

Integrating a Trimble Recon X 400 MHz Intel PXA255 Xscale CPU® Mobile Field Data Collection System Using Differentially Corrected Global Positioning System Technology and a Real-Time Bidirectional Actionable Platform within an ArcGIS Cyberenvironment for Implementing Mosquito Control

Benjamin G. Jacob, Robert J. Novak

Department of Global Health, University of South Florida, Tampa, USA Email: <u>bjacob1@health.usf.edu</u>, <u>rnovak@health.usf.edu</u>

Received 12 July 2014; revised 10 August 2014; accepted 5 September 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY). http://creativecommons.org/licenses/by/4.0/

Abstract

In this research, we determined the feasibility of using a Personal Digital Assistant (PDA) as a mobile field data collection system by monitoring mapping and regressing digitized sub-meter resolution polygons of multiple, malaria, mosquito, *Anopheline arabiensis s.s.*, aquatic, larval, habitat covariates. The system employed QuickBird raster imagery displayed on a Trimble Recon X 400 MHz Intel PXA255 Xscale CPU[®]. The mobile mapping platform was employed to identify specific geographical locations of treated and untreated seasonal *An. arabiensis s.s.* aquatic larval habitats in Karima rice-village complex in the Mwea Rice Scheme, Kenya. As data pertaining to *An. arabiensis s.s.* larval habitats were entered, all treated and untreated rice paddies within a 2 km buffer of the agro-village, riceland-complex, epidemiological, study site were viewed and managed on the PDA.

Keywords

Trimble Recon X, Vector Control Management System (VCMS), *Anopheles arabiensis*, QuickBird Imagery

How to cite this paper: Jacob, B.G. and Novak, R.J. (2014) Integrating a Trimble Recon X 400 MHz Intel PXA255 Xscale CPU[®] Mobile Field Data Collection System Using Differentially Corrected Global Positioning System Technology and a Real-Time Bidirectional Actionable Platform within an ArcGIS Cyberenvironment for Implementing Mosquito Control. *Advances in Remote Sensing*, **3**, 141-196. <u>http://dx.doi.org/10.4236/ars.2014.33012</u>

1. Introduction

Currently, very few mosquito control programs have collected valid outcome metrics that enable public health officials and funding agencies to quantitatively determine the degree and magnitude of their impact on time series, field or remote-sampled aquatic, larval, habitat, endemic, transmission-oriented, explanatory, parameter estimators. A principle reason for this deficiency is the absence of sufficient management systems to capture, manage and analyze large amount of diverse data being generated by various mosquito control activities. Existing management information is neither accurate nor comprehensive, a vital necessity for efficient, effective and economic mosquito management [1]. Geographic location, combined with knowledge of ecological features of multiple seasonal-sampled georeferenced vector mosquito aquatic larval habitat sites may be a potential key element for implementing efficient and cost effective larval control measures. A remote evaluation of mosquito treatment data offers a systematic way of comparing the costs and consequences of interventions in order to improve the allocation of resources by prioritizing georeferenced vector aquatic mosquito habitats based on seasonal larval/pupal productivity [2]. Treatments or habitat perturbations should be based on surveillance of larvae in the most productive areas of an ecosystem [3]. Additionally, a remote surveillance system can be used more widely to improve the uptake of existing effective control measures and ensure the maximum impact from the introduction of new technologies.

Personal Digital Assistant (PDA) technologies offer a unique way to collect and manage multiple, ecological, seasonal-sampled, vector, mosquito-related, endemic, transmission-oriented data variables. Developing and implementing streamlined data collection, aggregation and reporting methodologies using PDA handheld computers equipped with global positioning systems (GPS) can provide geographically detailed, real-time information when conducting surveys in remote villages [4]. While proper identification of mosquito species and knowledge of their bionomics focuses on control efforts, remote inspection of georeferenced larval habitats and weekly trapping for adult mosquitoes ensures knowledge of the mosquito population in a given municipality [3]. Accurate, georeferenced, collection data is crucial for understanding mosquito biogeography, ecology, and the impact of environmental changes, as well as for species distribution modeling, planning mosquito surveys, and for determining disease risk [2]. Field teams require a simple visual mechanism to help designate priority treatment of seasonal-sampled, georeferenced, vector, mosquito, aquatic, larval habitats within a project area, in order to promote on-going tracking in remote communities, which can maximize treatment efficiency. A PDA mobile collection system combined with newer GIS software could help prioritize health interventions based on field or remote specified, vector, mosquito-related, georeferenced, explanatory, endemic, transmission-oriented, covariate coefficients in a more standardized fashion.

The evolution of PDA technology, geographic information systems (GIS), GPS, and remote sensing (RS) data may enable spatiotemporal-sampled field collections and data analysis of ecological georeferenced vector mosquito aquatic larval habitat data in ways that were not possible before. Unfortunately, the rapid development and integration of remote technologies has created many new tools which has also widened the "digital divide", leaving many vector biologists, epidemiologists and other data analysts with little understanding of the technology and potential applications of current geospatial techniques for robust, seasonal, vector, mosquito, aquatic, larval habitat predictive, endemic, transmission-oriented, epidemiological, risk modeling. For example, subresolution RS (e.g., panchromatic QuickBird at 0.61 m pixel spatial resolution) data can be used to provide unique thematic information regarding a variety of time series georeferenced explanatory georeferencable vector mosquito aquatic larval habitat potential explanatory endemic transmission-oriented geo-biophysical characteristics, including surface temperature, imperviousness, water habitat clarity, evapotranspiration, vegetation pigments, biomass, canopy structure, leaf area index (LAI), and soil moisture [2]. Leaf Area Index (LAI) is a dimensionless quantity that characterizes plant canopies which is defined as the one-sided green leaf area per unit ground surface area (LAI = leaf area/ground area, m^2/m^2) in broad leaf canopies [5].

There are only limited contributions in literature of robust, biophysical, geo-spatiotemporal, endemic, transmission-oriented, vector, mosquito-related, predictive, epidemiological, risk models using data collected employing PDAs, however in these contributions biophysical technologies have provided detailed, local terrestrial geographic data [e.g., vegetation land use land cover (LULC), meandering riverine pathways)]. For example, Thorp *et al.* (2012) employed remote sensing technology to provide spatial information on crop growth status of durum wheat (*Triticum durum*) study by inverting a radiative transfer model and then generating multiple time series, eco-physiological, sub-models for estimating the crop's biophysical properties. An Atmospheric radiative transfer model, code or simulator calculates radiative transfer of electromagnetic radiation through a planetary atmosphere [1]. The outcome of the model inversion procedure was influenced by the timing and availability of remote sensing data, the spectral resolution of the data, the types of models implemented, and the choice of parameters that were adjusted in the submodel residually forecasted derivatives. The objective was to investigate seasonal phenological issues by inverting a linked radioactive transfer and then quantitating the eco-physiological model derivatives to estimate LAI, canopy weight, plant nitrogen content, and yield for the durum wheat (Triti*cum durum*). The study was conducted in central Arizona over the winter of 2010-2011. Observations of crop canopy spectral reflectance between 268 and 1095 nm were obtained weekly using a GER 1500 spectroradiometer. Other field measurements were regularly collected to describe plant growth characteristics and plant nitrogen content. Linkages were developed between the DSSAT Cropping System Model (CSM) and the PROSAIL radiative transfer model (CSM-PROSAIL) and between the DSSAT-CSM and an empirical model relating vegetation LULC parameter (CSM-Choudhury). A parameter estimation algorithm (i.e., PEST) was implemented to adjust the leaf area growth parameters of the CSM by minimizing error between measured and simulated seasonal canopy spectral reflectance. A genetic algorithm was implemented to identify the optimum combination of the remote sensing observations and to optimize simulations of multiple vegetation indices through model inversion. For example, the relative root mean squared error (RRMSE) between measured and simulated LAI was 24.1% for the CSM-PROSAIL model, whereas the stand-alone PROSAIL and CSM models simulated LAI with RRMSEs of 40.7% and 27.8%, respectively. Wheat yield was simulated with RRMSEs of 12.8% and 10.0% for the lone CSM model and the CSM-PROSAIL model, respectively. Optimized leaf area growth parameters for CSM-PROSAIL were different among cultivars (p < 0.05), while those for CSM-Choudhury were not. Only two observations, one at mid-vegetative growth and one at maximum vegetative growth, were required to optimize LAI simulations for CSM-PROSAIL, whereas CSM-Choudhury required four observations. Inverting CSM-PROSAIL using hyperspectral data offered several advantages as compared to the CSM-Choudhury inversion using a simple vegetation index, including better estimates of crop biophysical properties, different leaf area growth parameter estimates among cultivars (p < 0.05), and fewer required remote sensing observations for optimum LAI simulation.

Methods for constructing a seasonal explanatory geobiophysical geographic predictive real-time vector-related remotely sensed mosquito models include: Principal Component Analysis, Unsupervised Classification, Map Algebra, Band Ratios, and Regression of reflection values [2]. Imaging georeferenced, seasonal-sampled, vector, mosquito-related, spectrometry data, in conjunction with suitable analysis techniques in a cyber-environment, may provide a basis for quantitatively measuring phenological change in seasonal, terrain-related, georeferenced, vector, mosquito, aquatic, larval habitats which results from changes in primary productivity and canopy vegetation vigor. The phenological changes of seasonal, vector, mosquito, aquatic, larval habitats may be in response to regional and/or global-scale environmental or climatic changes.

Two assumptions must be satisfied if PDA technology, GIS, GPS, and imaging spectrometry are to be useful in remote biophysical analysis of seasonal, vector, mosquito, aquatic, larval habitats. First, there must be a strong correlation between georeferencable, explanatory, operationizable, canopy-shaded, larval, habitat characteristics and the rates at which processes important to the biosphere occur. Secondly, these canopy-shaded characteristics must be successfully measured using high spectral resolution remote-sensing data. These include the measurement of visible and near infra-red (NIR) spectral position for constructing normalized difference vegetation index (NDVI), moisture stress index and shortwave infrared reflectance parameters related to canopy chemistry parameters and terrestrial imaging spectrometry.

The NDVI is a simple graphical indicator that can be used to analyze remote measurements, typically but not necessarily from a space platform, and assess whether the target being observed contains live green vegetation or not [1]. Leaf cells have also evolved to scatter solar radiation in the NIR spectral region (which carries approximately half of the total incoming solar energy), because the energy level per photon in that domain (wavelengths longer than about 700 nanometers) is not sufficient to be useful to synthesize organic molecules. Although electromagnetic waves have wavelengths that range from 1 nanometer (nm) (10^{-9} meter) for x-rays to 1 kilometer (10^3 meters) for radio waves wavelengths of visible light are in the range of 400 to 700 nanometers. Jacob *et al.* [2] employed the equation E = (hc)/L in a MapTP Mobile, web-based mapping applications where E was the energy of photon in a seasonal, vector, malarial, mosquito-related (e.g., *Anopheles arabiensis s.s.*), aquatic, larval, habitat, spectral, endmember, forecasting, predictive, epidemiological, risk model, h was the Planck's

constant, 6.626×10^{-34} J·s, c was the speed of light, 2.998×10^{-8} m/s and L was wavelength for converting nanometers to units of joules (J), whereby, 715 nm = 7.15×10^{-7} m and then E = $(6.626 \times 10^{-34} \text{ J} \cdot \text{s})^{*}(2.998)$ $\times 10^{8}$ m/s)/(7.15 $\times 10^{-7}$ m) = 2.778 $\times 10^{-19}$ J. The authors noted that live green plants associated with the seasonal-sampled Anopheles arabiensis s.s. aquatic larval habitats appeared relatively dark in the photosynthetically active radiation (PAR) and relatively bright in sub-meter resolution (panchromatic QuickBird 0.61 m resolution) NIR region of the electromagnetic spectrum. Photosynthetically active radiation designates the spectral range (wave band) of solar radiation from 400 to 700 nanometers that photosynthetic organisms are able to use in the process of photosynthesis [1]. PAR explanatory covariate coefficients were employed for constructing multiple, biophysical, seasonal, An. arabiensis s.s.-related, endemic, transmission-oriented, predictive, explanatory, eco-epidemiological risk models where the forecasted derivatives were quantified as μ mol photons m⁻²·s⁻¹, a measure of the photosynthetic photon flux (area) density, or PPFD. The derivatives were also expressed as Einstein units (*i.e.*, $\mu E \cdot m^{-2} \cdot s^{-1}$). PAR action spectrum, revealed absorption spectra for chlorophyll-A, chlorophyll-B, and carotenoids related to the sampled aquatic larval habitats. Differential reflection in the red and infrared (IR) bands enabled remotely monitoring density and intensity of green vegetation growth using the spectral reflectivity of solar radiation. By contrast, clouds tended to be rather bright in the red (as well as other visible wavelengths) for some of the imaged, georeferenced, (e.g., Anopheles arabiensis s.s.), aquatic, larval, habitat objects and quite dark in the NIR.

Optimally, a seasonal, vector, mosquito-related, aquatic, larval habitat image should have less than 20% cloud coverage for constructing forecasting, epidemiological, endemic, transmission-oriented, risk-related, predictive models [2]. The pigment in plant leaves, chlorophyll, strongly absorbs visible light (from 0.4 to 0.7 μ m) for use in photosynthesis [1] which has been known to be associated with seasonal-sampled malaria mosquito aquatic larval habitats. For example, Mwangangi *et al.* (2007) conducted investigations using laboratory and field data on experimental riceland plots in the Mwea, Rice Scheme, Kenya. In these trials, chlorophyll derivatives were added to the infested plots which were subsequently ingested by the *An. arabiensis s.s.* larvae and the accumulated photoactive compound (photosensitizer) inside the larvae body induced sunlight exposure which created an oxidation stress that resulted in organism death.

The cell structure of the leaves, on the other hand, strongly reflects near-infrared light (from 0.7 to 1.1 μ m). The more leaves a plant associated to a georeferenced vector mosquito-aquatic larval habitat has, the more these wavelengths of light are affected, respectively. Since early instruments of Earth Observation, such as NASA's ERTS and NOAA's AVHRR, acquired data in visible and NIR, it was natural to exploit the strong differences in plant reflectance to determine their spatial distribution in these satellite images. Dube et al. (2011) studied repellent properties of volatiles emitted from ethnomedicinal plant leaves against malaria and yellow fever vectors in Ethiopia. In their research plant-based mosquito repellents, volatile emanations were investigated from five plant species, Corymbia citriodora, Ocimum suave, Ocimum lamiifolium, Olea europaea and Ostostegia integrifolia, traditionally used in Ethiopia as protection against mosquitoes. The behaviour of two mosquitoes, the malaria vector An. arabiensis and the arbovirus vector Aedes aegypti, was assessed towards volatiles collected from the headspace of fresh and dried leaves, and the smoke from burning the dried leaves in a two-choice landing bioassay and in the background of human odour. Volatile extracts from the smoke of burning dried leaves were found to be more repellent than those from fresh leaves, which in turn were more repellent to mosquitoes than volatiles from dried leaves. Of all smoke and fresh volatile extracts, those from Co. citriodora (52% - 76%) and Oc. suave (58% - 68%) were found to be the most repellent, Os. integrifolia (29% - 56%) to be intermediate while Ol. europaea (23% - 40%) and Os. integrifolia (19% - 37%) were the least repellent. One volatile present in each of the fresh leaf extracts of Co. citriodora, Oc. suave and Os. integrifolia was β -ocimene. The levels of β -ocimene reflected the mosquito repellent activity of these three fresh leaf extracts. Female hostseeking mosquitoes responded dose-dependently to β -ocimene, both physiologically and behaviourally, with a maximal behavioral repulsion at 14% β -ocimene. β -ocimene (14%) repelled mosquitoes in 6-minute landing assays comparable to the synthetic insect repellent N, N-diethyl-m-toluamide (10% DEET). Volatiles in the smoke of burning as well as fresh leaves of Co. citriodora and Oc. suave have significant repellent properties against host seeking An. arabiensis and Ae. aegypti mosquitoes.

The dependence of seasonal vector mosquito aquatic larval habitat canopy vegetation reflectance on sun and sensor geometry can potentially provide information on canopy properties of georeferenced seasonal-sampled vector mosquito aquatic larval habitats, but also may be a source of unmodeled systematic error in single-angle

remote sensing measurements. Asner *et al.* (2002) investigated view angle effects on canopy reflectance and spectral mixture analysis (SMA) of coniferous forests using AVIRIS and the angular variability of reflectance measurements from the NASA Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), and the resulting impact on spectral mixture analysis (SMA) using both full-range (400 - 2500 nm) and shortwave-infrared wave-lengths (2080 - 2280 nm; *AutoSWIR*). The study was conducted in coniferous forests in Central Oregon using five AVIRIS overpasses to generate multiple view angle measurements. Canopy reflectance was highly anisotropic, with the strength of the angular signal controlled by species type, canopy cover and soil reflectance. Canopy cover estimates from full-range SMA averaged only slight decreases (~6% relative) toward the retro-solar direction for 16 field plots in the study region. *Auto SWIR* was even less influenced by view angle, exhibiting changes only for large differences in view angle. In addition, *AutoSWIR*'s ability to accommodate spatiotemporal endmember variability led to stronger agreement with field cover values than full-range SMA. The results suggested that while view angle can significantly affect reflectance measurements from AVIRIS, the consequent variability in vegetation cover estimates from various LULC objects (e.g., georeferenced *An. arabiensis* aquatic larval habitats) SMA and *AutoSWIR* is low.

In addition, although many seasonal, vector, mosquito-related, endemic, transmission-oriented studies have investigated the view-angle dependence of vegetation metrics like the NDVI, the effect of viewing geometry on other remote sensing signatures and methods has not been adequately addressed. One of the most recent applications in vegetation, LULC-related, vector, mosquito-related, remote sensing is unmixing algorithms which may be used to estimate the fractional cover of major scene components, or endmembers (e.g., green vegetation, soil and shade), in an extracted spectrally decomposed image pixel. Fractional extent of vegetation is important from biophysical and biogeochemical perspectives, as well as for investigating secondary sources of spectral variation [1]. Ideally, the inclusion of photometric spectrally quantitated shade as an endmember compensates for view angle effects by modeling the portion of each pixel that is blocked from direct sunlight. However, wavelength dependent multiple scattering in vegetation canopies related to georeferenced, seasonal, vector, mosquito-related, explanatory biophysical, time series, simulation models results in shade that is not efficiently spectrally quantifiable. Therefore, there is some variability to be expected in decomposing, seasonal vector, mosquito-related products due to viewing and solar geometry.

Seasonal, canopy-shaded effects may be spectrally quantified by employing multiple unmixing algorithms and a time series, stochastic/deterministic, interpolator for generating robust, seasonal, endemic, transmissionoriented, endmember, eco-epidemiological, risk maps. For example, Jacob et al. (2011) spectrally decomposed a sub-meter spatial resolution (*i.e.*, QuickBird) riceland, An. arabiensis s.s., aquatic, larval, habitat pixel for forecasting productive habitats based on spatiotemporal, field-sampled, count data in Karima, riceland, agro-village, complex in the Mwea Rice Scheme, Kenya using a stochastically interpolated reference signature. Initially, the authors constructed a regression model which revealed that georeferenced, tillering An. arabiensis s.s. aquatic larval habitats were the most productive. Individual decomposed, sub-pixel, endmember, spectral, reflectance estimates from a QuickBird visible and NIR at 0.61 m pixel, spatial, resolution, of a paddy-preparation, aquatic, larval habitat were then extracted by using a Li-Strahler geometric-optical model. The model used three scene components: sunlit canopy (C), sunlit background (G) and shadow (T) generated from the riceland image. The G, C, T components' classes were estimated using ENVI, an object-based classification algorithm. In the object-based technology the Digital Number (DN) of the pixel in every QuickBird band was viewed using the z-profile from a spectral library. After making an atmospheric correction from the image for the epidemiological study site, the DN was converted into ground reflectance. A geo-spatiotemporal, convex, geometrical, explanatory, predictive model was also employed for endmember validation of the spectrally, decomposed, tillered habitat. An Ordinary kriged-based interpolation was performed in GIS using the reference signature generated from the unmixing models. The basic idea of kriging is to predict the value of a function at a given point (e.g., georeferenced, vector, mosquito, aquatic, larval habitat), by computing a weighted average of the known values of the function in the neighborhood of the point. Ordinary kriging assumes stationarity of the first moment (i.e., mean) of all random variables:

$E\left\{Z\left(x_{i}\right)\right\}=E\left\{Z\left(x_{0}\right)\right\}=m$

where m is unknown [1] Linear, unbiased, explanatory, predictors and variance, estimates were derived for all seasonally, productive, georeferenced, immature, *An. arabiensis s.s.* habitats in the epidemiological study site based on the extracted, unmixed, endmember, interpolated, spectral reflectance estimates.

Spectral unmixing tools and other algorithms [e.g., linear, non-linear, LULC] may be employed in PDA-GISdifferentially corrected GPS (DGPS)-RS technologies, to spatially/spectrally quantitate and interpolate visible and NIR pixel reflectance data (target signature) of a productive, seasonal-sampled, empirical, georeferenced, vector, mosquito, aquatic, larval habitat. Thereafter, the data can be employed as a dependent variable in a robust stochastic/deterministic interpolator within a remote cyber-environment for targeting other seasonal high density foci habitat sites along with their spatially/spectrally associated continuous explanatory variables, which can help implement and target larval control strategies. By so doing, the applications of PDA-GIS-DGPS-RS technologies would also enable vector biologists, epidemiologists and other research collaborators to gain a further understanding of state-of-the-art technologies currently available for implementing control operations [e.g., Integrated Vector Management (IVM)]. DGPS uses a network of fixed, ground-based reference stations to broadcast the difference between the positions indicated by the satellite systems and the known fixed positions for broadcasting the difference between measured satellite pseudoranges and actual (*i.e.*, internally computed) pseudo-ranges, and receiver stations (www.amsa.gov.au/navigation).

Integrated Vector Management is a decision-making process for the management of vector populations, so as to reduce or interrupt transmission of vector-borne diseases (http://www.ivmproject.net/); the strategy is based on the premise that effective control is not the sole preserve of the health sector but of various public and private agencies, including communities. Salient attributes of IVM include methods based on knowledge of factors influencing local vector biology; disease transmission and morbidity; use of a range of interventions, often in combination and synergistically; collaboration within the health sector and with other public and private sectors that impact on vectors; engagement of local communities and other stakeholders; and, a public health regulatory and legislative framework (www.who.int/mediacentre/factsheets/). Driving forces behind a growing interest in IVM for malaria eradication would include the need to overcome challenges experienced with conventional single-intervention approaches to targeted vector control as well as recent opportunities for promoting multi-sectorial approaches to human health. IVM would encourage then effective coordination of the control activities of all sectors that would have an impact on seasonal disease transmission including health, housing agriculture and others. Commensurate benefits for non-health-sector partners make it more likely that IVM approaches will be effective (http://www.ivmproject.net/). Alternate wet/dry intermittent irrigation, combined with other vector control methods, has been effective in controlling the vectors of Japanese encephalitis in China, India, Indonesia and Sri Lanka.

Implementation of IVM using PDA-GIS-DGPS-RS cyber-analytical data, however, requires explicit understanding of geo-spatiotemporal, georeferenced, patterns of vector, mosquito-borne, immature habitats in specific ecosystems (e.g., a riceland agro-village complex), for enabling vector biologists epidemiologists, and other data analysts to characterize various seasonal environmental conditions (e.g., flooding, droughts) and link endemic, transmission-oriented, covariate coefficients with other important data (e.g., demographic, meteorological) in both space and time. Spatial autocorrelation is the correlation among values of a single variable strictly attributable to their relatively close locational positions on a two-dimensional surface, introducing a deviation from the independent observations assumption of classical statistics. Positive geo-spatial autocorrelation in a PDA-GIS-DGPS-RS derived time-series, cyber-predictive, risk map means that the empirical-sampled nearby covariate coefficients values (e.g., aggregation of georeferenced vector mosquito aquatic larval habitats with high seasonal larval counts) tend to be similar on the risk map. Habitat-based, endemic, transmission-oriented, eco-epidemiological, risk modeling of mosquito larval interventions employing spatiotemporally quantitated entomological inoculation rates has determined that not all georeferenced aquatic habitats have a significant impact on incidence and prevalence [3].

Real-time bidirectional PDA-GIS-DGPS-RS technologies may measure geo-spatial autocorrelation and other bio-ecological trends in a regressed, time series, empirical datasets of quantitated Euclidean habitat distance measurements, while optionally creating line graphs of those distances and their corresponding z-scores. A z-score is a statistical measurement of a score's relationship to the mean in a group of scores. Commonly, in geo-spatiotemporal, vector, mosquito-related, endemic, transmission-oriented, statistical, risk-based data analyses, a z-score is tabulated from an empirical dataset of georeferenced, explanatory, field or remote-specified, regressors which may be positive or negative, indicating whether a residually forecasted derivative is above or below the mean, and by how many standard deviations [2]. Thereafter, an incremental, geospatial, autocorrelation tool (e.g., weighted matrix, eigenfunction decomposition, algorithm) in a real-time, bidirectional, PDA-GIS-

DGPS-RS cyber-environment may be employed for robustly quantitating a series of georeferencable, seasonal-sampled, larval habitats using a weighted matrix.

In the PDA-GIS-DGPS-RS cyber-environment, a vector, mosquito-related, seasonal, Hadamard matrix, would be a square matrix whose entries are either +1 or -1 and whose rows are mutually orthogonal. In geometric terms, this means that every two different rows in the Hadamard matrix would represent two perpendicular vectors, while in combinatorial terms, it means that every two different rows would have matching entries in exactly half of their columns and mismatched entries in the remaining columns. It is a consequence of this definition that the corresponding properties hold for columns as well as rows in the matrix. The n-dimensional parallelotope spanned by the rows of an $n \times n$ Hadamard matrix in the cyber-environment would then have the maximum possible n-dimensional volume among parallelotopes spanned by vectors whose entries are bounded in absolute value by 1 in the risk model. Equivalently, a Hadamard matrix would have a maximal determinant among matrices with entries of absolute value less than or equal to 1 in the risk model residually forecasted derivatives and so, the extremal solution of a Hadamard's maximal determinant matrix The analogous question for matrices with elements equal to 0 or 1 would then be equivalent since the maximal determinant of a $\{1, -1\}$ matrix of size n would be 2^{n-1} times the maximal determinant of a $\{0, 1\}$ vector, mosquito-related, seasonal, georeferenced, matrix of size n-1. The intensity of the geo-spatiotemporal autoregressive clustering would then be determined by the z-score returned in a real-time, bidirectional, PDA-GIS-DGPS-RS, cyber-endemic, transmission-oriented, risk map.

Typically, as the seasonal-sampled, georeferenced, time-series, aquatic, vector, mosquito-related, larval habitat, Euclidean, distance-based explanators increases, so would the z-score, indicating intensification of clustering in a specified geographic location in an epidemiological interventional study site. At some particular georeferenced habitat distance measurement, however, (e.g., Euclidean distance from an epidemiological capture point to a seasonal, prolific, georeferenced ,malaria-related, aquatic, larval habitat), the z-score would peak in a realtime bi-directional PDA-GIS-DGPS-RS cyber-environment. Furthermore, a vector biologist, epidemiologist and/ or other research collaborator may encounter multiple peaks in an ecological dataset of georeferenced, residually, forecasted, derivatives rendered from a seasonal, predictive, georeferenced, explanatory, vector, mosquito, aquatic, larval, habitat, risk model constructed in real-time, bidirectional, PDA-GIS-DGPS-RS architecture. When regressively, quantitating an empirical-sampled, time series, explanatory, dataset of non-linear georeferenced vector mosquito aquatic, larval, habitat, field and/or remote specified, endemic, transmission-oriented, georeferenced, data, feature attributes, commonly latent autocovariate errors occur. One strategy may be then to identify an appropriate scale of the predictive, geo-spatial, autocorrelation analysis employing bidirectional, PDA-GIS-DGPS-RS, cyber-technologies so as to select the optimal, seasonal, Euclidean, habitat, distance measurements as defined from an and eco-epidemiological capture point. By so doing, seasonal-sampled, georeferenced, vector, mosquito, aquatic, larval habitats that are associated with the statistically significant peaks may be identified and quantified. The temporal scale of the satellite image in a georeferenced cluster may then be determined in a robust, real-time, bidirectional, PDA-GIS-DGPS-RS cyber-environment.

Further, geo-spatiotemporally quantizing z-scores in real-time, bidirectional, PDA-GIS-DGPS-RS technologies can render seasonal, cartographic, interpolations of explanatory, georeferencable, vector, mosquito-related, risk, model, residual forecasts from a regressed, empirical, sampled, dataset of endemic, transmission-oriented, explanatory coefficients. Interpolation is a method of constructing new, explanatory, time series, georeferenced data points within the range of a discrete set of known data points [1]. These seasonal, vector, mosquito-related, interpolated, explanatory coefficients may reflect the intensity of georeferencable time series, clustering, in specific field or remote specified, ecological, empirical datasets.

In empirical, georeferencable, explanatory datasets of seasonal, vector, mosquito-related, endemic, transmission-oriented, risk-based, data points obtained by georeferenced targeted sampling, aquatic larval habitats can represent the values of a function for a limited number of independent variables values. It is often required to interpolate the value of that function for an intermediate value of the independent variable when time series regressive modeling georeferenced immature seasonal vector mosquito-related data feature attributes [2]. This may be achieved by curve fitting or regression analysis. A different problem which is closely related to cartographic interpolation of seasonal-sampled vector mosquito-related georeferenced data is the approximation of a complicated function by a simple function. Suppose the formula for some given function for a predictive, time series, geo-spatiotemporal, malarial-related, epidemiological, risk model is known, but too complex to evaluate efficiently. A few known seasonal-sampled, larval, habitat, georeferenced ,data points from the original function in a real-time, bidirectional, PDA-GIS-DGPS-RS, cyber-environment may be used to create a time series, cartographic interpolation based on a simpler function. Of course, when a simple function is employed to estimate productive, georeferenced, seasonal-sampled, larval, habitat, data points from the original, interpolation errors are usually present; however, depending on the problem domain and the time series, explanatory, interpolation method employed (e.g., block kriging in a PDA-GIS-DGPS-RS cyber-environment using sub-meter resolution satellite data), the gain in simplicity may be of greater value than the resultant loss in accuracy.

Gaussian processes in a predictive, and eco-epidemiological risk model can be powerful, explanatory, nonlinear, interpolation tool in a robust, real-time, PDA-GIS-DGPS-RS cyber-environment. Although Gaussian processes have a long history in the field of seasonal, vector, mosquito-related, regression statistics, they seem to have been employed extensively only in niche areas. With the advent of kernel machines in the machine learning community, seasonal, vector, mosquito-related, forecasting, risk models based on Gaussian processes have become commonplace for problems of time series regression (e.g., kriging) and classification as well as a host of more specialized applications [e.g., spectral endmember decomposition algorithm employed for interpolating canopied-shaded larval habitat objects]. Many interpolation tools in real-time, bidirectional, PDA-GIS-DGPS-RS cyber-environments can be constructed equivalent to particular Gaussian processes. For example, Gaussian, seasonal, quantification of georeferenced, vector, mosquito, larval, habitat, remote processes can be used not only for fitting an interpolant that passes exactly through given empirical, seasonal-sampled, georeferenced, data points but also for regression analyses (e.g., for fitting a curve through noisy data).

In seasonal-sampled, empirical, ecological, georeferencable datasets of explanatory, vector, mosquito data analyses, kriging or Gaussian process regression is a method of interpolation for which the values (e.g., total, larval, habitat, seasonal, density counts) are modeled by a Gaussian process governed by prior covariances, as opposed to a spline chosen to optimize smoothness of the fitted values. Quantitating time series, explanatory, covariance estimates in a robust, real-time, PDA-GIS-DGPS-RS cyber-environment can provide a measure of the strength of the correlation between two or more sets of random variates in an empirical dataset of seasonal-sampled, vector, mosquito-related, biophysical, time series explanators. In mathematics, a spline is numeric function that is piece-wise-defined by polynomial functions, and which possesses a sufficiently high degree of smoothness at the places where the polynomial pieces connect (*i.e.*, knots) [1]. A polynomial may have an expression consisting of explanatory georeferencable seasonal, vector, mosquito-related determinates which can involve operations of addition, subtraction multiplication for deriving explanatory, non-negative, integer exponents in a robust, real-time, bidirectional, PDA-GIS-DGPS-RS cyber-environment. For example, a polynomial of a single, seasonal, malaria, mosquito, (e.g., Anopheles gambiae s.l.)-related, indeterminate or variable tabulation could include x, is $x^2 - 4x + 7$, where x is a single-, sampled, explanatory, covariate, coefficient value (e.g., amount of daily precipitation in an eco-epidemiological, interventional, georeferenced, urbanized, study site), which then could be expressed mathematically as a quadratic polynomial in a robust, real-time, bidirectional, PDA-GIS-DGPS-RS cyber-environment.

A quadratic function in a robust, real-time, bidirectional, PDA-GIS-DGPS-RS cyber-environment, vector, mosquito-related, predictive, and eco-epidemiological, risk model would be a polynomial function of the form $f(x) = ax^2 + bx + c$, $a \neq 0$. By so doing, the expression $ax^2 + bx + c$ may define a quadratic function for constructing a robust, seasonal, vector, mosquito-related, predictive, endemic, transmission-oriented, risk model in a robust cyber-environment. The model residual forecasted derivatives would then be a polynomial of degree 2, or a 2nd degree polynomial, (*i.e.*, a quadratic polynomial). If the quadratic function in a seasonal, explanatory, georeferencable, vector-related, mosquito, larval, aquatic, habitat, seasonal, predictive, endemic, transmissionoriented, risk model in a robust real-time bidirectional PDA-GIS-DGPS-RS cyber-environment, is set equal to zero, then the result is a quadratic equation. In elementary algebra, a quadratic equation is any equation having the form $ax^2 + bx + c = 0$ where x represents an unknown, and a, b, and c are constants not equal to 0. If a = 0 in a robust, quadratic, seasonal, vector, mosquito-related, predictive, endemic, transmission-oriented, seasonal, ecoepidemiological. risk model constructed in a real-time bidirectional PDA-GIS-DGPS-RS cyber-environment, a quadratic equation is a second-order polynomial equation in a single variable x where $ax^2 + bx + c = 0$, with $a \neq 0$ and because it is a second-order polynomial equation, the fundamental theorem of algebra (*i.e.*, every polynomial equation having complex coefficients and degree ≥ 1 has at least one complex root) guarantees that it has two solutions [3]. The parameters a, b, and c are called, respectively, the quadratic coefficient, the linear coefficient and the constant or free term [3]. Because the quadratic equation involves only one unknown, it is called "univariate" [2]. A seasonal, vector, mosquito-related, real-time, bidirectional, PDA-GIS-DGPS-RS cyber-environment-derived explanatory, predictive, quadratic equation would contain powers of x that are non-negative integers, and therefore would be a polynomial equation. Quadratic, seasonal vector, malaria, mosquito-related, predictive equations can then be solved by a process known as factoring by completing the square, by using the quadratic formula, or by graphing the residually forecasted derivatives in the PDA-GIS-DGPS-RS cyber-environment.

Under suitable assumptions on the priors, kriging vector seasonal georeferenced mosquito-related data in a robust real-time bidirectional PDA-GIS-DGPS-RS cyber-environment would render the best linearized explanatory unbiased prediction of endemic transmission oriented intermediate values (e.g., larval habitat counts). In Bayesian statistical inference, a prior probability distribution, often called simply the prior, of an uncertain p is the probability distribution that would express one's uncertainty about p before some evidence is taken into account [3]. For example, p in a real-time, bidirectional, PDA-GIS-DGPS-RS cyber-environment, could be the proportion of residually forecasted number of larvae in a georeferenced habitat for an upcoming seasonal sampling frame. The prior in a geo-spatiotemporal, vector, mosquito-related, endemic, transmission-oriented, explanatory, georeferenced, risk model is meant to attribute uncertainty rather than randomness to an uncertain quantity [2]. The unknown quantity in the real-time, bidirectional, cyber-environment may then be expressed as a sampled parameter or late explanatory, predictor, variable employing a Bayesian, probabilistic, estimation, probabilistic matrix.

Bayes rule can be derived from more basic axioms of probability, specifically conditional probability [3]. Let (Ω, F, P) be a measure space with $P(\Omega) = 1$, then (Ω, F, P) is a probability space, (e.g., a mathematical construct that models prolific, georeferenced, vector, mosquito-related habitats) with sample space Ω , event space F and probability measure P. When applied, the probabilities involved in Bayes' theorem may have any of a number of probability interpretations for seasonal, georeferenced, vector, mosquito-related, forecasting, endemic, transmission-oriented, explanatory, risk models. In one of these interpretations, the theorem may be used directly as part of a particular approach to statistical inference of the residually forecasted derivatives. In particular, the Bayesian interpretation of probability, the theorem can express how a subjective degree of belief should rationally change to account for evidence: this is Bayesian inference, which is fundamental to Bayesian statistics. Bayes' theorem has applications in a wide range of calculations involving probabilities for seasonal, vector, mosquito-related, georeferencable, predictive, endemic, transmission-oriented, explanatory, risk model derived in a cyber-environment, not just in Bayesian inference. For instance, since individual, sampled, georeferenced, aquatic, larval, habitat, vector, mosquito-related, risk outcomes might be of little practical use, more complex events (e.g., aggregation of flooded larval habitats) may be employed to characterize groups of endemic, transmission-oriented, explanatory, time series outcomes. One applies Bayes' theorem, multiplying the prior by the likelihood function and then normalizing, to get the posterior probability distribution, which is the conditional distribution of the uncertain quantity given the data [3]. A prior for a seasonal, vector, mosquito-related, predictive, eco-epidemiological, risk model is often the purely subjective assessment of an experienced expert.

Theoretically, a conjugate prior may be employed in a real-time, bidirectional, PDA-GIS-DGPS-RS, cyberenvironment to make calculations of the posterior distribution in seasonal-sampled, empirical, regressed, mosquito-related, explanatory dataset easier. In Bayesian probability theory, if the posterior distributions $p(\theta|x)$ are in the same family as the prior probability distribution $p(\theta)$ in the prior and posterior are then called conjugate distributions, and the prior is called a conjugate prior for the likelihood function [3]. In seasonal, vector, mosquito-related statistics, a likelihood function (often simply the likelihood) is a function of the sampled estimators [2]. The likelihood of a set of seasonal-sampled, georeferencable, vector, mosquito-related, parameter values, θ given outcomes x, can be equal to the probability of those observed outcomes given those parameter estimators value in a PDA-GIS-DGPS-RS cyber-environment. A Gaussian prior over the mean in a vector, mosquito-related, endemic, transmission-oriented, seasonal, predictive, epidemiological, risk model may ensure that the posterior distribution is also Gaussian. This means that the Gaussian distribution would be a conjugate prior for the likelihood which would also be Gaussian in a robust, seasonal, vector, mosquito-related model employing real-time bidirectional PDA-GIS-DGPS-RS cyber-tools.

Parameters of prior distributions in a seasonal, vector, mosquito-related, endemic, transmission-oriented, predictive, epidemiological, risk model are called hyperparameters, to distinguish them from parameters of the model of the underlying-sampled data [2]. In Bayesian statistics, a hyperparameter is a parameter of a prior distribution; the term is used to distinguish them from parameters of the model for the underlying system [3]. For instance, if a vector biologist, epidemiologist or research collaborator is employing a beta distribution to model the distribution of a seasonal-sampled, vector, mosquito-related parameter p of a Bernoulli distribution in a real-time, bidirectional, PDA-GIS-DGPS-RS cyber-environment, then p would be a parameter of the underlying system (*i.e.*, Bernoulli distribution), and α and β a would be the sampled parameters of the prior distribution (*i.e.*, beta distribution), hence hyperparameters. In probability theory and statistics, the Bernoulli distribution is the probability distribution of a random variable which takes value 1 with success probability p and value 0 with failure probabity. This is vital in cyber-environments since traditionally seasonal-sampled, georeferenced, vector, arthropod-related aquatic larval habitat values based on non-cyber-environmental derived data has not yielded the most likely intermediate, total, larval, habitat, regressable, georeferencable, explanatory, covariate, coefficient, seasonal, density, count, values.

A real-time, bidirectional, PDA-GIS-DGPS-RS, cyber-environment may utilize a geospatial analytical domain for interpolating seasonal-sampled, empirical, vector, mosquito-related, explanatory dataset. The technique is also known as Kolmogorov Wiener prediction. Linear filters are ubiquitous in applied vector mosquito research ranging from simple differencing operations, mechanical detrending devices and seasonal adjustment to autoregressive models. Many of these filters have very long and often infinite impulse response sequences, such that some approximation procedure is necessary for finite samples. There are also very different kinds of seasonal, explanatory, interpolations for statistically targeting georeferenced, vector, mosquito, aquatic, larval, habitats in PDA-GIS-DGPS-RS cyber-environments. For instance, in the domain of seasonal, vector, mosquito-related, digital, signal processing, interpolations could be performed for converting a geographically, sampled, digital signal to a higher sampling rate (e.g., upsampling) employing various digital filtering techniques (e.g., convolution with a frequency-limited impulse signal). In this application there would be a specific requirement in the real-time, bidirectional, cyber-environment that the harmonic content of the original signal be preserved without creating aliased harmonic content of the original signal (*i.e.*, the original Nyquist limit of the signal above fs/2 of the original signal sample rate). In signal processing, the Nyquist rate is twice the bandwidth of a band limited function or a band limited channel [2].

In a robust, real-time, bidirectional, PDA-GIS-DGPS-RS cyber-environment, seasonal-sampled, georeferenced, vector, mosquito, aquatic, larval, habitat, explanatory, covariate coefficients rigorously regressed based on a response variable representing total, quantized, larval/pupal, density counts can provide additional autoregressive insights for geospatially, targeting, endemic, transmission-oriented zones (e.g., hyperendemic foci). For example, statistically significant peak z-scores computed employing real-time, bidirectional, PDA-GIS-DGPS-RS, technologies can indicate whether a georeferenced, dataset, of explanatory, aquatic, larval, habitat, Euclidean distances has endemic, transmission-oriented, geo-spatiotemporal processes promoting positive/negative, autocovariate, time series clustering. In a real-time, bidirectional, PDA-GIS-DGPS-RS cyber-environment, the seasonal specified autocovariate coefficients derived from regressing an empirical sampled dataset of explanatory, field and/or remote specified, georeferencable, covariate coefficients may be employed in autonormal, autopoisson or autologistic probabilistic, endemic, transmission-oriented, epidemiological risk model frameworks.

There are different ways to formulate the autonormal, autopoisson or autologistic model in a real-time bidirectional PDA-GIS-DGPS-RS cyber-environment depending on the biological context of the empirical sampled georeferenced vector mosquito-related field and/or remote specified explanatory endemic transmission parameter estimator datasets. For example, in a autologistic regression model framework employed in the analysis of vector mosquito species' spatial distributions, constructed in a robust real-time bidirectional PDA-GIS-DGPS-RS cyber-environment, an additional explanatory variable, (*i.e.*, the autocovariate), may be used to correct the latent effects of the geo-spatial autocorrelation coefficients. In statistics, given a real stochastic process X(t), the autocovariance is the covariance of the variable against a time-shifted version of itself. If the process has the mean $E[X_t] = \mu_t$, then the autocovariance is given by

$$C_{XX}(t,s) = E\left[\left(X_t - \mu_t\right)\left(X_s - \mu_s\right)\right] = E\left[X_t X_s\right] - \mu_t \mu_s$$

where E is the expectation operator [6]. The values of the time series, georeferencable, explanatory, vector mosquito-related, geo-spatiotemporal, quantitated autocovariate coefficient would then depend on the sampled field or remote-specified, parameter estimator values of the response variable in the neighborhood as defined by the PDA-GIS-DGPS-RS cyber-tools. This approach has not been employed for biogeographical, risk-based, explanatory, seasonal, real-time, vector, mosquito-related, epidemiological data analyses. Nor has it been assessed for its validity and performance against artificial, simulation, endemic, transmission-oriented, explanatory, georeferenced, field and remote-specified, data variable, with known properties. A geo-spatiotemporal, robust, explanatory, PDA-GIS-DGPS-RS cyber-assessment may quantitate varying range and strength of seasonal, latent, autocorrelation coefficients in sampled, vector, mosquito data as well as the prevalence of the focal species.

Autologistic regression models consistently underestimate the effect of the environmental variables in forecasting models while rendering biased, explanatory, estimates compared to a non-spatial, explanatory, logistic, regression method.

Historically, a drawback of autopoisson, or autologistic, geo-spatiotemporal, eco-epidemiological, forecasting, vector, mosquito-related, risk models constructed employing empirical-sampled, ecological, datasets of georeferenced, explanatory, endemic, transmission-oriented, field and/or remote-specified, covariate coefficients is that for positive autocorrelation the likelihood of the models is not available in closed form. A vector biologist, epidemiologist or a data analyst may investigate how this restriction can be avoided by right truncating a geo-spatiotemporal, explanatory, vector, mosquito-related, time series distribution datasets in a real-time, bidirectional PDA-GIS-DGPS-RS cyber-environment. Truncation is the term for limiting the number of digits right of the decimal point, by discarding the least significant ones [4]-[6]. In a robust, time series, explanatory, PDA-GIS-DGPS-RS cyber-environment, different parameter estimation techniques may be then reviewed which may be deemed applicable to seasonal, vector, mosquito-related, explanatory, field and/or remote-specified, georeferencable, simulation, explanatory, auto-models, for parameter, estimator significance testing. The residually, forecasted, time series derivatives may then be compared with other environmental, simulation, model outputs (e.g., generalized, Bayesian, hierarchical, probabilistic, estimation paridyms). The method may be easily implemented in a robust real-time, bi-directional, geo-spatiotemporal, PDA-GIS-DGPS-RS cyber-environment, digitized, orthogonal, grid matrix via standard statistical software packages and a maximum pseudo-likelihood estimate which may then subsequently render unbiased, explanatory, interpolated ,point estimates in an ecological, explanatory, dataset of residually, forecasted, vector, mosquito-related, time series derivatives.

An alternative method, Monte Carlo maximum likelihood in a PDA-GIS-DGPS-RS cyber-environment may also mathematically contribute to robust, risk-related, cartographic delineations of explanatory, georeferenced, seasonal-sampled, predictive, vector, mosquito-related, real-time, bidirectional operationalizable, eco-epidemiological, endemic, transmission-oriented, risk model residually forecasted derivatives rendered from a time series regression equation. The Monte Carlo maximum likelihood method can render unbiased, explanatory, continuous, estimators of the likelihood function for robustly constructing a family of diffusion-related, seasonal, vector, mosquito-related, forecasting, risk models within a computationally efficient real-time, bidirectional, PDA-GIS-DGPS-RS cyber-infrastructural, regression-based framework. Furthermore, the bidirectional cyber-environment could utilize an exact simulation of diffusions for constructing the predictive, vector, mosquito-related, real-time, seasonal, endemic, transmission-oriented, operationizable, explanatory, predictive, risk model with limited discretization error. A vector biologist, research epidemiologist or data analyst can then show that under regularity conditions, the Monte Carlo maximum likelihood estimation converges to the true likelihood estimation in a forecasting probabilistic, real-time, bidirectional, endemic, transmission-oriented, regression-based vector, mosquito-related, eco-epidemiological, risk model output. For datasize $n \to \infty$, a vector biologist, epidemiologist and/or a data analyst can then employ that the number of Monte Carlo iterations in the PDA-GIS-DGPS-RS cyber-environment which may be tuned as O(n1/2). Monte Carlo iterations can demonstrate the consistency properties of the Monte Carlo likelihood as an observational predictor of the true parameter estimator value in a geospatiotemporal, endemic, transmission-oriented, vector, mosquito-risk-related, explanatory, predictive, ecoepidemiological, risk model using O(n1/2) (2).

In mathematics, big O notation describes the limiting behavior of a function when the argument tends towards a particular value or infinity, usually in terms of simpler functions. It is a member of a larger family of notations that is called Landau notation, Bachmann-Landau notation or asymptotic notation. In computer science, big O notation is employed to classify algorithms by how they respond (e.g., in their processing time or working space requirements in a real-time, bidirectional, PDA-GIS-DGPS-RS cyber-environment to changes in input size. In analytic number theory, it is used to estimate the "error committed" while replacing the asymptotic size, or asymptotic mean size, of an arithmetical function, by the value, or mean value, it takes at a large finite argument. A famous example is the problem of estimating the remainder term in the prime number theorem. Big O notation can characterize seasonal, vector, mosquito-related functions according to their immature growth rates employing different functions in a real-time, bidirectional, PDA-GIS-DGPS-RS cyber-environment. The letter O is used because the growth rate of a function is also referred to as order of the function [3]. A description of a function in terms of big O notations, using the symbols o, Ω , ω , and Θ which may describe other kinds of bounds on asymptotic, vector, mosquito-related, environmental, explanatory, y growth rates [e.g., seasonally quantized land use land cover (LULC) change levels] in robust, real-time, bidirectional, PDA-GIS-DGPS-RS

cyber-environments.

Therefore, if a vector biologist, epidemiologist or a data analyst lets df and g be two functions in a robust explanatory, vector, mosquito, larval, habitat, forecasting, georeferencable, risk model in a real-time, bidirectional, PDA-GIS-DGPS-RS cyber-environment, the residually, forecasted, estimates targeting seasonal, endemic, transmission-oriented areas of interest (e.g., hyperendemic transmission foci) may be defined employing a subset of the sampled, endemic, transmission-oriented explanatory values (e.g., total larval habitat density count values). A vector ecologist, or a data analyst may then use f(x) = O(g(x)) as $x \to \infty$ if and only if there is a positive constant M such that for all sufficiently large values of x, f(x) is at most M multiplied by the absolute value of g(x) [2]. That is, f(x) = O(g(x)) in a robust, georeferencable, autoregressive, vector, mosquito-related, predictive, PDA-GIS-DGPS-RS cyber-environment derived real-time, bidirectional, endemic, transmission-oriented, risk model if and only if there exists a positive sampled real number M and a real number x_0 such that $|f(x)| \le M |g(x)|$ for all $x \ge x_0$.

In many seasonal, explanatory, predictive, vector, mosquito-related, model contexts, the assumption that is of interest is the growth rate (e.g., seasonal larval counts) as the variable *x* goes to infinity is left unstated, thus f(x) = O(g(x)) may be more viable when regressing seasonal-sampled, endemic transmission-oriented, explanatory, covariate coefficients in a real-time, bidirectional, PDA-GIS-DGPS-RS cyber-environment. The notation can also be used to describe the behavior of *f* near some seasonal-sampled, georeferenced, operationizable, vector, mosquito-related, habitat, immature, count values in a robust, PDA-GIS-DGPS-RS, cyber-environment using *a* (e.g., a = 0 during the dry season). As such, f(x) = O(g(x)) as $x \to a$ in a bidirectional, PDA-GIS-DGPS-RS, cyber-environment if and only if there exists positive sampled, vector, mosquito-related, density, count numbers δ and *M* where $|f(x)| \le M |g(x)|$ for $|x-a| < \delta$. If g(x) is non-zero for the sampled, vector, mosquito-related, seasonal, endemic, transmission-oriented, georeferencable, coefficient values of *x* which are sufficiently close to *a*, both of these definitions can be unified in a real-time bidirectional PDA-GIS-DGPS-RS cyber-environment

using the limit superior: f(x) = O(g(x)) as $x \to a$ if and only if $\limsup_{x \to a} \left| \frac{f(x)}{g(x)} \right| < \infty$. The formal defi-

nition of *O* notation is not used directly; rather, the *O* notation for a function *f* is derived by the following simplification rules. If f(x) is a sum of several terms in a robust, real-time, bidirectional, PDA-GIS-DGPS-RS cyber-environment, the one with the largest growth rate may be considered an explanatory georeferenced representative sample, and all others omitted. If f(x) is a product of several factors, any constants (terms in the product that do not depend on x in the cyber-environment may be also omitted.

Conversely, if $f(x) = 6x^4 - 2x^3 + 5$, in an explanatory, time series, quadratic, real-time, bidirectional, malaria mosquito-related, forecasting, epidemiological, risk model, for instance, and suppose a vector biologist, epidemiologist and/or another data analyst wishes to simplify this function, *O* notation may be employed to forecast, seasonal larval/pupal growth rate as *x* approaches infinity to parsimoniously temporally quantitate the statistically significant time series parameter estimators at a certain threshold. This function then could be timely quantitated in a robust, real-time, bidirectional PDA-GIS-DGPS-RS cyber-environment employing the sum of three terms: $6x^4$, $-2x^3$, and 5. Of these three terms, the one with the highest growth rate would be the one with the largest exponent as a function of *x*, namely $6x^4$ in the endemic, transmission-oriented, explanatory, risk model, residually, forecasted derivatives. Thereafter, $6x^4$ may be defined as a product of 6 and x^4 in which the first factor does not depend on x) in the cyber-environment. Omitting this factor will then result in the simplified form x^4 for the risk model derivatives. Thus, f(x) may be of "big-oh" of (x^4) in the bidirectional, real-time, PDA-GIS-DGPS-RS cyber-environment.

Mathematically, $f(x) = O(x^4)$ would be the model output variable in a vector mosquito-related predictive e endemic transmission-oriented eco-epidemiological explanatory risk model. A vector biologist, an epidemiologist or a research collaborator may then confirm this calculation using the formal definition: let $f(x) = 6x^4 - 2x^3 + 5$ and $g(x) = x^4$ in the cyber-environment. Applying, the statement that $f(x) = O(x^4)$ would be then equivalent to the expansion of $|f(x)| \le M |g(x)|$ for some suitable choice of x_0 and M and for all $x > x_0$. This equation may be proven by letting $x_0 = 1$ and M = 13 in a robust vector mosquito-related seasonal risk model constructed in a real-time bidirectional PDA-GIS-DGPS-RS cyber-environment. Then, for all $x > x_0$ the output would be tabulated as: $|6x^4 - 2x^3 + 5| \le 6x^4 + |2x^3| + 5$ so $|6x^4 - 2x^3 + 5| \le 13|x^4|$.

Hence, a vector biologist, epidemiologist or a data analyst can introduce and summarize the notion of time series auto-models in real-time bidirectional PDA-GIS-DGPS-RS cyber-environments for deriving robust, geospatiotemporal specifications for risk modeling empirically, georeferencable, explanatory, operationizable sets of time series field or remote-specified, vector, mosquito, endemic, transmission-oriented, data that contains a response variable. This is the model specification that was recently addressed by Jacob *et al.* [7] for empirically regressing an eco-epidemiological dataset of randomized explanatory Gaussian geo-spatiotemporal field and remote specified sub-meter resolution immature seasonal-sampled *An. arabiensis*-related endemic transmission oriented georeferenced explanatory covariate coefficients employing an auto-normal model. The authors revealed that the operationalizable residually forecasted derivatives rendered from the model using the georeferenced explanatory generalized auto-model specifications, including the auto-Poisson and auto-negative binomial probability specifications, robustly risk mapped the time series malarial empirical-sampled *An. arabiensis* s.s.-related ecological datasets while simultaneously accommodating positive spatial autocorrelation uncertainty coefficients.

The central role of the Poisson distribution with respect to the analysis of the georeferencable, seasonal, vector, mosquito-related, immature counts is analogous to the position of the normal distribution in the context of models for continuous data [7]. Accordingly, when seasonally regressed datasets of vector, mosquito, larval/ pupal data comprises of counts, especially for rare events, the optimal probability estimation eco-epidemiological, endemic, transmission-oriented, risk model in a PDA-GIS-DGPS-RS cyber-environment would be one based upon an auto-Poisson specification, which may be written in the form of approximations in the cyber-environment. The auto-log-Gaussian approximation attained from the regression of the seasonal-sampled, georeferenced, vector, mosquito data would circumvent the auto-Poisson's intractable normalizing factor, and both it and the auto-logistic approximation would circumvent the auto-Poisson's restriction to only situations involving negative spatial autocorrelation, a restriction at odds with the real world vector mosquito-related data analyses, since most georeferenced data exhibit positive spatial autocorrelation. The intractable normalizing constant can be resolved in a real-time bidirectional PDA-GIS-DGPS-RS cyber-environment using Markov Chain Monte Carlo (MCMC) procedures.

The negative autocorrelation restriction in a seasonal, explanatory, georeferencable, vector, mosquito-related, endemic, transmission-oriented, eco-epidemiological, risk model can also be resolved through Windsorizing the sampled georeferenced data feature attributes in a real-time, PDA-GIS-DGPS-RS cyber-environment. Winso-rizing is the transformation of statistics by limiting extreme values in the statistical data to reduce the effect of possibly spurious outliers [1]. Negative geo-spatial autocorrelation in many instances may be inconspicuous in empirical datasets of georeferencable, seasonal, vector, mosquito-related, predictive, autoregressive, eco-epidemiological, endemic, transmission-oriented, risk model residually forecasted derivatives. On occasion geo-spatiotemporal, latent, autocovariate traits may be quantitated parsimoniously in a geographically, weighted, regression-based, endemic, transmission-oriented matrix. By quantitating residual, explanatory, autocorrelated, explanators, and robust confidence intervals may be plotted reflecting seasonal distributions of productive, aquatic, larval habitats based on geo-spatiotemporal field-sampled count data in an epidemiological interventional study site (e.g., total weekly collected larval/pupal malarial mosquito data in a riceland agro-ecosystem).

Furthermore, due to the intractability of the normalizing constants, non-normalized explanatory randomized geo-spatiotemporal parameter estimates in an empirical-sampled georeferencable dataset of seasonal vector mosquito-related observational predictors in a real-time bidirectional PDA-GIS-DGPS-RS cyber-environment could be parsimoniously regressed. MCMC technique would be required although attention had been restricted to the natural exponential family of statistical distributions (e.g., auto-normal, auto-logistic, auto-binomial, auto-Poisson, and auto-negative binomial). In statistics, MCMC methods are a class of algorithms for sampling from probability distributions based on constructing a Markov chain that has the desired distribution as its equilibrium distribution [2].

Quasi-maximum likelihood estimate (QMLE, also known as a "pseudo-likelihood estimate" or a "composite likelihood estimate") is an estimate of a parameter θ in a seasonal real-time bidirectional vector mosquito-related endemic transmission oriented epidemiological predictive risk model that may be formed by maximizing a function that is related to the logarithm of the likelihood function in a robust PDA-GIS-DGPS-RS cyber-environment. In contrast, the maximum likelihood estimate would maximize the actual log-likelihood function for a seasonal-sampled vector mosquito-related georeferencable geo-spatiotemporal endemic transmission-oriented epidemio-logical risk model. The function that is maximized to form a QMLE in seasonal real-time bidirectional PDA-GIS-DGPS-RS cyber-environment can be a simplified form of the actual log likelihood function. A common way to

form such a simplified function may be to employ the log-likelihood function of a misspecified mosquito-related model that treats certain sampled georeferenced vector data feature attribute values in a PDA-GIS-DGPS-RS cyber-environment as being independent, even when in actuality they may not be. This would remove any time series parameter estimators from the vector mosquito-related seasonal real-time bidirectional risk model that are used to characterize the dependencies. However, performing this maximum likelihood estimation only makes sense for quantitating time series parameter estimator statistical significance in a PDA-GIS-DGPS-RS cyber-environment, if the dependency structure is a nuisance parameter with respect to the goals of the analysis. As long as the quasi-likelihood function that is maximized in the vector mosquito-related endemic transmission-oriented epidemiological risk model is not overly simplified, the QMLE or composite likelihood estimate generated would be consistent and asymptotically normal. This likelihood estimate would be less efficient than the maximum likelihood estimate in a robust seasonal real-time bidirectional PDA-GIS-DGPS-RS cyber-environment, but may only be slightly less efficient if the quasi-likelihood is constructed so as to minimize the loss of information relative to the actual likelihood. Standard approaches to statistical inference that are used with explanatory vector seasonal mosquito-related maximum likelihood estimates, such as the formation of confidence intervals, and statistics for a seasonal robust model can then be generalized to the quasi-maximum likelihood setting.

On the other hand, an exact maximum likelihood method developed for the parameter estimator statistical significance testing in a seasonal real-time bidirectional PDA-GIS-DGPS-RS cyber-environment employing a non-Gaussian vector mosquito-related nonlinear log-density function would depend on a latent Gaussian dynamic process with long-memory properties. This method would rely on the method of importance sampling based on time series explanatory linearized Gaussian approximating models constructed within a robust seasonal real-time bidirectional PDA-GIS-DGPS-RS cyber-environment from which latent processes can be simulated. Given the presence of a latent long-memory process, a vector biologist, epidemiologist or data analyst could require a modification of the sampling technique in these models. In particular, the long-memory in a seasonal real-time bidirectional PDA-GIS-DGPS-RS cyber-environment would need to be approximated by a finite dynamic linearized process. By so doing, approximations (e.g., multi-seasonal parameter estimators of a prolific georeferenced malaria-related mosquito aquatic larval habitat) may be compared within the cyber-environment derived residualized model outputs. For instance, it may be shown that auto-regression coefficients obtained from minimizing mean squared forecasting errors in a robust seasonal vector mosquito-related endemic transmission oriented seasonal real-time bidirectional risk model can lead to an effective and feasible method for rendering unbiased residual forecasts targeting prolific vector aquatic larval habitats based on geo-spatiotemporal field and/ or remote regressed georeferenced data feature attributes. Thus, in an empirical study quantitating explanatory time series dependent vector mosquito field and/or remote specified endemic transmission dynamics, a vector biologist, epidemiologist or other data analyst may quantitate log-return time series data by univariate and multivariate long-memory stochastic volatile and autoregressive models in a robust seasonal real-time bidirectional PDA-GIS-DGPS-RS cyber-environment employing distance-weighting schemes and Monte Carlo maximum likelihood estimation for constructing robust, generalized, long-memory, time series, explanatory, vector, mosquito-related, endemic, transmission-oriented, epidemiological, predictive risk models.

The Global Moran's *I* tool in Geospatial Analyst in seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environments may measure seasonal geo-spatial autocorrelation based on empirical datasets of regressed, georeferenced, seasonal, vector, mosquito-related, aquatic, larval habitat feature locations and feature values. Indicators of spatial association are statistics that evaluate the existence of clusters in the spatial arrangement of a given variable [6]. For instance, suppose a vector biologist, epidemiologist or other data analyst is studying seasonal tabulated malarial-related explanatory regressed monthly prevalence rates among census tracts in a given city constructs a real-time seasonal explanatory predictive epidemiological risk model in a real-time bidirectional PDA-GIS-DGPS-RS cyber-environment, employing the rates mean could efficiently reveal where areas have higher or lower rates of endemic transmission foci occur, than is to be expected by chance alone (*i.e.*, geolocalized georeferenced cluster); that is, the values occurring above or below those of a random distribution in space.

Global spatial autocorrelation is a measure of the overall clustering of the data [2]. One of the statistics used to evaluate global spatial autocorrelation is Moran's *I*, defined by:

$$I = \frac{\frac{N}{S_0} \sum_{i} \sum_{j} W_{ij} Z_i Z_j}{\sum_{i} Z_i^2}$$

where Z_i is the deviation of the variable of interest with respect to the mean; W_{ij} is the matrix of weights that in some cases is equivalent to a binary matrix with ones in position *i*, *j* whenever observation *i* is a neighbor of observation *j*, and zero otherwise; and $S_0 = \sum_i \sum_j W_{ij}$. The matrix *W* is required because in order to address ex-

planatory geo-spatial autocorrelation and also model spatial interaction, a structure to constrain the number of neighbors to be considered needs to be imposed. This is related to Tobler's first law of geography which states that *Everything depends on everything else, but closer things more so*-in other words, the law implies a spatial distance decay function, such that even though all sampled vector mosquito-related observational predictors have an influence on all other sampled observations, after some distance threshold (e.g., Euclidean distance from a georeferenced epidemiological capture point to a prolific sampled aquatic larval habitat) that influence can be neglected.

Distance decay in a geo-spatiotemporal explanatory seasonal vector mosquito-related field and/or remote georeferenced endemic transmission oriented predictive risk model may be graphically represented by a curving line that swoops concavely downward as distance along the x-axis increases in a seasonal residual scatter plot in a real-time bidirectional PDA-GIS-DGPS-RS cyber-environment. Distance decay can be then mathematically represented by the expression $I = 1/d^2$, where I is interaction and d is distance, among other forms. Global spatial analysis or global explanatory geo-spatial autocorrelation analysis would yield one statistic to summarize the whole study area in the seasonal real-time bidirectional cyber-environment. In other words, the global analysis in the PDA-GIS-DGPS-RS cyber-environment would assume homogeneity in the empirical-sampled georeferenced mosquito data. If that assumption does not hold in a vector mosquito-related endemic transmission-oriented ecoepidemiological predictive risk model, then having only one statistic to represent the geo-spatiotemporal autocorrelation coefficients in the real-time bidirectional PDA-GIS-DGPS-RS cyber-environment would not make sense, as the statistic would differ over geographical space. But if there is no global autocorrelation or no clustering in a regressed explanatory dataset of georeferenced field and/or remote seasonal-sampled vector mosquitorelated endemic transmission oriented aquatic larval habitat risk model parameter estimators, a vector biologist, epidemiologist or data analyst could still find georeferencable clusters at a local level using geo-localized geospatial autocorrelation coefficients.

Moran's I is a summation of individual cross products which may be exploited by the "Local indicators of spatial association" (LISA) to evaluate the clustering in those individual units (e.g., prolific georeferenced vector mosquito aquatic larval habitats) by calculating Local Moran's I for each spatial unit and evaluating the statistical significance for each I_i . From the previous equation a geo-spatiotemporal real-time bidirectional vector mosquito-related endemic transmission-oriented autoregressive risk model:

$$I_i = \frac{Z_i}{m_2} \sum_j W_{ij} Z_j$$

may be robustly constructed employing PDA-GIS-DGPS-RS cyber-tools where: $m_2 = \frac{\sum_i Z_i^2}{N}$. By so doing,

 $I = \sum_{i} \frac{I_{i}}{N}$ I would be the Moran's I measure of global autocorrelation in the cyber-environment where, I_{i} is local,

and *N* would be the number of analysis units in the endemic transmission-oriented georeferenced time series explanatory predictive risk map. Thereafter, a vector biologist, epidemiologist and/or other data analyst may employ a default conceptualization of the geospatial explanatory seasonal vector mosquito related endemic transmissionoriented field and/or remote specified relationships in a real-time bidirectional PDA-GIS-DGPS-RS cyber-environment for performing a robust Hot Spot Analysis tool to geographically specify all the geographically and non-geographically sampled time series larval/pupal habitat explanatory Euclidean distance values.

A Hot Spot Analysis tool embedded in a seasonal real-time bidirectional PDA-GIS-DGPS-RS cyber-environment could calculate the Getis-Ord Gi* statistic for each time series sampled explanatory field or remote specified vector mosquito aquatic larval habitat endemic transmission-oriented georeferencable data feature attribute in an empirical dataset. By so doing, the resultant z-scores and p-values could then determine where the sampled georeferenced vector mosquito larval habitat data features attributes with either high or low values (e.g., total seasonal larval habitat density counts) cluster geospatially. This seasonal real-time bidirectional PDA-GIS-DGPS-RS cyber-infrastructure tool would work by invasively examining each sampled georeferenced time series explanatory aquatic larval habitat feature attribute within the context of neighboring sampled aquatic larval habitat features. A georeferenced aquatic larval habitat explanatory data feature attribute with a high larval count value may reveal the importance of a sampled field and/or remote specified parameter estimator, but the geographic location of the sampled covariate may not be a statistically significant hot spot. To be a statistically significant hot spot, a georeferenced explanatory data feature attribute in a seasonal real-time bidirectional PDA-GIS-DGPS-RS cyber-environment would have to have a high value (e.g. total seasonal larval density count) and be surrounded by other geographically sampled features with high values as well (e.g. positive autocorrelation). The local sum for a georeferenced field and/or remote time series specified explanatory endemic transmission-oriented data feature attribute and its neighbors could then be compared proportionally to the sum of all the georeferenced vector mosquito data feature attributes in the cyber-environment.

When the local sum is very different from the expected local sum, and that difference is too large to be the result of random chance in an explanatory geo-spatiotemporal vector mosquito-related field and/or remote sampled endemic transmission-oriented eco-epidemiological predictive risk model, a statistically significant z-score would result. For instance, a time series Gi* statistic quantized in a robust seasonal explanatory real-time bidirectional PDA-GIS-DGPS-RS cyber-environment could return a z-score for each empirical sampled vector mosquito-related field and/or remote specified georeferenced endemic transmission oriented data feature attribute in an empirical regressed dataset of explanatory covariate coefficients. For statistically significant positive z-scores, the larger the z-score is, the more intense the clustering of high values (*i.e.*, hot spot), while for statistically significant negative z-scores, the smaller the z-score is, the more intense the clustering of low values (*i.e.*, cold spot) [2]. This PDA-GIS-DGPS-RS cyber-tool could then create a new output feature class with a z-score and p-value for each sampled georeferenced aquatic larval habitat explanatory data feature attribute in an input feature class.

If there is a selection set applied to the input feature class in a seasonal real-time bidirectional explanatory PDA-GIS-DGPS-RS cyber-environment, then only selected sampled georeferenced aquatic larval habitat endemic transmission oriented data feature attributes would be analyzed, and only selected attributes would appear in the output feature class. This cyber-tool would also optimally return the z-score and p-value field names as derived output values for potential use in other custom time series, explanatory, vector, mosquito-related, field and/or remote specified, risk models and scripts derived in the cyber-environment. When this cyber-tool runs, the output georeferenced habitat feature class would automatically be added to the table of contents with default rendering applied to the z-score field. The hot to cold spot rendering applied to the geo-spatiotemporal vector mosquito-related endemic transmission-oriented risk model would then be defined by a layer file in the cyber-environment. Thereafter, a vector biologist, epidemiologist and/or other research collaborator could reapply the default rendering, if needed, by importing a template layer of symbology. The recommended and default conceptualization of spatial relationships for the Hot Spot Analysis (e.g., Getis-Ord Gi*) tool would then be expressed as a fixed distance band in the seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment.

Fixed distance band method works well for point data as it is often a good option for polygonizing time series georeferencable data when there is a large variation in polygon size (e.g., very large seasonal epidemiological vector mosquito-related field and/or remote specified time series polygons at the edge of the epidemiological interventional study site and very small polygons at the center of the study site) (www.esri.com). Peak distances may then be appropriate for parsimoniously quantitating explanatory georeferenced time series larval habitat sampled values to use for tools with a Distance Band or Distance Radius parameters in seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment when conducting geo-spatiotemporal endemic transmission oriented vector mosquito risk-based data analyses. Relationships between the geo-spatiotemporally quantitated georeferenced data feature attributes may then be defined using both space and time, employing several new explanatory field and/or remote specified parameter estimators. Thereafter, by performing a robust space/time analyses, a vector biologist, epidemiologist and/or research collaborator may select an option in the cyber-environment for determining the conceptualization of the geo-spatiotemporal relationships in an ecological empirical dataset of seasonal-sampled parameter estimators selected for regression in the cyber-environment. By so doing, a vector biologist, epidemiologist and/or other research collaborator could define space in a seasonal explanatory georeferenced vector mosquito-related field and/or remote specified endemic transmission oriented predictive epidemiological risk model by specifying a threshold distance value; [e.g., specifying a Date/Time Field and both a Date/Time Type (such as HOURS or DAYS) and a Date/Time Interval Value], all of which may be independent interpolatable explanatory parameter estimators in the risk model. Commonly in geo-spatiotemporal

vector arthropod-related seasonal endemic transmission-oriented explanatory autoregressive covariate coefficients, the Date/Time Interval Value is an Integer [2]. For instance, if 1000 feet in entered in a robust PDA-GIS-GPS-RS cyber-environment, for an elevational-related explanatory covariate associated to a particular georeferenced vector mosquito-related larval/pupal aquatic larval habitat, selecting HOURS can provide Date/Time Interval Value features within 1000 feet and those occurring within three hours of each sampled georeferenced aquatic larval habitat other would be considered neighbors.

For performing a time series Hot Spot Analyses in a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment, seasonally-derived vector mosquito-related epidemiological forecasting risk model, an output source of identification may be added for use in for joining an output georeferencable explanatory dataset to the original dataset. Given a set of weighted vector geo-spatiotemporal-sampled georeferenced mosquito data attribute features, a PDA-GIS-GPS-RS cyber-environment can statistically identify significant hot spots and cold spots using the Getis-Ord Gi* statistic. This tool can identify statistically georeferenced seasonal vector mosquito geospatial clusters of high values (i.e., hot spots) and low values (i.e., cold spots). It can create a new Output Feature Class with a z-score and p-value for each feature in the Input Feature Class in the cyber-environment. The statistic can also return the z-score and p-value field names as robust derived output values for potential use in customized vector mosquito-related real-time predictive risk models and scripts. The z-scores and p-values are measures of statistical significance which tell you whether or not to reject the null hypothesis, feature by feature [2]. In effect, a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment can indicate whether the observed spatial clustering of high or low explanatory, georeferencable, operationizable, vector mosquito data values is more pronounced than one would expect in a random distribution of those same values. A high z-score and small p-value for a feature indicates a spatial clustering of high values. A low negative z-score and small pvalue indicates a spatial clustering of low values [3]. The higher (or lower) the z-score of the seasonal georeferenced vector mosquito-related explanatory endemic transmission-oriented covariate coefficients the more intense the clustering. A z-score near zero of a regressed dataset of vector related would then indicate no apparent spatial clustering. The z-score is based on the randomization null hypothesis computation [3].

Seasonal, explanatory, georeferencable, real-time, bidirectional, vector, mosquito-related data calculations based on either Euclidean or Manhattan distance require projected data to accurately measure habitat distances. For line and polygon features, feature centroids may be employed in distance computations in a robust real-time bidirectional PDA-GIS-DGPS-RS cyber-environment. For multipoints, polylines, or polygons with multiple parts, the centroid is computed using the weighted mean center of all feature parts [2]. The weighting for georeferenced point-related seasonal larval habitat attribute features in the cyber-environment would thus be 1 whereby, line features would be the length of the habitat and the polygon features would be the sample area. The Input Field in the cyber-environment would contain a variety of seasonal-sampled values. The math for this statistic would require some variation in the sampled variable being analyzed; it cannot solve if all input seasonal real-time bidirectional georeferenced vector mosquito-related endemic transmission-oriented explanatory covariate coefficient values are 1. If a vector biologist, epidemiologist or data analyst wants to use this tool to delineate the spatial trend pattern of incident data in a robust PDA-GIS-GPS-RS cyber-environment. This would involve aggregating the incident data into empirical sampled geographic datasets. The conceptualization of the georeferenced explanatory vector mosquito seasonal-sampled data feature attribute spatial relationship parameter estimators would then reflect inherent relationships among the sampled habitat features analyzed.

Within the seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment the FIXED_DISTANCE_ BAND configuration would be able to quantitate the default Distance Band or Threshold Distance which could then subsequently ensure each sampled georeferenced habitat feature attribute has at least one neighbor. But often, this default will not be the most appropriate distance to employ for robustly spatially quantitating geospatial georeferenced explanatory vector mosquito-related covariate coefficients. Additional strategies for selecting an appropriate scale (*i.e.*, distance band) for the analysis may be outlined in selecting a fixed distance band value in a robust seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment.

The INVERSE_DISTANCE or INVERSE_DISTANCE_SQUARED in a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment may be employed for seasonal vector mosquito-related explanatory parameter estimator statistical significance estimation. When zero is entered for the Distance Band or Threshold Distance parameter in the cyber-environment, all the sampled features may be considered neighbors of all other georeferenced data feature attributes; when this parameter is left blank, the default distance will be applied au-

tomatically in the cyber-environment. Weights for distances less than 1 become unstable when they are inverted [2]. Consequently, the weighting for seasonal-sampled vector mosquito-related explanatory georeferenced habitat feature data attributes separated by less than 1 unit of distance (common with Geographic Coordinate System Vector mosquito-related projections) may be given a weight of 1 in a real-time bidirectional PDA-GIS-GPS-RS cyber-environment.

Analysis on features with a Geographic Coordinate Systemvector mosquito-related projections is not recommended when the inverse distance-based spatial conceptualization methods (INVERSE DISTANCE, INVERSE_DISTANCE_SQUARED, or ZONE_OF_INDIFFERENCE) is selected in a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment. For the inverse distance options (INVERSE DISTANCE, INVERSE DISTANCE SQUARED, or ZONE OF INDIFFERENCE), any two seasonal sampled vector mosquito-related habitat points that are neigbours will be given a weight of one to avoid zero division. This would assure the sampled georeferenced larval habitat features are not excluded from the analysis. Additional options for the conceptualization of the seasonal geo-spatiotemporal relationships in the regressed vector mosquito-related parameter estimators, including space-time relationships, may be also quantitated in a robust seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment using the Generate Spatial Weights Matrix (SWM) or Generate Network Spatial Weights tools. To take advantage of these additional options, a vector biologist, epidemiologist or a data analyst may employ one of these tools to construct the time series spatially autoregressive vector mosquito-related weighted matrix file prior to analysis; by selecting GET SPATIAL WEIGHTS FROM FILE for parsimoniously quantitating the conceptualization of spatial relationships parameter; and for the Weights Matrix File parameter in the seasonal real-time bidirectional real-time cyber-environment while, specifying the path to the spatial weights file created.

Seasonal vector mosquito-related map layers can be used to define the Input Feature Class in the regressionrelated predictive risk model constructed in a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment. When employing a layer with a selection in the cyber-environment, however, only the selected features may be included in the analysis. If a vector biologist, epidemiologist or data analyst provides a Weights Matrix File with an spatial autoregressive extension in the cyber-environment, this tool would then be expected to conduct robust parameter estimator significance level quantification in a robust geo-spatiotemporal weights matrix file created using either the Generate Spatial Weights Matrix or Generate Network Spatial Weights tools; otherwise, this tool would be expected as an ASCII formatted spatial weights matrix file. In some seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment derived vector mosquito-related predictive seasonal risk models, the expected behavior of the residually forecasted derivatives may be different depending on which type of spatial weights matrix file is employed for the regressive quantization process. ASCII-formatted spatial weights matrix files apply specific time series weights when constructing seasonal explanatory geo-predictive predictive vector mosquito-related endemic transmission-oriented epidemiological risk models. Missing georeferenced data feature-to-feature attribute relationships in seasonal vector mosquito-risk models may then be treated as zeros. The default weight for self potential is zero, unless you specify a Self Potential Field value, or include self potential weights explicitly [2]. If the seasonal real-time bidirectional weights are row standardized, results will likely be incorrect for analyses on selection sets. If a vector biologist, epidemiologist or a data analyst runs an analysis on a selection set, a robust parameter estimation may be conducted by converting the ASC11spatial weights file to a SWM file by reading the empirical sampled vector mosquito-related log-transformed ASCII data into a table and then converting the table using the CONVERT TABLE option within a robust Spatial Weights Matrix in the cyber-environment. If the weights are row standardized, they will be restandardized for selection sets; otherwise, weights would be the default weight for self-potential.

Running a seasonal vector mosquito-residual risk based explanatory data analyses within an ASC11formatted SWM in a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment would be memory intensive. For example, if there exists an empirical sampled dataset of seasonal georeferencable vector data analyses on more than 5000 attribute features, a vector biologist, epidemiologist or data analyst may consider converting the ASCII-formatted SWM file into an SWM formatted file in the cyber-environment. This can be performed by putting ASCII weights into a formatted table (using Excel, for example) in the real-time bidirectional PDA-GIS-GPS-RS cyber-environment. Next, the Generate Spatial Weights Matrix tool using CONVERT_TABLE for the Conceptualization of Spatial Relationships parameter would be run. The output would be an SWM-formatted spatial weights matrix file. When this tool runs in the cyber-environment, an Output Feature Class would be automatically added to the table of contents with default rendering applied to the z-score field. The hot-to-cold rendering

applied would then be defined by a layer file in the cyber-environment. The default rendering may be re-applied, if needed, by importing the template layer symbology. The Output Feature Class in the seasonal PDA-GIS-GPS-RS cyber-environment would then include a SOURCE_ID field which would allow joining the Input Feature Classes.

When using shapefiles, however, a seasonal real-time bidirectional cyber-environment may not be able to store null values. Tools or other procedures that create shapefiles from nonshapefile inputs in the cyber-environment may store or interpret seasonally regressed vector mosquito-related null values as zero. In some cases, nulls are stored as very large negative values in shapefiles (www.esri.com). However, this can lead to unexpected seasonal residually forecasted misspecified derivatives. Conversely, a new parameter may be added to any output report file that allows the vector biologist, epidemiologist and/or other research collaborators to specify the location of the optional to PDF report. The seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment could then include new seasonal vector mosquito field and/or remote specified data to the report and diagnostics from the many regression matrices [e.g., Ordinary Least Square (OLS)] analysis, as well as several graphics to help interpretation.

A exploratory regression tool in a robust seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment could evaluate all possible combinations of input candidate explanatory time series georeferenced vector mosquito-related field and/or specified endemic transmission oriented predictor variables for constructing OLS models that best explain a dependent variable (e.g., total seasonal-sampled aquatic larval habitat density count data) within the context of user-specified criteria. In statistics, OLS or linear least squares is a method for estimating the unknown parameters in a linear regression model [6]. This method would minimize the sum of squared vertical distances between the observed responses in an empirical-sampled dataset of georeferenced vector mosquito related field or remote specified endemic transmission oriented parameter estimators and the responses predicted by the linear approximation. The resulting estimator may then be expressed by a simple formula in the seasonal real-time bidirectional cyber-environment, especially in the case of a single repressor on the right-hand side. For instance, suppose the resulting regressed field or remote specified vector mosquito-related endemic transmission oriented explanatory risk model consists of no bservations $\{y_i, x_i\}$ $n_i = 1$ in a robust PDA-GIS-GPS-RS cyber-environment. Each seasonal real-time bidirectional observation could then include a scalar response y_i and a vector of p endemic transmission oriented repressors x_i . In a linear vector mosquito related regression-based risk model constructed in the cyber-environment, the response variable would be a linear function of the regressors: $y_i = x'_i \beta + \varepsilon_i$ then where β is a $p \times 1$ vector of unknown field and/or remote parameters; $\varepsilon_i s$ are unobserved scalar random variables (i.e., errors). The residually forecasted regression derivatives in the cyber-environment could then account for the discrepancy between the actually observed endemic transmission oriented responses y_i and the "predicted outcomes" $x'_i\beta$; while simultaneously denoting a matrix transpose, so that $x'\beta$ is the dot product between the vectors x and β .

In mathematics, the dot product, or scalar product (or sometimes inner product in the context of Euclidean space), is an algebraic operation that takes two equal-length sequences of numbers (usually coordinate vectors) and returns a single number [6]. This operation in a robust, time series, explanatory, vector, mosquito-related, field and/or remote specified, endemic, transmission oriented, epidemiological risk model may be defined either algebraically or geometrically in a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment. This model can also be written in matrix notation as $y = X\beta + \varepsilon$, where y and ε are $n \times 1$ vectors, and X is an $n \times p$ matrix of regressors, which is also sometimes called the design matrix [6]. As a rule, in a robust explanatory vector arthropod-related field and/or remote specific endemic transmission oriented risk model the constant term is always included in the set of regressors X, by taking $x_{i1} = 1$ for all $i = 1, \dots, n$ [2]. The coefficient β_1 corresponding to this regressor would then be the intercept. There may be some relationship between the time series dependent vector arthropod-related field and/or remote specified explanatory endemic transmission oriented regressors. For instance, the third regressor may be the square of the second regressor in the seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment derived epidemiological, time series, predictive, risk model. In this case, assuming that the first regressor is constant, a vector biologist, epidemiologist and/or other research collaborators may generate a quadratic model specification in the second regressor. The model which would still be considered linear as it would be linear in the β seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment. In mathematics, the term quadratic describes something that pertains to squares, to the operation of squaring, to terms of the second degree, or equations or formulas that involve such terms [6].

The OLS estimator generated from a robust seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment would be consistent when the explanatory georeferenced field or remote specified endemic transmission oriented time series regressors are exogenous and there is no perfect multicollinearity, and are optimal in the class of linear unbiased estimators when the errors are homoscedastic and serially uncorrelated. Multicollinearity refers to a situation in which two or more explanatory variables in a multiple regression model are highly linearly related [6]. In perfect multicollinearity in a seasonal vector arthropod-related regression-based empirical dataset of residually forecasted explanatory georeferencable field and/or remote specified derivatives, the correlation between two independent variables is equal to 1 or -1 [2]. In practice, vector biologists, epidemiologists and/or other research collaborators rarely face perfect multicollinearity in empirically regressed datasets of explanatory time series georeferenced field and remote specified vector mosquito-related endemic transmission oriented explanatory covariate coefficients [2]. More commonly, the issue of multicollinearity would arise in a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment derived vector mosquito-related endemic transmission oriented risk models when there is an approximate linear relationship among two or more independent variables in the model.

Mathematically, an empirical dataset of time series, explanatory, vector, mosquito-related, field or remote, explanatory, endemic, transmission oriented, predictive variables would be perfectly multicollinear in a robust PDA-GIS-GPS-RS cyber-environment if there exists one or more exact linear relationships among some of the sampled variables. For instance, if there exists $\lambda_o + \lambda_1 X_{1i} + \lambda_2 X_{2i} + \dots + \lambda_k X_{ki} = 0$ holding for all sampled malaria mosquito-related time series field and remote specified endemic transmission oriented observations *i*, in an empirical sampled dataset where λ_j are constants and X_{ij} is the *i*th observation on the *j*th explanatory variable. Then multicollinearity may be examined in a real-time bidirectional PDA-GIS-DGPS-RS cyber-environment by attempting to obtain estimates for the parameters of the multiple regression equation $Y_i = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki} + \varepsilon_i$ [1.1]. The OLS estimates would then involve inverting the geo-spatiotemporal vector malarial mosquito-related

matrix $X^{T}X$ where $X = \begin{bmatrix} 1 & X_{11} & \cdots & X_{k1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1N} & \cdots & X_{kN} \end{bmatrix}$ in the PDA-GIS-GPS-RS cyber-environment. If there is an

exact linear relationship (e.g., perfect multicollinearity) among the independent variables in the cyber-environment, the rank of X (and therefore of $X^T X$) would then be less than k + 1, and the matrix $X^T X$ would not be invertible. For instance, suppose that instead of the above equation 1.1, a vector biologist, epidemiologist and/or other research collaborators construct a vector time series malaria-mosquito (e.g., *Anopheles gambiae s.l.*) related endemic transmission oriented forecasting equation in a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment employing a modified form with an error term $v_i: \lambda_o + \lambda_1 X_{1i} + \lambda_2 X_{2i} + \dots + \lambda_k X_{ki} + v_i = 0$. In this case, there would be no exact linear relationship among the seasonal –sampled field and/or remote specified time series explanatory endemic transmission oriented variables, but the X_j variables would be nearly perfectly multicollinear if the variance of v_i is small for some set of the sampled covariate coefficient values for the λ 's. In this case, the matrix $X^T X$ would have an inverse residually forecasted derivative, but it would be ill-conditioned so that a given computer algorithm in a robust PDA-GIS-GPS-RS cyber-environment, may or may not be able to compute an approximate inverse, and if it does the resulting computed inverse may be highly sensitive to slight variations in the sampled data (e.g., magnified effects of rounding error); thus, rendering misspecifications in the output.

In statistics, a sequence or a vector of random variables is homoscedastic if all random variables in the sequence or vector have the same finite variance [2]. The variance is a parameter that describes, in part, either the actual probability distribution of an observed population of numbers, or the theoretical probability distribution of a sample (a not-fully-observed population) of numbers [6]. In the latter case, an empirical regressed sampled dataset of georeferenced explanatory field and/or remote specified explanatory endemic transmission oriented covariate coefficient distributions tabulated in a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment may be used to construct an estimate of its variance: in the simplest cases this estimate can be the sample variance (e.g., homogeneity of variance). The complementary notion is called heteroscedasticity. The assumption of homoscedasticity simplifies mathematical and computational treatment in a time series, linearized, vector, mosquito-related, field and/or remote, specified, explanatory, endemic, transmission-oriented, epidemiological risk model. Serious violations in homoscedasticity (assuming a distribution of vector mosquito-related sampled data is homoscedastic when in actuality it is heteroscedastic) may result in overestimating the goodness of fit in an empirical dataset of sampled field and remote explanatory specified covariate coefficients as measured by the Pearson coefficient in a robust seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment. In statistics, the Pearson product-moment correlation coefficient (sometimes referred to as the PPMCC or PCC or Pearson's r) is a measure of the linear correlation (*i.e.*, dependence) between two variables X and Y, giving a value between +1 and -1 inclusive, where 1 is total positive correlation, 0 is no correlation, and -1 is total negative correlation [6].

Under some conditions, the method of OLS provides time series minimum-variance mean-unbiased estimation when the errors have finite variances. For example, under the additional assumption that the errors be normally distributed in a robust real-time bidirectional PDA-GIS-GPS-RS cyber-environment, OLS would be the maximum likelihood estimator for a seasonal filarsis mosquito-related (e.g., *Culex quinquefasciatus*) predictive endemic transmission-oriented explanatory epidemiological risk model. The OLS tool in the cyber-environment would produce an output feature class and optional tables with coefficient information and diagnostics for rendering robust inferences. All of these would be accessible in the robust real-time bidirectional cyber-environment Results window. The output feature class would be automatically added to the table of contents, with a hot/cold rendering scheme applied to the endemic transmission-oriented explanatory risk model forecasted derivatives. A full explanation of each output would be provided when interpreting the OLS results in the land cover data, for example, in the seasonal PDA-GIS-GPS-RS cyber-environment, to update information about the environment (e.g., natural or urban land cover) in seasonal mosquito larval control operational areas at a resolution that may not be obtained on the ground. Such data can be processed to provide potentially relevant information about seasonal environmental conditions critical for IVM implementation.

Further, to ensure a consistent scale of analysis Space-Time Window, Zone of Indifference, Contiguity, K Nearest Neighbor, and Delaunay Triangulation may also work well within a robust seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment for parsimoniously quantitating topographic trends in an empirical ecological dataset of geo-spatiotemporal sampled vector mosquito-related explanatory specified field and/or remote endemic transmission-oriented georeferencable feature data attributes. In pattern recognition, the k-Nearest Neighbors algorithm (or k-NN for short) is a non-parametric method used for classification and regression [1]. In both cases, the input would consist of the k closest training examples in the feature space in the constructed time series vector mosquito-related predictive endemic transmission-oriented explanatory risk model. The output in the bidirectional PDA-GIS-GPS-RS cyber-environment would depend on whether k-NN is used for classification or regression. In k-NN classification, the output is a class membership [1]. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor [3]. In k-NN regression, the output is the property value for the object. This value in a time series vector mosquito-related epidemiological predictive risk model parameter estimator quantitation process in the cyber-environment would be the tabulated average of the sampled field and/or remote explanatory coefficient values of its k nearest neighbors. k-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification [1]. The k-NN algorithm is among the simplest of all machine learning algorithms

In mathematics and computational geometry, a Delaunay triangulation for a set P of points in a plane is a triangulation DT(P) such that no point in P is inside the circumcircle of any triangle in DT(P) [3]. Delaunay triangulations maximize the minimum angle of all the angles of the triangles in the triangulation; they tend to avoid skinny triangles. For a set of seasonal-sampled vector mosquito-related aquatic larval habitat points on the same line there is no Delaunay triangulation (the notion of triangulation is degenerate for this case). For four or more points on the same circle (e.g., the vertices of a rectangle) in a real-time bidirectional PDA-GIS-GPS-RS cyber-environment, the Delaunay triangulation is not unique: each of the two possible triangulations that split the quadrangle into two triangles may not satisfy the "Delaunay condition", (*i.e.*, the requirement that the circumcircles of all triangles have empty interiors). By considering circumscribed spheres, the notion of Delaunay triangulation in a robust PDA-GIS-GPS-RS cyber-environment would extend to three and higher dimensions in a vector mosquito-related epidemiological predictive risk model residually forecasted derivatives. Generalizations are possible to metrics other than Euclidean in a vector mosquito-related predictive risk model [2]. However in these cases a Delaunay triangulation is not guaranteed to exist or be unique.

Attention to time series spatial patterns in an empirical sampled dataset of georeferenced explanatory specified endemic transmission oriented covariate coefficients in a robust real-time bidirectional PDA-GIS-GPS-RS cyber-environment can lead to insights that may have been otherwise overlooked, while ignoring space may lead to false conclusions about ecological relationships in the sampled data. Jacob *et al.* [7] employed multiple time series Gaussian geo-spatiotemporally autoregressive georeferencable explanatory operationalizable field and remote specified endemic transmission-oriented predictive epidemiological risk models, fit within ArcGIS Spatial Analyst, to examine relationships in geographic space of multiple georeferenced sub-meter resolution imaged aquatic larval habitats of *An. arabienisis*. The time series field and remote explanatory endemic transmission oriented covariate coefficients were sampled in the Karima agro-village complex, Mwea Rice Scheme, Kenya. Interestingly, in preliminary models that ignored space, the abundance of the sampled time series vector malarial mosquito larval habitat species was correlated with both local and landscape-scale seasonal-sampled variables. These models were then modified to account for quantitated broad scale surface spatial trends employing a time series analysis and fine-scale autocorrelation in an autoregressive covariance uncertainty matrix. A covariance matrix for a seasonal, *An. arabienisis*, aquatic, larval habitat, field and/or remote specified, endemic, transmission oriented risk model was constructed as a square matrix whose diagonal entries were the variances of, and whose off diagonal entries which were the covariance's between, the row/column labeling variables.

Residuals from OLS regression models then revealed positive autocorrelation (*i.e.*, georeferenced habitats of similar seasonal *An. arabienisis* aquatic larval habitat density counts aggregated in geographic space), indicating that the assumption of independent errors was violated in the forecasted model derivatives. In contrast, the time series autoregressive residually forecasted derivatives from the explanatory non-linear models showed little spatial pattern, suggesting that these models were appropriate. The magnitude of the sampled explanatory quantitated aquatic larval habitat seasonal effects tended to decrease, and the relative importance of different explanatory georeferenced biophysical explanatory field and remote specified georeferenced endemic transmission-oriented observational predictors shifted when the authors incorporated broad scale predictor variables and then fine-scale space into the analysis. The degree to which the random sampled *An. arabiensis* aquatic larval habitat effects changed when space was added to the models was roughly correlated with the amount of spatial structure in the empirical sampled aquatic larval habitat explanatory predictor variables.

Within a robust seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment, latent time series spatial patterns in empirical explanatory datasets of regressed vector mosquito-related endemic transmission oriented residually forecasted derivatives rendered from OLS models may be parsimoniously quantitated. Failure to include or adequately measure autocorrelation in an empirical dataset of regressed georeferenced field or remote specified explanatory endemic transmission oriented geo-spatiotemporal sampled *An. arabienisis* aquatic larval habitat variables may lead to stochastic/deterministic interpolation misspecifications. Time series seasonal real-time bidirectional *An. arabienisis* species abundance and distribution data may not be explained by a simple logistic regression-related endemic transmission oriented field and/or remote specified aquatic larval habitat risk model.

As such, given a set of time series sampled explanatory seasonal real-time bidirectional vector mosquito-related feature data attribute (e.g., total aquatic larval habitat density count), a robust PDA-GIS-GPS-RS cyberenvironment could determine whether a seasonally regressed spatial pattern expressed in an empirical dataset of endemic transmission oriented georeferencable explanatory risk model residually forecasted derivatives (e.g., locational covariates of an aggregation of productive georeferenced aquatic larval habitats in an urban ecosystem) has latent autocorrelation traits. For example, a geo-spatial autocorrelation tool in a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment could test the null hypothesis that in an endemic transmission oriented a time series malaria mosquito (urban Anopheles gambiae s.l.) explanatory endemic transmissionoriented epidemiological risk model there is no spatial clustering of the georeferenced habitats based on total seasonal-sampled larval density count values for an interventional study site. When the p-value is small and the absolute value of the z score in the cyber-environment is large enough that it falls outside of the desired confidence level, the null would be rejected [3]. If the index value tabulated is greater than 0, a set of sampled georeferenced explanatory aquatic larval habitat endemic transmission-oriented field or remote specified data feature attributes would then exhibit a clustered pattern. If the value is less than 0, the set of mosquito-related data feature attributes would exhibit a dispersed pattern. A PDA-GIS-GPS-RS cyber-tool would then calculate the Moran's I value employing both a z-score and p-value to evaluate the significance of the clustering, dispersion, or randomization in the empirical dataset of georeferenced time series regressed explanatory endemic transmission oriented covariate coefficients. p-values are generally numerical approximations of the area under the curve for a known distribution, limited by the test statistic [2]. The Global Moran's I tool calculates a z-score and p-value to indicate whether or not you can reject the null hypothesis [5]. In the case of a time series robust explanatory

georeferenced vector mosquito-related endemic transmission oriented risk model constructed in a PDA-GIS-GPS-RS cyber-environment, the null hypothesis would state that the sampled larval habitat sampled data feature attribute values, for instance are randomly distributed across an interventional study site area. The z-score would thus be based on the randomization null hypothesis computation in the cyber-environment. The 90 percent confidence interval for treatment effect consists of all the hypothesized values of an effect that are not rejected [1].

For each test, in a robust seasonal explanatory vector mosquito-related explanatory spatiotemporal field and/or remote specified endemic transmission-oriented risk model, a vector biologist, epidemiologist and/or other data analyst could reject the null hypothesis, if the one-tailed p-value in the most extreme direction is less than 0.05 in a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment. One-tailed tests are used for asymmetric distributions that have a single tail, such as the chi-squared distribution, which are common in measuring goodness-of-fit in seasonal vector arthropod-related linearized risk models, or for one side of a distribution that has two tails, such as the normal distribution, which is common in estimating georeferenced larval habitat geolocation [2]; this corresponds to specifying a direction in the cyber-environment. Measures of goodness of fit may then typically summarize the discrepancy between georeferenced observational values and the values expected in a PDA-GIS-GPS-RS cyber-environment under the seasonal predictive vector mosquito-related endemic-transmission oriented epidemiological risk model in question which can be used in statistical hypothesis testing, (e.g. to test for normality of forecasts in a regressed dataset of field or remote sampled endemic transmission oriented covariate coefficients), to determine whether two seasonal real-time bidirectional samples are drawn from identical distributions (e.g., Kolmogorov-Smirnov test), or whether outcome frequencies follow a specified distribution (Pearson's chi-squared test).

In georeferencable seasonal real-time bidirectional vector mosquito-related endemic transmission oriented risk-based predictive time series analyses, the chi-squared distribution (also chi-square or χ^2 -distribution) with k degrees of freedom is the distribution of a sum of the squares of k independent standard normal random variables. A special case of the gamma distribution may be employed for quantitating probability distributions for aggregating inferential statistics and hypothesis testing seasonal georeferenced vector mosquito-related endemic transmission oriented epidemiological risk model confidence intervals) [2]. In a robust seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment, the gamma distribution would be quantitated as a two-parameter family of continuous probability distributions which may be expressed commonly employing different parameterizations in a vector mosquito-related regression-based risk model constructed in a PDA-GIS-GPS-RS cyber-environment. For example, with a shape parameter k and a scale parameter θ in a robust PDA-GIS-GPS-RS cyber-environment and a shape parameter $\alpha = k$ and an inverse scale parameter $\beta = 1/\theta$, a rate parameter for a seasonal dengue-related mosquito (Aedes egypti)-related risk model may be efficiently quantitated. With a shape parameter k and a mean parameter $\mu = k$ quantified a rejection would occur for any observed explanatory field and/or remote sampled endemic transmission oriented value employing a specified threshold estimate region within a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment. For example, a vector biologist, epidemiologist and/or other research collaborator would have 0.10 probability of rejecting any hypothesized effect value in a time series predictive dengue mosquito-related eco-epidemiological risk model if it was true. The seasonal real-time bidirectional model would thus render a 1 - 0.10 = 0.90 confidence level for a robust seasonal dengue mosquito-related epidemiological field and remote specified explanatory predictive risk model constructed in a real-time bidirectional PDA-GIS-GPS-RS cyber-environment.

Optimally, a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-infrastructure for archiving seasonal vector mosquito-related regressed data collections would be a combination of data resources, network protocols, computing platforms and computational services. Most time series explanatory georeferenced vector mosquito-related ecological domains adopt intrinsic geospatial principles such as quantitated seasonal constraints in phenomena evolution (e.g., larval development) for determining robust randomized sample frames. Generally these environments require large amounts of time series data processing for performing robust optimal empirical geospatial explanatory georeferencable risk-based endemic transmission oriented risk-based analyses parsimoniously. These processes commonly include operationalized geospatial analysis of seasonal-sampled explanatory larval habitat data feature attribute relationship calculations, meteorological time series parameter estimator quantitation, land cover observational predictor estimator geovisualization, and geomorphological ecohydrological terrain-related risk-related catchment modeling. The algorithms for these data processing mechanisms may be integrated within a real-time PDA-GIS-GPS-RS cyber-environment seamlessly regardless of the size of the empirical sampled explanatory mosquito-related endemic transmission oriented field and remote specified cova-

riate coefficient geospatial explanatory ecological datasets.

The studies of transmission dynamics based on regressed empirical-sampled explanatory georeferenced field and/or remote specified ecohydrological and vegetation-related LULC covariate coefficients for timely seasonal vector mosquito-related risk-based eco-epidemiological data analyses have been historically more focused on the understanding of mechanisms and controls at the point/plot spatial scale rather than at the scale of topographically complex epidemiological landscapes. For example, although in riceland agro-village ecosystems where topographic vegetation and climatic gradients play a crucial role in controlling time series explanatory ecohydrological irrigated and biogeochemical explanatory land cover covariate coefficient behaviors, the effects of spatially varying georeferenced seasonal real-time bidirectional vector mosquito-related aquatic larval habitat structures based on environmental specifications such as, the role of irrigated water redistribution and terrainrelated ecogeomorphological processes have not been qualitatively nor quantitatively robustly regressed at the watershed scale.

Land degradation in riceland agro-village environments can alter seasonal vector mosquito-related larval habitat productivity through time series quantitated irrigated water erosion variables driven by ecogeomorphological processes which may also alter seasonal-sampled georeferenced aquatic habitat immature density counts along transfer paths at hill slopes [8]. Seasonal soil-hydraulic conditions of the upper soil layers and the vegetation structure of the hill slope can also influence immature vector mosquitoes. These LULC processes may be geospatiotemporally interlinked in seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environments with each other by augmenting feedback mechanisms in such a way that a small change in land use (e.g. temporary overgrazing, cattle trails) may result in real-time robust quantitation of seasonal immature density counts. A canopied forestland-shrub land LULC to rice agriculture LULC transition zone in an eco-epidemiological interventional georeferenced agro-village complex may then be rigorously analyzed in a PDA-GIS-GPS-RS cyberenvironment for generating optimal time series regressively quantitated soil-vegetation-transfer feedback mechanisms that influence seasonal aquatic larval habitat productivity. By so doing, an explanatory eco-geo-morphological time series process-based georeferenced endemic transmission oriented field or remote specified endemic transmission-oriented eco-epidemiological risk model may be developed which simulates the redistribution of sediments and nutrients during high-intensity rainstorm in 1-sec time steps in specific urban environments such as riceland agroecosystems. The soil moisture and transpiration dynamics in daily time steps, and the vegetation dynamics (*i.e.*, establishment, immature growth, mortality) in 14-day time steps for a high-resolution grid of 1×1 m^2 , for instance, may then be quantitatively/qualitatively geo-spatiotemporally regressed in a robust explanatory seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment. Furthermore, through long-term seasonal riceland malaria mosquito (An. arabiensis s.s.)-related predictive epidemiological risk modeling and the modeling of extreme environmental conditions (e.g., prolonged droughts), numerical approaches may be employed in a in a PDA-GIS-GPS-RS cyber-environment for parsimoniously analyzing which type of feedbacks may occur and may trigger persistent seasonal vegetation change and land degradation of hill slopes associated to productive georeferencable aquatic larval habitats based on geo-spatiotemporal field and/or remote immature sampled count data. Employing such models in a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment would then make it possible to couple the occurrence of geospatial time series explanatory spatial patterns of soil moisture of a hill slope georeferenced vector mosquito-related aquatic larval habitat during high-intensity rice irrigated flooding cycles. By so doing, the seasonal vector malarial mosquito field or remote specified endemic transmission oriented risk model could either investigate only individual extreme events or, model the long-term dynamics of remote classified LULC geographic sub-zones in a urbanized environment in a PDA-GIS-GPS-RS cyber-environment for spatially targeting prolific riceland An. arabiensis aquatic larval habitats in an epidemiological study site without including the detailed erosion processes.

At the plot scale, numerous theoretical field experimental predictive seasonal explanatory geo-spatiotemporal autoregressive epidemiological risk-related modeling approaches have been presented in GIS literature. For instance, recently, there has been an emergence of interest in extending time series explanatory eco-hydrological real-time predictive endemic transmission-oriented field and/or remote specified risk models for monitoring and evaluating specific ecosystems by robustly spatially quantitating larger scale georeferencable seasonal explanatory biophysical empirical vector mosquito-related endemic transmission oriented data feature attributes in Arc-GIS Spatial Analyst. Unfortunately, the two-way coupling between seasonal vegetation dynamics and seasonal irrigated rice water cycles was not studied in these models. By not seasonally quantitating vital seasonal relationships during specific temporally dictated sample frames (e.g., tillering in the rice cycle) in a time series vec-

tor mosquito-related endemic transmission oriented field and/or remote specified epidemiological risk model, a robust seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment, geospatially and temporally heterogeneous conditions of watersheds in catchment sites may not be revealed in the log-transformed residually forecasted linear/non-linear regression-based derivatives thus, creating misspecifications in stochastically/deterministically seasonal real-time bidirectional interpolated model. A distributed representation of explanatory georeferencable time series geo-spatiotemporal ecohydrological dynamics and a subsequent aggregation to the watershed scale are thus warranted for investigating time series sampled vector immature mosquito-related parameter estimators in riceland and other urban epidemiological study sites. Extending time series explanatory eco-geohydrological analysis of riceland malaria larval habitat georeferenced data feature attributes then to larger spatial scales within a robust PDA-GIS-GPS-RS cyber-environment may help resolve quantitating important uncertainty related parameter estimators (e.g., heteroscedastic residual, latent auto-correlated coefficients) explanatory endemic transmission oriented covariate coefficients.

Seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environments have domains that can collect, archive, share, analyze, visualize and simulate vector mosquito-related field and remote sampled data, information, and knowledge. Many of these domains can generate valuable predictive seasonal risk-related endemic transmission oriented epidemiological data and information for a malaria-sampled mosquito aquatic larval habitat geographic reference point. Georeferenced or geospatial data feature attributes in a real-time bidirectional robust PDA-GIS-GPS-RS cyber-environment can have spatial/spectral inter-connections that can follow geospatial time series principles/constraints, such as those required for efficient robust, explanatory, time series, endemic, transmission-oriented, vector, mosquito-related, predictive, seasonal risk modeling. A cross-cutting infrastructure that can support geospatial seasonal vector mosquito-related data processing within and across scientific domains is desirable [5]. A robust explanatory seasonal real-time bidirectional PDA-GIS-GPS-RS cyber environment infrastructure can support the collection, management, and utilization of geospatial georeferenced seasonal vector mosquito-related data, information, and knowledge for generating multiple relational seasonal geodatabases and domains.

Historically, the realization of the importance of time series cyber-infrastructures can be traced back conceptually to 1884 when the national program for topographic mapping was started, and formally to 1994 when the US Federal Geographic Data Committee (FGDC) was established to build a cross-agency National Spatial Data Infrastructure (NSDI). Since then, much progress has been made in defining standards by the Open Geospatial Consortium (OGC) and the International Organization for Standardization (ISO), implementing prototypes through various test beds, popularizing industry products through seed funding, and building applications for this infrastructure through several initiatives. For instance, in 2007, the Infrastructure for Spatial Information in the European Community (INSPIRE) directive entered into force and laid down a general framework for a Spatial Data Infrastructure (SDI) to support European Community environmental policies and activities. These initiatives presently support each other with their own unique emphases: for instance, the NSDI focuses on spatial data collection, sharing, and service, and their geodata.gov provides optimal geospatial aspects supplemented by geodata.gov.

Digital Earth is a vision popularized for advancing remote sensing technology to store, integrate, and utilize georeferenced vector seasonal sampled mosquito related endemic transmission-oriented data feature attributes to build a virtual world for multiple epidemiological risk modeling applications employing predictive seasonal real-time bidirectional PDA-GIS-GPS-RS technologies. Digital Earth is seen as a global strategic contributor to scientific and technological developments, and is considered a catalyst in finding solutions to international scientific and societal issues (www.digitalearth-isde.org/). Digital Earth can play a strategic and sustainable role in addressing such challenges to human society as natural resource depletion, food and water insecurity, energy shortages, environmental degradation, natural disasters response and infectious disease vector arthropod-related risk modeling (http://earth.google.com/outreach/showcase.html). Thus, Digital Earth may be an integral part of other advanced technologies in a robust seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment for constructing datasets of robust, georeferencable seasonal, vector, mosquito-related, endemic, transmission-oriented, field and/or remote specified explanatory eco-epidemiological risk modeling seasonal-sampled covariate coefficients including: earth observation, geo-information systems, communication networks, sensor webs, electromagnetic identifiers, virtual reality, grid computation and others. Grid computing may be focused on distributed computers in a PDA-GIS-GPS-RS cyber-environment to optimize distributed computing while Cloud

computing is focused on data, platforms, infrastructure, and software as services for end-users [5]. The Global Earth Observation System of Systems (GEOSS) is an initiative to build a system of systems for global Earth observations focused initially on nine societal benefit areas. These data systems and others then could be part of a robust PDA-GIS-GPS-RS architecture for capturing, managing, scoring and quantitating seasonal-sampled field and/or remote specified explanatory vector mosquito-related endemic transmission oriented risk variables.

Furthermore, over time, the amount and availability of geographic information has grown exponentially, and thus seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-infrastructures is needed to process and integrate seasonal-sampled vector mosquito larval habitat explanatory field or remote specified endemic transmission oriented geospatial information to supply real time risk-based eco-epidemiological analyses and to support scientific and application problems solving across multiple geographic regions. Robust PDA-GIS-GPS-RS cyber-environments would utilize geospatial principles and geospatial information to transform how vector mosquito research, development, and education are conducted within and across other disciplines such as Earth sciences and linear algebra for implementing IVM.

Optimally, seasonal mosquito vector biology-related cyber-environments would be based on recent advancements in GIS, information technology, PDA technology computer networks, sensor networks, Web computing, e-research/e-science and other solutions for optimizing seasonal explanatory observational georeferencable predictive vector mosquito-related data feature attribute geovisualizations for end-users (e.g., malarialogists) to utilize and contribute to the time series cyber-infrastructures. As such, robust real-time bidirectional PDA-GIS-GPS-RS cyber-environment functions for seasonal vector mosquito-related eco-spatiotemporal data analyses, management and storage could then include the following: (a) a middleware layer to bridge time series geospatial functions, resource management, monitoring, scheduling, and other system-level functions, (b) a geospatial information integration layer to parsimoniously quantitate explanatory geospatial explanatory seasonal mosquito sampled data, information, and knowledge flow as supported by multiple field and/or remotes ampled observations, geospatial processing, and knowledge mining; and, (c) geospatial functions to provide various analytical functions for end-users. Vector seasonal mosquito-related time series explanatory predictive geospatial information has a geographic connection that spans multiple domains [9]; thus, the framework of qualitatively/quantitatively regressing empirical geographically-sampled mosquito-related time series explanatory field and/or remote specified endemic transmission oriented datasets can be leveraged to facilitate the sharing of data, information, and knowledge within and across domains [2]. The evolution of cyberspace has resulted in an increasing number of applications to support seasonal vector mosquito-related research, development, and decision making for generalized seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-infrastructure transforms. The introduction of mobile devices will therefore broaden access to time series explanatory georeferenced endemic transmission-oriented field and/or remote specified cyber-infrastrures thereby improving the rate of sharing of geospatial information for robust mosquito-related location intelligence gathering.

Location Intelligence is the capacity to organize and understand complex phenomena through the use of graphical relationships inherent in all information [5]. The goal of location intelligence in a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment for seasonal malarial risk-based endemic transmission oriented field and/or remote specified exploratory data analyses would be to select the best "geographies of opportunity" (e.g., aggregation of georeferenced prolific larval habitats based on geo-spatiotemporal field and/or remote sampled data) by providing a fit-to-purpose geospatial time series analytic solution. Deploying location intelligence by analyzing seasonal data using a PDA-GIS-GPS-RS time series cyberinfrastructural explanatory endemic transmission oriented predictive georeferencable predictive risk model would then be a critical for success in seasonal mosquito control operations. Geolocational or seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-infrastructure tools could then enable vector biologists epidemiologists or other research collaborators to collect, store, analyze and visualize explanatory georeferenced field and/or remote specified endemic transmission oriented data feature attributes (e.g., terrestrial vegetation LULC attributes, aquatic chemistry of larval mosquito abundance in catch basins) accurately. Location intelligence experts are defined by their advanced education in spatial technology and applied use of spatial methodologies [4]. Location intelligence experts can use a variety of seasonal analytical tools to in a PDA-GIS-GPS-RS cyber-environment to measure productive vector mosquito aquatic larval habitat geographical locations for treatment. An important determinant of mosquito-borne pathogen transmission is the spatial distribution of vectors [2]. Optimally, these seasonal prolific larval habitat epidemiological risk mapping applications would transform large amounts of regressed time series dependent explanatory vector mosquito-related endemic transmission-oriented georeferencable field and/or remote specified data feature attributes in a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment into digital and cartographic color-coded visual representations. By so doing, seasonal trends and meaningful intelligence [e.g., levels of yearly change in forested canopied LULC to urban agro-village complex LULC in an epidemiological interventional study site] may be parsimoniously derived.

Land use land cover seasonal vector mosquito-related explanatory field and/or remote specified forecasting endemic transmission oriented georeferencable risk models constructed in a robust seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment would ensure fitness for use; would draw on research into the cognitive issues that surround increasingly personalized and flexible possibilities for risk map use with an expanded range of map forms; would respond to the state of the art in the realm of interfaces; and would drive and respond to developments in the field of vector databases and geocomputation. The creation of time series explanatory vector mosquito-related georeferenced aquatic larval habitat geographic and non-geographic location intelligence in a PDA-GIS-GPS-RS cyber-environment would be directed by domain knowledge, formal frameworks with a focus on decision support using remote sensing technologies.

It is possible to regard explanatory, predictive, seasonal, vector, time-series, mosquito-related, endemic, transmission-oriented, epidemiological risk modeling itself as a form of representation, although it may be assumed to be the seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment data-handling step prior to subsequent geographic representation. Modeling of seasonal geospatialized explanatory vector mosquito-related endemic transmission oriented data is undertaken to ensure that data are captured, held, managed and manipulated in a suitable way for the application in hand [2]. The application may involve representation of all explanatory georeferenced operationalizable field and/or remote seasonal-sampled mosquito aquatic larval habitat types in a PDA-GIS-GPS-RS cyber-environment for archival, communication and/or analytical purposes. Regardless, the visual representation of geospatial seasonal-sampled eco-geomorphological bidirectional mosquitorelated field and/or remote specified endemic transmission oriented real-time data and, in particular, the interaction of contemporary aspects of visualization and geospatial endemic transmission oriented visual time series representation can be addressed in a robust PDA-GIS-GPS-RS cyber-environment. All references for explanatory, georeferencable, time series, field and/or remote specified, endemic, transmission-oriented, malarial mosquito (e.g., immature An. gambiae s.l.) aquatic habitat representations should be broadly interpreted as referring to visual representation unless otherwise qualified [2]. Furthermore, a wide variety of methods for employing multi-source geo-ecomorphological data in digital explanatory vector mosquito-related aquatic larval habitat image classification have been developed which may be implemented in robust seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environments. These data sources may be viewed as attempts to exploit the value of amalgamating multiple sources of geo-spatiotemporal explanatory field and remote sampled data (e.g., ancillary data and/or multiple types of imagery) for extraction of vital information from seasonal georeferenced vector mosquito-related larval habitat images through image interpretation.

Hallmarks of multi-parameter explanatory regressed arthropod-related larval habitat image interpretation logic include (1) use of a systematic strategy that proceeds from "knowns" to "unknowns" (e.g., stochastic and/or deterministic time series interpolated larval habitats) and use of inference and "convergence of evidence" exploiting observed relationships between multiple data types (*i.e.*, image and ancillary data) [2]. For instance, Jacob *et al.* [7] developed a spectral endmember time series predictive autoregressive spectral risk model, using decomposed 0.61 m^2 sub-pixel reflectance estimates in an explanatory multi-temporal QuickBird visible and NIR image which were extracted from a georeferenced Simulium damnosum sensu lato species riverine larval habitat using a Li-Strahler geometric-optical model. In Africa, the parasite primarily transmitted by black flies of the S. damnosum s.l. species complex, a vector of onchocerciasis ("River Blindness") which develop as larvae in fast running rivers and streams. Onchocerciasis, or river blindness, has historically been one of the most important causes of blindness worldwide (http://www.who.int/topics/onchocerciasis). This procedure allowed for the creation of a spectral signature of a unit of habitat. The model employed three scene components: sunlit canopy (C), sunlit background (G) and shadow (T) generated from the QuickBird image, to determine the sub-pixel endmember spectra associated with the known habitats. The C, G and T component classes were estimated using the ENVI software package (Exelis Visual Information Solutions, Boulder, CO) which employs an object-based classification algorithm. In the geometric-optical approach, the bidirectional reflectance emitted from the georeferenced S. damnosum s.l. riverine larval habitat was modeled as a purely geometric phenomenon that resulted when scenes of discrete, three-dimensional objects were illuminated and viewed from different positions in the hemisphere. The shapes of the within canopied shaded objects (e.g., Precambrian rock) and their immature count densities were the

driving variables of the model along with the mixture of sunlit and shaded objects and background that was observed from a particular viewing direction, given a direction of illumination which in turn, controlled the brightness in the black fly riverine larval habitat image.

In Jacob *et al.* [7] non-parametric spatiotemporal estimators from the decomposed endmember spectra and the geometric-optical model were then used to construct a Boolean model that generated a spectral reference target signature in an ArcGIS database specific for the verified immature S. damnosum s.l. habitats. In probability theory, the Boolean-Poisson model or simply Boolean model for a random subset of the plane (or higher dimensions, analogously) is one of the simplest and most tractable models in stochastic geometry. In mathematics and telecommunications, stochastic geometry models of wireless networks refer to mathematical models based on stochastic geometry that are designed to represent aspects of wireless networks [1]. The related research consisted of analyzing these models with the aim of better understanding wireless communication networks in order to predict and control various network performance metrics. The immature S. damnosum s.l. habitats endemic transmission-oriented eco-epidemiological risk model required using techniques from stochastic geometry and related fields including point processes, spatial statistics, geometric probability, as well as methods from more general mathematical disciplines such as geometry, probability theory, stochastic processes, queuing theory, information theory, and Fourier analysis. Commonly, in seasonal vector arthropod-related field and/or remote endemic transmission-oriented eco-epidemiological risk modeling, stochastic geometry is employed to study of random spatial patterns in georeferenced larval habitat whereby the dependent variable is the immature sampled total density count data [9].

The explanatory linear and non-linear geo-spatiotemporal endmember identifiers of the S. damnosum s.l. riverine larval habitat spectral signatures were then used to predict georeferenced aquatic larval habitats along unsurveyed rivers in both Togo and Uganda. To assess the time series predictive S. damnosum s.l. habitats spectral endmember eco-epidemiological risk model's ability to forecast georeferencable prolific riverine larval habitat sites that may have become temporarily active under varying flow or flooding conditions, a second geographical model was developed. The strategic approach taken was to overlay a Digital Elevation Model (DEM) with signals characteristic of Precambrian rock plus white water, or Precambrian rock alone. The Voltaian system consists of Precambrian to Paleozoic sandstones along with joints, fractures, and quartz veins which had formed earlier in the rocks which has ecological associations with onchocerciasis-related Simulium vectors [2]. Digital elevation model is a digital model or 3D representation of a terrain's surface—commonly for a planet (including Earth), moon, or asteroid—created from terrain elevation data. In order to accomplish this, the authors employed PCI Geomatics software (PCI Geomatics, Toronto, Canada), which supported an automatic overlay of an interpolated wet and dry Precambrian rock signature along the river course. This analysis revealed the geographic locations of both active habitats (i.e. those with water flowing over Precambrian rock) as well as georeferencable sampled sites that might become active under increased flow or flooding conditions. The 3-dimendional (d), DEM is a simple tool to locate differences in elevation that would show areas where such fast flowing water could occur during different river flow conditions [5].

In Jacob et al. [7], of the 30 sites along the Sarakawa River in Northern Togo predicted to be larval habitats by the time series S. damnosum s.l. habitats spectral epidemiological endmember risk model, all (100%) were found to contain larvae. In contrast, none of the 52 sites not predicted by the model, but deemed to be potential habitat by the entomologist accompanying the verification team contained S. damnosum s.l. larvae. Together, these data suggested that the model exhibited a sensitivity and specificity approaching 100% for the prediction of S. damnosum s.l. riverine larval sites in Togo. A second model was then developed to predict such seasonally active sites. There was a complete correspondence between the sites forecasted by the first and the second model, suggesting that the model had identified all active and potentially active S. damnosum breeding sites in the savanna epidemiological riverine ecosystem epidemiological study site. To test the generality of the model, it was applied to predict S. damnosum s.l. riverine sites in northern Uganda. A total of 25 potential S. damnosum s.l. larval habitat sites were predicted. Of the 25 sites predicted to be suitable S. damnosum s.l. larval habitats by the model, 23 [92%; 95% (Cl) 81% - 100%] were found to contain S. damnosum s.l. larvae. In contrast, just 2/10 (20%; 95% CI 0% -45%) sites examined which were not predicted to represent S. damnosum s.l. aquatic habitats by the model were found to contain larvae. The spectral endmember model thus exhibited a sensitivity of 80% and a specificity of 92% when applied in Uganda, a performance that was statistically significant (p < 0.0001; Fisher's Exact test). The two sites that were not predicted by the model which nonetheless were found to contain larvae consisted of low hanging streamside vegetation immersed in fast flowing water. Simulium sibling species identified cytotaxonomically from onchocerciasis focuses consistently harbor much higher densities of larvae than the other, with three annual peaks of abundance: during the dry season and at the beginning and end of the rainy season with larvae most abundant on submerged rocks and fallen leaves, in small shallow areas characterized by high water current, high conductivity and sparse terrestrial vegetation cover [2]. The mean number of larvae found at the sites predicted by the endemic transmission oriented spectral spatiotemporal endmember risk model (21.91) was significantly greater that the mean number of larvae at the sites consisting of immersed overhanging vegetation (4.0; p < 0.001, Mann Whitney U test). Statistical methods may be used to develop predictive time series explanatory georeferencable epidemiological endmember equations in a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment, and these equations may then applied to the unsampled areas (e.g., georeferenced vector arthropod-related larval habitats) to generate a robust endmember predictive map.

Regression is a multivariate analysis that is available in ArcGrid Workstation which may be embedded in a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment for parsimoniously quantitating seasonally vector mosquito-related georeferenced spectral explanatory endemic transmission-oriented data attributes (e.g., QuickBird visible and NIR-based 0.61m resolution reference signature). The central role of the Poisson distribution, for instance, with respect to the analysis of count-related seasonal vector mosquito endemic transmission oriented georeferencable explanatory data is analogous to the position of the normal distribution in the context of models for continuous data. Accordingly, when time series field and/or remote sampled geospatial explanatory endemic transmission-oriented data comprise immature counts, especially for rare events, the probability seasonal real-time bidirectional vector mosquito-related aquatic larval habitat georeferencable explanatory risk model may be based upon an auto-Poisson specification, which may be written in the form of approximations in a PDA-GIS-GPS-RS cyber-environment. The auto-log-Gaussian approximation in a robust seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment could then circumvent the auto-Poisson's intractable normalizing factor, while the auto-logistic approximation would circumvent the auto-Poisson's restriction to only situations involving negative spatial autocorrelation in the sampled field and/or remote sampled endemic transmission oriented mosquito vector data. Commonly georeferencable vector mosquito data exhibit positive spatial autocorrelation [2]. The intractable normalizing constant can then be resolved using Markov Chain Monte Carlo (MCMC) procedures seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment.

In statistics, Markov chain Monte Carlo (MCMC) methods (which include random walk Monte Carlo methods) are a class of algorithms for sampling from probability distributions based on constructing a Markov chain that has the desired distribution as its equilibrium distribution [1]. Markov chain (discrete-time Markov chain or DTMC) is a mathematical system that undergoes transitions from one state to another on a state space whereby a random process is characterized as memoryless: the next state depends only on the current state and not on the sequence of events that preceded [2]. The negative autocorrelation restriction can then be resolved through Windsorizing sampled georeferenced vector mosquito related field and remote endemic transmission oriented data in a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment. Winsorized mean is a measure of central tendency, much like the mean and median, and even more similar to the truncated mean which involves the calculation of the mean after replacing given parts of a probability distribution or sample at the high and low end with the most extreme remaining values [3]. The Winsorized mean in a seasonal vector mosquito-related endemic transmission-oriented eco-epidemiological risk model uses more information from the distribution or sample than the median [2]. However, unless the underlying distribution is symmetric, the Winsorized mean of a sample in a robust PDA-GIS-GPS-RS cyber-environment is unlikely to produce an unbiased sampled vector mosquito-related parameter estimator for either the mean or the median.

Furthermore, since the process of interpolated seasonal arthropod-related image data interpretations often involves use of heuristics and/or "rules" based on expert entomological knowledge and observation (e.g., rules about biogeographic relationships between vegetation zonation, elevation, slope and other spatiotemporal ecogeomorphological explanatory terrain-related georeferenced vector mosquito-related aquatic larval habitat covariates) regression residuals rendered from probabilistic interpolators [e.g., Inverse Distance Weighted (IDW) Matrices] in seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environments, may render optimal unbiased estimators.

In the mathematical field of numerical analysis, stochastic/deterministic interpolation is a method of constructing new data points within the range of a discrete set of known data points [2]. Within Geostatistical Analyst in a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment, a vector biologist, epidemiologist or other research collaborator could easily create a continuous surface risk-related endemic transmissionoriented explanatory map from measurements stored in a point feature layer or raster layer or by using georeferenced polygon centroids [5]. The sample larval habitat points in the endemic transmission oriented explanatory georeferenced field and/or remote specified risk model may then render seasonal measurements such as elevation, depth to the water table, or levels of pollution which also may be kriged. The Interpolation tools are generally divided into deterministic and geostatistical methods. IDW, Spline, and Trend are deterministic, while Kriging is a geostatistical method (www.esri.com). Topo to Raster and Topo to Raster by File in a robust explanatory seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment would then employ an interpolation algorithm designed for creating continuous georeferencable larval habitat surfaces from contour lines [2]. These methods may contain properties favorable for creating surfaces for conducting robust eco-geohydrologic analysis of seasonal-sampled vector mosquito-related endemic transmission oriented field and/or remote specified explanatory covariate coefficients.

When used in conjunction with ArcMap and Geostatistical Analyst, a comprehensive set of PDA-GIS-GPS-RS cyber-tools may provide explanatory, georeferencable, interpolatable, larval habitat surfaces which may be employed to visualize, analyze, and understand spatial phenomena (e.g., prolific larval habitat development). This would include generating watershed and streams (e.g., ecohydrologic analysis), generating slope coefficients, delineating drainage networks, creating contours, working with low-pass and high-pass filters and analyzing the result associated within a geophysical urbanized environment. Further, drainage patterns (e.g., a dendritic pattern) associated with physical environments (landscape and geology map) be cartographically displayed. These models may be able to describe regions underlain by homogeneous material (e.g., flooded soils) that may help target georeferenced vector aquatic larval habitats. That is, the subsurface geology associated with georeferencable explanatory vector mosquito aquatic larval habitats may have a similar resistance to weathering so there no apparent control over the direction the tributaries take.

Surface interpolation functions create a continuous (or prediction) surface from sampled point values [5]. The continuous seasonal georeferenced vector mosquito larval habitat surface representation of an explanatory raster dataset can represent height, concentration, or magnitude—for instance, elevation of a sampled productive habitat [2]. In mathematical predictive time series risk modeling vector mosquito-related endemic transmission oriented explanatory field and/or remote specified data, bicubic interpolation would extend cubic interpolation for interpolating data points (sampled aquatic larval habitats) on a two dimensional regular grid [2]. The interpolated surface is smoother than corresponding surfaces obtained by bilinear interpolation or nearest-neighbor interpolation [1].

Bicubic interpolation of georeferenciable operationzable explanatory time series mosquito data can be accomplished using either Lagrange polynomials, cubic splines, or cubic convolution algorithm in a real-time bidirectional PDA-GIS-GPS-RS cyber-environment. Given a set of k + 1 vector mosquito-related data points where no

two x_j are the same, the interpolation polynomial in the Lagrange form is a linear combination $L(x) = \sum_{j=0}^{n} y_j \ell_j(x)$

of Lagrange basis polynomial

$$\ell_{j}(x) := \prod_{\substack{0 \le m \le k \\ m \ne j}} \frac{x - x_{m}}{x_{j} - x_{m}} = \frac{(x - x_{0})}{(x_{j} - x_{0})} \cdots \frac{(x - x_{j-1})}{(x_{j} - x_{j-1})} \frac{(x - x_{j+1})}{(x_{j} - x_{j+1})} \cdots \frac{(x - x_{k})}{(x_{j} - x_{k})},$$

where $0 \le j \le k$. In image processing, bicubic interpolation is often chosen over bilinear interpolation or nearest neighbor in image resampling, when speed is not an issue. In contrast to bilinear interpolation, which only takes 4 pixels (2 × 2) into account, bicubic interpolation considers 16 pixels (4 × 4). Images resampled with bicubic interpolation are smoother and have fewer interpolation artifacts. Surface interpolation functions make geographic time series predictions from sample measurements for all geographic locations in a raster dataset whether or not a measurement has been taken at the location (www.esri.com). Each model would produce predictions using different calculations. Regardless, robust seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environments may help improve interpolated interpretations from seasonal vector mosquito-related regressed data and georeferenced larval habitat digital images. These include image stratification, classification modification, postclassification sorting, and advanced methods for multisource data analysis.

In this study, QuickBird imagery from Digital Globe [7] was employed to display an empirical dataset of explanatory seasonal real-time bidirectional time series ecological sampled georeferenced endemic transmission oriented field and remote specified data feature attributes associated to multiple georeferenced aquatic larval aquatic habitats of *An. arabiensis*. The research was conducted in a riceland agro-village complex in the Mwea Rice Scheme, Kenya. The malarial vectorial system in Africa is complex, comprising typically of *Anopheles gambiae*, *An. arabiensis*, and *An. funestus* as the primary vectors and a number of complementally vectors including *An. pharoensis*, *An. coustani* and *An. rivuro*. *An. arabiensis* is considered a species of dry, savannah environments and sparse woodland yet it is known to occur in deforested areas (e.g., irrigated rice fields), but only where there is a history of recent land disturbance or clearance [10].

An. arabiensis aquatic larval aquatic habitats are generally small, temporary, sunlit, clear and shallow fresh water pools, although the vector mosquito species is able to utilize a variety of habitats including slow flowing, partially shaded streams and a variety of large and small natural and man-made habitats [8]. Immature *An. arabiensis* have been found in turbid waters and, on occasion, in brackish habitats. It readily makes use of irrigated rice fields where larval densities are related to the height of the rice, peaking when the plants are still relatively short and then dropping off substantially as the rice plants mature [2]. Such density fluctuations are also reflected in the adult population, which also peak when rice stalks are small and decline as the plants mature [8]. These patterns may be due to a preference for sunlit areas of water with relatively limited emergent vegetation, with densities decreasing as shade from the growing plants increases. There is evidence that *An. arabiensis* may be attracted by the application occurs at the start of plant cultivation, and dissolved oxygen content is related to sunlight exposure (e.g. via increasing photosynthesis), so which factor is the primary oviposition attractant in rice fields is uncertain for anopheline immatures. *An. arabiensis* is described as a zoophilic, exophagic and exophilic species [10]. Importantly, *An. arabiensis* have a wide range of feeding and resting patterns, depending on geographical location of the aquatic larval habitats.

Georeferenced entomological surveys in riceland agroecosystems are often biased by locations with known high vector densities using sub-meter resolution technologies [2]. For remote identification of georeferenced vector mosquito aquatic larval habitats, the first step is often to construct a discrete tessellation of the region [5]. Results of previous research have demonstrated time series digitized grid cell data to be superior to orthogonal grid cell data for identification of riceland *Anopheles* aquatic habitats [9]. Building a time series explanatory GIS digitized non-orthogonal gridded geodatabase can identify relationships between riceland mosquito aquatic habitats, rice plant stage of development and agro-village-complex within the framework of georeferenced relational database technology. For example, Jacob *et al.* [2] constructed multiple gridded seasonal QuickBird visible and NIR classified land cover maps for optimizing grid area surveys in an agro-village complex. The classified land cover maps were generated from mosaicked 2012 and 2013 multispectral QuickBird imagery that was employed for robustly quantitating remotely sensed change detection LULC data using the back propagation artificial neural network (ANN) in IDRISI.

In computer science and related fields, ANN are computations that are capable of machine learning and pattern recognition which are usually presented as systems of interconnected "neurons" that can compute values from inputs by feeding information through the network [5]. ANN consists of artificial neurons inspired from biological neurons. Artificial neurons are typically organized in layers, so that each ANN includes an input layer. This layer can have as many neurons as the ANN's input variants in robust seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environments. Input layer neurons are then connected to the neurons of the hidden layers or to neurons in the next layer. Hidden layers may also exist in ANN in a PDA-GIS-GPS-RS cyber-environments. Each hidden layer can have neurons linked in different ways to the other hidden layers or to the output layer. Hidden layer. Finally, an output layer may also be generated through which the output vector passes PDA-GIS-GPS-RS cyber-environments. This layer has as many neurons as the ANN's output variants. Output layer neurons can get their input through the input layers or the hidden layers [5].

IDRISI is the premier system for integrated neural network and machine learning solutions as it employs an advanced Multi-Layer Perceptron (MLP) neural network classifier that offers a variety of innovations (<u>www.esri.com</u>). Multilayer perceptron (MLP) can be feed-forward ANN models that may risk map empirical datasets of input seasonal-sampled *An. arabiensis* related field and remote specified data geo-spatiotemporally sampled in riceland environments onto a set of appropriate outputs in robust PDA-GIS-GPS-RS cyber-environments. Key features of IDRISI seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environments for risk modeling explanatory *An. arabiensis* related georeferenced endemic transmission oriented field and remote specified data feature attributes spatiotemporally sampled in riceland environments include: 1) a complete GIS analysis package for basic and advanced spatial analysis, including tools for surface and statistical analysis, 2) decision support, 3) land change and prediction, and image time series analysis, 4) a complete Image Processing system with extensive hard and soft classifiers, including machine learning classifiers such as neural networks and classification tree analysis, as well as image segmentation for classification, 5) integrated modeling environments including the Earth Trends Modeler for image time series quantitation of environmental trends and Land Change Modeler for LULC analysis and prediction [5]. IDRISI is an integrated GIS and remote sensing software which relies on a PC grid-based system that offers tools for researchers and scientists engaged in analyzing earth system dynamics for effective and responsible decision making for environmental management, sustainable resource development and equitable resource allocation (www.esri.com).

ANN algorithms have been employed for vector mosquito-related endemic transmission-oriented risk modeling. In Jacob et al. [2] training sites for "water", "grass/bare soil", "built", and "paved" classes were developed using polygons digitized from visual interpretation of the 0.61 m QuickBird panchromatic band. Thereafter a robust ANN algorithm produced a geo-spatiotemporal explanatory LULC classification with an overall accuracy of 81% and Kappa of 0.74, which was more accurate than those produced by other classification algorithms evaluated, such as maximum likelihood. The maximum likelihood classification tool considered both the variances and covariances of the class signatures for assigning each cell to one of the classes represented in a signature file with the assumption that the distribution of a class sample was normal. The "built" class has 24% errors of omission and 20% errors of commission, while the "bare soil" class had 7% errors of omission and 10% errors of commission. The kappa measure of agreement is the ratio K = P(A) - P(E)/P(E) where P(A) is the proportion of times the k raters agree, and P(E) is the proportion of times the k raters are expected to agree by chance alone [5]. Complete agreement in time series calculated kappa coefficients from the seasonally regressed time series explanatory malaria mosquito-related field and remote specified endemic transmission oriented risk model parameter estimators corresponded to K = 1, and lack of agreement (e.g., purely random coincidences of prevalence rates) corresponded to K = 0. Negative values of kappa in a seasonal explanatory seasonal real-time bidirectional predictive malaria-related mosquito risk model would mean negative agreement—that is, the propensity of raters to avoid assignments made by other raters [2].

In this research we tested the Trimble Recon[®] X-Series handheld computer as part of a seasonal real-time bidirectional PDA-GIS-GPS-RS cyber-environment for robustly and parsimoniously capturing analyzing and archiving multiple georeferenced *An. arabiensis* larval aquatic habitats and their geo-spatiotemporally sampled explanatory covariates in Karima rice-village agro-complex in the Mwea Rice Scheme in Kenya. Our Trimble Recon 400× outdoor rugged computer had a 400 MHZ processor 64 RAM/256 flash memory 4000 MAH battery bluetooth 802.11 yellow sunlight readable VGA display PD. Recon runs the Windows Mobile 6 operating system, ensuring broad support for software offered by Solution Providers world-wide (http://www.trimble.com/). Based on state-of-the-art hardware and software (200 and 400 MHz XScale processors, reflective color display, Windows Mobile), the Recon uses an innovative modular design with a Power Boot module that integrates the battery and ports into a replaceable component (http://ruggedpcreview.com/3 handhelds tds recon.html). The difference between the 200 and 400 models is in processor speed and amount of non-volatile NAND storage (http://www.trimble.com/). An important goal of NAND flash development has been to reduce the cost per bit and increase maximum chip capacity so that flash memory can compete with magnetic storage devices like hard disks. NAND flash has found a market in devices to which large files are frequently uploaded and replaced. MP3 players, digital cameras and USB drives use NAND flash (http://whatis.techtarget.com/definition/NAND-flash-memory).

The Recon meets rigorous MIL-STD-810F military standards for impact, vibration, humidity, altitude and extreme temperature. The X-series also comes with an IP67 rating. Timely sampling and archiving of seasonal explanatory predictive field and/or remote sampled endemic transmission oriented malarial data is vital for robust predictive risk modeling [2]. Further, the new X-Series gives users (e.g., vector biologists, epidemiologists) the option of integrated Bluetooth and 802.11 g wireless capabilities. The system presently employs Windows Mobile 5.0. Other new features included a built-in microphone, a redesigned keypad and an improved color touch screen. Recently, TDS announced a slightly more powerful battery (4000 vs. 3800 mAH) and the switch to a fully sunlight viewable LCD display. Power sources are vital for proper maintenance of PDA-GIS-GPS-RS cyber-environments [5].

QuickBird visible and near-infra red (NIR) imagery combined and displayed within a PDA-GIS-GPS-RS cy-

ber-environment employing a Trimble Recon X 400MHz Intel PXA255 Xscale CPU[®] may help visually determine the geographic location of field technicians in relation to the seasonal vector mosquito-related aquatic larval habitats being treated. Incorporating an effective results-based digital management system, can become a key component for ongoing seasonal malaria treatment and monitoring programs [4]. Riceland data acquisitions, within a GPS-equipped PDA employing wireless communications may optimize a geodatabase with a related Web application for optimal *An. arabiensis* mosquito data analyses for developing an IVM. As such, in this research, a Trimble Recon[®] X PDA (400 MHz Intel PXA255 Xscale CPU), incorporating digitized rice paddy polygons and QuickBird satellite within a PDA-GIS-GPS-RS cyber-environment was tested for use in identifying treated and untreated seasonal *An. arabiensis* aquatic habitats in Karima rice-village complex, Mwea Rice Scheme, Kenya. Although the focus in this research is on *An. arabiensis* in riceland agroecosystems, the PDA-GIS-GPS-RS cyber-environment architecture may be applicable to aggregate, store and analyze geo-spatiotemporal sampled field and/or remote specified endemic transmission-oriented explanatory covariate coefficients associated to vector and nuisance arthropos-related (e.g., immature *S. damnosum s.l.* black fly) aquatic larval habitats.

2. Material and Methodology

2.1. Study Site

The study was conducted 112 km northeast of Nairobi, Kenya in Karima rice-village complex within the Mwea Rice Scheme. Mwea occupies the lower altitude zone of the Kirinyaga District, in an expansive low-lying, formally wet savannah ecosystem. The scheme is situated on the foot hills of Mount Kenya at 37°20'E and 0°, 41'S. The Mwea Rice Scheme is located in the west central region of Mwea Division and covers an area of approximately 13,640 hectares. More than 50% of the scheme area is used for rice cultivation. The remaining area is used for subsistence farming, grazing, and community activities. The mean annual precipitation is 950 mm, with maximum rainfall occurring in April-May and October-November. The average temperatures range from 16°C to 26.5°C. Relative humidity varies from 52% to 67%. According to the 2010 Kenyan national census, the Mwea Rice Scheme has a population of 170,000, occupying 26,000 households. The Karima study site is located at the central-west region of the scheme and has 168 homesteads, with approximately 1250 residents.

In Mwea, the beginning of each cropping or growing cycle is scheduled according to the water availability through the irrigation water distribution scheme. The schedule of individual farmers' rice planting also differs within this time when water is available. Most fields are cultivated once a year, although some farmers cultivate a second crop. The typical cultivation cycle includes a sowing-transplanting period (June-August), a growing period (August-November) and a post-harvest period (November-December). The second crop is cultivated prior to the short rainy period between January and May. The duration of the rice cycle varies between 120 and 150 days, depending on the rice variety planted. After harvesting, active mosquito habitats may persist in shallow puddles left after harvest [8].

The start of the rice growing cycle begins with paddy preparation by pre-flooding the paddy with about 12 cm of water. Farmers sow rice seeds into a small section (nursery) of their paddy to obtain single stemmed plants for transplanting throughout the paddy. For transplanting, the number of seedling plants is 3 - 4 seedlings per hill with a planting density is 20 - 30 hills per m². Fertilizers in the form of Ammonium sulphate and triple super phosphate are applied at a rate of about 50 kg/ha and 125 kg/ha, respectively, about five days before transplanting. Studies by Muturi *et al.* [8] showed that dose amount of ammonium sulfate accounted for up to 40% mortality rate and one week delay in development time to *An. arabiensis* larvae. The actual amount of water used by farmers for land preparation, and during the crop growth period, is much higher than the actual field requirement. Paddy farmers in Mwea often store water in their fields as a back-up safety measure due to the unreliability in supply of water for irrigation [2]. This leads to a high amount of surface runoff, seepage and percolation, accounting for about 50 - 80 percent of the total water input to the field creating active mosquito aquatic larval habitats.

2.2. Aquatic Habitat Sampling

Base maps were generated in ArcGIS of the study site (Figure 1). Field sampling was conducted from July 2005 to July 2007. One hundred and fifty two temporary, permanent and semi-permanent *An. arabiensis* aquatic larval habitats in the Karima epidemiological study site were mapped and classified using a CSI-DGPS Max diffe-



Figure 1. QuickBird visible and near-infra-red data with ecologically sampled *An. arabiensis* aquatic larval habitats at the Karima study site.

rentially corrected global positioning systems (DGPS) receiver and OmniSTAR L-Band satellite signal. Within the DGPS MAX, CSI Wireless has integrated the CSI Wireless SLX receiver, a tri-purpose GPS/WAAS/L-band receiver, and the CSI Wireless SBX, a high performance DGPS beacon receiver for optimal positional accuracy (http://www.omnistar.com/Portals/). We employed three internal differential correction services. Binary messages may be output from the DGPS MAX receiver, the Mini MAX receiver, the PowerMAX receiver, the Vector, the VectorPRO, the Vector Sensor, the Vector Sensor PRO and the Vector OEM along with NMEA 0183 data (http://www.gpsdgps.com/download/m_program.pdf). In this research, the Binary messages of the sampled *An. arabiensis* aquatic larval habitats were supported by the receivers which were in an Intel Little Endian format for direct read in the PDA-GIS-GPS-RS cyber-environment architecture. Each binary message begins with an 8-byte header and ended with a carriage-return line-feed pair (0x0D, 0x0A). The first four characters of the header was the ASCII sequence BIN.

The beacon receiver successfully obtained DGPS beacon signals for the Karima agro-village complex epidemiologist study site. The Wide Area Augmentation System (WAAS) demodulator decoded the corrected georeferenced *An. arabiensis* aquatic larval habitat data from the WAAS, as the L-band satellite differential receiver obtained the corrections from the Omni STAR Worldwide DGPS service. The WAAS is an air navigation aid developed by the Federal Aviation Administration (prime contractor Raytheon Company) to augment the Global Positioning System (GPS), with the goal of improving its accuracy, integrity, and availability (http://www8.garmin.com/aboutGPS/waas.html). The WAAS is an air navigation aid developed by the Federal Aviation Administration (prime contractor Raytheon Company) to augment the Global Positioning System (GPS), with the goal of improving its accuracy, integrity, and availability. Essentially, WAAS is intended to enable aircraft to rely on GPS for all phases of flight, including precision approaches to any airport within its coverage area. WAAS uses a network of ground-based reference stations, in North America and Hawaii, to measure small variations in the GPS satellites' signals in the western hemisphere. Measurements from the reference stations are routed to master stations, which queue the received Deviation Correction (DC) and send the correction messages to geostationary WAAS satellites in a timely manner (every 5 seconds or better).

In addition to real-time DGPS, the DGPS MAX also supports post-processing of malaria-related data [2]. A vector biologist, epidemiologist and/or research collaborators may configure the DGPS MAX for output of binary measurement vector mosquito geo-spatiotemporal sampled field and remote specified endemic transmission oriented data for logging with the use of an external device. By so doing, a conversion utility would be available from CSI Wireless for translation from the proprietary binary format into the Receiver Independent Exchange format (RINEX). Receiver Independent Exchange Format is a data interchange format for raw satellite navigation system data which allows the user to post-process the received data to produce a more accurate result—usually with other data unknown to the original receiver, such as better models of the atmospheric conditions at time of measurement [5]. RINEX version enables storage of our sampled seasonal field and remote specified measure-

ments from pseudorange, carrier-phase and Doppler systems for to GPS (including modernization signals (e.g. L5 and L2C), GLONASS, Galileo, Beidou, along with data from EGNOS and WAASsatellite based augmentation systems (SBAS), OZSS, simultaneously (ftp://igscb.jpl.nasa.gov/pub/data/format/rinex210.txt). The final output of a navigation receiver is usually its position, speed or other related physical quantities. The base station then transmits the range error corrections to remote receivers in real-time. The remote receiver corrects its satellite range measurements using these differential corrections, yielding a much more accurate position (http://www.omnistar.com/Portals/). This was the predominant DGPS strategy used for our real-time An. arabiensis aquatic larval habitat-related field sampled applications. The OmniSTAR L-Band signal is a wide-area DGPS service that uses data from a worldwide network of reference stations, which are combined with precise orbit and clock information for every satellite in the NAVSTAR GPS satellite constellation. The OmniSTAR services, OmniSTAR-VBS (Virtual Base Station), OmniSTAR-HP (High Performance), OmniSTAR-XP (Extended Performance) and OmniSTAR-G2 (GPS and GLONASS), were specifically developed to satisfy the need for high accuracy positioning systems and services for land-based applications (http://www.omnistar.com/Portals/). The addition of GLONASS to the solution significantly increases the number of satellites available which is useful when faced with conditions that limit satellite visibility, such as terrain, vegetation or buildings. OmniSTAR G2 service provides short-term accuracy of 1 - 2 inches and long term repeatability of better than 10 centimerers, 95% CEP. OmniSTAR-HP/XP/G2 is a dual frequency premium augmentation solution in the OmniSTAR family of High Performance services. OmniSTAR-HP/XP/G2 delivers centimeter accuracy derived by a number of independent positioning technologies: OmniSTAR's primary network of dual frequency reference stations provide for measurements that eliminate errors caused by signal delays in the ionosphere. The positional accuracy of the CSI DGPS Max receiver, using an OmniSTAR L-Band satellite signal, was 0.179 (+/-0.392) [9]. Jacob et al. [2] described the purpose of DGPS for mosquito aquatic larval habitat mapping as providing global absolute positioning capability with respect to a consistent terrestrial reference frame

The Karima rice-village complex epidemiological study site was sampled for immature *Anopheles*, along various cross-sectional transects of potential habitats, using standard collecting techniques [11]. Areas with high and low larval productivity were categorized by habitat type; paddy, irrigation or drainage canal, marshland, riparian, ephemeral pool, seep, or other natural or man-made aquatic sites. Specific physical properties of these habitats were recorded. Temporary aquatic habitats were defined as habitats with water for 2 weeks, semi-permanent were habitats with water for 3 months and permanent habitats had water for 6 months or longer.

Water bodies were inspected for mosquito larvae using standard dipping techniques with a 350-ml dipper to collect the mosquito larvae [11]. The number of dips per habitat was 20. All data from the habitat characterization of each aquatic larval habitat was recorded on a field sampling form. Larvae and a sample of water from each larval habitat were placed in whirl-pack bags and transported to the Mwea Research Station for further processing. All 3rd and 4th in star larvae were immediately preserved in 95% ethanol and later identified morphologically to species using taxonomic keys [12]. The 1st and 2nd instars were reared in plastic pans under semi-field conditions and those that survived to 3rd in star were also preserved and identified morphologically to species. The pupae were kept in mosquito emergent cages [13] and the resultant emergent mosquitoes were identified morphologically. Anopheline larvae were separated from Culicine larvae for microscopic identification. Culicine larvae were not identified to genus and species levels as the focus of this study was on anophelines.

2.3. Aquatic Habitat Characterization

Meteorological data was acquired from a Davis Instruments 6153-IP Ethernet Wireless Weather Station placed in a secured homestead within the study site. Internet Protocol or IP Network Weather Stations update weather servers real-time, providing up to the second data without the use of a PC [5]. The VantagePro2 console and wireless receiver provided forecasting, on-screen graphing, and much more (http://www.ambientweather.com/dain61ipvafa.html). Quick view icons in the station revealed daily forecast (e.g., sunny, partly sunny, cloudy or, rain) at the Karima epidemiological rice agro-village complex while a moving ticker-tape displayed gave more specified meteorological variables (e.g., solar radiation). The integrated sensor suite combined a rain collector, temperature and humidity sensors, and anemometer into one package—making the setup for improving performance and reliability. For optimizing the weather monitoring at the study site, we added Weather Link to our Vantage Pro. The Weather Link data logger fit neatly into the Vantage Pro console storing our weather data even when the computer was off. The Weather Link software created graphs and generated summaries of meteorological variables at the Karima epidemiological rice agro-village complex. The Davis VantagePro2 employed frequency hopping

spread spectrum radio to transmit and receive data up to 1000' (300 m) line of sight. The fan aspirated radiation shield improved accuracy of the seasonal meteorological georeferenced explanatory endemic transmissionoriented covariate coefficients at the Karima epidemiological study site by circulating fresh air into the temperature and humidity shield.

Environmental variables recorded for each habitat were number of aquatic animals, depth, and distance to the nearest house, canopy coverage, shade, and turbidity. Distance to the nearest house was measured with a tape when it was shorter than 100 m. When the distance exceeded 100 m, it was estimated visually. The distance to the nearest house was categorized into 7 classes (e.g., 1: 50 - 100 m, 2: 101 - 200 m, and so on, and 7 for distances greater than 600 m). Canopy cover was defined as the amount of terrestrial vegetation and other objects in the habitat. The non-mosquito aquatic invertebrates collected in the dipper samples during the sampling occasion were preserved in 100% ethanol and later identified to family using taxonomic keys of [14]. The number of individuals of each family identified were counted and recorded. Shade coverage of a habitat was measured in percentage of water surface covered by placing a square frame (1 m²) with grids (100 cm²) above the habitat. Turbidity was measured by placing water samples in glass test tubes and holding against a white background, and classified into 2 levels: clear and turbid.

2.4. Remote Sensing Data

Raster image data from the Digital Globe QuickBird satellite service were acquired for the periods of: 15 July 2012 February 20, and 15 July 2013 within the agro-village complex epidemiological study site area. The QuickBird image data were delivered as pan-sharpened composite products in infra-red (IR) colors. The QuickBird raster image products provided four discrete non-overlapping spectral bands, in the 0.45 μ m to 0.72 μ m range, with an 11-bit collected information depth. The clearest, cloud-free images available of the contiguous sub-areas of the epidemiological study site were used to identify land cover and other spatial features associated with *An. arabiensis* aquatic larval habitats. The Order Polygon contained 5 vertices consisting of longitude/latitude (decimal degrees) geographic coordinates using a WGS84 ellipsoid. The satellite data contained 64 km² of the land cover in the study site.

The QuickBird imagery was classified using the Iterative Self-Organizing Data Analysis Technique (ISODATA) unsupervised routine in ERDAS Imagine v.8.7TM [15]. Unsupervised classifications are commonly used for the identification of land covers and mosquito habitats associated with intermediate hosts and disease vectors [16]-[18]. A total number of 15 Ground Control Points (GCPs) within the Karima study site were measured. A coordinate accuracy of 0.2 - 0.4 cm in the horizontal, and 0.6 - 0.8 cm in the vertical, was obtained. Twelve of the collected GCPs were applied as part of the image rectification process; while the remaining three were used as reference coordinates for the evaluation of QuickBird data.

Non-parametric approaches were used to build a correlation matrix between QuickBird pixels in the Karima agro-village study site image and its ground locations, based on two cubic polynomials algorithms. A univariate cubic polynomial has the form $f(x) = a_3x^3 + a_2x^2 + a_1x + a_3$ [2]. An equation involving a cubic polynomial is called a cubic equation [5]. A closed-form solution known as the cubic formula exists for the malaria-related solutions of an arbitrary cubic equation [0]. A cubic polynomial is a polynomial of degree 3 [5]. In this research we derived a univariate cubic polynomial from the empirical sampled dataset of remotely sensed spatiotemporal *An arabiensis* related endemic transmission oriented explanatory covariate coefficients employing the form $f(x) = a_3x^3 + a_2x^2 + a_1x + a_3$. An equation involving a cubic polynomial is called a cubic equation [5]. A closed-form solution known as the cubic formula also exists for the solutions of arbitrary cubic equations for quantitating uncertainty terms in latent forecasted derivatives from remotely sensed forecasting models. Non-parametric methods were then applied for the rectification of the acquired Standard type imagery using 1) Simple linear polynomial correction with 3-Dimensional (D) ground control point GCPs, 2) rational polynomial coefficients (RPC) correction without refinement; and, 3) RPC correction with linear refinement using 3D GCPs. These non-parametric methods were then applied for the rectification of the satellite imagery using a rational polynomial coefficient correction procedure with linear refinement utilizing the GCPs.

Thereafter, the rational function model (RFM) was used as an alternative to physical sensor models for 3D ground point determination with the high-resolution QuickBird imagery. However, owing to the sensor orientation bias inherent in the vendor-provided RPCs, the geopositioning accuracy obtained from these RPCs was limited. In this paper, the performances of two schemes for orientation bias correction (*i.e.*, RPCs modification and RPCs regeneration) were performed based on one separate-orbit QuickBird stereo image pair in the Karima

rice-village complex, and four cases for bias correction, including shift bias correction, shift and drift bias correction, affine model bias correction and second-order polynomial bias correction. A 2-step least squares adjustment method was then adopted for correction parameter estimation with a comparison with the RPC bundle adjustment method. The experiment results demonstrated that in general the accuracy of the 2-step least squares adjustment method and minimal one GCP, the modified RPCs improved the accuracy from the original 23 m to 3m in planimetry and 17 m to 4 m in height in the Karima agro-village complex epidemiological study site. With the shift bias correction yielded a geopositioning accuracy of better than 0.50 m in planimetry and 1 m in height with 3 well-positioned GCPs. Further, tests with the second-order polynomial bias correction model indicated the existence of potential high-order error signals in the vendor-provided RPCs. Thereafter, based on the condition that an adequate redundancy in GCP number was available, an accuracy of 0.42 m in planimetry and 0.80 m in height of the data was attainable.

Since biases or errors may still have existed after applying the original RPC parameters generated from the QuickBird image further refinement was performed using linear equations an sampled 3-dimensional (D) GCPs. Root Mean Square Error was then calculated for the linear polynomial rectification using the 3-D GCP's sampled in the study site (**Table 1**). The root-mean-square deviation (RMSD) or root-mean-square error (RMSE) is a frequently used measure of the differences between values predicted by a model or an estimator and the values actually observed. Basically, the RMSD represented the sample standard deviation of the differences between predicted values and observed values [5]. In this research, these individual differences were represented as the spatiotemporal vector mosquito larval habitat field and remote sampled endemic transmission oriented regression-based residuals when the calculations are performed over the data sample that was used for estimation, and are called prediction errors when computed out-of-sample. The use of additional 3-D GCPs for the refinement of the model allowed accurate dereferencing of the ecological sampled *An. arabiensis* aquatic habitat data. These types of process ensured that accurate geographic extents of the study site were able to encompass all georeferenced data.

In this research, the RMSD of a seasonal An. arabiensis larval habitat parameter estimator $\hat{\theta}$ with respect to an estimated parameter θ was defined as the square root of the mean square error:

 $RMSD(\hat{\theta}) = \sqrt{MSE(\hat{\theta})} = \sqrt{E((\hat{\theta} - \theta)^2)}$. For an unbiased estimator, the RMSD was then the square root of the

variance (*i.e.*, standard error in the QuickBird *An. arabiensis* aquatic larval habitat regression-based risk model). The RMSD of the predicted time series field and/or remote endemic transmission oriented explanatory larval habitat values \hat{y}_{th} for times *t* of a regression's dependent variabley was computed for *n* different predictions as

the square root of the mean of the squares of the deviations employing $\text{RMSD} = \sqrt{\frac{\sum_{t=1}^{n} (\hat{y}_t - y_t)^2}{n}}$. In some

disciplines, the RMSD is used to compare differences between two things that may vary, neither of which is accepted as the "standard" [5]. For instance, when measuring the average difference between two time series $x_{1,t}$

and
$$x_{2,t}$$
, the formula becomes $\text{RMSD} = \sqrt{\frac{\sum_{t=1}^{n} (x_{1,t} - x_{2,t})^2}{n}}$

 Table 1. Calculated Root Mean Square Error (RMSE) for a linear polynomial rectification using

 3-Dimensional GCP's sampled in the Karima study site.

GCP	XRi	YRi	Ri	Rx	Ry	Т
1	0.451	0.322	0.592	0.512	0.162	0.892
2	0.637	-1.271	1.263			
3	0.333	-0.626	0.721			

i: number of Ground Control Points (GCP), XRi and YRi: the X and Y residual for GCPi, Ri: the RMS error for GCPi, Rx and RY: the X and Y RMS error, T: the total RMS error.

2.5. Habitat Mapping

Each *An. arabiensis* aquatic habitat in the study site, with their corresponding data attributes, was entered into the Vector Control Management System (VCMS[®]) relational database software product [19]. The VCMS provides public health professionals involved in mosquito/vector control and pest management with a comprehensive database tracking and reporting system (<u>http://store.elecdata.com/field_data_collection/vcms.aspx</u>). Detailed information on larval habitat site locations and data collection and trapping activities comprised the core of our data system. VCMS is an international model for public health agencies that standardizes data collection/reporting activities and can aggregate data on various types of vector borne diseases.

In this research the VCMS[®] database supported a mobile field data acquisition component module, called Mobile VCMS that synchronized the seasonal sampled *An. arabiensis* aquatic larval habitats field data from industry standard Microsoft Windows MobileTM devices which also supported our add-on DGPS data collection. Mobile VCMS and its corresponding FieldBridge[®] middleware software component were then used to support both wired and wireless synchronizing of the field data for conducting the larval aquatic habitat monitoring. Field data collected with Mobile VCMS was synchronized directly into a centralized VCMS relational database repository. Additional geocoding and spatial display of the ecological data was handled using the embedded VCMS GIS Interface KitTM that was developed utilizing ESRI's MapObjectsTM 2 technology. The VCMS database supported the export of all field data, using any combination of *An. arabiensis* aquatic habitats sampled at the study site, in order to further process and spatially display specific data attributes in a standalone desktop GIS software package. VCMS[®] can interface with MapInfo GIS files via a MapX programming interface for display and manipulation of spatial data or imagery (www.esri.com). The MapInfo TAB format is a popular geospatial vector data format for geographic information systems software which is developed and regulated by MapInfo Corporation as a proprietary format (MITAB-MapInfo.TAB Read/Write Library).

ESRI ArcPad[®] software was installed on the Recon X[®] and we created an ArcPad folder on the C drive of our field computer (laptop PC) to connect and synchronize data with ArcPad on the PDA. The ArcPad[®] Data Manager for ArcGIS Desktop, the Datum Configuration Tool and the ArcPad[®] Deployment Manager were copied into this folder, along with other ArcPad[®] modules. ArcPad has traditional GIS functionality such as map navigation, layering, querying, hyperlinks, and so forth [5]. The key features of ArcPad include support for industry-standard data formats; display and query functionality; editing and data capture; support for optional GPS receivers; and ArcPad tools for ArcGIS Desktop, which are used to prepare data for use in ArcPad (www.esri.com). A key feature of ArcPad in seasonal mosquito data analyses is the ability to use data directly from an individual's field-level desktop or an organization's enterprise GIS system without the need to convert it to unique portable formats [2]. ArcPad uses vector data in industry-standard shape file format (as used by ArcInfo[®], ArcEditorTM, ArcView[®], ArcIMS[®], and other ESRI software programs. In this research ArcPad directly supported the following raster image formats (world file required): JPEG, PNG, CADRG, MrSID (compressed images), and Windows[®] bit map (BMP) generated from the multivariate time series empirical datasets of *An. arabiensis* related field and remote specified endemic transmission oriented covariate coefficients. This software was used to display and manipulate the QuickBird imagery on the Recon PDA field computer.

2.6. Grid-Based Algorithm

A polygon layer outlining each *An. arabiensis* aquatic habitat was created by digitizing the georeferenced QuickBird imagery in GeoGrid[®]. Our matrix was based on GEO grid framework based on grid technology, remote sensing data, and geographic information that was initially developed at National Applied Research Laboratories (NARL). The approach was initiated by synergy of NARLpsilas core competence on environment monitoring and disaster reduction techniques which included high-resolution satellite image processing, virtual reality visualization, grid computing, and disaster mitigation technology along with the advanced cyber-infrastructure environment established within NARL. The framework was constituted by three layers, *i.e.* application module, service interface, and computing/data/sensor grids. A prototype platform entitled 3D GIS was constituted by using 2 m resolution FORMOSAT-2 data and 5 m resolution digital terrain model that was modeled and then displayed in 3D stereo visualization and in Web pages. The presented approach emphasized the synergy of multiple discipline with cross-field cooperation for geoscience application which then becomes a benchmark in echo

to implementation of GEOSS.

Determining spatial heterogeneity in larval habitat distribution using a digitized grid cell database can have important operational significance because vector control operations can be designed to target zones where high larval densities occur [2]. Treatments or habitat perturbations should be based on surveillance of larvae in the most productive areas of the agro-ecosystem and adjacent village [25]. The impact of larval control using new formulations of insecticides should be rigorously tested using a modified longitudinally designed study employing information from digitized grid cells [7]. Laboratory studies should test *Bacillus thuringiensis ssp. israelensis* (Bti), *Bacillus sphaericus* (Neide) (Bsph) and Bti:Bsph ratios may be determined employing lethal concentration parameters on all highly productive riceland *An. arabiensis* aquatic habitats as identified in a stratified digitized grid matrix [9]. Controlling a small proportion of productive riceland *An. arabiensis* aquatic habitats may yield significant reductions in a rice environment. For larval control, we assume that treatments applied to individual habitats are 100% effective in eliminating all immatures, *i.e.* treated habitats produce zero contribution to the total productivity.

In this research, a cell within the grid matrix contained an attribute value, as well as a georeferenced aquatic larval habitat location coordinate, and was joined relationally to other databases (**Figure 2**). The spatial location of each cell was implicitly contained within the ordering of the matrix. The habitats were then characterized in relation to the ecological attributes sampled of an aquatic habitat. Each *An. arabiensis* aquatic larval habitat/polygon was assigned a unique identifier. Field attribute tables were then linked to the polygons. The polygons were used to define the sampling frame, which extended to include a 2 km buffer from the external boundary of the Karima rice-village study site [20]. A mark-release-recapture study in an area near Bamako, Mali, showed that *An. arabiensis* generally does not disperse further than 1 km [21]. In this research, digitized grid cells were stratified based on LULC transition throughout the rice cycle and defined as: ploughed, flooded, post-transplanted, tillering, and fallow/post-harvest.

1) Ploughing: field preparation prior to transplanting of rice seedlings.

2) Flooding: comprised of areas of intensive use with much of the land covered by water.

3) Post-transplanting: a period following rice seedlings transplanting.

4) Tillering: extends from the appearance of the first tiller until the maximum tiller (5 - 9) number is reached. Stem elongation occurs and the tillers continue to increase in number and height, with increasing ground cover and canopy formation.

5) Fallow/post-harvest: after harvesting, the land is left bare waiting the next crop cycle.



Figure 2. A stratified digitized gridded matrix overlaid onto a QuickBird visible and near infra-red of the spatially targeting quadrants for insecticide treatment at the Karima epidemiological study site.

Error coefficients were calculated for each LULC class and for the overall mapped area. Land use/cover types cannot be effectively extracted when different land cover types bear similar spectral characteristics or the same land cover classification has different spectral responses [22]. Importantly, there are many sources of both conservative and optimistic bias in seasonal mosquito land cover classification accuracy assessment, many of which are impossible to avoid [2]. Bias occurs when a classification estimate is optimistic or conservative [5]. Interestingly, there are at least three significant sources of conservative bias in time series vector arthropod-related image processing: errors in reference data, positional errors, and minimum mapping unit of reference grid [2]. Further, there are at least three significant sources of optimistic bias: use of training data for accuracy assessment, restriction of reference data sampling to homogeneous areas, and sampling of reference data not independent of training data. The magnitude and direction of bias in seasonal mosquito related classification accuracy estimates we assumed was not enough for accurate seasonal malarial mosquito risk mapping the Karima epidemiological study site. The error matrix is practically meaningless unless methods are reported in sufficient detail to enable readers to assess the potential for bias in the classification accuracy estimates [5].

In this research, we evaluated the results using pixel-based contingency matrices. Change maps for identifying seasonal malaria-related larval habitat 'hot spots" generated from satellite data, and change overlays created from traditional stand maps updated with aerial photography may be site-specifically cross-verified using contingency time series error matrices and GIS modeling [2]. It has been demonstrated in GIS literature clearly that a majority of the so-called digital change classification inconsistencies are not errors, but embody instead a powerful source of sub-stand information with the potential to significantly impact the sustainability of the resource management [5]. While our QuickBird pixel approach did not directly encompass the stand concept (stand representing the actual field sampled georeferenced *An. arabiensis* aquatic larval habitat), qualitative and quantitative seasonal malaria-related mosquito larval habitat data on change at the stand level was readily extracted from the satellite imagery [2]. Error matrices were then generated for our larval habitat classifications with the purpose of evaluating the map product for thematic accuracy. The associated error matrix provides the user with values indicating how true the map product is with respect to actual conditions on the ground [23] [24]. Errors of omission and commission can be calculated to update and refine the LULC classification for riceland environments or for estimating confidence on a per-class basis [25].

In this research overall accuracy and class-specific user and producer accuracies were calculated for each of the resultant land cover classes. For each mapping region, time series stratified sampling formulas were applied to estimate the error matrix cell proportions [22], and consequently, the estimates of overall and class-specific user's and producer's accuracy [23]. An overall accuracy and a producer and user accuracies for each LULC class was calculated using these methods. The overall accuracy was calculated by dividing the number of pixels correctly classified (*i.e.*, the sum of the diagonal axis of the matrix) by the total number of pixels included in the evaluation process. The producer accuracy is a measure of the omission error and indicates the percentage of pixels of a given land cover type that are correctly classified [5]. The user accuracy is a measure of the commission errors and indicates the probability that a pixel classified into a given class actually represents that class on the ground [5]. These validation techniques were calculated by dividing the number of pixels of the *i*th class correctly classified by the total number of pixels of the *i*th class.

In this research, the producer's accuracy related to the probability that a land cover class was correctly mapped and measured using the errors of omission (1-producer's accuracy). In contrast, the user's accuracy indicated the probability that a geo-spatiotemporal sampled *An. arabiensis* aquatic larval habitat from the land cover map actually matched the information from the georeferenced empirical datasets by measuring the error of commission (1-user's accuracy). Accuracy results were computed by weighting the proportions of each land cover, within the Karima epidemiological study site, against total land cover area used in the sampling frame. Specifically, the overall accuracy (\hat{P}) and producer's accuracy (\hat{P}_{Aj}) were estimated using post-stratified formulas [22]. Post-stratified estimators use the known pixel totals for each land-cover class (Ni+), treating the sample as a stratified random sample of ni+ pixels from the Ni+ pixels in that class [5]. In this research the user's accuracy (\hat{P}_{Ji}) was based on the simple random sampling formulas:

(1)
$$\hat{P} = \frac{1}{N} \sum_{k=1}^{q} \frac{N_{k+}}{n_{k+}} n_{kk}$$

(2)
$$P_{Ui} = n_{ii}/n_{i+}$$

(3) $\hat{P}_{Aj} = \frac{\left(N_{j+}/n_{j+}\right)n_{jj}}{\sum_{k=1}^{q} \left(N_{k+}/n_{k+}\right)n_{kj}}$

The spatial distribution of the sampled georeferenced aquatic larval habitats were created as a layer and overlaid on the land cover layer, and the number of *An. arabiensis* aquatic habitats in each land cover class was then calculated using Zonal stat, an avenue script in ArcGIS[®] [6]. Jacob *et al.* [9] created a 2 km buffer zones around Karima villages and then uses zonal statistics to estimate the mean of normalized difference vegetation index (NDVI), average precipitation, temperature, elevation and slope within each buffer zone to determine which factors are associated with high mosquito densities and high rates of malaria. In this research, the chi-square analysis for the sampled field and remote predictor variables was used to examine whether there were significant differences in proportions of positive and negative sampled aquatic habitats in each grid cell located in different LULC classes. Tukey style multiple comparison of proportions were used for post-hoc analyses. Each predicted estimate in the LULC classification was ground verified which revealed all classes were within correct statistical limitations. Unless each and every minimum mapping unit, which may be a pixel, polygon, or any other feature with spatial properties, has been verified on the ground, there is still unknown and un-quantified error [5].

2.7. Object Oriented Classification

 $\text{ENVI}^{\text{(B)}}$ spectral tools were used to analyze the QuickBird visible and NIR data, in order to confirm the location of varying states of rice crops. In ENVI 4.6^(B), a spectrum plot, known as a *Z*-profile, of the pixel under the cursor, was run through all bands of the QuickBird image of the *An. arabiensis* aquatic habitats in the Karima agro-village complex epidemiological study site. The basic workflow involved importing the data collected in the field from the study site into a spectral library. The content of the spectral library was then used in the Endmember Collection workflow to perform a supervised classification, based on the empirical ecological covariates sampled at each individual georeferenced aquatic larval habitat. Binary Encoding, Spectral Angle Mapper (SAM), and spectral feature fitting were used to rank and match any unknown spectrum to the materials in the spectral library. SAM is a method to compare known spectral target properties to every pixel in the scene [23]. Land cover class representative pixels were selected and compared to a reference dataset.

After pixels in a raster dataset have been assigned to categories, post-classification methods were used to clean up the resulting raster class image in preparation for conversion to vector data. Raster classification of riceland *An. arabiensis* aquatic habitat results often contain scattered individual pixels of one LULC class surrounded by a larger area of another LULC class (*i.e.*, islands) [25]. ENVI used sieving (*i.e.*, removal of isolated aquatic larval habitat pixels from our QuickBird image) and clumping (clumping adjacent, similar larval habitat pixels together) to minimize database clutter in our GIS analyses.

During the segmentation procedure, image objects from the QuickBird scenes were generated based on color, shape, and texture. Optimum segmentation parameters were generated for extracting different land cover classes associated with the sampled *An. arabiensis* aquatic larval habitats. Image segmentation is the key technology from image processing to image analysis, which divides an image into non-overlapping homogeneous regions and the combined heterogeneous adjacent regions [22]. In this research, image classification was done using FastLine-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASHTM) using a spatial model (GMD file) that converted the image's digital numbers (DN) associated to *An. arabiensis* aquatic habitat surface components to at-sensor radiance and computed at-sensor reflectance while normalizing solar elevation angle.

Interestingly, FLAASH is a first-principles atmospheric correction tool that corrects wavelengths in the visible through near-infrared and shortwave infrared regions, up tµm (http://www.exelisvis.com/docs/FLAASH.html). FLAASH works with most hyperspectral and multispectral sensors. Water vapor and aerosol retrieval are only possible when the image contains bands in appropriate wavelength positions. FLAASH can correct images collected in either vertical (nadir) or slant-viewing geometries (www.esri.com). Previously Yuan *et al.* [22] employed the capability of FLAASH in ENVI software to make atmospheric correction for EO-1 Hyperion hyperspectral image. Hyperion hyperspectral image of Zhangye city in Heihe River valley of Gansu province, China was acquired on September 10, 2007. Canopy spectra, biochemical component and GPS data of 41 plots were measured in near real-time during the satellite overpass. Hyperion hyperspectral image was geometrically cor-

rected using Lansat-7 ETM+ image. Thereafter, the digital numerical coefficients were transformed to radiance and apparent reflectance, and atmospheric correction of Hyperion image was made using FLAASH. The resulting radiance, apparent reflectance and reflectance after FLAASH of four typical objects, including corn, water body, desert and building, were compared. ASD spectra of corn were resampled to Hyperion corresponding bands using Gaussian filter function. The comparison between ASD resampled spectra and Hyperion spectra after FLAASH demonstrated that the atmospheric correction using FLAASH is very effective and these two spectra are consistent with each other and the correlation coefficient reached 0.987. In this research data pre-processing involved converting digital numbers to radiance, employing FLAASH atmospheric corrections and the equation:

$$\rho_{\text{BandN}} = \frac{\pi \left(L_{\text{BandN}} \times Gain_{\text{BandN}} + Bias_{\text{BandN}} \right) \times D^2}{E_{\text{BandN}} \times \left(\text{COS} \left((90 - \theta) \times \pi / 180 \right) \right)}$$

where,

 ρ_{BandN} = Reflectance for Band N; L_{BandN} = Digital Number for Band N; D = Normalized Earth-Sun Distance; E_{BandN} = Solar Irradiance for Band N.

2.8. Regression Analyses

A Poisson regression with a 95% confidence level was then constructed in PROC NL MIXED MOD using the spatiotemporal sampled field and remote specified *An. arabiensis* larval habitat predictors. In actuarial literature, researchers have suggested various Poisson regression model, which is also known as the Generalized Linear Model (GLM) with Poisson error structure [6] We expected the seasonal sampled *An. arabiensis* aquatic habitat larval/pupal count in the Karima study site to follow a Poisson distribution, as was the case in previous research in irrigated riceland areas. Poisson process is a commonly used starting point for modeling stochastic variation of ecological count data around a theoretical expectation [7]. In this research, PROC GENMOD used a class statement for specifying our sampled categorical indicator variables. The parameter $\lambda_i(X_i)$ is both the mean and the variance of the Poisson distribution for a sampled anopheline larval habitat *i* [2]. The dependent variable was the total *An. arabiensis* larval count in the sampled larval habitat.

We generated regression models for comparing transmission intensity at the Karima epidemiological study site. The regression analyses assumed independent counts (*i.e.*, n_i), taken at the sampled larval habitat locations $i = 1, 2, \dots, n$, where each of the explanatory predictor covariates represented in the linear framework was from a Poisson distribution. Our Poisson regression assumed the response variable Y had a Poisson distribution and assumed the logarithm of its expected value was modeled by a linear combination of the sampled larval habitat sampled parameter estimates. This expression was written more compactly as $\log(E(Y|x))$ where x was an n+ 1-dimensional vector consisting of n independent variables concatenated to 1 and θ , which in this research was simply expressed as a linearly linked to b. Thus, in our Poisson regression model θ was an input vector x and, the predicted mean of the associated Poisson distribution was given by $E(Y|x) = a^{\sigma x}$, if $x \in \mathbb{R}^n$ was a vector of the sampled independent variables. Thereafter, the Poisson models took the from $\log(E(Y|x)) = a'x + b$

where $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Positing salient estimators using the maximization of an auto-Gaussian log-likelihood function and a set of eigenvectors where lambda is a subspace of \mathbb{R}^n can identify important explanatory predictor covariates associated to productive habitats based on spatiotemporal field-sampled count data [1]-[5].

After the regression, we attempted to ascertain whether the proportions of the sampled An. arabiensis immature data differed by digitized grid cells in the study site. The model assumed a random sample between Y_i , (*i.e.*, sampled habitat total larval count density data), and the regressors X_{i1}, \dots, X_{ip} . A disturbance term, ε_i , which was a random variable, was added to this assumed relationship to capture the influence of all the parameters estimators sampled on Y_i other than X_{i1}, \dots, X_{ip} . The random error term, ε , in the regression models was assumed to be normally distributed with mean zero and variance σ^2 . Statistical characteristics of the sampled data were examined in PROC UNIVARIATE. The PLOT option in the PROC UNIVARAITE statement generated histograms and boxplots. The NORMAL option was used to test whether the field and remote-sampled explanatory field and remote specified endemic transmission oriented covariate coefficients had a normal distribution. The regression equation for the study sites was: $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_1 X_{2i1}$, $i = 1, 2, \dots, n$. It was important to distinguish the geo-spatiotemporal-sampled *An. arabiensis* aquatic larval habitat models in terms of random variables and the observed values of the random variables. Thus, we determined p + 1 parameters β_0, \dots, β_p . In order to estimate the sampled parameters, it was also useful to employ the matrix notation $Y = X \beta + \varepsilon$, where *Y* was a column vector that included the geo-spatiotemporal-sampled *An. arabiensis* aquatic larval habitat count values of Y_1, \dots, Y_n , sampled in the study site, which in this research included the unobserved stochastic components $\varepsilon_1, \dots, \varepsilon_n$ and the matrix *X*. This matrix was the observed larval habitat parameters $\begin{pmatrix} 1 & x_{1i} & \cdots & x_{1p} \end{pmatrix}$

ter values of the regressors expressed as: X =

$$= \begin{pmatrix} 1 & x_{2i} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}$$

Thereafter, the sampled explanatory predictor covariate coefficient values were log-transformed to normalize the distribution and minimize standard error in the residual forecasts. Multicollinearity diagnostics from the COLLIN option in SAS were estimated. Residual-based diagnostics for univariate and multivariate conditional heteroscedastic models previously constructed from clustering field and remote-sampled vector insect habitat parameter estimates have revealed that errors in variance uncertainty estimation can substantially alter numerical predictions models due to multicollinearity [5]. The SAS COLLIN option produced eigenvalues and condition index, as well as proportions of variances with respect to individual-sampled *An. arabiensis* aquatic larval habitat explanatory predictor covariates in the models. The conditional index scores indicated no significant multicollinearity in either model output.

In this research extra-Poisson variation was detected in the residual variance estimates of both An. arabiensis aquatic larval habitat models. Extra-binomial (i.e., extra-Poisson) variation occurs when discrete data comes in the form of counts or proportion that display greater variability than would be predicted when fitting a model [7]. When sampled vector mosquito aquatic larval habitat data are overdispersed, the square root and logarithmic transformations may be less effective at making the mean and variance independent [1]-[5]. As such, we employed a negative binomial regression with a gamma distributed non-homogenous mean constructed in PROC GENMOD to account for the overdispersion in both the geo-spatiotemporal An. arabiensis larval habitat models. The negative binomial distribution arises as a continuous mixture of Poisson distributions in a vector mosquito larval habitat distribution model where the mixing distribution of the Poisson rate is a gamma distribution [2]. The negative binomial model is a quadratic function of the mean and the variance which commonly affect the weights in the iteratively weighted least-squares algorithm for fitting spatiotemporal-sampled vector mosquito aquatic larval habitat data [4]. We used the negative binomial regression model with a non-homogenous gamma distributed mean to linearly adjust the sampled data in the Karima rice agro-village complex epidemiological study site. In this research, the negative binomial distribution [*i.e.*, NB(r, p)] was represented as a compound Poisson distribution by letting Y_n , $n \in N_0$ denote a sequence of the independent and identically distributed An. *arabiensis* aquatic larval habitat randomized variables, each one having the logarithmic distribution Log(p). The negative binomial for a predictive linear vector mosquito aquatic larval habitat model is a Poisson (λ) distribution, where λ is itself a random variable, distributed according to Gamma(r, p/(1-p)) [2].

Steps in constructing the negative binomial models in SAS was similar to our description of the Poisson *An. arabiensis* larval habitat time series predictive regression models, that is, the log of the mean, μ , was a linear function of the independent variables, $\log(\mu) = \text{intercept} + b_1 \times X_1 + b_2 \times X_2 + \dots + b_3 \times X_m$, which implied that μ was the exponential function of independent variables, $\mu = \exp(\text{intercept} + b_1 \times X_1 + b_2 \times X_2 + \dots + b_3 \times X_m)$. Instead of assuming as before that the distribution of Y, (*i.e.*, sampled *An. arabiensis* aquatic larval habitat parameters) were Poisson, we now assumed that Y had a negative binomial distribution. We relaxed the assumption about equality of mean and variance (*i.e.*, Poisson distribution property), since in our models the variance of negative binomial was equal to $\mu + k\mu^2$, when $k \ge 0$ was a dispersion parameter. The maximum likelihood estimation method was used to estimate k as well as the parameters of the regression models for $\log(\mu)$. Since the Poisson regression model can be generalized by introducing an unobserved heterogeneity term for observation i [5], individual sampled *An. arabiensis* larval habitat data at the Karima study site were assumed to differ randomly in a manner that was not fully accounted for by the observed covariate coefficient indicator values. In this research this was formulated where the unobserved heterogeneity term $\tau_i = e^{\varepsilon_i}$. I was independent of the vector of regressors X_i . Then the distribution of y_i conditional on x_i and τ_i was Poisson with conditional mean and conditional variance (*i.e.*, $\mu_i \tau_i$) which in this research was expressed as

$$f\left(y_{i} \middle| X_{i}, \tau_{i}\right) = \frac{\exp\left(-\mu_{i}\tau_{i}\right)\left(\mu_{i}\tau_{i}\right)^{y_{i}}}{y_{i}!}$$

We then let $g(\tau_i)$ be the probability density function of τ_i in the linear vector time series explanatory field and remote specified endemic transmission oriented *An. arabiensis* aquatic larval habitat models. Then, the distribution $f(y_i|X_i)$ was no longer conditional on τ_i , which was obtained by $f(y_i|X_i)$, τ_i with respect to τ_i : $f(y_i|X_i) = \int_0^{\infty} f(y_i|X_i,\tau_i)g(\tau_i)d\tau_i$ in both larval habitat models. Our autoregressive algorithms had an analytical solution to this integral which existed when τ_i was assumed to follow a gamma distribution. This solution was the negative binomial distribution. When the model contains a constant term, it is necessary to assume that $E(e^{\varepsilon_j}) = E(\tau_i) = 1$, in order to identify the mean of the distribution [7]. Thus, it was assumed that τ_i generated from the *An. arabiensis* aquatic larval habitat explanatory predictor covariates sampled in the Karima epidemiological rice agro-village complex study site followed a gamma (θ, θ) distribution with $E(\tau_i) = 1$ and $V(\tau_i) = 1/\theta$:

$$g(\tau_i) = \frac{\theta^{\theta}}{\Gamma(\theta)} \tau_i^{\theta-1} \exp(-\theta\tau_i)$$

where $\Gamma(\theta) = \int_0^\infty Z^{x-1} \exp(-z) dz$ was the gamma function and θ was a sampled field and/or remote sampled endemic transmission oriented parameter estimates. Then, the density of given X_i was derived as $f(y_i | X_i)$

which rendered
$$\int_{0}^{\infty} f\left(y_{i} | X_{i}, \tau_{i}\right) g\left(\tau_{i}\right) \mathrm{d}\tau_{i}, \quad \frac{\theta^{\theta} \mu_{i}^{y_{i}}}{y_{i}! \Gamma(\theta)} \int_{0}^{\infty} \mathrm{e}^{-(\mu_{i}+\theta)\tau_{i}} \tau_{i}^{\theta+y_{i}-1} \mathrm{d}\tau_{i}, \quad \frac{\theta^{\theta} \mu_{i}^{y_{i}} \Gamma\left(y_{i}+\theta\right)}{y_{i}! \Gamma(\theta) (\theta+\mu_{i})^{\theta+y_{i}}} \text{ and}$$

 $\frac{\Gamma(y_i + \theta)}{y_i ! \Gamma(\theta)} \left(\frac{\theta}{\theta + \mu_i}\right)^{\nu} \left(\frac{\mu_i}{\theta + \mu_i}\right)^{y_i}$. Making the substitution $\alpha = \frac{1}{\theta} (\alpha > 0)$, the distribution in both larval habitat

models was then rewritten as $f(y_i | X_i) = \frac{\Gamma(y_i + \alpha^{-1})}{y_i ! \Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu_i}\right)^{\alpha} \left(\frac{\mu_i}{\alpha^{-1} + \mu_i}\right)^{y_i}$, $y_i = 0, 1, 2, \cdots$ Thus, the negative

tive binomial distributions of the sampled aquatic larval habitats in the Karima study site were derived as a gamma mixture of Poisson random variables. The models had a conditional mean $E(y_i|X_i) = \mu_i = e^{x_i \beta}$ and conditional variance $V(y_i|X_i) = \mu_i \left[1 + \frac{1}{\theta}\mu_i\right] = \mu_i \left[1 + \alpha \mu_i\right] > E(y_i|X_i)$. The conditional variance of the negative of the negative variables.

tive binomial distribution exceeded the conditional mean. Overdispersion in linear vector mosquito aquatic larval habitat models results from neglected unobserved heterogeneity [7]. The negative binomial model with variance function (*i.e.*, $V(y_i|X_i) = \mu_i + \alpha \mu_i^2$) is quadratic in the mean, which is referred to as the NEGBIN2 model [7]. To estimate this model in this research, we specified DIST = NEGBIN (p = 2) in the MODEL statements. A test of the Poisson distribution was carried out by testing the hypothesis that $\alpha = \frac{1}{\theta_i} = 0$. The Poisson distribution was carried distribution where $\alpha = 0$ [5]. A World test of this have

distribution is a special case of the negative binomial distribution where $\alpha = 0$ [5]. A Wald test of this hypothesis was provided. In this research this test was the reported statistic for the estimated α in the linear study site models. The log-likelihood function of the models was then given by

$$\sum_{\ell i=1}^{N} \left\{ \sum_{j=0}^{y_{i}-1} \ln\left(j+\alpha^{-1}\right) - \ln\left(y_{i}!\right) - \left(y_{i}!\right) - \left(y_{i}+\alpha^{-1}\right) \ln\left(1+\alpha\exp\left(X_{i}'\beta\right)\right) + y_{i}\ln\left(\alpha\right) + y_{i}X_{i}'\beta \right\}$$

and the equation $\Gamma(y+a)/\Gamma(a) = \prod_{j=0}^{y-1} (j+a)$. The gradient in both our time series explanatory field and remote specified endemic transmission oriented *An. arabiensis* aquatic larval habitat pre-models was

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^{N} \frac{y_i - \mu_i}{1 + \alpha \mu_i} X_i \quad \text{and} \quad \frac{\partial \ell}{\partial \alpha} = \sum_{i=1}^{N} \left\{ -\alpha^{-2} \sum_{j=0}^{y_i - 1} \frac{1}{(j + \alpha^{-1})} + \alpha^{-2} \ln(1 + \alpha \mu_i) + \frac{y_i - \mu_i}{\alpha(1 + \alpha \mu_i)} \right\}.$$

Thereafter, we used the seasonal sampled immature count data and standard deviations of the log number of larval/pupal counts collected in the Karima study site to determine sample size requirements. We applied a sampling intensity formula for determining the number of immature *Anopheles* samples to collect from a finite population where $n = (t_s/E)^2$ where t = t value $(t \approx 2)$, s = the standard deviation of log-larval/pupal count values observed in the study site(s = 0.889) and E was the desired half-width of the confidence interval around the mean expressed in same units as standard deviation where $E = \ln(1.25)$. Applying this formula and assuming that the geo-spatiotemporal larval/pupal production was similar for the sampled *An. arabiensis* aquatic larval habitats in the study site, we determined 152 samples were required. We overlaid vector images of the sampling scheme with the raster image of the study site to identify AOI's within the sampling frame for field crews to visit. All potential aquatic georeferenced aquatic larval habitat sites were identified and data relative to species composition and abundance, predators, water quality and environmental parameters were collected longitudinally by the field crews.

Field data parameters were entered in Microsoft Excel files and analyzed using and SAS 9.1.3[®] (SAS Inc. Carey, NC, USA). Before analyses the data was tested for collinearity using design matrix from a Poisson regression model and run through SAS PROCREG/VIF (Variant Inflation Factor) procedure which indicated the absence of problematic correlation among the predictors. ANOVA test was used to compare the differences in the An. arabiensis larval/pupal abundance among the rice stages. The analyses revealed that post-transplanting was the rice stage associated with the highest larval/pupal count throughout the rice cycle. Differences in immature count between the An. arabiensis aquatic habitats were compared using Student's t-test. Forward multiple regression analysis was then used to obtain the best predictor variables explaining the abundance of mosquito immatures. In probability and statistics, Student's t-distribution (or simply the t-distribution) is a family of continuous probability distributions that arise when estimating the mean of a normally distributed population in situations where the sample size is small and population standard deviation is unknown [5]. A widely used statistical analyses in time series vector mosquito larval habitat endemic transmission oriented risk modeling is the Student's t-test for assessing the statistical significance of the difference between two sample means, the construction of confidence intervalsand for parsimoniously quantitating the difference between two linearized population means [2]. The relative abundance of mosquitoes was expressed as the number of mosquito larvae/pupae per 20 dips because the number of larvae/pupae sampled was low. Statistical analyses was done using log-transformed ($\log_{10}n + 1$) larval/pupal counts to normalize the data. Results were considered significant at P < 0.05.

3. Results

The overall accuracy of the land cover classification from the QuickBird image was Kappa statistic, 0.91 for the non-riceland site and 0.94 for the riceland epidemiological study site. Generally, the frequency of confusion of the LULC classification was low (Table 2). Overall, instances of confusion were minimal and did not affect user's classification accuracy for the study sites. The user's accuracy ranged between 92% and 97%, with relatively low errors of commission (excesses) and the producer's accuracy ranged between 91% and 98%.

All QuickBird polygons generated were conformed to larval habitat boundaries within the 1 km buffer of the study site. Multiple data layers of spatial and attribute data were created, using different coded values for various rice field attributes, which were linked to the same grid cell. This digitized grid-based algorithm allowed retrie-

.

study site.					
Land Cover Type	Producer's Accuracy (%)	Omission Errors (%)	User's Accuracy (%)	Commission Errors (%)	
Plowing Flooding	91	3	96	3	
	96	5	97	9	
Post-harvested	98	2	94	6	
Tillering	91	8	92	5	

val and transformation of seasonal larval habitat data, regardless of spatial dimensionality of a riceland *An. arabiensis* aquatic habitat, in the study site.

In this research, the relationship between prevalence of each individual potential predictor variable sampled in the epidemiological study site was investigated by single variable regression analysis in PROC NL MIXED. We used the regression line $(y_i - \overline{y}) = (\hat{y}_i - \overline{y}) + (y_i - \hat{y}_i)$ to generate a pseudo R² value where the first term was the total variation in the response y (i.e., total density larval count of An. arabiensis aquatic habitats) and the second term was the variation in mean response based on the sampled parameters. The third term was the residual value in the risk model estimates. Squaring each of these terms and adding over all of the sampled observations generated the equation $\sum (y_i - \overline{y})^2 = \sum (\hat{y}_i - \overline{y})^2 + \sum (y_i - \hat{y}_i)^2$. This equation was then written as SST = SSM + SSE, where SS was notation for sum of squares and T, M, and E were the notation for total, model, and error, respectively. The square of the sample correlation was equal to the ratio of the estimates while the sum of squares was related to the total sum of squares: $r^2 = SSM/SST$. This formalized the interpretation of R^2 as explaining the fraction of variability in the sampled An. arabiensis aquatic habitat data explained by the regression model. The sample variance s_y^2 was equal to $\sum (y_i - \overline{y})^2 / (n-1)$, which in turn was equal to the SST/df, the total sum of squares divided by the total df. A regression equation was then constructed using the mean square model (*i.e.*, MSM) = $\sum_{i=1}^{\infty} (\hat{y}_i - \overline{y})^2 / (l)$, which in this research was equal to the SSM/df. The corresponding mean square error (*i.e.*, MSE) was $\sum (y_i - \hat{y}_i)^2 / (n-2)$, then was equal to SSE/df and the estimate of the variance about the regression line (*i.e.*, σ^2). The MSE is an estimate of σ^2 for determining whether or not the null hypothesis is true [24]. For the geo-spatiotemporal sampled explanatory field and remote specified endemic transmission oriented predictive variables, (p) the An. arabiensis aquatic larval habitat modeled the DFM which in this research was equal to p, and the error degrees of freedom (dfe) which was equal to (n - p - 1), and the total degrees of freedom (dft) which was equal to (n-1), the sum of DFM and DFE. Explanatory and response variables were numeric. The relationship between the mean of the response variable (i.e. larval count) and the level of the sampled explanatory predictor covariates in the regression equation were assumed to be approximately linear (*i.e.*, straight line). The corresponding table generated classified each the field and remote-sampled An. arabiensis aquatic habitat larval habitat parameters in SAS as in Table 3.

In the multiple regression analyses, the test statistic MSM/MSE had an F(p, n - p - 1) distribution. In this research, the null hypothesis was $\beta_1 = \beta_2 = \cdots = \beta_p = 0$, and the alternative hypothesis was at least one of the sampled *An. arabiensis* aquatic habitat larval habitat parameters $\beta_j \neq 0$, $j = 1, 2, \cdots, p$. The F test did not indicate which of the parameters $\beta_j \neq$ was not equal to zero, only that at least one of them was linearly related to the response variable. The ratio SSM/SST = R^2 (*i.e.*, squared multiple correlation coefficient) was the proportion of the variation in the response variable that was explained by the immature sampled habitat data. The square root of R^2 (*i.e.*, the multiple correlation coefficient) was the correlation between the sampled *An. arabiensis* aquatic larval habitat observations (*i.e.*, y_i) and the fitted values (*i.e.*, \hat{y}_i). Additionally, from the sampling distribution, generated from the sampled *t* parameters, the probability of obtaining an F was large or larger than the one was calculated. We noted that, the *t*-test and the F-test were equivalent; the relation between ANOVA and *t* was given by $F = t^2$. Significant differences by ANOVA were noted for mean numbers of the sampled *An. arabiensis* aquatic larval habitat collected throughout the sampling frame (F = 44.7, df = 1).

A Poisson regression analyses was then constructed in PROC NL MIXED to determine the relationship between the sampled *An. arabiensis* aquatic habitat larval count data and the sampled habitat characteristics. The Poisson models were built using the time series field and remote-sampled larval habitat data. A negative binomial regression had to be employed, however, as an examination of the data indicated that overdispersion was

Table 3. The An. arabiensis aquatic habitat larval habitat regression-based model parameters.					
Source	Degrees of Freedom	Sum of Squares	Mean Square		
Model	p MSM/MSE	$\sum \left(\hat{y}_i - \overline{y} \right)^2$	SSM/DFM		
Error	n-p-1	$\sum \left(y_i - \hat{y}_i \right)^2$	SSE/DFE		
Total	n-1	$\sum \left(y_i - \overline{y}\right)^2$	SST/DFT		

a significant problem in the Poisson model. The Poisson distribution is a special case of the negative binomial distribution, where the mean approximates the standard deviation [24]. We assumed that the log of the mean, μ , was a linear function of independent variables, $\log(\mu) = \text{intercept} + b_1 \times X_1 + b_2 \times X_2 + \dots + b_3 \times X_m$ in the model which implied that μ was the exponential function of independent variables when multiple independent geospatiotemporal $\mu = \exp(\text{intercept} + b_1 \times X_1 + b_2 \times X_2 + \dots + b_3 \times X_m)$ [25]. Therefore, instead of assuming that the distribution of the sampled *An. arabiensis* aquatic habitat larval habitats parameter estimates (*i.e.*, *Y*) was Poisson, we were able to assume that Y had a negative binomial distribution. We relaxed the assumption about equality of mean and variance (*i.e.*, Poisson distribution property), since the variance of negative binomial was equal to $\mu + k\mu^2$, where $k \ge 0$ was a dispersion parameter. The maximum likelihood method was used to estimate *k*, as well as the sampled larval habitat parameters of the regression model for $\log(\mu)$. For the negative binomial distribution reduced to Poisson when *k* was 0.

In the regression analyses the null hypothesis was: H₀: k=0 and the alternative hypothesis was: H_a: k > 0. We recorded the log-likelihood (*i.e.*, LL) for the time series field and remote specified endemic transmission oriented *An. arabiensis* aquatic habitat larval habitat models. We used the likelihood ratio (LR) test to compute the LR statistic using -2(LL) (Poisson) and the LL (*i.e.*, negative binomial). The asymptotic distribution of the LR statistic had probability mass of one half at zero and one half-chi-square distribution with 1 df. To test the null hypothesis at the significance level α , we employed the critical value of chi-square distribution corresponding to significance level 2α , that was rejection of H₀, if LR statistic > $\chi^2_{(1-2\alpha, 1 \text{ df})}$. We generated the log of the mean, μ , which in this research was a linear function of independent variables, $\log(\mu) = \text{intercept} + b_1 \times X_1 + b_2 \times X_2 + \dots + b_3 \times X_m$, in the larval habitat model which implied that μ was the exponential function of the independent variables when $\mu = \exp(\text{intercept} + b_1 \times X_1 + b_2 \times X_2 + \dots + b_3 \times X_m)$.

Table 4 lists the dependent and independent variables collected in the Karima epidemiological riceland agro-village complex study site. The count of *An. arabiensis* larvae/pupae collected at a habitat had a mean of 4.7, with a standard deviation of 5.6. The median count was 5 *Anopheles* larvae/pupae and the count range from 0 to 41 *Anopheles* larvae. The distribution was right skewed with 75% of the sampled aquatic habitats having 4 or less larvae. Significant correlations existed between some of the independent variables including study site and number of tillers, (r = -0.23, p = 0.004). The table reveals the abundance of riceland *Anopheles* larvae/20 dips collected in the paddy and canal habitats at the Karima study site. In the study site, the difference in the abundance of pupae and 1st, 2nd and 3rd instars larvae collected in paddy and canal habitats was not significant (P > 0.05), while that of 4th instars larvae was significantly higher in the paddy habitats than in the canal habitats. (t = 5.19, df 179, P < 0.05).

The regression analyses identified rice height, levels of turbidity and number of tillers as significantly influencing the count of *An. arabiensis* mosquitoes in the study site. Muturi *et al.* [8] reported a positive association between turbidity and *An. arabiensis* in the Mwea rice fields. In the rice growing cycle, rice plants increase in height and tiller numbers and affect the microhabitat conditions of mosquito larval habitats [2].

The Feature Extraction Module in ENVI allows quick and accurate extraction of an *An. arabiensis* aquatic habitat from the QuickBird visible and NIR data. All post-transplanting habitat pixels were isolated from a LULC map. The endmember determination estimated the set of distinct spectra that comprised the mixed QuickBird pixels in the riceland scene. The produced abundance planes provided estimates of the fractional

Table 4. The mean number of An. arabiens	is larvae/pupae o	collected (mean ±	E SE) per 20 dip	s in paddies con-
taining different stages of rice growth usin	g field sampled	and QuickBird 0	.61 m visible a	nd near infra-rec
(NIR) data in Karima rice-village complex.				

	<u> </u>						ĺ
Paddy category	Number of habitats	1 st instars	2 nd unstars	3 rd instars	4 th instars	Pupae	
Ploughed	25	1.41	0.95	0.09	0.00	0.36	
Flooded	23	1.67	1.10	0.30	0.00	0.36	
Post-transplanting	30	8.02	9.00	8.89	3.22	0.99	
Tillering	28	2.00	6.67	2.00	1.20	0.67	
Fallow	27	1.00	0.67	0.00	0.00	0.01	

abundances for the endmembers in each QuickBird visible and NIR pixel. The ENVI Feature Extraction Module had the ability to extract a variety of *An. arabiensis* aquatic habitat features. The feature extraction workflow guides you through the process of segmenting the image by dividing the image into discreet, real-world, objects through the identification of edges and the grouping of pixels into regions (http://www.exelisvis.co.uk/). Prior to the final classification and extraction, the ENVI preview portal window allowed us view the preliminary results, so that we could make adjustments to the analysis before processing the entire image data. ENVI also provided us tools to polish and refine our extraction results. We chose to annotate the results, count individual features, edit or smooth vectors, assess classification accuracy, and generate summary statistics about extracted features. Following image segmentation, the workflow provided us with several options for defining the time series *An. arabiensis* aquatic larval habitat remote features that need to be extracted We chose example segments in the larval habitat image that were representative of each habitat feature class or create rule sets that must be met in order to assign segments to a feature class.

Vector map layers of rice fields were outlined as polygons, using ESRI shape files and MapInfo tab-file formats. By tapping on the PDA screen of a rice paddy polygon layer in the Karima study site, the end user had the ability to toggle a particular polygon as ON/OFF or TREATED/NOT TREATED, by changing the polygon line color, fill pattern, or both. The Recon $X^{(0)}$ supported both the raster image layer display and a vector-based data layer used for mapping rice paddy polygons. There was a minimal requirement to be able to output a simple text file of information indicating the paddy had been treated versus not treated (**Figure 3**). Spatial data sets consisting of a raster image layer and vector line work describing the QuickBird polygons were stored and displayed on the Recon $X^{(0)}$. When uploaded into a Recon X, the polygons displayed all treated and untreated paddies within a 1 km buffer of the study site.

4. Discussion

Initially, Poisson regression was used to model the geo-spatiotemporal larval productivity of the spatiotemporalsampled *An. arabiensis* aquatic larval habitats in the Karima agro-village riceland epidemiological study site in our customized real-time bidirectional PDA-GIS-DGPS-RS cyber-environment. An examination of the data, however, indicated that significant overdispersion was present. Therefore, a negative binomial was used to model the overdispersed Poisson larval count data. Negative binomial regression models estimate a dispersion parameter that is used to remove the effects of overdispersion and provide more accurate estimates of standard error [5]. The negative binomial was derived as a Poisson-gamma mixture and as a GLM. PROC NL MIXED expressed the variance of the response for the negative binomial as variance $(y) = \mu + k\mu^2$ as opposed to the more



Figure 3. QuickBird visible and near infra-red data of treated and untreated paddy habitats Karima epidemiological study site displayed in the Trimble Recon X, 400 MHz Intel PXA255 Xscale CPU.

common notation $(y) = \mu + \mu^2 / \nu \mu$ [2]. The difference in notation was trivial $(k = 1/\nu)$. The straightforward derivation of the models, from the negative binomial probability distribution function, did not, however, equate with the Poisson–gamma mixture-based version of the negative binomial. Rather, canonical link and inverse canonical link were converted to log form in the real-time PDA-GIS-DGPS-RS cyber-environment. A GLM-based negative binomial was produced that yielded identical parameter estimates of *An. arabiensis* aquatic larval habitats to those calculated by the mixture-based model for the Karima epidemiological study site.

As a non-canonical linked model, however, the standard errors did differ slightly from the mixture model. A maximum likelihood estimator was employed by default as an observed information matrix, to produce standard errors in the cyber-environment. The GLM algorithm produced standard errors based on the expected information matrix using the difference in standard errors between the two versions of negative binomial analyses. The GLM negative binomial algorithm was amended to allow production of standard errors based on the field-sampled larval count data of *An. arabiensis* aquatic larval habitats at the epidemiological study site. The amended GLM-based negative binomial rendered identical estimates and standard errors to that of the mixture-based negative binomial analyses. The log-negative binomial data was then imported into an SAS database within our real-time bidirectional PDA-GIS-DGPS-RS cyber-environment.

A time series stratified digitized grid-based algorithm in the real-time PDA-GIS-DGPS-RS cyber-environment identified the field and remote-sampled *An. arabiensis* aquatic habitats in the Karima epidemiological rice agro-village complex study site. By digitally tracing riceland *An. arabiensis* aquatic habitats using QuickBird visible and NIR data in a GeoGrid database, polygons were generated that approximated the curvature of the habitat boundaries. All visible and NIR polygons were conformed to the sampled aquatic habitat boundaries within the 2 km buffer of the epidemiological study site. Field attribute tables were then associated to the polygons.

Building a digitized grid geodatabase PDA-GIS-DGPS-RS cyber-tools can identify relationships between riceland mosquito aquatic habitats, rice plant stage of development and agro-village-complex within the frame-work of relational database technology. Thus, by digitally tracing riceland *An. arabiensis* aquatic habitats using QuickBird visible and NIR data in a PDA-GIS-DGPS-RS cyber-environment; a vector biologist, an epidemiologist, or other data analysts can generate polygons that approximate the curvature of the paddy and canal habitat boundaries. Field attribute tables can then be associated to the polygons. Multiple data layers can be created thereafter using different coded values for geographically representing various field attributes to the same digitized grid cell which can allow for multiple interactions between compatible ecological geodatabases enabling retrieval and transformation of seasonal larval habitat data efficiently regardless of spatial dimensionality of a same digitized grid cell which can allow for multiple interactions between compatible ecological databases for the same digitized grid cell which can allow for multiple interactions between compatible data layers can then be created PDA-GIS-DGPS-RS cyber-environment using different coded values for various field attributes for the same digitized grid cell which can allow for multiple interactions between compatible ecological databases while enabling retrieval and transformation of seasonal collected habitat data efficiently regardless of spatial dimensionality of a sampled habitat.

In this research, ENVI software in the real-time bidirectional PDA-GIS-DGPS-RS cyber-environment automatically categorized individual sampled An. arabiensis aquatic habitats pixels into separate spectral classes, converted remotely sensed raster layers to vector coverage and classified the layers as ESRI shape files. In ENVI the sampled habitat data and corresponding LULC classification generated in ArcGIS was evaluated using spectral and spatial browsing, and using color composites to characterize spectral variability of the sampled An. arabiensis aquatic habitat explanatory covariates. ENVI's built-in spectral libraries provided quick reference to the QuickBird visible and NIR data, giving well-characterized reference sets of habitat spectral values based on the LULC classifications. This was important as there was significant emissivity spatial heterogeneity within the georeferenced empirical ecological dataset based on individual sampled habitat reflectance estimates. Therefore several LULC feature data corresponding to the sampled An. arabiensis aquatic habitats had to be included in the spectral library. ENVI's spectral analysis tools also used an adjacency correction method to classify the images based on LULC reflectance spectral estimates (http://www.ittvis.com). For example, the MODTRANbased Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) technique in ENVI removed most of the solar and atmospheric effects, transforming the georeferenced LULC classified An. arabiensis aquatic larval habitat data from radiance to surface reflectance values. However, the most important tool in the PDA-GIS-DGPS-RS cyber-environment was the Feature Extraction module which helped automate the process of performing accurate segmentations for image classification of the sampled An. arabiensis aquatic larval habitats by LULC, thus accounting for larval/pupal development by rice stage. For instance, the FLAASHTM model included a method for retrieving an estimated haze amount from selected "dark" land covers pixels of sampled *An. arabiensis a*quatic larval habitats in the QuickBird scene. When looking for a single target, the result is a grayscale image where the pixels that best match the target properties are the darkest (<u>www.ittvis.com</u>). ENVI computed the texture features in the PDA-GIS-DGPS-RS cyber-environment of each individual sampled *An. arabiensis* aquatic larval habitat attribute in the epidemiological riceland agro-village epidemiological study site and converted raw texture descriptors into spectrally quantifiable data.

The spatiotemporal-sampled georeferenced explanatory field and remote specified *An. arabiensis* aquatic larval habitat data from ENVI was entered into the Recon X[®] in the PDA-GIS-DGPS-RS cyber-environment. The Recon X[®] utilized a touch screen for entering spatiotemporal explanatory field and remote data of *An. arabiensis* aquatic habitats. All map and data storage in the cyber-environment of the sampled data was done on removable Compact Flash (CF) storage cards. The PDA was configured with 64M RAM, 256M Flash memory, and a 2 GB SanDisk Extreme IV CF card. This process provided adequate backup protection to reload programs using the ecological time series sampled *An. arabiensis* aquatic habitat data. A CF card reader PDA-GIS-DGPS-RS cyber-environment was then specified to assist the field research team in quickly populating the CF cards with project maps and to transfer data with laptop PCs.

Live data entry pertaining to riceland agroecosystem treatment data of the sampled *An. arabiensis* aquatic habitats in our bidirectional PDA-GIS-DGPS-RS cyber-environment was performed on the Recon $X^{\text{(B)}}$ and, was transferable to a laptop computer via a wired (or optional wireless) connection. For this study, transfers were done using ActiveSync software that came with the Windows Mobile powered Recon $X^{\text{(B)}}$. The Mobile VCMSTM software was used to record field data on the Trimble Recon $X^{\text{(B)}}$ which was synchronized automatically with a centralized VCMS database in the PDA-GIS-DGPS-RS cyber-environment running on a laptop computer. For this research project, some of the PDAs also supported an integrated digital camera, allowing the field staff to geocode, map and take pictures of *An. arabiensis* aquatic larval habitat locations in the study site. Interestingly, the Recon $X^{\text{(B)}}$ included an optional wireless LAN connection feature that we used with Trim PixTM image technology to connect a Wi-Fi capable Nikon digital camera for automated capturing and geocoding of digital images of the sampled aquatic larval habitats in the PDA-GIS-DGPS-RS cyber-environment. Local area network (LAN) supplies networking capability to a group of computers in close proximity to each other [5]. A LAN in turn connected to the Internet.

Ecological sampled information stored on a Recon $X^{(0)}$ in the PDA-GIS-DGPS-RS architecture then directly forwarded and/or synchronized with other existing information systems to support further time series *An. arabiensis* aquatic larval habitat data analyses. A dedicated server was used as the back-end data repository which in this research we used to store collected ecological sampled habitat data, as it is aggregated from multiple Recon PDAs. The back-end dedicated server scenario supported the dissemination of the larval habitat information within the PDA-GIS-DGPS-RS cyber-environment via web-based reporting systems when connected to the Internet.

Web-based technologies have made anonymous, automated, and integrated reporting systems a reality [5]. Moreover, the use of Web-based technologies for time series endemic transmission oriented An. arabiensis aquatic larval habitat data in a robust PDA-GIS-DGPS-RS cyber-environment ensures reporting consistency and data quality, as well as the ability to survey and detect adverse events. This, in turn, provides data that may be measured quantitatively and, analyzed for trends, then used to generate feedback for the quality assurance process. Thus, web-based technology has the potential to improve workflow and the fluid transmission of vital information on errors that surface from a system perspective, without placing blame on fallible individuals and institutions. Safety may be improved in an ongoing manner with the use of technology, particularly as the effects of systemic changes are being assessed. The use of information technology for oriented time series An. arabiensis aquatic larval habitat data PDA-GIS-DGPS-RS cyber-environment can reduce errors in three ways: (1) adverse events (e.g., flooding, droughts) and errors associated with these events may be prevented before they can occur; (2) the response time needed to resolve the root cause of adverse events can be decreased when implementing gridded sub-meter resolution sampled data feature attributes, thus preventing uncertainty reoccurrence; and (3) LULC trends can be tracked and pertinent feedback about errors and adverse events then can be disseminated to other vector biologists, epidemiologists or research collaborators. Remote monitoring of An. arabiensis aquatic larval habitat data can help simplify therapeutic procedures through a user-friendly mobile device supplemented with interactive audiovisual cues which can help keep track of exchange details using electronic records [8].

While time series web-based reporting systems within PDA-GIS-DGPS-RS cyber-environment are becoming more common in public health for disease surveillance and for tracking interventions such as field sampled An. arabiensis aquatic habitat data, their use for program evaluation is relatively new. Alternatively, a broadband satellite access point could be employed to wirelessly link multiple mobile field data collection devices to a server, via the Field Bridge middleware software. For example, Hughes Network Systems' (HNS) 9201 Broadband Satellite IP Modem and Wi-Fi Access Point can be utilized as part of the field data communications within a robust PDA-GIS-DGPS-RS architecture for data transfer of seasonal-sampled field and/or remote An. arabiensis aquatic larval habitat endemic transmission oriented data to any Internet connected, back-end VCMS database server. The HNS9201 can support multiple field teams to simultaneously send and receive IP packet data of sampled An. arabiensis aquatic habitat data in PDA-GIS-DGPS-RS cyber-environment via USB, Ethernet, ISDN and Wi-Fi interfaces, using the Inmarsat BGAN satellite network and the Recon X^{\otimes} (Figure 4). This type of satellite link can provide 432 kbps IP data (transmit and receive) capability, Integrated Services Digital Network (ISDN) voice (4kbps) ISDN data (64 kbps) and Wi-Fi access point with multi-user capability. Firewall protection and data encryption techniques (e.g., 128 bit Advanced Encryption Standard (AES) Encryption) can prevent hackers from accessing the sampled field/remote endemic transmission oriented data, while firewall and password protection/encryption measures, at the database level can protect the transmission and storage of sampled An. arabiensis aquatic habitat data PDA-GIS-DGPS-RS cyber-environment. All the data managed by such a remote surveillance system can be spatially displayed via the use of GIS software, both as an embedded component of the VCMS database, or in a standalone desktop GIS using industry standard shape files. Traditional shapefiles contain cryptic field names, no rendering information and no metadata [5].

By using a PDA-GIS-DGPS-RS web-based reporting system connected to the Internet, the Trimble Recon X[®] enables the export or import of data layers completely-with labels, colors, rendering and metadata. Compatible geospatial software such as ESRI ArcGIS Interoperability tools, allows for a seamless transfer of spatiotemporally sampled field and remote specified explanatory *An. arabiensis* larval habitat endemic transmission oriented information between users of the software. Therefore, a Trimble Recon X[®] web-based reporting system can easily import complete packages of data and meta-data while viewing all field information sampled. This system will allow a field manager to easily distribute their data in a single file with meaningful variable names and descriptions; metadata and full map color and symbol rendering.

Furthermore, add-in technology can extend the functionality of a remote bidirectional PDA web-based reporting system by allowing vector biologists, epidemiologists and other\data analysts to have the ability to develop and run their own custom add-ins modules while providing a suitable environment to run complex spatial predictive eco-epidemiological risk models. For example, an empirical ecological dataset of georeferenced explanatory spatiotemporal sampled field and remote specified endemic transmission-oriented georeferenced *An arabiensis* aquatic larval habitat covariate coefficients values in a PDA-GIS-DGPS-RS cyber-environment can be exported from different GIS compatible software, such as spdep from R[®], GeoDa[®], D2K[®] Python[®] and from workflow environments from the National Center for Supercomputing Applications (NCSA)'s Cyber-Integrator



Figure 4. A bidirectional PDA web-based reporting system using a broadband satellite access point and the Internet.

which can then be analyzed and imported into a compatible field databases. A web-based joint database can then combine the spatiotemporal field and remote-sampled *An. arabiensis* aquatic habitat data and relevant programmatic information can be used to facilitate evidence-based decision-making, by geographically/non geographically spatiotemporally modeling key covariate coefficient indicators in robust PDA-GIS-DGPS-RS cyber-environments (e.g., parameter estimators of prolific larval habitats based on field and/or remote seasonal sampled data). The key advantages of a web-based joint database within a PDA-GIS-DGPS-RS cyber-infrastructure is: 1) the ability to integrate data (demographic, social, economic weather, etc.) with data that are critical to a project, 2) Ability to share data easily with others by sending with the specific *An arabiensis* aquatic larval habitat risk map view, the symbology and all data in one file—so they can quickly understand the issue under discussion, 3) Ability for data recipients to repurpose, redisplay and remix the data and, 4) All-time series field and remote data stored and analyzed is in GIS shape file format.

In the foreseeable future, GIS software in a PDA-GIS-DGPS-RS cyber-environment will continue to play essential roles for breaking through scientific challenges in numerous spatiotemporal explanatory georeferencable field and remote sampling frames for parsimoniously quantitating *An. arabiensis* aquatic larval habitat field and remote sampled endemic transmission oriented data for improving decision-making practices with broad societal impacts (e.g., implementation of IVM). Cyber GIS is fundamentally new software framework under development at Georgia Tech that can provide seamless integration of cyber-infrastructure for GIS, and spatial analysis and malarial time series risk modeling of geo-spatiotemporal sampled *An. arabiensis* data in a PDA-GIS-DGPS-RS cyber-environment.

Critical pieces of the cyber-infrastructure include supercomputers, high-capacity, mass-storage systems, system software suites and programming environments, scalable, interactive, visualization tools, productivity-enhancing software libraries and tools, large-scale data repositories and digitized scientific data management systems [5], However, fulfilling such roles in a PDA-GIS-DGPS-RS cyber-environment for efficient predictive risk modeling empirical sampled georeferenced field and remote specified *An. arabiensis* larval habitat explanatory data sets would be dependent on the ability to handle very large geo-spatiotemporal and complex analysis software based on synthesizing computational and spatial thinking enabled by the cyber-infrastructure, which conventional GIS-based software approaches do not provide.

An epidemiologist, vector biologist or other data analyst may establish CyberGIS as a fundamentally new software framework comprising a seamless integration of GIS, and spatial analysis/risk modeling capabilities in the PDA-GIS-DGPS-RS cyber-infrastructure. Specifically, such a project may robustly spatially target seasonally prolific georeferenced field and remote specified An. arabiensis larval habitat explanatory covariate coefficients in the cyber-environment by: 1) engaging in multidisciplinary tactics [e.g., field and remote GIS science, validation metrics derived from real time geostatistical non-linear algorithm] employing evolving CyberGIS software requirements; 2) integrating and sustaining a core set of composable, interoperable, manageable, and reusable CyberGIS software elements based on epidemiological interventional study site -driven and open source strategies; 3) empowering high-performance and scalable CyberGIS by exploiting spatial characteristics of the sampled endemic transmission oriented data and analytical operations for achieving unprecedented capabilities for geospatial risk malarial risk mapping; 4) enhancing an online geospatial problem solving environment to allow for the contribution, sharing, and learning of CyberGIS software by numerous users (e.g., vector biologists, epidemiologists) for fostering the development of education, outreach, and IVM crosscutting multiple disciplines; 5) deploy and test CyberGIS software by linking with national and international cyber-infrastructure to achieve scalability to significant sizes of geospatial problems, and vector malarial mosquito cyber infrastructure resources, and user communities; and 6) evaluating and improving a time series endemic transmission oriented CyberGIS framework through domain science applications and vibrant partnerships to gain better understanding of the complexity of coupled natural and anthropogenic influenced epidemiological ecosystems.

A time series endemic transmission oriented explanatory field and remote specified Cyber GIS software framework in a robust PDA-GIS-DGPS-RS cyber-environment will shift the current paradigm of GIS and associated spatial analysis/modeling software to create further scalable and sustainable software while achieving groundbreaking scientific advances in understanding human-natural ecosystems associated to immature *An. arabiensis* that would be impossible otherwise. These advances will dramatically advance the understanding of flooding preparedness and response and impacts for droughts. This framework will empower high-performance and collaborative geospatial problem solving and serve as a key driver for the interoperability of international cyber-infrastructure based on broad engagement of user communities related to GIS for both research and edu-

cation purposes for risk mapping seasonal An. arabienesis larval habitat field and remote specified endemic transmission oriented data.

In our PDA-GIS-DGPS-RS cyber-environment, the Trimble Recon handheld features a high-performance 400 MHz processor, built-in Bluetooth wireless technology, and built-in wireless LAN. The system provided two Compact Flash (CF) slots, letting us add peripherals such as GNSS cards, barcode scanners, or memory cards. We also had the ability to employ Bluetooth to connect wirelessly to other devices such as a laser rangefinder, a mobile phone for connection to the Internet, or Trimble's GPS Pathfinder[®] receivers. And since we were within range of a Wi-Fi network, the built-in wireless LAN in the Trimble Recon handheld in the cyber-environment made it very efficient to send and receive spatiotemporal-sampled field and remote sampled *An. arabiensis s.s.* larval m data. As soon as a Wi-Fi hotspot was discovered we could quickly and securely transfer large amounts of time series dependent endemic transmission oriented field and/or remote geosampled aquatic larval habitat data into the network.

Developing and implementing streamlined data collection, aggregation and reporting methodologies, emploving a PDA-GIS-DGPS-RS cyber-environment can provide geographically detailed, real-time information which can lower overall seasonal An. arabiensis aquatic habitat treatment costs. For example, a bidirectional PDA-GIS-DGPS-RS web-based reporting system using a broadband satellite access point and the Internet can provide optimally efficient and timely amounts of relevant field level information. This is important, because riceland An. arabiensis aquatic habitat treatment and management information has the most value when it is quick, clear, easy to understand, and relevant to decisions that need to be made immediately. An adaptable modular information surveillance system in the cyberinfrastrure can help insure that the right decisions are made to reduce the parasite, the vector or both in riceland areas. This system can support dissemination of information such as data about field activities and costs as they are compiled to provide an expedited understanding of how much is being spent in riceland An. arabiensis aquatic habitat treatment, and to determine whether the intended outcomes are being realized. This will allow both field managers and health ministries timely and optimal information to make field operational adjustments and maintain the most efficient and economically feasible pressure on riceland An. arabiensis aquatic habitat populations. For example, once patterns and correlations of spatiotemporal sampled An. arabiensis aquatic habitats are elucidated, field management practices can be modified to optimize applications of fertilizers and pesticides yielding lower overall costs, and minimizing environmental impacts caused by excessive insecticide applications. In addition, a PDA-GIS-DGPS-RS cyber-environment web-based reporting system can provide a module that is dedicated to measuring the economic status of the riceland community and clusters of communities, in order to monitor and measure the impacts of malaria and the economic benefits of malaria control interventions. These optimizations are a critical factor for malaria eradication.

Globally accessible, PDA-GIS-DGPS-RS cyberinfrastructural web-based *An. arabiensis* aquatic larval habitat maps, that are generated from near real-time map and field data updates, can be incorporated into other ecological sampled riceland datasets for fast and timely analyses. Although this project did not fully utilize web-based data dissemination techniques, the capability are there using the same off-the-shelf technologies employed for our field studies. The logical extension of this field research is to continue to expand the use of mobile field data collection and wireless communication technologies in a PDA-GIS-DGPS-RS cyber-environment to provide a single, seamless and cost effective information clearing house for IVM. Such a system can leverage GIS-mapping functionality, wireless communications and web-based publications. This study provides an important starting point for the creation of standardized techniques for mobile field data collection, global data transmission, data aggregation and display, and analyses of multivariate *An. arabiensis* aquatic larval habitat models within a robust cyber-environment. The standardization of a logical geodatabase design for implementing IVM will ultimately support linkage with other relational geodatabases, statistical packages and software applications.

The PDA-GIS-DGPS-RS cyberinfrastructural architecture, deployed as part of this project, is the first step in seamlessly sharing geospatially enabled *An. arabiensis* aquatic habitat data between end users, including field teams, project managers and other researchers. The envisioned web application of this technology would enable: the sourcing and integration of time series *An. arabiensis* aquatic larval habitat data, supporting the creation of new data sets, and the coordinated delivery of focused decision-supported information for generating reports, charts, maps, tables, using common data exchange formats (e.g., Excel, PowerPoint, SPSS, SAS). Additionally, revisal software can also be installed in the Trimble Recon $X^{\text{(e)}}$ in a PDA-GIS-DGPS-RS cyber-environment which can create *An. arabiensis* aquatic habitat maps from spreadsheet or database data which can be integrated with data from weather and other scientific studies, *i.e.* Virtual EarthTM. The initial deployment of the custom

application described here can ensure rapid deployment and proper site configuration needed to support local decision-making and malaria control efforts. A web-based shared database project PDA-GIS-DGPS-RS cyberenvironment can combine a variety of sampled field and remote-sampled mosquito data and relevant programmatic information, in order to facilitate evidence based decision-making for control operations in a real-time or near real-time environment.

Finally, using spatial modeled epidemiological time series data, acquired through a PDA-GIS-DGPS-RS cyberinfrastructural remote bidirectional surveillance system, can simulate disease transmission among An. arabiensis aquatic habitat based on larval/pupal productivity. For instance, Markov simulations models can be created from field sampled data using WinBUGS and its spatial module GeoBUGS which can be facilitated as part of the end-user repository database. Error propagation in these non-linear model estimates can also be spatially quantified using an autocovariate matrix [25]. The first prototype implementation of the Java virtual machine, done at Sun Microsystems, Inc., emulated the Java virtual machine instruction set in software hosted by a handheld device that resembled a contemporary PDA. Thus, accurate models can be developed by sampled immature An. arabiensis aquatic habitat data which can define a set of probabilistic environmental interactions that may affect transitions in aquatic habitats in riceland areas. These riceland An. arabiensis aquatic habitat models in a robust PDA-GIS-DGPS-RS cyberinfrastructural can be made to analysis and display data for predicting habitat locations and for testing and improving various research hypotheses. These models can also provide information about realistic sampling grids, ecological covariates that are involved in high larval/pupal population levels and other transmission dynamics from simple, transparent simulated representations. Geostatistical models of An. arabiensis aquatic habitats derived from PDA-GIS-DGPS-RS cyberinfrastructural data cannot replace mental, qualitative intuition and other intervention techniques, but models can expand into more formal and quantitative realms in which field components and their environmental interactions with productive aquatic habitats are made more specific which can help reduce the overall cost of implementing mosquito control strategies in riceland environments [25].

Geographic information science for seasonal malarial risk mapping in a PDA-GIS-DGPS-RS cyberinfrastructural is a dynamic area in which there are continuous innovations. In remote sensing, for example, increasing use of high resolution imagery (e.g., panchromatic QuickBird), and time-series data present both opportunities (e.g., improved characterization of land cover) and challenges (e.g., handling large data volumes, image understanding) for data fusion and integrated data analysis. Enhancements in internet, wireless and satellite communications, and innovations in *in-situ* sensors, for PDA-GIS-DGPS-RS cyber-infrastructures are paving the way for increasingly robust "real time" applications of remote sensing and GIS, (*i.e.* telegeoprocessing) for accurate seasonal malarial predictive risk mapping. Web-based tools, such as Google Earth (<u>http://earth.google.com/</u>) and Internet Map Service (IMS) applications PDA-GIS-DGPS-RS cyber-environments, now provide an increasingly larger audience of vector ecologists, epidemiologists and other research collaborators with ready access to geospatial malaria-related data while allowing elementary integration of imagery and graphics in PDAs, but more sophisticated implementation of web-based integrated geospatial analysis will require resolution of issues related to metadata standards, data transmission formats, client/server computation and communication protocols for advanced quantification of seasonal sampled malaria-related explanatory covariates.

In conclusion, QuickBird visible and NIR image datasets in a customized PDA-GIS-DGPS-RS cyber-environment containing digitized habitat boundaries, with unique identifiers displayed all sampled habitats within a 1 km buffer of the study site. An object-oriented classification in ENVI technology separated the sampled habitats by larval/pupal productivity. Thereafter, a regression analyses revealed significantly higher *An. arabiensis* larval/pupal counts in the post-transplanting stage of rice development in the cyber-environment. All field and remote-sampled explanatory variables were entered into the PDA-GIS-DGPS-RS cyberinfrastructural architecture. The Recon X displayed all treated and untreated *An. arabiensis* aquatic habitats in the study site. Recon X integrated solutions can enable field-workers to capture observation and location data at the same time. GIS/GPS/and Recon X technological integration may be used to build new databases, maintain database accuracy, and conduct spatial analysis in the field. A Recon X unit can be attached to a laptop for direct data input into a PDA-GIS-DGPS-RS cyberinfrastructural server. The solution can be further synchronized with applications that include digital photography and date and time of data capture. A web-based joint database within a robust PDA-GIS-GPS-RS cyber-environment can ensure timely deliverable of explanatory seasonal mosquito sampled data. This platform lends itself to a larger public health capability and has the ability to scale-in data, providing an easily accessible tool linking data from handheld to desktop to web. This database extends current field sam-

pling technology by providing the ability to transfer geographic information from either a centralized repository or directly between users. Data-based location intelligence aggregated and analyzed in PDA-GIS-GPS-RS cyber-environment servers can provide effective dissemination of information which will lead to improved program management and therefore impact.

Continued research should include field surveillance of highly productive riceland An. arabiensis aquatic larval habitats within PDA-GIS-DGPS-RS cyber-environments using the Recon X[®] can be mapped and monitored based on LULC transition throughout the crop season. The Recon X[®] has the ability to scale-in field sampled data, including gridded climate surfaces, soils, population density, elevation and its derivatives, land cover, and an extensive collection of human, agricultural and livestock census and production information. These data can also be combined with infrastructure data (e.g., roads, rivers, towns, political units) in a seamless, stand-alone Windows-based application or into a web-based environment that would express data in simple and engaging ways across the malaria sector. Unlike simple habitat mapping that just plots points on a map; a PDA-GIS-DGPS-RS cyberinfrastructural web-based reporting system can create dynamic maps that are interactive. Trimble Recon X[®] technology can plot results-based information on a collective mapping system designed to help malaria researchers assess and plan improved programs. Mobile field data collection and reporting, using PDA-based technologies, should be part of a tailored approach for reducing the immature population of An. arabiensis aquatic larval habitats through well-timed/targeted spraying. Off-the-shelf mobile field computing technologies, such as the Windows Mobile powered Trimble Recon X[®] configured with user-friendly data entry software and spatial data such as OuickBird visible and NIR data can play an important role in supporting malaria control field activities.

References

- [1] Sattler, M.A., Mtasiwa, D., Kiama, M., Premji, Z., Tanner, M., Killeen, G.F. and Lengeler, C. (2005) Habitat Characterization and Spatial Distribution of Anopheles sp Mosquito Larvae in Dar es Salaam (Tanzania) during an Extended Dry Period. *Malaria Journal*, 4, 4-11. <u>http://dx.doi.org/10.1186/1475-2875-4-4</u>
- [2] Jacob, B.G., Muturi, E.J., Funes, J., Shililu, J., Githure, J.I., Regens, J.L. and Novak, R.J. (2007) Using Imaging Technologies to Control Malaria. *Imaging Notes*, 22, 17-22.
- [3] Gu, W. and Novak, R.J. (2005) Habitat-Based Modeling of Impacts of Mosquito Larval Interventions on Entomological Inoculation Rates, Incidence, and Prevalence of Malaria. *American Journal of Tropical Medicine and Hygiene*, 11, 546-552.
- [4] CDC (Centers for Disease Control and Prevention). The President's Malaria Initiative (IPM) Rolls into Action. http://www.cdc.gov.
- [5] Jensen, J.R. (2005) Introductory Digital Image Processing: A Remote Sensing Perspective, 3rd Edition, Prentice-Hall, Upper Saddle River.
- [6] Environmental Systems Research Institute. http://www.esri.com
- [7] Digital Globe LTD. <u>http://digitalglobe.com</u>
- [8] Muturi, E.J., Mwangangi, J., Shililu, J., Muriu, S.M., Jacob, B.G., Kabiru, E.W., Gu, W., Mbogo, C.M., Githure, J. and Novak, R.J. (2007) Mosquito Species Succession and Physicochemical Factors Affecting Their Abundance in Rice Fields in Mwea, Kenya. *Journal of Medical Entomology*, 44, 336-344. http://dx.doi.org/10.1603/0022-2585(2007)44[336:MSSAPF]2.0.CO;2
- [9] Jacob, B.G., Nelson, P., Lampman, R., Morris, J., Raim, A., Funes, J., LaPointe, C. and Novak, R.J. (2006) Comparing GPS Technology for Identifying Spatial Ecological Variation for Urban Mosquito Management. *Wing Beats*, 16, 30-33.
- [10] Gillies, M.T. and Coetzee, M. (1987) A Supplement to the Anophelinae of Africa South of the Sahara. South African Institute for Medical Research, **55**, 1-143.
- [11] Service, M.W. (1993) Mosquito Ecology: Field Sampling Methods. 2nd Edition, Elsevier Publishers, Essex.
- [12] Edwards, F.W. (1941) Notes on British Mycetophilidae.
- [13] Bioequip Products Inc. www.bioquip.com
- [14] Meritt, R.W. and Cummins, K.W. (1996) An Introduction to the Aquatic Insects of North America. 3rd Edition, Kendall Hunt Editor, Dubuque.
- [15] ERDAS. http://geospatial.intergraph.com/Homepage.aspx
- [16] Rejmankova, E., Roberts, D., Pawley, A., Manguin, A. and Polanco, S. (1995) Predictions of Adult Anopheles albi-

manus Densities in Villages Based on Distances to Remotely Sensed Larval Habitats. American Journal of Tropical Medicine and Hygiene, 53, 482-491.

- [17] Beck, L.R., Rodriguez, M.H., Dister, S.W., Rodriguez, A.D., Rejmankova, E., Ulloa, A., Meza, R.A., Roberts, D.R., Paris, J.F. and Spanner, M.A. (1994) Remote Sensing as a Landscape Epidemiologic Tool to Identify Villages at High risk Malaria Transmission. *American Journal of Tropical Medicine and Hygiene*, **51**, 271-280.
- [18] Wood, B.R., Washino, R., Beck, L., Hibbard, K., Pitcairn, M., Roberts, D., Rejmankova, E., Paris, J., Hacker, C., Salute, J., Sebesta, P. and Legters, L. (1991) Distinguishing High and Low Anopheline-Producing Rice Fields Using Remote Sensing and GIS Technologies. *Preventive Veterinary Medicine*, **11**, 277-288. http://dx.doi.org/10.1016/S0167-5877(05)80014-5
- [19] Clarke Mosquito Control Products, Inc., Roselle, IL for the VCMS, Mobile VCMS and FieldBridge Software. <u>https://www.clarkemosquito.com/vcms_tech_support</u>
- [20] Jacob, B.G., Muturi, E.J., Funes, J.E., Shililu, J.I., Githure, J.I., Kakoma, I.I. and Novak, R.J. (2006) A Grid-Based Infrastructure for Ecological Forecasting of Rice Land Anopheles arabiensis Aquatic Larval Habitats. *Malaria Journal*, 5, 941-947. <u>http://dx.doi.org/10.1186/1475-2875-5-91</u>
- [21] Dolo, G., Briët, O.J., Dao, A., Traoré, S.F., Bouaré, M., Sogoba, N. and Touré, Y.T. (2004). Malaria Transmission in Relation to Rice Cultivation in the Irrigated Sahel of Mali. *Acta Tropica*, **89**, 147-159. <u>http://dx.doi.org/10.1016/j.actatropica.2003.10.014</u>
- [22] Green, K., Kempka, D. and Lackley, L. (1994) Using Remote Sensing to Detect and Monitor Land Cover and Land Use Changes. *Photogrammetric Engineering and Remote Sensing*, **60**, 331-337.
- [23] Congalton, R.G. (1991) A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. Remote Sensing of Environment, 37, 35-46. <u>http://dx.doi.org/10.1016/0034-4257(91)90048-B</u>
- [24] Arbia, G. (1988) Spatial Data Configuration in the Statistical Analysis of Regional Economics and Related Problems. Kluwer, Dordrecht.
- [25] Jacob, B.G., Griffith, D., Gunter, J., Muturi, E.J., Caamano, E., Shililu, J., Guthure, J., Regens, J. and Novak, R.J. (2009) A Spatial Filtering Specification for an Auto-Negative Binomial Model of *Anopheles arabiensis* Aquatic Habitats. *Transactions in GIS*, **12**, 515-539. <u>http://dx.doi.org/10.1111/j.1467-9671.2008.01110.x</u>



IIIIII II

 \checkmark

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.

