



## Linear regression models to predict the tropospheric parameters at the Global Positioning systems' sites over the East African region

<sup>1,2</sup>\*SSENYUNZI C R., <sup>1</sup>ORURU B., <sup>1</sup>D'UJANGA M F

<sup>1</sup>Department of Physics, Makerere University, P. O. Box 7062, Kampala, Uganda.

<sup>2</sup>Department of Physics, Busitema University, P. O. Box 236, Tororo, Uganda.

\* Corresponding author: [ssecliffe@gmail.com](mailto:ssecliffe@gmail.com)

### Abstract

Atmospheric water vapour is a very important component of the atmosphere that plays a crucial role in climate and atmospheric studies. The increasing demand for this component has led to the establishment of the Global Positioning systems (GPS) networks over the world. With access to the GPS observations, accurate surface pressure and the weighted mean temperature ( $T_m$ ) data, the accurate and reliable atmospheric water vapour can be computed as Precipitable Water Vapour (PWV). However, the large data gaps in GPS observations and the lack of reliable surface meteorological data at most of the GPS stations in East Africa has affected the quality and the applications of the PWV data. In this study, the PWV, pressure and the  $T_m$  linear models have been developed. The models were developed utilizing one year (2013) GPS PWV and European Centre for Medium-Range Weather Forecasts (ECMWF) 5th Re- Analysis (ERA5), Total Column Water vapour (TCWV), surface pressure and 2 meter (2m) temperature data. The developed models are to predict PWV over stations with the GPS data gaps and to generate pressure and the  $T_m$  so crucial in PWV computations. The PWV accuracy obtained with the developed models shows an average RMSE of 1.54 mm and MnB of 0.32 mm in comparison to the measured GPS PWV data. The pressure accuracy obtained with the developed models shows an average RMSE of 0.26 hPa and average MnB of 0.03 hPa in comparison to the sensor pressure data while average RMSE and MnB are 0.15 hPa and 0.03 hPa, respectively, in comparison to the ERA5 pressure data. The  $T_m$  accuracy obtained with the developed models shows an average RMSE of 1.32 K and MnB of -0.01 K. Based on the RMSE, the site-specific models developed can be utilized to supplement the GPS and weather stations data over the thirteen stations since they can provide estimates of nearly a similar degree of precision compared to the measured values.

**Keywords:** *Precipitable water vapour; total column water vapour; global positioning systems*

**Cite as:** *Ssenyunzi et al., (2021). Linear regression models to predict the tropospheric parameters at the Global Positioning systems' sites over the East African region. East African Journal of Science, Technology and Innovation 2(3).*

Received: 05/01/21  
Accepted: 09/06/21  
Published: 25/06/21

## Introduction

Atmospheric water vapour is of a great importance to the atmosphere due to its impact on weather and climate. Its observations have numerous applications including, tracking on weather and climate alterations and assimilation into mathematical climate forecast models to improve climate prediction precision (Li and Deng, 2013; Hong *et al.*, 2015). Though, the spatial and temporal variability of this component in the atmosphere makes it a completely hard parameter to measure. This problem was somewhat settled when radiosondes were presented in 1940's (Durre *et al.*, 2006).

The estimation of the PWV involves different observational standard methods which require observations from GPS, radiosondes, Light Detection and Ranging (LIDAR) systems, Water Vapour Radiometers (WVR) among others (Singh *et al.*, 2014; Chen *et al.*, 2018; Wang *et al.*, 2018; Hicks-Jalali *et al.*, 2020). Likewise calculations and procedures to estimate PWV using linear regression methods have been explained e.g. (Maghrabi, 2009; Maghrabi and Clay, 2010). However, many equipments for PWV estimation are excessively costly and are lacking in numerous parts of Africa.

The accurate water vapour estimations in time and space for better climate predictions can additionally be accomplished by the utilization of GNSS/GPS and this is alluded to as GNSS meteorology. The ground-based GNSS receivers offer various favorable circumstances which incorporate the ability to give information of good quality under any climate condition (Hagemann *et al.*, 2003), extraction of information for any point independent of nation limits, lower cost and observations throughout enormous territories in time. In any case, East Africa faces more difficulties because of the scattered GNSS reference sites, deficient and inconsistent information, absence of precise and dependable surface meteorological information and device failures, among others.

However in regions much like East Africa where the necessary PWV information are scant, there has been advancement and use of observational/empirical models (Hussain, 1984; Guerova *et al.*, 2005; Maghrabi and Clay, 2010; Maghrabi and Al Dajani, 2012). These empirical models are based totally at the

statistical fit among the PWV statistics and different meteorological parameters, which includes specific humidity, dew point temperature, air temperature, and relative humidity. A few of these parametric models of PWV have been created in various locales of the world dependent on experimental information (Tuller, 1977; Viswanathan, 1980; Hsu and Blanchard, 1989; Gueymard, 1994; Adeyemi, 2009a, b; Maghrabi and Al Dajani, 2012). These models are created from various relations between various meteorological parameters as proven through Reitan (1963), who found a linear connection between monthly mean surface dew point temperature and the natural logarithm of the PWV. Likewise Hussain (1984), advanced an empirical version that parameterized PWV as a component of the air temperature and relative humidity. Ruckstuhl *et al.*, (2007), proposed a linear connection between the GPS-inferred PWV and specific humidity at a few locations in Europe. Information on formula and overall performance evaluation of PWV correlations in the East Africa, is extremely urgent since this region has a high potential for the usage of PWV for weather alternate and variability.

Several research have been performed about empirical models predicting precipitable water vapour, for example Maghrabi and Al Dajani (2012). In their investigation, the predictive powers of 15 experimental models were examined towards the measured PWV for the duration 1985 to 2007 utilizing radiosonde information for the metropolis of Riyadh in focal Saudi Arabia. The performances of these models indicated RMSE in the range 3.54 mm to 10.78 mm. Using upper-air data from Thailand, Phokate, (2017) formulated an empirical model associated with vapour pressure. In this study, a comparison study between this research with other researches such as that of Okulov *et al.*, (2002), Adeyemi (2009a), and Leckner (1978), were carried out. The RMSE values between 5.3 mm and 11.0 mm were observed. Additionally Falaiye *et al.*, (2018), performed a study on empirical models to predict precipitable water vapour throughout West Africa. Following this study, the statistical correlations among the derived empirical models, NWP models and the Aerosol Robotic Network (AERONET PWV), indicated the RMSE values in the range between 3.3 mm to 8.6 mm. There are also empirical models developed for the prediction

of pressure and the weighted mean temperature. These include the Global Pressure and Temperature (GPT) model series such as GPT2w (Bohm *et al.*, 2015), GPT3 (Landskron and Bohm, 2018), and a global model by Sun *et al.*, (2019) among others. Most of these empirical models require information about time and location to predict the pressure and weighted mean temperature. However, due to spatial and temporal variability of water vapour, its derivation becomes difficult because such measurements alone are not sufficient for its accurate measurement. If surface pressure measurements from weather sensor or from Numerical Weather Prediction (NWP) datasets are available, PWV can however be predicted with higher accuracy (Huang *et al.*, 2021).

The inadequate and inconsistent GPS observations, weather parameters and the limited tropospheric water vapour monitoring instruments have greatly affected the application of PWV in weather and climate predictions over the East African region. The inconsistency of GPS information from the various stations could probably be due to equipment failure, or in a few instances

electrical power problems and poor internet connectivity. Because of these factors, linear regression models have been developed to predict PWV, pressure and weighted mean temperature at the available GNSS stations. The models can generate alternative data to fill up the gaps and to also improve on the continuous availability of the required data sets for computation of the PWV. These models use ERA5, surface total water vapour, surface pressure and 2m temperature as data inputs. These ERA5 surface parameters are readily available from the ECMWF global data archives and can be obtained at any point of interest defined by the longitude and latitude values.

## Material and methods

### Study area

The study area exists in the range of latitude 12° S to 4° N and longitude 28° E to 44° E. Thirteen stations from Uganda, Kenya, Tanzania and Rwanda were utilized for this study and the information about these stations are provided in Table 1. The region is part of the world traversed by the Equator and open to the Indian Ocean.

Table 1. GPS stations coordinates

Station Id and City	Country	Longitude (E) degrees	Latitude (N) degrees	Ellipsoidal height (m)
ARSH (Arusha)	Tanzania	36.698	-3.387	1345.161
DODM (Dodoma)	Tanzania	35.748	-6.187	1122.595
KFNY (Kifanya)	Tanzania	35.103	-9.547	1714.834
MAL2 (Malindi)	Kenya	40.194	-2.996	-20.937
MBAR (Mbarara)	Uganda	30.738	-0.602	1337.53
MBEY (Mbeya)	Tanzania	33.459	-8.912	1700.514
MOIU (Eldoret)	Kenya	35.29	0.288	2201.511
MTDK (Mtandika)	Tanzania	36.421	-7.544	548.175
MTVE (Mtwara)	Tanzania	40.166	-10.26	-11.414
NURK (Kigali)	Rwanda	30.09	-1.945	1483.809
RCMN (Nairobi)	Kenya	36.894	-1.221	1591.967
SNGC (Songea)	Tanzania	35.673	-10.69	1181.221
TNDC (Tunduru)	Tanzania	37.341	-11.063	672.948

### Computation of GPS PWV

The GPS observations gathered from 13 stations for a duration of four years from 2013 – 2016 were utilized. In computation of GPS

PWV, the Zenith Total Delay (ZTD) is calculated first. Applying the GPS observation, ZTD was estimated based on Precise Point Positioning (PPP) mode from goGPS v1.0 beta1

software package (Herrera *et al.*, 2016). In ZTD computation, the following corrections were performed, corrections for antenna phase wind up, center offset and variations, Earth and polar motion tides, ocean Tide Loading (OTL) and relativistic effects. The OTL information was acquired using coefficients from FES 2004 model from (<http://holt.oso.chalmers.se/loading>). Further details about the processing options for goGPS are explained by Ssenyunzi *et al.*, (2019).

The ZTD is the sum of two components which include the Zenith Hydrostatic Delay, (ZHD) and the Zenith Wet Delay, (ZWD) i.e.

$$ZTD = ZHD + ZWD \quad (1)$$

The Zenith Hydrostatic Delay, (ZHD) at the site can be accurately determined by the Saastamoinen model (Saastamoinen, 1972);

$$\begin{aligned} ZHD &= 0.0022768 \\ &\cdot \frac{P_s}{1 - 0.00266 \cdot \cos(2\varphi) - 0.00000028 \cdot H_s} \end{aligned} \quad (2)$$

where  $H_s$  is the altitude above sea level and  $\varphi$  is the latitude of the station and  $P_s$  is the air pressure at the station in hPa. Because of the spatial and temporal variability of water vapour, determination of the ZWD becomes problematic due to the fact that surface estimations alone are not adequate for its precise estimation. With the assistance of GPS ZTD data and the computed ZHD, ZWD was estimated as unknown parameter (Ssenyunzi *et al.*, 2020);

$$ZWD = ZTD - ZHD. \quad (3)$$

The GPS PWV therefore can be computed from ZWD using the conversion factor,  $K$  (Bevis *et al.*, 1994);

$$GPS\ PWD = K \cdot ZWD \quad (4)$$

Askne and Nordius (1987), estimated  $K$  as;

$$K = \frac{10^6}{\rho_v R_v \left( \frac{k_3}{T_m} + k_2 \right)} \quad (5)$$

where  $\rho_v$  is the density of liquid water and  $R_v$  is the specific gas constant of water vapour. The  $T_m$  is the weighted mean temperature of the atmosphere and the only unknown parameter in Equation 5 given by Davies *et al.*, (1985) as;

$$T_m = \frac{\int_h \frac{e}{T} dh}{\int_h \frac{e}{T^2} dh} \quad (6)$$

The  $T_m$  is computed from  $e$ , the partial water vapour pressure and  $T$ , the air temperature in K, all from the vertical profiles of ERA5 as explain in section 2.3. below.

#### **Determination of ERA5 Pressure and Weighted Mean Temperature at the GPS Stations**

The ERA5 datasets are the latest generation of ECMWF global atmospheric reanalysis from 1950 within 5 days of real time, delivered by the ECMWF (Hersbach & Dee, 2016). The ECMWF presents 1-h and 6-h estimates of a massive range of atmospheric, land and oceanic weather/climate variables. The data cover the Earth on a 30km grid and resolve the atmosphere utilizing 137 levels from the surface of the Earth up to a height of 80 km. ERA5 incorporates data at decreased spatial and temporal resolutions (<https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5>). In this study, the 4 years data (2013 to 2016) at the overall ERA5 spatial resolution of  $0.25^\circ \times 0.25^\circ$  and a 6-h temporal resolution was applied. The retrieval of the required parameters at the GPS sites from ERA5 data involves acquiring the following data sets first. The temperature, and relative humidity at model levels, the log of surface pressure and the surface geopotential are downloaded from the ECMWF using a Web API. The log of surface pressure and the surface geopotential data is used in the reconstruction of the full level pressure and the geopotential height associated with each model levels. By using the longitude and latitude of the point of interest and at the four nearest grid points, the required variables are bilinearly interpolated to the point of interest. Given that ECMWF model levels are characterized by geometric heights, the target variables are vertically linearly interpolated to the geometric heights of the GPS stations. Since the GPS stations heights are given as the ellipsoidal height of the World Geodetic System, 1984 (WGS-84), they were converted to the corresponding orthometric heights using the Earth Gravitational Model 1996 (EGM96) and the geoid undulation model. These heights are then compared to the geometric heights corresponding to the ECMWF model levels. The geometric height  $h$ , at the model levels is given as

$$h = \frac{H(h, \varphi) \cdot R(\varphi)}{G \cdot R(\varphi) - H(h, \varphi)} \quad (7)$$

where  $h$ , is the geometric potential in km,  $H(h, \varphi)$  is the geopotential height in km corresponding to  $h$ ,  $R(\varphi)$  is the radius of the earth in km at latitude  $\varphi$ .

The radius of the earth  $R(\varphi)$  is given by

$$R(\varphi) = \left( \frac{\cos^2(\varphi)}{R_{max}^2} + \frac{\sin^2(\varphi)}{R_{min}^2} \right)^{1/2}$$

where  $R_{max} = 6378.137$  km is the earth's equatorial radius and  $R_{min} = 6356.752$  km is the earth's polar radius.

The ratio,

$$G = \frac{g(h, \varphi)}{g_o}$$

where  $g_o$  is the normal gravity at a latitude of  $45^\circ$  whose value is  $9.80665 \text{ ms}^{-2}$ ,  $g(h, \varphi)$ , is normal gravity at a point  $(h, \varphi)$ , given by

$$g(h, \varphi) = \left[ \frac{1 + 0.000193185 \cdot \sin(\varphi)^2}{1 + 0.00669435 \cdot \sin(\varphi)^2} \right]^{1/2}$$

Then  $T_m$ , is calculated from the vertical profiles of  $e$ , the partial pressure of water vapour and  $T$ , the temperature using the integration method in Eq.6. The partial pressure of water vapour is given by

$$e = rh \cdot 6.1078 \exp\left(\frac{17.22T}{T + 237.5}\right)$$

where  $rh$ , is the relative humidity.

#### ***Obtaining ERA5 surface parameters and weather sensor pressure***

The study additionally applied ERA5, surface pressure, 2m temperature and TCWV estimates required in the development of linear regression models. These parameters are readily available from the ECMWF global data archives and can be obtained at any point of interest defined by the longitude and latitude values by bilinear interpolation only.

The weather sensor which provide the pressure at the GPS stations are only available at five (ARSH, DODM, MBey, MTVE, NURK) stations of the 13stations considered. The weather sensor data is available at <ftp://data-out.unavco.org/pub/rinex/met/>.

#### ***Linear regression models***

In order to develop the linear regression models to predict parameters at the GPS sites, the GPS PWV and the ERA5 parameters at the GPS sites obtained as explained in sections 2.2 and 2.3 together with the surface ERA5 parameters obtained as in section 2.4 are

required. The linear regression models are of the form,

$$y = ax + b \quad (8)$$

where  $y$  and  $x$  are the dependent and independent variables,  $a$  and  $b$  are the  $y$ -intercept and the slope, respectively. For this study,  $y$  represents the parameters at the GPS sites and  $x$  represent the ERA5 surface parameters. To obtain the linear regression models, the linear regression analysis was performed using the Microsoft excel on the selected  $y$  and  $x$  parameter.

#### **Results**

##### ***Relationship between GPS PWV and TCWV***

In this study, the accuracy of GPS PWV data computed as explained in section 2.2 at the station height was first compared to TCWV data obtained as explained in section 2.4. Table 2 shows the correlation, RMS errors of GPS PWV minus TCWV, and MnB of the estimates at each one of the 13 GPS stations for the years 2013-2016. In formulating the linear regression

equations, daily averages of GPS PWV and TCWV for 1-year (2013) for every one of the 13 stations have been applied independently. Table 3 shows the coefficients of determination, the slopes and y- intercepts for the regression equations for each site. The accuracy of PWV obtained from the site-specific linear models using TCWV as the input was assessed using GPS PWV datasets for the years 2014 - 2016 for each location separately. Figures 1, 2, and Table 4

show the model computed PWV accuracy achieved with the site-specific models. For clear diagrams, the area under study has been divided into two regions basing on rainfall regimes as explained in Ssenyunzi et al. (2020). The region characterized by single-peak rainfall regimes, that is, the region between latitudes 12° S and 4° S and the region with double-peak rainfall regimes between latitudes 4° S and 4° N.

Table 2. Correlation, MnB and RMSE between the GPS PWV and ERA5 surface TCWV (mm) data at the 13 stations

Site	GPS PWV - TCWV		
	Corr	MnB (mm)	RMSE (mm)
ARSH	0.953	1.90	2.49
DODM	0.980	-0.39	1.53
KFNY	0.987	-1.86	2.31
MAL2	0.971	1.19	2.07
MBAR	0.947	0.45	1.5
MBEY	0.982	-1.57	2.15
MOIU	0.945	-0.33	1.45
MTDK	0.982	6.63	6.98
MTVE	0.985	1.36	2.32
NURK	0.943	-0.05	1.55
RCMN	0.951	1.02	1.83
SNGC	0.987	-2.6	3.09
TNDC	0.986	-1.54	2.28
<b>Mean</b>		<b>0.32</b>	<b>2.43</b>

Table 2 shows that the distinction between the GPS PWV and TCWV has RMSE estimates between 6.98 mm and 1.45 mm and an average estimation of 2.43 mm. The MnB values range among - 2.60 mm and 6.63 mm with a general mean value of 0.32 mm. The estimates profoundly correlate, with correlation coefficients somewhere in the range of 0.943 and 0.987 at the 13 stations. For all correlations, MTDK was seen to have the highest RMSE (6.98 mm) and MnB (6.63 mm) values. According to Table 3, the coefficients of determination ( $R^2$ ) resulting from correlating the GPS PWV with the TCWV for the 13 stations are in the range

0.898-0.978.

It is seen in Figures 1, 2 and Table 4 that site specific PWV models show the RMSE ranging between 1.33 mm and 1.91 mm with a general average value of 1.54 mm. The MnB errors are very small with overall average value of less than  $\pm 0.32$  mm. The correlations between PWV computed from the models and GPS PWV for the individual stations range between 0.949 and 0.989. The highest MnB and RMSE values are observed at Mtandika (MTDK) station with values of 0.74 mm and 1.91 mm, respectively.

Table 3. The slopes (a), y-intercepts (b) and coefficients ( $R^2$ ) of determination for the PWV regression equations

Site	GPS PWV			
	$R^2$	a	b	No. data
ARSH	0.949	0.975	2.269	308
DODM	0.974	0.967	0.269	365
KFNY	0.978	0.969	-1.201	300
MAL2	0.932	0.932	3.814	360
MBAR	0.914	0.939	1.935	347
MBEY	0.965	0.978	-1.399	286
MOIU	0.898	0.996	-0.481	281
MTDK	0.967	1.136	2.283	228
MTVE	0.972	0.981	2.045	365
NURK	0.922	0.893	2.714	358
RCMN	0.91	0.928	2.398	268
SNGC	0.943	0.894	0.524	218
TNDC	0.966	0.926	1.126	222

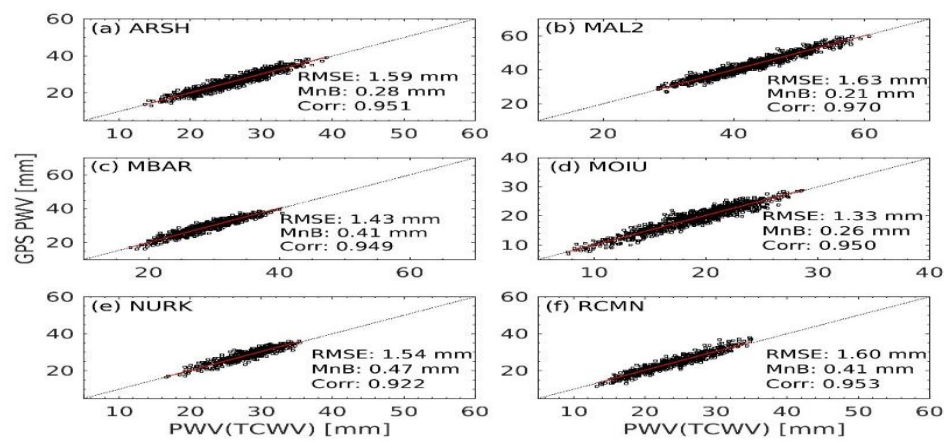


Figure 1. Scatter plots of GPS PWV versus modeled PWV (PWV (TCWV)) for stations in the in the region between latitudes  $4^{\circ}$  S and  $4^{\circ}$  N.

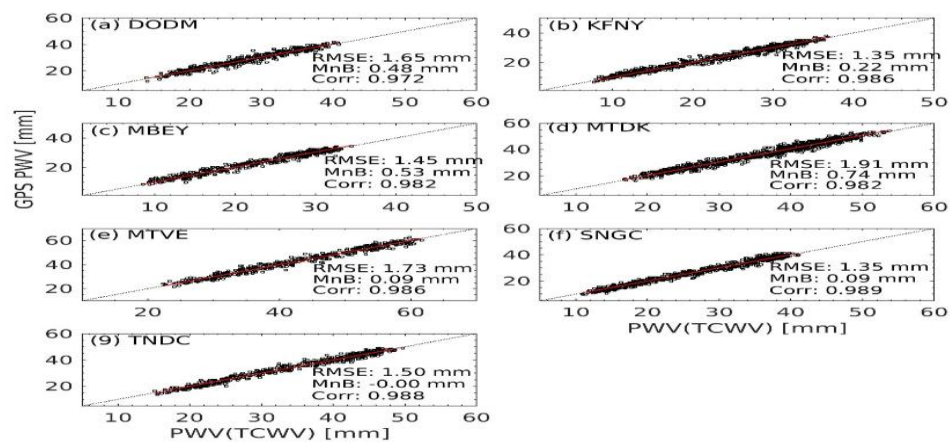


Figure 2. Scatter plots of GPS PWV versus modeled PWV (PWV (TCWV)) for stations in the in the region between latitudes  $12^{\circ}$  S and  $4^{\circ}$  S.

Table 4. Modeled PWV accuracy achieved with site specific linear models (2014-2016)

Site	GPS PWV - Modeled PWV			
	Corr	Bias (mm)	RMSE (mm)	sample data
ARSH	0.951	0.28	1.59	818
DODM	0.972	0.48	1.65	450
KFNY	0.986	0.22	1.35	658
MAL2	0.970	0.21	1.63	1063
MBAR	0.949	0.41	1.43	823
MBEY	0.982	0.53	1.45	515
MOIU	0.950	0.26	1.33	930
MTDK	0.982	0.74	1.91	856
MTVE	0.986	0.09	1.73	460
NURK	0.922	0.47	1.54	485
RCMN	0.953	0.41	1.60	660
SNGC	0.989	0.09	1.35	727
TNDC	0.988	0.00	1.50	509
<b>Mean</b>		<b>0.32</b>	<b>1.54</b>	

**Relationship between Weather Sensor Pressure and the ERA5 surface Pressure.**

The pressure estimates from the weather sensors available at the 5 stations were first compared with the surface pressure from ERA5. Table 5 shows the correlation, MnB and RMSE estimations of the differences between

the ERA5 surface values and the sensor values. Considering Table 5 and taking the pressure from sensors as reference data, the average MnB error and RMSE for surface pressure estimates are 3.03 hPa and 9.88 hPa, respectively, and the correlation coefficients are between 0.976 and 0.997.

Table 5. Correlation, MnB and RMSE between Sensors Pressure (hPa) at the 5 GPS station and the ERA5 surface pressure

Site	Pressure (hPa)		
	Corr	Bias	RMSE
ARSH	0.990	24.43	24.43
DODM	0.992	-2.68	2.69
MBEY	0.987	-12.43	12.44
MTVE	0.997	7.84	7.85
NURK	0.976	-1.99	2.00
<b>Mean</b>		<b>3.030</b>	<b>9.88</b>

The pressure from sensors co-situated at the GPS stations and the ERA5, surface pressure for 1-year (2013) for each one of the five stations, ARSH, DODM, MBEY, MTVE and NURK have been applied in developing the linear regression models. The regression equations' slopes, y-intercepts and the coefficients of determination are presented in Table 6. Figure 3 and Table 7 show the modeled pressure

accuracy achieved when compared with the weather sensor pressure values for the years 2014 to 2016.

From Table 6, the coefficient of determination ( $R^2$ ) for linear regression equations range between 0.953 and 0.995.



Table 6. The sensor pressure regression equations parameters

Site	Pressure (hPa)		
	R <sup>2</sup>	a	b
ARSH	0.975	1.182	-128.483
DODM	0.984	1.000	-1.728
MBEY	0.970	1.004	15.448
MTVE	0.995	1.020	-11.942
NURK	0.953	1.031	-28.926

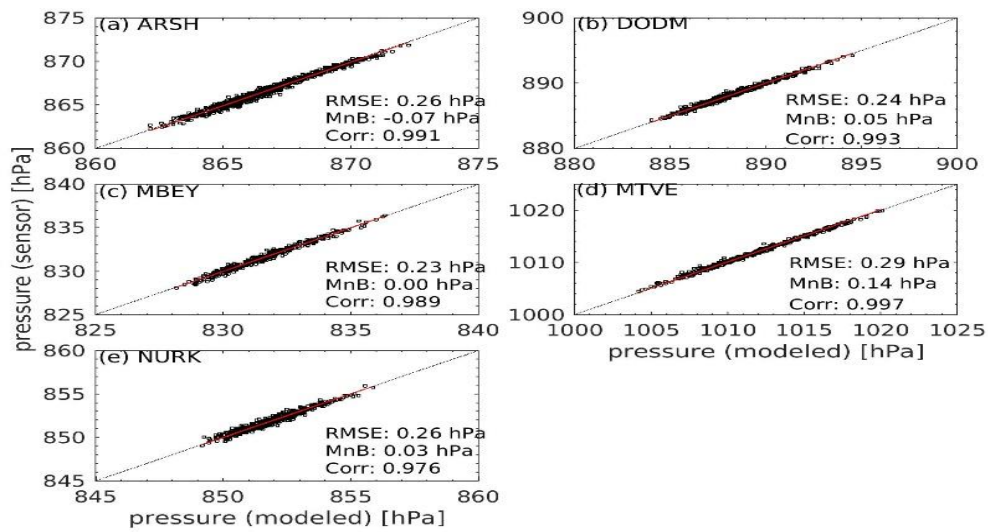


Figure 3. Scatter plots of sensor pressure versus modeled pressure from the regression equation

It is observed in Figure 3 and Table 7 that all the site-specific pressure models show RMSE between 0.23 hPa and 0.26 hPa with an average correlation ranges between 0.973 and 0.997.

estimation of 0.26 hPa. The MnB errors are very small with an average value of 0.03 hPa and the

Table 7. Modeled Pressure (hPa) accuracy achieved when compared to sensor pressure (2014-2016)

Site	Pressure (hPa)		
	Corr	Bias	RMSE
ARSH	0.991	-0.07	0.26
DODM	0.993	0.05	0.24
MBEY	0.989	0.00	0.23
MTVE	0.997	0.14	0.29
NURK	0.976	0.03	0.26
<b>Mean</b>		<b>0.03</b>	<b>0.26</b>

**Relationship between the ERA5 pressure at the GPS stations and ERA5 surface pressure.**

The estimates of pressure from ERA5 at the 13 stations obtained from the ERA5 model levels data as explained in section 2.3 were first

assessed using the ERA5 surface pressure obtained as in section 2.4. Table 8 presents the correlation coefficients, MnB and RMSE estimates of the differences between the ERA5 model level pressure values and the ERA5

surface pressure data of 4 years (2013 to 2016).

From Table 8, the difference between the ERA5, model level pressure and surface pressure estimates at the sites presents that the average MnB and RMSE estimates are between 4.30 hPa

and 13.54 hPa, respectively. The MnB errors at specific stations vary between -20.91 hPa and 61.64 hPa. The RMSE estimates at different sites vary between 1.52 hPa and 61.64 hPa, with correlation coefficients greater than 0.986.

Table 8. Correlation, MnB and RMSE between ERA5 model level Pressure (hPa) at the 13 GPS station and the ERA5 surface pressure

Site	Pressure (hPa)		
	Corr	Bias	RMSE
ARSH	0.995	25.43	25.43
DODM	0.993	-3.41	3.42
KFNY	0.999	14.86	14.86
MAL2	0.996	2.68	2.69
MBAR	0.986	3.16	3.16
MBEY	0.997	-11.48	11.48
MOIU	1.000	6.77	6.77
MTDK	0.993	61.64	61.64
MTVE	0.998	7.92	7.92
NURK	1.000	-1.52	1.52
RCMN	1.000	8.35	8.35
SNGC	0.999	-20.91	20.91
TNDC	1.000	-7.89	7.89
<b>Mean</b>		<b>4.30</b>	<b>13.54</b>

In order to develop pressure linear regression models, a 1-year (2013) ERA5 model level and the corresponding ERA5 surface pressure data have been selected from each of the 13 GPS stations. The regression equations slopes and the y-

intercepts are shown in Table 9. The pressure computed with the developed models have been compared with ERA5 model level pressure data for years 2014-2016 of each station separately.

Table 9: The slopes and y-intercepts for the ERA5 pressure regression equations parameters

Site	Pressure (hPa)		
	R <sup>2</sup>	a	b
ARSH	0.987	1.03	-2.79
DODM	0.986	0.99	3.30
KFNY	0.999	0.93	45.14
MAL2	0.992	0.94	58.75
MBAR	0.967	0.93	66.38
MBEY	0.993	1.00	-11.33
MOIU	0.999	1.02	-4.61
MTDK	0.986	1.19	-104.47
MTVE	0.997	1.01	-2.48
NURK	0.999	1.00	0.40
RCMN	0.999	1.03	-14.64
SNGC	0.998	0.91	59.88
TNDC	1.000	0.98	14.52

According to Table 9,  $R^2$  is observed to range between 0.953 and 0.995. It has also been observed from Figures 4, 5 and Table 10 that the modeled pressure values when compared to ERA5 pressure values show RMSE values ranging between 0.03 hPa and 0.30 hPa at

individual stations with an overall average estimate of 0.15 hPa. The MnB error values are observed to be exceptionally small with average overall value of 0.03 hPa. The comparisons have also exhibited high correlation coefficients between 0.987 and 1.000.

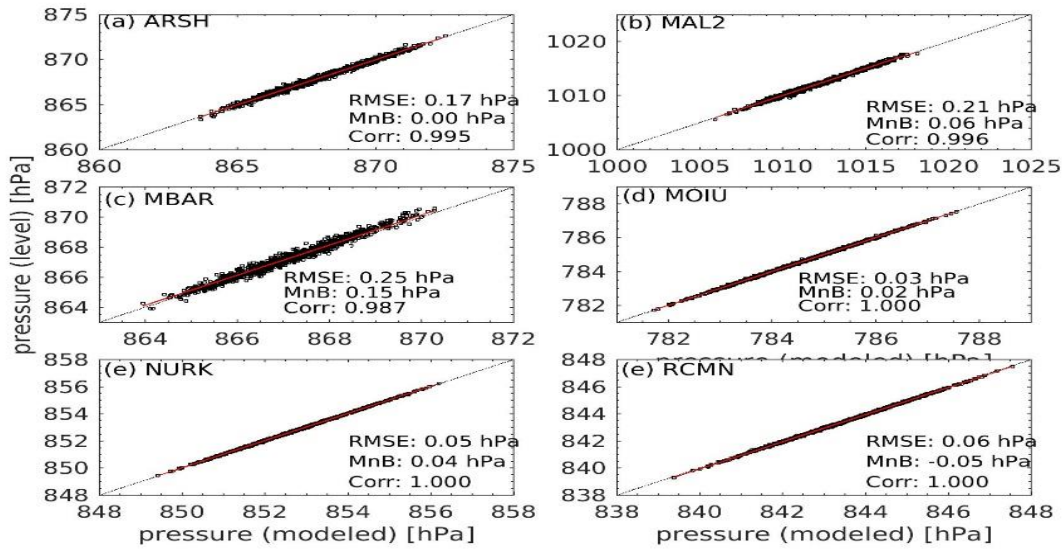


Figure 3. Scatter plots of modeled ERA5 pressure versus ERA5 model level pressure at the GPS stations in the region between latitudes  $4^{\circ}$  S and  $4^{\circ}$  N

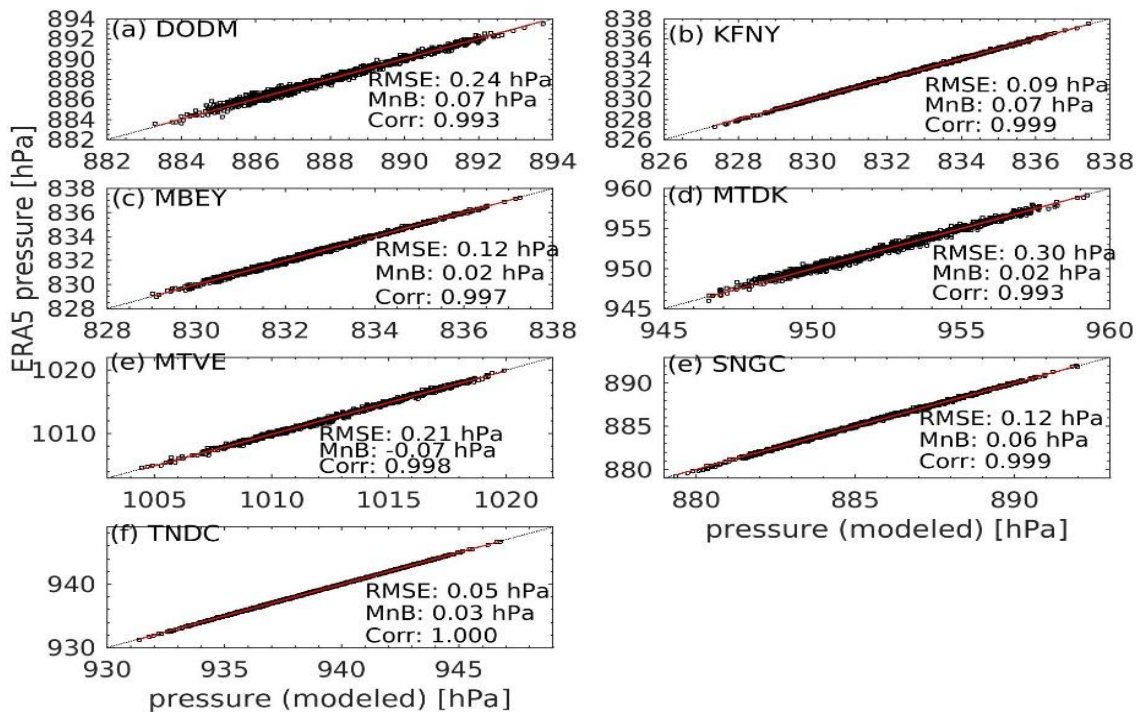


Figure 4. Scatter plots of modeled pressure versus ERA5 model level pressure at the GPS stations in the region between latitudes  $12^{\circ}$  S and  $4^{\circ}$  S

Table 10. ERA5 modeled pressure accuracy achieved with site specific models (2014-2016)

Site	Pressure (hPa)		
	Corr	Bias	RMSE
ARSH	0.995	0.00	0.17
DODM	0.993	0.07	0.24
KFNY	0.999	0.07	0.09
MAL2	0.996	0.06	0.21
MBAR	0.987	0.15	0.25
MBEY	0.997	0.02	0.12
MOIU	1.000	0.02	0.03
MTDK	0.993	0.02	0.30
MTVE	0.998	-0.07	0.21
NURK	1.000	0.04	0.05
RCMN	1.000	-0.05	0.06
SNGC	0.999	0.06	0.12
TNDC	1.000	0.03	0.05
<b>Mean</b>		<b>0.03</b>	<b>0.15</b>

**Correlation between the Weighted Mean Temperature at the GPS Stations and 2m Temperature**

In the development of the site specific ERA5  $T_m$  models for the 13 GPS stations, a 1-year (2013) ERA5  $T_m$  estimated using Eq. 6 and the corresponding 2m temperature data have been applied. The total number of samples used for each station is 365 for the one year. A linear regression between the two datasets was performed for each location. The regression equation coefficients are shown in Table 11. The linear regression analysis for  $T_m$  models

presents the correlation coefficients of determination between -0.002 and 0.464 as observed in Table 11. The accuracy of  $T_m$  with the site-specific models was validated with ERA5  $T_m$  for the years 2014 to 2016 for each GPS station and Table 12 shows the  $T_m$  accuracy achieved.

It is observed from Table 12 that all the  $T_m$  models show RMSE values ranging between 1.18 K and 1.69 K at specific stations with a general average value of 1.37 K. The MnB error values are seen to be very trivial with an average overall value of 0.01 k for the models.

Table 11: The slopes and y-intercepts for the weighted mean temperature regression equations

Site	R <sup>2</sup>	a	b
ARSH	0.231	0.353	179.05
DODM	0.033	0.16	237.66
KFNY	0.015	0.10	252.76
MAL2	0.260	0.67	88.14
MBAR	0.464	1.01	15.89
MBEY	0.033	0.17	230.93
MOIU	0.418	0.83	39.10
MTDK	0.227	0.28	203.18
MTVE	-0.002	-0.03	298.12
NURK	0.354	0.78	51.47
RCMN	0.327	0.55	121.82
SNGC	0.178	0.31	191.98
TNDC	0.224	0.40	167.16

Table 12. ERA5  $T_m$  accuracy achieved with site specific  $T_m$  models (2014-2016)

Site	Bias (K)	RMSE (K)
ARSH	0.05	1.18
DODM	-0.05	1.40
KFNY	0.03	1.55
MAL2	0.02	1.27
MBAR	-0.27	1.26
MBEY	-0.05	1.69
MOIU	-0.08	1.21
MTDK	0.03	1.18
MTVE	0.09	1.52
NURK	-0.01	1.28
RCMN	0.03	1.18
SNGC	-0.01	1.53
TNDC	0.14	1.52
<b>Mean</b>	<b>-0.01</b>	<b>1.37</b>

## Discussion

The results show that the RMSE and MnB of the PWV estimates derived from the linear regression models are much smaller than those obtained when GPS PWV was compared to the surface TCWV. Consequently, the results reveal that the site-specific models developed from surface TCWV data can estimate the GPS PWV at the stations sites with millimeter (mm) level of accuracy. The highest RMSE and MnB values for MTDK are credited to the vertical profiles extrapolation errors in the data utilized (Ssenyunzi *et al.*, 2019). The formulation of the linear regressions between the GPS PWV and the TCWV exhibited high coefficients of determination. The high coefficients of determination of the empirical relation connecting the PWV and TCWV were also reported by Acheampong and Obeng (2019).

Comparing empirical models that correlate the measured PWV data with surface meteorological parameters (Okulov *et al.*, 2002; Adeyemi, 2009b; Phokate, 2017; Falaiye *et al.*, 2018), the accuracy in PWV achieved when GPS PWV is correlated with TCWV has been improved. For example, the statistical correlations among the derived empirical models by Falaiye *et al.*, 2018, presented the RMSE values in the range between 3.3 mm to 8.6 mm. The empirical models predicting precipitable water vapour by Maghrabi and Al

Dajani (2012) presented RMSE in the range 3.54 mm to 10.78 mm. In this study, an average RMSE of 1.54 mm was presented by the developed linear regression equations. Therefore, the developed site specific models can be used as an alternative to predict PWV data in cases of station outages to supplement the existing GPS stations.

The results show that the RMSE and MnB of the pressure estimates derived from the modeled data sets are very low than those derived from the comparisons of the ERA5 model level and sensor pressure values with ERA5 surface pressure. The observed low RMSE values for pressure found in the tropical zone using the ERA5 data have so far been observed by Mateus *et al.*, (2020) and Ssenyunzi *et al.*, (2020). The results further reveal that MTDK has the highest RMSE values and this could be ascribed to the vertical profiles extrapolation and horizontal interpolation errors in data. It could likewise be ascribed to the uncompensated station elevations for the surface information. It is generally observed that the ERA5 pressure developed models can generate the pressure values with acceptable accuracies and they can fill in as supplements to the GPS observation networks referenced in this investigation.

It is however observed that the correlation coefficients of determination are below 0.5. This indicates that the relationships between  $T_m$  and

$T_s$  in the East African region is very low. Ross and Rosenfeld (1997), also reported low correlation at tropical radiosonde stations where the correlations were found to be less than 0.5. Also Mateus et al. (2020) established that the  $T_m$  and  $T_s$  values in the tropic zone have low coefficients of determination and this attributed to the lower temperature variations. The Intertropical Convergence Zone (ITCZ) is mainly the cause of this behavior in tropical regions. The ITCZ transfers ocean heat and moisture from the lower levels of the atmosphere to the upper levels of the troposphere and to medium and high latitudes and this creates prompt fluctuations in the air temperature and moisture with a less influence on the surface air temperature. Based on the low average RMSE, it is observed that the  $T_m$  models have produced  $T_m$  values suitable for this region. The results show that the site-specific  $T_m$  models can be used to provide  $T_m$  estimates of almost the same level of accuracy compared to ERA5  $T_m$  values at the 13 stations.

## Conclusion

In the study, site-specific PWV, pressure and weighted mean temperature models were developed from GPS PWV, pressure from sensors co-located and ERA5 model level pressure at the GPS stations and ERA5, surface pressure, 2 m temperature and the total column

## References

- Acheampong, A. & Obeng, K. (2019). Application of gnss derived precipitable water vapour prediction in West Africa. *J. Geod. Sci.*, 9:41-47.
- Adeyemi, B. (2009a). Empirical modeling of layered integrated water vapor using surface mixing ratio in Nigeria. *J. Appl. Meteor. Climatol.*, 48 (2), 369-380.
- Adeyemi, B. (2009b). Empirical formulations for inter-layer predictable water vapour in Nigeria. *Pacific J. Sci. Technol.* 10 (2), 35-45.
- Askne, J., & Nordius, H. (1987). Estimation of tropospheric delay for microwaves from surface weather data. *Rad. Sci.* 22, 379-386.
- Bevis, M., Businger, S., & Chiswell, S. (1994). GPS Meteorology: Mapping Zenith Wet Delays onto Precipitable Water. *J. Appl. Meteorol.* 33, 379-386.
- Bohm, J., Möller, G., Schindelegger, M., Pain, G., & Weber, R., (2015). Development of an improved blind model for slant delays in the troposphere (GPT2w). *GPS Solut.* 19:433.
- Chen, B., Dai, W., Liu, Z., Wu, L., & Xia, P. (2018). Assessments of GMI-Derived Precipitable Water Vapor Products over the South and East China Seas Using Radiosonde and GNSS. *Adv. Meteorol.* 7161328, 1-12.
- Davies, J. L., Herring, T. A., Shapiro, I. I., Rogers, A. E. E., & Elgered, G. (1985). Geodesy by radio interferometry: effects of atmospheric modeling on estimates of baseline length. *Rad. Sci.*, 20(6), 1593-1607.
- Durre, I., Vose, S., & Wuertz, D. (2006). Overview of the Integrated Global

water vapour data. The linear regression models were developed by performing a linear regression analysis using the Microsoft excel.

The developed PWV models using one-year data (2013) showed a general average RMSE of 1.54 mm when the modeled PWV estimates were compared to the corresponding GPS PWV data for years 2014-2016. The pressure estimates from models presented an average RMSE of 0.26 hPa when compared to pressure from the weather sensor. The estimates from ERA5 pressure models presented an overall average RMSE of 0.15 hPa, compared to ERA5 model level pressure values. The estimates for ERA5  $T_m$  models presented the average RMSE of 1.37 K when compared to the computed ERA5  $T_m$  values.

The developed PWV, pressure, and weighted mean temperature models can be used to predict the daily average values of the mentioned parameters with high accuracy. The developed linear regression models only require ERA5 surface pressure, 2m temperature and the total column water vapour datasets as inputs to generate the required parameters at the 13 stations under study.

## Acknowledgments

Ssenyunzi R. C. is thankful to Busitema University for the financial support.

- Radiosonde Archive. J. Clim. 19, 53-68.
- Falaiye, O. A., Abimbola, O. J., Pinker, R. T., Pérez-Ramírez, D., & Willoughby, A. A. (2018). Multi-technique analysis of precipitable water vapor estimates in the sub-Sahel West Africa. *Heliyon*, 4, e00765.
- Guerova, G., Brockmann, E., Schubiger, F., Morland, J., & Matzler, C. (2005). An integrated assessment of measured and modelled integrated water vapour in Switzerland for the period 2001-03. *J. Appl. Meteorol.* 44 (7), 1033-1044
- Gueymard, C. (1994). Analysis of monthly average atmospheric precipitable water and turbidity in Canada and northern United States. *Solar Energy* 53, 57-71.
- Hagemann, S., Bengtsson, L., & Gendt, G. (2003). On the determination of atmospheric water vapor from GPS measurements. *J. Geophys. Res.*, 108.
- Herrera, A. M., Suhandri, H. F., Realini, E., Reguzzoni, M., & Clara de Lacy, M. (2016). gogps: open-source matlab software. *GPS solut.* 20, 595-603.
- Hersbach, H., Dee, D. (2016). ERA5 reanalysis is in production. *ECMWF Newsletter* No. 147, 7.
- Hong, L., Yunchang, C., Xiaomin, W., Zhifang, X., Haishen, W., & Heng, H. (2015). Meteorological applications of precipitable water vapor measurements retrieved by the national GNSS network of China. *Geod. Geody.* 6(2), 135-142.
- Hsu, S. A., & Blanchard, B. W. (1989). The relationship between total precipitable water and surface level humidity over the sea surface. A further evaluation. *J. Geophys. Res.* 94, 14539 - 14545.
- Huang, L., Mo, Z., Liu, L., Zeng, Z., Chen, J., Xiong, S., & He, H. (2021). Evaluation of hourly PWV products derived from ERA5 and MERRA-2 over the Tibetan Plateau using ground based GNSS observations by two enhanced models. *Earth and Space Science*, 8, e2020EA001516. <https://doi.org/10.1029/2020EA001516>
- Hussain, M. (1984). Estimation of global and diffuse irradiation from sunshine duration and atmospheric water vapour content. *Solar Energy*, 33 (2), 217-220.
- Landskron, D., & Böhm, J., 2018. VMF3/GPT3: refined discrete and empirical troposphere mapping functions. *J. Geod.* 92:349-360.
- Leckner, B. (1978). The spectral distribution of solar radiation at the earth's surface elements of a model. *Solar Energy*, 20, 143-150.
- Li, G., & Deng, J. (2013). Atmospheric Water Monitoring by Using Ground-Based GPS during Heavy Rains Produced by TPV and SWV. *Adv. Meteorol.* 1-12. <http://dx.doi.org/10.1155/2013/793957>.
- Maghrabi, A. H. (2009). Parameterization of a simple model to estimate monthly global solar radiation based on meteorological variables, and evaluation of existing solar radiation models for Tabouk, Saudi Arab Energy Conserv. Manage. 50, 2754-2760.
- Maghrabi, A. & Al Dajani, H. M. (2012). Estimation of precipitable water vapour using vapour pressure and air temperature in an arid region in central Saudi Arabia, *J. Assoc. of Arab Universities for Basic and Applied Sciences*, 14, 1 - 8.
- Maghrabi, A. H., & Clay, R. (2010). Precipitable water vapour estimation on the basis of sky temperatures measured by a single-pixel IR detector and screen temperatures under clear skies. *J. Meteorol. Appl.* 17, 279-286.
- Mateus, P., Catalao, J., Mendes, B., & Nico, G. (2020). An era5-based hourly global pressure and temperature (hgpt) model. *Remote Sens.*, 12(1098), 1-18.
- Okulov, O., Ohvril, H., & Kivi, R. (2002). Atmospheric precipitable water in Estonia, 1990 - 2001. *Boreal Env. Res.* 7, 291 - 300.
- Phokate, S. (2017). Atmospheric water vapor: Distribution and Empirical estimation in the atmosphere of Thailand. *J. Phys.: Conf. Ser.*, 901, 012051.

- Reitan, C. H. (1963). Surface dew point and water vapor aloft. *J. Appl. Meteorol.* 2, 776- 779.
- Ross, R., & Rosenfeld, S. (1997). Estimating mean weighted temperature of the atmosphere for global positioning system applications. *J. Geophys. Res.*, 102(D18), 21719-21730.
- Ruckstuhl, C., Philipona, R., Morland, J., & Ohmura, A. (2007). Observed relationship between surface specific humidity, integrated water vapor, and longwave downward radiation at different altitudes. *J. Geophys. Res.* 112, D03302.
- Saastamoinen, J. (1972). Atmospheric correction for the troposphere and stratosphere in radio ranging of satellites. *Geophysical Monograph Series. American Geophysical Union, Band 15, Washington, D.C., pp. 247-251.*
- Singh, D., Kumar, G. J., & Deepak, K. (2014). Precipitable water vapor estimation in India from GPS-derived zenith delays using radiosonde data. *Meteorol. Atmos. Phys.* 123:209–220. DOI 10.1007/s00703-013-0293-1.
- Ssenyunzi, R.C., Oruru, B., Dujanga, F.M., Realini, E., Barindelli, S., Tagli aferro, G., & van de Giesen, N. (2019). Variability and accuracy of Zenith Total Delay over the East African Tropical Region. *Adv. Space Res.*, 64, 900 - 920.
- Ssenyunzi, R. C., Oruro, B., D’ujanga, F.M., Realini, E., Barindelli, S., Tagliaferro, G., van de Giesen, N., & von Engeln, A. (2020). Performance of ERA5 data in retrieving Precipitable Water Vapour over East African tropical region. *Adv. Space Res.* 65, 1877-1893. doi:<https://doi.org/10.1016/j.asr.2020.02.003>.
- Tuller, S. (1977). The relationship between precipitable water vapour and surface humidity in New Zealand. *Arch. Meteor. Geophys. Bioklim.* A26, 197-212.
- Wang, Y., Tang, L., Gao, T., Wang, Q., Lu, C., Song, Y., & Hua, D. (2018). Investigation and Analysis of All-Day Atmospheric Water Vapor Content over Xi’an Using Raman Lidar and Sunphotometer Measurements. *Remote Sens.* 10, 951. Doi 10.3390/rs10060951.