology (QUT), Australian Space Research Institute (ASRI) and other national and international partners, i.e. Australian Cooperative Research Centre for Satellite Systems (CRCSS), Kayser-Threde GmbH, Aerospace Concepts and Auspace which contribute to this project. The JAESat project is conducted under the leadership of the Queensland University of Technology. The main objectives of the JAESat mission are the design and development of a micro satellite in order to educate and train students and also to generate a platform in space for technology demonstration and conduction of research on a low cost basis. The main research objectives of JAESat are the in-orbit test and validation of the SPARx receiver and its performance, the performance of the on-board Orbit Determination (OD) concept, the test of an integrated GPS-Star Sensor system concept for a 3-axis Attitude Determination (AD) and its related algorithms and also various aspects of Relative Navigation. The aspects of atmospheric research will not be addressed within this article. This article will describe the overall JAESat concept and concentrate on the QUT space applications receiver SPARx and related GPS software concepts for OD and AD. The test environment for the development of GNSS space applications will be outlined and finally simulations and respective results including GPS hardware in the loop will be presented and discussed.

Key words: Spacecraft navigation, attitude determination, orbit determination, GPS space receiver, GNSS space applications, micro satellites

1 Introduction

The JAESat mission outlined in Enderle (2002, 2004) will ultimately consist of two micro-satellites (see Figure 1) which will fly in a formation. The JAESat micro-satellite itself will have two components, a master satellite and a so-called slave satellite. The components of JAESat will be attached to each other during the launch phase and will be separated in space, after the release of JAESat from the launcher. The JAESat mission is designed to conduct a variety of experiments based on the mode of interoperation between the payloads on-board the two satellites. A communication link between the two satellites will be established in the form of a RF Inter-Satellite Link (ISL). It is anticipated that JAESat will be launched in 2006. Negotiations with a launch provider for a piggy back launch are ongoing. For this reason the final orbit is not definite yet. However, it is intended to have a circular, nearly polar orbit with an orbit altitude of about 800 km. The operational life time of JAESat is expected to be round 12 months. After the separation of the slave from the master satellite the two satellites will drift away from each other with a low drift rate. JAESat is designed to have a high degree of on-board autonomy. The operations will be conducted via a ground station located at the Queensland University of Technology in Brisbane, Australia.

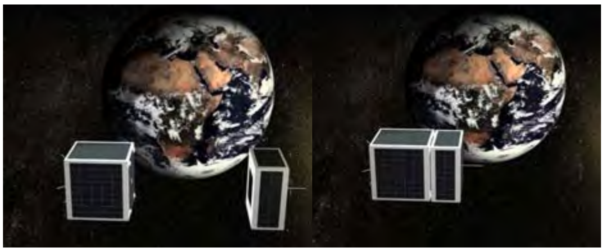


Fig 1 JAESat master and slave satellite concept before and after separation in space

2 JAESat Mission Concept

The JAESat micro-satellite project is an educational and GNSS technology demonstration mission, which will also generate data for scientific use. JAESat's high level mission objectives are:

- Design, develop, manufacture, test, launch and operate the educational/research micro-satellite JAESat
- Develop payloads with a technological and scientific relevance
- Use JAESat as a sensor in space and GNSS technology demonstrator mission

As can be seen in the high level mission objectives, the education and training aspects play an important role in the JAESat mission. The GNSS mission objectives are driven by the SPARx (SPace Applications GPS Receiver), a development by the Cooperative Research Centre for Satellite Systems/Queensland University of Technology. Functions and performance of SPARx will be tested and validated in space within the JAESat mission. A key element of the GNSS activities will be the testing of a new sensor concept for attitude determination, based on Star Sensor and GPS attitude information.

As already noted, JAESat will consist of two micro-satellites (master and slave) flying in a formation. JAESat master and slave satellites will fly in the same orbital plane. After the split between master and slave, the two micro satellites will then slowly drift away from each other. The master satellite will be a cube with a side length of 390mm. The slave satellite will have the following dimensions 390mm x 390mm x 195mm, which is half the height of the master satellite. The JAESat master satellite will be 3-axis stabilized, whereas the JAESat slave satellite will be gravity gradient stabilized. The mass of the slave satellite will be around 10kg, and the mass of the master will be around 30 kg, so that the total mass of JAESat will be around 40 kg. The orbit of JAESat will be a Low Earth Orbit (LEO) with an expected altitude of 800 km and an orbit inclination of around 90 deg. The ground track of JAESat is outlined in Figure 2.

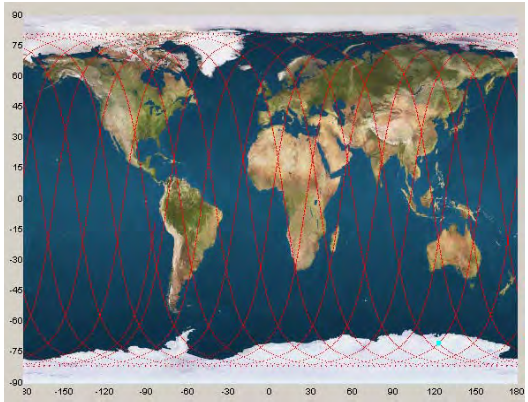


Fig 2 JAESat ground track

3 JAESat Satellite System

3.1 Structure

The JAESat structure concept is based on a tray approach. JAESat will have a total of eight trays. Figure 3 provide an overview of the JAESat structure concept.

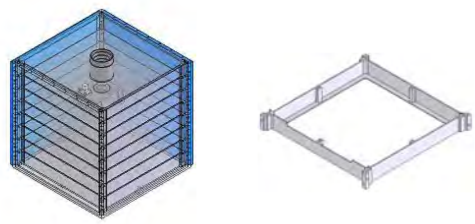


Fig 3 JAESat structure – tray concept

3.2 Power

The JAESat power system will consist of batteries and solar cells. The available power at the master satellite will be around 22 Watts (min) orbit average, based on the current JAESat design and the chosen solar cells.

3.3 Flight Computer

The JAESat on-board flight computer will be an Intrinsyc CerfBoard. The operating system is based on LINUX. The processor is an Intel XScale PXA255 microprocessor @ 400 MHz. The size of the board is:57.2mm x 69.9mm x 25.4mm. Power: 5V DCregulated, 400mA with no

CompactFlash device; peak of 1.1A with CompactFlash. The board has also a Battery Backed Real-Time Clock

3.4 Communication

The overall communication concept is outlined in Figure 4. The receiver is capable of 9600 baud and has low power and space requirements.

Fig 4 JAESat Communication Concept

3.5 Attitude Control System (ACS)

The master satellite will be 3-axis stabilized by using magnet torquers (air coils). The slave satellite will be gravity gradient stabilized without using a boom. Instead the moments of inertia will be designed so that a gravity gradient stabilization will be the result. After separation from the launcher, the JAESat master and slave will still be attached to each other. JAESat master ACS will than reduce the rotation rates around each axis and finally orient the satellite in such an orientation that the slave will be in its gravity gradient orientation and than JAESat will split into two satellites. After splitting into two satellites, JAESat will drift away from each other. The ACS of the master will be used for controlling the orientation and changing of rotation rates, necessary for testing of the new integrated Star Sensor GPS attitude sensor concept. Only one requirement for the attitude accuracy of the master has been derived, resulting from the need to have an Inter Satellite Link established. The master attitude accuracy requirement is 5 deg. The overall ACS concept is described in Figure 5.

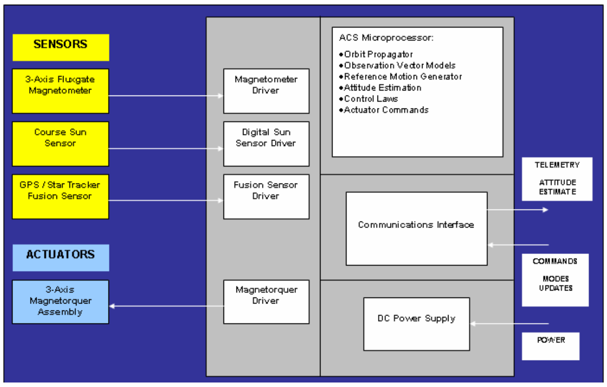


Fig 5 JAESat ACS Concept

4 JAESat Payloads

The JAESat payloads concept is driven by simplicity. The payload itself will be distributed between the JAESat master and slave satellite. One of the positive aspects of this distributed concept is that in the event of problems on one of the two satellites, or in the worst case scenario, the loss of the slave satellite, significant research can still be conducted.

The JAESat master satellite will have the following payloads on board: 1. The SPARx – GPS receiver for positioning (absolute and relative), on-board orbit determination and 3-axis attitude determination. 2. The Star Sensor- KM 1301 for 3-axis attitude determination. 3. Specific antennas for conduction of atmospheric research.

The JAESat slave satellite will have the following payloads on board: 1. SPARx - GPS receiver for positioning (absolute and relative). 2. A mini Video Camera (Web camera type). 3. Specific antennas for conduction of atmospheric research.

4.1. GPS Receiver – SPARx

The CRCSS/QUT GPS SPARx (see Figure 6) development is based on the MITEL GP2021, GP2015 and GP2010 Chip set and is a modification of the MITEL Orion GPS receiver demonstrator. The base for the development of the source code is the MITEL GPS Architect development kit. The source code modifications are specifically targeted towards robust and accurate operation on-board a satellite (see Enderle and Roberts 2003). Key elements of functionality are positioning and timing for satellite applications. Ongoing R&D activities are the development of a GPS receiver for on-board orbit determination (SPARx-OD) and also GPS receiver for satellite attitude determination (SPARx-AD). On-board the JAESat master satellite, either one SPARx-AD or three SPARx will fly with the capability of performing 3-

axis attitude determination. The on-board orbit- and attitude determination will be performed within the Flight Computer. The main characteristics of a GPS SPARx are given in Table 1.

Fig 6 Orion GPS demonstration board

Tab. 1 SPARx – physical characteristics of GPS receiver and interface board

Physical Characteristic – SPARx GPS Receiver Board	
Power (Vdd into receiver board)	+5 volts DC, +/- 10%
ACTIVE-ANTENNA POWER	+5 volts DC available
POWER CONSUMPTION (Vdd)	GPS Receiver Board only: 370mA 1.85W With Antenna: 395mA 1.98W
SIZE: (GPS Receiver Board only)	95mm x 50mm x 20mm
SIZE: (GPS Receiver Board plus Interface Board)	95mm x 50mm x 30mm
CONNECTORS	RF: SMA socket I/O: 9-pin(1x9), 0.1” pitch plug
Physical Characteristic - SPARx Interface Board	
INPUT POWER SUPPLY	+8 volts to +30volts DC
MEMORY / RTC BACKUP BATTERY	+3.6 v NiCad, 110mAhrs
BATTERY CHARGE RATE TOTAL RECHARGE TIME	7.5mA typical 18 hours approx
BATTERY BACKUP LIFE (when supplying Vdr pin on Orion receiver board)	8 weeks approx
DATA I/O LEVELS	RS232 levels (±10v)

A conceptual diagram for the JAESat attitude determination based on GPS is outlined in Figure 7. In addition to orbit and attitude determination, it is also intended to perform relative navigation between the JAESat master and slave satellite. Finally, SPARx will be used for collecting data from specific GPS antennae attached to the sides (looking to the horizon) of JAESat in order to perform atmospheric research.

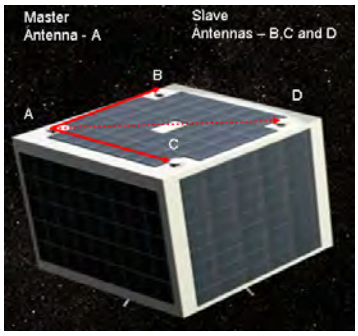


Fig 7 SPARx antenna array for JAESat 3-axis attitude determination. antenna D will be optional

4.2 Star Sensor – KM 1301

The Star Sensor KM-1303 is a contribution of the German Aerospace Company Kayser-Threde GmbH towards the JAESat project. . The Star Sensor KM 1301 is a low-cost single-package design for star recognition, relative and inertial attitude determination. This Star Sensor will be used in the testing of a new integrated attitude determination sensor concept. Star Sensors are the most accurate sensor types for satellite attitude determination. However, these types of sensors have some limitations/restrictions related to their range of operations. One limiting factor is the sensitivity of such a sensor systems with respect to rotation. If the satellite rotation rates are too high (the actual rates depending on the individual model, for the KM-1301 the angular rates are around 5 deg/sec), the Star Sensor will have difficulties with the identification of the stars and therefore an attitude determination can not be performed. However, this problem can be solved by using additional, external information from the GPS attitude sensor. The probability of achieving a star identification and an attitude solution, resulting from an attitude determination process will for this reason increase substantially.

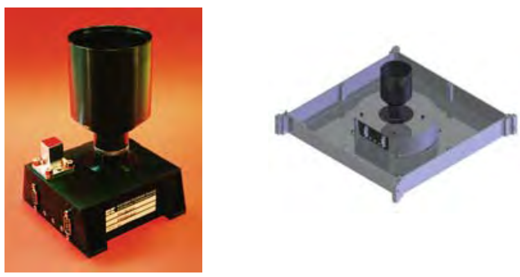


Fig 8 Star Sensor KM 1301 and its integration into the tray structure

5 JAESat Experiments, Testing and Simulations Environment

5.1 Experiments

The JAESat main experiments can be summarized in the following way; First, testing and evaluation of CRCSS/QUT GPS SPARx, including Attitude capability. Second, testing of a new integrated Star Sensor/GPS navigation sensor concept for 3-axis attitude determination. Third, Relative Navigation between JAESat Master and Slave satellite. Fourth, testing and evaluation of different Orbit Determination concepts, On-Ground - Precise orbit determination based on GPS Code and Carrier phase measurements and On-board orbit determination based on GPS receiver position solutions. Fifth, establishment of stable RF inter satellite links and seventh, atmospheric research.

5.2 Testing and Simulations Environment

The testing and simulations environment (shown in Figure 9) at QUT involves several aspects. The core of the testing environment is the Welnavigat GPS signal simulator. The signal simulator is capable of simulating the entire GPS satellite constellation and transmitting RF signals on six channels. Various simulations are possible including a satellite orbit scenario. An important feature in this context is the option to import scenario files, generated by the user.

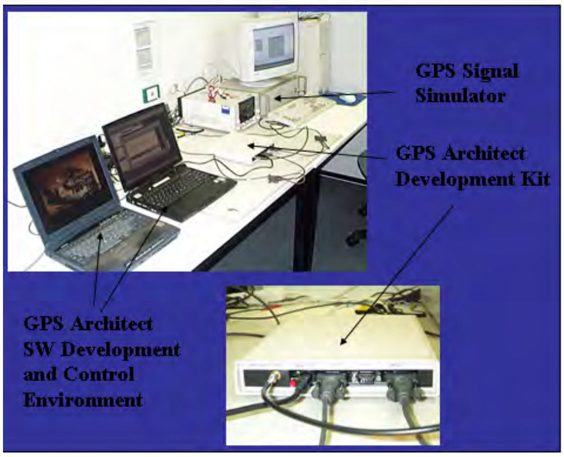


Fig 9 GPS testing and simulation environment

6 JAESat Operations

JAESat is designed for a high degree of on-board autonomy. However, the operations of JAESat will be conducted via a ground station located at the QUT in Brisbane, Australia. The number of ground station over flights is in the order of 4 – 6 per day. JAESat will be visible by the ground station for a period of 8 – 12 min per overflight. In Figure 10, the ground antenna, ground equipment and visibility plots are shown.

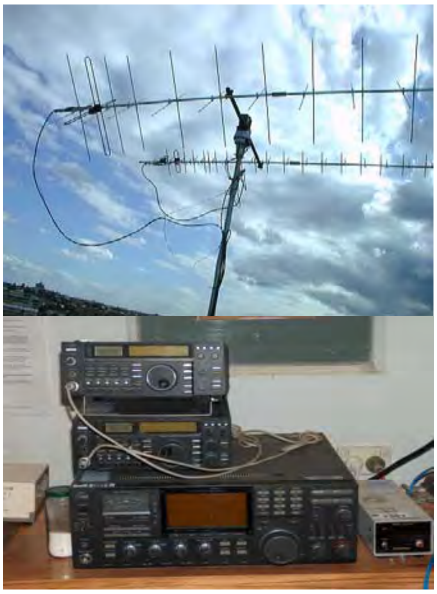


Fig 10 JAESat ground station at the Queensland University of Technology in Brisbane, Australia

7. JAESat – Testing, Simulations and Results

Testing and simulations are ongoing activities in order to demonstrate the feasibility of the proposed GNSS experiments and also obtain information about functionality and performance. The results shown here are related to the expected performance of the SPARx in space, the position, velocity and time solution, the capability of Orbit Determination (OD) and also the performance related to the Attitude Determination (AD) based on GPS measurements. Test and simulations have been conducted with and without Hard Ware in the loop.

7.1 Results from GPS Signal Simulator and SPARx

The main objective of these tests was to identify the GPS signal acquisition and tracking performance of SPARx based on different code versions. Results of tests conducted for JAESat including the GPS signal simulator and SPARx are given in Figure 11 and Figure 12. As it can be seen, the acquisition, reacquisition and the tracking behaviour of SPARx is good. SPARx generates a 3D position solution for more than 80% of the time per orbit within this HW in the loop simulation. In this context it is also important to understand that the GPS signal simulator only provides a total of six simulated channels. This means that these tests can be seen as a kind of worst case scenarios. One of the test objectives was also the testing of the time synchronisation performance and the generation of a Hard Ware Pulse Per Second (PPS) output based on code modifications (Bruggemann and Enderle 2004). Currently SPARx generates an HW PPS with an accuracy of better than 50nsec (1 Sigma) steered towards UTC. SPARx performance for a 3D absolute position solution in space is better than 25 m (1 Sigma).

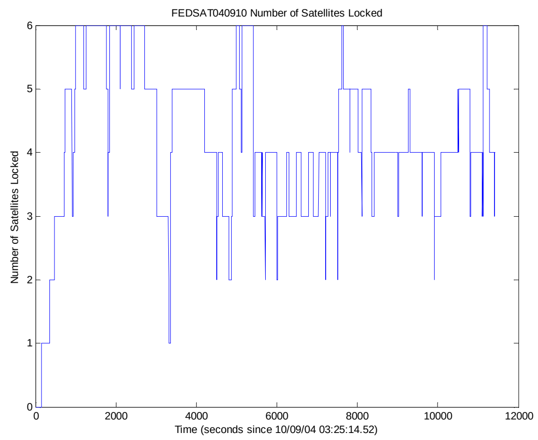


Fig 11 SPARx tracked GPS signals for a typical orbit scenario

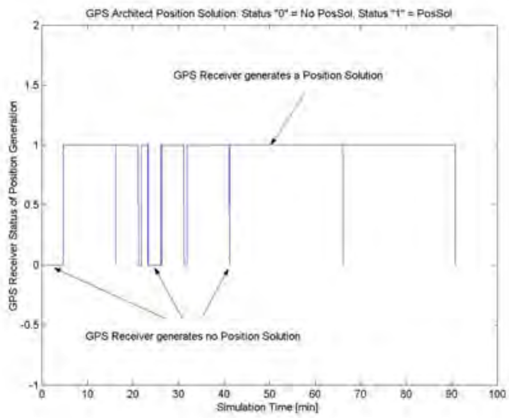


Fig 12 SPARx acquisition and tracking performance

7.2 Simulations for Orbit and Attitude Determination based on GPS

Simulations for the Orbit and Attitude Determination have been conducted. The results are presented in Table 2 (see also Figure 13). For the attitude simulations, the orientation of the antenna array was anti nadir. The baseline length was 36 cm for baseline AB, AC and 51 cm for baseline AD. The multipath error was assumed to be 3mm on the carrier Single Difference (SD). The Orbit Determination concept is based on GPS position solutions Enderle, Feng and Zhou. (2003), a total data arc of 6 hours was used with a position solution every 10 sec. For the generation of a reference orbit, an Earth gravity field with order and degree of 20 was used and numerically integrated. No other forces have been taken into account.

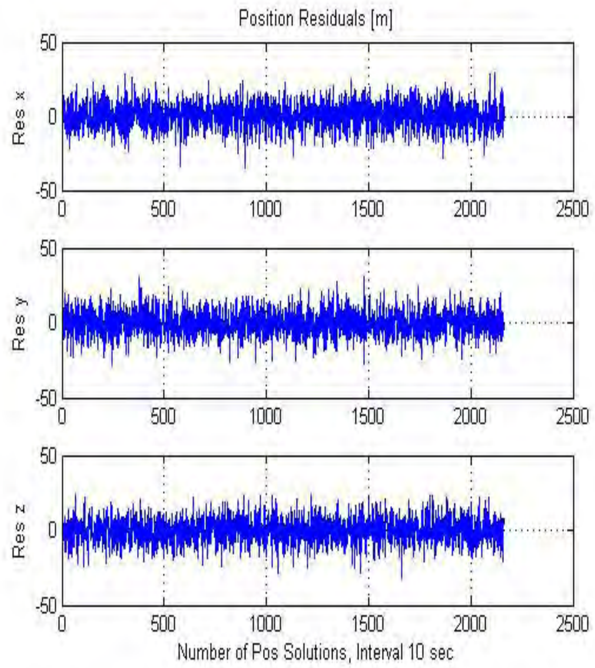


Fig 13 Position residuals for a simulated 6 hour orbit scenario

Tab. 2 Orbit Determination (OD) based on GPS position solutions. The results are obtained from a simulation. The position solution error was simulated with a 1 Sigma value of 25 m.

	Estimated State Vector	**Error wrt Reference State Vector**
Position x [m]	-3283178.224	-17.819
Position x [m]	-3652268.345	3.330
Position x [m]	5251321.252	-4.850
Velocity x [m/s]	2323.540285	0.002362
Velocity y [m/s]	5037.512456	-0.000726
Velocity z [m/s]	-4954.785328	0.000864

The results in Table 2 show that a satellite 3D position accuracy of around 18m can be achieved. The implementation of the code onto the on-board flight computer is one of the next steps. The implementation of the OD code will than be evaluated and optimised for robustness and performance.

The results for the JAESat Attitude Determination (AD) based on GPS, presented in Figure 16 are highlighting that GPS based AD would already by sufficient to comply with the JAESat attitude accuracy requirements. In Figure 14, it can also be clearly seen that the number of visible GPS satellites for JAESat lies between 4 SV and 12 SV with an average of about 8 SV. This means a substantial higher number of visible GPS satellites for the JAESat as have been used within the simulations using the GPS signal simulator (only 6 physical channels can be simulated).

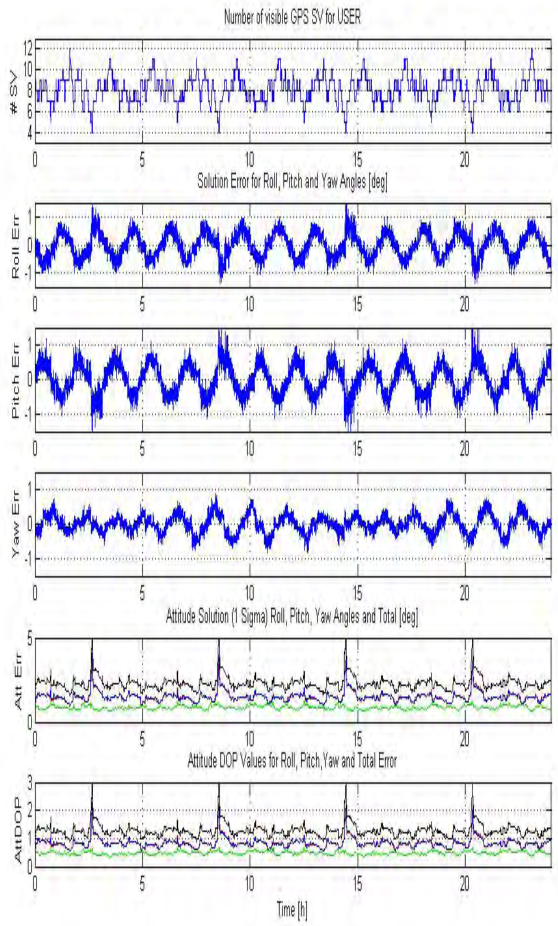


Fig 14 Attitude determination results for a simulated JAESat scenario

8. Conclusions

The tests and simulations have clearly demonstrated the feasibility of the JAESat GNSS experiments in terms of absolute positioning, timing, on board orbit and attitude determination. Further development and testing will be undertaken in order to cover the relative navigation aspects. Special emphasis will be given in the near future for the testing of the integrated attitude sensor based on Star Sensor and GPS attitude information. Between 2003 and 2004 a total of 24 students have worked on the JAESat project in various areas. This means that one of the high level mission objectives has already been achieved – the use of JAESat as an education and research platform.

References

Enderle W. (2002) : ***QUT - JAESat Project Documentation, JAESat – GPS - PO-0001***, QUT Project Documentation

Enderle W. (2004): ***JAESat Mission Concept Description***, QUT Project Documentation (Internal Info)

Enderle W.; Feng Y.; Zhou N. (2003):***Orbit Determination of FedSat based on GPS Receiver Position Solutions – First Results***, Proceeding SatNav2003, Melbourne, Australia, 22–25 July 2003

Enderle W.; Roberts P. (2003**): *GPS Receiver Development for Space Applications at Queensland University of Technology (QUT*),** Proceeding of the 10th Australian International Aerospace Congress and 14th National Space Engineering Symposium, Brisbane, Australia, 29 July – 1 August 2003

Bruggemann T.; Enderle W. **(**2004**): *Time Synchronization Capability of SPARx – the Queensland University of Technology (QUT) GPS Receivers*,** GNSS 2004, The 2004 International Symposium on GNSS/GPS, Sydney, Australia, 6–8 December 2004

Mitel Semiconductor.; DS4605 (1997): ***GPS Architect 12 Channel GPS Development System***

Mitel Semiconductor, AN4808 (2001): ***GPS Orion 12 Channel GPS Reference Design***

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2: 284-290

Journal of Global Positioning Systems
2005
© CPGPS

Grid Residual Tropospheric Corrections for Improved Differential GPS Positioning Over the Victoria GPS Network (GPSnet)

Yi Zheng, Yanming Feng, Zhengdong Bai

Cooperative Research Centre for Satellite System (CRCSS), Queensland University of Technology (QUT), Brisbane, Australia
e-mail: y.zheng@qut.edu.au Tel: +61-7-38641362; Fax: +61-7-38641517

Received: 8 December 2004 / Accepted: 12 October 2005

Abstract. Tropospheric delay is one of the major error sources in GPS positioning. The delay of radio signals caused by the troposphere can range from 2 m at the zenith to 20 m at lower elevation angles. In a wide area differential system, tropospheric delays are corrected locally by users using an empirical tropospheric model, with or without meteorological observations. This can easily result in residual tropospheric errors of several centimetres to a few decimetres in positioning solutions. In this paper, the residuals between GPS-derived zenith tropospheric delays (ZTD) and model-computed ZTDs at reference stations of a continuously operating network are obtained. From these residuals, grid residual ZTDs are generated over the network coverage through Ordinary Kriging interpolation. Users can obtain the additional residual corrections for troposphere errors for improved differential positioning over a regional area. Experimental data from one week ZTD estimates from 17 GPSnet reference stations were analysed. Results show that the RMS ZTD accuracy of about 1cm is generally achievable over the Land Victoria GPS network coverage, using the proposed grid tropospheric correction strategies, which can support centimetre level positioning in the region.

Key words: Residual zenith tropospheric delay, Ordinary Kriging, Tropospheric grid model, Wide Area DGPS

1 Introduction

Tropospheric delay is currently one of the major error sources in satellite-based positioning. The delay of radio signals caused by the troposphere can range from 2 m at the zenith to 20 m at lower elevation angles (e.g. below 10 degrees), depending on the pressure, temperature and humidity along the path of the signal transmission. If the tropospheric delay is not properly modeled, positioning accuracy can be degraded significantly.

In the current code-based Wide Area Differential GPS (WADGPS) positioning services, unlike ionospheric corrections, the tropospheric delay corrections are not broadcast to the users. Instead, the delays are supposed to be corrected locally by using an empirical tropospheric model and a mapping function adopted by the users. Wide Area Augmentation System (WAAS) and European Geostationary Navigation Overlay System (EGNOS) specifications recommend the application of an empirical model correction algorithm, based on the estimates of five meteorological parameters: pressure, temperature, water vapour pressure, temperature lapse rate and water vapour lapse rate. These meteorological parameter estimates are dependent on the receiver's height, latitude and day-of-year, using yearly average and associated seasonal variation data (Collins et al., 1997, Dodson et al., 1999). However, the residual delays after modelling are at a level of a few centimetres at the zenith, which may lead to a single point positioning (SPP) error of up to a few decimetres.

GPS geodetic processing techniques have enabled precise estimation of zenith tropospheric delays (ZTD) with an accuracy on the order of 1 mm for every 30 minutes. Recent studies have demonstrated that the relative ZTDs over a regional network can be estimated in real time (Fang et al., 2003). This provides the technical possibility of using real time ZTD solutions for real time, or near real time precise positioning.

The precise ZTDs estimates derived from ground-based GPS measurements are considered as the "true delays", while the differences between the values computed from a tropospheric delay model and GPS-derived ZTD estimates are defined as the "residual ZTD" estimates. The remaining ZTD model errors are usually ignored by the current WADGPS services. We propose to interpolate the residual ZTDs estimated from reference stations of a

GPS network to generate a residual tropospheric correction grid. The users within the coverage of the network can interpolate residual ZTD from the grid and add it to the ZTD calculated by the empirical model locally.

In our previous research, the ZTD estimates from a network of 129 International GPS Service (IGS) sites across Europe for over 3 months were collected, and the residual ZTDs of each station were directly interpolated by Ordinary Kriging (OK) method using the residual ZTDs from all other stations. From the statistical analysis of each station, the results have concluded that interpolating residual zenith tropospheric delays could be an efficient way to improve user-end ZTD estimation and the precision of differential GPS positioning (Zheng, 2004).

In this paper, additional efforts are made to apply the proposed strategy to the Victoria GPS Network (GPSnet). Further to the previous work, this research will use OK method to interpolate the residual ZTDs of GPSnet stations and to generate a residual tropospheric correction grid at the first place, then obtain the residual ZTDs at given stations interpolated from the grid using the user interpolation algorithm. The efficiency and accuracy of the interpolation procedures will also be analysed.

2 GPS ZTD Estimation

2.1 Victoria GPS Network (GPSnet)

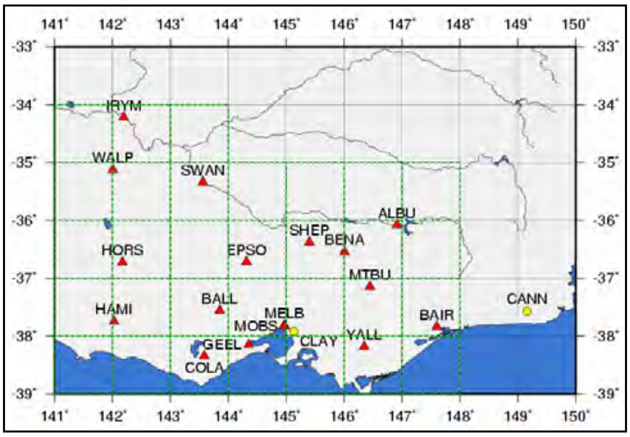


Fig. 1 Victoria GPS Network (GPSnet) site map

The Victoria's GPS Network – GPSnet as shown in Figure 1 is the network of permanent GPS base stations and supporting infrastructure that has been established co-operatively with a range of Industry partners, hosts and contributors; facilitated and operated by Land Victoria. The network records, distributes and archives GPS satellite correction data for accurate position determination, 24 hours a day, state-wide. GPSnet is designed to international and national standards to meet the specific needs of Victorian GPS users. One can refer to GPSnet official website (http://www.land.vic.gov.au) for the current details of base station and operational information.

2.2 Data Collection and Processing

Since the zenith hydrostatic delay can be modelled and removed with an accuracy of a few millimetres or even better using a surface troposphere model, the residual tropospheric delay remaining after applying a standard model is mostly due to the wet component. Therefore, in order to have better prospect of the improvement of ZTD estimation by applying interpolated residual ZTDs, we intentionally chose a recent week from 8th to 14th September, 2004, during which the Victoria state was experiencing an extensive rainfall with the maximum weekly amounts up to 150 mm, as shown in Figure 2.

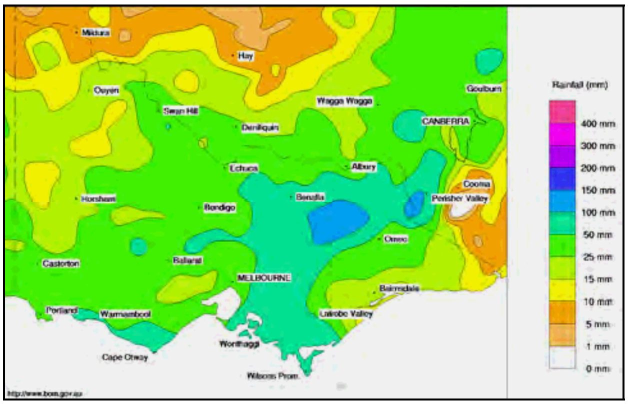


Fig. 2 One week rainfall amounts of Victorian region, from 8th to 14th September 2004, product of the National Climate Centre, Australia, retrieved from www.bom.gov.au)

There are currently 19 base stations within the whole GPSnet. GPS raw measurements from 17 stations, as marked by red triangle in Figure 1, were collected and processed. The other two stations, CANN and CLAY, marked as yellow circle in Figure 1, were excluded from our analysis because of lack of data during the week we chose. The coordinates of all 17 stations of the network were precisely determined and tightly constrained to 5 mm, 5 mm, 10 mm in north, east, and up components, respectively. The corresponding IGS SP3 final orbits were also retrieved and held fixed during the processing.

GAMIT software (King et al., 2000) was used to process the data, which parameterizes ZTD as a stochastic variation from the Saastamoinen model. The variation is constrained to be a Gauss-Markov process with a power density of $2\ cm/\sqrt{h}$). ZTD estimates at 15-minute interval were solved with piecewise linear interpolation (PWL). A sliding window processing strategy (Fang et al., 1998) with a 12-hour window length and a 4-hour forward step was used. However, the Gauss-Markov process provides an implicit constraint on the ZTD estimate at a given epoch from observations at preceding and following epochs, which means that the accuracy is expected to be lower at the beginning and end of each window. Therefore, to avoid the border effects of the Gauss-Markov filter, the central 4-hour ZTD estimates of each 12-hour session were extracted before moving the window forward.

3 Interpolating Residual Zenith Tropospheric Delays

3.1 Tropospheric Delay Model

To obtain the residual ZTDs, the modeled ZTD values need to be calculated first, and then removed from the GPS-derived ZTD estimates. The empirical tropospheric delay model used in our research is the Saastamoinen (SAAS) model (Saastamoinen, 1973), which reads:

$$\begin{aligned} &\mathrm{ZTD} = \mathrm{ZHD} + \mathrm{ZWD} \\ &\mathrm{ZHD} = 0.0022767 \frac{\mathrm{P}}{\mathrm{F}(\varphi, \mathrm{H})} \\ &\mathrm{ZWD} = 0.0022767 \frac{\mathrm{e}}{\mathrm{F}(\varphi, \mathrm{H})}\left(\frac{1225}{\mathrm{t} + 273.15} + 0.05\right) \\ &\mathrm{F}(\varphi, \mathrm{H}) = 1 - 0.00266 \cdot \cos 2\varphi - 0.00000028 \cdot \mathrm{H} \\ &\mathrm{e} = \mathrm{RH} \cdot 6.11 \cdot 10^{\frac{7.5\mathrm{t}}{\mathrm{t}+237.3}} \end{aligned} \tag{1}$$

where ZTD, ZHD and ZWD are zenith tropospheric delay, zenith hydrostatic (dry) delay and zenith wet delay, in m, respectively; φ is the ellipsoidal latitude of the station, H is station height in m; P is surface pressure in mbar, t is temperature in degree Celsius and RH is relative humidity in %, all at station height level.

If the meteorological parameters are given in reference to mean sea level (MSL), we need to convert them to the station level (SL), using the following relationships (Klein Baltink et al., 1999):

$$\begin{aligned} &\mathrm{P}_{\mathrm{SL}} = \mathrm{P}_{\mathrm{MSL}}\left(1 - 2.26 \times 10^{-5} \cdot \mathrm{H}\right)^{5.225} \\ &\mathrm{T}_{\mathrm{SL}} = \mathrm{T}_{\mathrm{MSL}} - 0.0065 \cdot \mathrm{H} \\ &\mathrm{RH}_{\mathrm{SL}} = \mathrm{RH}_{\mathrm{MSL}} \cdot \exp\left(-6.396 \times 10^{-4} \cdot \mathrm{H}\right) \end{aligned} \tag{2}$$

Since no real-time meteorological data were applied to estimate ZTD using the SAAS model in our research, the standard values of surface pressure of 1013.25 mbar, temperature of 18°C, and relative humidity of 50%, at mean sea level (MSL), were used to calculate ZTD for all stations and all epochs. To consider the effect of height, we applied Equation (2) to convert the standard meteo-rological data at MSL to station level for every station.

3.2 Ordinary Kriging Interpolation

Ordinary Kriging (OK) is an interpolation procedure used in geostatistics, using known values in the neighbourhood and a variogram to determine the unknown values of the location being estimated. The variogram is based on spheroidal distance between points. A maximum range value a can be used to limit the distance that vanishes the covariance. Three kinds of variograms are suggested by Wackernagel (1998) to calculate the weight:

Linear Variogram: $$\gamma_{ij} = \begin{cases} \dfrac{d_{ij}}{a}, & \text{for } a \le d_{ij} \\ 1, & \text{for } a > d_{ij} \end{cases}$$

Exponential Variogram: $$\gamma_{ij} = 1 - \exp\left(-d_{ij}/a\right)$$

Spherical Variogram: $$\gamma_{ij} = \begin{cases} \dfrac{3}{2}\dfrac{d_{ij}}{a} - \dfrac{1}{2}\dfrac{d_{ij}^3}{a^3}, & \text{for } a \le d_{ij} \\ 1, & \text{for } a > d_{ij} \end{cases}$$

After the variogram is calculated, the weights of the sample points can be obtained as follows:

$$\boldsymbol{\omega} = \mathbf{A}^{-1}\mathbf{B}\,;\ \mathbf{A} = \begin{bmatrix} \gamma_{11} & \gamma_{21} & \cdots & \gamma_{n1} & 1 \\ \gamma_{12} & \gamma_{22} & \cdots & \gamma_{n2} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{1n} & \gamma_{2n} & \cdots & \gamma_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix};\ \mathbf{B} = \begin{bmatrix} \gamma_{1i} \\ \gamma_{2i} \\ \vdots \\ \gamma_{ni} \\ 1 \end{bmatrix}$$

With the weights determined, the value at point i can be calculated from the known values **K** at the sample points $V_i = \boldsymbol{\omega}\mathbf{K}$.

3.3 Tropospheric Grid Model User Algorithm

To obtain the residual ZTD corrections from the correction grid at a user location, a user's algorithm similar to the WAAS ionospheric grid model (Chao, 1997) was adopted. Figure 3 gives the user interpolation algorithm description. At each user, the residual tropospheric correction (T_u) is interpolated from the four grid points, so-called Tropospheric Grid Points (TGPs), surrounding that user:

$$T_u = \sum_{i=1}^{4} \omega_i \cdot T_i \tag{3}$$

The weights (ω_i) in the interpolation are functions of TGP location and are calculated as recommended by WAAS MOPS Appendix A (RTCA SC-159, 1999). The general weighting funtion is:

$$\omega(x, y) = x^2 y^2 (9 - 6x - 6y + 4xy),$$

$$\text{for } 0 \le x \le 1,\ 0 \le y \le 1 \tag{4}$$

The x and y parameters are calculated from

$$x = \frac{\Delta\lambda_{IPP}}{longitude\ \ grid\ \ \operatorname{int} erval}$$
$$y = \frac{\Delta\varphi_{IPP}}{latitude\ \ grid\ \ \operatorname{int} erval} \tag{5}$$

So, the weight of each corner T_u is:

$$\omega_1(x,y) = \omega(x,y) \qquad \omega_2(x,y) = \omega(1-x,y)$$
$$\omega_3(x,y) = \omega(1-x,1-y) \qquad \omega_4(x,y) = \omega(x,1-y)$$

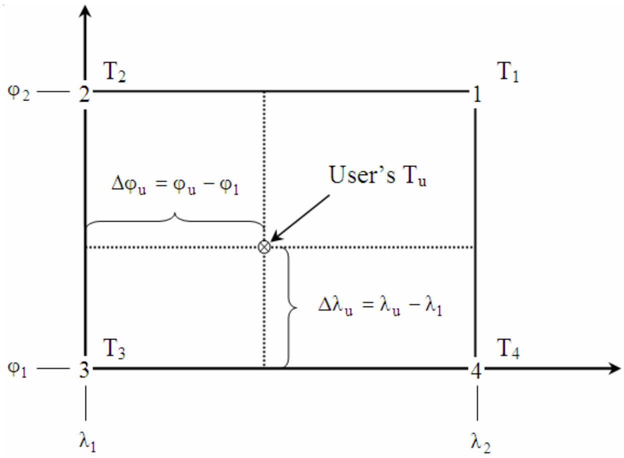


Fig. 3 User interpolation algorithm definition (Chao, 1997)

4 Results and Analysis

First of all, the error of the SAAS model itself against the GPS-derived ZTDs was examined to provide a criterion of the improvement after applying the interpolated residual corrections.

As mentioned before, for each station, identical meteorological data were used and converted from mean sea level to the station level to calculate ZTDs using the SAAS model. Therefore, the computed values for each station remain constant (straight line) during the whole week. Figure 4 illustrates the residuals of SAAS ZTD values with respect to GPS ZTD estimates for the station BAIR with the smallest difference, while Figure 5 illustrates the residuals for the station MTBU, having the largest errors. These residuals are considered as "true error" of the SAAS model at these stations with the given standard meteorological parameters.

After subtracting the SAAS modelled ZTD values from the GPS ZTD estimates, the residual ZTDs can be obtained and then interpolated. Two strategies are employed to compare and access the performance of OK interpolation. The first one is the identical strategy applied in our previous research, in which the residual ZTD of certain station at certain epoch is directly interpolated by OK method using the residuals of all other stations at the same epoch (as the block diagram enclosed by dash-line in Figure 6). The interpolated residual ZTDs obtained by this direct interpolation strategy are so-called "direct residual corrections".

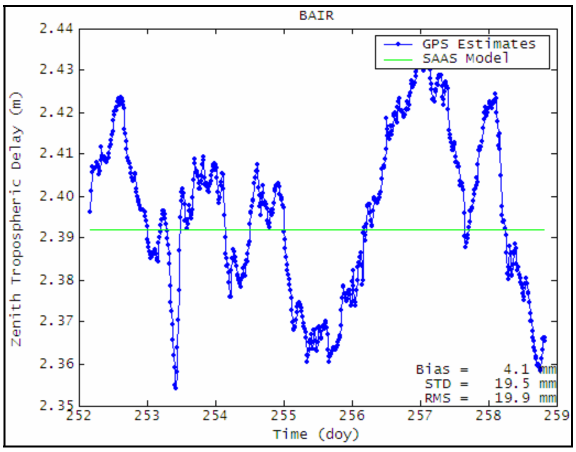


Fig. 4 SAAS ZTD values plotted against
GPS ZTD estimates for the station BAIR

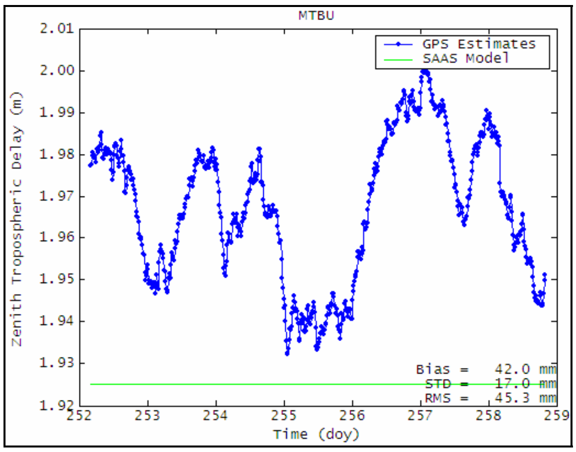


Fig. 5 SAAS ZTD values plotted against
GPS ZTD estimates for the station MTBU

This direct strategy is suitable for error analysis, but as discussed, users will use a simple interpolation model to compute the correction from four grid points, instead of all the points of the network. To reflect this situation, the grid strategy is used. The grid strategy generates the corrections for each tested site using residual ZTDs of the network stations excluding the testing station through OK interpolation, and then interpolates the residual ZTD of the excluded station from the grid using the user interpolation algorithm, as depicted in the block diagram enclosed by solid-line in Figure 6. The interpolated residual ZTDs obtained by this strategy may be called "grid residual ZTD corrections". Within the geographic span of GPSnet, a 1°×1° correction grid is created, as the green dash-line squares shown in Figure 1.

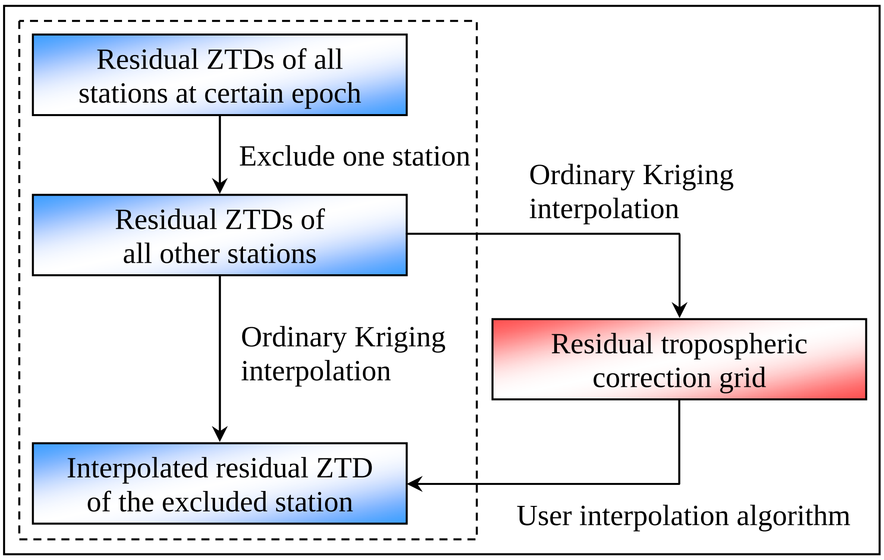


Fig. 6 Block diagram of residual zenith tropospheric delay interpolation strategies. The part enclosed by dash-line is the strategy of generating direct residual corrections, while the whole part enclosed by solid-line is the strategy of generating grid residual corrections.

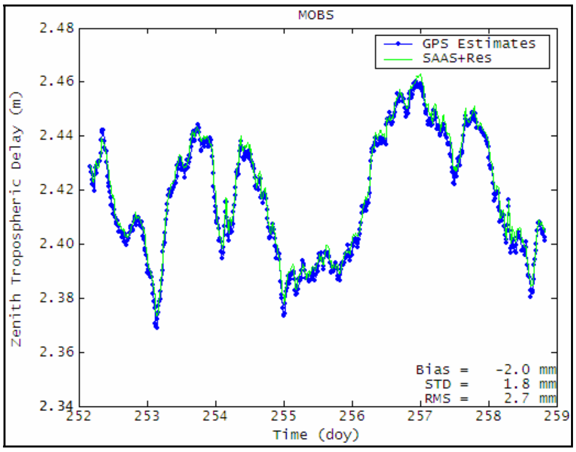


Fig.7 Comparison of GPS ZTD estimates and SAAS ZTD values with direct residual corrections for the station MOBS

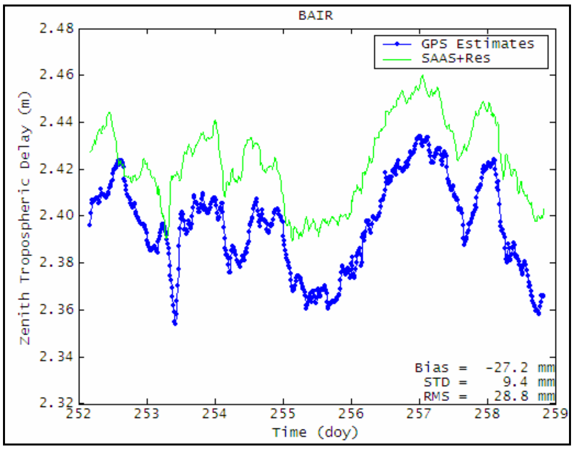


Fig.8 Comparison of GPS ZTD estimates and SAAS ZTD values with direct residual corrections for the station BAIR

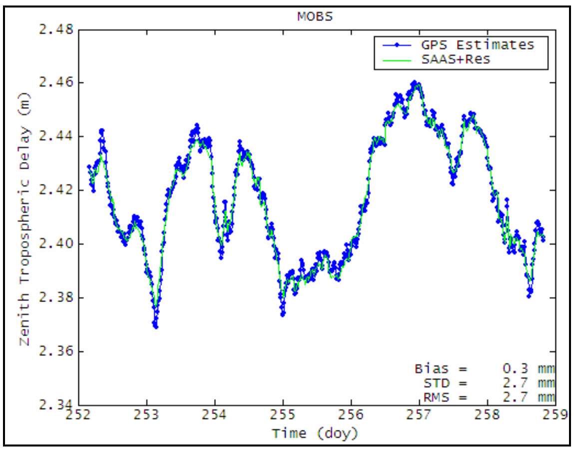


Fig.9 Comparison of GPS ZTD estimates and SAAS ZTD values with grid residual corrections for the station MOBS

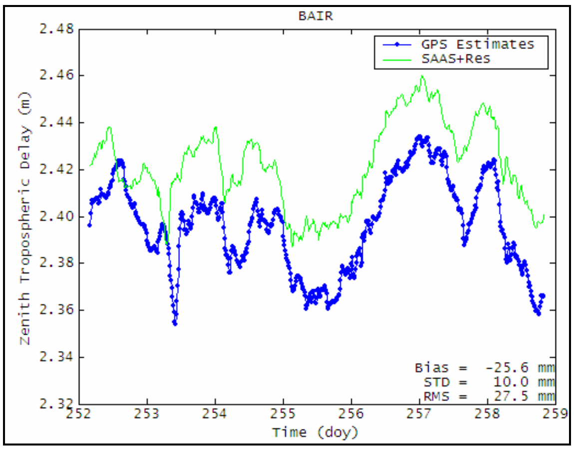


Fig.10 Comparison of GPS ZTD estimates and SAAS ZTD values with grid residual corrections for the station BAIR

Figures 7 and 8 plot the GPS-derived ZTDs against SAAS modelled ZTDs corrected with the direct residual corrections (SAAS+Res) for two different stations, MOBS and BAIR, where the best and the worst consistencies were found, respectively. Figures 9 and 10 present the similar results for the SAAS model with the grid residual corrections (SAAS+Grid Res) for the same stations, MOBS and BAIR, respectively.

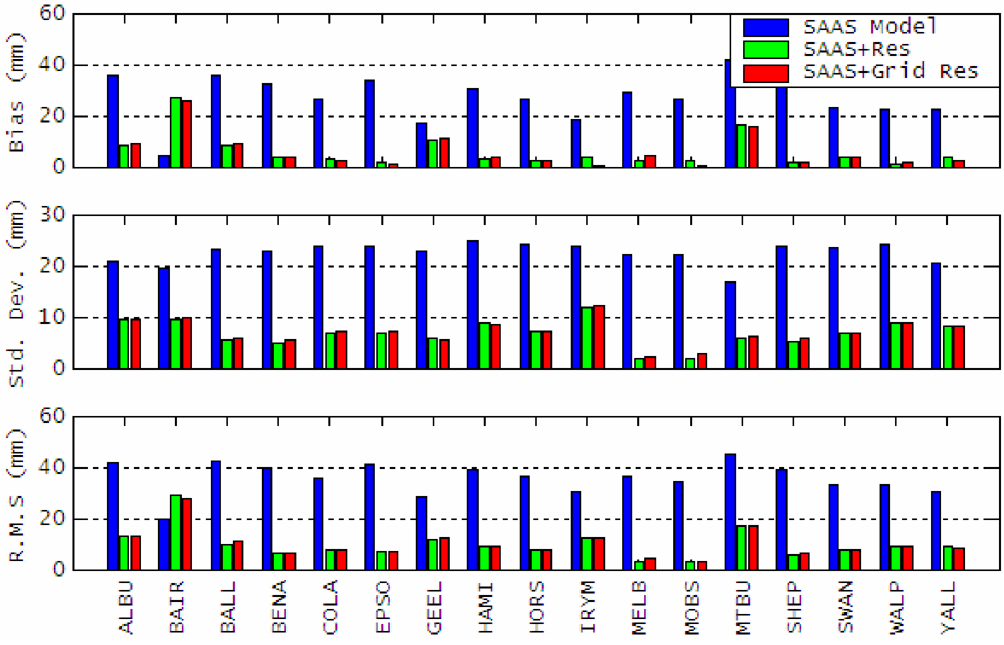


Fig. 11 Statistical comparisons of modeled ZTDs before corrections (SAAS model), after direct corrections (SAAS+Res), and after grid corrections (SAAS+Grid Res) with respect to GPS derived estimates for all stations. The upper figure shows the ***absolute values*** of biases, while the middle and bottom figures plot the standard deviation (STD) and root-mean-square (RMS) values, respectively.

Tab. 1 Overall statistics of SAAS ZTDs with respect to GPS estimates before corrections, after direct residual corrections, and after grid residual corrections (units in mm)

	SAAS	SAAS+Res	SAAS+Grid Res
Bias	26.8	-0.5	0.0
STD	24.2	11.4	11.4
RMS	36.1	11.4	11.4

Figures 11 compares the statistics for all 17 stations, showing the significant improvements by applying direct and grid residual corrections to the SAAS model values, respectively. Table 1 compares the overall bias, STD and RMS values before and after applying both interpolated residual corrections, demonstrating significant overall improvements after corrections. The biases are improved from 27 mm to zero, while the STDs are improved from 24 mm to 11 mm for both kinds of residual corrections. As it can be seen from the figures and the table, both interpolation strategies produce similar results of the interpolated corrections. Moreover, the overall statistic results of the grid residual corrections are even slightly better than the direct ones, which means the proposed tropospheric grid model and the user interpolation algorithm can be used to generate residual tropospheric corrections for near-real time or real-time applications without accuracy degradation from grid interpolation.

It is also interesting to notice that at station BAIR, the bias value is the smallest among all the stations before interpolation but reaches the largest after interpolation, while the STD value has still been improved after interpolation. The reason of causing the large bias at station BAIR is still under closer investigation.

7 Conclusions

Considering spatial and temporal variability of the meteorological parameters, tropospheric delays in GPS differential positioning are usually corrected locally at the user end with an empirical model and a mapping function. In this paper, we have evaluated an approach to providing additional tropospheric delay corrections for improved differential positioning.

One week GPS measurements of the GPSnet were processed and precise ZTD estimates were estimated every 15 minutes. After subtracting SAAS modeled values from GPS estimates, the residual ZTDs were obtained and then interpolated using Ordinary Kriging method by two approaches. One is “direct interpolation”, and the other is “grid interpolation”. However, slight differences were found between the results obtained from two approaches. After correcting with interpolated residual ZTDs, the bias caused by SAAS model is significantly improved from 27 mm to nearly zero, while the STD is improved from 24 mm to 11 mm for the tested data sets. Therefore, we can generally conclude that the RMS ZTD accuracy of about 1 cm is achievable over the Land Victoria GPS network coverage, using the proposed grid tropospheric correction strategy, which can support centimetre level positioning in the regional.

Finally, we summarise the interpolation procedures as follows: 1) reference stations and users use the same tropospheric model with the standard meteorological parameters converted from the mean sea level to the station level, to compute nominated troposphere delays in zenith directions; 2) the reference stations can use OK method to interpolate residual ZTDs and generate a tropospheric grid model, then broadcast it to the users within the service coverage area; 3) A user interpolates its residual ZTD correction from the grid model using the user interpolation algorithm. In all the process, no spatially and temporally variation of surface meteorological data needs to be involved.

Acknowledgements

This work was carried out in the Cooperative Research Centre for Satellite Systems (CRCSS) with financial support from the Commonwealth of Australia through the CRC program and the Australian Research Council linkage project.

Computational resources and services used in this work were provided by High-Performance Computing (HPC) and Research Support Unit, QUT, Brisbane, Australia.

The authors also gratefully acknowledge Mr. Peter Oates, the Victoria GPSnet Operations Manager, for kindly providing the raw GPS observation data and the antenna and receiver specifications of GPSnet stations. They are grateful to Dr. Peng Fang for fruitful discussions about GAMIT data processing strategies.

References

Chao Y. C. (1997): ***Real Time Implementation of the Wide Area Augmentation System for the Global Positioning System with an Emphasis on Ionospheric Modeling.*** Ph.D. dissertation, Stanford University, Stanford, CA.

Collins J. P.; Langley R. B. (1997): ***A Tropospheric Delay Model for the User of the Wide Area Augmentation System.*** Final contract report prepared for Nav Canada, Department of Geodesy and Geomatics Engineering Technical Report No.187, University of New Brunswick, Fredericton, N.B., Canada.

Dodson A. H.; Chen W.; Baker H. C.; Penna N. T.; Roberts G. W.; Jeans R. J.; Westbrook J. (1999): ***Assessment of EGNOS Tropospheric Correction Model.*** Proceedings of ION GPS-99, Nashville, TN, September 14-17, 1401-1407.

Fang P.; Bock Y. (1998): ***Sliding Window Procedure for Super Near Real-time Continuous GPS Water Vapor Estimation Using Predicted Orbits.*** Supp. To Annals Geophysicae, 16, C379.

Fang P.; Bock Y.; Gutman S. (2003): ***Evaluation of real-time ground-based GPS meteorology.*** Proceedings of EGS-AGU-EUG Joint Assembly, Nice.

King R. W.; Bock Y. (2000): ***Documentation for the GAMIT GPS analysis software Version 10.03.*** Massachusetts Institute of Technology.

Klein Baltink H.; Derks H. J. P.; van Lammeren A. C. A. P.; Ambrosius B. A. C.; van der Hoeven A. G. A.; van der Marel H.; Kleijer F.; Kosters A. J. M. (1999): ***Chapter 2: Water Vapour from GPS Tropospheric Delay Estimates. in***, GPS Water Vapour Meteorology, Beleids Commissie Remote Sensing (BCRS), 3-11.

RTCA SC-159 (1999): ***Minimum Operational Performance Standards for Global Positioning System/Wide Area Augmentation System Airborne Equipment.*** RTCA/DO-229B, RTCA Inc., Washington D. C.

Saastamoinen J. (1973): ***Contributions to the Theory of Atmospheric Refraction, Part II Refraction Corrections in Satellite Geodesy.*** Bulletin Geodesique, 10713-34.

Wackernagel H. (1998): ***Multivariate Geostatistics: An Introduction with Applications.*** Springer-Verlag, 291 pp.

Zheng Y. (2004): ***Interpolating Residual Zenith Tropospheric Delays for Improved Wide Area Differential GPS Positioning.*** Proceedings of ION GNSS 2004, Long Beach, LA, September 21-24.

Journal of Global Positioning Systems (2005)
Vol. 4, No. 1-2:291-295

Journal of Global Positioning Systems
2005
© CPGPS

Triple-Frequency Method for High-Order Ionospheric Refractive Error Modelling in GPS Modernization

Zemin Wang[1], Yue Wu[2], Kefei Zhang[3], Yang Meng[4]

[1]School of Geodesy and Geomatics (SGG), Wuhan University, China
e-mail: zmwang_whu@163.com

[2]School of Geodesy and Geomatics (SGG), Wuhan University, China
e-mail: wuyuewangyan@163.com

[3]School of Mathematical and Geospatial Sciences, RMIT University, Australia
e-mail: kefei.zhang@rmit.edu.au

[4]School of Geodesy and Geomatics (SGG), Wuhan University, China
e-mail: mengyang19@163.com

Received: 6 December 2004 / Accepted: 12 July 2005

Abstract. New opportunities for the refinement of ionospheric modelling and reduction of the ionospheric error in GPS measurements arise since a third-frequency will be introduced for the modernised GPS system. This paper investigates theoretical models of the ionospheric refractive error. A triple-frequency method of correcting the 1st and 2nd order ionospheric refraction is presented and a triple-frequency ionosphere-free combination method is proposed for GPS modernization. These new methods can be equally applied to the European GALILEO system. In addition, typical effects of the 2nd and 3rd order ionospheric effects are also investigated, and a correct formula for the 3rd order ionospheric error is derived in a simple way for easy implementation.

It is anticipated that the proposed refraction correction methods will play an important role in both the modernized GPS and GALILEO systems. Results show that the proposed triple-frequency methods can correct the ionospheric refraction effects to a millimetre level. Since the models are given in simple forms, these corrections can be easily implemented in many real-time applications and the triple-frequency methods for correcting high-order ionospheric can significantly eliminate the error remained in the current ionospheric models. The method developed will potentially contribute to a better long-range baseline ambiguity resolution and an accuracy improvement in precise point positioning.

Key words: GPS Modernization, Ionospheric Refractive Error, Triple-Frequency Methods, GALILEO.

1 Introduction

It is well known that ionospheric refractive error is one of the main factors that restrict the GPS positioning accuracy. The equivalent distance error of the ionospheric effect is typically about 15 meters in daytime, 3 meters in the evening, and a maximum of 50 meters in the zenith and 150 meters in the horizon (Zhou, 1995). Therefore, the ionospheric effect must be corrected for. Existing routine procedures to tackle ionospheric effect are to use dual-frequency observations. Most of the error can be eliminated due to the frequency-dependent nature of the ionospheric effects. This is why both GPS and GLONASS systems use two frequencies to transmit signals. However, the dual-frequency method can only account for the linear component of the ionospheric effect (i.e. the 1st order refractive error), which is sufficient for most navigation applications. There are some applications that require accounting for higher order ionospheric refractive error. The imminent GPS modernization (McDonald, 2002) and GALILEO (European Commission, 2003) deployment provides the potential to further refine the ionospheric models due to the introduction of a third-frequency.

During the periods of high Total Electron Content (TEC), higher-order ionospheric effects can reach several tens of centimetres in range error, which adversely affect the accuracy of GPS observables. Brunner and Gu (1991) use the full form of the refractive index given in Eq. (3) to calculate the residual range error from the first–order form of refractive index. Their model also includes the effects of the Earth's geomagnetic field and ray bending at both GPS frequencies, L_1 and L_2. However, in order to achieve millimetre-level of accuracy for ionospheric error correction, both the actual maximum electron density, N_m, and the average value of the longitudinal component of the Earth's magnetic field along the ray path are required. For practical usages, however, it is difficult to either access or estimate these parameters.

Bassiri and Hajj (1993) carry out a similar research on high-order GPS ionospheric range errors. The electron density profile shape is required in the model and higher-order ionospheric corrections are not as good as those claimed by Brunner and Gu (1991). However, their models are much easier to be implemented since the actual ionospheric data is required.

In this paper, a triple-frequency method of correcting both the 1st and the 2nd order ionospheric refraction is presented. In addition, a triple-frequency ionosphere-free combination method is also developed for the imminent GPS modernization and GALILEO systems. The 3rd order ionospheric refraction correction is also investigated in a simplified way. It is expected that a millimetre level of ionospheric correction accuracy can be achieved using the methods developed.

2 Ionosphere delay error

The ionosphere is a dispersive media that is ionized by the ultraviolet radiation from the sun. The TEC value of the ionosphere varies mainly due to day-to-night variations. However, it also depends on a user's location (i.e. the geomagnetic latitude), time of year, and sunspot cycle. The ionosphere affects the GPS signals propagating through it. The propagation of GPS signals has the characteristics that there is a phase advance of the carrier and a group delay of the code. The phase velocity v_p and the velocity of the group of frequencies v_g have the following relationship:

$$v_g = v_p - \lambda \frac{dv_p}{d\lambda}$$

Incorporating $d\lambda / \lambda = - df / f$ in to the above equation, we then have

$$v_g = v_p + f \frac{dv_p}{df} \tag{1}$$

The refractive index of GPS signals is defined as:

$$n = \frac{c}{v} = \frac{\lambda_{vac}}{\lambda}$$

Where c and λ_{vac} are the velocity and wavelength of GPS signals in vacuum respectively. v and λ are the velocity and wavelength of GPS signals in the ionosphere respectively. Substitute the above equation into equation (1), then we obtain

$$n_g = n_p / \left[1 - (\frac{dn_p}{df}) / (\frac{n_p}{f}) \right]$$

Develop the above equation into a series form of $(1 - x)^{-1} = 1 + x - x^2 + \cdots$, and take the first two terms only. The relation of n_g and n_p can then be expressed as

$$n_g = n_p + f \frac{dn_p}{df} \tag{2}$$

According to Brunner and Gu (1991), the diffused refractive index can be expressed as

$$n_p = 1 - C_X N_e f^{-2} / 2 \pm C_X C_Y N_e (H_0 \cos\theta) f^{-3} / 2 - C_X{}^2 N_e^2 f^{-4} / 8 \tag{3}$$

Where

$$\begin{cases} C_X = \dfrac{e^2}{4\pi^2 \varepsilon_0 m} \\ C_Y = \dfrac{\mu_0 e}{2\pi m} \end{cases}$$

and e represents the power of an electron, m is the mass of an electron, N_e is the electron density in the atmosphere, H_0 is the quantity of the geomagnetic field, Θ is the angle between H_0 and the direction of signal transmitting, f is the frequency, ε_0 and μ_0 are physical constants. Equation (3) defines the refractive index up to an accuracy of 10^{-9} (Brunner and Gu, 1991), so it ensures that the optical path length is of millimetre accuracy.

Equation (3) can then be expressed as

$$n_p = 1 + a_1 / f^2 + a_2 / f^3 + a_3 / f^4 \tag{4}$$

Where a_1, a_2 and a_3 are the simplified coefficients. Substituting equation (4) into equation (2), we have

$$n_g = 1 - a_1 / f^2 - 2a_2 / f^3 - 3a_3 / f^4 \tag{5}$$

When GPS signals penetrate through the ionosphere, the distance and phase errors of the transmission paths due to the variation of the refractive index:

$$\begin{cases} \delta\rho = \int_s (n-1)ds \\ \delta\varphi_p = \frac{f}{c}\int_s (n_p-1)ds \end{cases} \tag{6}$$

Substituting equations (4) and (5) into equation (6) gives

$$\begin{aligned} \delta\rho_p &= \int_s (n_p-1)ds \\ &= \int_s (a_1/f^2 + a_2/f^3 + a_3/f^4)ds \\ &= \int_s (a_1/f^2)ds + \int_s (a_2/f^3)ds + \int_s (a_3/f^4)ds \end{aligned} \tag{7}$$

$$\begin{aligned} \delta\rho_g &= \int_s (n_g-1)ds \\ &= \int_s (-a_1/f^2 - 2a_2/f^3 - 3a_3/f^4)ds \\ &= -\int_s (a_1/f^2)ds - 2\int_s (a_2/f^3)ds - 3\int_s (a_3/f^4)ds \end{aligned} \tag{8}$$

Again, ionospheric refractive error is generally about 15 meters in the daytime, and about 3 m in the evening, and maximum 50 m in the zenith, 150 m in the horizon (Zhou, 1995). If only the 1st-order ionospheric refractions are corrected, the remaining error is usually about 1.5 centimetres, and its maximum is about 5~15 centimetres. If better than 1 centimetre accuracy is required, we must correct the 2nd order ionospheric refraction in theory. Moreover, if you want to achieve millimetre-level of precision, the 3rd order ionospheric refraction must be corrected for. This also suggests that the remaining error after applying the 1st-order ionospheric refraction correction is about centimetre-level for phase measurements and about meter-level for code pseudoranges. This may be much higher when there are severe sunspot activities. It is concluded that for precise positioning applications the 2nd order and the 3rd order ionospheric refractions must be corrected for.

3 Triple-frequency method for 2nd order ionospheric refraction correction

It is well known that the 1st order ionospheric refractions can be eliminated by current dual-frequency method using its frequency dependent nature. With GPS modernization, it becomes possible to correct 2nd order ionospheric refraction using three frequency measurements. Below we present the procedure of resolving the 2nd order ionosphere effects.

From equation (7), we have (only the 1st- and 2nd-order ionospheric refractions are taken into account here)

$$\delta\rho_p = \int_s (a_1/f^2 + a_2/f^3)ds = A_1/f^2 + A_2/f^3$$

Where,

$$A_1 = \int_s a_1 ds\,,\quad A_2 = \int_s a_2 ds$$

To distinguish the effects from different frequencies, the above equation can be re-written as

$$\delta\rho_p(f_i) = A_1/f_i^2 + A_2/f_i^3 \qquad i = 1,2,3 \tag{9}$$

The 1st order ionospheric refraction removed combinations can be obtained as

$$\delta\rho_p(f_1)\bullet f_1^2 - \delta\rho_p(f_2)\bullet f_2^2 = A_2(1/f_1 - 1/f_2)$$
$$\delta\rho_p(f_1)\bullet f_1^2 - \delta\rho_p(f_3)\bullet f_3^2 = A_2(1/f_1 - 1/f_3)$$

Then the 2nd-order ionospheric refraction removed combinations is

$$\frac{(\delta\rho_p(f_1)\bullet f_1^2 - \delta\rho_p(f_2)\bullet f_2^2)(f_1 f_2)}{f_2 - f_1} - \frac{(\delta\rho_p(f_1)\bullet f_1^2 - \delta\rho_p(f_3)\bullet f_3^2)(f_1 f_3)}{f_3 - f_1} = 0$$

This can be simplified as

$$\begin{aligned} & f_1^3(f_3 - f_2)\bullet\delta\rho_p(f_1) + f_2^3(f_1 - f_3)\bullet\delta\rho_p(f_2) \\ &\qquad + f_3^3(f_2 - f_1)\bullet\delta\rho_p(f_3) \\ &= B_1\bullet\delta\rho_p(f_1) + B_2\bullet\delta\rho_p(f_2) + B_3\bullet\delta\rho_p(f_3) \\ &= 0 \end{aligned}$$

Where,

$$\begin{cases} B_1 = f_1^3(f_3 - f_2) \\ B_2 = f_2^3(f_1 - f_3) \\ B_3 = f_3^3(f_2 - f_1) \end{cases}$$

Given $m_i = f_1 \cdot B_i / (B_1+B_2+B_3)$, i=1,2,3, the above equation becomes:

$$m_1\bullet\delta\rho_p(f_1) + m_2\bullet\delta\rho_p(f_2) + m_3\bullet\delta\rho_p(f_3) = 0$$

This means that the 2nd order ionospheric refractions have been corrected and this procedure is also suitable for code measurements. Considering the following relationship

$$\delta\varphi_p = \frac{f}{c}\delta\rho_p\,,$$

we obtain:

$$\delta\varphi_p(f_1)\bullet m_1/f_1+\delta\varphi_p(f_2)\bullet m_2/f_2+\delta\varphi_p(f_3)\bullet m_3/f_3$$
$$=\delta\varphi_p(f_1)\bullet m+\delta\varphi_p(f_2)\bullet n+\delta\varphi_p(f_3)\bullet k$$
$$=0$$

where $m = m_1/f_1$, $n = m_2/f_2$, $k = m_3/f_3$.

Then the 1st-order and 2nd order ionospheric refraction removed combination of three-frequency can be expressed as:

$$\varphi = m\varphi_1 + n\varphi_2 + k\varphi_3 \tag{10}$$

The corresponding phase combination is

$f = m f_1 + nf_2 + kf_3$

$\lambda = c/f$

The relations between the ionospheric correction and phase pseudorange observation are given below. The phase pseudorange observation equations are

$$\rho_i = \rho_0 + \delta\rho_p(f_i) = \rho_0 + A_1/f_i^2 + A_2/f_i^3 \quad (i = 1, 2, 3) \tag{11}$$

From the above equations we obtain

$$A_1 = \frac{\rho_{12} f_1^3(f_3^3 - f_2^3) - \rho_{23} f_3^3(f_2^3 - f_1^3)}{f_1^3(f_2 - f_3) + f_2^3(f_3 - f_1) + f_3^3(f_1 - f_2)} \tag{12}$$

$$A_2 = -\frac{\rho_{12} f_1^3 f_2 f_3(f_3^2 - f_2^2) - \rho_{23} f_1 f_2 f_3^3(f_2^2 - f_1^2)}{f_1^3(f_2 - f_3) + f_2^3(f_3 - f_1) + f_3^3(f_1 - f_2)} \tag{13}$$

In GPS modernization, the three GPS signal frequencies are L1 ： 1575.42MHz （154 × 10.23MHz） ； L2 ： 1227.60MHz（120×10.23MHz）and L5：1176.45MHz （115 × 10.23MHz） respectively. Substituting these values into equations (12), (13) and (9) gives

$$\begin{cases}\delta\rho_p(f_1) = -6.080583\rho_{12} + 20.049766\rho_{23} \\ \delta\rho_p(f_2) = -7.080583\rho_{12} + 20.049766\rho_{23} \\ \delta\rho_p(f_3) = -7.080583\rho_{12} + 19.049766\rho_{23}\end{cases} \tag{14}$$

where, $\rho_{12} = \rho_1 - \rho_2$, $\rho_{23} = \rho_2 - \rho_3$. These are the formulas of calculating the value of ionospheric refraction by three-frequency phase measurements. GPS code pseudorange measurements can be obtained in a similar way.

The 1st order ionospheric refraction takes up 99% of the total effects. Table 1 lists the remaining ionospheric refraction with both the 1st order and 2nd order terms corrected. It is demonstrated that correction of the 2nd order term is necessary for a precision of better than one centimetre-level. Correcting 3rd-order term is also necessary for the precision of more than one millimetre level. Here we assume that the sum of 4 terms at the right hand side of equation (3) is the total of the ionospheric refraction.

Table 1. Maximum vertical ionospheric range error [unit=m]

CASE 1 : TEC=4.55e18m^{-2} N_m =20.0e12m^{-3}			
Frequency	1st-order effect (1/f^2)	2nd-order effect (1/f^3)	3rd-order effect (1/f^4)
L1	73.8428	0.0818	0.0079
L2	121.6150	0.1729	0.0215
L5	132.4201	0.1964	0.0254
L1/L2	0	0.0590	0.0130
L1/L2/L5	0	0	0.0054
CASE 2 : TEC=1.38e18 m^{-2} N_m =6.0e12 m^{-3}			
Frequency	1st-order effect (1/f^2)	2nd-order effect (1/f^3)	3rd-order effect (1/f^4)
L1	22.3963	0.0248	0.0007
L2	36.8854	0.0524	0.0020
L5	40.1626	0.0596	0.0023
L1/L2	0	0.0179	0.0012
L1/L2/L5	0	0	0.0005

Note that CASE 1 corresponds to an absolute maximum solar cycle condition that rarely happens, while CASE 2 is typical for high N_m values that are frequently observed (Klobuchar, 1983).

4 Third order ionospheric refraction correction

From Table 1, it can be seen that the 3rd order effect is still as large as 25mm in carrier L_5. The ionospheric free linear combination of the phase observation on L_1, L_2, and L_3 still remains 5mm in CASE 1. Thus, we must correct the 3rd order effects.

In order to calculate the 3rd order term, an integration of N_e^2 along the signal path is required. The integration is very difficult and involves a complicated process. Hartmann and Leitinger (1984) have suggested a shape parameter η through the following relation:

$$\eta \equiv \int_s N_e^2 ds \Big/ N_m \int_s N_e ds$$

Where N_m is the maximum value of the electron density N_e. The value of the shape parameter only slightly varies with the elevation angle and the maximum electron density. This is proven by Hartmann and Leitinger (1984) and Brunner (1991). According to their research, η can be assigned an approximate value of 0.66 and this represents η for any profile of the electron density distribution in the ionospheric with sufficient accuracy. The 3rd order effect of ionosphere can be expressed as:

$$\delta\rho_p^{3rd-order} = -C_X{}^2\eta N_m/(8f^4)\cdot\int_s N_e ds \tag{15}$$

If we select the semi-empirical ionospheric model developed by Anderson et al. (1987) who gives an electron density distribution in the ionosphere. The approximate relation of N_m and TEC can be expressed as

$$\mathrm{TEC}=k\cdot N_m \tag{16}$$

Where, $k\approx 2.27\times 10^5$m.

TEC can be obtained in several ways; one convenient way is to obtain it from the difference between ρ_1 and ρ_2. Neglecting high-order effects, we have

$$\delta\rho_p(f_1) = -1.54573\rho_{12} = -C_X\cdot TEC/(2f_1{}^2)$$

We then have

$$\mathrm{TEC}=9.51768\times 10^{16}\rho_{12} \tag{17}$$

$$\delta\rho_p^{3-order} = -C_X{}^2\eta/(8kf^4)\cdot(TEC)^2$$

$$= -2.13964\times 10^{31}\rho_{12}{}^2/f^4 \tag{18}$$

Using the difference between ρ_1 and ρ_2, the 3rd order ionospheric effect can be easily calculated using equation (18). Although some approximate parameters are adopted in equation (18), it is still of high accuracy. Its relative error is less than ±5%. Considering the fact that the 3rd order effects are generally less than 10mm, the absolute accuracy of equation (18) is better than ±1mm.

5 Conclusions

The availability of the third frequency from the modernized GPS and GALILEO systems provides an opportunity to eliminate the ionospheric propagation effects more efficiently. In this paper, a three frequency ionosphere free combination is given, which has effectively eliminated the 1st order and the 2nd order ionospheric effects. Furthermore, formulas of calculating the value of ionospheric refraction by differencing the carrier phase measurements in three frequencies are derived. Finally, an equation calculating the 3rd order ionospheric effects using the difference between ρ_1 and ρ_2 is given. Although some approximate parameters are adopted in the equation, it is still of high accuracy. The absolute accuracy of the 3rd order ionospheric effects is better than ±1mm.

The advantage of the triple frequency methods for correcting high-order ionospheric refraction is apparent to the imminent GPS modernization program. It is concluded that the proposed triple frequency methods can correct the ionospheric refraction at millimetre-level. Since the corrections are given in simple forms, these methods can be easily implemented in many real-time applications to considerably eliminate the error in current ionospheric models. We believe that this will potentially contribute to a better long-range baseline ambiguity resolution and an accuracy improvement in precise point positioning. In addition, the sensitivity of detecting cycle slips by using a refined ionospheric refraction can be improved. These new methods are also applicable to GALILEO system as well.

Acknowledgement

Partial financial support from the Australia Research Council Linkage project (LP0455170) endorsed to a research consortium led by A/Prof Kefei Zhang is highly appreciated.

References

Anderson DN.; Mendillo M.; Herniter B. （1987）: ***A semi-empirical low latitude ionospheric model,*** Radio Science, 15,1009-1016.

Bassiri S.; Hajj G.A. (1993): ***Higher-order ionospheric effect on the Global Positioning System observables and means of modeling them***. Manuscripta Geodaetica, 18:280-289.

European Commission. (2003): ***The Galilei Project, Galileo Design Consolidation***, http://europa.eu.int/comm/dgs/ energy_transport/galileo/doc/galilei_brochure.pdf.

Fritz K.; Brunner.; Min Gu. (1991): ***An improved model for the dual frequency ionospheric correction of GPS observations,*** Manuscripta Geodaetica, 16:205-214.

Hartmann GK.; Leitinger R.（1984）: ***Range errors due to ionospheric and tropospheric effects for signal frequencies above 100 MHz,*** Bull. Geod.,58: 109-136

Klobuchar JA. (1983): ***Ionospheric Effects on Earth-Space Propagation***. AGFL-tr-84-0004.

Klobuchar JA. (1996): ***Ionospheric effect on GPS. Global Positioning System: Theory and Applications.*** Edited by B.W.Parkinson. American Institute of Aeronautics and Astronautics

Langley, RB. (2001): *Satellite Navigation: GPS Modernization and R&D in the Academic Sector.* National Sector Team for Space Annual Meeting, Canadian Space Agency. 3-4 July

McDonald, K. (2002): ***The Modernization of GPS: Plans, new capabilities and the future relationship to Galileo***, Journal of Global Positioning Systems, 1(1), 1-17.

Zhou Zhongmo. (1995): ***Principle and Practice of GPS Satellite Surveying.*** BEIJING: Press of Surveying and Mapping.