

Towards Increasing Data Availability for Meteorological Services: Inter-Comparison of Meteorological Data from a Synoptic Weather Station and Two Automatic Weather Stations in Kenya

Richard Muita^{1*}, Paul Kucera², Stella Aura¹, David Muchemi¹, David Gikungu¹, Samuel Mwangi¹, Martin Steinson², Paul Oloo¹, Nicholas Maingi¹, Ezekiel Muigai¹, Mwaura Kamau¹

¹Kenya Meteorological Department, Institute for Meteorological Training and Research, Nairobi, Kenya

²University Corporation for Atmospheric Research/National Centre for Atmospheric Research (UCAR/NCAR), Boulder, USA

Email: *rukwaro2003@yahoo.co.uk

How to cite this paper: Muita, R., Kucera, P., Aura, S., Muchemi, D., Gikungu, D., Mwangi, S., Steinson, M., Oloo, P., Maingi, N., Muigai, E., & Kamau, M. (2021). Towards Increasing Data Availability for Meteorological Services: Inter-Comparison of Meteorological Data from a Synoptic Weather Station and Two Automatic Weather Stations in Kenya. *American Journal of Climate Change*, 10, 300-316.

<https://doi.org/10.4236/ajcc.2021.103014>

Received: May 7, 2021

Accepted: September 15, 2021

Published: September 18, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution-NonCommercial International License (CC BY-NC 4.0).

<http://creativecommons.org/licenses/by-nc/4.0/>



Open Access

Abstract

Meteorological data is useful for varied applications and sectors ranging from weather and climate forecasting, landscape planning to disaster management among others. However, the availability of these data requires a good network of manual meteorological stations and other support systems for its collection, recording, processing, archiving, communication and dissemination. In sub-Saharan Africa, such networks are limited due to low investment and capacity. To bridge this gap, the National Meteorological Services in Kenya and few others from African countries have moved to install a number of Automatic Weather Stations (AWSs) in the past decade including a few additions from private institutions and individuals. Although these AWSs have the potential to improve the existing observation network and the early warning systems in the region, the quality and capacity of the data collected from the stations are not well exploited. This is mainly due to low confidence, by data users, in electronically observed data. In this study, we set out to confirm that electronically observed data is of comparable quality to a human observer recorded data, and can thus be used to bridge data gaps at temporal and spatial scales. To assess this potential, we applied the simple Pearson correlation method and other statistical tests and approaches by conducting inter-comparison analysis of weather observations from the manual synoptic station and data from two Automatic Weather Stations (TAHMO and 3D-PAWS) co-located at KMD Headquarters to establish existing consistencies and variances in several weather parameters. Results show there is com-

parable consistency in most of the weather parameters between the three stations. Strong associations were noted between the TAHMO and manual station data for minimum ($r = 0.65$) and maximum temperatures ($r = 0.86$) and the maximum temperature between TAHMO and 3DPAWS ($r = 0.56$). Similar associations were indicated for surface pressure ($r = 0.99$) and RH ($r > 0.6$) with the weakest correlations occurring in wind direction and speed. The Shapiro test for normality assumption indicated that the distribution of several parameters compared between the 3 stations were normally distributed ($p > 0.05$). We conclude that these findings can be used as a basis for wider use of data sets from Automatic Weather Stations in Kenya and elsewhere. This can inform various applications in weather and climate related decisions.

Keywords

Meteorological Data, Manual Weather Station, Automatic Weather Station, Correlation

1. Introduction

Meteorological data is useful for varied applications across many socio-economic sectors. They can be used in weather and climate forecasting, disaster risk reduction and water resources management, landscape planning, and many others. However, the availability of meteorological data requires a good network of manual observation stations at the surface, upper-air, and on the ocean as well as other support systems which facilitate the collection, recording, processing, archiving, and other data management operations. In sub-Saharan Africa, such networks are limited due to low investment and capacity (Dupar et al., 2021). Such situations constrain the development, provision, and maintenance of quality climate services and their application.

In Kenya and the East African region, meteorological data is a very important resource considering that weather and climate variability are driven by several global influences including the El Niño and La Niña phenomena in the tropical Pacific, the Congo air mass, the Inter-Tropical Convergence Zone, the Indian Ocean temperatures and local climatic-factors such as the lake circulation effects among others (Marchant et al., 2007; Berhane & Zaitchik, 2014) which require regular monitoring and evaluation. The region has had its fair share of severe weather and extreme climate impacts such as flooding, hailstorms, droughts which have caused loss of human life, and other adverse socio-economic and environmental impacts. To monitor and evaluate the weather patterns in Kenya, the National Meteorological Service (KMD) operates and controls 40 Synoptic Stations spread across the country (Figure 1, <https://meteo.go.ke/>) and about 600 rain gauge stations operated by private observers. These stations are operated manually by KMD personnel on a continuous basis. Generally, the station network is sparse compared to the World Meteorological Organization's (WMO)

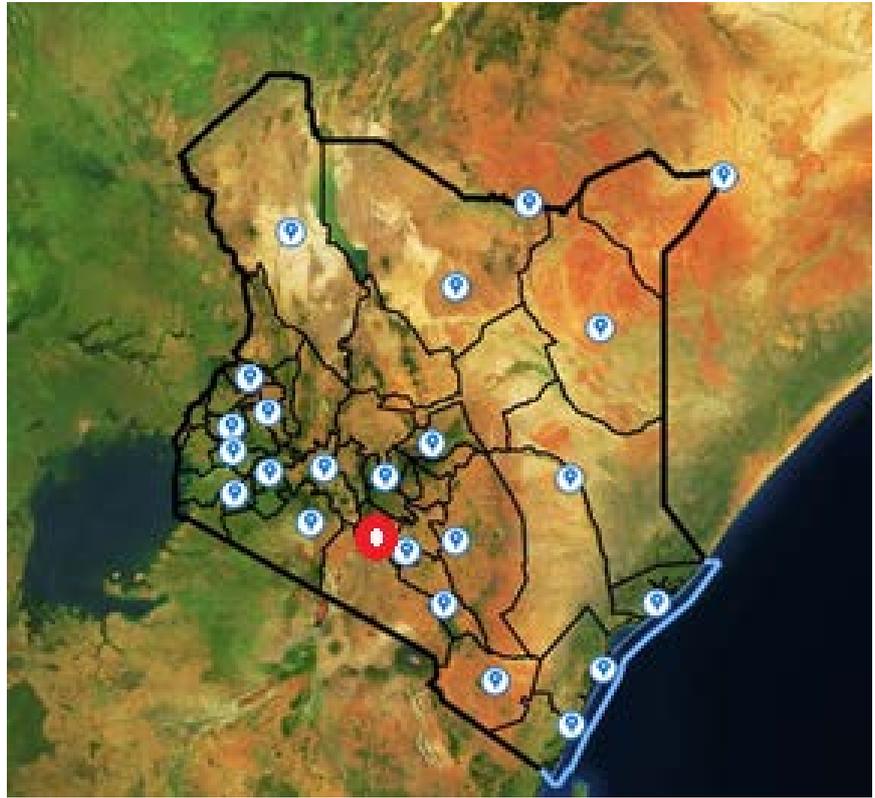


Figure 1. Map of Kenya showing location of some of the Synoptic Stations (ringed symbols) operated by KMD. The study location is indicated in red. Source: <https://meteo.go.ke/> .

recommended practice regarding the spacing between neighboring stations of 20 km. Although the KMD has in the last few years installed a number of Automatic Weather Stations (AWS), the data is not yet fully integrated into meteorological applications or shared globally through the Global Telecommunication System (GTS) of the World Meteorological Organisation (WMO) as required. This is mainly because the quality of the AWS datasets is not yet well known.

Most studies have carried out inter-comparisons of meteorological data but focused more on satellite based weather parameters and gridded data (Ayasha, 2021; Rivoire et al., 2021; Schumacher et al., 2020; Ford & Quiring, 2019; Zeng et al., 2018). These have largely ignored the significant biases that can be addressed by data from Automatic weather stations relative to surface observation stations.

The key advantage of comparisons between ground station observations and datasets from AWSs is that the datasets can provide more coverage in time and space and hence a better description of the weather and climate of a given area. Further, since ground station observations may have some uncertainties (especially when some data are missing) comparing with AWS data may bridge the gap and hence improve its quality and use. This study, therefore, provides the means to enhance the quality and quantity of available observational datasets through the calibration of the AWS data leading to improvements in early warning services. Recently in 2019, the WMO's HIGH Impact Weather LAKE

System (HIGHWAY) project funded by the United Kingdom (UK) Department for International Development (DFID) have been promoting early warning systems (EWS) to improve resilience to weather and climate extremes for the local communities around the Lake Victoria region by exploring the potential of using AWS data sets in Kenya.

Within the above context, this study aims at 1) Inter-comparison of the ground based observational data sets (rainfall, temperature, wind speed and direction, surface pressure, and relative humidity) from a KMD synoptic weather station with data from the Trans-African Hydro-Meteorological Observatory (TAHMO) AWS and the 3D-Printed Automatic Weather Station (3D-PAWS), co-located at KMD; 2) Carrying out inter-comparison of data between the two AWSs (TAHMO and 3D-PAWS).

This study is organized as follows. Section 2 describes the data sets used in the study including details of the study area. In section 3 statistical methods used in the analysis and comparison of the different data sets are presented. The results from the analyses are given in Section 4 while the discussion of the results is in Section 5.

2. Data and Study Area

2.1. Study Area Description

The datasets used in this study are from the Dagoretti Corner Meteorological Station which is located in Nairobi Kenya. All three stations are located in separate sites within the Meteorological Station compound. The manual synoptic weather station is located at Lon. 36.75°E and Lat. 1.3°S. The TAHMO AWS is at Lon. 36.7602°E and Lat. 1.3018389°S. The 3D-PAWS AWS is located at Lon. 36.7601°E and Lat. 1.30172°S. The three stations are located at an average altitude of 1790 m above mean sea level. The geographic positions of the AWSs and the synoptic weather stations clearly show that these stations are more or less collocated.

The weather regime of the study area is semi-humid tropical with average annual rainfall of 1060 mm mean annual temperature of 17.8°C, and maximum temperatures reaching about 25.5°C. The rainfall distribution is bimodal occurring in two seasons: March-May (long rains) and October-December (short rains) season. The January-February period is generally dry while June-September is cool and dry with occasional rains.

2.2. Data

The data sets used in this study are from the manual synoptic weather station, TAHMO AWS and the 3D-PAWS at the Dagoretti Corner Meteorological Station in Nairobi. The manual synoptic station data sets of daily rainfall, daily temperature at 06Z and 12Z, daily minimum and maximum temperature, daily relative humidity, hourly wind speed, hourly wind direction, surface pressure and solar radiation were acquired from the National Climate Database at KMD.

Thus the data observations from the manual synoptic weather station are on daily and hourly intervals. The TAHMO observations were obtained from KMD and are at 5 minutes intervals while 3D-PAWS observations at 1-minute intervals were obtained from the University Corporation for Atmospheric Research (UCAR).

Quality controls and checks were carried out on the data from the three sources and examined for consistency. Subsequently, the parameters selected for the comparative analysis were Rainfall, Temperature, Pressure, Relative Humidity, Solar Radiation, Wind speed and Direction. The data used in the analysis covered the period 2016 to 2018 and part of 2019.

2.3. Data Structure

Data sets for all the parameters from the manual station were prepared in a single “.csv” file with daily values in a cross tab format while the TAHMO data were organized in five-minute values in a list in multiple files summarized in one “.csv” file per day. The 3D-PAWS data in ASCII format were stored in separate files for each sensor (e.g., humidity and temperature, rainfall, wind direction and speed, and surface pressure) with a resolution of 1-min records. All the AWS data from 3D-PAWS and TAHMO were processed to match the temporal resolution of the manual station by aggregating the minute and hourly data into daily values using the R-statistical software (R Core Team, 2017) and Excel. To enable the analysis, the specific parameters were prepared as follows.

2.4. Rainfall

To match the daily observation period (24 hours) of the manual synoptic weather station, the rainfall data from TAHMO and 3D-PAWS were accumulated to daily values starting from 0600Z of the current day to 0600Z of the next day and cast back by one day.

2.5. Surface Pressure

The observed surface pressure from the manual synoptic station is at hourly timescale and therefore the AWSs surface pressure data from TAHMO and 3D-PAWS were processed to extract the observation from the top of each hour to match the observation time of the manual synoptic station.

2.6. Relative Humidity

The manual synoptic weather station data were at 0600Z and 1200Z and therefore AWSs observations from TAHMO and 3D-PAWS were matched for the 0600Z and 1200Z times to match the observations from the manual synoptic weather station.

2.7. Solar Radiation

The hourly manual synoptic weather station daily total radiation data were not

considered in the analysis since they could not be matched with the observations from the 1-minute and 5-minute temporal resolutions from the 3D-PAWS and TAHMO stations.

2.8. Wind Speed

The wind speed data for the manual synoptic weather station was at the hourly time stamp and hence the AWSs data from TAHMO and 3DPAWS were matched at the top of each hour with the manual synoptic station observations. However, due to its structure, the 3D-PAWS data were aggregated to average the 10-minute observations before the top of the hour so as to match the procedure used for the manual observations.

2.9. Wind Direction

Similar to other hourly data sets the wind direction data from the manual synoptic weather station were matched with the AWSs data from TAHMO and 3D-PAWS at the top of the hour. The 3D-PAWS data were also processed to match the manual observations using similar procedure as that of the wind speed.

2.10. Temperature (Dry Bulb, T_{\max} , T_{\min})

The dry bulb temperature data from the manual synoptic weather station were at 0600Z and 1200Z where maximum (T_{\max}) and minimum (T_{\min}) temperature data were extracted as single values for each day. To match, the AWSs temperature observations from TAHMO and 3D-PAWS were extracted at 0600Z and 1200Z on each day of the inter-comparison. Similar to solar radiation, the AWSs T_{\max} and T_{\min} were extracted from the 1-minute and 5-minute temperature values from 3D-PAWS and TAHMO, respectively for any given day. However, the T_{\max} and T_{\min} values for the TAHMO and 3D-PAWS were only extracted and matched with the manual data if a complete record was available for a given day.

2.11. Data Processing, Quality Checks and Controls

Missing data were not indicated after the initial extraction of the different parameters from the 3 data sources, leading to data sets of varying lengths. To correct this, identification and insertion of gaps were done in all the parameters until they all had equal lengths. However, the rainfall data from the manual synoptic station was in a cross tab table format (**Figure 2**) which was converted to a list to enable inter-comparison with the AWS datasets before the gaps were inserted. The gaps that were identified in all the datasets were filled using “NA”.

Further, the 3D-PAWS data files were in “.xml” format which was converted to the “.csv” format prior to extraction and aggregation. In addition, the 3D-PAWS data posed some challenges in aggregating and selecting the right values since they contained multiple values for each top of the hours. This led to

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Stn ID	Name	Lat	Long	Year	Month	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	9136164	DAGORET	-1.3013	36.75968	2015	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	9136164	DAGORET	-1.3013	36.75968	2015	2	0	0	0	0	0	0	0	0	0	0	0	0	0	5.5
4	9136164	DAGORET	-1.3013	36.75968	2015	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	9136164	DAGORET	-1.3013	36.75968	2015	4	0	29.1	13.4	25.9	3.1	1.9	0	27.3	0	0	0	0	20.4	1.5
6	9136164	DAGORET	-1.3013	36.75968	2015	5	34.1	0	0	1.7	5	2.3	0.2	5.1	0	0	73.1	116	10.7	0
7	9136164	DAGORET	-1.3013	36.75968	2015	6	25.6	12.6	12.2	1.3	100.1	8.7	0	0	0	0	0	0	0	5.6
8	9136164	DAGORET	-1.3013	36.75968	2015	7	0	0	0	0	0	9.7	0	0	0	0	0	0	0	0
9	9136164	DAGORET	-1.3013	36.75968	2015	8	0	0	0	0	0	0.1	1.5	1.3	0.4	0	0	0	0	93
10	9136164	DAGORET	-1.3013	36.75968	2015	9	1.3	0	1.6	0	0	0	0	0	0	0	0	9.3	0	0
11	9136164	DAGORET	-1.3013	36.75968	2015	10	0	0	0	0	0	0	0	0	0	0	0	9.5	0	0
12	9136164	DAGORET	-1.3013	36.75968	2015	11	2.5	1.3	0	15	6.4	0	0	0	32.6	31.6	5.4	0	5.2	15
13	9136164	DAGORET	-1.3013	36.75968	2015	12	0	15.6	0.3	0	0	0	0	0	0	0	0	16.9	6.1	10.7
14																				

Figure 2. Sample of rainfall data from the manual station in a crossbar format.

the 3D-PAWS data being reprocessed into consistent, continuous ASCII formatted files that included missing records and bad data flags. Once this was achieved, it was possible to consistently compare the reformatted 3D-PAWS data (see Figure 3) and TAHMO data files in similar formats. From this process, the gaps in the AWSs datasets should be considered to be potential sources of errors in comparison analysis and data checks and tests of meteorological observations.

2.12. Analysis Methods

To address the objectives of this study, we used several statistical methods to assess and carry out inter-comparison analyses. Graphical methods were used to visualize the data through line and scatter plots. To visually define a matching point between the different distributions of a data series from the three data sources, we checked on concurrent points that did not overlap. To gain better comparisons, each variable (e.g. rainfall) was plotted for all the three sources (TAHMO, 3D-PAWS and Manual) on one graph.

We used the simple Pearson correlation method to assess the strength of association between the variables from the three sources (Zou et al., 2003). The Pearson correlation method measures the linear correlation between two variables X and Y using the correlation coefficient (r) and is given by the equation:

$$r_{XY} = \frac{N \sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{N \sum X_i^2 - (\sum X_i)^2} \sqrt{N \sum Y_i^2 - (\sum Y_i)^2}} \tag{1}$$

where, r_{XY} or r is the correlation coefficient, N is the number of observations, and $\sum X$ is the sum of x scores (values), $\sum Y$ is the sum of y scores (values), while $\sum XY$ is the sum of the products of x and y values. The values of r range between +1 and -1, with 1 showing that there is a perfect/positive linear correlation, 0 showing no linear correlation, and -1 showing there is a negative linear correlation. The higher the value of r is, the stronger the association between two variables.

We tested the statistical significance of the correlation coefficients (r) using the Shapiro-Wilk normality test (Emerson, 2015) and the Anderson-Darling normality test (Liebscher, 2016). Shapiro-Wilk’s method is based on the correlation between the data and the corresponding normal scores and is widely

1	%Month	Day	Year	Tmax	Tmin	RHmax	RHmin	T_00	RH_00	T_06	RH_06	T_12	RH_12	T_18	RH_18
2	03	16	2017	-998.0	-998.0	-998.0	-998.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0
3	03	16	2017	-998.0	-998.0	-998.0	-998.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0
4	03	17	2017	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0
5	03	18	2017	-998.0	-998.0	-998.0	-998.0	-999.0	-999.0	-999.0	-999.0	29.0	29.6	23.0	29.6
6	03	19	2017	29.9	17.8	88.0	21.3	18.3	86.1	20.5	71.6	28.9	26.0	22.5	26.0
7	03	20	2017	30.2	17.2	90.3	14.6	18.3	85.2	20.2	70.8	29.4	21.8	22.6	21.8
8	03	21	2017	31.0	16.4	88.5	25.4	16.9	70.4	21.2	68.6	29.6	26.1	22.6	26.1
9	03	22	2017	29.7	16.4	92.4	27.0	18.5	87.3	21.0	68.6	28.3	29.8	16.8	29.8
10	03	23	2017	29.7	16.2	94.7	22.7	16.4	93.7	20.0	74.5	26.6	30.6	22.0	30.6
11	03	24	2017	31.3	16.2	90.8	12.5	17.9	66.7	20.5	64.6	30.0	18.5	22.6	18.5
12	03	25	2017	31.1	17.1	91.2	16.3	18.3	79.8	22.4	67.1	30.2	21.3	23.7	21.3
13	03	26	2017	31.5	16.3	95.2	22.6	18.6	76.5	20.8	68.1	31.2	24.0	23.8	24.0
14	03	27	2017	-998.0	-998.0	-998.0	-998.0	16.8	94.6	19.5	79.5	29.3	28.0	21.2	28.0
15	03	28	2017	28.0	13.7	97.3	35.5	16.4	93.7	18.5	83.7	27.6	36.3	18.2	36.3
16	03	29	2017	27.7	16.1	94.7	38.2	16.3	91.8	18.4	85.5	26.2	43.2	19.2	43.2
17	03	30	2017	-998.0	-998.0	-998.0	-998.0	17.4	87.9	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0
18	03	31	2017	-998.0	-998.0	-998.0	-998.0	-999.0	-999.0	-999.0	-999.0	27.2	31.7	20.5	31.7
19	04	01	2017	28.5	15.8	91.8	19.2	16.9	70.7	20.5	65.6	27.6	29.0	21.0	29.0
20	04	02	2017	29.7	16.2	87.8	9.2	17.8	67.7	21.1	60.9	28.6	12.0	22.7	12.0
21	04	03	2017	29.1	16.5	90.8	24.6	17.5	85.2	19.8	69.6	29.1	24.6	21.8	24.6
22	04	04	2017	29.5	16.9	87.6	24.1	17.2	85.0	19.4	77.1	28.7	28.4	22.2	28.4
23	04	05	2017	29.8	15.5	96.5	24.3	16.9	87.5	21.6	63.0	28.9	28.3	21.2	28.3

Figure 3. Screen shot of reformatted 3D-PAWS data (for Temperature and RH).

used and recommended for normality tests because it provides better estimates compared to the Kilmogorov-Smirnov method. However, the Anderson-Darling method is efficient in analysis of samples with $N > 5000$ compared to the Shapiro Wilk method ($N < 5000$).

These normality tests are done with the assumptions that for a test that is not significant it satisfies the condition of the null hypothesis for normally distributed sample (p value > 0.05). Alternatively, if the test is significant, the distribution is considered to be non-normal. All correlations are considered to be significant at the 0.05 level.

To support the analysis of normality tests we used the graphical method to visualize the associations between the variables.

In particular, we used Q-Q (quantile-quantile) plots to assess how best the compared data set samples associated with the normal distribution. Overall, we summarized and displayed the coefficients from the comparisons using a correlation matrix.

Lastly, we used regression method to estimate the best line of fit from the correlations between the variables from the three data sources.

3. Results

3.1. Pre-Analysis Results

The initial analysis of the 3 datasets for the period 2017-2018, indicated that there was good agreement between the temperatures readings of the manual station and TAHMO but comparatively lower agreement with the 3D-PAWS. Comparisons for rainfall and relative humidity were largely variable. One reason for this might be that the TAHMO AWS had a broken rain-gauge sensor, for a brief period between 2017 and June 2018.

3.2. Time Series Comparisons

Figure 4 displays the variability in the time series of the T_{\max} , T_{\min} and T06, RH06 for the manual station and 3DPAWS from 2017-2019. The graph shows that only few of the values of T_{\max} , T_{\min} , and T06, RH06 overlap for both stations. This was apparently due to battery failure at night and thus more observations for temperature and RH at 12Z from the 3D-PAWS station which was mostly operational during daytime hours.

Figure 5 compares the time series of minimum temperatures and maximum temperatures for 3 data sources for the periods with consistent data. The temporal patterns indicated in the three data sets show reasonable agreements although the 3D-PAWS minimum temperature indicates a high bias of about 10°C higher compared to the TAHMO and manual station series. As earlier mentioned, the battery on the 3D-PAWS was not working properly during the non-daylight hours of the inter-comparison period. This problem calls for continuous monitoring and maintenance of AWS and all other stations to ensure measurements continue. Due to this, the reprocessing and computing of the temperature for the 3D-PAWS significantly reduced the number of matching records but also produced fair comparison between the stations. Similarly, the patterns of the maximum temperature indicate some disparities in some periods where all the three data sets are observed. For instance, there are considerable agreements for the maximum temperature between the manual station and 3D-PAWS save for the manual temperatures indicating some spikes and missing data. These results show that the consistent concurrence for the maximum temperature between 3D-PAWS and the manual station was because 3D-PAWS was working more efficiently during daytime hours relative to night-time hours.

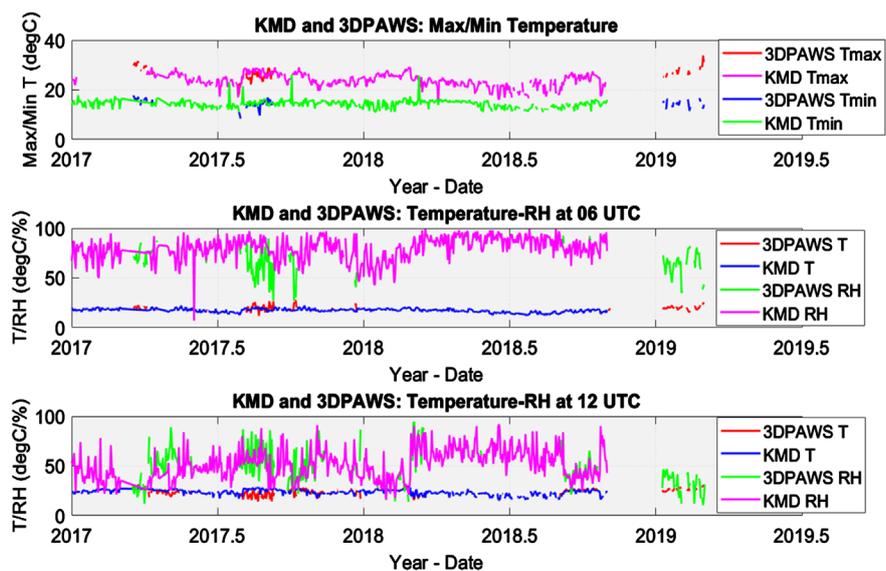


Figure 4. Time series of Tmax/Tmin from KMD and 3D-PAWS (top panel), temperature and RH at 06 UTC for KMD and 3D-PAWS (middle panel), and temperature and RH at 12 UTC for KMD at 3D-PAWS.

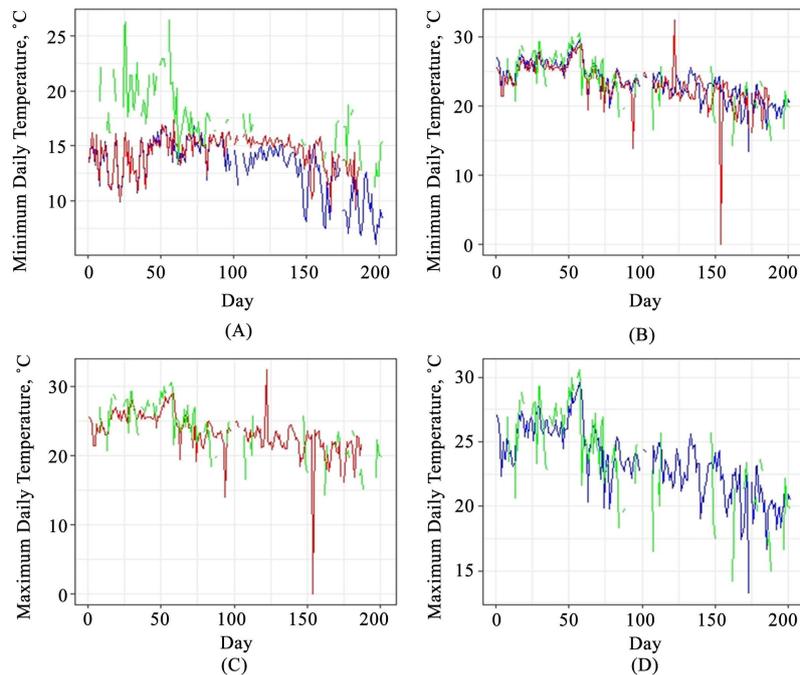


Figure 5. Comparisons for (A) Daily minimum temperature (Dagoretti (red), TAHMO (blue), 3D-PAWS (green)) within the period 2017/2018; (B) Daily maximum temperature between Dagoretti, TAHMO and 3D-PAWS; (C) Daily maximum temperature between Dagoretti and 3D-PAWS; (D) Between TAHMO and 3D-PAWS.

3.3. Correlations Results

Modestly higher correlations were observed for the minimum temperature between TAHMO and manual station ($r = 0.65$) and for the maximum temperature ($r = 0.61$ to $r = 0.86$) (Figure 6). The correlations for maximum temperature for TAHMO and 3D-PAWS were modest ($r = 0.56$).

The Shapiro Wilk normality test showed that the distributions of the minimum and maximum temperatures are not significantly different from normal distribution and hence normally distributed (p -value < 0.05). For example, the maximum temperature for the manual station had the best normal distribution pattern compared to the other data sets.

Most of the other meteorological variables from the 3 different stations indicated positive correlations. Strong correlations were indicated in the relative humidity at 1200Z between the manual and 3D-PAWS ($r = 0.59$) with lowest correlations being observed at 0600Z. The surface pressure and relative humidity displayed normal distribution compared to the other variables such as rainfall which is not normally distributed for all the 3 stations and can be described using non-linear distribution (Figure 7). Rainfall in Kenya has largely followed a log-normal distribution and other exponential distributions. Subsequently, surface pressure at the manual stations was highly correlated with the TAHMO ($r = 0.67$, $p < 0.05$) and 3D-PAWS ($r = 0.65$, $p < 0.005$) and between the two AWS datasets (Figure 9, $p = 0.99$). Similarly, several variables indicated normal distribution patterns which were statistically significant (< 0.05 , Table 1).

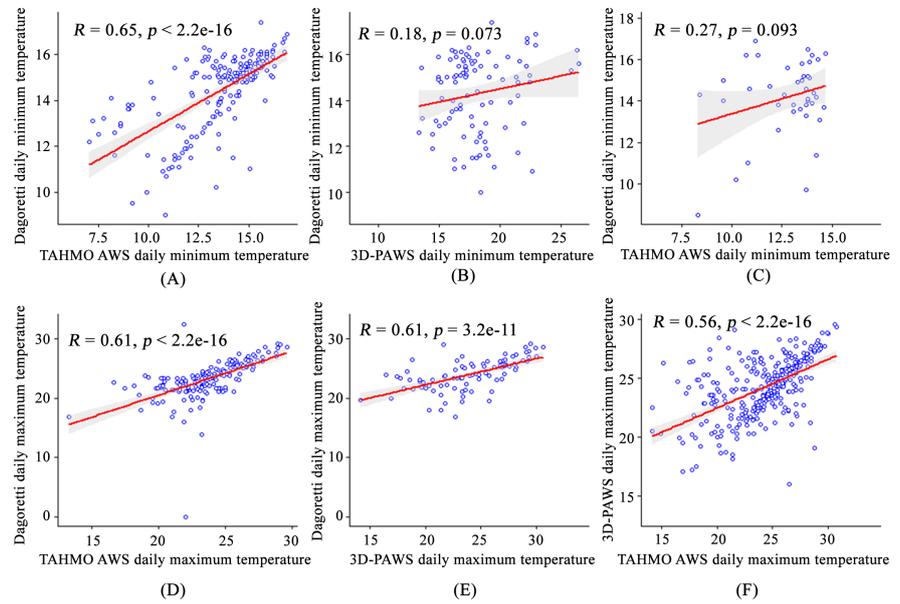


Figure 6. Correlations for daily minimum and maximum temperature between manual and TAHMO ((A) and (D)), manual and 3D-PAWS ((B) and (E)) and 3D-PAWS and TAHMO ((C) and (F)). The minimum and maximum temperatures are fitted with a linear regression line, while the gray envelope is the 95% confidence interval.

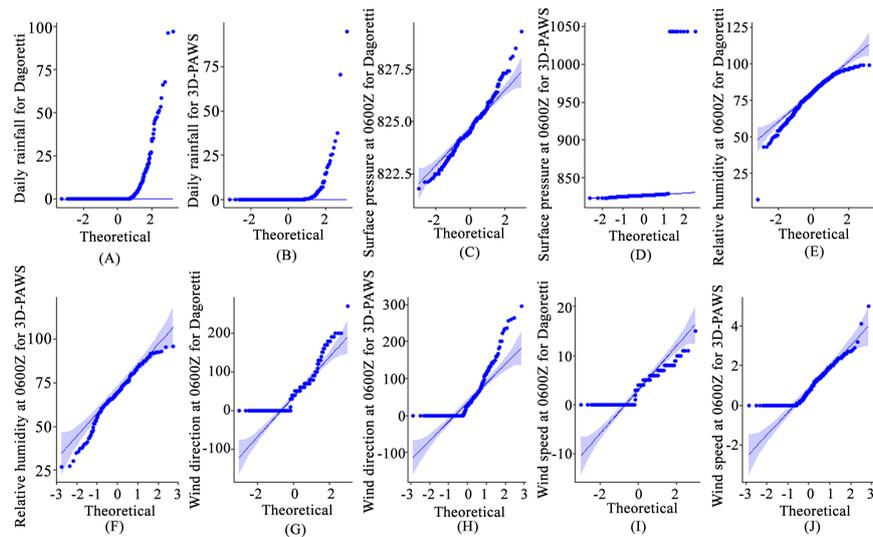


Figure 7. Q-Q plots for normality test for daily rainfall ((A) and (B)), surface pressure at 006Z ((C) and (D)), relative humidity at 006Z and at 12Z ((E) and (F)), wind direction and wind speed at 006Z between manual (Dagoretti) and 3D-PAWS (RH_3D_12Z) stations (G, H, I, and J).

The correlations for wind speed and direction for the manual, 3D-PAWS and TAHMO were fairly strong. To validate this, the wind rose between 3D-PAWS and the manual station indicated consistent and strongly NE winds at the manual station location (Figure 8). However, the wind speeds were higher at the manual station than at the AWSs possibly due to height differences. This may need to be further examined.

Some earlier results had indicated that there was a very strong positive correlation for surface pressure between manual and TAHMO ($r = 0.67$), between TAHMO and 3D-PAWS ($r = 0.96$) and between manual and 3D-PAWS ($r = 0.65$) (Table 2 and Figure 9).

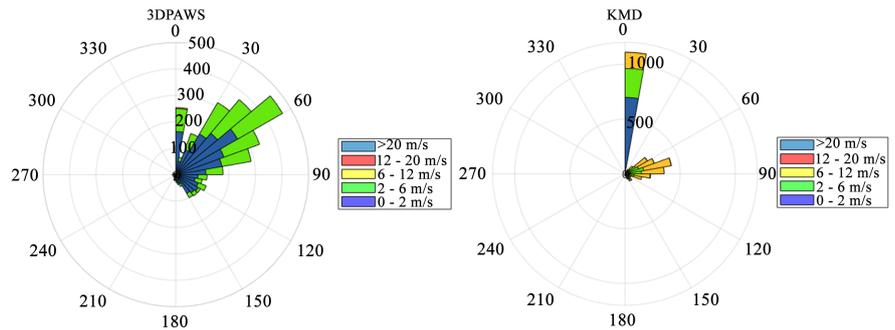


Figure 8. Windrose for 3D-PAWS (left panel) and KMD (right panel)

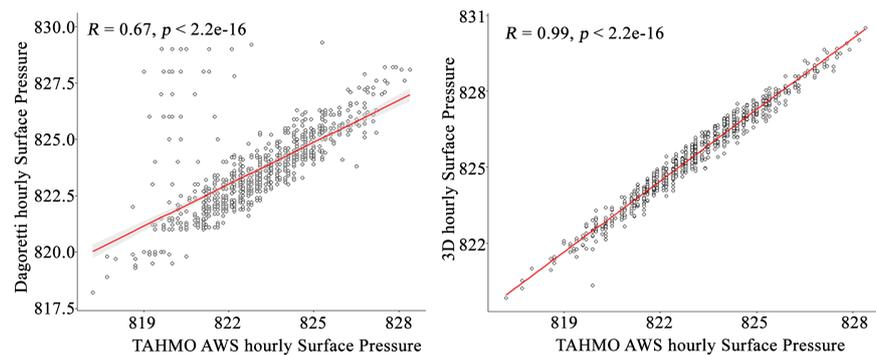


Figure 9. Correlation for hourly surface pressure between manual station and TAHMO AWS observations (left) and between the TAHMO and 3D-PAWS (right) and fitted with a linear regression line at the 95% confidence interval.

Table 1. Shapiro-Wilk normality test for comparisons between manual station, TAHMO and 3D-PAWS meteorological variables.

Variable	W	p-value
Dagoretti daily rainfall	0.32531	<2.2e-16
3D-PAWS daily rainfall	0.18067	<2.2e-16
Dagoretti surface pressure	0.98414	0.0004962
3D-PAWS surface pressure	0.35564	<2.2e-16
Dagoretti relative humidity at 006Z	0.94811	5.348e-14
3D-PAWS relative humidity at 006Z	0.96837	0.0008388
Dagoretti wind direction at 006Z	0.81444	<2.2e-16
3D-PAWS wind direction at 006Z	0.81444	<2.2e-16
Dagoretti wind speed at 006Z	0.84307	<2.2e-16
3D-PAWS wind speed at 006Z	0.87698	5.147e-13

Table 2. Summary of correlation coefficients for Dagoretti minimum and maximum temperatures between manual, TAHMO AWS and 3D-PAWS observations.

Parameter (variable)	Correlation variables	Correlation coefficient (r)
1) Minimum temperature	Manual vs TAHMO AWS	0.651
	Manual vs 3D-PAWS	0.180
	TAHMO AWS vs 3D-PAWS	0.270
2) Maximum temperature	Manual vs TAHMO AWS	0.610
	Manual vs 3D-PAWS	0.610
	TAHMO AWS vs 3D-PAWS	0.560
3) Relative humidity at 06Z	Manual vs TAHMO AWS	0.832
	Manual vs 3D-PAWS	0.220
	TAHMO AWS vs 3D-PAWS	0.621
4) Relative humidity at 12Z	Manual vs TAHMO AWS	0.765
	Manual vs 3D-PAWS	0.590
	TAHMO AWS vs 3D-PAWS	0.858
5) Wind direction	Manual vs TAHMO AWS	0.0793
	Manual vs 3D-PAWS	0.0004
	TAHMO AWS vs 3D-PAWS	0.495
6) Wind speed	Manual vs TAHMO AWS	0.0054
	Manual vs 3D-PAWS	0.3636
	TAHMO AWS vs 3D-PAWS	0.0144
7) Surface pressure	Manual vs TAHMO AWS	0.6713
	Manual vs 3D-PAWS	0.6500
	TAHMO AWS vs 3D-PAWS	0.9859
8) Rainfall	Manual vs 3D-PAWS	0.3764

4. Discussion and Conclusions

Rainfall and other meteorological parameters such as temperature are invaluable for not only the monitoring and forecasting of the weather but also management of climate related disasters. However, sparseness of observation stations for measurements and collection of these data especially in sub-Saharan Africa and other developing countries is a major challenge. National Meteorological and Hydrological Services globally have been the main meteorological data collecting institutions. There is a need for enhancement of the observation network, through establishment of more automatic weather stations in most countries, as well as improved monitoring and prediction of weather and climate patterns.

Our study of manual and AWS data sets has demonstrated that the two modes of weather observations compare well and can therefore guide decision making. The analyses of meteorological data and comparisons between the manual station and AWSs revealed considerable agreements between most of the weather

parameters in spite of the low correlations found between the rainfall and wind observations compared to the other variables (**Table 2**). This is in agreement with other findings that have shown that meteorological parameters measured from a ground station can compare relatively well with other observations from a reference source e.g. AWS (Dombrowski et al., 2021).

Whereas the surface pressure was highly correlated between the three stations, some studies have found uncertainties in the comparisons between on ground station pressure and the reference data, e.g. (Dombrowski et al., 2021). The high variability between some meteorological variables at different stations could be due to several factors such as instrument error, change of location, damages or lack of station maintenance (Ford et al., 2020).

Overall, comparison of the manual station data and the TAHMO and 3D-PAWS observations showed that there was potential for concurrences between the different variables even at some small spatial co-locations of the different measurement instruments (**Table 2**).

Despite the strong correlations between the different variables, anomalies were present when assessing some parameters such as wind speed and solar radiation due to complexities in aggregating such observations between the manual and automatic stations. This provides a challenge where such parameters may be required to complement monitoring and forecasting of the weather.

Based on our findings, the different meteorological parameters compared reasonably well between the three stations. The correlation coefficients between the parameters from the manual station and the AWSs were within acceptable levels and can be used as a basis for validation and application of the data in forecasting and other uses. There is high potential from the findings that observations from the 3D-PAWS and TAHMO stations compared well when both were in operation and relative to the manual station.

In conclusion, manual station datasets can be used alongside observations from AWS after adequate assessment of the quality and agreements between the data sets have been done.

Acknowledgements

In addition to the WMO HIGHWAY project for facilitating the two workshops that enabled the organisation and analysis of the datasets, we recognize the UK Aid (UK Met Office) support. We would also like to thank the Kenya Meteorological Department, Trans African Meteorological and Hydrological Observatories (TAHMO), and the University Corporation for Atmospheric Research/National Centre for Atmospheric Research (UCAR/NCAR) for providing the data that was used in this study.

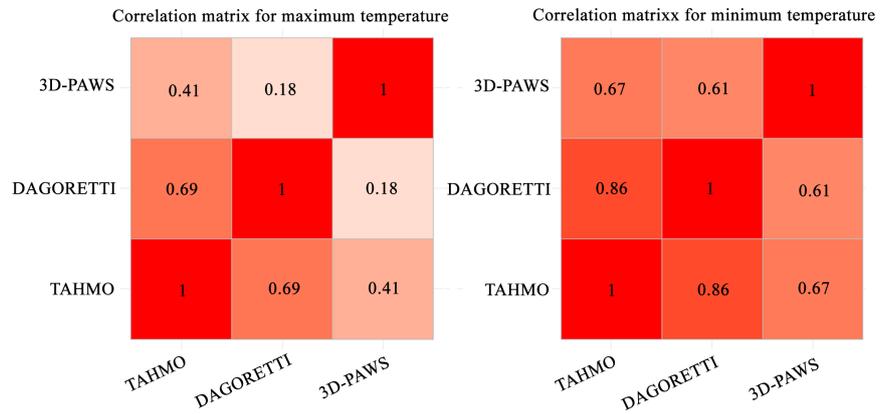
Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

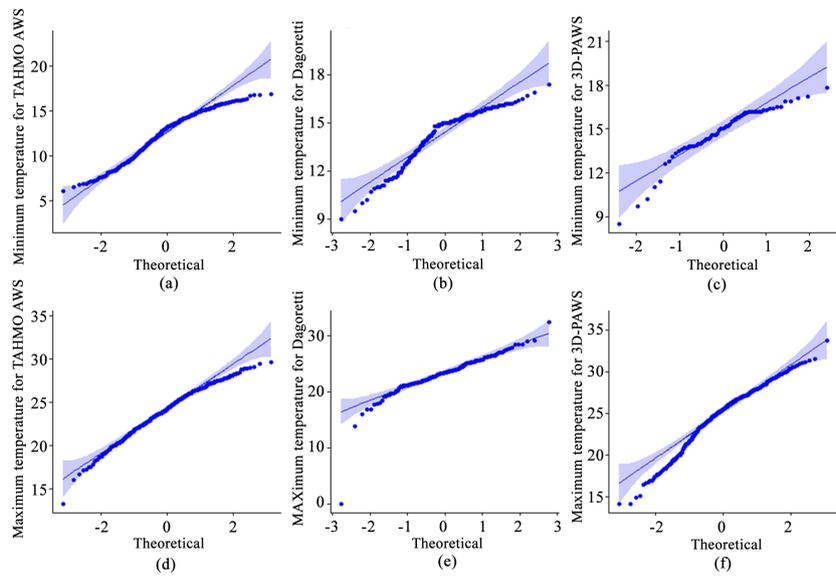
References

- Ayasha, N. (2021). A Comparison of Rainfall Estimation Using Himawari-8 Satellite Data In Different Indonesian Topographies. *International Journal of Remote Sensing and Earth Sciences (IJReSES)*, 17, 189-200. <https://doi.org/10.30536/ijreses.2020.v17.a3441>
- Berhane, F., & Zaitchik, B. (2014). Modulation of Daily Precipitation over East Africa by the Madden-Julian Oscillation. *Journal of Climate*, 27, 6016-6034. <https://doi.org/10.1175/JCLI-D-13-00693.1>
- Dombrowski, O., Hendricks Franssen, H. J., Brogi, C., & Bogena, H. R. (2021). Performance of the ATMOS41 All-In-One Weather Station for Weather Monitoring. *Sensors*, 21, Article No. 741. <https://doi.org/10.3390/s21030741>
- Dupar, M., Weingärtner, L., & Opitz-Stapleton, S. (2021). Investing for Sustainable Climate Services: Insights from African Experience. <http://www.indiaenvironmentportal.org.in/files/file/sustainability%20of%20climate%20services.pdf>
- Emerson, R. W. (2015). Causation and Pearson's Correlation Coefficient. *Journal of Visual Impairment & Blindness*, 109, 242-244. <https://doi.org/10.1177/0145482X1510900311>
- Ford, T. W., & Quiring, S. M. (2019). Comparison of Contemporary *in situ*, Model, and Satellite Remote Sensing Soil Moisture with a Focus on Drought Monitoring. *Water Resources Research*, 55, 1565-1582. <https://doi.org/10.1029/2018WR024039>
- Ford, T. W., Quiring, S. M., Zhao, C., Leasor, Z. T., & Landry, C. (2020). Triple Collocation Evaluation of *In Situ* Soil Moisture Observations from 1200+ Stations as part of the US National Soil Moisture Network. *Journal of Hydrometeorology*, 21, 2537-2549. <https://doi.org/10.1175/JHM-D-20-0108.1>
- Liebscher, E. (2016). Approximation of Distributions by Using the Anderson Darling Statistic. *Communications in Statistics—Theory and Methods*, 45, 6732-6745. <https://doi.org/10.1080/03610926.2014.966844>
- Marchant, R., Mumbi, C., Behera, S., & Yamagata, T. (2007). The Indian Ocean dipole—The Unsung Driver of Climatic Variability in East Africa. *African Journal of Ecology*, 45, 4-16. <https://doi.org/10.1111/j.1365-2028.2006.00707.x>
- R Core Team (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org/>
- Rivoire, P., Martius, O., & Naveau, P. (2021). A Comparison of Moderate and Extreme ERA-5 Daily Precipitation with Two Observational Data Sets. *Earth and Space Science*, 8, e2020EA001633. <https://doi.org/10.1029/2020EA001633>
- Schumacher, V., Justino, F., Fernández, A., Meseguer-Ruiz, O., Sarricolea, P., Comin, A. et al. (2020). Comparison between Observations and Gridded Data Sets over Complex Terrain in the Chilean Andes: Precipitation and Temperature. *International Journal of Climatology*, 40, 5266-5288. <https://doi.org/10.1002/joc.6518>
- Zeng, Q., Wang, Y., Chen, L., Wang, Z., Zhu, H., & Li, B. (2018). Inter-Comparison and Evaluation of Remote Sensing Precipitation Products over China from 2005 to 2013. *Remote Sensing*, 10, Article No. 168. <https://doi.org/10.3390/rs10020168>
- Zou, K. H., Tuncali, K., & Silverman, S. G. (2003). Correlation and Simple Linear Regression. *Radiology*, 227, 617-628. <https://doi.org/10.1148/radiol.2273011499>

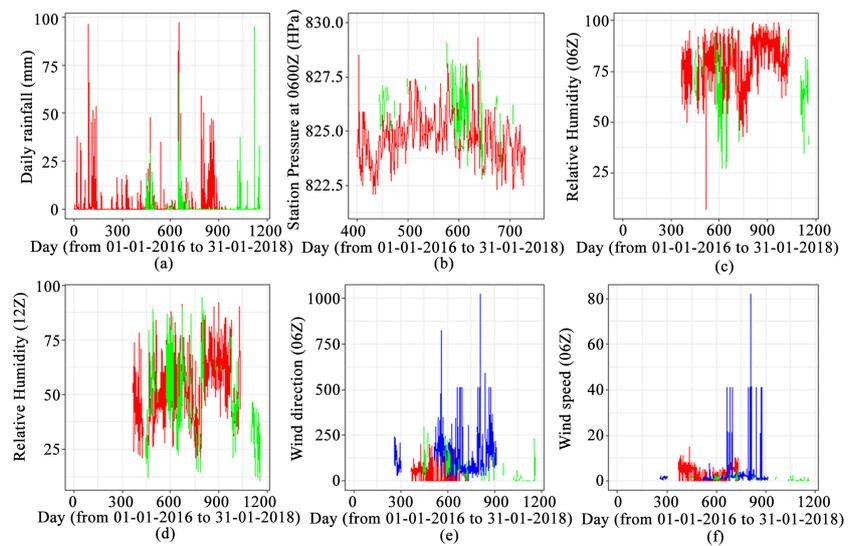
Supplementary Materials



(A)



(B)



(C)

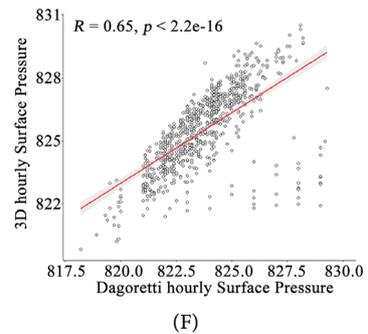
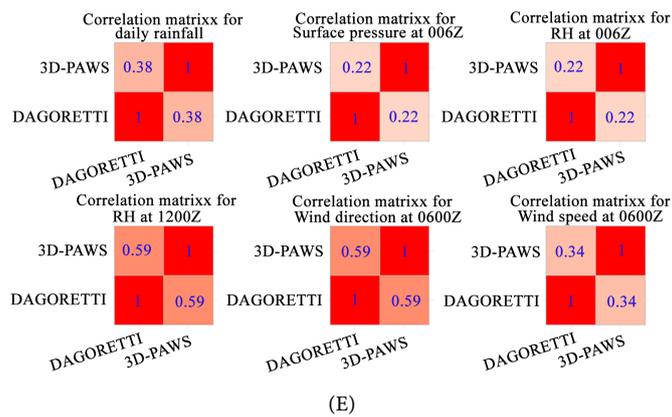
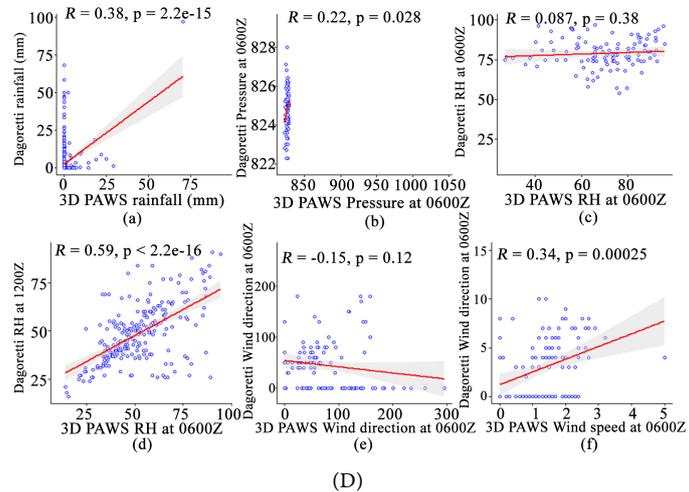


Figure S1. (A): Correlation matrix for minimum and maximum temperature for 3D-PAWS, Dagoretti (manual) and TAHMO; (B): QQ plots for minimum temperature for (a) TAHMO (b) Manual (Dagoretti) and (c) 3D-PAWS and for maximum temperature for (d) TAHMO (e) Manual (Dagoretti) and (f) 3D-PAWS; (C): Comparisons of daily rainfall (a), surface pressure at 0600Z (b), relative humidity at 0600Z (c), relative humidity at 1200Z, wind direction at 006Z and wind speed at 006Z between manual (Dagoretti) station and 3D-PAWS; (D): Comparisons of correlations for daily rainfall, surface pressure at 006Z, relative humidity at 006Z, relative humidity at 12Z, wind direction and wind speed at 006Z between manual (Dagoretti) and 3D-PAWS (RH_3D_12Z) stations; (E): Correlation matrix for daily rainfall, surface pressure at 006Z, relative humidity at 006Z and at 12Z, wind direction and wind speed at 006Z between manual (Dagoretti) and 3D-PAWS (RH_3D_12Z) stations; (F): Correlation for hourly surface pressure between manual station and 3D-PAWS observations fitted with a linear regression line at the 95% confidence interval.