Received: 9 September 2020 | Revised: 9 July 2021 | Accepted: 7 September 2021

DOI: 10.1002/met.2030

Meteorological Applications
Science and Technology for Weather and Climate
Open Access
RMetS

RESEARCH ARTICLE

Regression models for predicting daily IGS zenith tropospheric delays in West Africa: Implication for GNSS meteorology and positioning applications

Samuel Osah | **Akwasi A. Acheampong** | **Collins Fosu** | **Isaac Dadzie**

Department of Geomatic Engineering, College of Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Correspondence
Samuel Osah, Department of Geomatic Engineering, College of Engineering, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.
Email: osahsamuel@yahoo.ca

Abstract

The ability to precisely and accurately model and predict tropospheric delay is essential for precise global navigation satellite system (GNSS) and meteorological applications. The International GNSS Service (IGS) provides highly accurate and highly reliable daily time series zenith tropospheric delay (ZTD) products for all its member sites using data from each IGS site. Nevertheless, if for reasons such as poor internet connectivity, equipment failure, and power outages the IGS station is inaccessible, gaps are created in the data archive, resulting in degrading the quality of the ZTD estimation, as well as inhibits the quality of precipitable water vapour (PWV) estimation, needed for precise positioning applications, meteorological studies, and weather forecasting. To address this challenge, five regression models are proposed in this study to model and predict daily ZTDs using daily datasets from four IGS stations in West Africa over a period of 5 years (2015–2019). The site-specific Vienna Mapping Functions 3 (VMF3) products (ZTD, pressure, temperature, water vapour partial pressure) and stations' coordinates (latitudes and longitudes) are used as the predictors, while the IGS final ZTD product as the response variable in fitting the models. Several performance measures are calculated to compare the predictive performance of the models. The results show that the five regression models performed outstandingly and agree very well with the IGS-ZTD data, and hence provide a useful alternative for ZTD predictions and also in the event the West African IGS stations' ZTD data are unavailable. Nonetheless, the support vector regression model outperformed the remaining four models.

KEYWORDS

global navigation satellite system (GNSS), performance measures, regression models, Vienna Mapping Functions 3 (VMF3), zenith tropospheric delay (ZTD)

1 | INTRODUCTION

The ability to precisely and accurately model and predict tropospheric delays has become an important field of research in the last few decades (Li et al., 2020; Yang et al., 2020, 2021). Tropospheric delay induces an excess propagation path length on global navigation satellite system (GNSS) signals when propagating through the

troposphere due to the variations in refractivity along the signal transmission path (Mendes, 1999). The excess propagation path length introduces an error in the measured range, and the end degrades the positioning accuracy, particularly the vertical component (Nikolaidou et al., 2018).

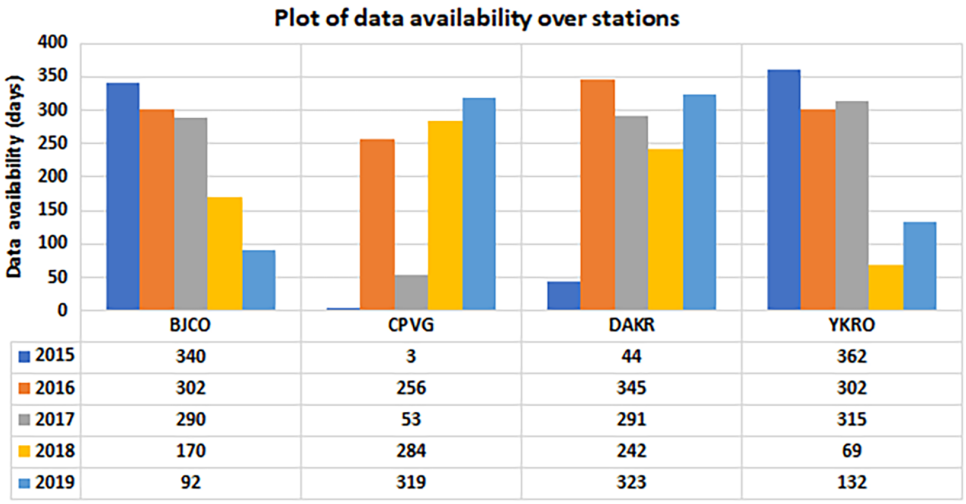
The impact of the tropospheric delay on GNSS signals ranges from approximately 2.0 to 2.6 m in the zenith direction to approximately 20–28 m near the horizon and at lower elevation angles (Sanlioglu & Zeybek, 2012). The fact that tropospheric delay cannot be removed by a combination of dual or multi-frequency GNSS observations owing to its non-dispersive nature, unlike the ionospheric delay, the tropospheric delay is considered as the major error source in GNSS applications and, therefore, remain a major challenge and a growing concern to GNSS users. Precise and accurate modelling and/or mitigation of tropospheric delay is, therefore, required for precise positioning and navigation applications. Moreover, precise estimates of tropospheric delay can as well be utilized to infer precipitable water vapour (PWV) in the atmosphere with high spatial and temporal resolution (Bevis et al., 1992, 1994), which plays a critical role in meteorological studies, weather forecasts, and climate monitoring.

With the growing demand for GNSS technology, a dense network of permanent and continuously operating IGS stations has been established and distributed globally with the aim of using GNSS for precise geodetic applications and a tool for remote sensing of water vapour in the atmosphere (Hadas et al., 2017). Owing to the non-uniform spatial and temporal distribution of water vapour content in the atmosphere, several techniques and models have evolved over the years towards the estimation and mitigation of tropospheric delay (Hofmann-Wellenhof et al., 2008; Jiang et al., 2020). The models are grouped into surface meteorological parameter-dependent models (Black, 1978; Hopfield, 1969; Saastamoinen, 1972) and user location and time-dependent models (Landskron & Böhm, 2018; Leandro et al., 2008; Penna et al., 2001). Recently, the application of different techniques such as precise point positioning (PPP) (Hadas et al., 2013; Zhao et al., 2018; Zhu et al., 2010), artificial neural network (ANN) (Ding & Hu, 2019; Yang et al., 2017; Zhang et al., 2020), and adaptive neuro-fuzzy inference system (ANFIS) (Suparta & Alhasa, 2013, 2015, 2016) has also proven useful in modelling and predicting tropospheric delay in the zenith direction (ZTD), since the empirical models cannot provide enough accuracy for precise positioning (Yang et al., 2017).

The IGS provides, among other products, high-quality continuous daily time series ZTD products at an accuracy of approximately 4 mm for all IGS ground sites using PPP technique and data in rinex format generated by each IGS site (Byun et al., 2005). Due to the high quality of the IGS-ZTD product, it is being utilized as the standard reference for validating the performances of other ZTD models or products and also the retrieval of PWV for research and meteorological applications (Ding & Hu, 2019; Yao et al., 2018). The quality of the IGS-ZTD estimates depends on the quality and continuous daily data archive at the various IGS ground stations (Byun et al., 2005). However, if for some reasons, such as poor internet connectivity, equipment failure, and electrical power problems or outages (Isioye et al., 2015; Ssenyunzi et al., 2019), the IGS ground sites are inaccessible or absent, gaps are created in the data archive owing to the inconsistency of data streaming, and the quality of the tropospheric delay estimation is degraded (Jgouta et al., 2016). This as well inhibits the quality and continuous retrieval of the PWV by the IGS sites. The African IGS ground sites, apart from being limited in number, are also noted for having large data gaps (Isioye et al., 2015; Ssenyunzi et al., 2020; Walpersdorf et al., 2007), as evidence from Figure 1, and for that reason lack providing continuous daily data, which are resourceful and adequate to provide high spatial and temporal resolution ZTD data. Thus, the accurate quantification of ZTD remains a challenge in West Africa. Therefore, the need to investigate alternative methods for predicting ZTD to augment and improve the continuous availability of ZTD data at the IGS stations in West Africa is very crucial.

This paper proposes to apply five statistical regression models, multiple linear regression (MLR), quadratic regression (QR), generalized linear model (GLM), generalized additive model (GAM), and support vector regression (SVR), to build a reliable and accurate prediction model for predicting daily IGS final ZTDs in West Africa. Moreover, if the under-studied IGS tracking stations' ZTD data are inaccessible or absent, the best prediction model can as well be utilized as an alternative model to predict the missing ZTD values as a means of improving continuous data availability at the stations. These models were selected following Clements et al. (2005), Sayegh et al. (2014)), Parveen et al. (2016, 2017, 2020), Acheampong and Obeng (2019)), and Zhang et al. (2019), as these models have been commonly used in other areas of study.

In recent years, numeric weather prediction (NWP) re-analysis data from worldwide used models such as those of European Centre for Medium-Range Weather Forecasts (ECMWF), the National Centres for Environmental Prediction (NCEP), and National Centre for Atmospheric Research (NCAR) have become a source of atmospheric data for ZTD modelling by many research institutes and groups as a means of improving space geodetic techniques such as GNSS (Hobiger et al., 2008; Li



	BJCO	CPVG	DAKR	YKRO
2015	340	3	44	362
2016	302	256	345	302
2017	290	53	291	315
2018	170	284	242	69
2019	92	319	323	132

FIGURE 1 Data availability (in days) for the IGS stations in West Africa from 2015 to 2019, indicating the stations' varying performances of data acquisition

et al., 2020; Urquhart et al., 2014; Zhou et al., 2020). The Vienna Mapping Functions (VMF: VMF1 and VMF3) service, formally the Global Geodetic Observing System (GGOS) atmosphere project developed at the Vienna Technical University (TU Wien), is one such institution. The VMF service, among other products, offers daily continuous site-specific ray-traced zenith tropospheric delays (ZHD, ZWD), surface meteorological parameters (P, T, and e), and mapping functions derived from the ECMWF ERA-40 (for VMF1) and ERA-Interim (for VMF3) reanalysis data at 6 h daily and 4 epochs (00, 06, 12 and 18 UT) temporal resolution for IGS stations worldwide. VMF1 is the most widely used product. The precision and accuracy of the VMF service products have been investigated and validated against the IGS zenith delays by a number of researchers (Yao et al., 2017, 2018; Osah et al., 2021a) to be accurate enough for tropospheric delay mitigation and model's evaluation. Tropospheric delay models have also been developed out of the service products. For example, Yao et al. (2013) analysed the temporal and spatial variations of the GGOS atmosphere data and built a new global zenith tropospheric delay model (GZTD) using a spherical harmonic function. Likewise, Sun et al. (2017) also derived a global zenith tropospheric delay simplified model (GZTDS) with the assumption that the troposphere is a nonlinear system using the GGOS atmosphere data. Ding and Hu (2019) and Osah et al. (2021b) also utilized the GGOS atmosphere and VMF3 data to develop an ANN and deep learning (DL) techniques-based ISAAS and DLztd models over Russia and West Africa, respectively. In addition, the VMF1 mapping functions have been reported as the most accurate mapping functions to date (Landskron & Böhm, 2018; Tesmer et al., 2007), and their use is recommended for high-precision geodetic applications by the International Earth Rotation and Reference Systems Service (IERS) (Feng et al., 2020; Nikolaidou et al., 2018; Urquhart et al., 2014), and also for all precise geophysical applications (Boehm & VanDam, 2009). All of these studies provide a clear indication and proof that the VMF service products are reliable, accurate, and suitable for use in precise GNSS applications and tropospheric delay modelling.

In this study, we adopted the recent version of the VMF service products, the site-specific operational VMF3-ZTD products (including ray-traced zenith delays, surface meteorological parameters (pressure (P), temperature (T), and water vapour partial pressure (e)) for global IGS stations, as well as the coordinates of the stations [latitude (φ) and longitude (λ)], as inputs or predictors for developing the regression models, and the models' prediction performances are evaluated using six different statistical evaluation metrics. This study will go a long way in addressing data gaps issues with the West African IGS stations and also enhance research on GNSS positioning and meteorological applications in the region, which has seen limited studies in these disciplines due to the limited GNSS network and large data gaps.

In the last decade, ANN and ANFIS techniques have been used extensively in various studies to model and predict ZTD, which have shown good results. For example, Pikridas et al. (2010) developed an ANN-based ZTD model for predicting ZTD over some selected EUREF permanent GNSS network (EPN) stations. The predicted results showed an MB ranging from −10.9 cm to 11.5 cm and a root mean square error (RMSE) of 3.6 cm. Zheng et al. (2015) also employed the same technique to develop a regional ZTD prediction model over Jiangsu Province of China, providing an RMSE of 0.004 m. Likewise, Ding et al. (2016) and Ding and Hu (2019) also established an ANN-based ZTD prediction model over Russia using

ground-based GNSS-ZTD data from 11 IGS stations. The predicted results gave a bias of −4.4 mm and RMSE of 20.4 mm. Correspondingly, Yang et al. (2017) used the backpropagation neural network (BP NN) technique to develop a regional ZTD model using 15 continuously operating reference stations (CORS) in Hong Kong. The accuracy of their model was reported to be 1.1 cm. Suparta and Alhasa (2013) proposed an ANFIS model (a technique that combines neural networks and fuzzy inference system) for predicting ZTD using five selected stations in Antarctica and three selected stations in Malaysia and Singapore regions. The results showed an RMSE of 0.0175 mm. Li et al. (2020) equally proposed a regional ZTD model based on generalized regression neural network (GRNN). The GRNN model was developed using meteorological data from the ECMWF and the United States National Centres for Environmental Prediction (NCEP). Using the IGS-ZTD product as a reference over 100 globally distributed IGS stations, the experimental results showed a mean bias (MB) of 9.5 mm and RMSE of 12.7 mm. Of all these studies, statistical regression methods have received little attention for modelling and predicting ZTD, as limited or no studies have been conducted so far to explore their applicability, especially in the West African region. Regression analyses, however, have been proven useful and effective in solving many problems in several fields of studies for modelling and making predictions. For instance, In the field of health, for the prediction of lung cancer rate (Clements et al., 2005), heart disease (Polaraju & Prasad, 2017), and CO_2 emissions (Libao et al., 2017), in the field of environmental science, for the prediction of PM_{10} concentration (Ul-Saufie et al., 2011; Sayegh et al., 2014), in the field of engineering, for the prediction of deformation behaviour of concrete-face rockfill dam (Wen et al., 2021) and inflow to the Ujjani dam reservoir (Rajmane & Waikar, 2020), in the field of hydrology, for the prediction of monthly streamflows (Zhang et al., 2019), development of flood damage risk map for forecasting rainfall data (Choi, 2021), in the field of meteorology (GNSS), for the prediction of weighted mean temperature (Bevis et al., 1992; Isioye et al., 2016; Schueler et al., 2001) and PWV (Acheampong & Obeng, 2019), in ionospheric studies, for the prediction of global position system (GPS) ionospheric delay and scintillation (Aon et al., 2018; Zhang et al., 2019), and many more. As noted above, regression analysis has been a widely used statistical technique for developing prediction models across various disciplines, largely due to its ability to analyse the relationship between response and predictor variables without requiring a large number of training datasets for optimal prediction compared with ANN, which requires a considerable number of training datasets. This study is aimed at investigating the applicability of regression models in modelling and predicting ZTD in West Africa for precise positioning and meteorological applications. The hypothesis (null (H_0) and alternate (H_1)) guiding this study is as follows:

H0. *There is no significant variation in the performance of the regression models.*

H1. *There is significant variation in the performance of the regression models.*

This paper is structured as follows: Section 2 describes the basic theory for the tropospheric delay modelling, while data presentation, the methodology of models' development, and performance evaluation between models' comparison are given in Section 3. Section 4 presents the experimental results and discussion for each regression model obtained. The conclusions are given in Section 5.

2 | TROPOSPHERIC DELAY

The troposphere forms the lower portion of the neutral atmospheric layer up to approximately 50 km (Meunram & Satirapod, 2019). The effect of the troposphere on GNSS signals is called tropospheric delay, which induces an extra delay on GNSS measurements. Tropospheric delay is a function of the total refractivity (N), which depends on pressure [P (mbar or hpa)], temperature [(T (K)] and relative humidity [RH (%)] or water vapour partial pressure [e (mbar or hpa)] along the signal transmission path as well the location of the receiver antenna. The total delay computed in the zenith direction is called the zenith tropospheric/total delay (ZTD), which is the sum of the zenith hydrostatic delay (ZHD) and the zenith wet delay (ZWD). The ZHD accounts for approximately 90% of the ZTD and can easily be modelled or predicted to sub-millimetre accuracy by empirical prediction models using surface meteorological observations, whereas the ZWD accounts for the remaining 10% of the ZTD. In contrast to the ZHD, the ZWD cannot be precisely modelled or predicted due to its large spatial and temporal variability (Younes, 2016; Zhang et al., 2016). The troposphere is characterized as being a non-dispersive medium for radio frequencies up to 15 GHz, hence its effect is independent of GNSS frequencies and, therefore, affects carrier phase and code measurements equally (Sanz et al., 2013).

Following Bevis et al. (1992), Mendes (1999)), and Kleijer (2004)), the excess propagation path (ΔL_r^t) can be

calculated by integration of the refractive index along the signal transmission path, s, from satellite, t to a receiver, r, at elevation angle (θ) via the troposphere as follows:

$$\Delta L_r^t(\theta) = \underbrace{\int_s (n(s)-1)ds}_{\text{Path delay}} + \underbrace{\left[\int_s ds - \int_G dG\right]}_{\text{Signal bending}}. \quad (1)$$

Equation (1) can also be written as follows:

$$\Delta L_r^t(\theta) = \int_s (n(s)-1)ds + [S-G], \quad (2)$$

where S is the curved signal path, G is the geometric signal path, $n(s)$ is the refractive index along the signal transmission path.

In the zenith direction where satellites are directly overhead the receiver (i.e., $\theta = 90^\circ$), the signal path is a straight line (i.e., S and G become identical), and the bending effect disappears ($S - G \approx 0$, Equation 2) with the assumption that the troposphere is horizontally stratified (Bevis et al., 1992; Mendes, 1999). Expressing the signal delay in the zenith direction (i.e., $\theta = 90^\circ$), Equation (2) now becomes:

$$\Delta L = \int_S [n(s)-1]ds. \quad (3)$$

Expressing n in terms of total refractivity N, $N = 10^6(n-1)$ where Equation (3) can be rewritten as follows:

$$\Delta L = 10^{-6}\int_s N(s)ds. \quad (4)$$

The refractivity (N) can be related to the meteorological parameters (P, T, e) as (Smith & Weintraub, 1953):

$$N = \left[k_1\frac{P_d}{T}\right] + \left[k_2\frac{e}{T} + k_3\frac{e}{T^2}\right] = N_h + N_w, \quad (5)$$

where k_1, k_2, k_3 are refractivity constants whose values can be inferred from Bevis et al. (1994) as 77.60 ± 0.05 K/hpa, 70.4 ± 2.2 K/hpa, and $3.739 \times 10^5 \pm 0.012\,\mathrm{K}^2$/hpa, respectively, $P_d = P + e$, is the partial pressure of the dry gases (in hpa), P is the total surface pressure (in hpa), T is the surface temperature (in kelvin (K)), and e, the water vapour partial pressure (in hpa).

Equation (5) can further be rewritten with the assumption that the air in the troposphere behaves as an ideal gas, Equation (5) becomes (Nafisi et al., 2012; Yuan et al., 2019)

$$N = \left[k_1\frac{P}{T}\right] + \left[k_2'\frac{e}{T} + k_3\frac{e}{T^2}\right] = N_h + N_w, \quad (6)$$

where $k_2' = k_2 - k_1\frac{M_w}{M_d} = 22.1 \pm 2.2\,\mathrm{K\ hPa^{-1}}$ (Bevis et al., 1994). M_w and M_d are the molar mass of water vapour and dry air, respectively.

For NWP models, N can be computed as (Chen et al., 2011; Hu & Yao, 2019; Jiang et al., 2020; Lou et al., 2018; Wilgan & Geiger, 2019; Yuan et al., 2019)

$$N = k_1\frac{P-e}{T} + k_2\frac{e}{T} + k_3\frac{e}{T^2}. \quad (7)$$

Here, P and T are based on pressure-level data, and e is computed as follows:

$$e = q \times P/0.622, \quad (8)$$

where q is the specific humidity (in kg/kg or unitless).

Re-writing Equation (4) in terms of ZTD,

$$\begin{aligned}\mathrm{ZTD} &= 10^{-6}\int_h^{TOA} N_d(h)dh + 10^{-6}\int_h^{TOA} N_w(h)dh \\ &= \mathrm{ZHD} + \mathrm{ZWD},\end{aligned} \quad (9)$$

where h is the height of the station or receiver antenna, and TOA is the top of the atmosphere.

For NWP models, the final ZTD is computed as (Chen et al., 2011; Hu & Yao, 2019; Jiang et al., 2020; Lou et al., 2018; Wilgan & Geiger, 2019; Yuan et al., 2019):

$$\mathrm{ZTD} = \mathrm{ZHD}_{\mathrm{top}} + \mathrm{ZTD}_{\mathrm{level}}, \quad (10)$$

$\mathrm{ZHD}_{\mathrm{top}}$ is the delay at the top level, computed using the Saastamoinen model as (Saastamoinen, 1972):

$$\mathrm{ZHD}_{\mathrm{top}} = 0.002277 \cdot \frac{P_{\mathrm{top}}}{1 - 0.00266\cdot\cos(\varphi) - 2.8\times10^{-7}h_{\mathrm{top}}}, \quad (11)$$

$$\mathrm{ZTD}_{\mathrm{level}} = 10^{-6}\int_h^{h_{\mathrm{top}}} Ndh = 10^{-6}\sum_{i=1}^{\mathrm{top}} N_i\Delta s_i, \quad (12)$$

where P_{top} is the pressure at the top level (in hpa), φ is the latitude of the station, h_{top} is the top-level altitude (in metres), N_i is the total refractivity at the ith level,

computed using Equation (7), top is the number of levels of the NWP model, and Δs_i is the geometric distance (in metres) between the ith and $(i + 1)$-th layer of the NWP model given as (Hu & Yao, 2019)

$$\Delta s_i = Z_{i+1} - Z_i, \tag{13}$$

Z represents the height of the pressure level (in metres).

With an appropriate mapping function and using the satellite elevation angle (θ) as input, the ZTD can be resolved into the slant tropospheric delay (STD) as follows:

$$\mathrm{STD}_r^t = \mathrm{ZHD} \times \mathrm{MF}_h(\theta) + \mathrm{ZWD} \times \mathrm{MF}_w(\theta), \tag{14}$$

where MF_h and MF_w represent the hydrostatic and wet mapping functions.

ZWD can then be converted to PWV for meteorological applications using a conversion factor pie (Π) as (Bevis et al., 1994):

$$\mathrm{PWV} = \Pi * \mathrm{ZWD}, \tag{15}$$

Π is dimensionless and can be determined from Askne and Nordius (1987)) as follows:

$$\Pi = \frac{10^6}{\left(\frac{k_3}{T_m} + k_2^{'}\right)\rho_w R_w}, \tag{16}$$

where R_w is the specific gas constant for water vapour (461.5 J kg^{-1} K^{-1}), ρ_w is the density of water (998 kg m^{-3}), T_m is the weighted mean temperature in kelvin (K), $k_2^{'}$ is refractivity constant (16.52 K/hpa).

3 | MATERIALS AND METHODS

3.1 | Study area

This study was conducted using four IGS stations located in four West African countries, namely, Benin (BJCO), Cote d'Ivoire (YKRO), Senegal (DAKR), and Cape Verde (CPVG). The West African region lies between latitudes 0° N and 25° N and longitudes 20° W and 20° E. The details of the stations and their visual locations are shown in Table 1 and Figure 2, respectively.

3.2 | Datasets

The study utilized the daily site-wise operational VMF3 ZTD products for IGS stations and the IGS final ZTD products for a period of 5 years spanning from 2015 to 2019. The following subsections summarize the datasets used in this study.

3.2.1 | VMF3-ZTD data

The VMF service (https://vmf.geo.tuwien.ac.at/) provides both gridded and site-wise ray-traced VMF1 and VMF3 tropospheric products as open access data for global and more than 500 IGS site users. They are generated based on ray-traced delays at 3° elevation using NWM data spanning 2001–2010 from the ECMWF ERA-40 and ERA-Interim re-analysis, respectively (Putri et al., 2020); and are provided every 6 h daily at four epochs (00, 06, 12 and 18 UT) and 23 h UT the next day (Boehm et al., 2009). VMF3 is the successor of VMF1 realized on both 1° × 1° and 5° × 5° global grids (Landskron & Böhm, 2018). The site-wise VMF3-ZTD data fields for the IGS stations include ZHD, ZWD, P, T, e, and mapping function coefficients ah and aw for both hydrostatic and wet components, respectively. The ZTD is obtained by adding the ZHD and ZWD (i.e., ZTD = ZHD + ZWD). The VMF3-ZTD data are available at https://vmf.geo.tuwien.ac.at/trop_products/GNSS/VMF3/.

3.2.2 | IGS-ZTD data

The IGS also publishes daily final ZTD products based on PPP techniques combined with IGS final orbit and clock solutions for all IGS ground-based tracking stations with an accuracy of approximately 4 mm. The zenith delays are computed at an interval of 5 min from the IGS stations and are released with 4 weeks latency (Byun et al., 2005). The IGS-ZTD data are available at ftp://cddis.nasa.gov/gnss/products/troposphere/zpd/.

3.3 | Data extraction

The downloaded daily datasets were extracted into Microsoft Excel for easy exploration and analysis. The extractions were done using coded scripts or m-files, '***read_vmfGNSS.m***' and '***ReadIGSTrop.m***' written in MATLAB programming language for this study. For the VMF3-ZTD data, the '***read_vmfGNSS.m***' m-file was used to extract all the needed parameters in the data file. Parameters extracted were ZHD, ZWD, P, T, and e for each epoch (00, 06, 12, and 18 UT) daily and the first epoch (00 h UT) the next day to obtain complete data for 24 or 23 h UT. The ZTD was obtained following Equation (9). No interpolation or extrapolation was performed, since the site-wise parameters are computed at station height and, therefore, no extrapolation was required to transfer from grid height to station height. To obtain the daily (24 hours) average datasets, the

TABLE 1 IGS stations in West Africa selected for this study

Site	City	Country	Latitude	Longitude	Ellipsoidal height	Period of data
BJCO	Cotonou	Benin	6.3847° N	02.4500° E	30.700 m	2015–2019
YKRO	Yamoussoukro	Cote d'Ivorie	6.8706° N	05.2401° W	270.000 m	2015–2019
DAKR	Dakar	Senegal	14.7212° N	17.4395° W	51.000 m	2015–2019
CPVG	Espargos	Cape Verde	16.7321° N	22.9349° W	94.089 m	2015–2019

FIGURE 2 Map of West Africa showing the locations of the selected IGS stations (GoogleMap, 2021). The red round markers indicate the IGS stations

mean of each daily parameter (ZTD, P, T, e) extracted was computed by summing all the daily epochs (00, 06, 12, and 18 UT) and the first epoch (00 h UT) the next day, and dividing by 5 as presented in Equation (17). The station coordinates for which these parameters are valid were as well extracted from the station coordinate file, '***gnss. ell***'.

$$V = \frac{v_{00} + v_{06} + v_{12} + v_{18} + v_{00\ \mathrm{nextday}}}{5}, \tag{17}$$

where V is the computed average parameter of interest, v is the parameter of interest, 00, 06, 12, 18 are the daily epochs, and 00nextday is the first epoch of the following day.

Similarly, the IGS-ZTD data were also extracted using the '***ReadIGSTrop.m***' m-file. Since the IGS-ZTD is computed at an interval of 5 min (temporal resolution) for each day, the mean ZTD was also calculated for the entire day (24 h) for each file, by summing all the ZTDs at each epoch and dividing by the total number of epochs in the file, presented as follows:

$$\mathrm{ZTD_{igs}} = \sum_{i=1}^{n} \frac{\mathrm{ZTD_E}}{n} \tag{18}$$

where $n = 288$ is the total number of epochs, $\mathrm{ZTD_E}$ is the ZTD at each epoch.

For each IGS station, the two datasets (VMF3 and IGS) were compared and the corresponding days of data availability were extracted and sorted for the models' development. The total number of days of data available for each station in a given year within the study period is shown in Figure 3.

3.4 | Development of regression models

Regression analysis is one of the most ordinarily utilized statistical modelling methods for modelling relationships between a dependent variable (often called the 'response variable') and independent variables (often called 'predictors', 'covariates', or 'features') (Yazar et al., 2017). It employs several models and methods (linear and non-linear) for fitting the relationships between the response variable and two or

more predictor variables, which can potentially be used for predicting the response variable given a set of predictor variables. Research has shown that linear regression analysis has often been preferred for fitting experimental data. Conversely, non-linear regression is a powerful alternative to linear regression, since it offers the most flexible functionality for curve fitting. Linear regression analysis may be inadequate in some instances to describe experimental data, and in these instances, non-linear regression offers the best performance for system description (Yazar et al., 2017). This paper considers five regression models, both linear (MLR, GLM,) and non-linear (QR, GAM, SVR), to relate the IGS final ZTDs (response variable) and the site-wise VMF3-ZTD products including P, T, e, and ZTD as well as the stations' latitudes φ, longitudes λ, and ellipsoidal heights (h) (predictor variables) using R studio statistical computing software. The 5 years of data (2015–2019) were split into training data (2015–2018) and test data (2019). The training data are used for developing the models, while the test data, which are an independent set (not used in the model development process), are used for assessing the models' predictive performance. Figure 3 provides a summary flow of the methodology.

3.4.1 | Detection of multicollinearity

Variables selection plays an essential role in fitting a regression model. Meteorological parameters (P, T, e) and ZTD extracted from the site-wise VMF3-ZTD data

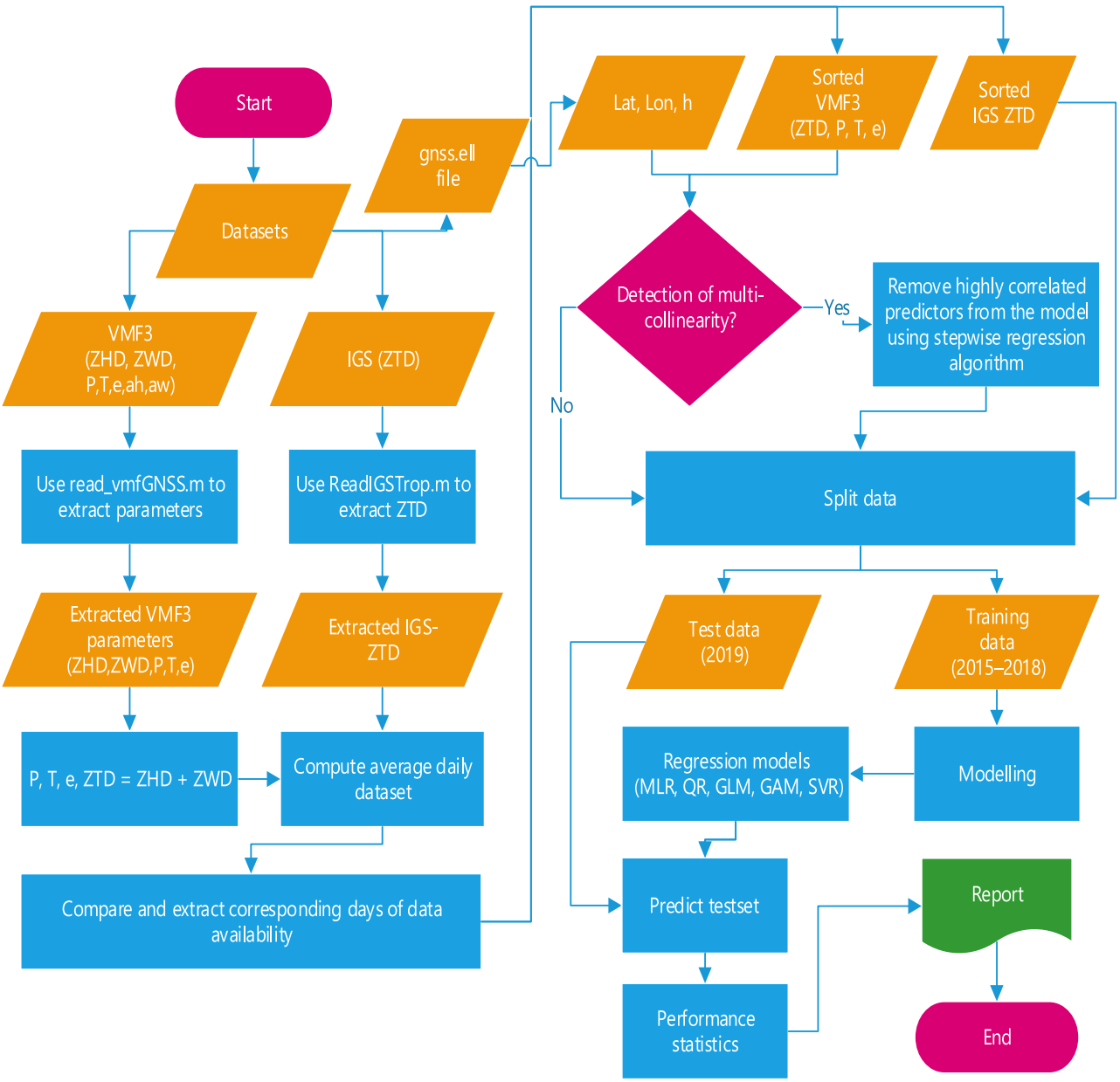


FIGURE 3 General flow of the methodology

fields as well as stations coordinates (φ, λ, h) as prior indicated were initially set for input following Zheng et al. (2015) and Ding et al. (2016). However, in multiple regression models, there are situations where two or more predictor variables will be highly inter-correlated, implying that they measure almost the same information (James et al., 2013). This extreme inter-correlation is known as multicollinearity. Multicollinearity causes the variances of the regression coefficients to be inflated, hence resulting in redundancy among the predictor variables. One of the ways to detect multicollinearity is to examine the correlation between each pair of predictor variables (i.e., pairwise correlation coefficient (R)). If two of the variables are highly correlated, then this may be a possible source of multicollinearity. Following Dormann et al. (2013), an absolute value of $R > 0.7$ ($|R| > 0.7$) among two or more predictors is an indication of collinearity or multicollinearity. Figure 4 provides a correlation plot showing collinearity between all the covariates and possible source of multicollinearity. For surety of multicollinearity, the variance inflation factor (VIF), in an R package, **mctest** (Imdadullah et al., 2016), was utilized to test for multicollinearity between the predictors used in this study. The VIF indicates how much the variance of a regression coefficient is inflated due to collinearity. VIF is expressed as (Imdadullah et al., 2016) follows:

$$\mathrm{VIF}_i = \frac{1}{1 - R_i^2}, \tag{19}$$

where R_i^2 is the coefficient of determination between the predictor variables. Following Imdadullah et al. (2016) value of VIF > 3, 5, 10 indicates the existence of multicollinearity among predictors. The result of the

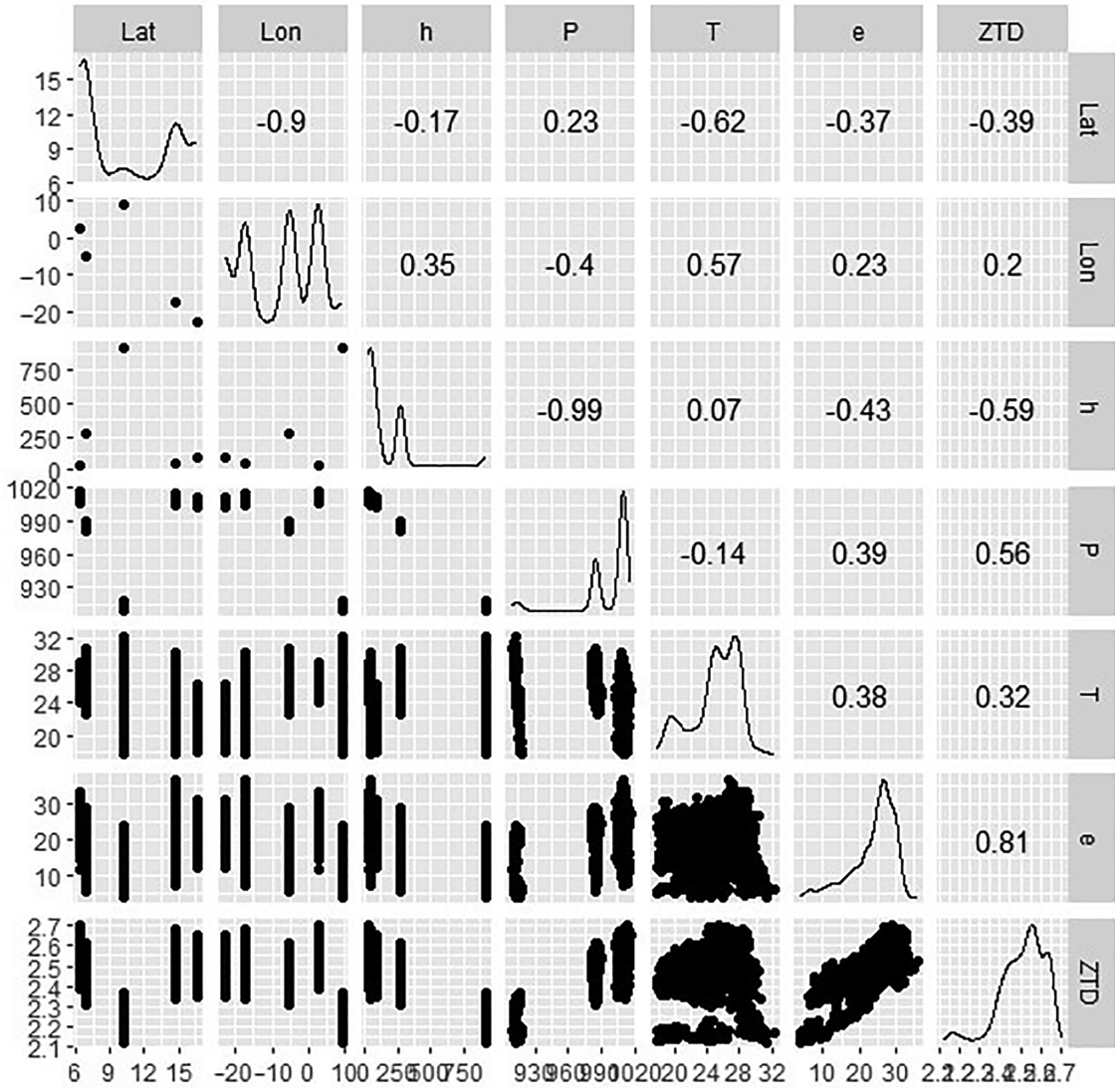


FIGURE 4 Correlation plot showing collinearity between all the predictor variables. Lat and Lon may be collinear as $|-0.898270| \geq 0.7$, h and P may be collinear as $|-0.994753| \geq 0.7$, e and ZTD may be collinear as $|0.814648| \geq 0.7$

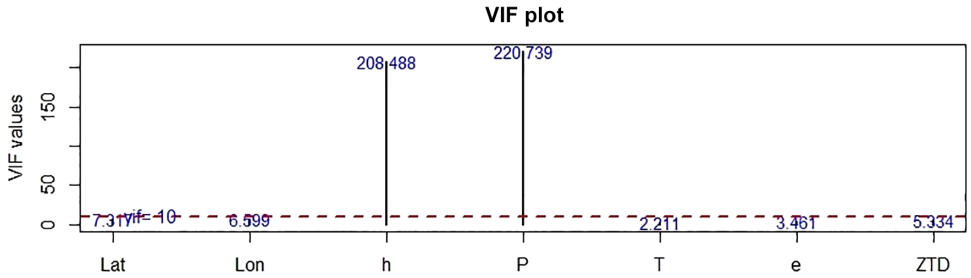


FIGURE 5 The VIF plot. VIF values of each predictor greater than 10 indicate the existence of multicollinearity among predictors

multicollinearity test is shown in Figure 5, indicating h and P as the two possible sources of multicollinearity (VIF > 10). As seen in Figure 4, pressure and altitude or height have a strong negative correlation as the two are inversely or exponentially related (Equation 20). This implies that pressure decreases with increasing altitude and vice versa, hence the possible reason the two are the potential cause of multicollinearity. Mathematically, pressure and altitude are related as follows (Liu et al., 2019; Simeonov et al., 2016; Sissenwine et al., 1962; Xu et al., 2020):

$$P = P_0 \exp\left(-\frac{g_m M_d}{R \cdot T_0 (1 + 0.6077 \cdot q)}(h - h_0)\right) \text{ or } P = P_0 \left(\frac{T}{T - L(h - h_0)}\right)^{\left(\frac{g_m M_d}{R \times L}\right)}, \quad (20)$$

where P_0 (1013.25 hpa), T_0 (288.16 K), and h_0 (0) are pressure, temperature, and altitude at sea level, g_m is the mean gravitational acceleration (9.80655 m/s^2), M_d is the molar mass of dry air (28.9644×10^{-3} kg/mol), R is the universal gas constant (8.3143 J K^{-1} mol^{-1}), L is the temperature lapse rate (6.5 K/km) for dry air.

To overcome multicollinearity, the following approaches are suggested (de Souza et al., 2018; Dunn & Smyth, 2018; Gwelo, 2019; Imdadullah et al., 2016):

- Remove highly correlated predictors from the model. If two or more predictors have a high VIF, one has to be removed from the model. Stepwise regression can be employed to aid in the removal.
- Use partial least squares regression (PLS) or principal components analysis, a statistical technique to change a set of possibly correlated predictors into a smaller set of linearly uncorrelated variables or components. Regression methods that cut the number of predictors to a smaller set of uncorrelated components.

This study considered the first approach, in which h is considered non-significant and removed after stepwise regression algorithm using the **stepAIC** function of the **MASS** package in R.

3.4.2 | Multiple linear regression model (MLRM)

MLR attempts to model the relationship between a response variable and a set of predictor variables or covariates by fitting a linear equation to the observed data using ordinary least squares (OLS). The fitted equation can then be used to predict the response variable as a function of several predictor variables. The MLR can be represented as (Gwelo, 2019; Sayegh et al., 2014) follows:

$$Y = \beta_0 + \sum_{i=1}^{N} \beta_i x_i + \varepsilon = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_N X_N + \varepsilon, \quad (21)$$

where Y is the response variable, X_1 through X_N are the predictor variables, β_0 is the constant term, β_1 through β_N are the regression coefficients, and ε is the residual or the fitting error.

3.4.3 | Quadratic regression model (QRM)

QR adds a quadratic term to the linear regression model. This is to check whether the quadratic model fits the data better than the linear model. The quadratic model takes the form (Yazar et al., 2017):

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_2^2 + \cdots, \quad (22)$$

MLRM and QRM are fitted using the **lm ()** function. The QRM adds the **poly ()** function to fit the quadratic term.

3.4.4 | Generalized linear model (GLM)

In GLM (Nelder & Wedderburn, 1972), the response variable is modelled by a linear function of the independent variates plus a link function, which allows for the response variable to relate to the linear predictor of the regressors. Thus, GLM consists basically of three components:

Random component, which specifies the probability distribution of the univariate response variable (Y) with the mean μ.
Systematic component, also known as the *linear predictor* (η), for example, $\eta = \beta_0 + \beta x_1 + \beta x_2$, which specifies the linear combination of the predictor variables in the model.
Link function, $g(\mu) = \eta$, specifying the link between the random and the systematic components.

Generally, the GLM is formulated as (Nelder & Wedderburn, 1972) follows:

$$E(Y) = \mu = g^{-1}(X\beta), \tag{23}$$

where $E(\boldsymbol{Y})$ is the expected value of $\boldsymbol{Y}$; $\boldsymbol{X\beta}$ is the linear predictor; $\boldsymbol{g}$ is the link function such that $E(Y \mid X) = \mu = g^{-1}(\eta)$.

GLM is fitted using the **glm ()** function with the Gaussian family and 'identity' link function.

3.4.5 | Generalized additive model (GAM)

GAM (Hastie and Tibshirani, 1986) is an extension of GLM in which the linear predictor, η relates to the covariates through smoothing functions, s_i applied to each covariate x_i. The general form of GAM is given as (Hastie and Tibshirani, 1986; Sayegh et al., 2014) follows:

$$\begin{aligned} Y_i &= \beta_0 + \sum_{i=1}^{N} s_i(x_i) + \varepsilon_i \\ &= \beta_0 + s_1(x_1) + s_2(x_2) + \cdots + s_N(x_N) + \varepsilon_i, \end{aligned} \tag{24}$$

GAM is fitted using the **gam ()** function from the R package **mgcv** with the **s ()** function used to fit the smoothing spline.

3.4.6 | Support vector regression model (SVRM)

SVR is the adaptation of the support vector machine (SVM) algorithm (Gunn and others, 1998; Vapnik et al., 1997) for regression technique, preserving the key features that characterize the SVM algorithm (find maximum margin) except for working with continuous data instead of classification for SVM. SVR recognizes the existence of non-linearity in the dataset and provides a proficient prediction model.

The general equation of the SVRM can be summarized as (Adnan et al., 2017; García et al., 2019) follows:

$$Y = f(x) = W^T \Phi(x) + b, \tag{25}$$

where W is the weight vector, $\Phi(x)$ is the feature function, which maps the predictor x to a high-dimensional feature space, b is constant and defines the bias term. A detailed explanation of SVR can be inferred from Parveen et al. (2016, 2017, 2020)).

SVR is fitted using the **svm ()** function from the R package '**e1071**'.

3.5 | Performance evaluation

In order to evaluate the performances of the developed regression models on the test data, it is necessary to measure the extent to which the predicted response values on the independent test data are close to the corresponding true response (IGS) values. Hence the predictive performances of the developed models, MLR, QR, GLM, GAM, and SVR on the test data are compared with the IGS final ZTD through the computations of the following selected most widely used statistical indicators: MB, RMSE, mean absolute percentage error (MAPE), Nash–Sutcliffe coefficient of efficiency (NSE), coefficient of determination (R^2), and the fraction of predictions within a factor of two (FAC2) (Ali & Abustan, 2014; Nash & Sutcliffe, 1970; Sayegh et al., 2014). Equations for the computation of these indicators are represented as follows:

$$\mathrm{MB} = \frac{1}{N}\sum_{i=1}^{N}(M_i - O_i), \tag{26}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(M_i - O_i)^2}, \tag{27}$$

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{M_i - O_i}{O_i}\right| \times 100, \tag{28}$$

$$\mathrm{NSE} = 1 - \left[\frac{\sum_{i=1}^{N}(O_i - M_i)^2}{\sum_{i=1}^{N}\left(O_i - \overline{O}\right)^2}\right], \tag{29}$$

$$R^2 = \frac{\left[\sum_{i=1}^{N}\left(M_i - \overline{M}\right) \times \left(O_i - \overline{O}\right)\right]^2}{\left[\sum_{i=1}^{N}\left(M_i - \overline{M}\right)^2\right]\left[\sum_{i=1}^{N}\left(O_i - \overline{O}\right)^2\right]}, \tag{30}$$

$$\mathrm{FAC2} = 0.5 \leq \frac{M_i}{O_i} \leq 2.0, \quad (31)$$

where M_i and O_i represent the ith modelled and observed (IGS) ZTDs, respectively, and $\overline{M}$ and $\overline{O}$ their respective means. N is the total number of observations.

To further investigate the performances of the regression models, a multiple comparison test (MCT) with the 'scheffe' method (McHugh, 2011; Lee & Lee, 2018) was conducted to test whether the performances of the regression models in relation to their RMSE values are significantly different from one another at 5% significant level using MATLAB ***multcompare function***. A similar test was also performed to test whether the differences in the mean values of the five regression models and the observed (IGS) ZTDs are significant. The null and alternate hypothesis (H_0 and H_a) tests are stated as follows:

H0. *The difference between the means is equal to* 0.

Ha. *The difference between the means is different from* 0.

4 | RESULTS AND DISCUSSION

The objective of this paper is to fit a regression model that will estimate daily IGS final ZTDs over four IGS stations in West Africa using P, T, e, ZTD, φ, and λ from the VMF3-ZTD products as independent variables and IGS final ZTD as the response variable. The results of the various statistical indicators, the MB, RMSE, MAPE, R^2, NSE, and FAC2 for each regression model at each and all the IGS stations are summarized in Tables 2 and 3, respectively. The MB provides an indication of whether the mean model predictions are under (value when negative) or over (value when positive) estimated, while the RMSE provides an overall measure of how close the model's predicted values are to the observed (IGS) values. Smaller values of the MB and RMSE indicate better model predictions. The results in Table 2 show smaller values of MB and RMSE at all the IGS stations for all the models, indicating smaller prediction errors, which demonstrates good agreement of the modelled and IGS-ZTD values, and hence better performance.

Additionally, $\boldsymbol{R^2}$ measures the goodness of fit, which indicates how well each model prediction explains the

TABLE 2 Statistical evaluation results of the prediction models w.r.t. IGS-ZTD values at each IGS station

IGS station	Model	MB (m)	RMSE (m)	MAPE (%)	R^2 (−)	NSE (−)	FAC2 (−)
BJCO	MLR	0.00053	0.00961	0.308	0.941	0.940	1.00
	QR	0.00183	0.00963	0.302	0.942	0.940	1.00
	GLM	0.00053	0.00961	0.308	0.941	0.940	1.00
	GAM	0.00136	0.00963	0.303	0.941	0.940	1.00
	SVR	0.00140	0.00952	0.294	0.943	0.941	1.00
CPVG	MLR	−0.00253	0.00763	0.233	0.989	0.987	1.00
	QR	0.00151	0.00686	0.217	0.990	0.990	1.00
	GLM	−0.00253	0.00763	0.233	0.989	0.987	1.00
	GAM	0.00155	0.00706	0.223	0.990	0.989	1.00
	SVR	0.00222	0.00728	0.238	0.990	0.989	1.00
DAKR	MLR	0.00264	0.00901	0.290	0.992	0.990	1.00
	QR	−0.00100	0.00858	0.259	0.992	0.991	1.00
	GLM	0.00264	0.00901	0.290	0.992	0.990	1.00
	GAM	−0.00107	0.00869	0.263	0.992	0.990	1.00
	SVR	−0.00061	0.00875	0.270	0.991	0.990	1.00
YKRO	MLR	0.00688	0.01169	0.378	0.913	0.869	1.00
	QR	0.00620	0.01145	0.370	0.909	0.873	1.00
	GLM	0.00688	0.01169	0.378	0.913	0.869	1.00
	GAM	0.00720	0.01188	0.386	0.912	0.865	1.00
	SVR	0.00528	0.01090	0.349	0.911	0.884	1.00

TABLE 3 Overall average of statistical evaluation results of the regression models over the year 2019

Model	MB (m)	RMSE (m)	MAPE (%)	R^2 (–)	NSE (–)	FAC2 (–)
MLR	0.00313	0.00948	0.302	0.959	0.947	1.00
QR	0.00263	0.00913	0.287	0.958	0.949	1.00
GLM	0.00313	0.00949	0.303	0.959	0.947	1.00
GAM	0.00283	0.00933	0.294	0.959	0.946	1.00
SVR	0.00238	0.00911	0.288	0.959	0.951	1.00

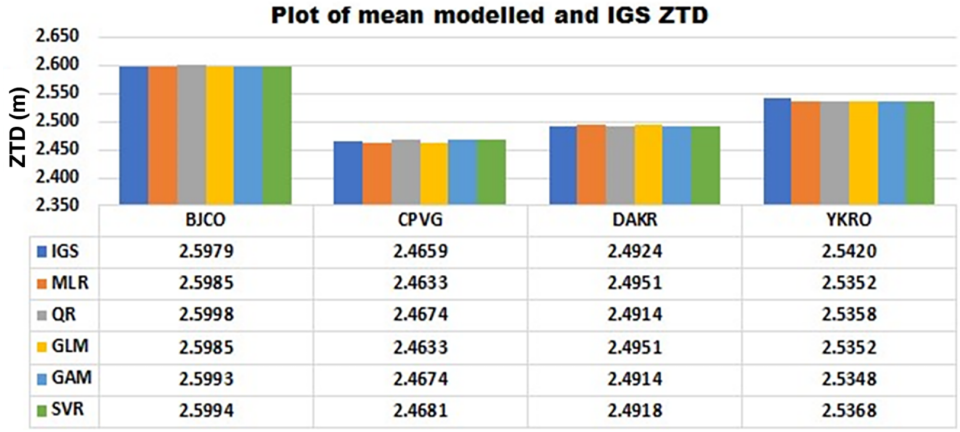


FIGURE 6 Comparison of mean predicted and IGS ZTDs over the four selected IGS stations in West Africa

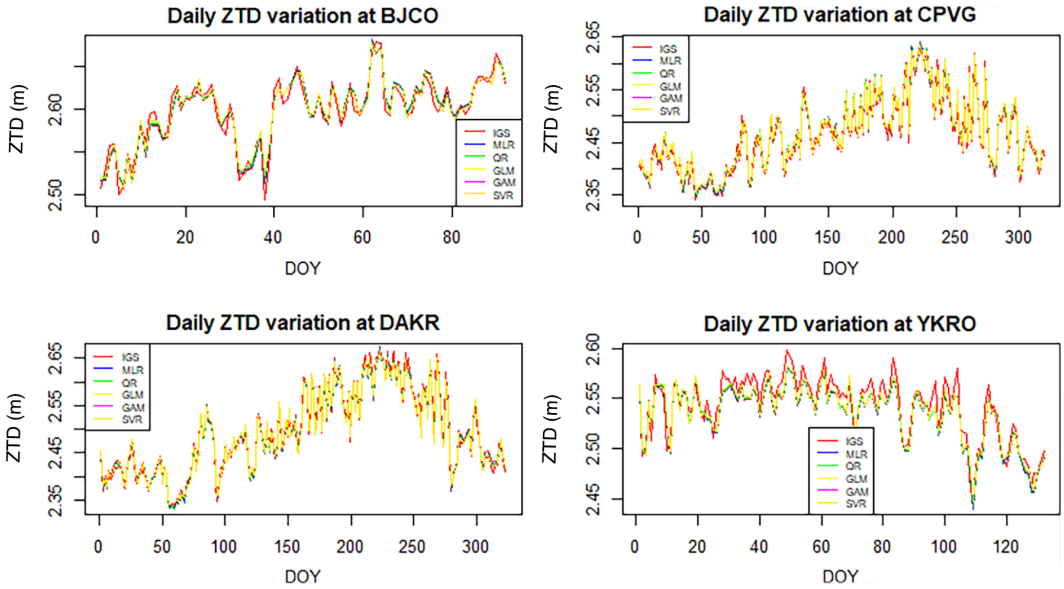


FIGURE 7 Daily variation of observed (IGS) and predicted ZTD values at each IGS station for the year 2019

variance in the response variable (IGS). R^2 ranges from 0 to 1, with higher values indicating better performance. The results in Table 2 clearly show that all the models at all the stations have R^2 values closer to 1, representing a high percentage (90.9%–99.2%) of the variation in the observed (IGS) data been explained by the prediction models, and thus implies the better and improved estimation of the dependent variable (IGS). IGS station DAKR had the highest R^2 ranging between 0.991 and 0.992 (99.1%–99.2% explained), while YKRO had the least, with R^2 ranging between 0.909 and 0.913 (90.9%–91.3% explained).

MAPE likewise indicates prediction accuracy and measures how much error in predicting compared with the observed (IGS) values expressed in percentage terms. The smaller the MAPE, the better the performance of the model. It is obvious from Table 2 that the MAPE for all the models at all the IGS stations is less than 0.5% (0.217%–0.386%), signifying a high accuracy of the proposed regression models' predictions. This means that the regression models can predict accurately to approximately 99.5%.

Moreover, **NSE** measures the relative magnitude of the residual variance compared with the variance of the observed data (Nash & Sutcliffe, 1970). According to Chiew and McMahon (1993)), NSE provides a more direct measure of the agreement between the observed (IGS) and predicted values than R^2. Legates and Mccabe (2013)) finally stressed that NSE is a necessary quality for any metric of model evaluation. A value of NSE closer to 1 signifies that the model can reproduce satisfactorily the observed (IGS) data, with NSE = 1.0 indicating a perfect match of the model predictions to the observed (IGS) values. It can further be seen from Table 2 that the regression models at all the stations produce NSE values closer to 1. IGS station DAKR recorded the highest NSE values ranging between 0.991 and 0.992, while YKRO recorded the least values ranging between 0.865 and 0.884. Adopting Chiew and McMahon (1993)) and Moriasi et al. (2015)) model classification criteria, a model is considered perfect for prediction if NSE ≥0.93, very good if NSE ≥0.80, and satisfactory if NSE ≥0.60. On the basis of these model performance criteria (rule of thumb), it can be stated that the regression models' predictions agree better with that of the IGS.

Furthermore, **FAC2** measures the fraction or percentage of the models' predictions that meet the condition in Equation (31). FAC2 is characterized as a robust performance measure since it is not overly affected by outliers (Chang & Hanna, 2004). FAC2 value closer to 1.0 implies a closer match between the modelled and observed (IGS) values and, therefore, suggest better model performance. A perfect model has a FAC2 = 1. From the results in Table 2, the models again showed excellent performance with 100 per cent of their predictions at all the stations within the factor of two (FAC2) of the observed (IGS) data.

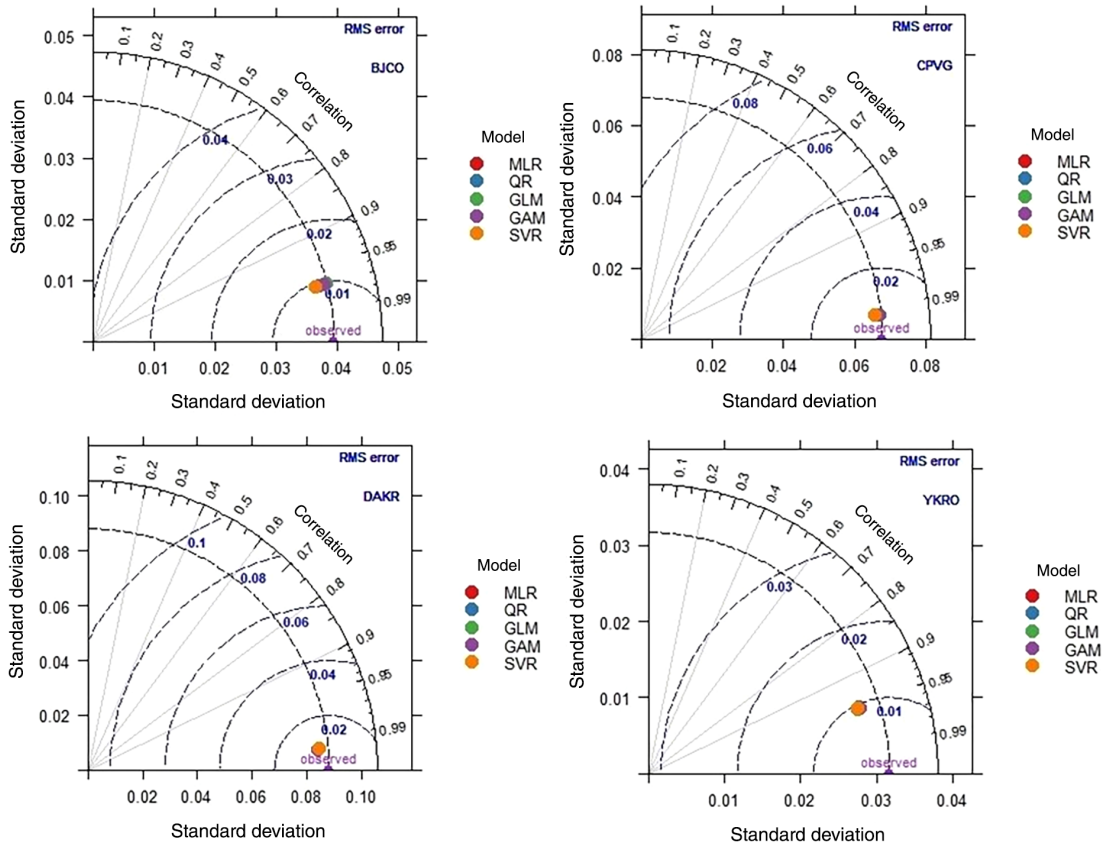


FIGURE 8 Taylor's diagram showing a statistical comparison between modelled and IGS values over the IGS stations. The coloured round markers represent the five regression models, the blue contour or arcs represent the centred RMSE, the black arc with the marked graduated values (0.1–0.99) represent the Pearson correlation coefficient (R), the point labelled 'observed' on the x-axis represents the observed (IGS) standard deviation

Figures 6 and 7 further show the daily and mean variations of the observed (IGS) and the predicted ZTD values at all the stations, which allow quantitative comparison between the predicted and the observed (IGS) ZTDs at each IGS station. From Figures 6 and 7, it can be seen that all the regression models produce a similar outcome, which is almost the same as or matches well with the observed (IGS) ZTD values; that is, the overall trend is completely consistent, showing a good correlation.

To further evaluate graphically the performance of the regression models, Taylor's diagram was employed and plotted for the five regression models at the four (4) IGS stations, as shown in Figure 8. Taylor diagram (Taylor, 2001) provides a visual comparison between modelled and observed values by graphically displaying a statistical summary of how well models' predictions match the observed (IGS) values in terms of their correlation coefficient (R), centred RMSE, and standard deviation (SD). Thus, for each of the five models, three performance statistics (R, centred RMSE, and SD) are plotted. The five models, each represented on the diagram by a different coloured symbol as shown in Figure 8, are compared and the distance between each model and the point labelled '***observed***' is a measure of how realistically each model reproduces or matches the observed values. From the various plots in Figure 8, the position of each coloured symbol (red, blue, green, etc.) quantifies how closely the models' predictions at each IGS station match the observed (IGS) values. The black dashed arc in the diagram represents the observed (IGS) SD at the point marked '***observed***' on the x-axis. Models that match well with the observed values will lie nearest or exactly on the black arc or the '***observed***' marked point having high correlation and low RMSEs. The blue contours indicate the centred RMSE between the modelled and observed (IGS) values, which is proportional to the distance to the marked point '***observed***' on the x-axis. It can be deduced from Figure 8 that the modelled values at all the stations agree best with the IGS-ZTD data, with IGS stations CPVG and DAKR having the best match with smaller RMSE and high R values. Station CPVG, however, has almost the same SD as the observed, whereas station DAKR has a slightly lower SD than the observed. IGS stations BJCO and YKRO, on the other hand, have about the same RMSE and R. BJCO, however, showed slightly lower SD than YKRO compared with the observed SD. Figure 8 again shows the comparable performance of the regression models under study.

Yet again, the results of the MCT are shown in Table 4 and Figures 9 and 10. The MCT is used to

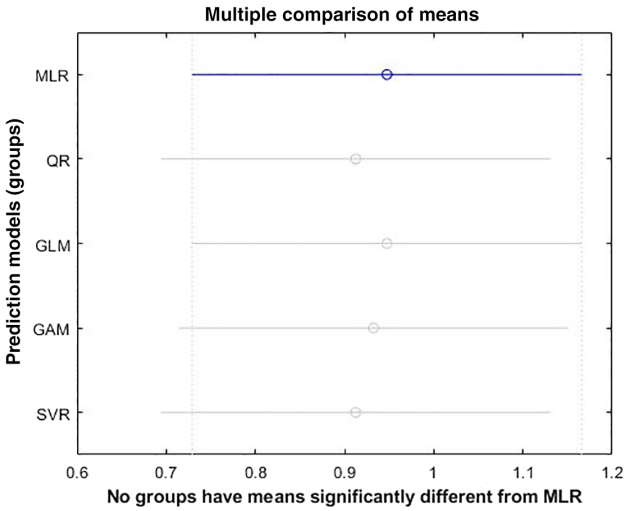


FIGURE 9 Multiple comparison of groups RMSE means (unit: cm)

TABLE 4 Result of the multiple comparison test (1 = MLR, 2 = QR, 3 = GLM, 4 = GAM, 5 = SVR)

Groups pair		Mean difference	95% confidence interval		
Level 1 (L1)	Level 2 (L2)	(L1–L2)	Lower limit	Upper limit	*p* value
1	2	0.035	−0.402	0.472	0.999
1	3	0.000	−0.437	0.437	1.000
1	4	0.015	−0.422	0.452	1.000
1	5	0.035	−0.402	0.472	0.999
2	3	−0.035	−0.472	0.402	0.999
2	4	−0.020	−0.457	0.417	1.000
2	5	0.000	−0.437	0.437	1.000
3	4	0.015	−0.422	0.452	1.000
3	5	0.035	−0.402	0.472	0.999
4	5	0.020	−0.417	0.457	1.000

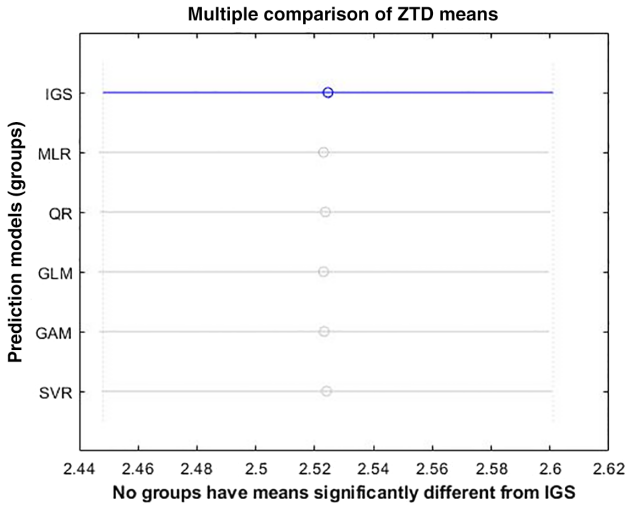


FIGURE 10 Multiple comparison of groups ZTD means (unit: m)

determine which pairs of means are significantly different, and which pairs are not. A small p value of about 0 (p value = 0.00) suggests that the mean response is significantly different. Figures 9 and 10 provide an interactive graph with each group mean represented by a symbol (small circle) and a comparison interval or bar represented by a line extending from the symbol (horizontal line). Two groups (e.g., MRL and QR in Figure 9) means are significantly different if their intervals are disjoint and significantly indifferent if their intervals overlap or intersect. For a selected group (e.g., MLR (Figure 9) or IGS (Figure 10)), the comparison bar is highlighted blue, and all other groups that are significantly different are highlighted red. The bars for the groups that are not significantly different are highlighted grey (e.g., QR, GAM), as shown in Figures 9 and 10. The results from the MCT as shown in Figure 9 and Table 4 indicate that the mean RMSE of group 1 (MLR) in blue is not significantly different from groups 2, 3, 4, and 5 (QR, GLM, GAM, and SVR) shown in grey and with p values >0 at 95% confidence level. Likewise, Figure 10 also indicates that the mean ZTD of group 1 (IGS) in blue is not significantly different from groups 2, 3, 4, 5, and 6 (regression models) at a 5% significance level.

Last of all, it is obvious from Tables 2 and 3 and Figures 6–8 that the five developed models, MLRM, QRM, GLM, GAM, and SVRM, yielded very compatible results and best-fit the IGS final ZTD data with a greater degree of accuracy. Nonetheless, the overall performance statistics in Table 3 further reveals that the SVR model slightly performed better than the QR model with an average bias of SVRM: 2.38 mm, QRM: 2.63 mm and RMSE of SVRM: 9.11 mm, QRM: 9.13 mm. This is followed by GAM with an MB of 2.83 mm and RMSE of 9.33 mm, then MLR and GLM, which produced the same average bias and RMSE of 3.13 and 9.48 mm, respectively. The results suggest that the MLR, QR, GLM, GAM, and SVR models can provide good prediction performance and could be successfully applied in GNSS and its related fields to establish a ZTD prediction model that could provide a reliable and accurate daily prediction of ZTD. These findings, especially the performance of the SVR model over the MLR, are in agreement with Parveen et al. (2020), in which the authors compared the performance of SVR and MLR models to predict cadmium removal from wastewater based on R^2, average absolute relative error (AARE), RMSE, mean relative error (MRE), SD, and so forth. The SVR model was preferred over the MLR model since it had the highest R^2 and lowest AARE values of 0.9997% and 0.67%, respectively, compared with R^2 of 0.2161 and AARE of 29.27% for the MLR model. Correspondingly, Abdel-Sattar et al. (2021) evaluated the applicability of the SVR, ANN, and MLR models to predict the mass of ber fruits. The results showed that the SVR model provides better predictive values for the mass of ber fruits than those obtained with the widely used MLR model in terms of their R^2 (SVR: 0.9947, MLR: 0.4614) and RMSE (SVR: 1.8814 g, MLR: 11.3742 g) values. Similar findings have also been obtained by Ahmadi and Rodehutscord (2017) and Parveen et al. (2016, 2017). Furthermore, the performance of the QR model over the MLR model is also consistent with the study carried out by Acheampong and Obeng (2019). In the study, the authors estimated the amount of PWV over four selected IGS ground stations in West Africa using QR, MLR, and exponential regression (ER) models. The estimated PWV were compared with their respective total column water vapour (TCWV) derived from the ECMWF ERA-Interim re-analysis data in terms of R^2 and RMSE statistical measures. The results showed that the QR model was the best-fitting regression model as it produced the highest R^2 and lowest RMSE values compared with the MLR and ER models. Moreover, the parallel performance of the MLR and GAM models is also in accordance with Sayegh et al. (2014). In the study, the performances of four statistical models for predicting PM_{10} concentrations were compared. The findings in terms of statistical metrics indicated that the MLRM and the GAM yielded very comparable performance.

As indicated above, studies on the application of regression techniques for modelling and predicting ZTD are very rare, unlike other techniques such as ANN and ANFIS that have been applied to model and predict ZTD. Although previous studies on ANN have attempted to model and predict ZTDs, direct comparisons with these studies cannot be made due to differences in the number of GNSS ground stations used, the amount of data used,

the number and types of predictors or input parameters used, and the time interval or duration used. However, if these conditions are ignored and the results of the regression models developed are compared with the results obtained with the ANN as discussed in the introductory section, then it can be stated that the regression models developed outclass those of the ANN-based ZTD models (Pikridas et al., 2010; Ding et al., 2016; Yang et al., 2017; Li et al., 2020). Again, these results, particularly the performance of the SVR model over the ANN-based ZTD models, are consistent with the results of several studies conducted in various fields where the SVR model has been compared with the ANN technique. For instance, He et al. (2014), Adnan et al. (2017)), and Bafitlhile and Li (2019)) compared the performances of ANN and SVR for streamflow or riverflow forecasting and recommended SVR instead of ANN. Likewise, Çimen (2008) compared the efficacy of ANN and SVR for the estimation of daily suspended sediments and recommended the SVR model as the best alternative to the ANN method. In addition, Parveen et al. (2017) evaluated the predictive performance of ANN and SVR for predicting the sorption capacity of Cr (VI). The result also showed the superiority of the SVR model over the ANN model in predicting the sorption capacity of Cr (VI). Last but not least, Valizadeh and Sohrabi (2018) also evaluated the application of ANNs and SVR for simultaneous spectrophotometric determination of commercial eye drop contents as well as recommended the SVR model over the ANN technique.

The efficacy of the SVR model over the ANN, the linear regression (LR) models, and possibly the remaining models developed in this study according to Parveen et al. (2017, 2020) can be attributed to the fact that the SVR model is based on the principle of structural risk minimization (SRM), whereas the ANN and LR models are based on the principle of empirical risk minimization (ERM). While in the SRM principle, the generalization accuracy is optimized over the empirical error and the smoothness of the regression function or the capacity of SVM, the ERM principle only minimizes the empirical error and does not consider the capacity of the learning systems, which eventually results in poor generalization performance. Ahmadi and Rodehutscord (2017) and Abdel-Sattar et al. (2021) further assert that the SVR model has a wide-ranging approximation capability that can practically approximate all forms of non-linear functions including quadratic functions, and is good at fitting functions and recognizing patterns in diverse kinds of data; whereas the LR models are based on the assumption of linearity, which is useful for only linear approximations, requiring linear function specification to be regressed, hence the flexibility of regression equation may be extremely inadequate. On the other hand, the performance of GAM compared with GLM is also a result of the fact that GAM, as already noted, is an extension of GLM, in which the linear predictor η is not limited to linearity in the covariates X_1, X_2, ..., X_n, but is the sum of smoothing functions s_i applied to each covariate X_1, X_2, ..., X_n. That is, in GLM, the linear predictor η is a weighted sum of the n covariates, $\eta = \sum_{i=1}^{n} \beta_i x_i$ whereas, in GAM, the linear predictor η is replaced by an additive predictor or smooth function of the form $\eta = \sum_{i=1}^{n} s_i(x_i)$ (Hastie & Tibshirani, 1986). With its additive predictive functions, GAM is designed to provide a more flexible technique for identifying and capturing non-linear effects or patterns of covariates on the dependent variable Hastie and Tibshirani (1986) than GLM, hence producing a more accurate fit than GLM in the study. Again, GLM and the MLR model having the same results, GLM is an extended part of the MLR (General Linear Model). The difference between GLM and the MLR model is in the link function. In fitting the GLM, the Gaussian family with the 'identity' link function was utilized. Gaussian functions are regularly used to characterize the probability density function of normally distributed random variables. The MLR model is also based on the assumption of normality, which for any given predictor variable X_1, X_2, ..., X_n, the response variable Y, is normally distributed. The Gaussian function with 'identity' link function in GLM effectively converts GLM to MLR. That is, GLM without Gaussian: $E(Y) = g^{-1}(X\beta)$, for GLM Gaussian with identity link, $E(Y) = X\beta$, which is equivalent to MLR: $Y\tilde{N}(X\beta, \sigma^2)$, thus, producing the same results as in Tables 2 and 3. Finally, yet importantly, QR is a second-order polynomial regression, which is an extension of linear regressions. It extends linear regression by adding a quadratic term to include non-linear relationships between response and predictors. In fact, QR can be seen as an example of an additive model, with a simple form of additive function of explanatory variables, that is, a second-order polynomial function, just like GAM with an additive spline smoothing function. The QR model with better performance than the GAM, GLM, and MRL models in this study could be that the QR model is better adapted to the effects of non-linearity or patterns of covariates than the GAM, GLM, and MRL models in this study, and therefore provides a better fit than the GAM, GLM, and MRL models. However, the comparable performance of the regression models in this study could be attributed to the adequacy and reliability of the datasets used as they are statistically well scattered or dispersed in both the input and output space.

5 | CONCLUSIONS

Precise modelling of tropospheric delay is of wide practical importance for precise geodetic, meteorological, and

climatological applications. Large data gaps resulting from inadequate and inconsistency of continuous data streaming from the West African IGS stations hinder daily continuous ZTD estimation by the IGS products centre, and in so doing, degrade the quality of the ZTD and PWV retrieval. To address this shortcoming and provide an effective prediction model for predicting ZTD in West African using regression methods, this study proposed five regression models, multiple linear regression (MLR), quadratic regression (QR), generalized linear model (GLM), generalized additive model (GAM), and support vector regression (SVR) for predicting daily IGS final ZTD in West Africa. The site-wise VMF3-ZTD products for IGS stations including P, T, e, ZTD, φ, and λ on daily basis over a period of 5 years from 2015 to 2019 were used as the covariates/predictors and the IGS final ZTD product over the same period represented as the response variable in fitting the regression models. Four years of data from 2015 to 2018 from four IGS station (BJCO, CPVG, DAKR, and YKRO) in West Africa were used in training the models and data of 2019 from the same IGS stations were used in validating the models. In order to compare the predictive performances of the developed models, six different statistical performance metrics, MB, RMSE, MAPE, R^2, NSE, and FAC2, non-parametric post-hoc or MCT, and graphical representations such as Taylor's diagrams and time variation plots were employed to determine the degree of agreement between the models' predictions and IGS final ZTD data. The results from the various analyses revealed that the performance of the developed models did not vary substantially (i.e., have nearly the same accuracies) and agreed best to the IGS final ZTD product; implying that they can successfully be applied to establish reliable and accurate ZTD predictions in GNSS and its related fields. The results also indicate that the five regression models (SVR, QR, GAM, GLM, and MLR) can as well provide a suitable alternative for predicting ZTD in West Africa in the event of inaccessibility or missing of ZTD data from the IGS stations being studied, as this will help improve the continuous availability of daily time series ZTD data in the region for any GNSS-related applications. However, a detailed comparison of the overall performance indicated that the SVR model performed slightly better than the regular QR model, followed by GAM, GLM, and MLR in the order given. The results also showed that there was no noticeable difference in performance between the MLR and GLM models as the two models produce the same average statistics. Moreover, the p values from the non-parametric MCT in Table 4 clearly suggest that the alternate hypothesis (H_1) of the study should be rejected, while we accept the null hypothesis (H_0) at 95% confidence level. The overall average performance statistics are provided as follows: MB (SVRM: 2.38 mm, QRM: 2.63 mm, GAM: 2.83 mm, GLM: 3.13 mm, MLR: 3.13 mm), RMSE (SVRM: 9.11 mm, QRM: 9.13 mm, GAM: 9.33 mm, GLM: 9.48 mm, MLR: 9.48 mm), MAPE (SVRM: 0.29% mm, QRM: 0.29%, GAM: 0.29%, GLM: 0.30%, MLR: 0.30%), R^2 (SVRM: 0.959, QRM: 0.958, GAM: 0.959, GLM: 0.959, MLR: 0.959), NSE (SVRM: 0.951, QRM: 0.949, GAM: 0.946, GLM: 0.947, MLR: 0.947), FAC2 (SVRM = QRM = GAM = GLM = MLR = 1.0). In general, the analysis presented in this study provides that of the five regression models, the choice of the SVR model will be optimum.

This study has demonstrated that regression techniques can offer good prediction performance and can be successfully applied in GNSS and related fields to develop a ZTD prediction model that can provide reliable and accurate ZTD predictions. Both linear (MLR, GLM) and non-linear regression (SVR, QR, and GAM) techniques have been investigated in this study to find a reliable model for predicting ZTD. Performance analysis of the models showed that the ZTD prediction performance of the non-linear regression models is more efficient than the linear models, although both model types provided comparable performance. The performance of the non-linear regression models over the linear models indicates that the dependent and independent variables showed a non-linear relationship. Despite the promising results obtained from the developed models, the findings of this study are limited to only the West African IGS stations employed. Further studies could extend these findings using stations' data outside of those used in this study to investigate if better predictions can be obtained. Likewise, the number of GNSS ground stations and probably a larger dataset could also be increased for the same purpose. Moreover, the study detected multicollinearity among the predictor variables based on the variance inflation factor (VIF) and was treated using a stepwise regression algorithm. This problem could also be solved by principal component analysis (PCA). PCA regression is an interesting area of research. For further studies, we hope to use PCA to treat multicollinearity, as this is also necessary to determine whether more accurate predictions can be made with PCA. Another future work is needed to compare with other soft computing methods like ANN, ANFIS, GRNN, LSTM, and so forth to test whether the proposed regression models can provide better results than those obtained by other researchers. Finally, the GLM gave the same results as the MLR model owing to the Gaussian family and identity link function used. Further study is again required to use other families (such as binomial, gamma, poison, etc.) and link functions (e.g., 'logit', 'inverse', 'log') other than the identity link function we used in this study. This will help know if still better results could be obtained by varying the families and link functions.

ACKNOWLEDGEMENTS

The authors wish to thank the Vienna Mapping Functions (VMF) and International GNSS Service (IGS) data centres for their open access tropospheric delay products.

AUTHOR CONTRIBUTIONS

Samuel Osah: Conceptualization (equal); data curation (lead); formal analysis (lead); funding acquisition (supporting); investigation (lead); methodology (lead); resources (supporting); software (lead); supervision (supporting); validation (lead); visualization (lead); writing – original draft (lead); writing – review and editing (lead). **Akwasi Afrifa Acheampong:** Conceptualization (equal); data curation (supporting); formal analysis (supporting); funding acquisition (supporting); investigation (supporting); methodology (supporting); resources (lead); software (supporting); supervision (lead); validation (supporting); visualization (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Collins Fosu:** Conceptualization (equal); data curation (supporting); formal analysis (supporting); funding acquisition (lead); investigation (supporting); methodology (supporting); resources (lead); software (supporting); supervision (lead); validation (supporting); visualization (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Isaac Dadzie:** Conceptualization (equal); data curation (supporting); formal analysis (supporting); funding acquisition (lead); investigation (supporting); methodology (supporting); resources (lead); software (supporting); supervision (lead); validation (supporting); visualization (supporting); writing – original draft (supporting); writing – review and editing (supporting).

ORCID

Samuel Osah https://orcid.org/0000-0002-6905-2082
Akwasi A. Acheampong https://orcid.org/0000-0003-1640-6307
Isaac Dadzie https://orcid.org/0000-0002-2146-6796

REFERENCES

Abdel-Sattar, M., Aboukarima, A.M. & Alnahdi, B.M. (2021) Application of artificial neural network and support vector regression in predicting mass of ber fruits (*Ziziphus mauritiana* Lamk.) based on fruit axial dimensions. *PLoS ONE*, 16, 1, e0245228–15. https://doi.org/10.1371/journal.pone.0245228

Acheampong, A. & Obeng, K. (2019) Application of GNSS derived precipitable water vapour prediction in West Africa. *Journal of Geodetic Science*, 9(1), 41–47. https://doi.org/10.1515/jogs-2019-0005

Adnan, R.M., Yuan, X., Kisi, O. & Yuan, Y. (2017) Streamflow forecasting using artificial neural network and support vector machine models. *American Scientific Research Journal for Engineering, Technology, and Sciences*, 29(1), 286–294. Available at: http://www.asrjetsjournal.org/index.php/american_scientific_journal/article/view/2814 [Accessed 24th September 2021].

Ahmadi, H. & Rodehutscord, M. (2017) Application of artificial neural network and support vector machines in predicting metabolizable energy in compound feeds for pigs. *Frontiers in Nutrition*, 4(27), 1–8. https://doi.org/10.3389/fnut.2017.00027

Ali, M.H. & Abustan, I. (2014) A new novel index for evaluating model performance. *Journal of Natural Resources and Development*, 4(1), 1–9. https://doi.org/10.5027/jnrd.vi0.01

Aon, E.F., Ho, Y.H., Othman, A.R. & Shaddad, R.Q. (2018) Modeling of GPS ionospheric scintillation using nonlinear regression technique. *Lecture Notes on Data Engineering and Communications Technologies*, 5, 180–188. https://doi.org/10.1007/978-3-319-59427-9_20

Askne, J. & Nordius, H. (1987) Estimation of tropospheric delay for microwaves from surface weather data. *Radio Science*, 22(3), 379–386.

Bafitlhile, T.M. & Li, Z. (2019) Applicability of ε-support vector machine and artificial neural network for flood forecasting in humid, semi-humid and semi-arid basins in China. *Water*, 11(1), 85. https://doi.org/10.3390/w11010085

Bevis, M., Businger, S., Herring, T.A., Rocken, C., Anthes, R.A. & Ware, R.H. (1992) GPS meteorology: remote sensing of atmospheric water vapor using the global positioning system. *Journal of Geophysical Research: Atmospheres*, 97(D14), 15787–15801.

Bevis, M., Businger, S., Chiswell, S., Herring, T.A., Anthes, R.A., Rocken, C. et al. (1994) GPS meteorology: mapping zenith wet delays onto precipitable water. *Journal of Applied Meteorology*, 33(3), 379–386.

Black, H.D. (1978) An easily implemented algorithm for the tropospheric range correction. *Journal of Geophysical Research: Solid Earth*, 83(B4), 1825–1828.

Boehm, J., Kouba, J. & Schuh, H. (2009) Forecast Vienna Mapping Functions 1 for real-time analysis of space geodetic observations. *Journal of Geodesy*, 83, 397–401. https://doi.org/10.1007/s00190-008-0216-y

Boehm, J. & VanDam, T. (2009) Modeling deficiencies and modeling based on external data. In: *Second GGOS Unified Analysis Workshop, IERS, Grand Hyatt, San Francisco, CA, December*, pp. 11–12.

Byun, S.H., Bar-sever, Y.E. & Gendt, G. Proceedings of the 18th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2005), Long Beach, CA

Chang, J.C. & Hanna, S.R. (2004) Air quality model performance evaluation. *Meteorology and Atmospheric Physics*, 87(1–3), 167–196. https://doi.org/10.1007/s00703-003-0070-7

Chen, Q., Song, S., Heise, S., Liou, Y.-A., Zhu, W. & Zhao, J. (2011) Assessment of ZTD derived from ECMWF/NCEP data with GPS ZTD over China. *GPS Solutions*, 15(4), 415–425. https://doi.org/10.1007/s10291-010-0200-x

Chiew, F.H.S. & McMahon, T.A. (1993) Assessing the adequacy of catchment streamflow yield estimates. *Soil Research CSIRO*, 31(5), 665–680. https://doi.org/10.1071/sr9930665

Choi, H.I. (2021) Development of flood damage regression models by rainfall identification reflecting landscape features in Gangwon Province, the Republic of Korea. *Land*, 10(2), 123. https://doi.org/10.3390/land10020123

Clements, M.S., Armstrong, B.K. & Moolgavkar, S.H. (2005) Lung cancer rate predictions using generalized additive models.

Biostatistics, 6(4), 576–589. https://doi.org/10.1093/biostatistics/kxi028

Çimen, M. (2008) Estimation of daily suspended sediments using support vector machines. *Hydrological Sciences Journal*, 53(3), 656–666. https://doi.org/10.1623/hysj.53.3.656

Ding, M. & Hu, W. (2019) Erratum and addendum to the paper "A new ZTD model based on permanent ground-based GNSS-ZTD data", Survey Review, 2016, 48(351), 385–391. *Survey Review*, 51 (366), 280–287. https://doi.org/10.1080/00396265.2017.1420585

Ding, M., Hu, W., Jin, X. & Yu, L. (2016) A new ZTD model based on permanent ground-based GNSS-ZTD data. *Survey Review*, 48(351), 385–391. https://doi.org/10.1179/1752270615y.0000000034

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G. et al. (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Dunn, P.K. & Smyth, G.K. (2018) *Generalized linear models with examples in R*. New York: Springer, p. 562. https://doi.org/10.1007/978-1-4419-0118-7

Feng, P., Li, F., Yan, J., Zhang, F. & Barriot, J.-P. (2020) Assessment of the accuracy of the saastamoinen model and VMF1/VMF3 mapping functions with respect to ray-tracing from radiosonde data in the framework of GNSS meteorology. *Remote Sensing*, 12(20), 3337. https://doi.org/10.3390/rs12203337

García, V., Sánchez, J.S., Rodríguez-Picón, L.A., Méndez-González, L.C. & Ochoa-Domínguez, H.d.J. (2019) Using regression models for predicting the product quality in a tubing extrusion process. *Journal of Intelligent Manufacturing*, 30(6), 2535–2544. https://doi.org/10.1007/s10845-018-1418-7

Gunn, S.R. & others. (1998) Support vector machines for classification and regression. *ISIS Technical Report*, 14(1), 5–16.

Gwelo, A.S. (2019) Principal components to overcome multicollinearity problem. *Oradea Journal of Business and Economics*, 4(1), 79–91. doi:10.47535/1991ojbe062

Hadas, T., Kaplon, J., Bosy, J., Sierny, J. & Wilgan, K. (2013) Near-real-time regional troposphere models for the GNSS precise point positioning technique. *Measurement Science and Technology*, 24(5), 55003. https://doi.org/10.1088/0957-0233/24/5/055003

Hadas, T., Teferle, F.N., Kazmierski, K., Hordyniec, P. & Bosy, J. (2017) Optimum stochastic modeling for GNSS tropospheric delay estimation in real-time. *GPS Solutions*, 21(3), 1069–1081.

Hastie, T. & Tibshirani, R. (1986) Generalized additive models. *Statistical Science*, 1(3), 297–318.

He, Z., Wen, X., Liu, H. & Du, J. (2014) A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *Journal of Hydrology*, 509, 379–386. https://doi.org/10.1016/j.jhydrol.2013.11.054

Hobiger, T., Ichikawa, R., Takasu, T., Koyama, Y. & Kondo, T. (2008) Ray-traced troposphere slant delays for precise point positioning. *Earth, Planets and Space*, 60(5), 1–4. https://doi.org/10.1186/bf03352809

Hofmann-Wellenhof, B., Lichtenegger, H. & Wasle, E. (2008) *GNSS—Global Navigation Satellite Systems GPS, GLONASS, Galileo, and more*. New York: Springer. https://doi.org/10.1007/978-3-211-73017-1

Hopfield, H. S. (1969) Two-quartic tropospheric refractivity profile for correcting satellite data. *Journal of Geophysical Research*, 74(18), 4487–4499. https://doi.org/10.1029/JC074i018p04487

Hu, Y. & Yao, Y. (2019) A new method for vertical stratification of zenith tropospheric delay. *Advances in Space Research*, 63(9), 2857–2866. https://doi.org/10.1016/j.asr.2018.10.035

Imdadullah, M., Aslam, M. & Altaf, S. (2016) Mctest: an R package for detection of collinearity among regressors. *The R Journal*, 8(2), 495–505.

Isioye, O.A., Combrinck, L., Botai, J.O. & Munghemezulu, C. (2015) The potential for observing African weather with GNSS remote sensing. *Advances in Meteorology*, 2015, 1–16.

Isioye, O.A., Combrinck, L. & Botai, J. (2016) Modelling weighted mean temperature in the West African region: implications for GNSS meteorology. *Meteorological Applications*, 23(4), 614–632. https://doi.org/10.1002/met.1584

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013) *An introduction to statistical learning*. New York: Springer.

Jgouta, M., Nsiri, B. & Marrakh, R. (2016) Usage of a correction model to enhance the evaluation of the zenith tropospheric delay. *International Journal of Applied Engineering Research*, 11(6), 4648–4654.

Jiang, C., Xu, T., Wang, S., Nie, W. & Sun, Z. (2020) Evaluation of zenith tropospheric delay derived from ERA5 data over China using GNSS observations. *Remote Sensing*, 12(4). 663. https://doi.org/10.3390/rs12040663

Kleijer, F. (2004) *Troposphere modeling and filtering for precise GPS leveling*. PhD thesis, Delft University of Technology.

Landskron, D. & Böhm, J. (2018) VMF3/GPT3: refined discrete and empirical troposphere mapping functions. *Journal of Geodesy*, 92(4), 349–360. https://doi.org/10.1007/s00190-017-1066-2

Leandro, R.F., Langley, R.B. & Santos, M.C. (2008) UNB3m_pack: a neutral atmosphere delay package for radiometric space techniques. *GPS Solutions*, 12(1), 65–70. https://doi.org/10.1007/s10291-007-0077-5

Lee, S. & Lee, D.K. (2018) What is the proper way to apply the multiple comparison test? *Korean Journal of Anesthesiology*, 71(5), 353–360. https://doi.org/10.4097/kja.d.18.00242

Legates, D.R. & Mccabe, G.J. (2013) A refined index of model performance: a rejoinder. *International Journal of Climatology*, 1056(April 2012), 1053–1056. https://doi.org/10.1002/joc.3487

Li, L., Xu, Y., Yan, L., Wang, S., Liu, G. & Liu, F. (2020) A regional NWP tropospheric delay inversion method based on a general regression neural network model. *Sensors*, 20(11), 3167. https://doi.org/10.3390/s20113167

Libao, Y., Tingting, Y., Jielian, Z., Guicai, L. & Yanfen, L. (2017) Prediction of CO_2 emissions based on multiple linear regression analysis. *Energy Procedia*, 105, 4222–4228. https://doi.org/10.1016/j.egypro.2017.03.906

Liu, Z., Chen, X. & Liu, Q. (2019) Estimating zenith tropospheric delay based on GPT2w model. *IEEE Access*, 7, 139258–139263. https://doi.org/10.1109/ACCESS.2019.2931984

Lou, Y., Huang, J., Zhang, W., Liang, H., Zheng, F. & Liu, J. (2018) A new zenith tropospheric delay grid product for real-time PPP applications over China. *Sensors*, 18(1), 1–14. https://doi.org/10.3390/s18010065

McHugh, M.L. (2011) Multiple comparison analysis testing in ANOVA. *Biochemia Medica*, 21(3), 203–209. https://doi.org/10.11613/BM.2011.029

Mendes, V.B. (1999) *Modeling the neutral-atmospheric propagation delay in radiometric space techniques*. Department of Geodesy and Geomatics Engineering Technical Report No. 199, University of New Brunswick, Fredericton, New Brunswick, Canada, 353 pp.

Meunram, P. & Satirapod, C. (2019) Spatial variation of precipitable water vapor derived from GNSS CORS in Thailand. *Geodesy and Geodynamics*, 10(2), 140–145. https://doi.org/10.1016/j.geog.2019.01.003

Moriasi, D.N., Gitau, M.W., Pai, N. & Daggupati, P. (2015) Hydrologic and water quality models: performance measures and evaluation criteria. *Transactions of the ASABE*, 58(6), 1763–1785. doi:10.13031/trans.58.10715

Nafisi, V., Urquhart, L., Santos, M.C., Nievinski, F.G., Bohm, J., Wijaya, D.D. et al. (2012) Comparison of ray-tracing packages for comparison of ray-tracing packages for troposphere delays. *IEEE Transactions on Geoscience and Remote Sensing*, 50(2), 469–481. https://doi.org/10.1109/TGRS.2011.2160952

Nash, E. & Sutcliffe, V. (1970) River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, 10(3), 282–290.

Nelder, J.A. & Wedderburn, R.W.M. (1972) Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3), 370–384. https://doi.org/10.2307/2344614

Nikolaidou, T., Balidakis, K., Nievinski, F., Santos, M. & Schuh, H. (2018) Impact of different NWM-derived mapping functions on VLBI and GPS analysis. *Earth, Planets and Space*, 70(1), 95. https://doi.org/10.1186/s40623-018-0865-x

Osah, S., Acheampong, A.A., Fosu, C. & Dadzie, I. (2021a) Evaluation of zenith tropospheric delay derived from ray-traced VMF3 product over the west African region using GNSS observations. *Advances in Meteorology*, 2021, 1–14. https://doi.org/10.1155/2021/8836806

Osah, S., Acheampong, A.A., Fosu, C. & Dadzie, I. (2021b) Deep learning model for predicting daily IGS zenith tropospheric delays in West Africa using TensorFlow and Keras. *Advances in Space Research*, 68(3), 1243–1262. https://doi.org/10.1016/j.asr.2021.04.039

Parveen, N., Zaidi, S. & Danish, M. (2016) Support vector regression model for predicting the sorption capacity of lead (II). *Perspectives in Science*, 8, 629–631. https://doi.org/10.1016/j.pisc.2016.06.040

Parveen, N., Zaidi, S. & Danish, M. (2017) Development of SVR-based model and comparative analysis with MLR and ANN models for predicting the sorption capacity of Cr (VI). *Process Safety and Environmental Protection*, 107, 428–437. https://doi.org/10.1016/j.psep.2017.03.007

Parveen, N., Zaidi, S. & Danish, M. (2020) Support vector regression: a novel soft computing technique for predicting the removal of cadmium from wastewater. *Indian Journal of Chemical Technology*, 27(1), 43–50.

Penna, N., Dodson, A. & Chen, W. (2001) Assessment of EGNOS tropospheric correction model. *The Journal of Navigation*, 54(1), 37–55.

Pikridas, C., Katsougiannopoulos, S. & Ifadis, I.M. (2010) Predicting zenith tropospheric delay using the artificial neural network technique. Application to selected EPN stations. *Journal of the National Cancer Institute*, 88(24), 1803–1805.

Polaraju, K. & Prasad, D.D. (2017) Prediction of heart disease using multiple linear regression model. *International Journal of Engineering Development and Research Development*, 5(4), 1419–1425.

Putri, N., Landskron, D. & Böhm, J. (2020) Assessing the performance of Vienna Mapping Functions 3 for GNSS stations in Indonesia using precise point positioning. *Advances in Geosciences*, 50, 77–86. https://doi.org/10.5194/adgeo-50-77-2020

Rajmane, D. & Waikar, M. (2020) Prediction of inflow to the Ujjani dam reservoir using linear. *International Journal of Civil Engineering and Technology (IJCIET)*, 11(12), 39–46. https://doi.org/10.34218/IJCIET.11.12.2020.003

Saastamoinen, J. (1972) Atmospheric correction for the troposphere and stratosphere in radio ranging satellites. In: *The use of artificial satellites for geodesy*. Vol. 15, pp. 247–251.

Sanlioglu, I. & Zeybek, M. (2012) Investigation on GPS heighting accuracy with use of tropospheric models in commercial GPS softwares for different heights. In: *FIG Working Week 2012, At Rome, Italy, Volume: TS09B—Precise Point Positioning, 6080*, pp. 6–10.

Sanz, J., Juan, J.M. & Hernández-Pajares, M. (2013) *GNSS data processing, Vol. I: Fundamentals and algorithms*. ESA Communications, TM-23/l.

Sayegh, A.S., Munir, S. & Habeebullah, T.M. (2014) Comparing the performance of statistical models for predicting PM10 concentrations. *Aerosol and Air Quality Research*, 14(3), 653–665. https://doi.org/10.4209/aaqr.2013.07.0259

Schueler, T., Pósfay, A., Hein, G.W. & Biberger, R. (2001) A global analysis of the mean atmospheric temperature for GPS water vapor estimation. In: *Proceedings of the 14th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 2001), Salt Lake City, UT, September 11–14*, pp. 2476–2489.

Simeonov, T., Sidorov, D., Teferle, F.N., Milev, G. & Guerova, G. (2016) Evaluation of IWV from the numerical weather prediction WRF model with PPP GNSS processing for Bulgaria. *Atmospheric Measurement Techniques Discussions*, 1–15. https://doi.org/10.5194/amt-2016-152

Sissenwine, N., Wexler, H. & Dubin, M. (1962) The U.S. standard atmosphere, 1962. *Journal of Geophysical Research*, 67(9), 3627–3630. https://doi.org/10.1029/JZ067i009p03627

Smith, E.K. & Weintraub, S. (1953) The constants in the equation for atmospheric refractive index at radio frequencies. *Journal of Research of the Notional Bureau of Standards*, 50(1), 39–41.

Ssenyunzi, R.C. et al. (2020) Performance of ERA5 data in retrieving precipitable water vapour over east African tropical region. *Advances in Space Research*, 65(8), 1877–1893. https://doi.org/10.1016/j.asr.2020.02.003

Ssenyunzi, R.C., Oruru, B., D'ujanga, F.M., Realini, E., Barindelli, S., Tagliaferro, G. et al. (2019) Variability and accuracy of zenith total delay over the east African tropical region. *Advances in Space Research*. https://doi.org/10.1016/j.asr.2019.05.027

Suparta, W. & Alhasa, K.M. (2013) Application of ANFIS model for prediction of zenith tropospheric delay. In: *2013 3rd International Conference on Instrumentation, Communications, Information Technology and Biomedical Engineering (ICICI-BME)*, pp. 172–177. https://doi.org/10.1109/ICICI-BME.2013.6698487

Suparta, W. & Alhasa, K.M. (2015) Modeling of zenith path delay over Antarctica using an adaptive neuro fuzzy inference system technique. *Expert Systems with Applications*, 42(3), 1050–1064. https://doi.org/10.1016/j.eswa.2014.09.029

Suparta, W. & Alhasa, K.M. (2016) *Modeling of tropospheric delays using ANFIS*. New York: Springer.

de Souza, J.B., Reisen, V.A., Franco, G.C., Ispány, M., Bondon, P. & Santos, J.M. (2018) Generalized additive models with principal component analysis: an application to time series of respiratory

disease and air pollution data. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 67(2), 453–480. https://doi.org/10.1111/rssc.12239

Sun, J., Wu, Z., Yin, Z. & Ma, B. (2017) A simplified GNSS tropospheric delay model based on the nonlinear hypothesis. *GPS Solutions*, 21(4), 1735–1745. https://doi.org/10.1007/s10291-017-0644-3

Taylor, K.E. (2001) Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, 106(D7), 7183–7192. https://doi.org/10.1029/2000JD900719

Tesmer, V., Boehm, J., Heinkelmann, R. & Schuh, H. (2007) Effect of different tropospheric mapping functions on the TRF, CRF and position time-series estimated from VLBI. *Journal of Geodesy*, 81 (6–8), 409–421. https://doi.org/10.1007/s00190-006-0126-9

Ul-Saufie, a.Z., Yahya, A.S. & Ramli, N.a. (2011) Improving multiple linear regression model using principal component analysis for predicting PM10 concentration in Seberang Prai, Pulau Pinang. *International Journal of Environmental Science*, 2(2), 403–410.

Urquhart, L., Santos, M., Nievinski, F. & Böhm, J. (2014) Generation and assessment of VMF1-type grids using North-American numerical weather models. *International Association of Geodesy Symposia*, 139, 3–9. https://doi.org/10.1007/978-3-642-37222-3_1

Valizadeh, M. & Sohrabi, M.R. (2018) The application of artificial neural networks and support vector regression for simultaneous spectrophotometric determination of commercial eye drop contents. *Spectrochimica Acta—Part A: Molecular and Biomolecular Spectroscopy*, 193, 297–304. https://doi.org/10.1016/j.saa.2017.11.056

Vapnik, V., Golowich, S.E. & Smola, A. J. (1997) Support vector method for function approximation, regression estimation and signal processing. In: *Advances in neural information processing systems*. Cambridge MA: MIT Press, pp. 281–287.

Walpersdorf, A., Bouin, M.-N., Bock, O. & Doerflinger, E. (2007) Assessment of GPS data for meteorological applications over Africa: study of error sources and analysis of positioning accuracy. *Journal of Atmospheric and Solar-Terrestrial Physics*, 69 (12), 1312–1330. https://doi.org/10.1016/j.jastp.2007.04.008

Wen, L., Li, Y. & Chai, J. (2021) Multiple nonlinear regression models for predicting deformation behavior of concrete-face Rockfill Dams. *International Journal of Geomechanics*, 21(2), 04020253. https://doi.org/10.1061/(asce)gm.1943-5622.0001912

Wilgan, K. & Geiger, A. (2019) High-resolution models of tropospheric delays and refractivity based on GNSS and numerical weather prediction data for alpine regions in Switzerland. *Journal of Geodesy*, 93(6), 819–835. https://doi.org/10.1007/s00190-018-1203-6

Xu, C., Yao, Y., Shi, J., Zhang, Q. & Peng, W. (2020) Development of global tropospheric empirical correction model with high temporal resolution. *Remote Sensing*, 12(4), 721. https://doi.org/10.3390/rs12040721

Yao, Y.-B., He, C.-Y., Zhang, B. & Xu, C.-Q. (2013) A new global zenith tropospheric delay model GZTD. *Chinese Journal Geophysics*, 56(7), 2218–2227. https://doi.org/10.6038/cjg20130709

Yao, Y., Xu, X. & Hu, Y. (2017) Precision analysis of GGOS tropospheric delay product and its application in PPP. *Acta Geodaetica et Cartographica Sinica*, 46(3), 278–287. https://doi.org/10.11947/j.AGCS.2017.20160383

Yao, Y., Xu, X., Xu, C., Peng, W. & Wan, Y. (2018) GGOS tropospheric delay forecast product performance evaluation and its application in real-time PPP. *Journal of Atmospheric and Solar-Terrestrial Physics*, 175(2018), 1–17. https://doi.org/10.1016/j.jastp.2018.05.002

Yang, F., Guo, J., Zhang, C., Li, Y. & Li, J. (2021) A regional zenith tropospheric delay (ZTD) model based on. *Remote Sensing*, 13(5), 838. https://doi.org/10.3390/rs13050838

Yang, L., Gao, J., Zhu, D., Zheng, N. & Li, Z. (2020) Improved zenith tropospheric delay modeling using the piecewise model of atmospheric refractivity. *Remote Sensing*, 12(23), 3876.

Yang, Y., Xu, T., & Ren, L. (2017) A new regional tropospheric delay correction model based on BP neural network. In: *Forum on Cooperative Positioning and Service (CPGPS)*. Harbin, China: IEEE, pp. 96–100. https://doi.org/10.1109/CPGPS.2017.8075104

Yazar, I., Yavuz, H.S. & Yavuz, A.A. (2017) Comparison of various regression models for predicting compressor and turbine performance parameters. *Energy*, 140, 1398–1406. https://doi.org/10.1016/j.energy.2017.05.061

Younes, S.A.-M. (2016) Modeling investigation of wet tropospheric delay error and precipitable water vapor content in Egypt. *The Egyptian Journal of Remote Sensing and Space Science*, 19(2), 333–342. https://doi.org/10.1016/j.ejrs.2016.05.002

Yuan, Y., Holden, L., Kealy, A., Choy, S. & Hordyniec, P. (2019) Assessment of forecast Vienna Mapping Function 1 for real-time tropospheric delay modeling in GNSS. *Journal of Geodesy*, 93, 1–14.

Zhao, Q., Yao, Y., Yao, W. & Li, Z. (2018) Real-time precise point positioning-based zenith tropospheric delay for precipitation forecasting. *Scientific Reports*, 8(1), 1–12. https://doi.org/10.1038/s41598-018-26299-3

Zhang, Z., Balay, J.W. & Liu, C. (2019) Regional regression models for estimating monthly streamflows. *Science of the Total Environment*, 706, 135729. https://doi.org/10.1016/j.scitotenv.2019.135729

Zhang, Q., Li, F., Zhang, S. & Li, W. (2020) Modeling and forecasting the gps zenith troposphere delay inwest antarctica based on different blind source separation methods and deep learning. *Sensors*, 20(8). https://doi.org/10.3390/s20082343

Zhang, H., Yuan, Y., Li, W., Li, Y. & Chai, Y. (2016) Assessment of three tropospheric delay models (IGGtrop, EGNOS and UNB3m) based on precise point positioning in the Chinese region. *Sensors*, 16(1), 122.

Zheng, D.Y., Hu, W.S., Wang, J. & Zhu, M.C. (2015) Research on regional zenith tropospheric delay based on neural network technology. *Survey Review*, 47(343), 286–294. https://doi.org/10.1179/1752270614Y.0000000130

Zhou, F., Cao, X., Ge, Y. & Li, W. (2020) Assessment of the positioning performance and tropospheric delay retrieval with precise point positioning using products from different analysis centers. *GPS Solutions*, 24(1), 12. https://doi.org/10.1007/s10291-019-0925-0

Zhu, Q., Zhao, Z., Lin, L. & Wu, Z. (2010) Accuracy improvement of zenith tropospheric delay estimation based on GPS precise point positioning algorithm. *Geo-Spatial Information Science*, 13(4), 306–310. https://doi.org/10.1007/s11806-010-0400-0

How to cite this article: Osah, S., Acheampong, A. A., Fosu, C., & Dadzie, I. (2021). Regression models for predicting daily IGS zenith tropospheric delays in West Africa: Implication for GNSS meteorology and positioning applications. *Meteorological Applications*, *28*(5), e2030. https://doi.org/10.1002/met.2030