



Article

Enhanced Neural Network Model for Worldwide Estimation of Weighted Mean Temperature

Fengyang Long ¹, Chengfa Gao ^{1,*}, Yuxiang Yan ¹ and Jinling Wang ²

¹ School of Transportation, Southeast University, Nanjing 211189, China; fylong1990@seu.edu.cn (F.L.); geodesic_y_yan@seu.edu.cn (Y.Y.)

² School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia; jinling.wang@unsw.edu.au

* Correspondence: gaochfa@seu.edu.cn

Abstract: Precise modeling of weighted mean temperature (T_m) is critical for realizing real-time conversion from zenith wet delay (ZWD) to precipitation water vapor (PWV) in Global Navigation Satellite System (GNSS) meteorology applications. The empirical T_m models developed by neural network techniques have been proved to have better performances on the global scale; they also have fewer model parameters and are thus easy to operate. This paper aims to further deepen the research of T_m modeling with the neural network, and expand the application scope of T_m models and provide global users with more solutions for the real-time acquisition of T_m . An enhanced neural network T_m model (ENNTm) has been developed with the radiosonde data distributed globally. Compared with other empirical models, the ENNTm has some advanced features in both model design and model performance. Firstly, the data for modeling cover the whole troposphere rather than just near the Earth's surface; secondly, the ensemble learning was employed to weaken the impact of sample disturbance on model performance and elaborate data preprocessing, including up-sampling and down-sampling, which was adopted to achieve better model performance on the global scale; furthermore, the ENNTm was designed to meet the requirements of three different application conditions by providing three sets of model parameters, i.e., T_m estimating without measured meteorological elements, T_m estimating with only measured temperature and T_m estimating with both measured temperature and water vapor pressure. The validation work is carried out by using the radiosonde data of global distribution, and results show that the ENNTm has better performance compared with other competing models from different perspectives under the same application conditions, the proposed model expanded the application scope of T_m estimation and provided the global users with more choices in the applications of real-time GNSS-PWV retrieval.



Citation: Long, F.; Gao, C.; Yan, Y.; Wang, J. Enhanced Neural Network Model for Worldwide Estimation of Weighted Mean Temperature. *Remote Sens.* **2021**, *13*, 2405. <https://doi.org/10.3390/rs13122405>

Academic Editors: Xingxing Li, Jacek Paziewski and Mattia Crespi

Received: 28 April 2021

Accepted: 13 June 2021

Published: 19 June 2021

Keywords: weighted mean temperature; neural network technique; ensemble learning; precipitation water vapor; zenith wet delay; troposphere; GNSS meteorology

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Detecting atmospheric water vapor with Global Navigation Satellite System (GNSS) has been paid more attention than traditional techniques (such as water vapor radiometer, radiosonde etc.), because it has the advantages of low-cost, all-weather, high-precision and high-resolution in space and time [1–5]. The key to realize real-time conversion from zenith wet delay (ZWD) to precipitable water vapor (PWV) when using GNSS for remote sense water vapor is to obtain accurate weighted mean temperature (T_m) in time. However, the T_m needs to be calculated by numerical integration based on the measured atmospheric profiles [6,7]; for example, the atmospheric profile data collected by radiosonde and GNSS radio occultation are the closest to the actual conditions and can be used to compute T_m [8–10], but they are rarely applied in the real-time retrieval of GNSS-PWV, because the cost of collecting such data is very high, which is difficult for general users to afford. The reanalysis data derived from the numerical weather models (NWMs) have very high

spatiotemporal resolutions, they generally contain the values of various meteorological elements at different heights from the ground to the tropopause and can also be used to compute T_m [10–12]. The reanalysis data have in fact been widely used for climatological studies of PWV rather than for real-time GNSS-PWV retrieval, because these data are produced using the assimilation system and there is a time lag from generation to release. The establishment of the empirical T_m model is an alternative way to obtain T_m values in real time, and the study on T_m modeling has become one of the most important topics in GNSS meteorology.

A large number of empirical T_m models have been developed in recent decades, and they can be divided into two categories according to their application conditions: one is the models without measured meteorological elements (NMTm models); and the other one is called MMTm model, which requires the measured meteorological elements in the T_m estimation. The NMTm models do not need the measured meteorological elements as input, and they use the trigonometric functions to simulate the periodicities (include annual, semi-annual and diurnal variation) of T_m in different locations [13–16]. The reanalysis data generated by NWMs are usually used as the data source in the development of NMTm models, and one of the most representative NMTm models is the Global Pressure and Temperature 2 wet (GPT2w) model, which is an empirical model providing empirical values of various meteorological elements near the Earth's surface including T_m [17]. In recent years, the research focus of NMTm models has shifted to the study on the impact of height differences on the model accuracy, and a series of NMTm models such as the GTm_R model [18], IGPT2w model [19,20], GWMT-D model [11] and GTrop-Tm model [21] have been developed. The application scope of NMTm models has even been extended from the ground to the tropopause [22]. The GWMT-D model and GTrop-Tm model, for example, can be used to globally compute T_m values at different heights in the near-Earth space (from the ground to a 10 km altitude) in real time.

The MMTm models, however, are usually established based upon the relations between T_m and measured meteorological elements such as the temperature (T_s) and water vapor pressure (P_w) of the site. The Bevis formula ($T_m = a \cdot T_s + b$) was first used to calculate T_m [6], its coefficients were fitted from the radiosonde data distributed in the North America, but it was found that the relation between T_m and T_s varies with the locations and seasons, so a Bevis formula with constant coefficients a and b would result in poor accuracy when estimating T_m in different locations and seasons. Many scholars have conducted the localization research on the Bevis formula and found that the model coefficients tailored to the specific locations and seasons can better fit the local conditions [15,23–26]. Some MMTm models for global GNSS-PWV retrieval have also been developed, such as the GTm-I model, PTm-I model [8] and the TVGG model [27]. The fundamental aim of these models is to use the Bevis formula to model the main part of T_m (T_{m0}), and the trigonometric function to simulate the periodicity of the residual caused by the Bevis formula (ΔT_m). Ding [9] first tried to develop the MMTm model with the neural network (NN) technique, and established the first-generation neural network T_m model (named NN-I model), the inputs of which contain the T_m value calculated by the GPT2w model, the latitude and measured T_s of the site. However, these MMTm models can only be used to estimate the T_m values near the Earth's surface, and the complexity of the topography is not considered in the modeling, so their application scope needs to be further extended. Yao et al. [28] first studied the applicability of Bevis formula at different altitudes and established an MMTm model based on the near-Earth atmospheric temperature. Ding [10] released the second-generation neural network T_m model (named NN-II model), where the altitude of the site is introduced as an additional input, and the T_m value derived from the GPT2w model is replaced by that from the GTrop-Tm model. NN-II model extends the T_m calculation from the Earth's surface to 10 km altitude, which greatly expanded the application scope of MMTm models.

The NN-I model and NN-II model all show better performance compared with other published T_m models on the global scale, which has been proved in the literature studies [9,10].

However, the studies on neural network-based T_m modeling can be further deepened. Firstly, the NN-I/NN-II models are all developed based on the neural network technique, but the results of neural network training may be greatly affected by the disturbance of training samples [29]. The NN-I/NN-II models only use the adjusted weight and bias values after one certain training mission as the final model parameters, and their generalization abilities need to be discussed and strengthened. Secondly, the global distribution of data samples used for developing NN-I and NN-II models is extremely uneven, and the sample size in the Northern Hemisphere is much larger than that in the Southern Hemisphere, which would inevitably affect the application of the neural network-based model on the global scale. Thirdly, the measured T_s is required when calculating T_m with the NN-I/NN-II models, but it is not applicable in some cases where the meteorological elements of the site are difficult to measure, so the development of neural network-based T_m models without measured meteorological elements can be taken as a direction of further study. Fourthly, many studies have demonstrated that the introduction of water vapor pressure as input into T_m modeling can help to improve the model accuracy [8,15,18], which is also a study idea worth trying.

In this work, T_m modeling based on neural network is further studied. An enhanced neural network T_m model (namely ENNTm for short) is developed with the help of the ensemble learning method within the framework of machine learning. The ENNTm provides three sets of model parameters corresponding to three empirical models, the ENNTm-A model, ENNTm-B model and ENNTm-C model, which can meet the needs of T_m estimation under three different application conditions, i.e., no measured meteorological element is needed (ENNTm-A model), only the air temperature of the site is required (ENNTm-B model) and both air temperature and water vapor pressure are needed (ENNTm-C model). In the following sections, the modeling process are introduced, their generalization ability of ENNTm is discussed, and the comparisons of model performance with those of other published models are also carried out.

2. The Determination of T_m

2.1. Computing T_m by Numerical Integration

The realization of real-time GNSS-PWV retrieval benefits from the linear relationship between zenith wet delay (ZWD) and precipitable water vapor (PWV) [1,6,7], which can be described by

$$\text{PWV} = \Pi \times \text{ZWD} \quad (1)$$

$$\Pi = \frac{10^6}{\rho_w R_v (k'_2 + k_3/T_m)} \quad (2)$$

where Π is the conversion coefficient to be estimated, and the ρ_w , R_v , k'_2 and k_3 are all constants which can be found in the literature studies [6,15,21–23]. However, the weighted mean temperature (T_m) is a variable that determines the final value of Π , thus plays a decisive role in the GNSS-PWV retrieval. The T_m is defined as the mean temperature of the atmospheric water vapor [30–33] that:

$$T_m = \frac{\int_{h_s}^{\infty} (P_w/T) dh}{\int_{h_s}^{\infty} (P_w/T^2) dh} \quad (3)$$

where P_w and T are the water vapor pressure and absolute temperature of the atmosphere along the vertical direction, respectively; h_s is the height of the site (the starting height for numerical integrating).

In practice, a series of discrete sampling data of the zenith atmospheric profile could be measured and used to calculate T_m by numerical integration. Multiple datasets including atmospheric profiles collected by sounding balloons and reanalysis data produced by NWMs could be used to compute T_m with Equation (3) and act as the data sources for T_m modeling. However, the meteorological elements extracted from reanalysis data are all

generated using assimilation systems [34,35], they do not quite fit the actual atmospheric conditions, and considerable uncertainties may exist in terms of establishing MMTm T_m models with them. For instance, the ECMWF re-analysis-interim (ERA-interim), a modern reanalysis data produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), it was found that there were great differences between the meteorological elements extracted from them and the actual measurements [4], and the performance of MMTm models developed with them are usually not as good as those with radiosonde data [9,15]. The radiosonde data are actually measured by sounding balloons, so the values of meteorological elements derived from them are closer to the actual atmospheric conditions. In addition, the Integrated Global Radiosonde Archive (IGRA) currently provides the measured atmospheric profile data of more than 1000 globally distributed radiosonde stations [36,37], which can fully meet the needs of modeling T_m at the global scale.

2.2. T_m Determination with Empirical Models

2.2.1. GPT2w Model Predictions

The GPT2w (Global Pressure and Temperature 2 wet) model is the basis of the first-generation neural network T_m (NN-I) model, so it is introduced first. The GPT2w model is an excellent blind model for tropospheric delay estimation and can also provide the empirical values of various meteorological elements such as temperature, water vapor pressure and T_m , etc. [17] It is a grid model and the seasonal periodicity of the T_m at each grid node is simulated with a trigonometric function, which can be described by

$$T_m^g = A_0 + A_1 \cos\left(\frac{2\pi \cdot \text{doy}}{365.25}\right) + A_2 \sin\left(\frac{2\pi \cdot \text{doy}}{365.25}\right) + A_3 \cos\left(\frac{4\pi \cdot \text{doy}}{365.25}\right) + A_4 \sin\left(\frac{4\pi \cdot \text{doy}}{365.25}\right) \quad (4)$$

where the T_m^g is the T_m estimate at a certain grid point, *doy* is the day of year, and two sets of coefficients ($A_i, i = 0, 1, \dots, 4$) with mesh resolution of $1^\circ \times 1^\circ$ and $5^\circ \times 5^\circ$ (lat. \times lon.) are provided. The T_m at the target location could be calculated via bilinear interpolating with four grid points closest to the target location.

2.2.2. GTrop-Tm Model Predictions

The GTrop (global troposphere) model is also a grid model and designed for zenith tropospheric delay and T_m estimation [21], and in this paper we gave the part for estimating T_m the name GTrop-Tm model. The GTrop-Tm model also uses the trigonometric function to model the T_m at the grid node:

$$T_m^g = A_0 + A_1(y - 1980) + A_2 \cos\left(\frac{2\pi \cdot \text{doy}}{365.25}\right) + A_3 \sin\left(\frac{2\pi \cdot \text{doy}}{365.25}\right) + A_4 \cos\left(\frac{4\pi \cdot \text{doy}}{365.25}\right) + A_5 \sin\left(\frac{4\pi \cdot \text{doy}}{365.25}\right) \quad (5)$$

where y is the year, $A_i (i = 0, 1, \dots, 5)$ are the coefficients with resolution of $1^\circ \times 1^\circ$ (lat. \times lon.).

The GTrop-Tm model not only further considers the inter-annual variation of T_m , but also has some innovations in modeling strategy compared with the GPT2w model, i.e., the GTrop-Tm model takes the height differences between the target location and the grid points for interpolation into account, and the lapse rate (δ) of T_m with the height at each grid is fitted with the data samples from the Earth's surface to the tropopause. Consequently, the T_m at the grid point should be corrected to the height level of the target location before they are used for bilinear interpolating, which can be expressed by

$$T_m^c = T_m^g + \delta \cdot (h_t - h_g) \quad (6)$$

where the h_t and h_g are the heights of the target location and the grid point for interpolation, respectively, and T_m^c is the corrected T_m of the grid point.

The δ of each grid node is also considered to be a variable with characteristics of inter-annual, annual and semiannual variations, which is also fitted by an equation which is the same as Equation (5). However, T_m is assumed to decrease approximately linearly with the height in the development of the GTrop-Tm model, which may be inconsistent with reality.

2.2.3. NN-I Model Predictions

The NN-I model is a typical case of applying the artificial intelligence to the GNSS meteorology. Ding [9] combined the GPT2w model with the measured surface air temperature (T_s) to develop a combined model for predicting the T_m , namely the first-generation neural network T_m model. The NN-I model makes use of the powerful nonlinear mapping ability of the neural network, and a multiple-layer feedforward neural network (MFNN) is used in the modeling. The model principle can be simply described as

$$T_m = f(T_{m_mGPT2w}, lat, T_s) \quad (7)$$

where the T_{m_mGPT2w} means the T_m derived from the GPT2w model, which is modified by Equation (6) with a lapse rate of -5.1 K/km, the lat and T_s are the latitude and surface air temperature of the site.

2.2.4. NN-II Model Predictions

The NN-I model only considers the correlation between T_m and air temperature near the Earth's surface, but Ding [10] found that the correlation coefficient between the T_m and air temperature below 10 km altitude reached 0.97. He further extended the application scope of the NN-I model to the near-Earth space from the surface to 10 km altitude, and developed the second-generation neural network T_m model (NN-II model), which can be described by

$$T_m = f(T_{m_GTrop}, lat, T_s, H_s) \quad (8)$$

The NN-II model and NN-I model have similar modeling principles and use the same modeling tool (MFNN). However, there are three very important differences between them. Firstly, one of the input parameters of the NN-II model, T_s , refers to the temperature of the site which may be at any height of the troposphere rather than the temperature near the Earth's surface used in NN-I model. Secondly, the NN-II model uses the T_m value derived from the GTrop-Tm model as the input rather than that from the GPT2w model, because the GTrop-Tm model extends the T_m calculation from the ground to the near-Earth space below 10 km altitude. Thirdly, the height of the site is taken as one of the inputs in the NN-II model, while in the NN-I model, the height of the station is taken as an indirect input to modify the GPT2w model (with a lapse rate of -5.1 K/km). The application scope of NN-II model is greatly extended compared with NN-I model.

3. Development of the ENNTm

3.1. Data and Study Area

The sounding balloon released from the ground-based radiosonde stations can directly measure the vertical profiles of meteorological elements. The Integrated Global Radiosonde Archive (IGRA) provides free access to the radiosonde data (<https://www.ncdc.noaa.gov/data-access/weather-balloon/integrated-global-radiosonde-archive>) (accessed on 26 April 2021), and version 2 of IGRA (IGRA V2) has been accessible since 2016, whilst approximately 1000 of the over 2700 IGRA stations are currently reporting data. The data are organized one file per station with two observations per day at UTC 00:00 and UTC 12:00, whereas the pressure, geopotential height, air temperature, and the vapor pressure computed from the temperature, pressure and dewpoint depression at the same level are reported in the IGRA-derived sounding parameters files.

The study area of this work is the whole near-Earth space from the surface to the tropopause, globally, for which a total of 878 IGRA stations were selected, of which 209 stations were used to develop the enhanced neural network T_m model (ENNTm), and the observations of another 569 IGRA stations were used to evaluate the model performances, which can be seen in Figure 1. The data for modeling were measured from 2011 to 2015, and the data for evaluation were from 2016 to 2018. The T_m value at different height layers for each measured atmospheric profile was calculated by using numerical integration method introduced in Section 2.1.

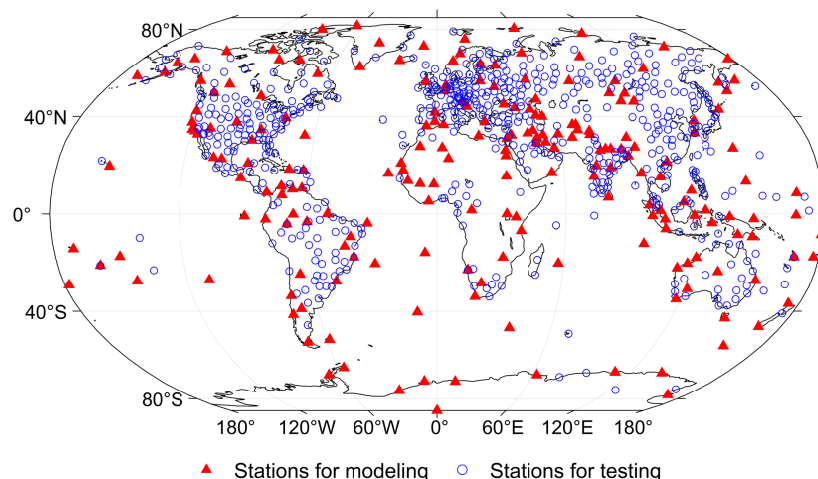


Figure 1. The distribution of IGRA stations used in this work.

It is worth mentioning that the distribution of radiosonde stations is actually extremely uneven due to the global land–sea distribution. The general situation is that the distribution of radiosonde stations is relatively dense in the Northern Hemisphere, such as in East Asia, Europe and North America, etc., while there are few radiosonde stations in the Southern Hemisphere, where only South America and Australia have relatively dense distributions of radiosonde stations. Such problems often exist in the modeling process of previous neural network-based T_m models. Taking the NN-I model and NN-II model, for instance, the distribution of training stations adopted in the modeling is extremely uneven, and the available data samples in the Northern Hemisphere are often far more numerous than those in the Southern Hemisphere. In order to weaken the impact of imbalanced data samples on the model performance at the global scale, this paper attempted to make the uniform distribution of modeling stations when selecting IGRA stations for training, and adopted the up-sampling and down-sampling methods to make the data volume uniform at each latitude band (with latitude interval of 20°). The data size distribution and sampling methods for each latitude band are shown in Table 1.

Table 1. The sampling method and size of the data samples for each latitude band.

Latitude Band	Samples before Preprocessing	Sampling Method	Samples after Preprocessing
90°S–70°S	207,022	up-sampling	621,066
70°S–50°S	313,368	up-sampling	726,736
50°S–30°S	591,581	up-sampling	691,581
30°S–10°S	1,087,641	down-sampling	887,646
10°S–10°N	1,611,588	down-sampling	900,000
10°N–30°N	1,219,985	down-sampling	900,000
30°N–50°N	2,328,141	down-sampling	900,000
50°N–70°N	2,289,975	down-sampling	800,000
70°N–90°N	722,951	up-sampling	722,951

3.2. The BPNN and Ensemble Learning

An artificial neural network (ANN) is a mathematical or computational model that mimics the structure and function of a biological neural network. Modern neural networks usually serve as nonlinear statistical data modeling tools, and have been widely used to solve various problems. For example, the multilayer feedforward neural network (MFNN) is very good for solving nonlinear optimization problems of multiple input parameters, which has been widely used in regression tasks [15,38]. Consequently, a three-layer feedforward neural network with one input layer, one hidden layer and one output layer was employed in this work. The gradient descent method was used to minimize the difference between network output and target output, and a most famous algorithm named back-propagation (BP) was usually adopted in the training of TFNN, so the TFNN was also called the BPNN [39]. Its mathematical principle can be described by

$$Y = f(b^{(2)} + \overrightarrow{W'_{(2)}} \cdot g(b^{(1)} + \overrightarrow{W'_{(1)}} \cdot \overrightarrow{X}) \quad (9)$$

where $\overrightarrow{W'_{(1)}}$ and $\overrightarrow{W'_{(2)}}$ denote the transpose of $\overrightarrow{W_{(1)}}$ and $\overrightarrow{W_{(2)}}$, i.e., the weight matrix to connect neurons of the input layer and hidden layer and weight matrix to connect neurons of the hidden layer and output layer, respectively; $b^{(1)}$ and $b^{(2)}$ indicate the offset from the input layer to the hidden layer and offset from the hidden layer to the output layer, respectively. The \overrightarrow{X} stands for the input variable. A hyperbolic tangent function:

$$g(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (10)$$

and a linear function:

$$f(x) = x \quad (11)$$

are usually adopted as the activation function in the hidden layer and output layer, respectively.

BPNN can well simulate the nonlinear relationship between the input and output. However, the initial weight values for a BPNN are usually randomly generated before training, so the results are usually different from each other after training even for the same training set. In the training process, there is a great risk of falling into the local minimum as a result of training with the gradient descent method, and the generalization ability of the BPNN model for regression may be very poor due to incorrect selection [15,29,40]. Ensemble learning is a theoretical framework of machine learning based on combination strategy which includes the boosting and bagging method. The boosting method usually contains a series of base learners which are trained in series, and there is strong correlation between the sequentially generated learners [40]. In contrast, the bagging method and its base learners are trained in parallel, and the training process could be seen as mutually independent. The bagging method was based on the bootstrap sampling, and the training set for each base learner was approximately independent from one other, to obtain multiple base learners with diversity. For regression tasks, a simple average method can be used to combine multiple base learners to obtain a combined model with better generalization performance:

$$T_m = \frac{1}{T} \sum_{i=1}^T h_i(x) \quad (12)$$

From the perspective of bias and variance, bagging pays more attention to reducing variance, so it is more effective on decision trees, neural networks and other learners that are susceptible to sample disturbance [29].

3.3. The Input Parameters

We computed the correlation coefficients between the T_m and the air temperature (T_s), water vapor pressure (P_w) with the data samples (during 2011 and 2018) derived from 878 IGRA stations, and found that they were 0.9812 and 0.7424, respectively, so the T_s ,

P_w can be considered as the indispensable input of the neural network-based T_m model. In addition, the GTrop-Tm model well simulated the spatiotemporal variations of T_m in the troposphere, even though it had some defects. We also found that the correlation coefficient between T_m calculated by the GTrop-Tm model (T_{m_GTrop}) and T_m extracted from the radiosonde data reached 0.9709. Therefore, in this study, the T_{m_GTrop} was also used as one of the input parameters of the neural network T_m model, and the flow of computing T_{m_GTrop} can be found in the literature study [21]. The height of the site is also an indispensable input of ENNTm due to the vertical variation characteristics of T_m , which is also the key to determining whether the new model can perfectly simulate the T_m in the entire troposphere. Many studies have found that the T_m and its correlation with T_s are significantly changed with the latitude (rather than the longitude) at the global scale, so the latitude (ϕ) of the site should be used as one of the inputs of the new model. The day of year (*doy*) was also adopted as the input to model the seasonal variation of T_m , as well as the $T_m - T_s, P_w$ relations. Therefore, six parameters (*doy*, T_{m_GTrop} , ϕ , h , T_s , P_w) were set as the input variables of ENNTm, while the year, month, date and longitude (λ) of the site served as indirect inputs.

3.4. Modeling Schemes

In the development of ENNTm, three modeling schemes were designed in consideration of the different application conditions that the global users may face. We called the first one the ENNTm-A model, which is applicable to the situation where the site meteorological elements are difficult to measure. The second one is named the ENNTm-B model, which is suitable for the situation that has easy access to the field temperature (T_s). The third one is the ENNTm-C model, which applies to the situation where the air temperature and water vapor pressure (P_w) can be conveniently measured. Their input and output and the model structures are shown in Figure 2.

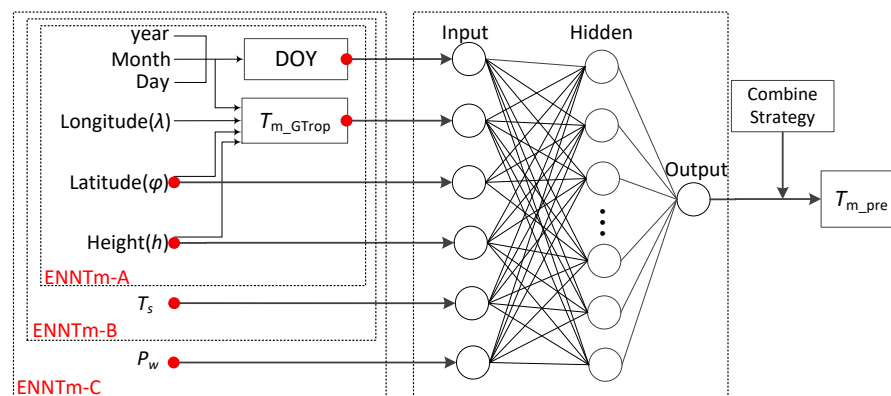


Figure 2. The model structures of three different modeling schemes. The red solid circles denote the direct input parameter of BPNN models.

For each modeling scheme, we designed a total of 10 kinds of network structures, of which the number of neurons in the hidden layer (N) ranges from 5 to 50 ($N = 5, 10, \dots, 50$). For each network structure, we conducted 50 independent training tasks, since the literature study [15] illustrated that an ensemble size of more than 10 can achieve a better generalization ability for BPNN models. As a result, a total of 1500 training tasks were carried out in the development of ENNTm. By extracting data from the radiosonde stations around the world, we processed and generated nearly 10 million pieces of available samples for modeling. However, we only randomly sampled 1 million pieces of data for each training task, so as to obtain diversified BPNNs with different performances via multiple training missions.

4. Results

Atmospheric profiles measured from 2016 to 2018 by 569 IGRA stations were utilized to verify the performance of the ENNTm proposed in this work. The mean deviation (MD), mean absolute deviation (MAD), root-mean-square error (RMSE), standard deviation (STD) and Pearson correlation coefficient (R) were used to evaluate the model accuracies:

$$MD = \frac{1}{n} \sum_{i=1}^n (T_{m,i}^{pre} - T_{m,i}^{obs}) \quad (13)$$

$$MAD = \frac{1}{n} \sum_{i=1}^n |T_{m,i}^{pre} - T_{m,i}^{obs}| \quad (14)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (T_{m,i}^{pre} - T_{m,i}^{obs})^2} \quad (15)$$

$$STD = \sqrt{\frac{1}{n} \sum_{i=1}^n (T_{m,i}^{pre} - T_{m,i}^{obs} - MD)^2} \quad (16)$$

$$R = \frac{\sum_{i=1}^n (T_{m,i}^{pre} - \overline{T_m^{pre}})(T_{m,i}^{obs} - \overline{T_m^{obs}})}{\sqrt{\sum_{i=1}^n (T_{m,i}^{pre} - \overline{T_m^{pre}})^2} \sqrt{\sum_{i=1}^n (T_{m,i}^{obs} - \overline{T_m^{obs}})^2}} \quad (17)$$

where n is the number of the test samples, $T_{m,i}^{pre}$ and $T_{m,i}^{obs}$ are the i -th T_m value computed with the empirical model and the T_m value derived from the radiosonde measurements, respectively. $\overline{T_m^{pre}}$ and $\overline{T_m^{obs}}$ are, respectively, the mean values of $T_{m,i}^{pre}$ and $T_{m,i}^{obs}$.

In this section, the generalization abilities of different modeling schemes were discussed, and the performance of the proposed models were evaluated in comparison with three published empirical models: the GTrop-Tm model, NN-I model and NN-II model. The features and operation conditions of different competing models are shown in Table 2.

Table 2. The features and operation conditions of different competing models.

Model	Application Scope	Modeling Strategy	Direct Inputs
GTrop-Tm	Below 10 km altitude	Least squares	$\phi, \lambda, h, doy, year$
NN-I	Near the Earth's surface	BPNN	ϕ, T_{m_GPT2w}, T_s
NN-II	Below 10 km altitude	BPNN	$\phi, h, T_{m_GTrop}, T_s$
ENNTm-A	Below 10 km altitude	BPNN, ensemble learning	$\phi, h, doy, T_{m_GTrop}$
ENNTm-B	Below 10 km altitude	BPNN, ensemble learning	$\phi, h, doy, T_{m_GTrop}, T_s$
ENNTm-C	Below 10 km altitude	BPNN, ensemble learning	$\phi, h, doy, T_{m_GTrop}, T_s, P_w$

4.1. The Generalization Ability of ENNTm

We first examined the generalization abilities of different modeling schemes, the atmospheric profiles of 569 radiosonde stations distributed globally measured in 2016 (a total of 8,065,535 samples) were employed in this subsection. The MD, MAD, RMSE, STD and R are shown in Figures 3–7. The ensemble size of different modeling schemes are all of 50.

From Figure 3, one can see that the MDs of different modeling schemes become smaller as the number of neurons in the hidden layer (N) increases. The MD changes greatly when N is not exceeding 10, and many more outliers could be found for the ENNTm-C model when N is 5, and with N rising, the MDs of different modeling schemes tend to be stable. The effect of ensemble learning is not obvious from the perspective of MD, since the ensemble learning pays more attention to reducing variance rather than expectation of when it is used for regression tasks, but we can verify the effect of ensemble learning through some other indicators such as MAD, RMSE, STD and R.

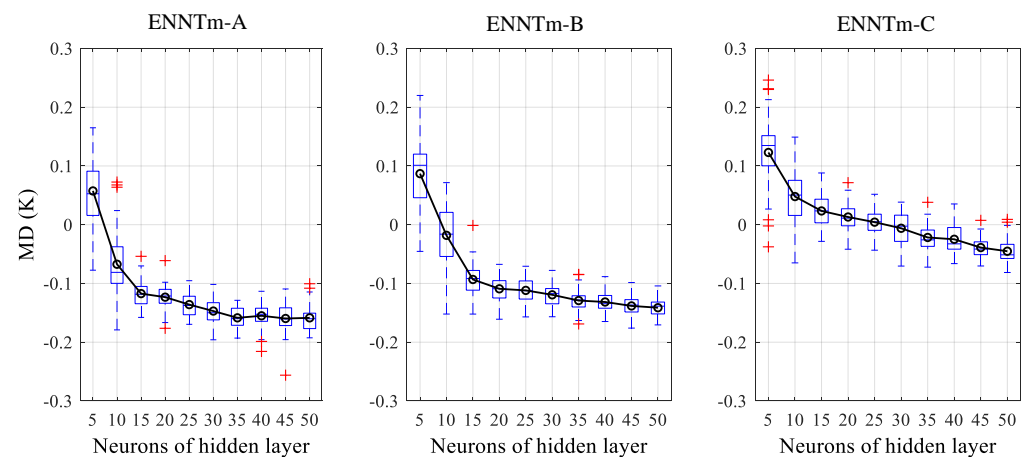


Figure 3. Box-whisker plot of MD for different modeling schemes with different BPNN structures. The small circle denotes the results after combination, the red '+' is the outlier.

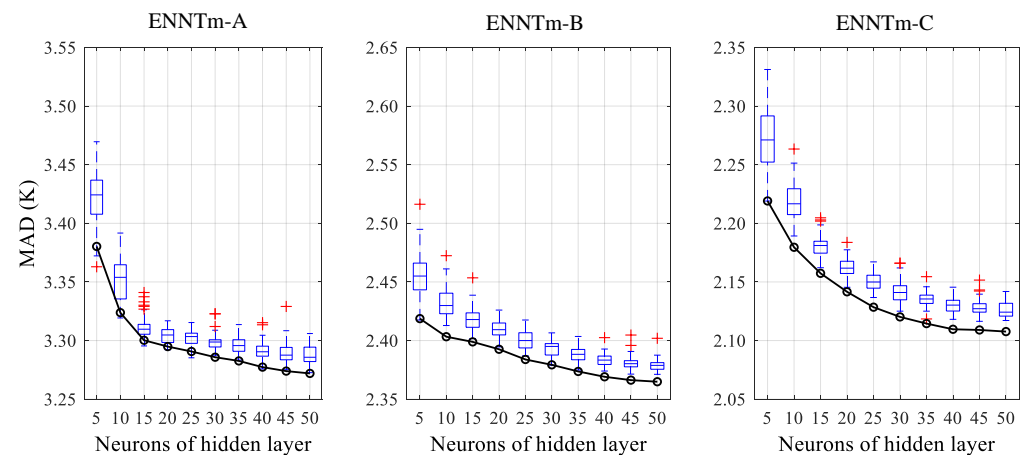


Figure 4. The MAD of different modeling schemes with different BPNN structures.

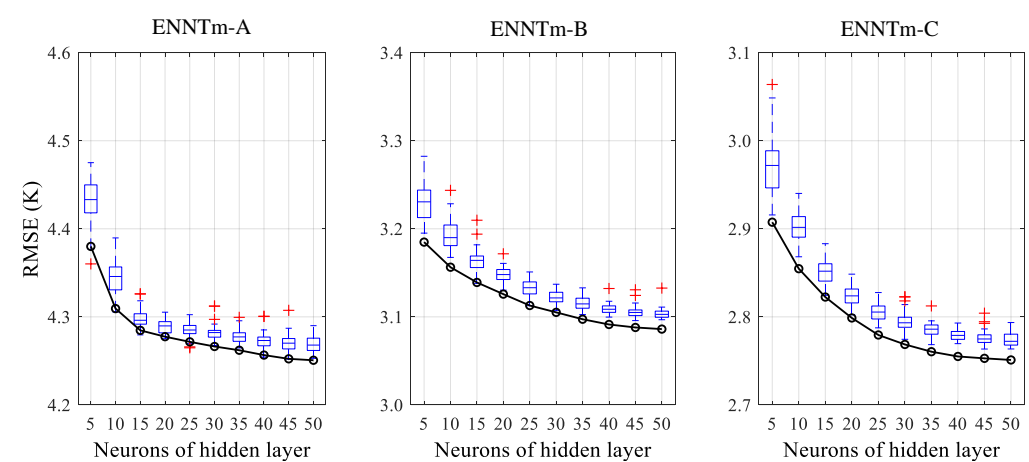


Figure 5. The RMSE of different modeling schemes with different BPNN structures.

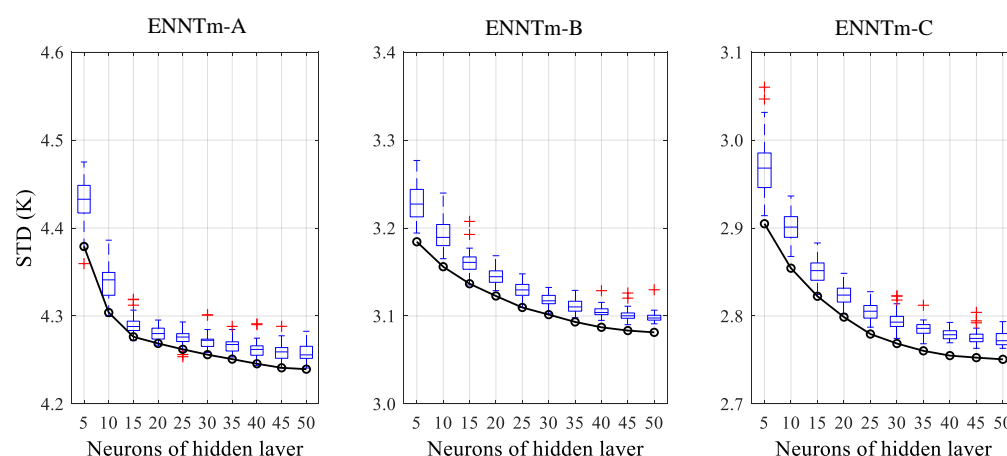


Figure 6. The STD of different modeling schemes with different BPNN structures.

As can be seen in Figures 4–6, the MAD, RMSE and STD decrease with the increase in N , however, when N is greater than 20, the downward trends of these indicators gradually slow down. Another significant and common phenomenon is that the performances of ENNTm models become stronger after combination, the MAD, RMSE and STD of different modeling schemes can even be better than the individual BPNN with the best performance. It can be concluded that the ensemble learning method can help the BPNN model achieve better generalization ability.

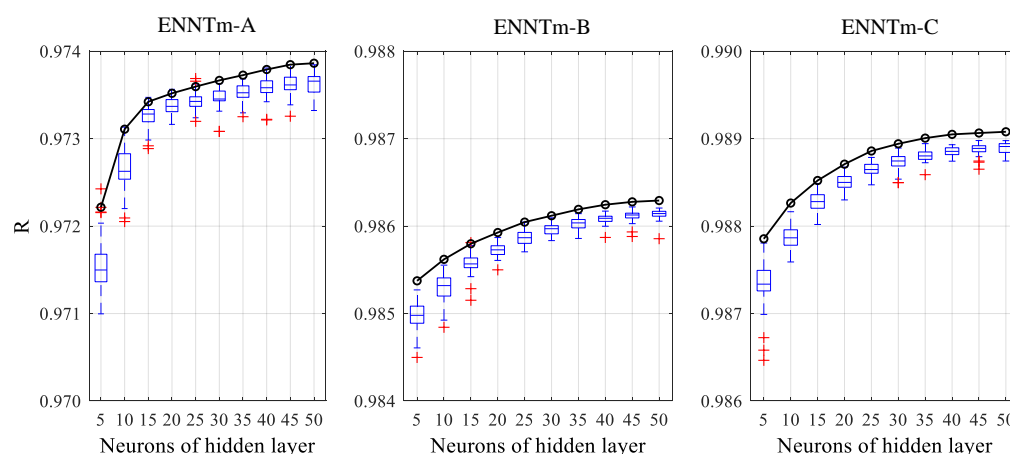


Figure 7. The Pearson correlation coefficient (R) of different modeling schemes with different BPNN structures.

We also examined the Pearson correlation coefficients (R) between the model values and T_m values derived from the radiosonde measurements. It can be seen from Figure 7 that the R becomes larger with the increase in N for different modeling schemes, and many more outliers can be found when N does not exceed 20. The effect of ensemble learning is also remarkable for all the modeling schemes, which is consistent with the results shown in Figures 4–6.

According to the analysis above, we found that the performances of BPNN T_m models become stronger as the number of neurons in the hidden layer (N) increases. However, for a BPNN model with only a three-layer network structure, the improvement of model performance was limited with the increase in model complexity. For example, the reduction rate of RMSE slows down as N is greater than 20. In addition, the model performances can be further improved and better generalization ability can be achieved by combining different training tasks. In conclusion, we believe that the model with N greater than 20 and

enhanced by ensemble learning would achieve better performance, so we also provide all the weight and bias values after training for the user's reference.

4.2. The Overall Performances of Different Models

We investigated the overall performances of ENNTm models in comparison with three published empirical models: the GTrop-Tm model, NN-I model and NN-II model. The atmospheric profile data measured from 2017 to 2018 were utilized in this subsection. The results of ENNTm models all correspond to the results when the number of neurons in the hidden layer (N) is 25.

We first evaluated the model accuracies of different models with the data near the Earth's surface (below 100 m above the Earth's surface), a total of 1,687,098 samples were used to compute the MD, MAD, RMSE, STD and R, and the results are shown in Table 3.

Table 3. The accuracies of different models validated by test samples below 100 m above the Earth's surface during 2017 and 2018.

Models	MD (K)	MAD (K)	STD (K)	RMSE (K)	R
GTrop-Tm	−0.41	3.04	3.97	4.00	0.9418
ENNTm-A	−0.12	3.00	3.96	3.96	0.9419
NN-I	−0.53	2.43	3.13	3.18	0.9640
NN-II	−0.24	2.50	3.26	3.27	0.9617
ENNTm-B	−0.18	2.39	3.11	3.11	0.9646
ENNTm-C	−0.07	2.27	2.94	2.94	0.9684

From Table 3, one can see that the GTrop-Tm model and NN-I model have a larger negative MD on the global scale, while the NN-II model and ENNTm models perform better, and the ENNTm-C model has the smallest MD (−0.07 K) among them. In terms of the MAD, STD and RMSE, the GTrop-Tm model and ENNTm-A model perform poorer than the other competing models since they are all NMTm models, but since others belong to MMTm models, numerous studies have demonstrated that the model accuracies of MMTm models are better than those of NMTm models. Among the published models, the NN-II model was designed for computing T_m from the Earth's surface to 10 km altitude, but its performance is not as good as that of the NN-I model in computing T_m near the Earth's surface. A tiny improvement can be found for the ENNTm-B model compared to the NN-I model. The ENNTm-C model, however, performs the best, since its MAD, STD and RMSE are significantly reduced compared to other competing models, which is consistent with the conclusions of the literature study [15]. We can also obtain similar conclusions from the perspective of R.

Since the ENNTm models are all designed for calculating T_m values in near-Earth space, we conducted the model accuracy comparisons with the data from the Earth's surface to 10 km altitude (a total of 25,373,469 samples), the results are shown in Table 4.

Table 4. The accuracies of different models validated by data from the Earth's surface to a 10 km altitude during 2017 and 2018.

Models	MD (K)	MAD (K)	STD (K)	RMSE (K)	R
GTrop-Tm	0.79	3.69	4.75	4.81	0.9701
ENNTm-A	−0.21	3.3	4.28	4.28	0.9736
NN-I	3.49	4.82	5.59	6.59	0.9763
NN-II	−0.39	2.56	3.28	3.34	0.9846
ENNTm-B	−0.1	2.35	3.08	3.08	0.9864
ENNTm-C	0.04	2.09	2.73	2.73	0.9893

In Table 4, one can see that the MD, MAD, STD and RMSE of NN-I model are far larger than those of other models. The NN-I model is only designed for the T_m estimation near the Earth's surface, and the comparisons between the NN-I model and NN-II model were carried out in the literature study [10], in which it was concluded that the NN-II model is far superior to the NN-I model in all aspects when validated with the data from the Earth's surface to the tropopause, so we did not consider the NN-I model in the following analysis. Among other competing models, the GTrop-Tm model has the largest positive MD, while marked negative MD could be found for the NN-II model and ENNTm-A model. The ENNTm-B model and ENNTm-C model all performed better than other models in terms of MD. In the MAD, STD and RMSE cases, significant improvement can be seen for ENNTm-A model compared to GTrop-Tm model, with an improvement of about 10% in MAD, STD and RMSE, respectively. The ENNTm-B model also improves much in MAD, STD and RMSE compared with the NN-II model, by 8%, 7% and 8%, respectively. The ENNTm-C model performs the best, improves by 12% over ENNTm-B model and 21% over NN-II model in RMSE, respectively. A similar phenomenon can be found in the case of R, but their R values are much larger than those shown in Table 3, since the models shown in Table 4 (except for the NN-I model) are all designed for near-Earth space. In general, the performances of MMTm models are much better than those of NMTm models, and the improvements of ENNTm models in comparison with other published models under the same application conditions are remarkable.

From the comparison and analysis above, we can obtain similar conclusions from the perspectives of MAD, STD, RMSE and R in the evaluation of model accuracy. Therefore, in the following analysis, we only consider the MD and RMSE as the performance indicators.

4.3. The Spatial Performances of Different Models

We further conducted the spatial performance comparisons for different models on the global scale. The testing dataset was divided into five groups according to the geographical zones, the definitions of different geographical zones are shown in Table 5.

Table 5. The definition of different geographical zones.

Geographical Zone	Latitude Band	Abbreviation
North frigid zone	66°34'N–90°N	NFZ
North temperate zone	23°26'N–66°34'N	NTZ
Tropical zone	23°26'S–23°26'N	TZ
South temperate zone	23°26'S–66°34'S	STZ
South frigid zone	66°34'S–90°S	SFZ

The MD and RMSE for different models at each test IGRA station are calculated and shown in Figures 8 and 9, and the statistics of MD and RMSE for each model in different geographical zones are also shown in the figure.

From Figure 8a,b, remarkable positive MD in the middle-to-high latitudes can be found for the GTrop-Tm model, but the IGRA stations from Australia to southern China show obvious negative MDs. The ENNTm-A model performs similarly but is a great improvement over the GTrop-Tm model in most regions. In Figure 8c–e, the NN-II model shows marked negative MDs in the Northern Hemisphere, such as in North America and southern China, but positive MDs could be found in north Asia, South Africa and South America, etc. The ENNTm-B model and ENNTm-C model, however, both show improvement compared to the NN-II model, especially in the middle-to-high latitudes. From Figure 8f, the detailed statistics of MDs for different models seem to be consistent with the results shown in Figure 8a–e. The GTrop-Tm model shows a large positive MD in the NFZ, the NT and the STZ, but a smaller negative MD can be found in the SFZ. In the TZ, there are no marked MDs for any of the competing models. The NN-II model shows a large positive MD in the SFZ and a marked negative MD in the NFZ, but it performs well

in other geographical zones. The ENNTm models, however, all show very small MDs in each geographical zone.

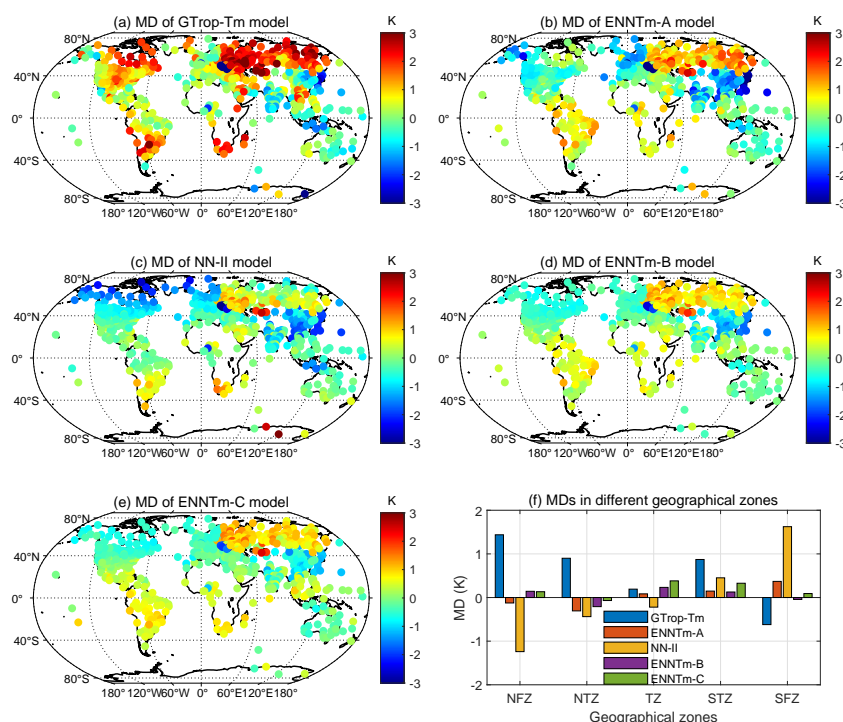


Figure 8. The scattergram of MD (mean deviation) for (a) GTrop-Tm model; (b) ENNTm-A model; (c) NN-II model; (d) ENNTm-B model and (e) ENNTm-C model at each test IGRA station; and (f) their statistics in different geographical zones validated by radiosonde data during 2017 and 2018.

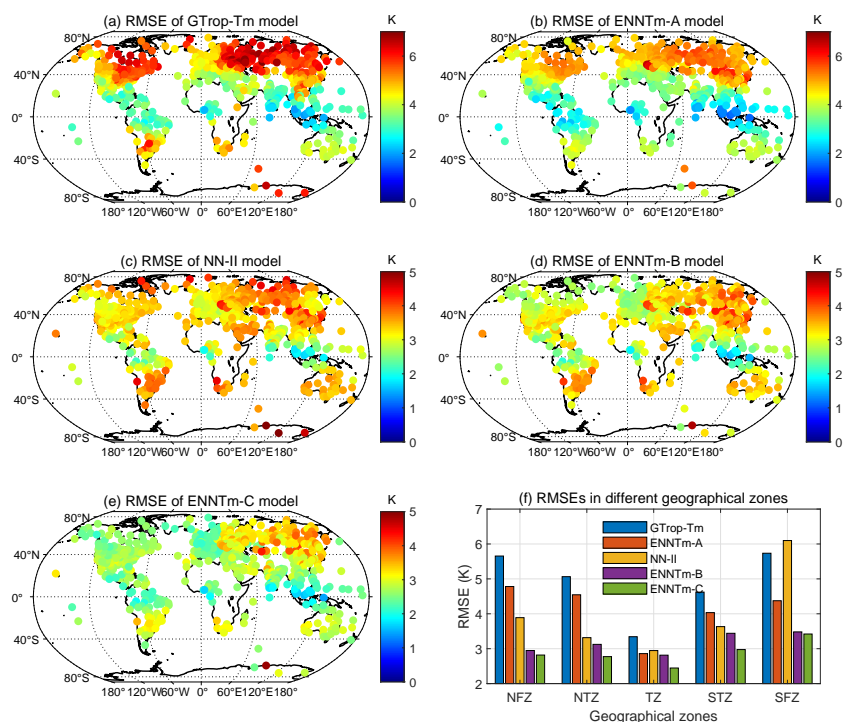


Figure 9. The scattergram of RMSE (root-mean-square error) for (a) GTrop-Tm model, (b) ENNTm-A model, (c) NN-II model, (d) ENNTm-B model and (e) ENNTm-C model at each test IGRA station, and (f) their statistics in different geographical zones validated by radiosonde data during 2017 and 2018.

With regard to the RMSE, one can see from Figure 9a,b,f that the GTrop-Tm model shows a larger RMSE on the global scale than other competing models, especially in the middle to high latitudes. However, significant improvement can be found for the ENNTm-A model in all of the geographical zones compared to the GTrop-Tm model, especially in the SFZ (reduced by 1.4 K) and NFZ (by 0.8 K). From Figure 9c–f, one can see that the values of RMSE for the NN-II model are almost all from 3.0 K to 5.0 K, except for a few IGRA stations distributed near the equator, where it is even inferior to that of the NMTm models (GTrop-Tm model and ENNTm-A model) in the SFZ. The ENNTm-B model and ENNTm-C model have greatly improved this defect, and the improvement in RMSE is very significant in the SFZ (by 2.6 K) and the NFZ (by 0.9 K). We can also see that the RMSEs of ENNTm-C model are smaller than those of the ENNTm-B model in the tropical zone and the temperate zones, even though the improvement is not significant.

We also computed the RMSE reductions for ENNTm models with respect to the GTrop-Tm model and NN-II model, and the detailed statistics of RMSE reductions are shown in Figure 10.

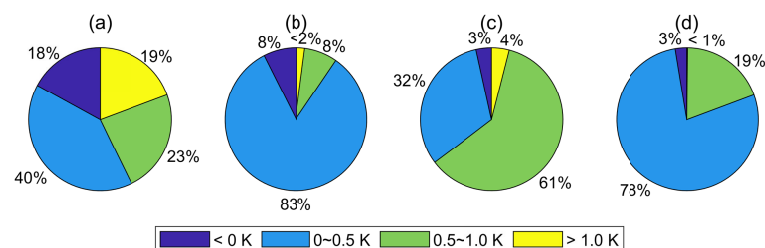


Figure 10. The statistics of RMSE reductions in (a) ENNTm-A w.r.t. GTrop-Tm; (b) ENNTm-B w.r.t. NN-II; (c) ENNTm-C w.r.t. NN-II; and (d) ENNTm-C w.r.t. ENNTm-B. w.r.t = with respect to.

One can see from Figure 10a that the RMSE of the ENNTm-A model reduces in over 80% of test IGRA stations with respect to GTrop-Tm model, and in the stations where the RMSE reduces by more than 0.5 K, this accounts for 42%. From Figure 10b, the RMSE of more than 90% stations reduces for the ENNTm-B model with respect to the NN-II model, and even though the RMSE of over 83% stations only reduces by 0–0.5 K, this can still illustrate the advantage of the ENNTm-B model over the NN-II model in terms of model accuracy. In Figure 10c, one can see that the RMSE reduction in the ENNTm-C model with respect to the NN-II model is significant enough that the test IGRA stations whose RMSE reduces by over 0.5 K account for 65% of the total. In the comparison between ENNTm-B model and ENNTm-C model shown in Figure 10d, the RMSE for the ENNTm-C model at over 97% of the test IGRA stations reduces with respect to ENNTm-B model, which further indicates that the accuracy of the T_m model can be improved by introducing water vapor pressure into modeling.

To investigate the performances of different models in the vertical direction, we divided the test samples in each geographical zone into 10 groups with an interval of 1.0 km altitude to compute the MD and RMSE. The results are shown in Figures 11 and 12.

Figure 11 shows that the MD for each model generally shows an increasing trend with the increase in height and the GTrop-Tm model is the most prominent one, just as that shown in Figure 11b–d. It seems that the GTrop-Tm model performs the poorest at the height layers from 4 to 10 km, except in the SFZ. The ENNTm-A model is shown to be much better than the GTrop-Tm model at any height layers, especially at the height layers above 4 km. The NN-II model has a very large positive MD at the height layers from 8 to 10 km, as shown in Figure 11d,e, however, negative MDs could be found for it at the height layers below 6 km in the NFZ and NTZ shown in Figure 11a,b. The ENNTm-B model and ENNTm-C model perform much better than the NN-II model at any height layers in most geographical zones, such as at the height layers below 5 km in NFZ and NTZ, as shown in Figure 11a,b, as well as the height layers above 7 km in STZ and SFZ shown in Figure 11d,e. Poor model performances could be found for the ENNTm-B model and ENNTm-C model at height

layers over 8 km near the equator, as shown in Figure 11c, but their MDs are not significant, and no more than 1.0 K.

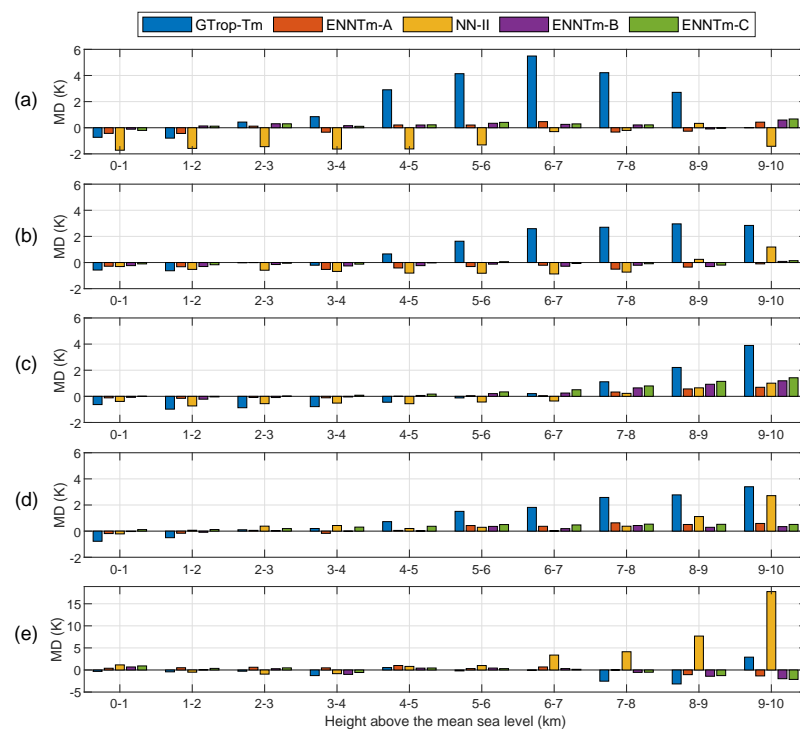


Figure 11. The statistics of MDs at different height layers in the (a) NFZ; (b) NTZ; (c) TZ; (d) STZ; and (e) SFZ.

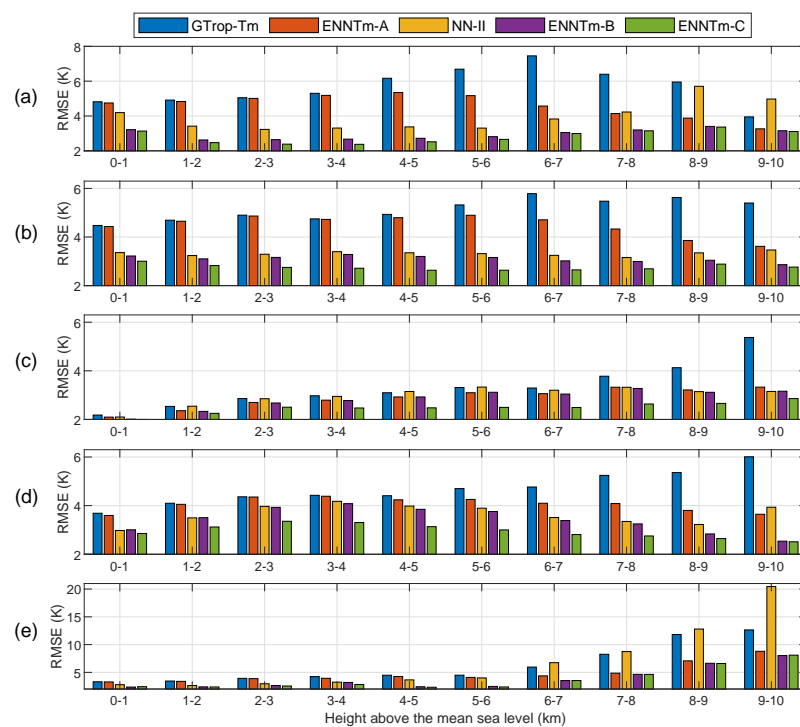


Figure 12. The statistics of RMSEs at different height layers in the (a) NFZ; (b) NTZ; (c) TZ; (d) STZ; and (e) SFZ.

With regard to the detailed statistics of vertical RMSEs in different geographical zones shown in Figure 12, one can see that the RMSE of the GTrop-Tm model increases with the height in the Southern Hemisphere, while its RMSE first increases then decreases with the height in the Northern Hemisphere. The ENNTm-A model also has such a rule but it is better than the GTrop-Tm model, and its RMSE reduces at any height layers compared to the GTrop-Tm model. The NN-II model performs well in the RMSE at any height layers from the NTZ to STZ, as shown in Figure 12b–d, however, in the SFZ shown in Figure 12e, the NN-II model shows a far larger RMSE compared to other competing models at the height layers above 6 km. The RMSE of the ENNTm-B model improves compared to the NN-II model at almost any height layers, even though the improvement is not sometimes significant. The ENNTm-C model, however, improves much over the NN-II model in the near-Earth space in any geographical zones, and it performs the best among the competing models.

4.4. The Temporal Performances of Different Models

We also investigated the temporal performance of different models. The test samples were divided into six groups according to the geographical zones, and we computed the MD and RMSE in different months for each group, for which the results are shown in Figures 13 and 14.

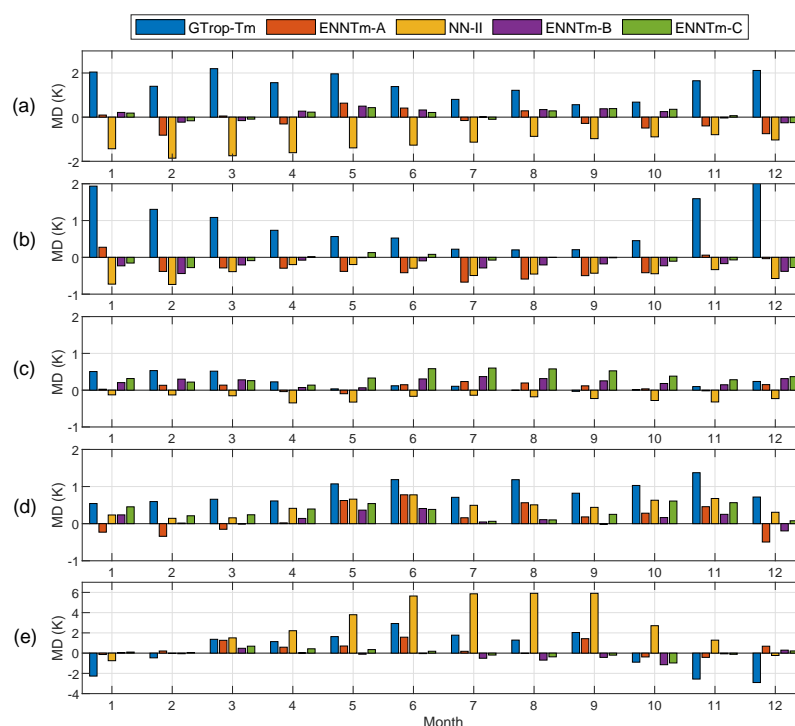


Figure 13. The statistics of MD for each model in different months in the (a) NFZ; (b) NTZ; (c) TZ; (d) STZ; and (e) SFZ.

Figure 13e shows that in the south frigid zone (SFZ), the NN-II model has a far larger positive MD than other competing models in the Southern Hemisphere during winter months, and large positive MDs could be found from April to October, while there is no significant bias in other months. The GTrop-Tm model, however, shows large negative MD in summer months (such as November, December and January) but a remarkable positive MD in winter months. In the statistics of the south temperate zone (STZ) shown in Figure 13d, a similar phenomenon can be seen for the MDs for the GTrop-Tm model and NN-II model in the winter months, which are larger than those in summer months. In the detailed statistics of MDs in the north frigid zone (NFZ) and the north temperate zone (NTZ) shown in Figure 13a,b, this phenomenon also exists as the MDs for the GTrop-Tm

model and NN-II model are larger during winter months (in the Northern Hemisphere) than those during the summer months. However, there are no significant biases which can be found for ENNTm models in any months throughout the year, except for the results in the Tropical Zone (TZ) presented in Figure 13c, where the ENNTm-C model shows a relatively larger positive MD (not more than 0.6 K) from June to September.

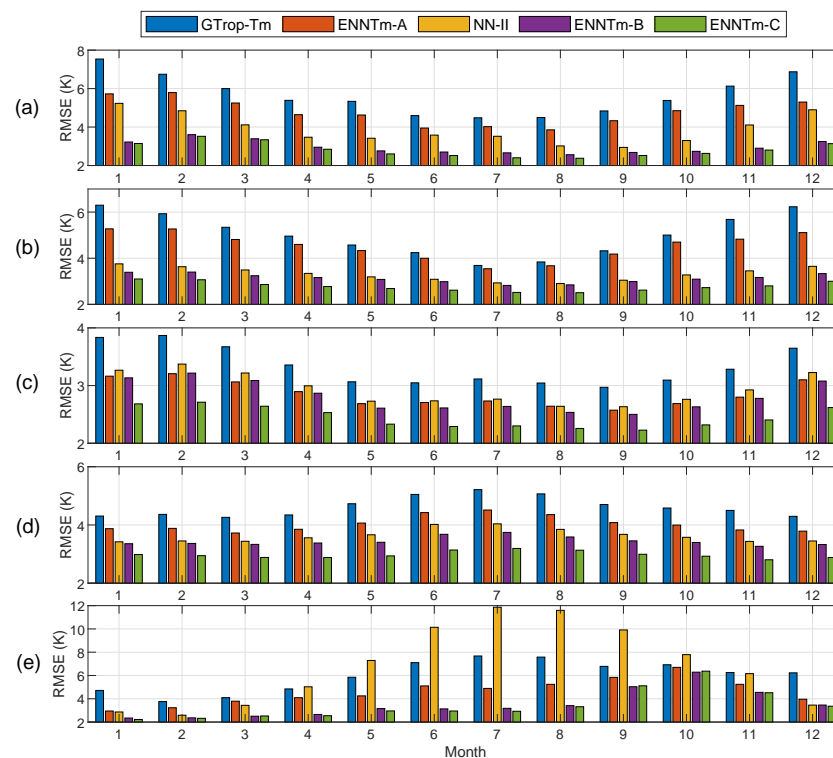


Figure 14. The statistics of RMSE for each model in different months in the (a) NFZ; (b) NTZ; (c) TZ; (d) STZ; and (e) SFZ.

In terms of the detailed statistics of RMSEs shown in Figure 14, a common phenomenon is that the RMSEs of all the competing models are larger in winter months and smaller in summer months, it is mainly caused by the fact that the T_m changes are larger during winter and smaller during summer months [41]. The RMSEs of GTrop-Tm model are always the largest among all the competing models in different months throughout the year, except for the results shown in Figure 14e. The ENNTm-A model is of a great improvement compared to the GTrop-Tm model, but these are still less accurate than those MMTm models. The NN-II model performs the poorest in the winter months of the STZ, with the maximum RMSE of nearly 12 K, but it performs much better than the GTrop-Tm model and ENNTm-A model in other zones. The improvement of ENNTm-B model over the NN-II model is significant in STZ, while in other zones, even though the accuracy of the ENNTm-B model has been consistently improved over the NN-II model in each month, therefore the improvement is not pronounced. The performance of the ENNTm-C model is further improved on the basis of the ENNTm-B model, which indicates that the water vapor pressure is an important factor to improve the model accuracy.

5. Discussion

In this work, the data samples from the Earth's surface to the tropopause were used to develop an enhanced neural network T_m model (ENNTm), which provides three sets of model parameters that can meet three different application conditions for global users (namely ENNTm-A, ENNTm-B and ENNTm-C, respectively). The BPNN shows a powerful capability and capacity to capture the spatiotemporal variations between the T_m and its

associated factors in the development of ENNTm models. The models proposed in this study were further optimized by ensemble learning, the results presented in Section 4.1 show that this method can efficiently strengthen the generalization abilities of BPNN models after multiple training sessions, as the BPNN model after combination showed better generalization ability than that after one certain training mission. From the comparative analysis in Section 4, the ENNTm models proposed in this work showed better performance than other published models listed in this paper from multiple perspectives. The ENNTm-A model improved by 10% over the GTrop-Tm model in RMSE, and had smaller bias σ at the global scale. The ENNTm-B model also improved much in overall MAD, STD and RMSE compared with the NN-II model, by 8%, 7% and 8%, respectively. The ENNTm-C model performed the best, as its RMSE was reduced by 21% and 12% compared to those of the NN-II model and ENNTm-B model, respectively. There are multiple reasons for the improvements of ENNTm models over other competing models (under the same application conditions), as follows.

Firstly, the data for modeling in this study were derived from the atmospheric profiles measured by the sounding balloons, while the GTrop-Tm model was developed with the data derived from the NWMs. These two data sources are different somehow, which leads to some differences between the GTrop-Tm model and ENNTm-A model. The ENNTm-A model performed better than the GTrop-Tm model at most IGRA stations for testing (account for more than 80%).

Secondly, T_m is assumed to have an approximately linear decrease with the increase in height in the vertical direction in the development of the GTrop-Tm model, which deviated from reality. This is exactly the reason why the MD and RMSE of the GTrop-Tm model at the higher height layers were much larger than other competing models, just as shown in Figures 11 and 12. The ENNTm models, however, well simulated the relation between T_m and height with the help of the powerful nonlinear mapping ability of the neural network. The results showed that the performance of the ENNTm-A model in the vertical direction was much better than that of the GTrop-Tm model.

Thirdly, the global distribution of data samples for developing NN-I/NN-II model is extremely uneven, as there are far more samples distributed in the Northern Hemisphere than that in the Southern Hemisphere, while the training results of the neural network are extremely dependent on the training samples. As a result, the performance of the NN-II model on the global scale is also very uneven, as it can be seen from Figures 11–14, since the performance of NN-II model in the frigid zones was very poor. In this study, we employed the up-sampling and down-sampling method to make the global distribution of training samples as uniform as possible. The results showed that the ENNTm-B model improved to some extent over the NN-II model, especially in the regions where the IGRA stations are rarely distributed.

Fourthly, the ENNTm-C model showed the best performance from all perspectives, as it even improved by 12% on the basis of the ENNTm-B model, which largely benefited from the introduction of water vapor pressure (P_w) into modeling. The T_m has a strong correlation with the P_w of the site, and the correlation coefficients between T_m and P_w computed by using the data in the troposphere (from the ground to the tropopause) reached 0.74. We considered the P_w as one of the inputs of the ENNTm models to improve the model accuracy and provide users with more choices.

6. Conclusions

In this study, we developed an enhanced neural network T_m model named ENNTm, in the aim of achieving the excellent capability of the neural network in dealing with nonlinear optimization problems of multiple input parameters. The ENNTm includes three empirical models, the ENNTm-A model, ENNTm-B model and ENNTm-C model, which provide solutions for global users to compute the weighted mean temperature under three different application conditions in real time. Their generalization abilities were enhanced with the ensemble learning method under the machine learning framework.

The comparative analysis showed that the accuracies of ENNTm models are significantly improved compared with the published models under the same application conditions: The ENNTm-A model was superior to the GTrop-Tm model and its accuracy was improved by 10% in RMSE; The ENNTm-B model showed better performance than both the NN-I model and NN-II model from different perspectives, and the RMSE at over 93% test stations reduced with respect to NN-II model; The ENNTm-C model performed the best among all of the competing models from different perspectives, it was 12% and 21% better than ENNTm-B model and NN-II model in RMSE, respectively.

We concluded that the ENNTm proposed in this work performed better than other published models listed in this paper (under the same application conditions) at different perspectives. It benefited from not only the strong capability of neural networks in capturing the characteristics of T_m variation and its relationship with the essential associated factors and the use of the ensemble learning method in strengthening the generalization abilities of new models, but also the use of the measured atmospheric profile data, as well as the reasonable pretreatment work before modeling. However, the variations of T_m and the relationship between T_m and meteorological elements are extremely complex at different spatial and temporal contexts, in addition to their own spatiotemporal distribution characteristics, weather and climate change of different scales can also affect this relationship, which cannot be perfectly reflected with an empirical model, so more research on the T_m modeling under different weather conditions is needed in the following study.

Author Contributions: Conceptualization, F.L. and C.G.; methodology, F.L. and J.W.; software, F.L.; validation, F.L. and Y.Y.; formal analysis, F.L. and Y.Y.; investigation, F.L. and Y.Y.; resources, Y.Y.; data curation, F.L.; writing—original draft preparation, F.L.; writing—review and editing, F.L., C.G. and J.W.; visualization, F.L.; supervision, C.G. and J.W.; project administration, F.L. and C.G.; funding acquisition, C.G. and F.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the National Natural Science Foundation of China, grant number 41974030; and the Postgraduate Research & Practice Innovation Program of Jiangsu Province, grant number KYCX17_0150.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The radiosonde data used in this study are available at the website: <https://www1.ncdc.noaa.gov/pub/data/igra/> (accessed on 20 April 2021) or <ftp://ftp.ncdc.noaa.gov/pub/data/igra> (accessed on 26 April 2021). The source code of GTrop model can be freely accessed at <https://github.com/sun1753814280/GTrop> (accessed on 26 April 2021) and the source code of the NN-II model is available at <https://link.springer.com/article/10.1007/s10291-020-0975-3> (accessed on 26 April 2021). The MATLAB source code of ENNTm is provided at <https://github.com/FengyangGeo/ENNTm> (accessed on 26 April 2021).

Acknowledgments: The author would like to thank the Integrated Global Radiosonde Archive (IGRA) for providing access to radiosonde data and Zhangyu Sun for free sharing the source code of GTrop model. The author would also like to thank Maohua Ding for their valuable discussion, the free MATLAB source code of NN-I model and some experimental data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rocken, C.; Ware, R.; Van Hove, T.; Solheim, F.; Alber, C.; Johnson, J.; Bevis, M.; Businger, S. Sensing atmospheric water vapor with the global positioning system. *Geophys. Res. Lett.* **1993**, *20*, 2631–2634. [CrossRef]
2. Rocken, C.; Hove, T.V.; Johnson, J.; Solheim, F.; Ware, R.; Bevis, M.; Chiswell, S.; Businger, S. GPS/STORM—GPS sensing of atmospheric water vapor for meteorology. *J. Atmos. Ocean. Technol.* **1995**, *12*, 468–478. [CrossRef]
3. Jin, S.; Luo, O.F. Variability and climatology of PWV from global 13-Year GPS observations. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 1918–1924.
4. Wang, X.; Zhang, K.; Wu, S.; Fan, S.; Cheng, Y. Water vapor-weighted mean temperature and its impact on the determination of precipitable water vapor and its linear trend. *Geophys. Res. Atmos.* **2016**, *121*, 833–852. [CrossRef]

5. Darrag, M.; AbouAly, N.; Mohamed, A.M.S.; Becker, M.; Saleh, M. Evaluation of precipitable water vapor variation for east mediterranean using GNSS. *Acta Geod. Geophys.* **2020**, *55*, 257–275. [\[CrossRef\]](#)
6. Bevis, M.; Businger, S.; Herring, T.A.; Rocken, C.; Anthes, R.A.; Ware, R.H. GPS meteorology: Remote sensing of atmospheric water vapor using the global positioning system. *J. Geophys. Res. Atmos.* **1992**, *97*, 15787–15801. [\[CrossRef\]](#)
7. Ross, R.J.; Rosenfeld, S. Estimating mean weighted temperature of the atmosphere for Global Positioning System applications. *J. Geophys. Res. Atmos.* **1997**, *102*, 21719–21730. [\[CrossRef\]](#)
8. Yao, Y.; Zhang, B.; Xu, C.; Yan, F. Improved one/multi-parameter models that consider seasonal and geographic variations for estimating weighted mean temperature in ground-based GPS meteorology. *J. Geod.* **2014**, *88*, 273–282. [\[CrossRef\]](#)
9. Ding, M. A neural network model for predicting weighted mean temperature. *J. Geod.* **2018**, *92*, 187–198. [\[CrossRef\]](#)
10. Ding, M. A second generation of the neural network model for predicting weighted mean temperature. *GPS Solut.* **2020**, *24*, 61. [\[CrossRef\]](#)
11. He, C.; Wu, S.; Wang, X.; Hu, A.; Wang, Q.; Zhang, K. A new voxel-based model for the determination of atmospheric weighted mean temperature in GPS atmospheric sounding. *Atmos. Meas. Tech.* **2017**, *10*, 2045–2060. [\[CrossRef\]](#)
12. Ildikó, J.; Szabolcs, R. Developing a global model for the conversion of zenith wet tropospheric delays to integrated water vapour. *Acta Geod. Geophys.* **2018**, *53*, 259–274. [\[CrossRef\]](#)
13. Yao, Y.; Xu, C.; Zhang, B.; Cao, N. GTm-III: A new global empirical model for mapping zenith wet delays onto precipitable water vapour. *Geophys. J. Int.* **2014**, *197*, 202–212. [\[CrossRef\]](#)
14. Long, F.; Hu, W.; Dong, Y.; Yu, L. A neural network method for estimating weighted mean temperature over China and adjacent areas. *J. Southeast Univ.* **2021**, *37*, 84–90.
15. Long, F.; Hu, W.; Dong, Y.; Wang, J. Neural Network-Based Models for Estimating Weighted Mean Temperature in China and Adjacent Areas. *Atmosphere* **2021**, *12*, 169. [\[CrossRef\]](#)
16. Mateus, P.; Catalão, J.; Mendes, V.B.; Nico, G. An ERA5-Based Hourly Global Pressure and Temperature (HGPT) Model. *Remote Sens.* **2020**, *12*, 1098. [\[CrossRef\]](#)
17. Böhm, J.; Möller, G.; Schindelegger, M.; Pain, G.; Weber, R. Development of an improved empirical model for slant delays in the troposphere (GPT2w). *GPS Solut.* **2015**, *19*, 433–441. [\[CrossRef\]](#)
18. Li, Q.; Yuan, L.; Chen, P.; Jiang, Z. Global grid-based Tm model with vertical adjustment for GNSS precipitable water retrieval. *GPS Solut.* **2020**, *24*, 73. [\[CrossRef\]](#)
19. Huang, L.; Liu, L.; Chen, H.; Jiang, W. An improved atmospheric weighted mean temperature model and its impact on GNSS precipitable water vapor estimates for China. *GPS Solut.* **2019**, *23*, 51. [\[CrossRef\]](#)
20. Yang, F.; Guo, J.; Meng, X.; Shi, J.; Zhang, D.; Zhao, Y. An improved weighted mean temperature (Tm) model based on GPT2w with Tm lapse rate. *GPS Solut.* **2020**, *24*, 46. [\[CrossRef\]](#)
21. Sun, Z.; Zhang, B.; Yao, Y. A global model for estimating tropospheric delay and weighted mean temperature developed with atmospheric reanalysis data from 1979 to 2017. *Remote Sens.* **2019**, *11*, 1893. [\[CrossRef\]](#)
22. Yao, Y.; Sun, Z.; Xu, C.; Xu, X.; Kong, J. Extending a model for water vapor sounding by ground-based GNSS in the vertical direction. *J. Atmos. Sol. Terr. Phys.* **2018**, *179*, 358–366. [\[CrossRef\]](#)
23. Li, L.; Wu, S.; Wang, X.; Tian, Y.; He, C.; Zhang, K. Seasonal multifactor modelling of Weighted-Mean temperature for Ground-Based GNSS meteorology in hunan, china. *Adv. Meteorol.* **2017**, *2017*, 3782687. [\[CrossRef\]](#)
24. Lan, Z.; Zhang, B.; Geng, Y. Establishment and analysis of global gridded Tm-Ts relationship model. *Geod. Geodyn.* **2016**, *7*, 101–107. [\[CrossRef\]](#)
25. Isioye, O.A.; Combrinck, L.; Botai, J. Modelling weighted mean temperature in the West African region: Implications for GNSS meteorology. *Meteorol. Appl.* **2016**, *23*, 614–632. [\[CrossRef\]](#)
26. Sun, Z.; Zhang, B.; Yao, Y. Improving the Estimation of Weighted Mean Temperature in China Using Machine Learning Methods. *Remote Sens.* **2021**, *13*, 1016. [\[CrossRef\]](#)
27. Jiang, P.; Ye, S.; Lu, Y.; Liu, Y.; Chen, D.; Wu, Y. Development of time-varying global gridded Ts-Tm model for precise GPS-PWV retrieval. *Atmos. Meas. Tech.* **2019**, *12*, 1233–1249. [\[CrossRef\]](#)
28. Yao, Y.; Sun, Z.; Xu, C. Applicability of Bevis formula at different height level and global weighted mean temperature model based on near-earth atmospheric temperature. *Acta Geod. Cart. Sin.* **2019**, *48*, 276–285.
29. Zhou, Z.; Wu, J.; Tang, W. Ensembling neural networks: Many could be better than all. *Artif. Intell.* **2002**, *137*, 239–263. [\[CrossRef\]](#)
30. Askne, J.; Nordius, H. Estimation of tropospheric delay for microwaves from surface weather data. *Radio Sci.* **1987**, *22*, 379–386. [\[CrossRef\]](#)
31. Liou, Y.A.; Teng, Y.T.; Van Hove, T.; Liljegren, J.C. Comparison of Precipitable Water Observations in the Near Tropics by GPS, Microwave Radiometer, and Radiosondes. *J. Appl. Meteorol.* **2001**, *40*, 5–15. [\[CrossRef\]](#)
32. Liu, J.; Sun, Z.; Liang, H.; Xu, X.; Wu, P. Precipitable water vapor on the Tibetan Plateau estimated by GPS, water vapor radiometer, radiosonde, and numerical weather prediction analysis and its impact on the radiation budget. *J. Geophys. Res.* **2005**, *110*, D17106. [\[CrossRef\]](#)
33. Vázquez B, G.E., Grejner-Brzezinska, D.A. GPS-PWV estimation and validation with radiosonde data and numerical weather prediction model in Antarctica. *GPS Solut.* **2013**, *17*, 29–39. [\[CrossRef\]](#)
34. Saha, S.; Nadiga, S.; Thiaw, C.; Wang, J.; Wang, W.; Zhang, Q.; Van den Dool, H.M.; Pan, H.-L.; Moorthi, S.; Behringer, D.; et al. The NCEP climate forecast system. *J. Clim.* **2006**, *19*, 3483–3517. [\[CrossRef\]](#)

-
35. Dee, D.P.; Uppala, S.M.; Simmons, A.J.; Berrisford, P.; Poli, P.; Kobayashi, S.; Andrae, U.; Balmaseda, M.A.; Balsamo, G.; Bauer, P.; et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *QJR Meteorol. Soc.* **2011**, *137*, 553–597. [[CrossRef](#)]
 36. Durre, I.; Vose, R.S.; Wuertz, D.B. Overview of the integrated global radiosonde archive. *J. Clim.* **2006**, *19*, 53–68. [[CrossRef](#)]
 37. Durre, I.; Yin, X.; Vose, R.S.; Applequist, S.J. Enhancing the data coverage in the integrated global radiosonde archive. *J. Atmos. Ocean. Technol.* **2018**, *35*, 1753–1770. [[CrossRef](#)]
 38. Hu, A.; Wu, S.; Wang, X.; Wang, Y.; Norman, R.; He, C.; Cai, H.; Zhang, K. Improvement of Reflection Detection Success Rate of GNSS RO Measurements Using Artificial Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 760–769. [[CrossRef](#)]
 39. Zhang, L.; Wang, F.; Sun, T.; Xu, B. A constrained optimization method based on bp neural network. *Neural Comput. Appl.* **2016**, *29*, 413–421. [[CrossRef](#)]
 40. Zheng, D.; Yao, Y.; Nie, W.; Chu, N.; Lin, D.; Ao, M. A new three-dimensional computerized ionospheric tomography model based on a neural network. *GPS Solut.* **2021**, *25*, 10. [[CrossRef](#)]
 41. Huang, L.; Jiang, W.; Liu, L.; Chen, H.; Ye, S. A new global grid model for the determination of atmospheric weighted mean temperature in GPS precipitable water vapor. *J. Geodesy.* **2018**, *93*, 159–176. [[CrossRef](#)]