IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 10 July 2024, accepted 7 August 2024, date of publication 9 August 2024, date of current version 23 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3441331

RESEARCH ARTICLE

New Models of Zenith Tropospheric Delay for Chinese Mainland and Surrounding Areas Based on Convolutional Neural Network and Random Forest

JIAHAO ZHANG[1], QIN LIANG[1,2], AND YUNQING HUANG[1,2]
[1]School of Mathematics and Computational Science, Xiangtan University, Xiangtan 411105, China
[2]National Center for Applied Mathematics in Hunan, Xiangtan 411105, China

Corresponding authors: Qin Liang (liangqin1997@xtu.edu.cn) and Yunqing Huang (huangyq@xtu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFA0713500, in part by the Key Project of Hunan Education Department under Grant 22A0126, and in part by the Natural Science Foundation of Hunan Province under Grant 2022JJ30555.

ABSTRACT Accurate models of zenith tropospheric delay (ZTD) is crucial in meteorology as well as in navigation and positioning. In this study, we employ Convolutional Neural Network (CNN) and Random Forest (RF) models to establish six direct or compensation models for estimating ZTD in Chinese mainland and surrounding areas. The modeling process utilizes ZTD data from 205 stations spanning the period 2013 to 2018. Model validity is assessed using ZTD data from 202 stations in 2019. Comparative analysis, considering the overall Root Mean Square Error (RMSE), is conducted between these newly proposed CNN/RF-based models and Saastamoinen, A&N, GPT3 and RF-based models constructed by the methods presented in the previous study (ZTD-RF1, ZTD-RF3). The results demonstrate the superiority of the six CNN/RF-based models over the previously proposed models. In general, compensation models exhibit an improvement over direct models, and models incorporating meteorological parameterisation outperform models without such parameterisation. When the meteorological data are available, our proposed model provided a good representation of the instability of water vapour pressure in the ZTD, especially in monsoon climates. The optimal model is identified as the RF-based compensation model (ZTD-RF4). The ZTD-RF4 model achieves an overall RMSE of 3.24 cm, representing a 29.47% reduction of the RMSE compared to the Saastamoinen model (4.60 cm), a 26.75% reduction compared to the A&N model (4.43 cm), and slightly superior to the ZTD-RF3 model (3.28 cm). When the meteorological data are unavailable, the optimal choice is the CNN-based compensation model (ZTD-CNN2), which exhibits an overall RMSE of 4.21 cm, indicating a 7.89% reduction compared to the GPT3 model (4.57 cm) and significantly superior to the ZTD-RF1 model (4.34 cm). In contrast to current machine learning (ML)-based ZTD calculation models, we introduce the idea of compensation based on traditional models and a new CNN structure is constructed, which all proved to be capable of better performance in ZTD modeling.

INDEX TERMS Convolutional neural network, Chinese mainland and surrounding areas, GPT3 model, random forest, Saastamoinen model, zenith tropospheric delay.

I. INTRODUCTION

Tropospheric delays caused by refraction in the neutral atmosphere are one of the major sources of error in Global Navigation Satellite System (GNSS) measurements. Beyond their impact on positioning accuracy, these delays serve as a

The associate editor coordinating the review of this manuscript and approving it for publication was Seifedine Kadry.

valuable data source for assessing atmospheric water vapor levels. The accuracy of zenith tropospheric delay models directly affects the accuracy of weather forecasts and the convergence rate of high-precision positioning [1], [2]. Zenith tropospheric delay consists of zenith hydrostatic delay (ZHD) caused by dry gases and induced polarized moment of water vapor, and zenith wet delay (ZWD) caused by water vapor. In this study, we focus on the accurate modeling of ZTD.

The exact ZTD value can be obtained by integrating refractivity in the zenith direction, from the station altitude to the top of the lower atmosphere [15]. However, due to the absence of data far above the station, this exact ZTD is difficult to obtain. Therefore, a series of approximation models [3], [4], [5], basing only on data of the station altitude, have been established. To estimate the ZTD with these models, the surface meteorological data should be provided. In cases where surface meteorological data are unavailable, the Global Pressure and Temperature (GPT) models [6], [7], [8], [9] can be utilized to provide an approximation of these meteorological data. In the main text, we call models that use surface meteorological data as ''models with requiring meteorological data''. Models, which incorperate GPT, that don't need direct meteorological data are called ''models without requiring meteorological data''.

In 1969, the Hopfield model was developed by analyzing and validating measured meteorological soundings from 18 stations around the world [3]. Then in 1973, the Saastamoinen model was constructed based on the U.S. Standard Atmosphere Model (SAM). This model divided the troposphere into two layers and integrated them separately, accounting for changes in gravitational acceleration in the calculations [4]. Modeling ZWD is exceptionally difficult due to the uneven and rapidly changing distribution of water vapor in the atmosphere over time. To address this, Askne and Nordius (A&N) developed a more accurate model by estimating mean temperature weighted with water vapor pressure (T_m) and water vapor decrease factor (λ) in 1987 [5]. These models mentioned in this paragraph are based on meteorological data at the stations. To use these models for ZTD estimation, it is necessary to obtain the information at the station including the float day of year (doy), the latitude (lat), the longitude (lon), the ellipsoid height (h_s) and the surface meteorological data at the station including the surface pressure (P_s), the surface temperature (T_s), the surface partial pressure of water vapor (e_s).

In the absence of surface meteorological data T_m, P_s, T_s, e_s, Boehm et al. [6] developed an empirical model called the Global Pressure and Temperature (GPT) model to predict these meteorological data. This model, constructed using the accumulation of meteorological data and employing ninth-order spherical harmonic functions, was developed based on 40 years of reanalysis data from the European Centre for Medium-Range Weather Forecasts (ECMWF). Following that, the GPT2 model [7] was realized using monthly mean profiles of ERA-Interim data, the GPT2w model [8] and the GPT3 model [9] were developed by adopting the functional formulations with the predefined periodic form to account for climatological mean and seasonal variations of meteorological parameters and their vertical distribution. They were realized mainly based on the monthly reanalysis data of the numerical weather model (NWM). Therefore, these models without requiring surface meteorological data are unsuitable for accurately forecasting the nonstationary patterns or short-term fluctuations in the neutral atmosphere. Moreover, due to the limited spatial resolution of NWM data and the usage of predefined functional formulation, the global empirical troposphere models cannot well represent complex variation in the tropospheric wet delay for the regions with specific climate patterns. When using these models for ZTD estimation, only the doy, lat, lon, h_s of the station are needed.

The models mentioned above all depend on using single functions for modeling and fitting, which may not deal with complex nonlinear variations of ZTD accurately. However, regression analysis based on machine learning methods excels in modeling nonlinear features through using a large amount of historical data. Hence, machine learning methods have been widely applied to various fields, including the fields of meteorology and satellite geodesy such as weather forecasting [10], ionospheric parameter modeling [11] and modeling ZTD. Xiao et al. used improved back propagation neural network (BPNN) to construct a tropospheric zenith total delay (ZTD) spatial forecast model for Japan based on 11-day tropospheric data provided by National Center Atmospheric Research(NCAR), achieved an root mean squared error(RMSE) of 8.52mm which significantly superior to four-parameter model the paper mentioned [12]. Lu et al. developed TropNet model based on a deep spatiotemporal neural network for 6 hours of ZWD forecasting by combining information provided by the Geostationary Operational Environmental Satellite-R series and the global forecast system (GFS), showing an overall improvement of 15.5% when compared to the GFS ZWD [13]. Li et al. developed BPNN-based and random forest(RF)-based ZWD models for Chinese mainland through regression analysis of seven-year radiosonde data, obtained an overall accuracy of 4.7cm for models without requiring surface meteorological data and an overall accuracies of less 3.7cm for models with meteorological parameterization [14]. Li et al. proposed improved GPT3 model for spatiotemporal ZWD forecasting in Japan by adopting radial basis function (RBF) to offset the error between the ZTD estimates calculated by GPT3 model and the GNSS-derived ZTD,obtained a mean RMSE of 37.8mm [15]. Li et al. analyzed the relationship between the residuals of GPT3-derived ZTD minus GNSS-derived ZTD and the spatiotemporal information by using RF to model ZTD for spatial ZWD forecasting in Chinese mainland, achieved an significant improvement compared to GPT3 model and BPNN-based model [16]. Crocetti et al. built a global spatial ZWD forecast model based on Extreme Gradient Boosting (XGBoost), which achieved an RMSE squared error of 8.1mm [17].

Up to now, few studied have focused ZTD modeling based on machine learning in Chinese mainland and surrounding areas. Therefore, for better estimating and forecasting the ZTD in Chinese mainland and surrounding areas, convolutional neural network (CNN) and random forest (RF) with a strong nonlinear ftting ability are adopted to model ZTD using historical radiosonde data at 205 stations from 2013 to 2018 in this study. Six different models, direct or compensated, with or without requiring surface meteorological data, are established to meet the demand for high-precision calculation of ZTD in Chinese mainland and the surrounding areas. Compared with the existing machine learning based ZTD calculation models [14], [18], the main advancement of this study can be summarised as: (1) we construct a new structure of CNN; (2) introducing CNN approach to the ZTD modeling problem; (3) ideas for the construction of ZTD compensation models based on machine learning (surface meteorological data is available or not) are provided; (4) our models can be used to forecast ZTD more accurately than developed models including Saastamoinen, A&N, GPT3 models and RF-based models constructed by the methods presented in the previous study [14]; (5) the established ZTD models was analysed for full performance in space, time and altitude.

The rest part of this paper is organized as follows. In Section II, we introduce the sources of data and some preprocessing and utilization methods. Subsequently, we detail the calculation of the ''exact'' ZTD value, which involves integrating radiosonde data to approximate the ZTD value. Following that, we present three traditional models (Saastamoinen, A&N, and GPT3) that solely rely on surface data at the station, yielding approximations of the ''exact'' ZTD. Section III outlines the construction of six new CNN/RF-based ZTD models. In Section IV, we validate the effectiveness of the proposed ZTD models to estimate ZTD at 202 stations for 2019. The results are compared with the Saastamoinen, A&N, GPT3, ZTD-RF1 and ZTD-RF3 models. Finally, we elaborate on and summarize the conclusions in Section V.

II. DATA AND CALCULATION OF ZTD

In this section, we describe the data source used to build the new ZTD models and some methods of data quality control. Then, we describe the process of calculating the ''exact'' ZTD by using radiosonde data. Moreover, ZTD determination using the GPT3 [9] and models proposed by Saastamoinen [4], and Askne and Nordius [5] is also presented.

A. DATA SOURCE AND PREPROCESSING

The radiosonde data used in this study were provided by the Integrated Global Radiosonde Archive (IGRA). The IGRA radiosonde data we used can be downloaded at https://www.ncei.noaa.gov/pub/data/igra/. The radiosonde data provides vertical profiles of meteorological variables, which include air temperature, total pressure, relative humidity, geopotential height, wind speed, etc., at a series of significant isobaric levels [19] covering the troposphere. This study selected 205 stations in the range of $8^\circ N - 62^\circ N$ and $68^\circ E - 142^\circ E$. Due to the partial absence of IGRA data, the following data quality control methods were used in this study: (1) the sounding records with two adjacent pressure levels exceeding 200 hPa were removed; (2) the sounding records containing a top pressure layer exceeding 300 hPa have been excluded; (3) the sounding records with a top partial pressure of water vapor exceeding 0.1 hPa have been excluded.

B. CALCULATION OF THE ''EXACT'' ZTD

ZTD value can be obtained by integrating refractivity in the zenith direction, from the station altitude to the top of the lower atmosphere [20]. ZTD can be divided into ZHD, caused by dry air, and ZWD, caused by water vapor. Since there is a gap between the top altitude that can be reached by the soundings and the top altitude involved in ZHD, ZHD can be calculated using the radiosonde data and using Saastamoinen model in the top layer of the radiosonde data. This calculation method was based on previous studies, which showed that the ZHD results of Saastamoinen model agreed with ray tracing results at the sub-millimetre level [21], [22]. Therefore, ZHD can be calculated by the formula:

$$ZHD_{radio} = 10^{-6} \cdot \int_{h_{geo}}^{h_{top}} N_h dh + \frac{0.0022768 \cdot P_{top}}{1 - 0.0026 \cdot \cos(2 \cdot lat) - 0.00028 \cdot h_{top}}. \tag{1}$$

Here h_{geo} and lat are the station's geopotential height (unit: m) and latitude, respectively. h_{top} and P_{top} represent geopotential height (unit: m) and pressure (unit: hPa) at the top layer of the radiosonde data, respectively. N_h is the hydrostatic refractivity in the lower atmosphere. It can be expressed approximately as [20]:

$$N_h = k_1 \cdot \frac{P_d}{T} \cdot Z_d^{-1}, \tag{2}$$

$$Z_d^{-1} = 1 + P_d \cdot \left[57.96 \cdot 10^{-8} \left(1 + \frac{0.52}{T}\right) - 9.4611 \cdot 10^{-4} \cdot \frac{T_C}{T^2}\right]. \tag{3}$$

Here P_d, T and $T_C = T - 273.15$ are the partial pressure of dry air (unit: hPa), the absolute temperature (unit: K) and Celsius temperature (unit:°C) at a certain sounding level of the radiosonde data, respectively. Since the partial pressure of water vapor at the top layer in the radiosonde data is very small, the part of the zenith wet delay above the top layer can be ignored. ZWD can be calculated using the radiosonde data by the formula [20]:

$$ZWD_{radio} = 10^{-6} \cdot \int_{h_{geo}}^{h_{top}} N_w dh, \tag{4}$$

$$N_w = \left[\left(k_2 - k_1 \cdot \frac{M_w}{M_d}\right) \cdot \frac{P_w}{T} + k_3 \cdot \frac{P_w}{T^2}\right] \cdot Z_w^{-1}, \tag{5}$$

$$Z_w^{-1} = 1 + \frac{1650P_w \cdot T_C^*}{T^3}. \quad (6)$$

Here N_w is wet refractivity in the lower atmosphere and $T_C^* = 1 - 0.01317T_C + 1.75 \cdot 10^{-4}T_C^2 + 1.44 \cdot 10^{-6}T_C^3$. P_w is the partial pressure of water vapor (unit: hPa) at a certain sounding level of the radiosonde data. $k_1 = 77.69$, $k_2 = 71.2952$ and $k_3 = 375463$ are atmospheric refraction coefficients, Z_d and Z_w are compressibility factors for dry air and water vapor, respectively. $M_d = 18.0152$ and $M_w = 28.9644$ are molar mass (unit: g/mol) for dry air and water vapor, respectively. Note in (2-6), data $P = P_d + P_w$, T, P_w are all related to the height h.where P is the total pressure (unit: hPa) at a certain sounding level of the radiosonde data. These data can be obtained from the radiosonde data.

Since the altitude used for the station height is different from the standard of geopotential height used in the radiosonde data, we convert the altitude of the station to the geopotential height by the formula [23]:

$$h_{geo} = \frac{Y_s(lat)}{Y_{45}} \cdot \frac{R(lat) \cdot h_s}{R(lat) + h_s}, \quad (7)$$

$$Y_s(lat) = 9.780325 \cdot \left[\frac{1 + 0.00193185 \cdot \sin^2(lat)}{1 - 0.00669435 \cdot \sin^2(lat)}\right]^{0.5}, \quad (8)$$

$$R(lat) = \frac{6378.137}{1.006803 - 0.006706 \cdot \sin^2(lat)}, \quad (9)$$

where lat and h_s are the latitude and the ellipsoid height at the station. $Y_s(lat)$ is the normal gravity (unit: m/s^2) on the surface of the ellipsoid of revolution for latitude, $R(lat)$ represents an effective radius (unit: m) of the earth for latitude lat, $Y_{45} = 9.80665$ is the normal gravity (unit: m/s^2) on the surface of the ellipsoid of revolution for latitude 45°. The surface meteorological data at the station can be derived from the radiosonde data if the station's geopotential height does not coincide with any layer in the radiosonde data [24].

Using numerical integration in the calculation process of (1) and (4), we can obtain the **"exact" ZTD value**, T_{radio}, from the radiosonde data by

$$T_{radio} = ZHD_{radio} + ZWD_{radio}. \quad (10)$$

C. TRADITIONAL MODELS

Here, we present the traditional models commonly used for calculating ZTD when only surface data at the station is provided. If meteorological data is available, the Saastamoinen or A&N models can be employed to estimate ZTD. The expression for ZTD in the Saastamoinen model can be written as [4]

$$T_{saas} = \frac{0.0022768 \cdot [P_s + (1255/T_s + 0.05) \cdot e_s]}{1 - 0.0026 \cdot \cos(2 \cdot lat) - 0.00028 \cdot h_s}. \quad (11)$$

Here P_s is the surface pressure, T_s is the surface temperature, e_s is the surface partial pressure of water vapor.

In order to establish a more accurate ZTD model, the A&N model was built through estimating water vapor decrease factor (λ) and the mean temperature (T_m), weighted with water vapor pressure. In the A&N model, ZHD and ZWD can be calculated by [5]:

$$ZHD_{A\&N} = \frac{0.0022768 \cdot P_s}{1 - 0.0026 \cdot \cos(2 \cdot lat) - 0.00028 \cdot h_s}, \quad (12)$$

$$ZWD_{A\&N} = 10^{-6} \cdot \left[\left(k_2 - k_1 \cdot \frac{M_w}{M_d}\right) + \frac{k_3}{T_m}\right] \cdot \frac{R_d}{(\lambda + 1) \cdot g_m} \cdot e_s, \quad (13)$$

where R_d and g_m are the gas constant for dry air and the gravitational acceleration (unit: m/s^2) at the mass center of the vertical column of the atmosphere, respectively. The ZTD value calculated by the A&N model, $T_{A\&N}$, can be written as

$$T_{A\&N} = ZHD_{A\&N} + ZWD_{A\&N}. \quad (14)$$

If the surface meteorological data is unavailable or incomplete, we can use the GPT3 model to estimate the surface meteorological data at the station. Substitute the estimated meteorological data into the A&N model, then we can get the ZTD estimation of the GPT3 model ($R_d = 287.0464\text{JK}^{-1}\text{kg}^{-1}$ and $g_m = 9.80665\text{m/s}^2$) [9]. The GPT3 model T_{GPT3} uses global grids of the mean and seasonal amplitude coefficients for meteorological parameters to estimate the surface meteorological data at a specific station. The horizontal resolution of the grid in the GPT3 model had been classified as 1° × 1° or 5° × 5°, this study adopted a 1° × 1° grid which is more accurate.

III. NEW ZTD MODELS BASED ON CONVOLTIONAL NEURAL NETWORK AND RANDOM FOREST

The main objective of this study is to build two types of ZTD models using 2013-2018 radiosonde data from 205 stations in Chinese mainland and surrounding areas: the CNN/RF-based ZTD direct models and the CNN/RF-based ZTD compensated models. CNN and RF are two well-known models for data regression analysis. One should provide a large group of dataset for training or testing. Each set of data should include the input variables and the corresponding target variable. For models with meteorological parameterization, the input variables include the float day of year (doy), the latitude (lat), the longitude (lon), the ellipsoidal height (h_s), the surface pressure (P_s), the surface temperature (T_s) and the surface partial pressure of water vapor (e_s). Meanwhile, It is known that P_s and e_s can be converted by the relative humidity (RH): $e_s = P_s \cdot RH$. However, since RH varies with time and location, if the estimation of RH is used to establish the relationship between P_s and e_s, choosing either P_s or e_s as an input parameter will introduce the estimation error of RH into the estimation of the ZTD. Therefore, we choose both P_s and e_s as the exact input parameters to avoid unnecessary errors caused by the conversion between the two during the modeling process. For models without meteorological parameterization, the input variables only include doy, lat, lon and h_s. For direct models, the target variable is set to be the "exact" ZTD, while for ZTD compensation model, the target variable is set to be the difference between the "exact"

ZTD and the ZTD estimates computed by the baseline model. In this study, six different models are constructed to meet the demands of different applications. The ZTD-CNN1 schemes are the direct models without requiring surface meteorological data based on CNN, can be expressed as

$$T_{ZTD-CNN1} = f_{ZTD-CNN1}(doy, lat, lon, h_s); \quad (15)$$

the ZTD-CNN2 and ZTD-RF2 schemes are the GPT3-compensated models without requiring surface meteorological data based on CNN and RF, can be expressed as

$$T_{ZTD-CNN2/RF2} - T_{GPT3} = f_{ZTD-CNN1/RF1}(doy, lat, lon, h_s); \quad (16)$$

the ZTD-CNN3 scheme is the direct models with requiring surface meteorological data based on CNN, can be expressed as

$$T_{ZTD-CNN3} = f_{ZTD-CNN3}(doy, lat, lon, h_s, P_s, T_s, e_s); \quad (17)$$

the ZTD-CNN4 and ZTD-RF4 schemes are the Saastamoinen-compensated models with requiring surface meteorological data based on CNN and RF, can be expressed as

$$T_{ZTD-CNN4/RF4} - T_{saas} = f_{ZTD-CNN4/RF4}(doy, lat, lon, h_s, P_s, T_s, e_s); \quad (18)$$

These schemes can be applied flexibly under different conditions. Moreover, the ZTD-RF1 and ZTD-RF3 schemes presented in the previous study [14] as a comparison. Involved input and target variables of these models are summarized in Table 1. Validation and comparison of these models are presented in Section IV.

A. CONVOLUTIONAL NEURAL NETWORK (CNN)

CNN, as one of the most widely used and well-established method in deep learning, was initially employed in the field of text recognition [25]. Deep learning algorithms can automatically extract features and patterns from data, enabling the accomplishment of more complex tasks [26], [27]. Therefore, we constructed a new multi-layer structure of CNN as a data regression analysis tool to calculate the ZTD if the ''Input variables'' are provided in this study.

Fig. 1 shows the basic structure of regression calculation using CNN. We need to point out that the ''Input layer'' corresponds to the column vector composed of ''The input variables'' in Table 1, the ''Output layer'' corresponds to the estimate of the ''Target variable'' in Table 1. n is the number of the input variables, m is the number of feature maps, p is the size of the convolution kernel. Key components of CNNs include convolutional layers, pooling layer and activation functions. The Convolutional layer primarily operates on the data feature matrix provided by the Input layer. It uses convolutional kernels to map these features to the subsequent layer, facilitating deep learning of data features and enabling profound exploration of their characteristics.The convolution layer can be expressed as:

$$x_j^K = f\left(\sum_{i \in M_j} x_i^{K-1} \cdot w_{ij}^K + b_j^K\right), \quad (19)$$

where w_{ij}^K and b_j^K are the weights and biases corresponding to the convolution filters, respectively. x_j^K and x_j^{K-1} represent the feature mapping of the previous layer and the current layer, respectively. M_j denotes the set of feature mappings, and $f(\cdot)$ represents the activation function. The activation function in the network can improve the nonlinear fitting ability of CNN. The ReLU activation function was chosen in this study. It can be expressed as [28]:

$$\mathrm{Re\,LU}(x) = \max(0, x). \quad (20)$$

Batch Normalization in CNN effectively addresses the issue where parameter updates and iterations impede the convergence speed of the training model [29]. The pooling layer is added in the network to preserve the essential information of input variables and compress data dimensions through the down-sampling. In this study, the max pooling

$$Pooling^{\max} = \max_{i,j \in \Omega} a_{i,j}, \quad (21)$$

was employed. Here $a_{i,j}$ refers to the output from the previous convolutional layer, and i and j denote the index, Ω denotes the pooling window. In the process of parameters training in CNN, we used the stochastic gradient descent with momentum (SGDM) algorithm [30].

The selection of optimal hyperparameters is beneficial to the improvement of the CNN regression model. Considering the substantial impact of the number of feature maps (m) and the convolutional kernel size (p) on model fitting, and acknowledging the constraint on p, we conducted a hyperparameter selection. For the ZTD-CNN3 and ZTD-CNN4 schemes, the number of kernels and the size of the convolutional kernel were set to the ranges of 1-50 and 1-7, respectively. While for the ZTD-CNN1 and ZTD-CNN2 schemes, the range of p was from 1 to 4. These selections were made to model the effects at 202 stations throughout the year 2019 in our study. As shown in Fig. 2, once p is fixed, the RMSE fluctuates very little when $m \geqslant 30$. So we set the hyperparameters for the ZTD-CNN1 and ZTD-CNN2 schemes to be m, $p = 40$, 3, and the hyperparameters for the ZTD-CNN3 and ZTD-CNN4 schemes to be m, $p = 40$, 5. Therefore, the internal CNN structure used for modeling ZTD in this study consists of the convolutional layer consists of 40 feature maps with kernel size (3×1 or 5×1), the max pooling layer which employs a pooling window of (2, 1) and a stride size of (1, 1), the convolutional layer consists of 80 feature maps with kernel size (3×1 or 5×1), fully connected layer with output size is 1 in this order.

B. RANDOM FOREST (RF)

RF is an ensemble learning algorithm proposed by Breiman [31]. It uses multiple decision trees and averages

TABLE 1. Basic training structure of new ZTD models based on CNN and RF (The ZTD-CNN1 and ZTD-RF1 schemes are direct models without meteorological parameterisation; The ZTD-CNN2 and ZTD-RF2 schemes are the GPT3-compensated models without meteorological parameterisation; The ZTD-CNN3 and ZTD-RF3 schemes are direct models with meteorological parameterisation; The ZTD-CNN4 and ZTD-RF4 schemes are the saastamoinen-compensated models with meteorological parameters).

Methods	Schemes	The input variables	Target variable
CNN	ZTD-CNN1	doy , lat , lon , h_s	T_{radio}
	ZTD-CNN2	doy , lat , lon , h_s	$T_{radio} - T_{GPT3}$
	ZTD-CNN3	doy , lat , lon , h_s , P_s , T_s , e_s	T_{radio}
	ZTD-CNN4	doy , lat , lon , h_s , P_s , T_s , e_s	$T_{radio} - T_{saas}$
RF	ZTD-RF1	doy , lat , lon , h_s	T_{radio}
	ZTD-RF2	doy , lat , lon , h_s	$T_{radio} - T_{GPT3}$
	ZTD-RF3	doy , lat , lon , h_s , P_s , T_s , e_s	T_{radio}
	ZTD-RF4	doy , lat , lon , h_s , P_s , T_s , e_s	$T_{radio} - T_{saas}$

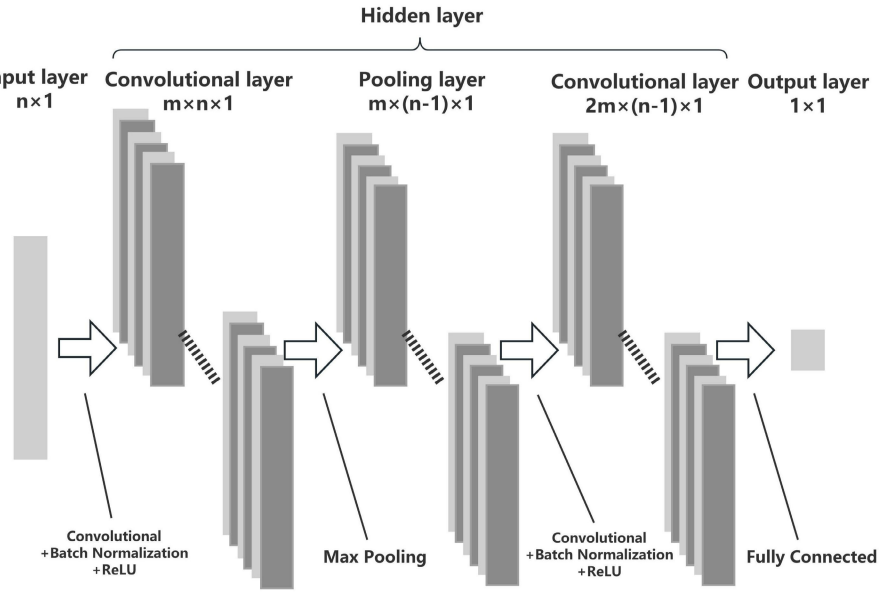

FIGURE 1. Adopted basic structure of regression calculation based on CNN.

the outputs of those trees to solve regression calculation problems. Each decision tree is trained by utilizing randomly selected feature subsets. Various training sets are created by randomly sampling from the dataset based on these feature subsets.

This strong randomness can reduce overfitting and enhance generalization capabilities. RF is widely applied to regression analysis problems due to its robustness and strong nonlinear ftting ability [32]. Therefore, we use RF for ZTD modeling to improve the nonlinear ftting ability. In this study, we adopted the classification and regression tree (CART) [33] as the decision tree algorithm. The basic structure of RF-based regression calculation adopted in our study is shown in Fig. 3. Note that ''Input'' and ''Output'' correspond to the vector of ''The input'' variables and the estimate of the ''Target variable'' in Table 1, respectively. t is the number of subdatasets, i.e., the number of decision trees, generated by put-back sampling of the original dataset, and l is the minimum number of samples required at the leaf nodes. In the process of constructing decision trees, the evaluation of the goodness of cutoff variables and cutoff points is crucial. We employ the weighted sum of the mean square error (MSE) of each child node as the criterion for this evaluation. The judgment criteria can be expressed as:

$$M = \frac{N_1}{N_a} \cdot \mathrm{MSE}(X_1) + \frac{N_2}{N_a} \cdot \mathrm{MSE}(X_2), \tag{22}$$

where X_1 and X_2 denote the subsets of training samples after slicing. N_1 and N_2 denote the number of samples for X_1 and X_2 respectively. N_a denote the total number of samples. The

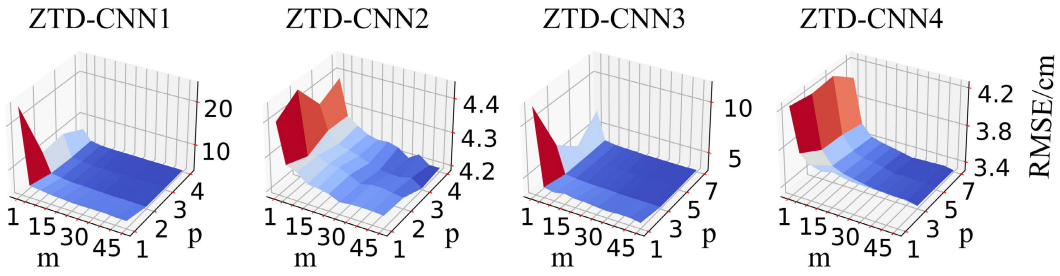

FIGURE 2. RMSE of different schemes based on CNN with different hyperparameters.

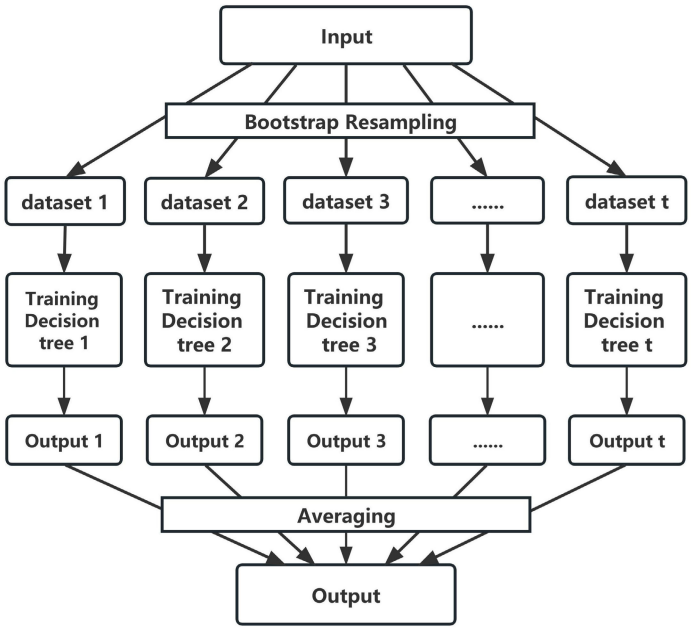

FIGURE 3. Adopted basic structure of regression calculation based on RF.

MSE of a set can be described as:

$$MSE(X) = \frac{1}{N} \cdot \sum_{i=1}^{N} (y_i - \bar{y})^2, \tag{23}$$

where $X = X_1, X_2$ and $N = N_1, N_2$. Let y_i and $\bar{y}$ denote the target variable values and target variable means in X, respectively. Finally, the regression calculation results for RF are obtained by simply averaging the outputs of all decision trees [34]:

$$F(I) = \frac{1}{t} \sum_{i=1}^{t} T_i(I). \tag{24}$$

Here t is the number of decision trees, I denotes the input variables, $F(I)$ denotes the final output of the RF, and $T_i(I)$ is the output of the i-th decision tree.

Similar to CNN, the careful selection of hyperparameters is crucial during the model training process using RF. Given that t is directly associated with the training model's feature extraction capability and l is directly linked to the training model's capacity to capture noise in the training data, this study involved a hyperparameter selection process. We varied t from 1 to 50 and l from 1 to 50, examining modeling effects across 202 stations during the year 2019. As shown in Fig. 4, the overall sensitivity of RF to the two selected hyperparameters is notably weaker compared to that of CNN.. It's worth noting that the RMSE remains almost constant when and.Consequently, we fix t, l = 30 for the ZTD-RF1, ZTD-RF2, ZTD-RF3 and ZTD-RF4 schemes.

IV. MOEDL VALIDATION AND ANALYSIS

In this section, we validate and analyze our models by comparing the corresponding RMSE with traditional models and the ZTD models (ZTD-RF1, ZTD-RF3) constructed by the methods presented in the previous study [14]. First of all, the "exact" ZTD values are calculated by the radiosonde data with the method mentioned in subsection II-B. Among all these models, ZTD-CNN3, ZTD-CNN4, ZTD-RF3, ZTD-RF4, Saastamoinen and A&N are models with requiring surface meteorological data. These surface meteorological

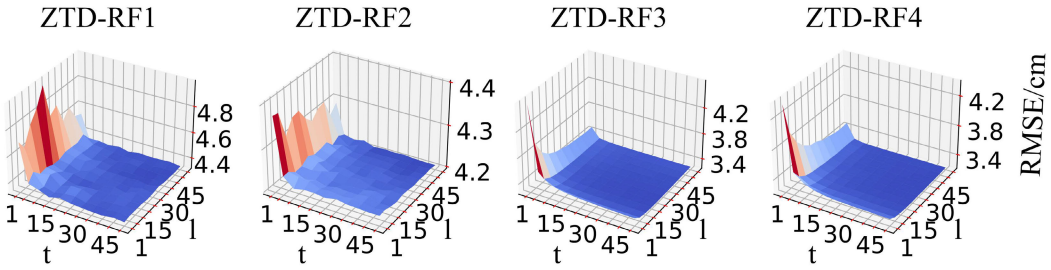

FIGURE 4. RMSE of different schemes based on RF with different hyperparameters.

data can be obtained by the radiosonde data. ZTD-CNN1, ZTD-CNN2, ZTD-RF1, ZTD-RF2, and GPT3 are models without requiring surface meteorological data. We analyze and compare our CNN/RF-based ZTD models with traditional models, ZTD-RF1 model and ZTD-RF3 model by applying them to 202 stations in Chinese mainland and surrounding areas during 2019. Considering applications on different scales, we validate and analyze the new models in overall, spatial, temporal and altitudinal aspects. The error criterions used in this study are the root mean square error (RMSE), mean error (ME) and mean absolute error (MAE). In most cases, RMSE was considered the primary criterion. These error criterions can be expressed as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (y_i - y_i^*)^2}, \tag{25}$$

$$\mathrm{ME} = \frac{1}{n} \cdot \sum_{i=1}^{n} (y_i - y_i^*), \tag{26}$$

$$\mathrm{MAE} = \frac{1}{n} \cdot \sum_{i=1}^{n} |y_i - y_i^*|, \tag{27}$$

here y_i is the estimation values, y_i^* is the ''exact'' values, i is the index, and n denotes the total number of data for estimation.

A. OVERALL PERFORMANCE AND ANALYSIS

Table 2 records the overall ME, MAE and RMSE of the ZTD calculated by different models at 202 stations in Chinese mainland and surrounding areas in 2019. Since the performance of ME and MAE is almost equivalent to that of RMSE, we exclusively compare these models based on the RMSE aspect in the following analysis.As shown in Table 2, the performance of the CNN/RF-based ZTD model has improved compared with these three traditional models.

For models without requiring surface meteorological data, the GPT3 compensationed model (ZTD-CNN2, ZTD-RF2) exhibit a slight improvement over their direct counterparts (ZTD-CNN1, ZTD-RF1). These four models also outperform the traditional GPT3. Based on the experimental results, we recommend the ZTD-CNN2 or ZTD-RF2 models when the surface meteorological data is unavailable or incomplete.

For models with requring surface meteorological data, the ZTD-CNN3, ZTD-CNN4, ZTD-RF3 and ZTD-RF4 models are superior to the Saastamoinen model and the A&N model, with improvements over 20%. Meanwhile, the Saastamoinen compensation models (ZTD-CNN4, ZTD-RF4) perform better than direct models (ZTD-CNN3, ZTD-RF3). Based on experimental results, we recommend the ZTD-CNN4 and ZTD-RF4 models for ZTD estimation when surface meteorological data is available.

Moreover, as shown in Table 2, the ZTD-CNN1, ZTD-CNN2, ZTD-RF1, and ZTD-RF2 models without requiring surface meteorological data outperform the Saastamoinen and A&N models with requiring surface meteorological data. This means that the CNN/RF-based ZTD models should be utilized regardless of the availability of meteorological data. Meanwhile, the GPT3 compensation models (ZTD-CNN2, ZTD-RF2) further improves accuracy compared to the direct models (ZTD-CNN1, ZTD-RF1). It should also be noted that, as shown in Table 2, the inclusion of meteorological parameters significantly enhanced the accuracy of the CNN/RF-based models, and the aforementioned results clearly demonstrated experimentally that there is a strong correlation between the exact values of the ZTD and the meteorological parameters. Fig. 5 illustrates the relationship between the ZTD values estimated by the models and the ''exact'' ZTD calculated by the radiosonde data at the 202 stations in Chinese mainland and surrounding areas during 2019. R^2 is an indicator that effectively evaluates the fit of the models.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i^* - y_i)^2}{\sum_{i=1}^{n} \left(y_i^* - \left(\sum_{j=1}^{n} y_j^* \right) \Big/ n \right)^2}. \tag{28}$$

The closer to 1 R^2 is, the better the model. As shown in Fig. 5. The R^2 values of the ZTD-CNN1, ZTD-CNN2, ZTD-CNN3, ZTD-CNN4, ZTD-RF1, ZTD-RF2, ZTD-RF3 and ZTD-RF4 models were 0.9702, 0.9707,0.9809,0.9815,0.9688, 0.9707, 0.9822 and 0.9826, respectively, exhibiting an improvement compared to those of the Saastamoinen (0.9651), A&N (0.9676) and GPT3 (0.9655) models. This phenomenon indicates that the new models fit significantly

TABLE 2. Overall estimation errors of all ZTD models for the 202 stations analyzed during 2019 and the improvement rates (the overall RMSE) of the new models compared to the traditional models.

Models	ME/cm	MAE/cm	RMSE/cm	Improvement rate compared to Saastamoinen	Improvement rate compared to A&N	Improvement rate compared to GPT3
ZTD-CNN1	-0.27	3.09	4.24	7.67%	4.11%	7.14%
ZTD-CNN2	-0.10	3.06	4.21	8.42%	4.89%	7.89%
ZTD-CNN3	-0.16	2.40	3.40	25.94%	23.08%	25.51%
ZTD-CNN4	-0.07	2.34	3.34	27.24%	24.43%	26.82%
ZTD-RF1	-0.08	3.18	4.34	5.54%	1.89%	4.99%
ZTD-RF2	-0.08	3.05	4.21	8.41%	4.88%	7.88%
ZTD-RF3	-0.11	2.25	3.28	28.62%	25.87%	28.21%
ZTD-RF4	-0.10	2.24	3.24	29.47%	26.75%	29.07%
Saastamoinen	-1.71	3.30	4.60	-	-	-
A&N	-1.32	3.17	4.43	-	-	-
GPT3	-0.29	3.37	4.57	-	-	-

better than the traditional models. It should be noted that the ZTD-RF1 model displays significantly irregular fluctuations at the Graphical end of the density scatters. These irregular fluctuations are attributed to models trained using RF are mathematically expressed as average of segmented functions. Furthermore, the density scatters of the ZTD-CNN1, ZTD-CNN2, and ZTD-RF2 models exhibited finer shapes compared to those of the ZTD-RF1 and GPT3 models, particularly in regions with higher data density, indicating superior error stability. So we continue to recommend employing the ZTD-CNN2 and ZTD-RF2 models when surface meteorological data is not available. Based on the R^2 values and the density scatters of the ZTD-CNN3, ZTD-CNN4, ZTD-RF3, ZTD-RF4, Saastamoinen and A&N models with requiring surface meteorological data, it was evident that the ZTD-CNN3, ZTD-CNN4, ZTD-RF3 and ZTD-RF4 models were more fitted than the Saastamoinen and A&N models. Additionally, there were some noticeable scatters situated above the density scatters for the Saastamoinen and A&N models, as showed in Fig. 5. The ZTD-CNN4 and ZTD-RF4 models, in contrast, effectively prevented the emergence of noticeable scatters. Consequently, they outperformed the Saastamoinen and A&N models in terms of error stability. Combining their error stability and RMSE, we still recommend adopting the ZTD-CNN4 and ZTD-RF4 models when surface meteorological data is available. It is important to acknowledge that Fig. 5 illustrates a scatters fault resulting from the inadequate densification of stations in the 2019 IGRA radiosonde data.

B. SPATIAL PERFORMANCE AND ANALYSIS

In order to show the spatial performance of these new models, we calculate the site-wise RMSE of each station. Table 3 presents the minimum RMSE (Min), maximum RMSE (Max) and mean RMSE (Mean) value at 202 stations in Chinese mainland and surrounding areas during 2019. From Table 3, it can be observed that for models without requiring surafce meteorological data, as demonstrated by the Max or Mean value, the ZTD-CNN1, ZTD-CNN2, ZTD-RF1, and ZTD-RF2 models exhibited greater stability than the GPT3 model. In particular, the ZTD-CNN1 and ZTD-CNN2 models exhibited a reduction of the Max value compared to the ZTD-RF1 and ZTD-RF2 models, suggesting that the ZTD-CNN1 and ZTD-CNN2 models effectively controlled the ZTD errors within a more optimal range. These suggested an overall improvement in the effectiveness of the models for regional applications. Therefore, in the absence of surface meteorological data for estimating ZTD in the Chinese mainland and surrounding areas, we recommend the ZTD-CNN2 model. While our previous recommendation include both the ZTD-CNN2 and ZTD-RF2 models based on the overall assessment, considering the stability of the spatial RMSE distribution, we suggest prioritizing the ZTD-CNN2 model over the ZTD-RF2 model. Note that ZTD-CNN1, ZTD-CNN2, ZTD-RF1, and ZTD-RF2 are models without requiring surface meteorological data. It is acceptable that traditional models with requiring surface meteorological data might outperform our models without requiring surface meteorological data. Hence we might observe negative improvement in Table 3.

Among the models with requiring surface meteorological data, the ZTD-CNN4 and ZTD-RF4 models exhibite improvements of 21.86% and 24.45%, respectively, compared to the A&N model, and meanwhile, the ZTD-RF4 model is slightly better than ZTD-RF3 model. Given the significant improvements in the accuracy of the ZTD-CNN4 and ZTD-RF4 models compared to the traditional models, and the comparable stability of the ZTD estimation between the ZTD-CNN4 model and the ZTD-RF4 model, we continue to recommend their utilization for ZTD estimation in Chinese mainland and surrounding areas when surface meteorological data is available.

Fig. 6, Fig. 7 and Fig. 8 display the site-wise RMSE, ME and MAE of different models. It is evident that our six new ZTD models and the ZTD-RF1 and ZTD-RF3 models

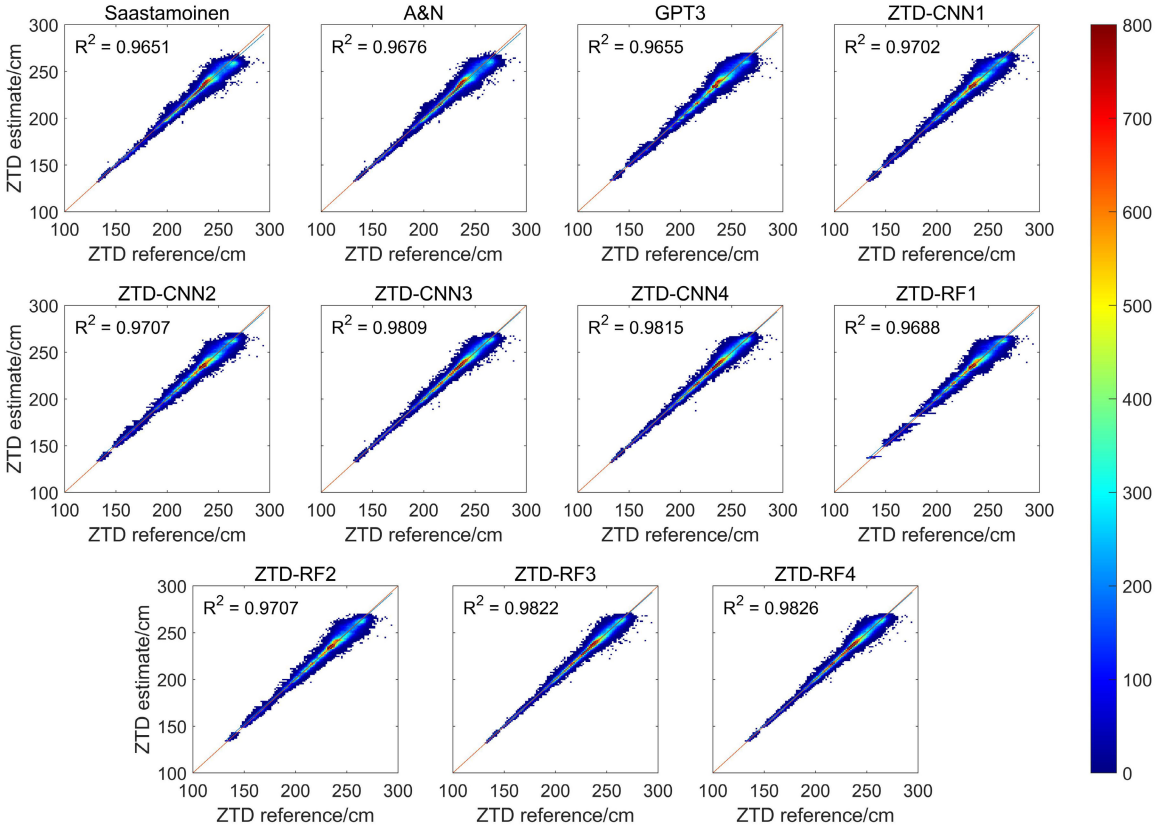

FIGURE 5. Density scatters of ZTD estimates using new models against the "exact" ZTD for the 202 stations analyzed during 2019 (ZTD reference denotes the "exact" ZTD, the color bar on the right represents the number of ZTD in equi-spaced range of values).

TABLE 3. Spatial RMSE and improvement rates (the mean RMSE) of new ZTD models compared to traditional models, for 202 stations analyzed during 2019 (Min, Max and Mean denotes the minimum, maximum and mean RMSE for 202 stations).

Models	Min/cm	Max/cm	Mean/cm	Improvement rate compared to Saastamoinen	Improvement rate compared to A&N	Improvement rate compared to GPT3
ZTD-CNN1	1.38	8.79	4.16	3.29%	-0.96%	8.29%
ZTD-CNN2	1.34	8.59	4.12	4.16%	-0.06%	9.11%
ZTD-CNN3	1.09	6.92	3.29	23.37%	20.00%	27.33%
ZTD-CNN4	0.91	6.86	3.22	25.15%	21.86%	29.02%
ZTD-RF1	1.76	9.15	4.25	1.01%	-3.35%	6.12%
ZTD-RF2	1.33	9.21	4.11	4.43%	0.22%	9.37%
ZTD-RF3	0.72	6.49	3.15	26.77%	23.54%	30.55%
ZTD-RF4	0.81	6.53	3.11	27.63%	24.45%	31.37%
Saastamoinen	0.87	12.15	4.30	-	-	-
A&N	0.90	11.16	4.12	-	-	-
GPT3	1.37	10.35	4.53	-	-	-

outperform three traditional models in all three metrics of RMSE, ME and MAE. When surface meteorological data is unavailable, the ZTD-CNN1, ZTD-CNN2, ZTD-RF1 and ZTD-RF2 models could be more effectively utilized than GPT3 in the test area, concerning the spatial distribution of RMSE and MAE. These four models fit especially well in the northern and southwestern areas of Chinese mainland, Mongolia, northern area of Japan and Russian regions near Mongolia. Meanwhile, their accuracy in Southeast China is comparable to the GPT3 model. It is noteworthy that, when applied in Southern India, the GPT3 model's effectiveness is inadequate. The ZTD-CNN1, ZTD-CNN2, ZTD-RF1, and ZTD-RF2 models notably enhanced the accuracy of ZTD estimation in Southern India. Furthermore, with regards to the ME, the GPT3 model exhibited poor performance in Japan, while the ZTD-CNN1, ZTD-CNN2, ZTD-RF1 and ZTD-

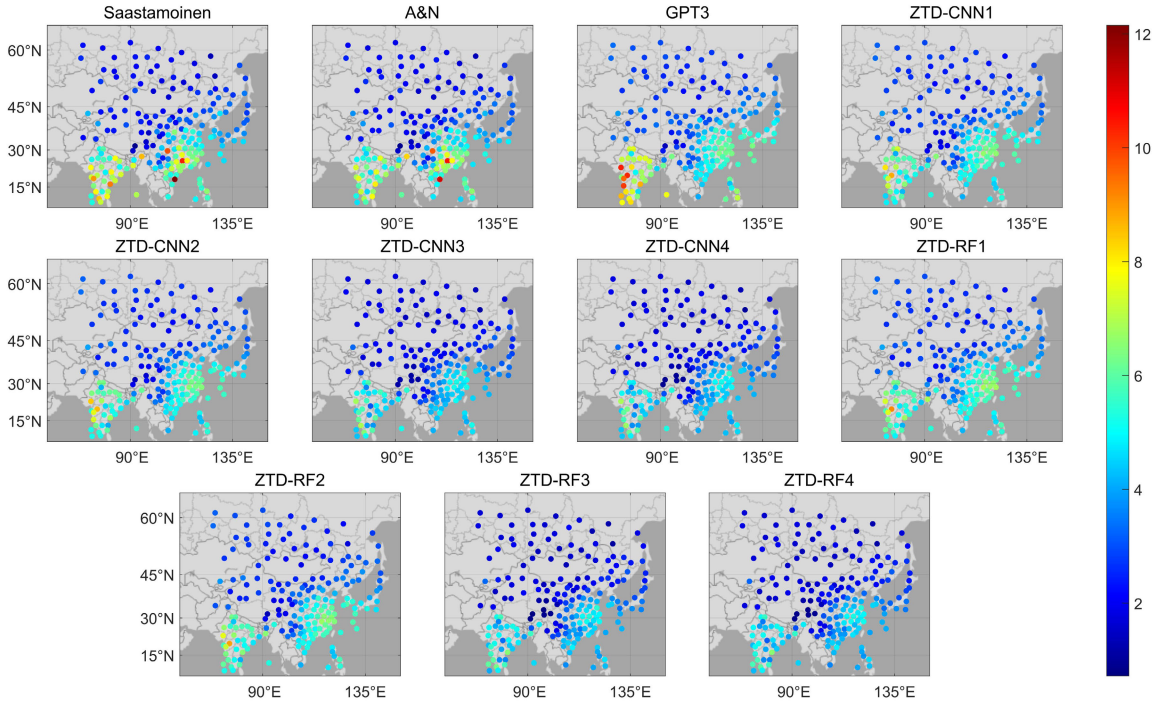

FIGURE 6. Site-wise RMSE (unit: cm) of different models for the ZTD estimation at the 202 stations analyzed during 2019.

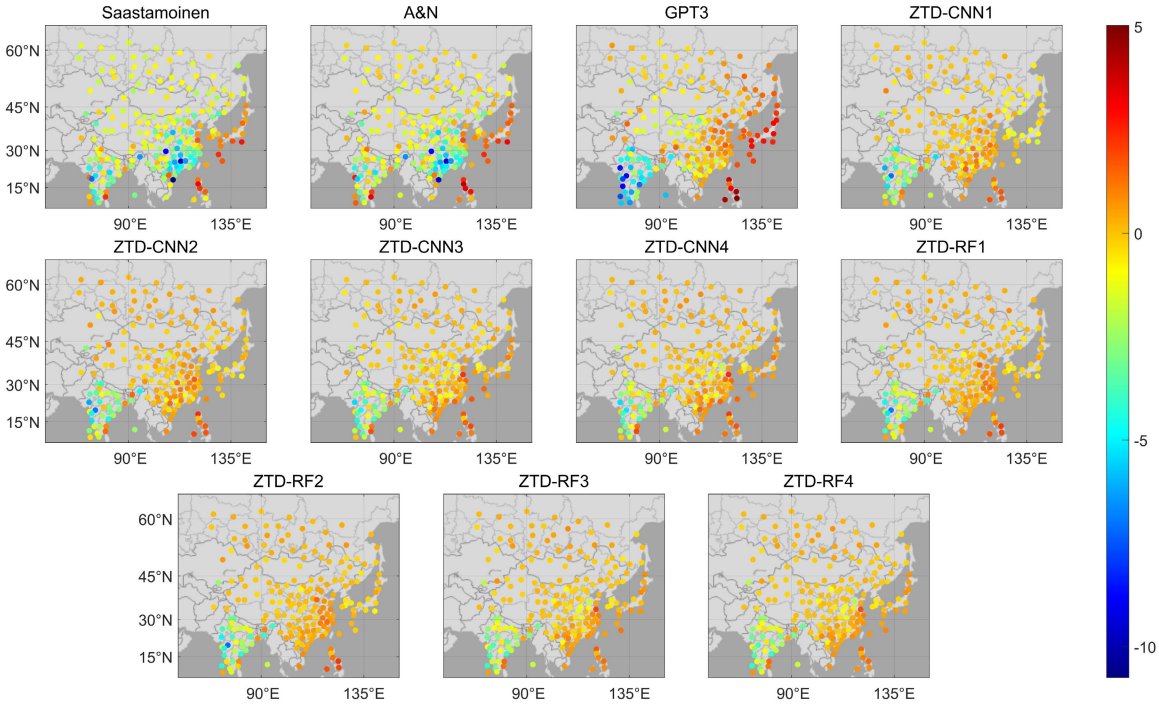

FIGURE 7. Site-wise ME (unit: cm) of different models for the ZTD estimation at the 202 stations analyzed during 2019.

RF2 models had undergone significant enhancements. When surface meteorological data was available, the ZTD-CNN3, ZTD-CNN4, ZTD-RF3 and ZTD-RF4 models obtained better accuracy compared to the Saastamoinen and A&N models at the 202 stations. In particular, compared with the traditional models, the ZTD-CNN3, ZTD-CNN4 ZTD-RF3 and ZTD-RF4 models had greatly enhanced the accuracy of the ZTD estimation in Southeast China, Japan

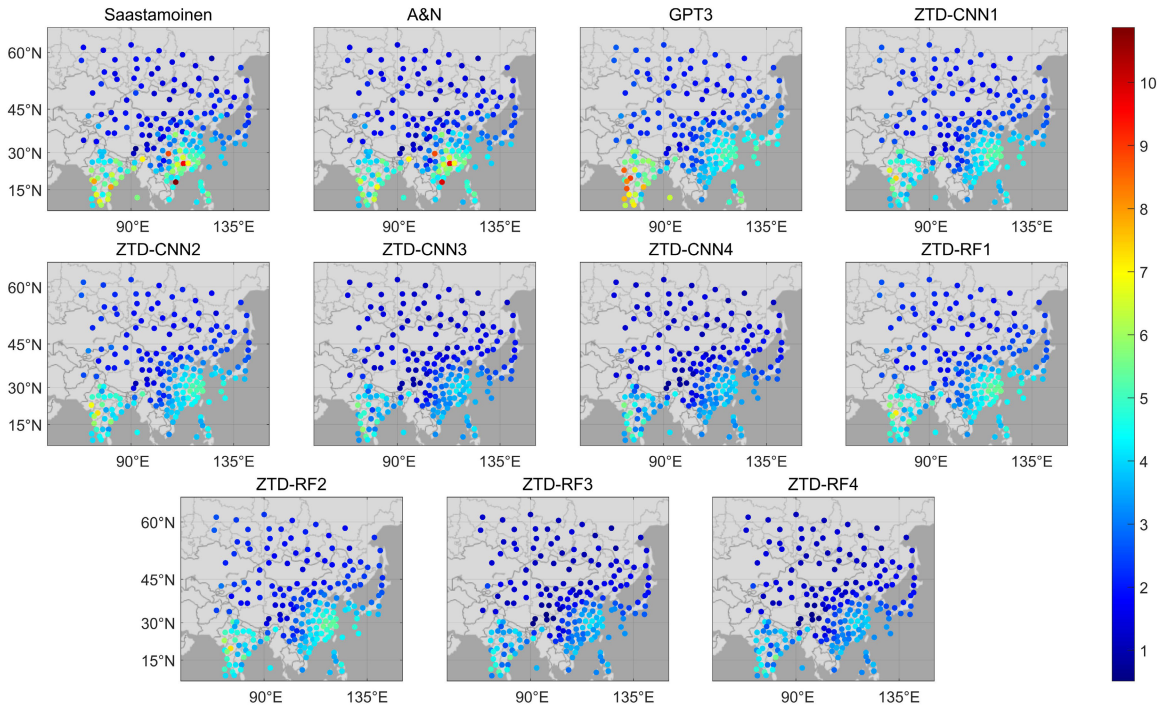

FIGURE 8. Site-wise MAE (unit: cm) of different models for the ZTD estimation at the 202 stations analyzed during 2019.

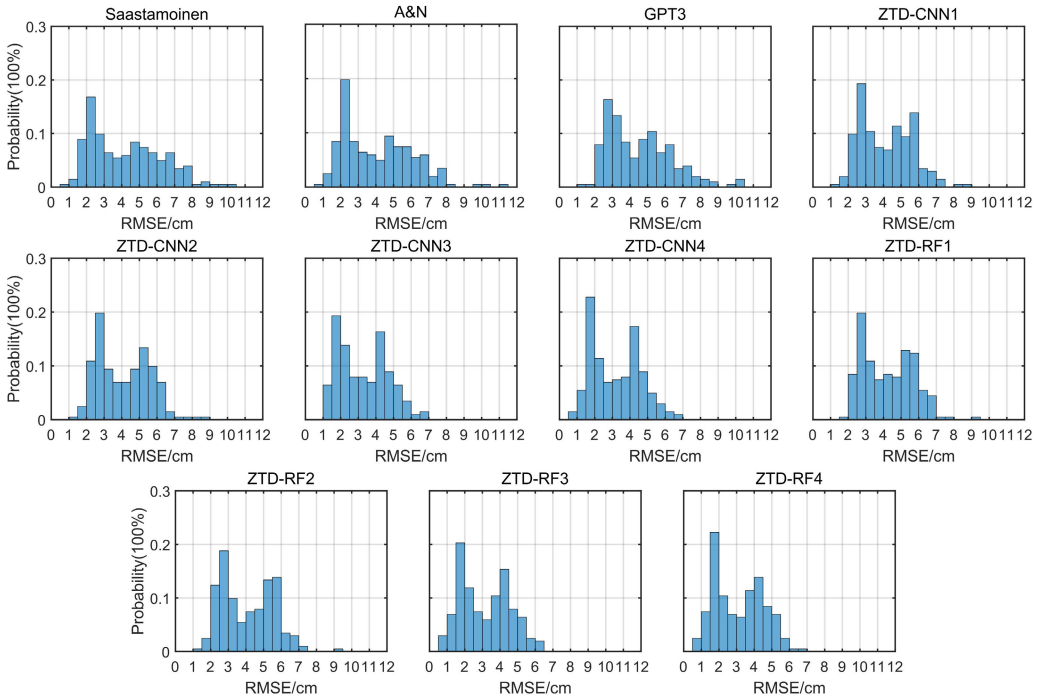

FIGURE 9. The RMSE distribution for 202 stations (Probability denotes the percentage of stations in a given RMSE range, e.g. when using the ZTD-RF1 model, 40 stations have the RMSEs between 2.5 cm and 3 cm, so the Probability value of 2.5 cm - 3 cm is 40/202 = 0.1980198 = 19.80198%).

and India. After comparing the new ZTD models with requiring meteorological data to those without, it is evident that the incorporation of meteorological parameters significantly improved the overall accuracy and stability of the models in estimating ZTD in Chinese mainland and surrounding areas.

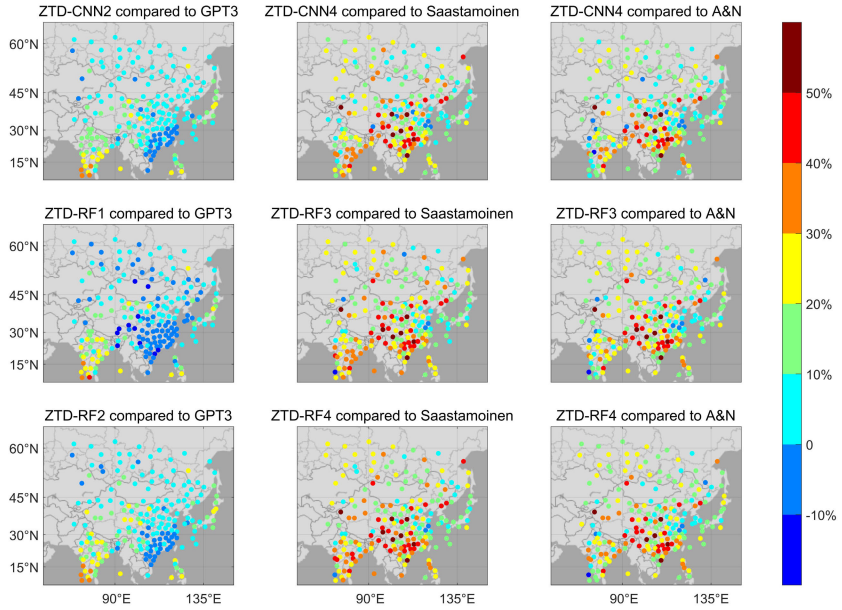

FIGURE 10. Improvement rates in RMSE for the new model compared to the traditional model for the 202 stations analysed in 2019.

TABLE 4. Percentages of stations with different ranges of improvement rates among the 202 stations during 2019 (the percentage range in the first row represents the range of improvement rates).

Comparison schemes	Less than -20%	-20%-0%	0%-20%	20%-40%	40%-60%	More than 60%	Maximum improvement rates
ZTD-CNN2 compared to GPT3	0.00%	15.84%	73.27%	10.89%	0.00%	0.00%	39.86%
ZTD-RF1 compared to GPT3	2.97%	38.61%	46.53%	11.39%	0.50%	0.00%	40.38%
ZTD-RF2 compared to GPT3	0.00%	17.82%	66.83%	15.35%	0.00%	0.00%	37.81%
ZTD-CNN4 compared to Saastamoinen	0.00%	2.97%	45.54%	38.61%	10.89%	1.98%	63.83%
ZTD-RF3 compared to Saastamoinen	0.00%	2.48%	39.60%	43.07%	13.37%	1.49%	66.64%
ZTD-RF4 compared to Saastamoinen	0.00%	1.49%	39.60%	42.08%	15.35%	1.49%	66.75%
ZTD-CNN4 compared to A&N	0.00%	5.45%	54.46%	30.69%	8.42%	0.99%	64.28%
ZTD-RF3 compared to A&N	0.00%	3.96%	47.52%	35.64%	11.39%	1.49%	67.05%
ZTD-RF4 compared to A&N	0.00%	2.48%	49.01%	36.14%	10.89%	1.49%	67.16%

Fig. 9 presents the histogram distribution of the site-wise RMSE for 202 stations in the Chinese mainland and surrounding areas during 2019. The figure clearly demonstrated that when the GPT3 and ZTD-RF1 models are used, the RMSE distribution at the 202 stations is more dispersed compared to the ZTD-CNN1, ZTD-CNN2, and ZTD-RF2 models. Across the 202 stations, the RMSEs of the ZTD-CNN1, ZTD-CNN2, and ZTD-RF2 models are overall smaller than that of the GPT3 and ZTD-RF1 models, especially for the ZTD-CNN2 and ZTD-RF2 models. This observation indicated that the ZTD-CNN2 and ZTD-RF2 models could enhance the accuracy and stability of ZTD calculation in comparison to the GPT3 and ZTD-RF1 models. When the ZTD-CNN3, ZTD-CNN4, ZTD-RF3 and ZTD-RF4 models were used for the 202 stations during 2019, most stations exhibited RMSEs of 1.5 cm - 6 cm, representing a reduction compared to the Saastamoinen and A&N models. Therefore, when compared with the Saastamoinen and A&N models requiring surface meteorological data, the ZTD-CNN3, ZTD-CNN4 ZTD-RF3 and ZTD-RF4 models demonstrated superior performance. Generally, the distribution of RMSEs across the 202 stations exhibits greater dispersion when employing the traditional models (Saastamoinen, A&N, and GPT3 models). In contrast, the RMSEs of our proposed new models (ZTD-CNN1, ZTD-CNN2, ZTD-CNN3, ZTD-CNN4, ZTD-RF2, and ZTD-RF4 models) at the 202 stations were mainly concentrated within a smaller range. This demonstrated that our new models are capable

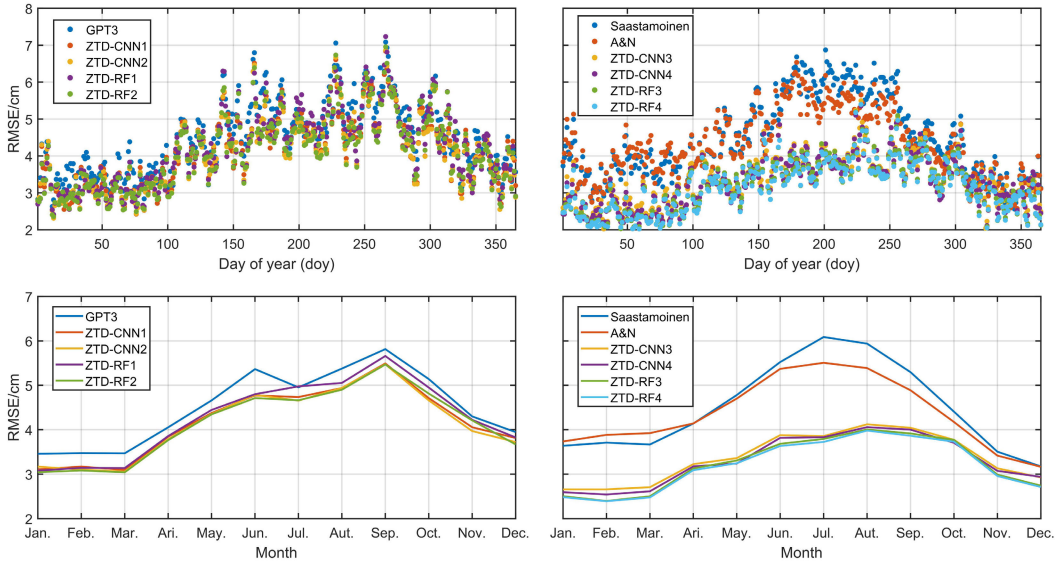

FIGURE 11. Daily and monthly variations in terms of RMSE for different models for the ZTD estimation in 2019.

of maintaining RMSE within a more favorable range at the 202 stations.

Fig. 10 illustrates the spatial distribution of improvement rates in RMSE for the ZTD-CNN2, ZTD-CNN4, ZTD-RF1, ZTD-RF2, ZTD-RF3 and ZTD-RF4 models compared to the traditional models at 202 stations in Chinese mainland and surrounding areas during 2019. The ZTD-RF1 performed worse than the GPT3 model in almost half of the region, which is not ideal. However, It was evident that the ZTD-CNN2 and ZTD-RF2 models exhibited an overall improvement compared to the GPT3 model in various regions, except for Southeastern China. ZTD-CNN2 and ZTD-RF2 models are significantly better than GPT3 models in most regions, but because surface meteorological data were not introduced, the models cannot reproduce the complex changes of atmospheric water vapor in the East Asian monsoon climate types, the forecast accuracy is decreased in Southeast China. Furthermore, evidently, the introduction of surface meteorological data for ZTD modeling greatly improves the accuracy of ZTD forecasting in Southeastern China. the ZTD-CNN3, ZTD-CNN4, ZTD-RF3 and ZTD-RF4 models exhibited almost an overall improvement compared to the Saastamoinen and A&N model in the test regions.

Table 4 presents the detailed distribution of improvement rates for the ZTD-CNN2, ZTD-CNN4, ZTD-RF1, ZTD-RF2, ZTD-RF3 and ZTD-RF4 models. Among the 202 stations, the ZTD-CNN2 and ZTD-RF2 models exhibited higher accuracy than the GPT3 model at 84.16% and 82.18% of the stations, repectively, and significantly outperformed the ZTD-RF1 model (58.42%). For the ZTD-CNN4, ZTD-RF3 and ZTD-RF4 models, 97.03%, 97.52% and 98.51% of the stations, respectively, demonstrated higher accuracy than the Saastamoinen model among the 202 stations. In addition, the

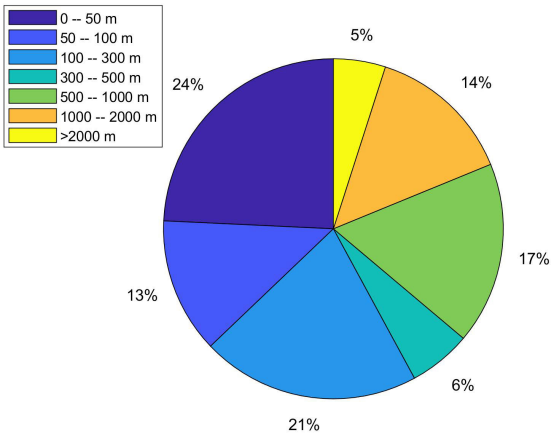

FIGURE 12. Percentages of stations within given altitude intervals.

ZTD-CNN4, ZTD-RF3 and ZTD-RF4 models achieved maximum improvement rates of 63.83%, 66.64% and 66.75%, respectively, compared to the Saastamoinen model. When compared to the A&N model, the ZTD-CNN4 ZTD-RF3 and ZTD-RF4 models outperformed it at 94.55%, 96.04% and 97.52% of the stations, respectively, with maximum improvement rates of 64.28%, 67.05% and 67.16%.

C. TEMPORAL PERFORMANCE AND ANALYSIS

Previous studies have demonstrated that ZTD exhibits pronounced seasonal variations [35], [36]. Therefore, in order to further explore the performance of the ZTD models involved in this study across time, we calculated the daily and monthly RMSEs for all models in 2019. The results are shown in Fig. 11.

As illustrated in Fig. 11, for models without requiring surface meteorological data, their daily accuracy in terms

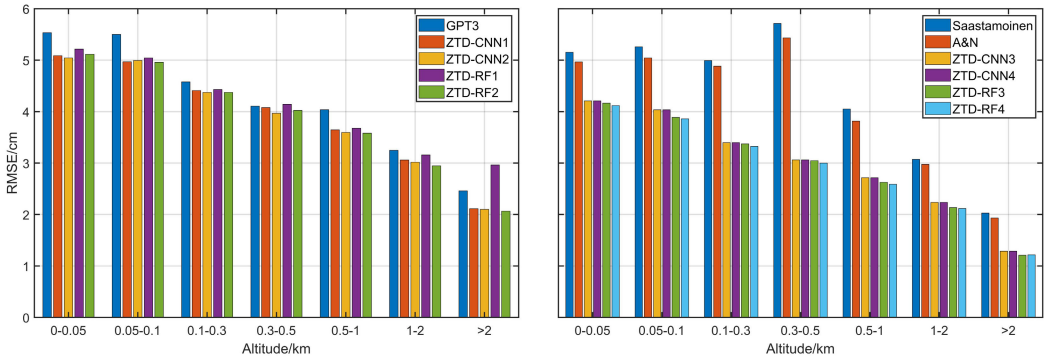

FIGURE 13. Altitudinal variations in terms of RMSE for different models for the ZTD estimation in 2019.

of RMSE mostly lies between 2-7cm. Further focusing on the monthly accuracy in terms of RMSE, we will see that they lies between 3-6cm, with ZTD-CNN1, ZTD-CNN2 and ZTD-RF2 models being more accurate compared to GPT3 and ZTD-RF1 models. For models with requiring surface meteorological data, we can find that the ZTD-CNN3, ZTD-CNN4, ZTD-RF3 and ZTD-RF4 (daily RMSE: 2-5cm, monthly RMSE: 2-4cm) models significantly outperform the Saastamoinen (daily RMSE: 2-7cm, monthly RMSE: 3-6cm) and A&N (daily RMSE:2-6.5cm,monthly RMSE:3-5.5cm) models, especially in summer. Furthermore, it is important to highlight that the estimation accuracy of all models is influenced by temporal factors. It can be observed that the models perform the worst in summer (June to August) and the best in winter (December to February). This is due to the fact that the water vapour content and dynamics are much greater in the summer than in the winter in the region under study. For models that do not require surface meteorological data, the ZTD-CNN2, ZTD-CNN1, and ZTD-RF2 models demonstrate superior performance in capturing the cyclic variations of ZTD over time in comparison to the GPT3 model. In the case of models requiring surface meteorological data, the ZTD-RF4 models offer a most comprehensive representation of ZTD, particularly in relation to rapid fluctuations in water vapour pressure, in comparison to the Saastamoinen, A&N, ZTD-CNN3, ZTD-CNN4, ZTD-RF3 models.

D. ALTITUDINAL PERFORMANCE AND ANALYSIS

Subsequently, as altitude has a significant effect on ZTD changes [37], [38], the relationship between the performance of the ZTD models included in this study and altitude was further investigated., and given that our test stations were distributed in an uneven manner across the altitude, we devised seven altitude intervals and calculated the RMSEs within each of these intervals for the purpose of exploration. Percentage of stations in the given seven altitude intervals is shown in Fig. 12.

From Fig. 13, it can be observed that for the models without requiring meteorological data, their estimation accuracies increase with altitude. The two GPT3 compensated models (ZTD-CNN2, ZTD-RF2) demonstrate the most accurate performance, outperforming the GPT3 and ZTD-RF1. However, the direct model ZTD-RF1 is less accurate than the GPT3 model in certain altitude intervals (>2 km and 0.3-0.5km). For models requiring meteorological data, the estimation accuracies of the ZTD-CNN4 and ZTD-RF4 models demonstrated an increase with increasing altitude. Furthermore, their estimation accuracies were found to be significantly superior to those of the Saastamoinen and A&N models, meanwhile, slightly superior to those of the ZTD-RF3 model. It is important to note that the estimation accuracies of the Saastamoinen and A&N models are not proportional to altitude at lower altitudes ($\leq$0.5 km). This is due to the fact that the low-altitude areas utilised in this study are predominantly situated in Southeastern China and India, which are characterised by East Asian and Indian monsoon climates with high water vapour content and active weather patterns. However, the Saastamoinen and A&N models are unable to adequately capture the intricate variations in ZTD.

V. CONCLUSION

In this study, we utilized convolutional neural network (CNN) and random forest (RF) to train two types of zenith tropospheric delay (ZTD) models: the CNN/RF-based direct models and the CNN/RF-based compensated models. Six new models were trained using the radiosonde data (2013-2018) provided by IGRA for 205 stations located in Chinese mainland and surrounding areas. The validation of these ZTD models was performed using radiosonde data (2019) for 202 stations situated in Chinese mainland and surrounding areas. Comparative analyses were then performed with the Saastamoinen, A&N, GPT3, ZTD-RF1 and ZTD-RF3 models to evaluate the effectiveness of these approaches. Generally, our new CNN/RF-based ZTD models demonstrated improvement compared to traditional ZTD models and existing RF-based ZTD models.

For models without requiring surface meteorological data, the optimal model is the CNN-based compensation model (ZTD-CNN2), which exhibits an overall RMSE of 4.21 cm, indicating a 7.89% reduction compared to the GPT3 model

(4.57 cm) and significantly superior to ZTD-RF1 model (4.34 cm). Furthermore, the ZTD-CNN2 model demonstrated superior error stability compared to the GPT3 and ZTD-RF1 models. Spatially, the ZTD-CNN2 model itself exhibited higher accuracy in estimating ZTD in the northern and southwestern areas of the Chinese mainland, as compared to the southeastern area. Additionally, the ZTD-CNN2 model effectively enhanced the accuracy of ZTD estimation in the northern and southwestern areas of Chinese mainland, India and Japan. Moreover, it successfully addressed the issue of excessive ME in the ZTD estimation in Japan.

With the introduction of surface meteorological data, based on a more accurate portrayal of water vapour pressure, both the CNN-based compensation model (ZTD-CNN4) and the RF-based compensation model (ZTD-RF4) demonstrated superior performance compared to the Saastamoinen model and the A&N model. The ZTD-RF4 model achieved an overall RMSE of 3.24 cm, representing a 29.47% reduction of the RMSE compared to the Saastamoinen model (4.60 cm) and a 26.75% reduction compared to the A&N model (4.43 cm). Additionally, the ZTD-RF4 model reached a mean RMSE of 3.11 cm, indicating a 27.63% reduction of the mean RMSE compared to the Saastamoinen model (4.30 cm) and a 24.45% reduction compared to the A&N model (4.12 cm). Compared with the Saastamoinen and A&N models, the ZTD-CNN4 and ZTD-RF4 models exhibited significantly better performance in terms of the accuracy and error stability. Spatially, the ZTD-CNN4 and ZTD-RF4 models demonstrated excellent accuracy across the Chinese mainland and surrounding areas. Particularly noteworthy was their ability to overcome the limitations of traditional models, which often resulted in poor accuracy in ZTD estimation for regions such as Southeastern China and India with monsoon climate types.

ACKNOWLEDGMENT

(Jiahao Zhang, Qin Liang, and Yunqing Huang are co-first authors.)

REFERENCES

[1] M. Bevis, S. Businger, T. A. Herring, C. Rocken, R. A. Anthes, and R. H. Ware, "GPS meteorology: Remote sensing of atmospheric water vapor using the global positioning system," *J. Geophys. Res., Atmos.*, vol. 97, no. 14, pp. 15787–15801, Oct. 1992, doi: 10.1029/92jd01517.

[2] C. Lu, X. Li, F. Zus, R. Heinkelmann, G. Dick, M. Ge, J. Wickert, and H. Schuh, "Improving BeiDou real-time precise point positioning with numerical weather models," *J. Geodesy*, vol. 91, no. 9, pp. 1019–1029, Feb. 2017, doi: 10.1007/s00190-017-1005-2.

[3] H. S. Hopfield, "Two-quartic tropospheric refractivity profile for correcting satellite data," *J. Geophys. Res.*, vol. 74, no. 18, pp. 4487–4499, Aug. 1969, doi: 10.1029/jc074i018p04487.

[4] J. Saastamoinen, "Atmospheric correction for the troposphere and stratosphere in radio ranging satellites," *Use Artif. Satell. Geod.*, vol. 15, pp. 247–251, Jan. 1972, doi: 10.1029/GM015p0247.

[5] J. Askne and H. Nordius, "Estimation of tropospheric delay for microwaves from surface weather data," *Radio Sci.*, vol. 22, no. 3, pp. 379–386, May 1987, doi: 10.1029/RS022i003p00379.

[6] J. Boehm, R. Heinkelmann, and H. Schuh, "Short note: A global model of pressure and temperature for geodetic applications," *J. Geodesy*, vol. 81, no. 10, pp. 679–683, Feb. 2007, doi: 10.1007/s00190-007-0135-3.

[7] K. Lagler, M. Schindelegger, J. Böhm, H. Krásná, and T. Nilsson, "GPT2: Empirical slant delay model for radio space geodetic techniques," *Geophys. Res. Lett.*, vol. 40, no. 6, pp. 1069–1073, Mar. 2013, doi: 10.1002/grl.50288.

[8] J. Böhm, G. Möller, M. Schindelegger, G. Pain, and R. Weber, "Development of an improved empirical model for slant delays in the troposphere (GPT2w)," *GPS Solutions*, vol. 19, no. 3, pp. 433–441, Jul. 2015, doi: 10.1007/s10291-014-0403-7.

[9] D. Landskron and J. Böhm, "VMF3/GPT3: Refined discrete and empirical troposphere mapping functions," *J. Geodesy*, vol. 92, no. 4, pp. 349–360, Sep. 2017, doi: 10.1007/s00190-017-1066-2.

[10] J. A. Weyn, D. R. Durran, R. Caruana, and N. Cresswell-Clay, "Sub-Seasonal forecasting with a large ensemble of deep-learning weather prediction models," *J. Adv. Model. Earth Syst.*, vol. 13, no. 7, Jul. 2021, Art. no. e2021MS002502, doi: 10.1029/2021ms002502.

[11] A. V. Zhukov, Y. V. Yasyukevich, and A. E. Bykov, "GIMLi: Global ionospheric total electron content model based on machine learning," *GPS Solutions*, vol. 25, no. 1, p. 19, Nov. 2020, doi: 10.1007/s10291-020-01055-1.

[12] G. Xiao, J. Ou, G. Liu, and H. Zhang, "Construction of a regional precise tropospheric delay model based on improved BP neural network," *Chin. J. Geophys.-Chin. Ed.*, vol. 61, no. 8, pp. 3139–3148, 2018, doi: 10.6038/cjg2018L0565.

[13] C. Lu, Y. Zheng, Z. Wu, Y. Zhang, Q. Wang, Z. Wang, Y. Liu, and Y. Zhong, "TropNet: A deep spatiotemporal neural network for tropospheric delay modeling and forecasting," *J. Geodesy*, vol. 97, no. 4, p. 34, Mar. 2023, doi: 10.1007/s00190-023-01722-4.

[14] Q. Li, L. Yuan, and Z. Jiang, "Modeling tropospheric zenith wet delays in the Chinese Mainland based on machine learning," *GPS Solutions*, vol. 27, no. 4, p. 171, Oct. 2023, doi: 10.1007/s10291-023-01507-4.

[15] S. Li, N. Jiang, T. Xu, Y. Xu, H. Yang, Z. Zhang, A. Guo, and Y. Wu, "A precipitation forecast model with a neural network and improved GPT3 model for Japan," *GPS Solutions*, vol. 27, no. 4, p. 186, Aug. 2023, doi: 10.1007/s10291-023-01526-1.

[16] J. Li, Q. Zhang, L. Liu, Y. Yao, L. Huang, F. Chen, L. Zhou, and B. Zhang, "A refined zenith tropospheric delay model for Mainland China based on the global pressure and temperature 3 (GPT3) model and random forest," *GPS Solutions*, vol. 27, no. 4, p. 172, Oct. 2023, doi: 10.1007/s10291-023-01513-6.

[17] L. Crocetti, M. Schartner, F. Zus, W. Zhang, G. Moeller, V. Navarro, L. See, K. Schindler, and B. Soja, "Global, spatially explicit modelling of zenith wet delay with XGBoost," *J. Geodesy*, vol. 98, no. 4, p. 23, Apr. 2024, doi: 10.1007/s00190-024-01829-2.

[18] Q. Li, J. Böhm, L. Yuan, and R. Weber, "Global zenith wet delay modeling with surface meteorological data and machine learning," *GPS Solutions*, vol. 28, no. 1, p. 57, Jan. 2024, doi: 10.1007/s10291-023-01595-2.

[19] Q. Li, L. Yuan, P. Chen, and Z. Jiang, "Global grid-based T_m model with vertical adjustment for GNSS precipitable water retrieval," *GPS Solutions*, vol. 24, no. 3, p. 73, May 2020, doi: 10.1007/s10291-020-00988-x.

[20] T. Nilsson, J. Boehm, D. D. Wijaya, A. Tresch, V. Nafisi, and H. Schuh, "Path delays in the neutral atmosphere," in *Atmospheric Effects in Space Geodesy*. Berlin, Germany: Springer, Jan. 2013, pp. 73–136, doi: 10.1007/978-3-642-36932-2_3.

[21] H. W. Janes, R. B. Langley, and S. P. Newby, "Analysis of tropospheric delay prediction models: Comparisons with ray-tracing and implications for GPS relative positioning," *Bull. Géodésique*, vol. 65, no. 3, pp. 151–161, Sep. 1991, doi: 10.1007/bf00806344.

[22] V. B. Mendes and R. B. Langley, "Tropospheric zenith delay prediction accuracy for high-precision GPS positioning and navigation," *Navigation*, vol. 46, no. 1, pp. 25–34, Mar. 1999, doi: 10.1002/j.2161-4296.1999.tb02393.x.

[23] X. Wang, K. Zhang, S. Wu, S. Fan, and Y. Cheng, "Water vapor-weighted mean temperature and its impact on the determination of precipitable water vapor and its linear trend," *J. Geophys. Res., Atmos.*, vol. 121, no. 2, pp. 833–852, Jan. 2016, doi: 10.1002/2015jd024181.

[24] W. Zhang, Y. Lou, J. Huang, F. Zheng, Y. Cao, H. Liang, C. Shi, and J. Liu, "Multiscale variations of precipitable water over China based on 1999–2015 ground-based GPS observations and evaluations of reanalysis products," *J. Climate*, vol. 31, no. 3, pp. 945–962, Jan. 2018, doi: 10.1175/jcli-d-17-0419.1.

[25] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.

[26] M. Reichstein, G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, and N. Carvalhais, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, Feb. 2019, doi: 10.1038/s41586-019-0912-1.
[27] H. Bi, L. Huang, H. Zhang, S. Xie, L. Zhou, and L. Liu, "A deep learning-based model for tropospheric wet delay prediction based on multi-layer 1D convolution neural network," *Adv. Space Res.*, vol. 73, no. 10, pp. 5031–5042, May 2024, doi: 10.1016/j.asr.2024.02.039.
[28] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Int. Conf. Mach. Learn.*, Haifa, Israel, 2010, pp. 807–814.
[29] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," Feb. 2015, *arXiv:1502.03167*.
[30] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *Proc. 30th Int. Conf. Int. Conf. Mach. Learn.*, Atlanta, GA, USA, 2013, pp. 1139–1147.
[31] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/a:1010933404324.
[32] P. Kontschieder, S. R. Bulo, H. Bischof, and M. Pelillo, "Structured class-labels in random forests for semantic image labelling," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 2190–2197, doi: 10.1109/iccv.2011.6126496.
[33] R. A. Berk, "Classification and regression trees (CART)," in *Statistical Learning from a Regression Perspective*. Berlin, Germany: Springer, Jan. 2008, pp. 1–65, doi: 10.1007/978-0-387-77501-2_3.
[34] A. Criminisi, "Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning," *Found. Trends Comput. Graph. Vis.*, vol. 7, nos. 2–3, pp. 81–227, 2011, doi: 10.1561/0600000035.
[35] S. Jin, J. Park, J. Cho, and P. Park, "Seasonal variability of GPS-derived zenith tropospheric delay (1994–2006) and climate implications," *J. Geophys. Res., Atmos.*, vol. 112, no. D9, Jun. 2007, doi: 10.1029/2006jd007772.
[36] L. He, Y. Yao, C. Xu, H. Zhang, F. Tang, C. Ji, Z. Liu, and W. Wu, "A new global ZTD forecast model based on improved LSTM neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 9606–9614, 2024, doi: 10.1109/JSTARS.2024.3391821.
[37] G. Zhu, L. Huang, Y. Yang, J. Li, L. Zhou, and L. Liu, "Refining the ERA5-based global model for vertical adjustment of zenith tropospheric delay," *Satell. Navigat.*, vol. 3, no. 1, p. 27, Nov. 2022, doi: 10.1186/s43020-022-00088-w.
[38] L. Huang, G. Zhu, H. Peng, L. Liu, C. Ren, and W. Jiang, "An improved global grid model for calibrating zenith tropospheric delay for GNSS applications," *GPS Solutions*, vol. 27, no. 1, p. 17, Jan. 2023, doi: 10.1007/s10291-022-01354-9.

JIAHAO ZHANG received the M.S. degree in management science from Xi'an University of Architecture and Technology, Xi'an, China, in 2019. He is currently pursuing the Ph.D. degree with Xiangtan University.

His main research interest includes application of mathematical methods to precise point positioning (PPP).

QIN LIANG received the Ph.D. degree in computational mathematics from Peking University, Beijing, China.

Her main research interest includes application of mathematical methods to precise point positioning (PPP).

YUNQING HUANG received the Ph.D. degree in computational mathematics from the Institute of Systems Science, Chinese Academy of Sciences.

His main research interest includes application of mathematical methods to precise point positioning (PPP).