Weighted Mean Temperature Hybrid Models in China Based on Artificial Neural Network Methods

Meng Cai, Junyu Li *, Lilong Liu, Liangke Huang, Lv Zhou ©, Ling Huang and Hongchang He

College of Geomatics and Geoinformation, Guilin University of Technology, Guilin 541004, China
* Correspondence: junyu_li@whu.edu.cn; Tel.: +86-133-0763-5838

Abstract: The weighted mean temperature \( T_m \) is crucial for converting zenith wet delay to precipitable water vapor in global navigation satellite system meteorology. Mainstream \( T_m \) models have the shortcomings of poor universality and severe local accuracy loss, and they cannot reflect the nonlinear relationship between \( T_m \) and meteorological/spatiotemporal factors. Artificial neural network methods can effectively solve these problems. This study combines the advantages of the models that need in situ meteorological parameters and the empirical models to propose \( T_m \) hybrid models based on artificial neural network methods. The verification results showed that, compared with the Bevis, GPT3, and HGPT models, the root mean square errors (RMSEs) of the new three hybrid models were reduced by 35.3%/32.0%/31.6%, 40.8%/37.8%/37.4%, and 39.5%/36.4%/36.0%, respectively. The consistency of the new three hybrid models was more stable than the Bevis, GPT3, and HGPT models in terms of space and time. In addition, the three models occupy 99.6% less computer storage space than the GPT3 model, and the number of parameters was reduced by 99.2%. To better evaluate the improvement of hybrid models \( T_m \) in the precipitable water vapor (PWV) retrieval, the PWVs calculated using the radiosonde \( T_m \) and zenith wet delay (ZWD) were used as the reference. The RMSE of PWV derived from the best hybrid model’s \( T_m \) and the radiosonde ZWD meets the demand for meteorological research and is improved by 33.9%, 36.4%, and 37.0% compared with that of Bevis, GPT3, and HGPT models, respectively. The hypothesis testing results further verified that these improvements are significant. Therefore, these new models can be used for high-precision \( T_m \) estimation in China, especially in Global Navigation Satellite System (GNSS) receivers without ample storage space.

Keywords: weighted mean temperature; global navigation satellite system meteorology; artificial neural network; empirical model; hybrid model

1. Introduction

Water vapor is an essential component of the Earth’s atmosphere and is crucial for global atmospheric radiation, water cycle, and energy balance [1,2]. Study on water vapor’s spatial and temporal distribution is quite important in weather and climate forecasting. The GNSS signals propagating through the troposphere are delayed and bent due to the nonvacuum conditions, known as tropospheric delays [3–5]. The tropospheric delay is the zenith total delay (ZTD) multiplied by the tropospheric mapping function, and ZTD consists of zenith hydrostatic delay and zenith wet delay. The GNSS zenith wet delay can be converted into precipitable water vapor using Weighted Mean Temperature \( (T_m) \) [6]. Therefore, obtaining a high-precision \( T_m \) is crucial for improving the accuracy of GNSS retrieving precipitable water vapor [6–12]. Previous studies have shown that GNSS-derived PWV is accurate and reliable, with RMSE of 1–3 mm [6,13].

An accurate way to obtain high-precision \( T_m \) is to integrate the vertical profiles of temperature and humidity. However, these profiles are often challenging to obtain in practical work, and thus currently the \( T_m \) models are usually used to calculate the \( T_m \). These models can be divided into two categories. The first category is the model that needs in situ
meteorological parameters. Bevis et al. [6] analyzed the correlation between the surface temperature \((T_s)\) and \(T_m\) of radiosonde data in North America. They found an excellent linear correlation between \(T_m\) and \(T_s\), thereby proposing a linear regression formula of \(T_m\) \((T_m = a + bT_s)\). Since the model coefficients have significant seasonal and local characteristics, to obtain high-precision \(T_m\) values in other regions, the coefficients of the Bevis formula need to be re-estimated from local radiosonde data \([14,15]\). Therefore, many scholars have established regional models for different areas \([16–18]\). These models have high accuracy; however, they are only applicable to certain areas and cannot reflect the delicate nonlinear relationship between \(T_m\) and meteorological factors. Another model \([19,20]\) is an empirical model driven by spatiotemporal information, which does without in situ meteorological parameters and can generate empirical \(T_m\) for large area or even global locations. Leandro et al. [21] used relative humidity to replace the water vapor pressure in the parameter table of the UNB3 model and established the UNB3m model. The UNB3m model considers the annual cyclic variation of meteorological parameters, has a resolution of 15 latitudes, and can be used to calculate \(T_m\). Yao et al. [22] used the GGOS grid \(T_m\) with a 6 h resolution to analyze the daily variation characteristics of \(T_m\), and they constructed a new global \(T_m\) model (GTm-III) that considers the diurnal variation of \(T_m\). Subsequently, Böhm et al. [19] established the GPT2w model, which can provide vital tropospheric parameters such as \(T_m\) and water vapor pressure with a horizontal resolution of up to \(1^\circ \times 1^\circ\). Landskron et al. [23] proposed the GPT3 model, which incorporated meteorological parameter data directly from the GPT2w model; hence, \(T_m\) in the GPT3 model is consistent with GPT2w. Sun et al. [24] used ERA5 data to establish a new model integrating tropospheric delay correction for GNSS positioning and \(T_m\) calculation for GNSS meteorology. The model has a spatial resolution of \(0.5^\circ \times 0.5^\circ\) and a temporal resolution of 1 h. Mateus et al. [25] developed an hourly global pressure and temperature (HGPT) model based on the full spatial and temporal resolution of the new ERA5 reanalysis. The HGPT model can provide hourly surface pressure, surface air temperature, zenith hydrostatic delay, and weighted average temperature information with a spatial resolution of \(0.25^\circ \times 0.25^\circ\). As the spatiotemporal factors considered by the empirical model increase, the models’ overall accuracy gradually improves, but it also increases the number of model parameters. In addition, scholars have also developed many similar empirical models \([20,26,27]\). Although these empirical models are driven by spatiotemporal information and convenient for use and have strong universality, their accuracy is inferior to the model that needs in situ meteorological parameters mentioned above, especially in areas with relatively sparse weather stations and large terrain fluctuations. Still, it is difficult to reflect the delicate nonlinear relationship between \(T_m\) and spatiotemporal factors; therefore, the accuracy of the \(T_m\) model needs to be further improved. Artificial neural network methods have been widely used in various industries due to their nonlinear fitting ability \([28–31]\).

In this study, we hope to establish several \(T_m\) models with higher accuracy than the approved \(T_m\) models, such as Bevis, GPT3, and HGPT, through artificial neural networks, and these models require fewer parameters. Consequently, the PWV converted by the \(T_m\) from the new models can meet the requirements of meteorological research. To achieve this goal, we used artificial neural network methods to fit the relationship between high-precision radiosonde \(T_m\), empirical \(T_m\) (provided by the UNB3m model with few parameters), meteorological parameters, and spatiotemporal factors, and then built high-precision \(T_m\) models.

2. Study Area and Methods for Calculating \(T_m\)

2.1. Study Area

The research area spans \(70^\circ\text{E}–135^\circ\text{E}\) and \(15^\circ\text{N}–55^\circ\text{N}\), which covers the land of China and some surrounding countries and regions (see Figure 1). In this study, we accessed radiosonde data from the Integrated Global Radiosonde Archive (IGRA). Data from 150 radiosonde stations in the study area were obtained for the experiments. The distribution of the stations is indicated by the red triangles in Figure 1.
2. Study Area and Methods for Calculating $T_m$

2.1. Study Area

The research area spans 70°E–135°E and 15°N–55°N, which covers the land of China and some surrounding countries and regions (see Figure 1). In this study, we accessed radiosonde data from the Integrated Global Radiosonde Archive (IGRA). Data from 150 radiosonde stations in the study area were obtained for the experiments. The distribution of the stations is indicated by the red triangles in Figure 1.

Figure 1. Study terrain and distribution of radiosonde stations (marked with red triangles).

2.2. Method of Calculating $T_m$

2.2.1. Calculation of $T_m$ Based on Radiosonde Data

Radiosonde data are important meteorological observation data. The IGRA has provided high-quality sounding observations from more than 1500 radiosondes and sounding balloons worldwide since the 1960s and launches radiosonde twice daily at 00:00 and 12:00 UTC. This study used data from 150 radiosonde stations (2007–2016) to calculate $T_m$ according to the following formula [6]. It is worth mentioning that during long-term continuous observations, the radiosonde data may have discontinuities and outliers. Since these values will affect the results of the new model establishment, this study used the interquartile range (IQR) method to remove outliers in the long-term series of radiosonde data [32].

$$T_m = \frac{\int (e/T)dH}{\int (e/T^2)dH}$$  

(1)

where $T$ is the absolute temperature (K). $e$ is the water vapor pressure (hPa), which is calculated from the relative humidity ($RH$) using Equations (2) and (3) [6,10]:

$$e_s = 6.11 \times 10^7 \left(\frac{7.5 + T_d}{237.3 + T_d}\right)$$  

(2)

$$e = RH \cdot e_s / 100$$  

(3)

where $e_s$ is the saturated vapor pressure (hPa) and $T_d$ is the dew point temperature (°C).

2.2.2. Calculating $T_m$ Based on the UNB3m Model

The UNB3m model is based on a look-up table with annual mean and amplitude for temperature, pressure, temperature lapse rate, and water vapor pressure height factor. These parameters are calculated for a particular latitude and day of year using a cosine
function for the annual variation and a linear interpolation for latitude. The annual average value of the meteorological parameters is calculated using the following formula:

$$AVG_{\phi} = \begin{cases} AVG_{15}, & \phi \leq 15^\circ \\ AVG_{75}, & \phi \geq 75^\circ \\ AVG_i + \frac{(AVG_{15} - AVG_i)}{15}(\phi - LAT_i), & 15^\circ < \phi < 75^\circ \end{cases}$$ (4)$$

where $\phi$ is the latitude of the target point ($^\circ$), $AVG_{\phi}$ is the annual mean value, $i$ is the latitude band closest to the target location that has a smaller value, and $LAT_i$ is the value of the corresponding latitude band. The formula for calculating the annual cycle amplitude is as follows:

$$AMP_{\phi} = \begin{cases} AMP_{15}, & \phi \leq 15^\circ \\ AMP_{75}, & \phi \geq 75^\circ \\ AMP_i + \frac{(AMP_{15} - AMP_i)}{15}(\phi - LAT_i), & 15^\circ < \phi < 75^\circ \end{cases}$$ (5)$$

where $AMP_{\phi}$ denotes the annual cycle amplitude. The meteorological parameter values at a specified time at the target point can be calculated by entering the mean yearly value and annual periodic amplitude value of the target parameter into the trigonometric function:

$$X_{\phi,doy} = AVG_{\phi} - AMP_{\phi} \cos\left(\left(doy - doy_0\right) \frac{2\pi}{365.25}\right)$$ (6)$$

where $X_{\phi,doy}$ is the annual periodic amplitude value, $doy$ is the day of the year, and $doy_0$ specifies the initial phase of the regular change, which is 28 in the Northern Hemisphere and 211 in the Southern Hemisphere.

The UNB3m model can calculate $T_m$ according to the following formula:

$$T_m = (T_0 - \beta H)\left(1 - \frac{\beta R}{g_m(\lambda + 1)}\right)$$ (7)$$

where $T_0$, $\beta$, and $\lambda$ are the meteorological parameters calculated according to procedures (4)–(6), which are temperature (K), temperature lapse rate (K/m), and water vapor pressure height factor, respectively; $H$ is the orthometric height (m); $R$ is the gas constant for dry air (287.054 J/kg/K); and $g_m$ is the acceleration of gravity at the atmospheric column centroid (m/s$^2$), which can be expressed as

$$g_m = 9.784\left(1 - 2.66 \times 10^{-3}\cos(2\phi) - 2.8 \times 10^{-7}H\right)$$ (8)$$

The above model has few parameters and is convenient and straightforward to use; however, the UNB3m model only considers the variation of $T_m$ with latitude, and his grid is too sparse, so the accuracy is limited [33].

2.2.3. Calculating $T_m$ Based on the GPT3 Model

The global pressure and temperature 3 (GPT3) model proposed by Landskron and Bohm is the latest version of the GPT series [23]. The meteorological parameters of the GPT3 model are the same as that of GPT2w with admirable performance. It is one of the most widely used models [10,31]. GPT3 characterizes $T_m$ seasonal variations based on the following Equation (9) and takes into account their geographical variations by $1^\circ \times 1^\circ$ or $5^\circ \times 5^\circ$ grids.

$$r(t) = A_0 + A_1\cos\left(\frac{2\pi t}{365.25}\right) + B_1\sin\left(\frac{2\pi t}{365.25}\right) + A_2\cos\left(\frac{4\pi t}{365.25}\right) + B_2\sin\left(\frac{4\pi t}{365.25}\right)$$ (9)$$
where \( r(t) \) is the meteorological parameters to be estimated; \( t \) denotes the day of year; \( A_0 \) represents its mean value; and \((A_1, B_1)\) and \((A_2, B_2)\) are their annual and semiannual amplitudes, respectively.

3. Construction of Hybrid Model

In this study, we use Backpropagation neural network (BPNN), Random Forest (RF), and Generalized Regression Neural Network (GRNN) in the neural network toolbox of MATLAB software to optimize UNB3m \( T_m \) by using high-precision radiosonde \( T_m \) data, then construct three hybrid models.

3.1. Three Artificial Neural Network Methods

3.1.1. BPNN

The BPNN, first proposed by Rumelhart et al. [34], is one of the most widely used ANNs. The network adopts the gradient descent method to minimize the differences between the network output and target output [35].

The BPNN consists of an input layer, a hidden layer, and an output layer. The number of neurons in the input layer of the BPNN is equal to the number of input variables, and the number of neurons in the output layer is equal to the number of output variables. Almost every bounded continuous function can be approximated with an arbitrarily minor error using a neural network with a single hidden layer [36]. Therefore, in this study, we chose the number of hidden layers as a single layer and determined the number of neurons in the hidden layer in Section 3.3.

Because the empirical model optimized in this study is the UNB3m model, UNB3m \( T_m \) must be used as an essential input for the artificial neural network model. UNB3m \( T_m \) is established considering the variations in the latitude, height, and day of year, so there must be a strong correlation between UNB3m \( T_m \) and latitude, height, and day of year. Simultaneously, some studies [4,19,24] found that there exist long-term interannual variations and diurnal variations in \( T_m \), so we take year, hour of day (hod), latitude (lat), height, and day of year (doy) as input variables.

It is well known that the surface temperature \( (T_s) \) and water vapor pressure \( (e_s) \) have strong relationships with the observation \( T_m \) [6,11], and the \( T_m \) models based on in situ \( T_s/e_s \) observations can reach higher precision than empirical models. Thus, the \( T_s/e_s \) were employed as inputs for the new model. Li et al. [5] suggested that UNB3m often causes great prediction tropospheric delay biases in some areas because it is only based on latitude–label meteorological parameters that are also needed for calculating \( T_m \). Moreover, many \( T_m \) models [11,19,25], considering the longitude variations, have excellent performance. Therefore, we also took longitude as an input variable. It should be noted that the widespread distribution of surface meteorological observation facilities, \( T_s \) and \( e_s \), can be obtained in real-time in most regions [37]. Therefore, it is not difficult to obtain the temperature and water vapor pressure to support the operation of the new model, and establishing the \( T_m \) model based on meteorological parameters has a good potential for real-time application. Table 1 shows all the input parameters and the output parameter. In addition, when training the model, the radiosonde-derived \( T_m \) was loaded into the output layer, and the output layer outputs the corrected \( T_m \) value when used. The structure of the BPNN \( T_m \) model is shown in Figure 2.

### Table 1. Main features of the hybrid models.

<table>
<thead>
<tr>
<th>Input parameters</th>
<th>Output parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>surface temperature ( (T_s) ), water vapor pressure ( (e_s) ), year, day of year (doy), and hour of day (hod), latitude, longitude, height, and UNB3m-( T_m )</td>
<td>( T_m )</td>
</tr>
</tbody>
</table>
Table 1. Main features of the hybrid models.

<table>
<thead>
<tr>
<th>Input parameters</th>
<th>Output parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>surface temperature ((T_s)), water vapor pressure ((e_s)), year, day of year (doy), and hour of day (hod), latitude, longitude, height, and UNB3m-(T_m)</td>
<td>(T_m)</td>
</tr>
</tbody>
</table>

Figure 2. Structural diagram of backpropagation neural network.

The process of training the BPNN includes forward and backpropagation. Each neuron in each layer of the BPNN model is directly connected to the neurons in the next layer and had an activation function. This study used the hyperbolic tangent function to activate the input and hidden layer neurons. A linear function was used to activate the neurons of the hidden and output layers. The equations are represented as:

\[
g(x) = \frac{2}{1 + \exp(-2x)} - 1 \quad (10)
\]

\[
f(x) = x \quad (11)
\]

Then the final output of BPNN can be expressed as:

\[
Y(X) = f(W_{3,2} \cdot g(W_{2,1} \cdot X + b_1) + b_2) \quad (12)
\]

where \(W_{2,1}\) and \(W_{3,2}\) represent the weight matrix, \(b_1\) and \(b_2\) represent the bias matrix. These four matrices store the coefficients of the BPNN model. \(X\) and \(Y\) are the input and output variables, respectively.

3.1.2. RF

Breiman and Cutler first proposed a random forest in 2001 [38]. RF is an ensemble learning method used for classification and regression. RF works by constructing many decision trees during training and then outputting the mode of the classes or the average prediction of the individual trees [38]. RF has the advantages of fast training speed and handling complex nonlinear relationships between the input and output variables. The structure of the RF \(T_m\) model is shown in Figure 3. The input data included time (year, day of year, and hour of day), location (latitude, longitude, and height), \(e_s\), \(T_s\), and UNB3m-\(T_m\).
In addition, the output data were radiosonde-derived $T_m$ or the corrected $T_m$. Because overfitting can occur with a single decision tree, RF overcomes this problem by introducing randomness into each decision tree and averaging the results. The result of the model was the mean of the consequences of all constructed decisions, as shown in the following formula:

$$Y(X) = \frac{1}{B} \sum_{b=1}^{B} T_b(X)$$

(13)

where $X$ represents the input variable, $Y$ is the final RF output value, $T_b$ denotes the output value of each decision tree, and $B$ represents the number of decision trees. The number of decision trees directly affected the accuracy of the RF model. Because the number of decision trees must be optimally selected, the enumeration method is generally used. The specific number value is determined in Section 3.3.

Figure 3. Structural diagram of random forest.

3.1.3. GRNN

Specht first proposed GRNN in 1991 [39]. The GRNN neural network is a radial basis function network based on mathematical statistics. The GRNN has a strong nonlinear mapping ability and learning speed, and the network can also handle unstable data. The GRNN consists of four layers: the input, pattern, summation, and output layers. The number of neurons in the input layer corresponds to the number of input variables; nine neurons correspond to the time information (year, day of year, and hour of day), location information (latitude, longitude, and height), $e_s$, $T_s$, and UNB3m-$T_m$. The number of neurons in the output layer corresponds to the number of output variables. In this experiment, the output layer had only one neuron corresponding to the radiosonde $T_m$ or the model-corrected $T_m$. The summation layer includes two types of summation neurons, which perform arithmetic summation and weighted summation of the output values of the
The number of neurons in the pattern layer corresponds to the number of training samples, and the transmission of neurons can be expressed as:

$$p_i = \exp\left\{-\frac{(X - X_i)^T(X - X_i)}{2\sigma^2}\right\}, i = 1, 2, \ldots, n$$  \hspace{1cm} (14)

where $p_i$ is the output of the $i$ neuron in the pattern layer, represented by the exponential function of the square of the Euclidean distance between the input variable $X_i$ (the $i$-th learning sample) and its corresponding test sample. $\sigma$ represents the spread parameter, the only unknown parameter in the network, and needs to be set first. It is necessary to determine the optimal spread parameter as $\sigma$, which, when too large, may make the estimation very smooth or, when too small, may result in a value too close to the sample value [39]. The specific value is defined in Section 3.3. The structure of the GRNN $T_m$ model is shown in Figure 4.

![Figure 4. Structural diagram of generalized regression neural network.](image)

3.2. Evaluation Indicators Adopted by the Model

This experiment adopted a 10-fold cross-validation method to evaluate the different artificial neural network models [40]. The basic principle of the 10-fold cross-validation technique is to randomly split the dataset into 10 groups and then select 9 groups as the training set and 1 group as the test set. This process was repeated 10 times, so each part of the dataset was tested once, trained 9 times, and all residuals were computed and saved. Note that it can ensure that more data is involved in the training so that the results are closer to the accuracy of the final model and can also prevent overfitting. We calculated five statistical indicators based on these residuals to evaluate model performance. These indicators are the bias, mean absolute error (MAE), standard deviation (STD), root
mean square error (RMSE), and Pearson correlation coefficient (R). The formulas of these indicators are as follows:

\[ \text{Bias} = \frac{1}{N} \sum_{i=1}^{N} (T_{hm,k,i} - T_{rs,k,i}) \]  
(15)

\[ \text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |T_{hm,k,i} - T_{rs,k,i}| \]  
(16)

\[ \text{STD} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (T_{hm,k,i} - T_{rs,k,i} - \text{Bias})^2} \]  
(17)

\[ \text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (T_{hm,k,i} - T_{rs,k,i})^2} \]  
(18)

\[ R = \frac{\sum_{i=1}^{N} (T_{hm,k,i} - \overline{T_{hm,k,j}})(T_{rs,k,i} - \overline{T_{rs,k,j}})}{\sqrt{\sum_{i=1}^{N} (T_{hm,k,i} - \overline{T_{hm,k,j}})^2 \sum_{i=1}^{N} (T_{rs,k,i} - \overline{T_{rs,k,j}})^2}} \]  
(19)

where \( N \) is the number of samples, \( T_{hm,k,i} \) is the \( T_m \) value output by the hybrid models, \( T_{rs,k,i} \) is the \( T_m \) values derived from the radiosonde data, and \( \overline{T_{hm,k,j}}, \overline{T_{rs,k,j}} \) are the mean values of \( T_{hm,k,i} \) and \( T_{rs,k,i} \), respectively.

After verifying the accuracy and reliability of the hybrid models using 10-fold cross-validation, we fitted all samples to generate the final model for subsequent \( T_m \) prediction. When users want to use the hybrid models to calculate \( T_m \), they only need to collect \( T_s \) and \( e_s \) and then input them with the time and location information into the models’ code.

3.3. Parameter Determination

In this experiment, the number of neurons in the BPNN hidden layer, the number of RF decision trees, and the GRNN spread value are the parameters that need to be set first. We followed Sun et al. [41] to set up the crucial parameters. For the BPNN, the optimal number of neurons in the hidden layer can be selected in the range of \( 2\sqrt{n} + \mu \) to \( 2n + 1 \) (\( n \) is the number of neurons in the input layer, \( \mu \) is the number of neurons in the output layer) [42,43]; therefore, the experiment used a 10-fold cross-validation technique to test the BPNN models with 7 to 19 hidden layer neurons. For RF, we set the number of decision trees between 5 and 95 with a step size of 10 and then used 10-fold cross-validation to test the performance of RF models with different numbers of decision trees. For GRNN, the spread value used is usually in the range of 0.01 to 1 [28]. After many tests, we found that the optimal spread value was between 0 and 0.1; therefore, this experiment narrowed the selection range from 0.01 to 0.1, with a step size of 0.01. Similarly, we used a 10-fold cross-validation technique to test the performance of GRNN models with different spread values. Finally, the root mean square errors (RMSE) calculated from the cross-validation residuals were used to evaluate the performance of the other models with different parameter settings. The results are shown in Figure 5.
Therefore, we finally chose 0.06 as the spread value. Thus, we set the number of decision trees to 55 to reduce the complexity of the model. In the RF model, when the number of decision trees increased from 5 to 55, the RMSE decreased significantly. But this decrease closed out after the number of decision trees exceeded 55. Thus, we set the number of decision trees to 55 to reduce the complexity of the model. In the GRNN model, when the spread value increased from 0 to 0.06, the RMSE gradually decreased; however, after the spread value exceeded 0.06, the RMSE gradually increased. Therefore, we finally chose 0.06 as the spread value.

4. Performance Analyses of Hybrid Models

In this study, the three hybrid models constructed using the three artificial neural network methods of BPNN, RF, and GRNN are named hybrid model 1, hybrid model 2, and hybrid model 3 (referred to as hm1, hm2, and hm3, respectively). The Bevis model \( T_m = 70.2 + 0.72T_s \) needs in situ meteorological parameters as input and is commonly used internationally. The GPT3 model is an empirical model that is widely used, and it is a grid model that also can reflect the characteristics of \( T_m \) in a certain area. Its accuracy performance in China can be used to represent mainstream empirical models. The HGPT model [25] is a recently released \( T_m \) model with open source code. Therefore, the Bevis, GPT3 \( (1° \times 1°) \), and HGPT models were chosen to compare with the new models.

4.1. Overall Performance

After determining the parameters of the three artificial neural network methods, the experiment selected 936,034 samples from 2007 to 2016 for training and obtained the corresponding models and results. The cross-validation and model fitting accuracy results are listed in Table 2. Note that since the 10-fold validation uses all the data to verify the model’s accuracy, the validation results here span from 2007 to 2016. Scatter plots between the \( T_m \) values obtained from different models and the \( T_m \) values derived from the radiosonde data are shown in Figures 6 and 7. In Figures 6 and 7, the color indicates data density.
Table 2. Relative radiosonde data, the accuracy evaluation results of different models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Hyperparameter</th>
<th>Bias (K)</th>
<th>MAE (K)</th>
<th>STD (K)</th>
<th>RMSE (K)</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNB3m</td>
<td></td>
<td>-1.97</td>
<td>8.41</td>
<td>10.78</td>
<td>10.955</td>
<td>0.540</td>
</tr>
<tr>
<td>Hm1</td>
<td>Cross-V</td>
<td>0.00</td>
<td>2.28</td>
<td>2.95</td>
<td>2.954</td>
<td>0.969</td>
</tr>
<tr>
<td></td>
<td>Fitting</td>
<td>0.00</td>
<td>2.28</td>
<td>2.95</td>
<td>2.953</td>
<td>0.969</td>
</tr>
<tr>
<td>Hm2</td>
<td>Cross-V</td>
<td>0.00</td>
<td>2.07</td>
<td>2.70</td>
<td>2.703</td>
<td>0.974</td>
</tr>
<tr>
<td></td>
<td>Fitting</td>
<td>0.00</td>
<td>1.62</td>
<td>2.10</td>
<td>2.096</td>
<td>0.984</td>
</tr>
<tr>
<td>Hm3</td>
<td>Cross-V</td>
<td>0.06</td>
<td>0.02</td>
<td>2.09</td>
<td>2.76</td>
<td>0.973</td>
</tr>
<tr>
<td></td>
<td>Fitting</td>
<td>0.01</td>
<td>1.59</td>
<td>2.10</td>
<td>2.101</td>
<td>0.984</td>
</tr>
<tr>
<td>Bevis</td>
<td></td>
<td>-0.80</td>
<td>3.53</td>
<td>4.49</td>
<td>4.563</td>
<td>0.931</td>
</tr>
<tr>
<td>GPT3</td>
<td></td>
<td>-0.48</td>
<td>3.35</td>
<td>4.31</td>
<td>4.340</td>
<td>0.932</td>
</tr>
<tr>
<td>HGPT</td>
<td></td>
<td>0.00</td>
<td>3.33</td>
<td>4.32</td>
<td>4.317</td>
<td>0.932</td>
</tr>
</tbody>
</table>

Figure 6. Scatter plots of estimated $T_m$ against $T_m$ derived from radiosonde data for different models: (a) hybrid model 1 cross-validation; (b) hybrid model 1 fitting; (c) hybrid model 2 cross-validation; (d) hybrid model 2 cross-validation; (e) hybrid model 3 cross-validation; (f) hybrid model 3 fitting.
were reduced by 31.6%, 37.4%, and 36.0% respectively. After improving the three RMSEs, the LS model is based on the same modeling data set as the hybrid model, and the specific formula is as follows:

\[ T_m = a_1 + a_2 \cdot T_s + a_3 \cdot r^2 + a_4 \cdot h + a_5 \cdot lon + a_6 \cdot lat \cos \left( \frac{DOY - a_{10}}{365.25} \cdot \frac{2\pi}{4\pi} \right) + a_{11} \cdot \cos \left( \frac{HOD - a_{12}}{24} \cdot \frac{2\pi}{2\pi} \right) + a_{13} \cdot \cos \left( \frac{HOD - a_{14}}{24} \cdot \frac{2\pi}{2\pi} \right) \]

(20)

Table 2 shows that the biases of the UNB3m model, the Bevis model, and the GPT3 model are \(-1.97\) K, 0.80 K, and \(-0.48\) K, respectively. The results indicate that there are systematic biases in the three models. However, the biases of the three hybrid models are all close to zero, which suggests that the artificial neural network method successfully corrects systematic biases. Because the artificial neural network methods successfully removed systematic biases, the STDs and RMSEs were nearly equal. After being corrected by the three hybrid models, the RMSEs are reduced to 2.954 K, 2.703 K, and 2.763 K, which are 73.1%, 75.3%, and 74.8% lower than the UNB3m model, respectively. By using the analysis of variance (ANOVA), it is further demonstrated that the \(T_m\) of the three hybrid models is not significantly different from the radiosonde \(T_m\) (\(p > 0.05\)). Meanwhile, relative to the Bevis model, the RMSEs were reduced by 35.3%, 40.8%, and 39.5%; for the GPT3 model, the RMSEs were reduced by 32.0%, 37.8%, and 36.4%; for the HGPT model, the RMSEs were reduced by 31.6%, 37.4%, and 36.0% respectively. After improving the three artificial neural network methods, the correlation coefficient \(R\) increased from 0.540 to 0.969, 0.974, and 0.973. The above analysis shows that all three artificial neural network methods significantly improved the accuracy of the UNB3m model in calculating \(T_m\). These improvements may be accounted for by the strong ability of artificial neural network methods to fit complex nonlinear relationships. (The discussion will be described in the following paragraph.) In addition, when comparing the three hybrid models, we can see that the accuracy validation result of hm2 is smaller than those of hm1 and hm3. Therefore, hm2 exhibited a more stable performance.

To illustrate whether the improvement in accuracy of the three hybrid models comes from the data source or the method, we developed a linear model named LS model, which has the same input parameters as the three hybrid models using the least-squares method. The LS model is based on the same modeling data set as the hybrid model, and the specific formula is as follows:

\[ T_m = a_1 + a_2 \cdot T_s + a_3 \cdot r^2 + a_4 \cdot h + a_5 \cdot lon + a_6 \cdot lat + a_7 \cdot \cos \left( \frac{DOY - a_{10}}{365.25} \cdot \frac{2\pi}{4\pi} \right) + a_{11} \cdot \cos \left( \frac{HOD - a_{12}}{24} \cdot \frac{2\pi}{2\pi} \right) + a_{13} \cdot \cos \left( \frac{HOD - a_{14}}{24} \cdot \frac{2\pi}{2\pi} \right) \]

(20)
Then, the accuracy of the hybrid models, LS model, Bevis model, GPT3 model, and HGPT model are compared. The results are shown in Table 3.

**Table 3.** The overall accuracy of different models.

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE (K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hm1</td>
<td>2.954</td>
</tr>
<tr>
<td>Hm2</td>
<td>2.703</td>
</tr>
<tr>
<td>Hm3</td>
<td>2.763</td>
</tr>
<tr>
<td>LS model</td>
<td>3.340</td>
</tr>
<tr>
<td>Bevis</td>
<td>4.563</td>
</tr>
<tr>
<td>GPT3</td>
<td>4.340</td>
</tr>
<tr>
<td>HGPT</td>
<td>4.317</td>
</tr>
</tbody>
</table>

Table 3 shows that the RMSE of the LS model has been improved to different degrees when compared with that of the Bevis, GPT3, and HGPT models, which should be due to the use of the $T_s$, $e_s$, and $T_m$ data in the study area to fit the relationship between them. Using the same data source, the RMSE of the three hybrid models increased by 13.2%, 23.7%, and 21.0% compared with the LS model. This improvement should be due to the method that can fit the nonlinear relationship between different parameters.

### 4.2. Spatiotemporal Performance of the Hybrid Models

In this section, we calculate the RMSEs of 150 radiosonde stations to analyze the spatial performance of the hybrid models. Figure 8 shows the specific values and distributions. Figure 9 shows the frequencies of the RMSEs for each interval. The numbers over the bars represent the number of stations within the corresponding RMSE range. Figures 8 and 9 also include the RMSEs of the UNB3m, Bevis, GPT3, and HGPT models at each radiosonde station.

**Figure 8.** Spatial distribution of the RMSE that are calculated cross-validation residuals at each radiosonde station for (a) UNB3m model, (b) Hybrid model 1, (c) Hybrid model 2, (d) Hybrid model 3, (e) Bevis model, (f) GPT3 model, and (g) HGPT model.
As shown in Figure 8, regardless of the model used, low latitudes show a smaller RMSE and high latitudes show a larger RMSE, which is consistent with the results of Sun et al. [24]. This phenomenon occurs because the seasonal variation at high latitudes is more substantial than that at low latitudes, which increases the difficulty of $T_m$ modeling and eventually leads to larger RMSEs at high latitudes. At the same time, it can be seen in the figure that the RMSEs of the UNB3m model exhibit apparent differences in different latitude ranges. In contrast, the three hybrid models are uniformly distributed and stable. The RMSEs of the Bevis, GPT3, and HGPT models are much larger than those of the three hybrid models. Therefore, the method proposed in this study improves the accuracy of the model. Further, it makes the model accuracy more evenly distributed in space, owing to the introduction of geographic information into the model input layer.

Figure 9 shows that the RMSEs of the three hybrid models are much smaller than the UNB3m model. Simultaneously, compared with the RMSEs of the Bevis model, the GPT3 model, and the HGPT model, the hybrid models also have a significant improvement. The number of sites with an RMSE value smaller than 3.0 K is 0 for the UNB3m model, 74 for the hybrid model 1, 101 for the hybrid model 2, 89 for the hybrid model 3, 22 for the Bevis model, 20 for the GPT3 model, and 21 for the HGPT model. The above results show that the method proposed in this study significantly improves the model's accuracy, and hm2 performs best in all models.

Because latitude is an essential factor affecting $T_m$ [6], we divided the study area into eight latitude bands with intervals of 5° to compare and analyze the performance of each model in different latitude bands. The results are shown in Figure 10.
the method proposed in this study significantly improves the model’s accuracy, and \( h_{m2} \) performs best in all models. Because latitude is an essential factor affecting \( T_m \) [6], we divided the study area into eight latitude bands with intervals of 5° to compare and analyze the performance of each model in different latitude bands. The results are shown in Figure 10.

**Figure 10. Biases and RMSEs of each model at different latitudes.**

The biases of the three hybrid models at different latitudes are all close to zero, which indicates that the artificial neural network method has a noticeable effect on correcting the systematic error of the UNB3m model (Figure 10). The stability of the hybrid models at different latitudes is better than that of the Bevis, GPT3, and HGPT models (Figure 10). The RMSEs of all models are generally smaller at low latitudes and larger at high latitudes and show an increasing trend with increasing latitude, which is consistent with the results presented in Figure 8. Regardless of the latitude band, the RMSEs of the hybrid models were much lower than those of the Bevis, GPT3, and HGPT models, indicating that the three hybrid models significantly improved the accuracy of \( T_m \). Furthermore, compared with the Bevis, GPT3, and HGPT models, the RMSEs of the three hybrid models are all within 2 to 3.6 K in different latitude bands, and the variation between various bands is slight and uniform, showing more notable advantages. Comparing the three hybrid models, the RMSEs of \( h_{m2} \) at different latitudes are slightly lower than those of the other two models. This result means that \( h_{m2} \) outperforms other models.

\( T_m \) is greatly affected by station altitude, and altitude differences may lead to uncertainties of \( T_m \) model accuracy [6]. Therefore, we divided the height into eight layers with an interval of 0.5 km and calculated RMSEs for each height layer. The results of the analysis are presented in Figure 11.
Figure 11 shows that the biases of the three hybrid models tend to be stable and remain near 0, indicating that the accuracies of the hybrid models at different heights are much better than the UNB3m model, Bevis model, GPT3 model, and HGPT model. The RMSEs of the hybrid models obtained after optimization by the artificial neural network methods are 30% lower, relative to the UNB3m model. The RMSEs of the three hybrid models decrease with the increase in altitude and are lower than that of the Bevis, GPT3, and HGPT models. Furthermore, most of the RMSEs of the three hybrid models were below 3 K and were much smaller than those of the UNB3m, Bevis, GPT3, and HGPT models. This result shows that the accuracies of the hybrid models are higher than those of the Bevis, GPT3, and HGPT models and have a more stable accuracy in the vertical direction.

Because $T_m$ has prominent seasonal characteristics [6], we calculated the biases and RMSEs for the four seasons to analyze the temporal variation of the performance of each model, as shown in Figure 12.

Figure 12 shows that the UNB3m model shows poor accuracy in all seasons, especially in summer. However, after the correction by the artificial neural network methods, the accuracy in each season has been greatly improved. The RMSE in spring, autumn, and winter dropped to more than 50% of the UNB3m model, and the RMSE in summer directly decreased to more than 80%. In addition, the RMSEs of the hybrid models in each period are smaller than that of the Bevis, GPT3, and HGPT models, indicating the superiority of the hybrid model. The hybrid models have higher and more uniform accuracy in time. When comparing the three hybrid models, we can see that the RMSEs of hm2 are smaller than those of hm1 and hm3. Therefore, hm2 performs best in each season.
Figure 12. Biases and RMSEs of each model in different seasons.

4.3. Occupancy of Hybrid Models

We also compared the computer storage space and the number of parameters for each model, and the results are presented in Table 4.

Table 4. The computer storage space and the number of parameters for each model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Computer Storage Space</th>
<th>Number of Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNB3m</td>
<td>104 KB</td>
<td>103</td>
</tr>
<tr>
<td>Hm1</td>
<td>104 KB</td>
<td>104</td>
</tr>
<tr>
<td>Hm2</td>
<td>104 KB</td>
<td>104</td>
</tr>
<tr>
<td>Hm3</td>
<td>104 KB</td>
<td>104</td>
</tr>
<tr>
<td>GPT3</td>
<td>29,081.6 KB</td>
<td>324,003</td>
</tr>
</tbody>
</table>

Table 4 indicates that the computer storage space occupied by the three hybrid models is small and is reduced by 99.6% compared to that occupied by the GPT3 model. Compared with GPT3 model, the number of parameters was reduced by 99.2%. Therefore, the new model has a tremendous advantage in storage.

5. Applications of Hybrid Models in Retrieving PWV

From the performance discussions of the three hybrid models, we recommend hybrid model 2 (Hm2). The formula for calculating PWV by combining ZWD and $T_m$ is as follows:

$$PWV = \Pi \times ZWD$$  \hspace{1cm} (21)

$$\Pi = \frac{10^6}{\rho_w R_v [k_3 / T_m + k_2]}$$  \hspace{1cm} (22)

where $\rho_w$ is the density of water and $R_v$ is the specific gas constant.

To evaluate the effect of error in $T_m$ on its synthesized PWV, a commonly used quantity is the relative error in PWVs calculated using the following formula:
\[
\frac{\sigma_{\text{pwv}}}{\text{PWV}} = \frac{\Pi(T_m + \sigma_{T_m}) - \Pi(T_m)}{\Pi(T_m)} = \frac{1 + \left(\frac{k'}{k_3}\right) T_m}{1 + \left(\frac{k'}{k_3}\right) (T_m + \sigma_{T_m})} \cdot \frac{T_m + \sigma_{T_m}}{T_m} - 1
\]  

(23)

where \(\sigma_{\text{pwv}}\) is the error in \(\text{PWV}\) caused by the error in \(T_m\) and \(\sigma_{T_m}\).

Since \(k_2'/k_3 \approx 5.9 \times 10^{-5} \text{ K}^{-1}\) and \(T_m\) is in the range from 220 K to 310 K in our experiment, Equation (23) can be simplified to [22,44]

\[
\frac{\sigma_{\text{pwv}}}{\text{PWV}} \approx \frac{T_m + \sigma_{T_m}}{T_m} - 1 = \frac{\sigma_{T_m}}{T_m}
\]  

(24)

Assuming that there is no error in the value of \(\text{ZWD}\), if the error of \(T_m\) is small, this will improve the converting accuracy of \(\text{ZWD}\) to \(\text{PWV}\). Therefore, our established hybrid model will indirectly enhance the accuracy of \(\text{PWV}\).

To further illustrate the improved accuracy of the hybrid model for \(\text{PWV}\) inversion, we selected four stations for analysis, and the station information is shown in Table 5. The \(\text{PWVs}\) calculated from the \(\text{ZWD}\) and the \(T_m\) provided by the four radiosonde stations in 2016 were employed as a reference to validate the \(\text{PWV}\) mapped from radiosonde \(\text{ZWD}\) using the hybrid model 2/Bevis/GPT3/HGPT model. The results are shown in Figure 13, Tables 6 and 7.

Table 5. The information of the four stations.

<table>
<thead>
<tr>
<th>Station Number</th>
<th>Latitude/(^{\circ})</th>
<th>Longitude/(^{\circ})</th>
<th>Altitude/m</th>
</tr>
</thead>
<tbody>
<tr>
<td>58,457</td>
<td>30.23</td>
<td>120.16</td>
<td>43.1</td>
</tr>
<tr>
<td>50,557</td>
<td>49.16</td>
<td>125.23</td>
<td>242.6</td>
</tr>
<tr>
<td>51,463</td>
<td>43.78</td>
<td>89.61</td>
<td>921.4</td>
</tr>
<tr>
<td>45,004</td>
<td>22.33</td>
<td>114.17</td>
<td>66.17</td>
</tr>
</tbody>
</table>

Table 6. The MAE of the differences between the computed and the reference PWV.

<table>
<thead>
<tr>
<th>Station Number</th>
<th>Hm2 Bevis Change in GPT3 Change in HGPT</th>
<th>MAE/(\text{mm})</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>58,457</td>
<td>0.199 0.293 46.9 0.430 115.4 0.428 114.4</td>
<td>0.150 0.224 49.3</td>
<td>0.332 121.2 0.332 121.2</td>
</tr>
<tr>
<td>50,557</td>
<td>0.101 0.182 80.6 0.204 102.5 0.240 137.8</td>
<td>0.067 0.134 99.8</td>
<td>0.137 104.2 0.157 133.5</td>
</tr>
<tr>
<td>51,463</td>
<td>0.111 0.248 124.5 0.217 96.5 0.206 92.5</td>
<td>0.080 0.169 112.0</td>
<td>0.166 108.2 0.154 92.5</td>
</tr>
<tr>
<td>45,004</td>
<td>0.219 0.454 107.1 0.380 73.3 0.308 72.0</td>
<td>0.169 0.381 124.7</td>
<td>0.307 81.3 0.308 82.0</td>
</tr>
</tbody>
</table>

Table 7. The RMSE of the differences between the computed and the reference PWV.

<table>
<thead>
<tr>
<th>Station Number</th>
<th>Hm2 Bevis Change in GPT3 Change in HGPT</th>
<th>RMSE/(\text{mm})</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>58,457</td>
<td>0.199 0.293 46.9 0.430 115.4 0.428 114.4</td>
<td>0.199 0.293 46.9</td>
<td>0.430 115.4 0.428 114.4</td>
</tr>
<tr>
<td>50,557</td>
<td>0.101 0.182 80.6 0.204 102.5 0.240 137.8</td>
<td>0.101 0.182 80.6</td>
<td>0.204 102.5 0.240 137.8</td>
</tr>
<tr>
<td>51,463</td>
<td>0.111 0.248 124.5 0.217 96.5 0.206 92.5</td>
<td>0.111 0.248 124.5</td>
<td>0.217 96.5 0.206 92.5</td>
</tr>
<tr>
<td>45,004</td>
<td>0.219 0.454 107.1 0.380 73.3 0.308 72.0</td>
<td>0.219 0.454 107.1</td>
<td>0.380 73.3 0.308 72.0</td>
</tr>
</tbody>
</table>

As shown in Figure 13, the PWV calculated by the hybrid model agrees better with the reference value, which shows that the hybrid model outperforms the Bevis, GPT3, and HGPT models in retrieving the PWV. The RMSEs of PWV derived from the Hm2 model are all less than 1 mm; the accuracy is very appreciable in weather research because the required RMS is 3 mm [45]. Tables 6 and 7 show that compared with the Bevis model, the retrieving PWV accuracy of the hybrid model is greatly improved, and its MAE is 96.5%.
Table 6. The MAE of the differences between the computed and the reference PWV.

<table>
<thead>
<tr>
<th>Station Number</th>
<th>Hm1 MAE/mm</th>
<th>Bevis MAE/mm</th>
<th>Change in %</th>
<th>GPT3 MAE/mm</th>
<th>Change in %</th>
<th>HGPT MAE/mm</th>
<th>Change in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>58,457</td>
<td>0.150</td>
<td>0.224</td>
<td>49.3</td>
<td>0.332</td>
<td>121.2</td>
<td>0.332</td>
<td>121.2</td>
</tr>
<tr>
<td>50,557</td>
<td>0.067</td>
<td>0.134</td>
<td>99.8</td>
<td>0.137</td>
<td>104.2</td>
<td>0.157</td>
<td>133.5</td>
</tr>
<tr>
<td>51,463</td>
<td>0.080</td>
<td>0.169</td>
<td>112.0</td>
<td>0.166</td>
<td>108.2</td>
<td>0.154</td>
<td>92.5</td>
</tr>
<tr>
<td>45,004</td>
<td>0.169</td>
<td>0.381</td>
<td>124.7</td>
<td>0.307</td>
<td>81.3</td>
<td>0.308</td>
<td>82.0</td>
</tr>
</tbody>
</table>

Table 7. The RMSE of the differences between the computed and the reference PWV.

<table>
<thead>
<tr>
<th>Station Number</th>
<th>Hm2 RMSE/mm</th>
<th>Bevis RMSE/mm</th>
<th>Change in %</th>
<th>GPT3 RMSE/mm</th>
<th>Change in %</th>
<th>HGPT RMSE/mm</th>
<th>Change in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>58,457</td>
<td>0.199</td>
<td>0.293</td>
<td>46.9</td>
<td>0.430</td>
<td>115.4</td>
<td>0.428</td>
<td>114.4</td>
</tr>
<tr>
<td>50,557</td>
<td>0.101</td>
<td>0.182</td>
<td>80.6</td>
<td>0.204</td>
<td>102.5</td>
<td>0.240</td>
<td>137.8</td>
</tr>
<tr>
<td>51,463</td>
<td>0.111</td>
<td>0.248</td>
<td>124.5</td>
<td>0.217</td>
<td>96.5</td>
<td>0.206</td>
<td>86.1</td>
</tr>
<tr>
<td>45,004</td>
<td>0.219</td>
<td>0.454</td>
<td>107.1</td>
<td>0.380</td>
<td>73.3</td>
<td>0.379</td>
<td>72.6</td>
</tr>
</tbody>
</table>

As shown in Figure 13, the PWV calculated by the hybrid model agrees better with the reference value, which shows that the hybrid model outperforms the Bevis, GPT3, and HGPT models in retrieving the PWV. The RMSEs of PWV derived from the Hm2 model are all less than 1 mm; the accuracy is very appreciable in weather research because the required RMS is 3 mm [45]. Tables 6 and 7 show that compared with the Bevis model, the retrieving PWV accuracy of the hybrid model is greatly improved, and its MAE is 96.5% lower on average than the Bevis model. The RMSE is 89.8% lower on average. The accuracy of the hybrid model in retrieving PWV has been dramatically improved in comparison with the GPT3 model and the HGPT model. Simultaneously, the MAE of the hybrid model decreased by 103.7% and 107.3% on average, and RMSE decreased by 96.9% and 102.7%, respectively. These proved that the hybrid model is superior to the Bevis, GPT3, and HGPT models in retrieving PWV.

To demonstrate the outperformance of the Hm2 for retrieving PWV more intuitively, Figure 14 gives RMSEs reduction at all radiosonde stations. Figure 14 illustrates that the Hm2 model shows smaller RMSE than other models at most stations. The mean RMSE reduction of the Hm2 model reaches up to 33.9% in comparison with Bevis, 36.4% in comparison with GPT3, and 37.0% in comparison with HGPT. It is interesting to note that the RMSE of the Hm2 PWV is reduced at all stations in China Mainland. Overall, the Hm2 model performs the best in mapping ZWD onto PWV.

To further verify the Hm2 has a significant improvement in accuracy compared to the comparison model, the RMSE of each model at each site was used as a sample for hypothesis testing. Set the null hypothesis H0: there is no significant difference between the RMSE of the comparison model at each site (sample X1) and the RMSE of the Hm2 model at each site (sample X2). Alternative hypothesis H1: there is a significant difference between X1 and X2. Due to the large sample size, the z-test [46] was used. Set the left-side confidence level to a = 0.05. Calculating the statistic z and looking up the table to get the corresponding p value, the result is shown in Table 8. By comparing the statistic z and the p value (p (|Z| > 1.64) = 0.05), we can see that the absolute value of the z of the three pairs of samples is larger than 1.64 in all cases, with the corresponding p less than 0.05. Therefore, the null hypothesis is finally rejected, indicating significant differences among the Hm2 RMSEs and other model’s RMSEs. This result suggests that the Hm2 RMSE is improved significantly compared to that of other models at the 95% confidence level.
6. Conclusions

To overcome the drawbacks of existing $T_m$ models with poor universality, profound local accuracy loss, and difficulty in reflecting the nonlinear relationship between $T_m$ and meteorological parameters, this study used artificial neural network methods for constructing $T_m$ hybrid models in China. Validated with radiosonde $T_m$, the accuracies of the three hybrid models were 2.954 K, 2.703 K, and 2.763 K in terms of RMSE. In view of RMSE, compared with the UNB3m model, the accuracies were improved by 73.1%, 75.3%, and 74.8%; for the Bevis model, accuracies were optimized by 35.3%, 40.8%, and 39.5%; and for the GPT3 model, accuracies were improved by 32.0%, 37.8%, and 36.4%, respectively. Moreover, the hybrid models effectively weakened the spatiotemporal variation in the accuracy of the UNB3m model and achieved higher and more uniform accuracy in space and time. Among the three hybrid models, hm2 exhibited the best performance, followed by hm3. The models constructed in this study had better accuracy than the three international models. Moreover, the computer storage space occupied by the new models is significantly lower than that of the GPT3 model, and the number of parameters is substantially reduced. The accuracy improvement in the best hybrid model’s $T_m$ on its resultant PWV at four exemplary radiosonde stations and the whole study area were investigated using the PWVs through the radiosonde ZWD and $T_m$. The RMSEs of PWV derived from the Hm2 model are all less than 1 mm at exemplary radiosonde stations; the accuracy meets the needs of weather research. The overall error of the best hybrid model’s $T_m$ in the resultant PWV is smaller than that of Bevis, GPT3, and HGPT models by 33.9%, 36.4%, and 37.0% in terms of RMSE. The results of hypothesis testing further proved that this accuracy improvement of the best hybrid $T_m$ relative to the compared models is significant. The new models can be widely used to calculate high-precision $T_m$ and are more suitable for GNSS receivers without large storage space.

However, the study area was mainly conducted in China, and the global regions need to be further examined to validate the new models. In addition, meteorological parameters (surface temperature and water vapor pressure) were considered when constructing the model in this study. In future research, we hope to develop a globally hybrid model that can calculate the $T_m$ only based on geographical information (latitude, longitude, height) and temporal information (year, day of year (doy), hour of day (hod)).
Author Contributions: Data curation, L.L., L.H. (Liangke Huang), L.Z. and H.H.: formal analysis, M.C. and J.L.; investigation, J.L.; methodology, M.C. and J.L.; software, M.C. and J.L.; validation, M.C., J.L. and L.L.; writing—original draft preparation, J.L.; writing—review and editing, J.L., M.C., L.H. (Liangke Huang), L.Z. and L.H. (Ling Huang). All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Guangxi Natural Science Foundation of China (2020GXNSFBA297145), Foundation of Guilin University of Technology (GUTQDJ6616032), National Natural Science Foundation of China (42074035, 41874033, 42064002), China Postdoctoral Science Foundation (2019T120687, 2018M630880), the Fundamental Research Funds for the Central Universities (2042020k0009), Guangxi Science and Technology Base and Talent Project (No.ADJ924 5060), and Innovation Project of Guangxi Graduate Education (YCSW2022 322).


Acknowledgments: The authors would like to thank the University of New Brunswick (UNB) for providing the UNB3m codes, Vienna University of Technology for providing the GPT3 codes, and IGRA (Integrated Global Radiosonde Archive) for providing the radiosonde data. We would like to thank Mateus et al. for providing the HGPT codes.

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

References

Remote Sens. 2022, 14, 5762


21. Leandro, R.F.; Langley, R.B.; Santos, M.C. UNB3m_pack: A neutral atmosphere delay package for radiometric space techniques. *GPS Solut.* **2008**, *12*, 65–70. [CrossRef]


