



WORLD
METEOROLOGICAL
ORGANIZATION



INTERCOMPARISON OF TAHMO, 3DPAWS AUTOMATIC WEATHER STATIONS AND KMD'S DAGORETTI CORNER, NAIROBI SYNOPTIC WEATHER STATION DATA



WMO HIGHWAY PROJECT

Contributors:

Stella Aura, Dr Richard Muita, Paul Oloo, David Muchemi, Dr Paul Kucera, Samuel Mwangi, Nicholas Maingi, Ezekiel Muigai, and Martin Steinson

KENYA METEOROLOGICAL DEPARTMENT
31st October 2019

Table of Contents

INTERCOMPARISON OF TAHMO, 3DPAWS AUTOMATIC WEATHER STATIONS AND KMD SYNOPTIC WEATHER STATIONS DATA AT DAGORETTI CORNER, NAIROBI	1
Executive Summary	4
1.0. Introduction	6
1.2. Workshop objectives	8
1.3. Justification	8
1.3.1 Study Area	8
2.0. Data Collection and Preparation	9
2.1. Data Collection	9
2.2. Data Presentation	9
i) Precipitation Data	10
ii) Surface Pressure	10
iii) Relative Humidity	10
iv) Radiation	10
v) Wind Speed	10
vi) Wind Direction	10
vii) Temperature (Dry bulb, Tmax, Tmin)	11
2.3. Data Processing, Reformatting and Gaps Insertion	11
3.0. Analysis Methods	13
4.0. Results of the Analysis	14
4.1. Pre-Analysis Results	14
4.2. Updated Results	15
4.3. Summary and Conclusion	25
4.4. Recommendations	25
5.0. References	26

List of Figures

Fig 1: Initial comparisons of daily relative humidity at 12Z for manual (Dagoretti) station (red), TAHMO AWS (blue) and 3D-PAWS observations for the period 2016 – 2018 (left) and between daily minimum temperature for manual (red) and TAHMO AWS (blue) observations for the period 2016 – 2018 (right)	15
Fig 2: Time series of Tmax/Tmin from KMD and 3DPAWS (top panel), temperature and RH at 06 UTC for KMD and 3DPAWS (middle panel), and temperature and RH at 12 UTC for KMD at 3DPAWS.....	17
Fig 3: Comparisons for (A) daily minimum temperature (Dagoretti (red), TAHMO (blue), 3DPAWS (green) for the period 2017/2018, (B) daily maximum temperature between Dagoretti, TAHMO and 3DPAWS, (C) daily maximum temperature between Dagoretti and 3DPAWS, and (D) between TAHMO and 3DPAWS.....	17
Fig 4: Correlations for daily minimum and maximum temperature between manual and TAHMO (A and D), manual and 3D-PAWS (B and E) and 3DPAWS and TAHMO (C and F). The minimum and maximum temperatures are fitted with a linear regression line (blue), while the gray envelope is the 95% confidence interval	18
Fig 5: Correlation matrix for minimum and maximum temperature for 3D-PAWS, Dagoretti (manual) and TAHMO	18
Fig 6: QQ plots for minimum temperature for (A) TAHMO (B) Manual (Dagoretti) and (C) 3DPAWS and for maximum temperature for (D) TAHMO (E) Manual (Dagoretti) and (F) 3DPAWS	19
Fig 7: 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 3DPAWS	20
Fig 8: 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	21
Fig 9: 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	21
Fig 10: 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).	22
Fig 11: Windrose for 3DPAWS (left panel) and KMD (right panel).....	22
Fig 12: Correlation for hourly surface pressure between manual (Dagoretti) and TAHMO AWS observations fitted with a linear regression line (blue) at the 95% confidence interval.....	24
Fig 13: Correlation for hourly surface pressure between TAHMO AWS and 3D-PAWS observations fitted with a linear regression line (blue) at the 95% confidence interval.....	24
Fig 14: Correlation for hourly surface pressure between manual (Dagoretti) and 3D-PAWS observations fitted with a linear regression line (blue) at the 95% confidence interval.....	25

Executive Summary

The World Meteorological Organization's (WMO) HIGH impact Weather LAke SYstem (HIGHWAY) Project is a United Kingdom (UK) Department for International Development (DFID) funded project, which is aimed at increasing the resilience of the communities around the Lake Victoria basin to extreme weather and events such as floods, storms and droughts by improving on the early warning systems (EWS).

Two workshops were conducted to carry out the inter-comparison of the data from the 3D-Printed Automatic Weather Stations (3D-PAWS and Trans-African Hydrological Observatories (TAHMO) Automatic Weather Stations (AWS) with the manual observations from the KMD Dagoretti manual station in Nairobi as part of the Lake Victoria HIGHWAY Field Campaign. This was aimed at using data from the KMD manual weather observations to validate the TAHMO and 3DPAWS data for application in forecasting and other early warning activities. The availability of consistent good quality datasets from manual and automatic weather observation stations will enable the HIGHWAY project to enhance the high-resolution regional Numerical Weather Prediction (NWP) model products for the Lake Victoria region.

The data from the manual station in Nairobi were obtained from the National Climate Database at KMD and included daily rainfall accumulation, daily temperature (T) (Tmax, Tmin, T (06 and 12 UTC), Relative Humidity (06 and 12 UTC), Surface Pressure (06 and 12 UTC), and Wind Speed and Wind Direction observed at the top of each hour. The data observations from the Dagoretti manual synoptic weather station are on daily and hourly intervals. TAHMO observations are at 5 minutes intervals while 3D-PAWS observations were collected at 1- minute intervals.

The various weather parameters compared very well between the three stations. However, there are some features that need to be investigated further. Analysis of the data showed that the 3D-PAWS and TAHMO data provided comparable values. For example, significant correlation was observed between the TAHMO and Dagoretti manual station data for both minimum and maximum temperatures ($r = 0.65$ and $r = 0.61$, respectively) while the correlation for maximum temperature between TAHMO and 3DPAWS was modest ($r = 0.56$). The highest correlation was for maximum temperature between Dagoretti and TAHMO ($r = 0.86$).

There is need to have consistent maintenance of the stations so that the downtime is minimized. These data sets from the AWSs once validated can be used in weather forecasting and other early warning applications in the region.

1.0. Introduction

The World Meteorological Organization's (WMO) HIGH impact Weather LAke SYstem (HIGHWAY) Project is a United Kingdom (UK) Department for International Development (DFID) funded project, which is aimed at improving early warning systems (EWS) and hence increasing resilience of communities around Lake Victoria to extreme weather and climate. To achieve this, the project will improve observation and collection of weather and climate information with the aim of providing a better understanding of the dynamics and circulation of the Lake Victoria.

The HIGHWAY project brings together the local National Meteorological and Hydrological Services (NMHSs) within the East African region namely; Kenya, Rwanda, Tanzania and Uganda. The other institutions involved in the HIGHWAY project are the East African Community (EAC), UK Met Office, and the University Corporation for Atmospheric Research/National Center for Atmospheric Research (UCAR/NCAR) of the United States of America (USA).

The available data from the manual Meteorological stations from the NMHSs around the Lake Victoria region are sparse and there is little known information on the quality. These NMHSs also have Automatic Weather Stations (AWS) whose data quality is not known. The AWS datasets are also not shared globally through the Global Telecommunication System (GTS).

Potentially, there exists a denser network of AWSs from the Trans-African Hydro Meteorological Observatory (TAHMO). There is also a smaller network of 3D-Printed Automatic Weather Stations (3D-PAWS). It is with this in mind that the WMO recommended that the Kenya Meteorological Department (KMD) undertakes an inter-comparison analysis of TAHMO, 3D-PAWS and the KMD synoptic weather station using the KMD Dagoretti Corner Meteorological Station in Nairobi as the reference point. The goal of this work is to explore the potential of using the TAHMO and 3D-PAWS networks in the early warning services around the Lake Victoria Basin by assessing the quality of data from these two AWS networks using the KMD manual station data as a reference. The availability of consistent good quality datasets will enable the HIGHWAY project enhance the high-resolution regional Numerical Weather Prediction (NWP) model products for Lake Victoria, which will

improve the accuracy of the climate and weather products that are disseminated to the end-users around the Lake Victoria Basin.

To accomplish the data intercomparison, a one-week workshop was held in Mombasa along the coastal region of Kenya from 4th to 7th February 2019. The findings from the workshop were not conclusive due to the limited available data from the 3D-PAWS station co-located with the KMD station at Dagoretti Corner. This data was later made available by NCAR/UCAR, and hence another workshop was organised on 6th to 8th August 2019 in Naivasha to finalize the inter-comparison analysis.



Workshop participants in Naivasha in Kenya

1.2. Workshop objectives

The inter-comparison analysis had three main objectives:

- i) To assess the quality of rainfall, temperature, wind speed and direction, pressure and relative humidity from TAHMO, 3D-PAWS and the KMD Synoptic Weather station at Dagoretti Meteorological Station in Nairobi.
- ii) Inter-comparison of data from TAHMO, 3D-PAWS and the KMD Synoptic Weather station at Dagoretti Meteorological Station in Nairobi using statistical methods.
- iii) The performance of the 3D-PAWS and TAHMO AWSs in comparison to the KMD Synoptic Weather station at Dagoretti Corner Meteorological Station in Nairobi.

1.3. Justification

The main issue for the Lake Victoria Basin is that the climate variability around it is complex and driven by several global influences including the El Niño and La Niña in the tropical Pacific as well as regional forcing such as the Congo air mass, Indian Ocean temperatures and local climatic-factors such as the lake circulation effects which are not well understood. The resultant effect is that local communities are exposed to severe weather and climate events including flooding, hailstorms and lightning strikes as well as drought which have often led to loss of livelihoods as well as property. This inter-comparison study will provide a means of improving availability of quality data by facilitating calibration of the AWS data. The use better quality data will lead to improved early warning services in the region.

1.3.1 Study Area

The testing datasets for the inter-comparison analysis are from the Dagoretti Corner Meteorological Station in Nairobi, Kenya located at various sites within the Meteorological Station compound. The manual synoptic weather station is located at a longitude of 36.75000°E and latitude of 1.30000°S. The TAHMO AWS is located at a longitude of 36.76020°E and latitude of 1.3018389°S. The 3D-PAWS AWS is located at a longitude of 36.7601°E, latitude of 1.30172°S. All the three stations lie at an average altitude of 1790m above the mean sea level. The geographic positions of the AWSs and the synoptic weather stations clearly show that these stations are nearly collocated to each other.

2.0. Data Collection and Preparation

2.1. Data Collection

The current exercise utilises AWS datasets from TAHMO, 3D-PAWS and manual synoptic weather observations from the Dagoretti Corner Meteorological Station in Nairobi, Kenya. The data sets were obtained from the National Climate Database at KMD and include daily rainfall accumulation, daily temperature at 06Z and 12Z, daily minimum and maximum temperature, daily relative humidity, hourly wind speed, hourly wind direction and station pressure. The data observations from the manual synoptic weather station are on daily and hourly intervals. TAHMO observations are at 5 minutes intervals. 3D-PAWS observations were collected at 1- minute intervals.

During this exercise, data from the manual synoptic weather station, TAHMO AWS and 3D-PAWS at Dagoretti Corner meteorological station were examined to assess quantity, quality, and the available observables for comparison. The relevant parameters that were chosen for inter-comparison were Pressure, Relative Humidity, Solar Radiation, Wind speed and Direction, Rainfall and Temperature. The initial data covered the period January 2018 to June 2018.

Snapshot of Dagoretti Manual Synoptic Weather Data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	Stn ID	Year	Month	Day	DB(0600)	WB(0600)	DP(0600)	RH(0600)	Tmax	Tmin	Radiation	Evaporation	DB(1200)	WB(1200)	DP(1200)	RH(%)						
2	9136164	2017	1	1	19.8	15.9	13.7	67	24	13.9	18.89	5.5	23.6	17.5	14.2	55						
3	9136164	2017	1	2	19	16.2	14.7	75	24.1	14.6	6.15	3.5	23.7	17.1	13.4	51						
4	9136164	2017	1	3	17.6	16.2	15.5	87	24.1	15.8	12.62	4.5	23.8	17.2	13.5	52						
5	9136164	2017	1	4	16.7	15.3	14.6	87	24.2	15	1.79	2.5	21.5	16.9	14.4	63						
6	9136164	2017	1	5	18.5	14.9	12.7	68	21.9	13.9	12.93	5	22	15.3	13.1	49						
7	9136164	2017	1	6	18.5	13.8	10.6	59	23	9.7	14.51	5	24.8	13.2	3.1	24						
8	9136164	2017	1	7	16.8	14.8	13.7	81	25.2	14.1	9.34	5.5	22.7	16	11.9	50						
9	9136164	2017	1	8	16.9	15.2	14.3	84	14.2	15.49	4.5	24.5	16.7	12.1	45							
10	9136164	2017	1	9	18	15.4	14	76	17.9	16.11	4.5	24.1	15.9	1.7	42							
11	9136164	2017	1	10	17.4	15.5	14.5	82	15.7	14.9	5	23.1	16.5	12.6	51							
12	9136164	2017	1	11	17.6	15.6	14.5	81	15	11.72	4.5	24.6	17.2	13	47							
13	9136164	2017	1	12	17.7	15.8	14.8	82	15.8	1.94	2.6	2.3	17.3	15.8	74							
14	9136164	2017	1	13	16.5	14.8	13.9	84	14.9	9.82	4.5	22.9	16.3	12.4	51							
15	9136164	2017	1	14	15.9	14.5	13.7	86	15	15	5	22.3	14.7	9.5	43							
16	9136164	2017	1	15	18.5	15.3	13.5	72	13	16.62	5	25.7	15.7	8.9	34							
17	9136164	2017	1	16	19.1	16.3	14.8	75	13.1	14.16	5.5	25.5	16.7	11.3	40							
18	9136164	2017	1	17	19	15.5	13.5	69	15	16.62	6.5	25.3	15	7.6	32							
19	9136164	2017	1	18	17.8	15	13.4	74	15.4	11.17	6	25	16.2	1.6	40							
20	9136164	2017	1	19	18.8	15	12.7	67	12.5	15.37	5.5	24.2	15.5	7.9	39							
21	9136164	2017	1	20	19.3	14.1	10.6	56	1.3	22.59	6	25.3	15.9	9.7	37							
22	9136164	2017	1	21	18.5	15.3	13.5	72	15.8	19.89	5.5	24.3	16.3	11.4	43							
23	9136164	2017	1	22	18.4	16.2	15.1	80	15.9	13.92	21.6	16.6	13.8	60								
24	9136164	2017	1	23	16.8	15.2	13.8	82	15.2	15.83	24	15.2	8.6	37								

2.2. Data Presentation

Manual Station data were organized in a single “.csv” file containing daily values in a crosstab format while TAHMO data had five-minute values in a list in multiple files, one “.csv” file per day. The 3D-PAWS data were organized in ASCII format that were stored in separate files for each sensor (e.g., humidity and temperature, rain gauge, wind direction and speed, and surface pressure) with a resolution of 1-min records. Data from 3D-PAWS and

TAHMO were then processed to match the temporal resolution (hourly or daily) of the manual station. The following weather variables (data) were considered in this inter-comparison analysis study.

i) Precipitation Data

The precipitation data from the 3D-PAWS and TAHMO AWSs were accumulated to daily values starting from 0600Z of the current day to 0600Z to the following day as Daily values and then cast back one day, to match the observation period of the manual synoptic weather station.

ii) Surface Pressure

The hourly manual synoptic weather data for surface pressure were obtained from the National Climate Database. 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.

iii) Relative Humidity

The manual synoptic weather station data were available at 0600Z and 1200Z 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.

iv) Radiation

The manual synoptic weather station daily total radiation data were found to be inappropriate for comparison with observations from the 1-minute and 5-minute temporal resolutions from the 3D-PAWS and TAHMO stations, respectively. Therefore, it was not possible to compare them with the other AWSs datasets.

v) Wind Speed

The hourly wind speed data for the manual synoptic weather station was obtained from the National Climate Database. The AWSs data from TAHMO and 3DPAWS were matched at the top of each hour with the manual synoptic station observations. The 3D-PAWS data were also processed to average the 10-minute observations before the top of the hour to match the procedure used for the manual observations.

vi) Wind Direction

The hourly wind direction data for the manual synoptic weather station were obtained from the National Climate Database. The AWSs data from TAHMO and 3D-PAWS were matched to the top of the hour observations from the manual synoptic weather stations. The 3D-PAWS

data were also processed to average the 10-minute observations before the top of the hour to match the procedure of the manual observations.

vii) Temperature (Dry bulb, Tmax, Tmin)

The manual synoptic weather station dry bulb temperature data were available at 0600Z and 1200Z. The manual synoptic weather Tmax and Tmin data were also obtained as single values for each day. The AWSs data from TAHMO and 3D-PAWS were extracted to obtain the temperature observations at 0600Z and 1200Z on each day of the inter-comparison. The AWSs Tmax and Tmin were extracted from the 1-minute and 5 minute temperature values from 3D-PAWS and TAHMO, respectively, to obtain the minimum and maximum temperature for a given day. For 3D-PAWS, the Tmax and Tmin values were only obtained if a complete record was available for a given day.

2.3. Data Processing, Reformatting and Gaps Insertion

Once the required data were extracted, it was discovered that all the data sources did not indicate the missing data, resulting in data records with different lengths. The first step in the data analysis was to identify and insert gaps, to make the lists of same lengths of data sets selected for inter-comparison.

The manual rainfall data were converted from a crosstab table (shown below) to a list to ease inter-comparison, and gaps inserted. Gaps were also inserted into the other manual data (Pressure, Relative Humidity, Radiation, Wind speed and Direction and Temperature).

Stn ID	Name	Lat	Long	Year	Month	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	9136164	DAGORET	-1.3013	36.75968	2015	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	9136164	DAGORET	-1.3013	36.75968	2015	2	0	0	0	0	0	0	0	0	0	0	0	0	5.5
3	9136164	DAGORET	-1.3013	36.75968	2015	3	0	0	0	0	0	0	0	0	0	0	0	0	0
4	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	20.4	1.5
5	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
6	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	5.6
7	9136164	DAGORET	-1.3013	36.75968	2015	7	0	0	0	0	0	9.7	0	0	0	0	0	0	0
8	9136164	DAGORET	-1.3013	36.75968	2015	8	0	0	0	0	0	0.1	1.5	1.3	0.4	0	0	0	93
9	9136164	DAGORET	-1.3013	36.75968	2015	9	1.3	0	1.6	0	0	0	0	0	0	0	0	9.3	0
10	9136164	DAGORET	-1.3013	36.75968	2015	10	0	0	0	0	0	0	0	0	0	0	0	9.5	0
11	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
12	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
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

The 3D-PAWS data presented a unique challenge as it was not a common “.csv” file but rather like “.xml” file. It had to be converted to a “.csv” file before data extraction and aggregation. It was also found that the data had multiple values for each top of the hour, differing by a few seconds, making it extremely difficult to select the right value programmatically. 3D-PAWS data were reprocessed into consistent, continuous ASCII formatted files that included missing records and bad data flags during the August 2019 workshop. This improved the ability to compare 3D-PAWS in a consistent and

repeatable framework. Below, are example screen shots of the 3D-PAWS and TAHMO data files formats.

Screen shot of 3D-PAWS data (original csv format)

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	# dataset: GeoCSV 2.0												
2	# field_unit: ISO_8601 C	%	mb	mb	C	C	m/s	deg	mm	w/m2	w/m2	w/m2	
3	# field_type: datetime	float	float	float	float	float	float	float	float	float	float	float	float
4	# attribution: 3d-kenya.chordsrt.com												
5	# delimiter:												
6	# device_information: Kenya_3DPAWS_05												
7	# creation_date: 2018-11-16 07:22:58 +0000												
8	# data collection site: Kenya Meteorological Department												
9	# data collection longitude: 36.7601												
10	# data collection latitude: -1.30172												
11	# data collection elevation: 1750.0 meters												
12	# Units of Measure: (milliseconds)	C	%	mb	mb	C	C	m/s	deg	mm	w/m2	w/m2	w/m2
13	# Measurements in File: 182500												
14	Time	HTU21D_T	HTU21D_RH	BMP180_SLP	BMP180_SP	BMP180_T	MCP9808	windspeed	winddirection	precipitation	SI1145_VIS	SI1145_IR	UV1145_UV
15	2017-03-15T11:30:40Z									0			
16	2017-03-15T11:45:19Z									0			
17	2017-03-15T12:00:57Z									0			
18	2017-03-15T12:15:25Z							0					

Screen show of reformatted 3D-PAWS data (for T and RH).

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	-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	

Screen shot of TAHMO multiple files

Name	Date modified	Type	Size
TA00025_20180604.data	04/06/2018 11:00	DATA File	27 KB
TA00025_20180605.data	05/06/2018 18:27	DATA File	24 KB
TA00025_20180606.data	06/06/2018 18:27	DATA File	22 KB
TA00025_20180607.data	07/06/2018 11:28	DATA File	22 KB
TA00025_20180608.data	08/06/2018 13:28	DATA File	23 KB
TA00025_20180609.data	09/06/2018 14:29	DATA File	26 KB
TA00025_20180610.data	10/06/2018 16:30	DATA File	22 KB
TA00025_20180611.data	11/06/2018 21:31	DATA File	23 KB
TA00025_20180612.data	12/06/2018 18:32	DATA File	23 KB
TA00025_20180613.data	13/06/2018 20:33	DATA File	26 KB
TA00025_20180614.data	14/06/2018 00:34	DATA File	23 KB
TA00025_20180615.data	15/06/2018 22:35	DATA File	22 KB
TA00025_20180616.data	16/06/2018 22:36	DATA File	22 KB
TA00025_20180617.data	17/06/2018 06:36	DATA File	22 KB
TA00025_20180618.data	18/06/2018 13:37	DATA File	22 KB
TA00025_20180619.data	19/06/2018 18:38	DATA File	22 KB

Much time was spent aggregating and filtering the AWSs data to hourly and daily time steps and identifying the gaps so that the data from all the sources would be of equal lengths, to ease inter-comparison. This involved writing and testing DOS, MS-Excel and R scripts for this exercise. Combining the multiple TAHMO AWS daily files also presented a challenge, resulting to having to write and test DOS and R scripts for processing the data into a single file.

Screen shots of raw 3-D Data (left) and gaps identified in the data indicated using 'NA' (right)

	A	B	C	D	E	F	G	H	I
36	Kenya_3D	2017	3	19	15:00	3.17			
37	Kenya_3D	2017	3	19	16:00	2.84			
38	Kenya_3D	2017	3	19	19:00	1.56			
39	Kenya_3D	2017	3	19	21:00	1.3			
40	Kenya_3D	2017	3	19	22:00	1.13			
41	Kenya_3D	2017	3	19	23:00	0.83			
42	Kenya_3D	2017	3	20	0:00	1.62			
43	Kenya_3D	2017	3	20	1:00	2.3			
44	Kenya_3D	2017	3	20	2:00	0.49			
45	Kenya_3D	2017	3	20	3:00	1.44			
46	Kenya_3D	2017	3	20	4:00	1.31			
47	Kenya_3D	2017	3	20	5:00	0.9			
48	Kenya_3D	2017	3	20	6:00	1.74			
49	Kenya_3D	2017	3	20	7:00	2.93			
50	Kenya_3D	2017	3	20	8:00	1.62			
51	Kenya_3D	2017	3	20	10:00	1.81			
52	Kenya_3D	2017	3	20	11:00	3.02			
53	Kenya_3D	2017	3	20	13:00	3.16			
54	Kenya_3D	2017	3	20	14:00	2.98			
55	Kenya_3D	2017	3	20	16:00	1.8			
56	Kenya_3D	2017	3	20	17:00	1.94			
57	Kenya_3D	2017	3	20	18:00	1.25			
58	Kenya_3D	2017	3	20	19:00	1.81			
59	Kenya_3D	2017	3	20	21:00	0.4			
60	Kenya_3D	2017	3	21	2:00	0.54			

	A	B	C	D	E	F	G	H
10649	2017	3	19	15	3.17			
10650	2017	3	19	16	2.84			
10651	2017	3	19	17	NA			
10652	2017	3	19	18	NA			
10653	2017	3	19	19	1.56			
10654	2017	3	19	20	NA			
10655	2017	3	19	21	1.3			
10656	2017	3	19	22	1.13			
10657	2017	3	19	23	0.83			
10658	2017	3	20	0	1.62			
10659	2017	3	20	1	2.3			
10660	2017	3	20	2	0.49			
10661	2017	3	20	3	1.44			
10662	2017	3	20	4	1.31			
10663	2017	3	20	5	0.9			
10664	2017	3	20	6	1.74			
10665	2017	3	20	7	2.93			
10666	2017	3	20	8	1.62			
10667	2017	3	20	9	NA			
10668	2017	3	20	10	1.81			
10669	2017	3	20	11	3.02			
10670	2017	3	20	12	NA			
10671	2017	3	20	13	3.16			
10672	2017	3	20	14	2.98			
10673	2017	3	20	15	NA			

In general, the gaps in the AWSs data should be noted as a potential source of errors in this exercise.

3. Analysis Methods

We employed several statistical methods to carry out inter-comparison analysis. Graphical methods were used to visualize the data through line and scatter plots. Any one variable e.g. minimum temperature was plotted for all the three sources (TAHMO, 3D-PAWS, Dagoretti (in situ)) on one graph using R-statistical program (R Core Team, 2017).

To assess degree of relationships between the varied data types for the three sources, we used simple Pearson correlation method, which measures the linear correlation between two variables X and Y using the correlation coefficient (r).

This is given by the equation:

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

Where, r_{xy} or r is the correlation coefficient, N is the number of observations, and ΣX is the sum of x scores, ΣY is the sum of y scores, while ΣXY is the sum of the products of x and y scores

The values of r range between +1 and -1, where 1 indicates perfect/positive linear correlation, 0 is no linear correlation, and -1 is total negative linear correlation.

To test the statistical significance of the correlation coefficients, we used the Shapiro-Wilk normality test (Emerson 2015) and the Anderson-Darling normality test (Liebscher, 2016). The Shapiro Wilk's method is widely 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. 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.

Further, we used graphical method to visualize the correlations between any two variables e.g. TAHMO and Dagoretti (insitu) data. We further applied regression method to estimate the best straight line that summarises the correlation between the variables.

Finally we used Q-Q (quantile-quantile) plots to see how best the data sample correlated with the normal distribution. Overall, we used a correlation matrix to display all correlations between the varied data sets

4.0 Results of the Analysis

4.1. Pre-Analysis Results

Initial assessment conducted in February 2019 showed that Dagoretti corner manual station, TAHMO and 3D-PAWS datasets required more harmonisation.

From the analysis of these datasets, preliminary results showed good agreement between the manual station and TAHMO temperatures and relatively lower agreement with the 3D-PAWS. Rainfall and relative humidity values were found to vary greatly. Further investigation found that the TAHMO AWS had a broken rain-gauge sensor, during the period between 2017 and June 2018, though the actual dates could not be confirmed. **Fig 1** shows an example of the graphical comparisons of relative humidity at 12Z and minimum temperature for the manual, TAHMO and 3D-PAWS.

As a result of the above pre-analysis results, it was decided that the 6 months' data was not likely to give reliable conclusion and that hourly data for the manual station needed to be obtained from the National Climate Database. It was also decided that the period be extended to cover 2016 to the end of 2018 where possible.

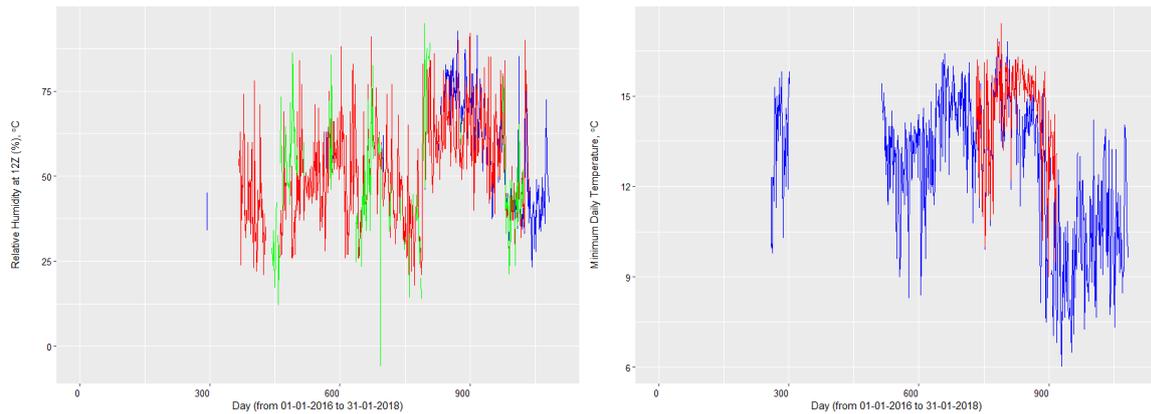


Fig 1: Initial comparisons of daily relative humidity at 12Z for manual (Dagoretti) station (red), TAHMO AWS (blue) and 3D-PAWS observations (green) for the period 2016 – 2018 (left) and between daily minimum temperature for manual (red) and TAHMO AWS (blue) observations for the period 2016 – 2018 (right)

4.2. Updated Results

The output shown in **Fig 2** is the time series of the Tmax, Tmin and T06, RH06 observations for Dagoretti and 3DPAWS from 2017 - 2019. There are only a few values of Tmax, Tmin, and T06, RH06 when both stations overlap. As mentioned previously, this was due to battery failure at night. There were more observations for temperature and RH at 12 considering that the 3D-PAWS station was often operational during daytime hours (results not shown).

The graphical outputs presented in **Fig 3** shows the comparisons between the minimum temperatures and maximum temperatures, respectively, for TAHMO, Dagoretti and 3DPAWS. Whereas the patterns in the three data sets agree, the comparison of minimum temperature indicates there is a general high bias with 3D-PAWS. Further investigation during the workshop indicated that the battery on the 3D-PAWS was not keeping charge during the inter-comparison period (it was working properly before the inter-comparison period). The station was shutting down during non-daylight hours and therefore, was not recording the true minimum daily temperature. In other words, it was recording the minimum temperature when it was operating during daylight hours, which on average was about 10⁰C higher than the minimum temperature observed with the manual station and TAHMO. The 3D-PAWS data were reprocessed to compute the minimum and maximum temperature when there was a complete, continuous record for that data. This significantly reduced the number of matching records, but resulted in a fair comparison between stations. This comparison identified the need for proper maintenance (as for any station) to ensure there is continuous measurement at a site. On the other hand, the maximum temperature largely seems to have similar patterns although with some disparities in some areas where the three data sets are

available. For example, the comparisons for the maximum temperature between manual station and 3DPAWS show some agreement (**Fig 3C**) although the manual temperature appears to have some spikes and missing data compared to the later. The more consistent agreement between the stations for maximum temperature indicates that 3D-PAWS station was operating properly during daytime hours.

The highest correlations of minimum temperature were indicated between TAHMO and Dagoretti station data ($r = 0.65$) with similar results being indicated for correlations of maximum temperature between TAHMO and Dagoretti station data ($r = 0.61$) (Figure 4, Table 1). The correlations for maximum temperature TAHMO and 3DPAWS were modest ($r = 0.56$). Nonetheless, considering the correlation matrix for the 3 temperature data sets (Figure 5), the highest correlations were found in the maximum temperature between Dagoretti and TAHMO ($r = 0.86$).

From figure 6 and table 2 the Shapiro Wilk normality test indicate that all the correlations for minimum and maximum temperatures had a p-value < 0.05 meaning that the distribution of the temperatures are not significantly different from normal distribution and they tend to be normally distributed.

The maximum temperature for the manual (Dagoretti) station appears to have the best normal distribution pattern compared to the other data sets (Figure 6E).

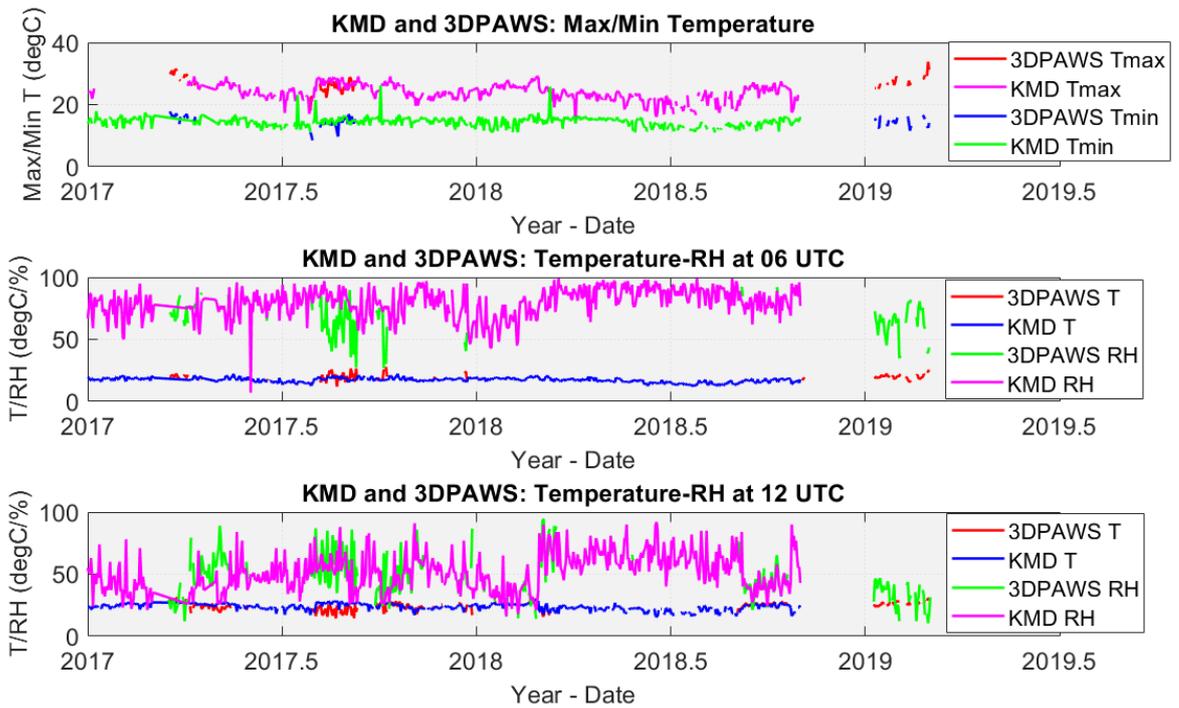


Fig 2: Time series of Tmax/Tmin from KMD and 3DPAWS (top panel), temperature and RH at 06 UTC for KMD and 3DPAWS (middle panel), and temperature and RH at 12 UTC for KMD at 3DPAWS.

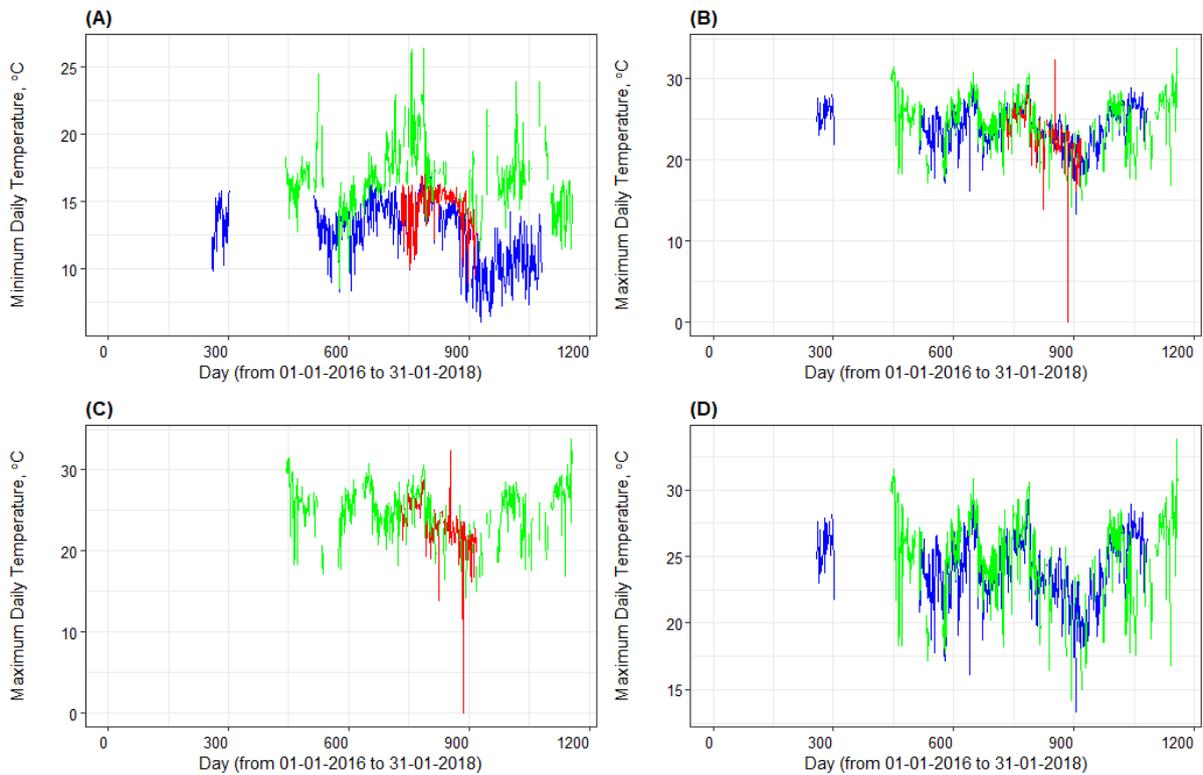


Fig 3: Comparisons for (A) daily minimum temperature (Dagoretti (red), TAHMO (blue), 3DPAWS (green) for the period 2017/2018, (B) daily maximum temperature between Dagoretti, TAHMO and 3DPAWS, (C) daily maximum temperature between Dagoretti and 3DPAWS, and (D) between TAHMO and 3DPAWS

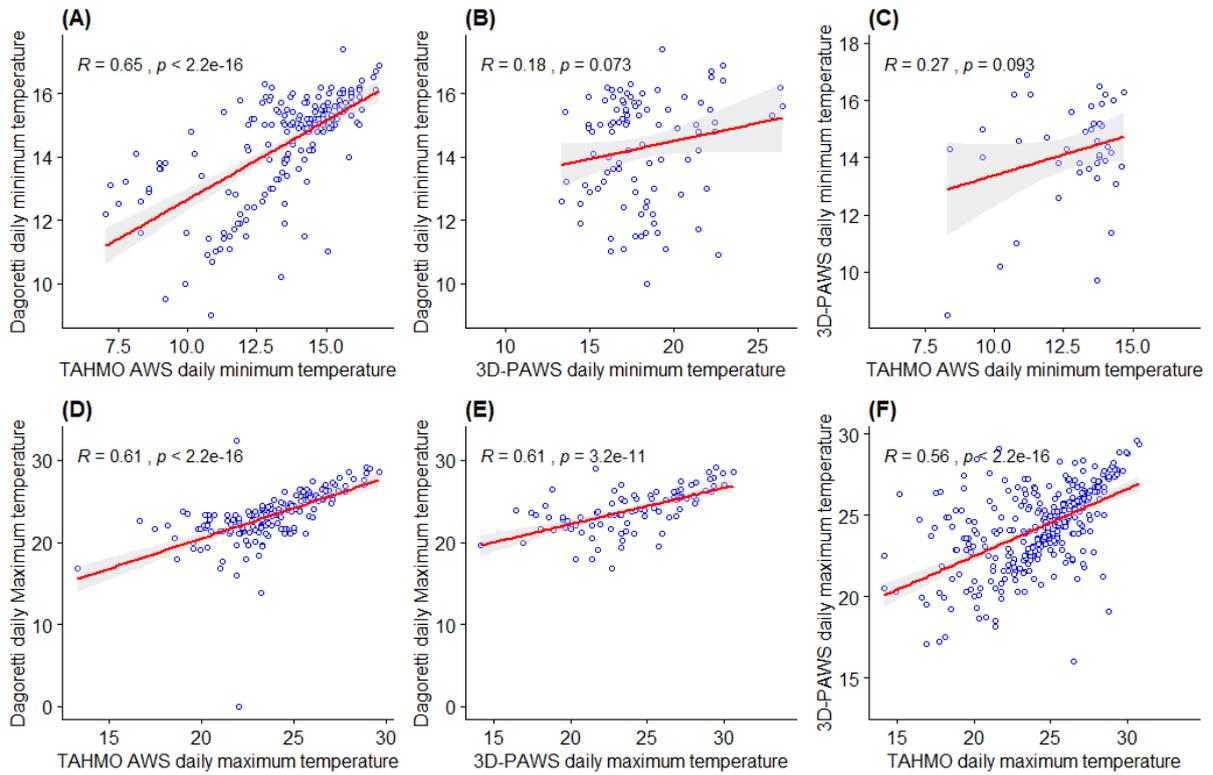


Fig 4: Correlations for daily minimum and maximum temperature between manual and TAHMO (A and D), manual and 3D-PAWS (B and E) and 3DPAWS and TAHMO (C and F). The minimum and maximum temperatures are fitted with a linear regression line (blue), while the gray envelope is the 95% confidence interval

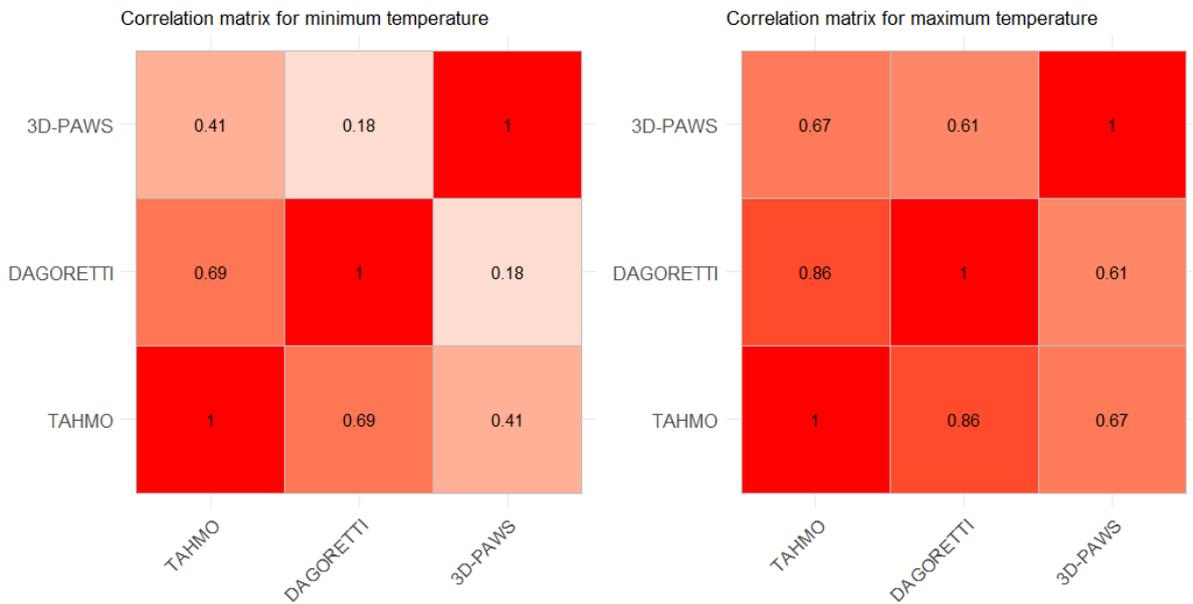


Fig 5: Correlation matrix for minimum and maximum temperature for 3D-PAWS, Dagoretti (manual) and TAHMO

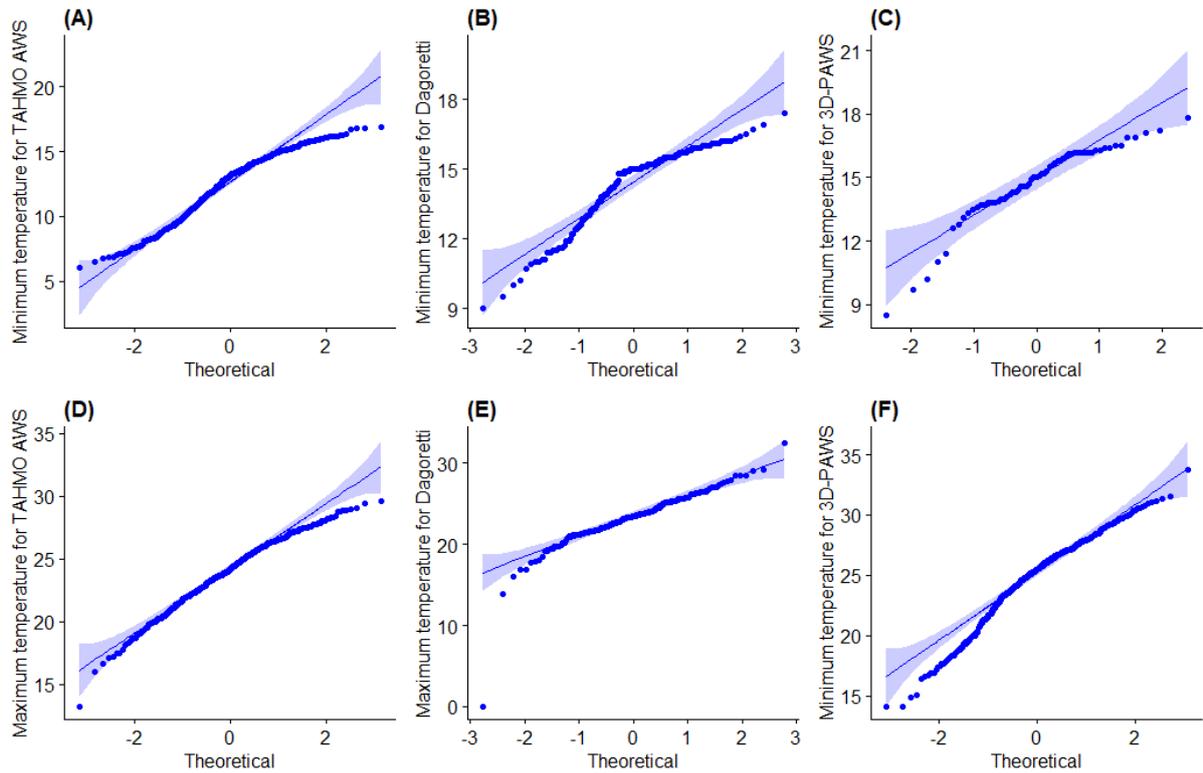


Fig 6: QQ plots for minimum temperature for (A) TAHMO (B) Manual (Dagoretti) and (C) 3DPAWS and for maximum temperature for (D) TAHMO (E) Manual (Dagoretti) and (F) 3DPAWS

Figure 7 shows comparisons of daily rainfall, surface pressure, relative humidity at 0600z and at 1200z and wind direction and speed at 0600z between manual station and 3DPAWS. Large differences are observed in when rainfall was recorded between the two stations especially the period where 3DPAWS recorded no rainfall while the manual station shows that some rainfall was recorded (Figure 7A). The period with no rainfall for 3D-PAWS is attributed to the lack of maintenance of the tipping bucket not being serviced. Wind direction and speed results indicate only a few records having similar wind speed and direction. This is likely attributed to the manual wind speeds being recorded on a 10 m mast when 3D-PAWS and TAMHO winds are measured at 2 m. Winds can be significantly different at the two levels because of surface roughness

The relationships between the different weather variables are largely positive (Figure 8 and Figure 9). The most significant correlations are indicated with relative humidity at 1200Z between the manual and 3D-PAWS ($r = 0.59$) while the lowest is indicated with the relative humidity at 0600Z with the manual station.

Normality test for the different weather variables indicate that surface pressure and relative humidity show some level of normal distribution compared to the other variables (Figure 10) and the normality test results indicated that all the distributions were significant (Table 2).

Rainfall for example, shows that the daily rainfall is not normally distributed and likely follows a non-linear distribution (Figure 10 (A) and (B)). Rainfall in Kenya has largely followed a log-normal distribution and other exponential distributions.

Although the comparisons and strength of correlation for wind speed and direction for the manual, 3DPAWS and TAHMO were fairly strong, comparison of wind rose between 3D-PAWS and the manual station shows consistency and a predominate NE wind in Nairobi (Figure 11). The wind speeds are much higher for the manual station, which is consistent with the difference in heights. One feature that needs to be investigated is the large distribution of winds near North for the manual station.

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 3DPAWS ($r = 0.65$) (Table 1, Figure 12, Figure 13 and Figure 14).

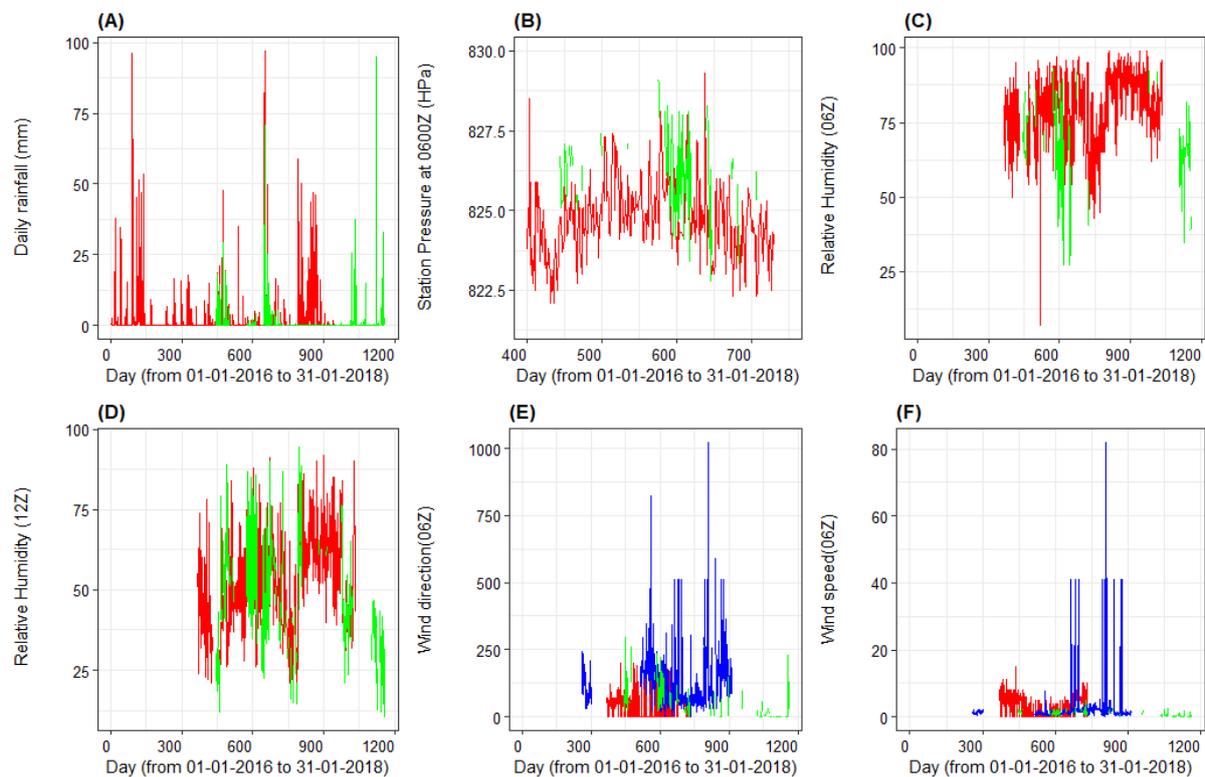


Fig 7: Comparisons of daily rainfall (A), surface pressure at 0600Z (B), relative humidity at 0600z (C), relative humidity at 1200z, wind direction at 006Z and wind speed at 006Z between manual (Dagoretti) station and 3DPAWS

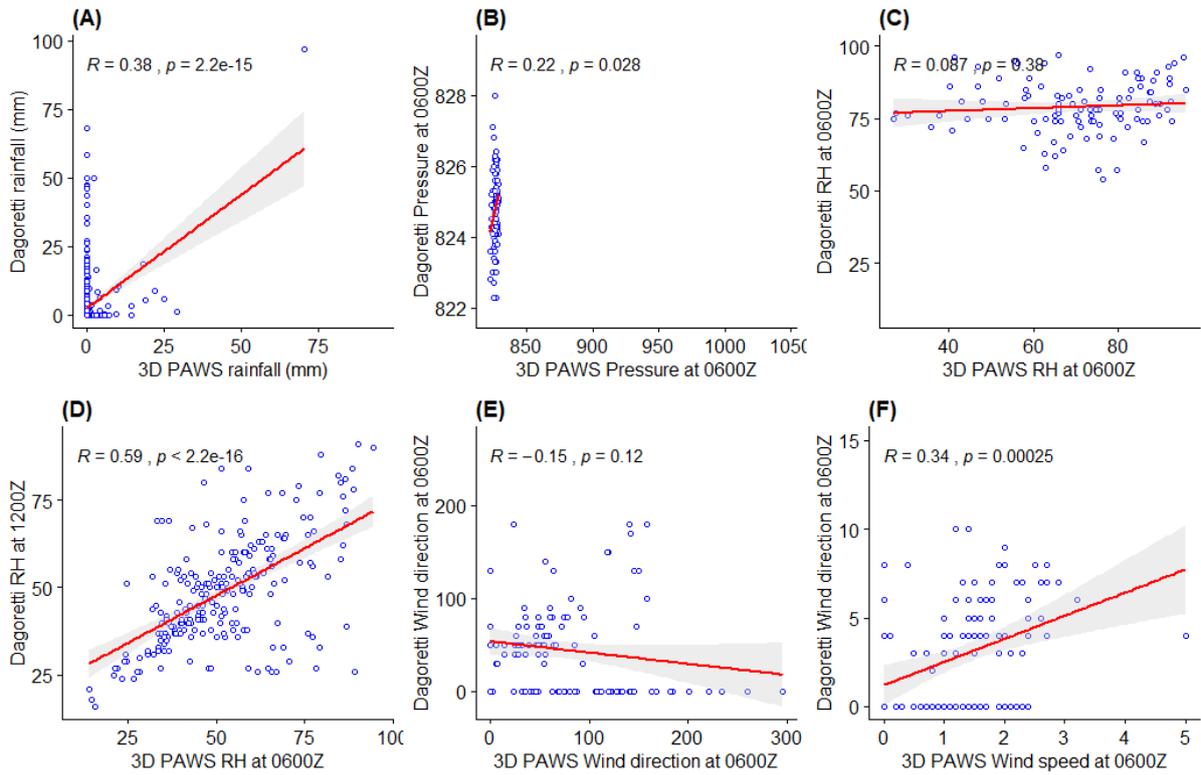


Fig 8: Comparisons of correlations for daily rainfall, surface pressure at 006Z, relative humidity at 006Z, relative humidity at 12Z, wind direction and wind speed at 006Z between manual (Dagoretti) and 3D-PAWS (RH_3D_12Z) stations



Fig 9: 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

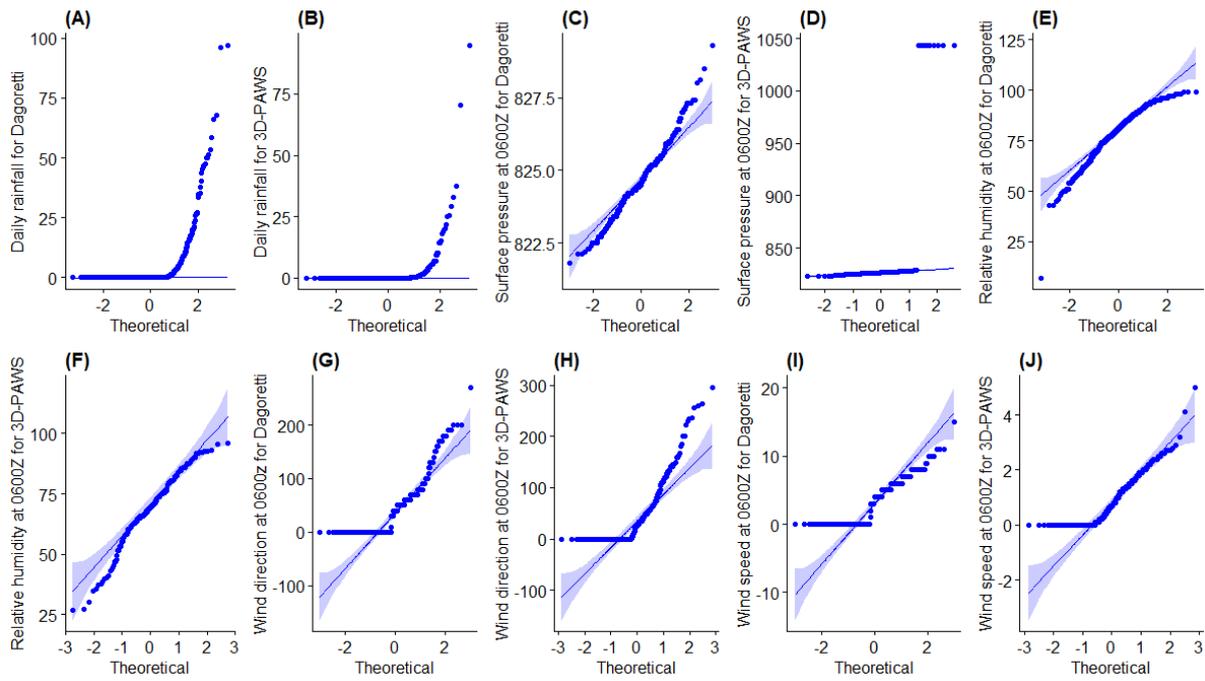


Fig 10: Q-Q plots for normality test for daily rainfall (A and B), surface pressure at 006Z (C and D), relative humidity at 006Z and at 12Z (E and F), wind direction and wind speed at 006Z between manual (Dagoretti) and 3D-PAWS (RH_3D_12Z) stations (G, H, I, and J).

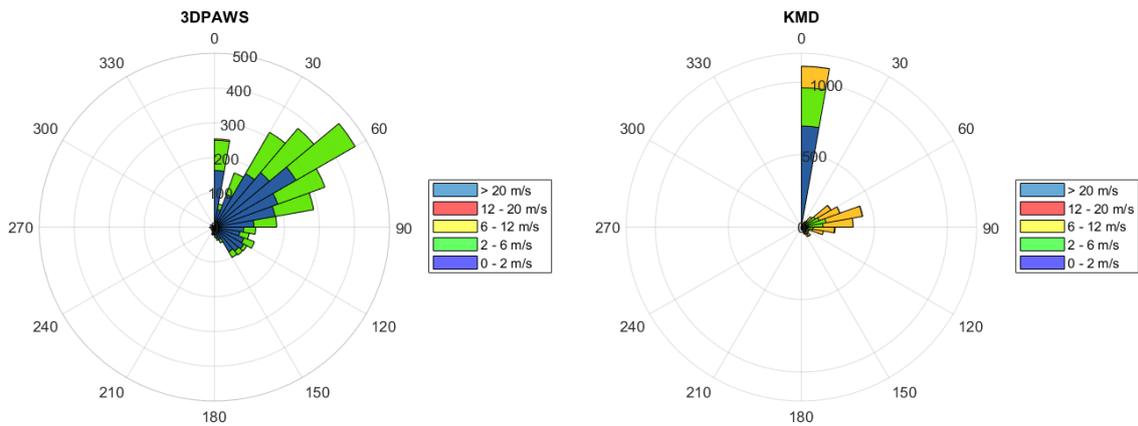


Fig 11: Windrose for 3DPAWS (left panel) and KMD (right panel)

Table 1: 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

Table 2: Shapiro-Wilk normality test for minimum and maximum temperature

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

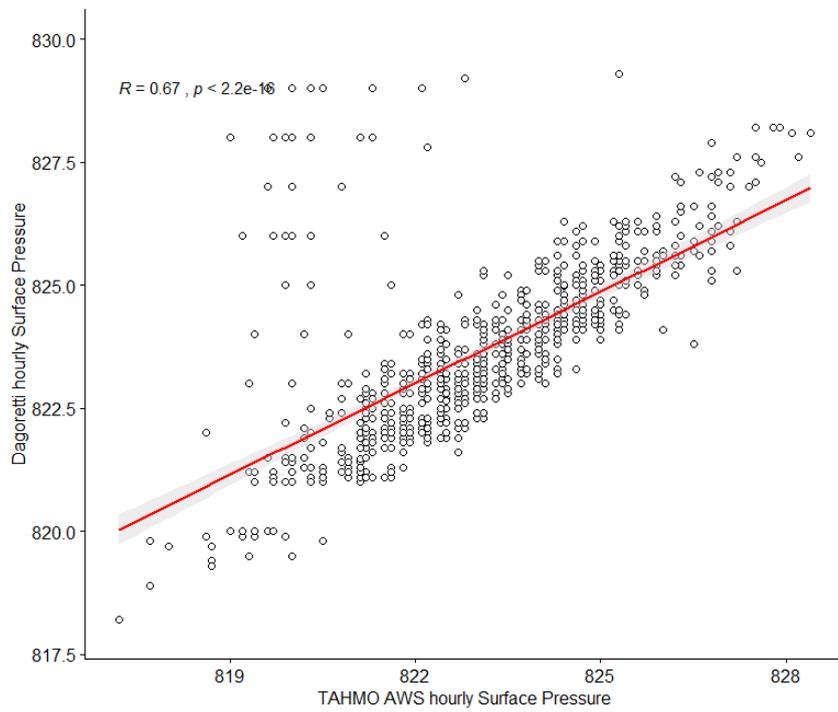


Fig 12: Correlation for hourly surface pressure between insitu (Dagoretti) and TAHMO AWS observations fitted with a linear regression line (blue) at the 95% confidence interval

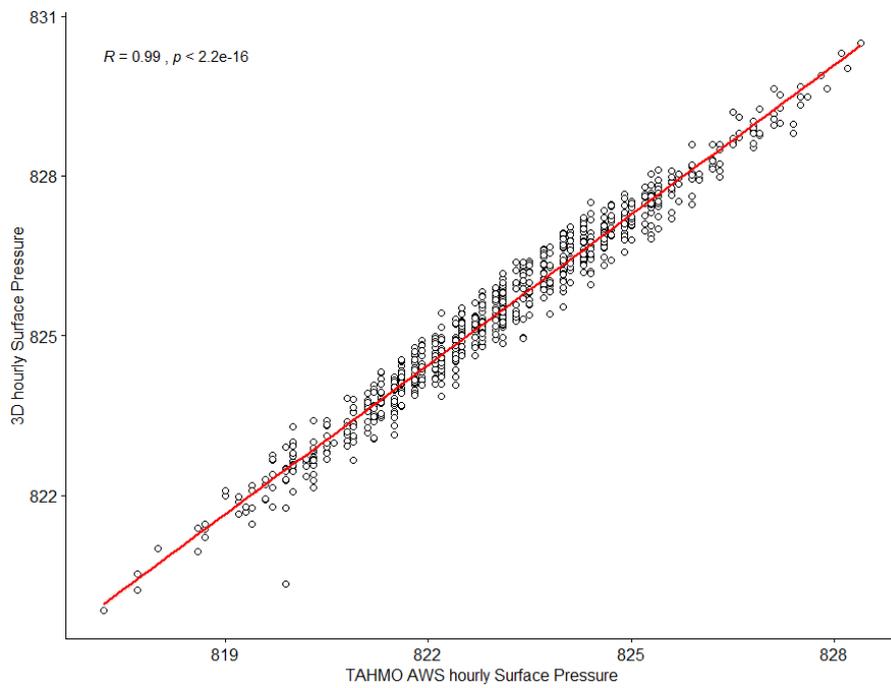


Fig 13: Correlation for hourly surface pressure between TAHMO AWS and 3D-PAWS observations fitted with a linear regression line (blue) at the 95% confidence interval

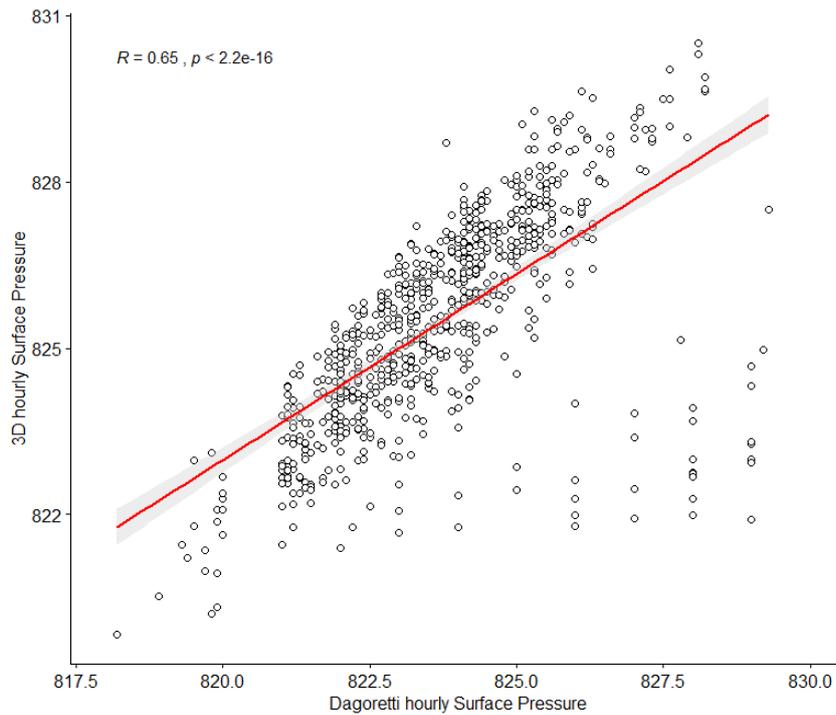


Fig 14: Correlation for hourly surface pressure between manual (Dagoretti) and 3D-PAWS observations fitted with a linear regression line (blue) at the 95% confidence interval

4.3. Summary and Conclusion

The measurements of the different weather parameters compare reasonably well. There are acceptable levels of correlations between the data values of the different weather parameters from the three stations. The analysis showed that 3D-PAWS and TAHMO stations provided comparable values when in operation. There are some features that need to be investigated further such as whether the AWS sensors are correctly calibrated; how wind and rainfall measurements from TAHMO and 3DPAWS can be harmonized with the manual measurements among other features. Therefore, there is a need for consistent maintenance of the stations to minimize the downtime and other issues such as faulty or dead batteries, problematic data loggers and broken radio links and any other physical damages.

4.4. Recommendations

- (i) This study should be used as a basis for upscaling inter-comparisons of all AWS data sets across the region.
- (ii) The up scaled and validated data sets should be used to improve the accuracy of the Numerical Weather Predictions for areas with minimal or scarce manual observations in the Lake Victoria basin region and used for early warning to minimize impacts of extreme climate

- (iii) There is need to provide consistent and regular maintenance of the AWS Network in the region to minimize loss and quality of data
- (iv) Capacity development in data management should be supported in order to improve processing and analysis of AWS datasets which are at finer resolution and with larger data quantity compared to the manual observations

5.0. References

Emerson, R. W. (2015). Causation and Pearson's correlation coefficient. *Journal of Visual Impairment & Blindness*, 109(3), 242-244.

Liebscher, E. (2016). Approximation of distributions by using the Anderson Darling statistic. *Communications in Statistics-Theory and Methods*, 45(22), 6732-6745.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.