

A Deep Learning–Based Regional Atmospheric Weighted Mean Temperature Model for China

Ruizhao Jiang[✉], Bo Li and Huizhong Zhu

School of Geomatics, Liaoning Technical University

47 Zhonghua Road, City of Fuxin, Liaoning, Liaoning, P.R. China, Postal code: 123000

✉ corresponding author

Abstract: To further enhance the accuracy of atmospheric weighted mean temperature (T_m) models in ground-based Global Navigation Satellite System (GNSS) retrieval of precipitable water vapor (PWV), we propose and develop a regionally adaptive Multi-Hidden-Layer neural network for T_m, hereafter referred to as MHL_T_m. A multiparameter cooperative T_m-modeling framework has been established using radiosonde observations from 65 launch sites across China during 2014–2018. We analyzed the nonlinear coupling between surface temperature (T_s), surface water-vapor pressure (e), latitude (Lat), elevation (H), and the temporal factor day of year (DOY) with radiosonde-derived integral T_m values. Radiosonde data from 2019 served as an independent reference to evaluate MHL_T_m's performance, which was then compared against the Bevis, GPT3, and Elastic Net models. Experimental results showed that the annual mean bias of MHL_T_m was −0.61 K, representing reductions of 30 % and 58 % relative to Bevis and GPT3, respectively, and slightly higher than Elastic Net (−0.11 K). The annual mean RMSE of MHL_T_m was 2.77 K, corresponding to improvements of 35 %, 62 %, and 18 % over Bevis, GPT3, and Elastic Net, respectively. Across different latitudinal and altitudinal zones in China, MHL_T_m exhibited superior accuracy and stability compared to Bevis, GPT3, and Elastic Net, demonstrating excellent regional applicability.

Keywords: MHL, atmospheric weighted mean temperature, machine learning, precipitable water vapor, GNSS.

1 Introduction

Precipitable water vapor (PWV) is a key physical parameter representing the total atmospheric water-vapor content. Its accurate monitoring is essential for deepening our understanding of climate-change mechanisms and improving early warning capabilities for extreme weather events [Rocken et al., 1997; Shi et al., 2023; Li et al., 2024; Ma et al., 2025; Jiang et al., 2025]. Variations in PWV are closely linked to cloud formation, precipitation processes, and extreme weather, making PWV an indispensable dataset for weather forecasting, climate-change studies, and hydrological-cycle analysis [Jiang et al., 2024; Jiang et al., 2024; Jiang et al., 2023]. Although the traditional observation techniques, such as radiosondes, sun photometers, water-vapor radiometers, and satellite remote sensing, can provide high-accuracy or large-scale PWV information, they are often constrained by temporal resolution, cost, and environmental conditions, and thus cannot satisfy the requirements for real-time, continuous, high-spatiotemporal-resolution monitoring [Kishore et al., 2011; Yao et al., 2022; Zhang et al., 2018]. Beyond its conventional applications in navigation, intelligent traffic management, and emergency rescue, a Global

Navigation Satellite System (GNSS) offers a unique advantage in that its microwave signals penetrate the atmosphere, rendering GNSS a low-cost, all-weather, high-spatiotemporal-resolution tool for atmospheric sensing [Ji & Shi, 2014; Sobrino & Romaguera, 2008; Vaquero-Martinez et al., 2008; Bevis et al., 1992]. Consequently, GNSS has become an indispensable technological support for meteorology and climatology research [Hagemann et al., 2003; Zhang et al., 2019; Liu et al., 2001]. In the GNSS-based retrieval of PWV, the zenith wet delay (ZWD) must be converted to PWV via a crucial water-vapor conversion coefficient (Π), for which the atmospheric weighted mean temperature (T_m) is the sole determining variable [Liu et al., 2001]. Therefore, considerable research has been devoted to developing and refining empirical and data-driven models for T_m .

Methods for calculating T_m can be divided into two main categories: models based on non-in situ meteorological parameters and models based on in-situ observations. The non-in situ models use atmospheric reanalysis data and therefore do not depend on local measurements. Böhm et al. [2007] first developed the Global Pressure and Temperature (GPT) model using ECMWF's ERA-40 reanalysis, employing ninth-order spherical harmonics to generate gridded surface temperature and pressure fields worldwide, thus establishing the meteorological foundation for GNSS water-vapor retrieval. Subsequent versions have further improved performance: GPT2w incorporates 37 wet-temperature layers from ERA-Interim and adds annual and semiannual harmonic terms to refine the lapse-rate calculation, reducing errors by 23 % in mid- to high-latitude and high-elevation regions (e.g., below 3000 m in China). Landskron et al. [2018] then introduced the GPT3 model, which integrates atmospheric-gradient information traced from numerical weather models and uses a $0.5^\circ \times 0.5^\circ$ grid to reduce T_m estimation errors to 1.47 K, with the successful validation over polar sea-ice regions. At the same time, the observation-based T_m models have trended toward regionalization and intelligent approaches. Bevis et al. [1992] established the classic

linear model that quantifies the relationship between surface temperature (T_s) and T_m . But this model exhibited significant limitations in complex terrain. The ET-TM model was proposed for China's complex climate regimes [Liu et al., 2001]. Based on data from 73 radiosonde stations, the region-specific T_m models were derived to incorporate surface temperature and saturation vapor pressure. The ET-TM model outperforms the Bevis and GPT-3 models overall and effectively mitigates the influences of latitude and elevation on T_m , making it well suited for PWV retrieval in China. Yao et al. [2014] systematically explored the global relationship between T_m and T_s and developed a latitude-dependent global T_m model. With the maturation of machine-learning methods, many researchers have widely applied these algorithms to modeling the atmospheric weighted mean temperature (T_m), owing to their strong function-fitting capabilities. Ding [2018] was the first to attempt a global T_m model using a backpropagation (BP) neural network and achieved the higher accuracy than the GPT2w model. Sun et al. [2021] used random forest (RF), generalized regression neural network (GRNN), and BP neural network (BPNN), respectively, to build T_m models for China that rely on measured meteorological inputs; their results showed that all three machine-learning methods outperformed the GPT3 model. In brief, constructing T_m models using machine-learning approaches is feasible.

Given that T_m is the cornerstone parameter in GNSS meteorology and reliable PWV retrieval demands high accurate T_m , traditional methods (e.g., the Bevis formula) face a challenge of their underlying linear assumptions and limited capacity to capture nonlinear effects in complex terrain. This motivates the introduction of deep - learning algorithms to elucidate the intricate relationships among meteorological variables and T_m .

2 Study Area and Data

2.1 Study area

The study area encompasses mainland China,

spanning latitudes 4° N to 53° N and longitudes 70° E to 135° E. It exhibits a pronounced east–west heterogeneity: from the low-lying southeastern coastal plains to the high-altitude northwestern interior plateaus, forming a multi-tiered geomorphological staircase. Climatically, the region is influenced by both monsoonal and continental systems, with the southeast being humid and rainy while the northwest remains arid and dry, and with the marked diurnal and seasonal temperature contrasts. The complex geo-climatic patterns in this region limit the applicability of global meteorological models (e.g., the GPT series and ERA5), particularly introducing systematic biases in estimates of moisture-related parameters such as T_m . Hence, the development of more appropriate T_m models is necessary. Fig. 1 shows the locations of 65 radiosonde stations in the study area.

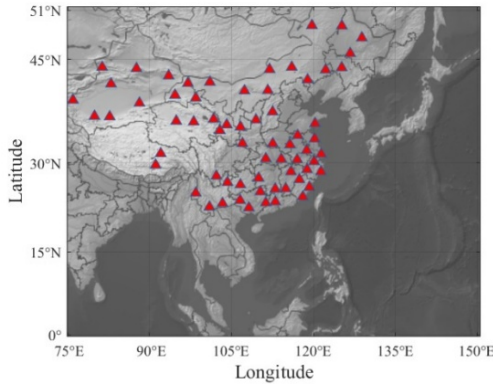


Fig. 1 Radiosonde Station Distribution Map

2.2 Data Source

A dataset comprising radiosonde measurements from 65 stations within the study area was selected, covering the five-year period from 2014 to 2018, with 282,143 soundings in total. Data from 2019 were reserved for the accuracy assessment. The raw observations (their station latitude and longitude, temperature, dew-point temperature, and relative humidity) were downloaded from the University of Wyoming Upper-Air Sounding Archive (<http://weather.uwyo.edu/upperair/sounding.html>).

The atmospheric weighted mean temperature was then computed through numerical integration. T_m as the temperature integral ratio weighted by water

vapor pressure is given as follows:

$$T_m = \frac{\int \frac{e}{T} dz}{\int \frac{e}{T^2} dz} \quad (1)$$

where e denotes the water-vapor pressure aloft at a station, T is the absolute temperature in K (Kelvin), and z is the vertical height above the station. In practice, the integral in (1) is discretized to

$$T_m = \frac{\sum_{i=1}^n \left(\frac{e_i}{T_i} \right) \Delta h_i}{\sum_{i=1}^n \left(\frac{e_i}{T_i^2} \right) \Delta h_i} \quad (2)$$

where e_i and T_i are, respectively, the mean water-vapor pressure and mean temperature of the i th atmospheric layer, and Δh_i is the thickness of the layer. The vapor pressure e at each level can be derived from the dew-point temperature T_d (in °C) provided by the radiosonde data [Kraus, 2004]:

$$e = 6.11 \times 10^{\left(\frac{7.5 \times T_d}{T_d + 237.7} \right)} \quad (3)$$

The weighted-mean based on radiosonde profiles in (2) is widely recognized as the most accurate method for computing T_m , and adopted in this study.

3 MHL_ T_m Model Construction in China

To further improve the accuracy of the T_m model, a regionalized T_m model (MHL_ T_m) incorporating a multi-hidden-layer neural network algorithm was developed. The traditional T_m models often suffer from limited local adaptability when applied to China's complex and highly variable climatic and topographic conditions. In contrast, the MHL algorithm, with its superior nonlinear fitting capability and robustness, provides an ideal solution for handling multidimensional and high-variable relationships. In comparison with the traditional models as well as other machine learning approaches, this study comprehensively evaluates the performance differences and applicable scenarios of various methods in term of predicting T_m . This section specifically describes the model selection:

3.1 Classical Models

The Bevis and GPT3 models were chosen as reference benchmarks because they have been widely used in the GNSS/atmospheric literature and serve as well-established, reproducible baselines for T_m estimation.

(1) Bevis Formula

Bevis's empirical formula serves as a representative traditional model, estimating T_m through a linear regression relationship with surface temperature (T_s) and is widely used in GNSS meteorology. Its principal advantages lie in its clear physical interpretation and low computational complexity on account of no complex parameter tuning, making it a standardized empirical tool in the early stage of the GNSS water-vapor retrieval. It provides a performance benchmark for subsequent analysis, facilitating verification of whether the data-driven models could achieve a significant improvement in the predictive accuracy [Bevis et al., 1992]:

$$T_m = A_0 + A_1 \cos \frac{2\pi DOY}{365.25} + B_1 \sin \frac{2\pi DOY}{365.25} + A_2 \cos \frac{4\pi DOY}{365.25} + B_2 \sin \frac{4\pi DOY}{365.25} \quad (5)$$

where DOY stands for the day of year (unit: day), A_0 is the annual mean T_m , A_1 and B_1 are the annual-cycle coefficients, A_2 and B_2 are the semiannual-cycle coefficients, and A_0, A_1, B_1, A_2, B_2 are in K [Landskron et al., 2018].

3.2 Machine Learning Models

(1) Elastic Net

Elastic Net regression combines L_1 and L_2 regularization to retain the feature-selection capability while mitigating multicollinearity (e.g., the seasonal correlation between T_s and DOY), wherein L_1 regularization adds a penalty equal to the sum of the absolute values of the model parameters to the loss function, whereas L_2 regularization adds a penalty equal to the sum of the squared model parameters in model training. If the linear multi-feature model significantly outperforms the Bevis formula, this would indicate that incorporating

$$T_m = 0.72 \times T_s + 70.2 \quad (4)$$

where T_m and T_s denote the weighted atmospheric mean temperature and surface temperature, respectively, in K .

(2) GPT3 Grid Model

Based on global meteorological reanalysis data, GPT3 grid model employs spatiotemporal interpolation to extend the discrete station observations onto a regular grid, and applies pressure-level height corrections and vertical-layer weighting to spatially extrapolate T_m at different altitudes. Its core advantage lies in overcoming the limitations of observational station distribution, providing reliable prior constraints for unobserved regions. Studies have indicated that when computing T_m over China, its results at a spatial resolution of $1^\circ \times 1^\circ$ outperformed those at $5^\circ \times 5^\circ$ [Cai, et al., 2022]. To highlight the superiority of the MHL_ T_m model, the GPT3-derived T_m results at $1^\circ \times 1^\circ$ resolution was used for model performance comparison:

additional predictors could enhance the predictive skill. Moreover, if it has limited performance, nonlinear methods may be employed to capture more complex dependencies.

Let the data $X \in R^{n \times p}$ denote the design matrix (n : the sample size, p : the number of predictors), $y \in R^n$ (the response vector), and $\beta \in R^p$ (the coefficient vector). The objective of the elastic net is to minimize the following expression:

$$\frac{1}{2n} \|y - X\beta\|_2^2 + \lambda \left(\alpha \|\beta\|_1 + \frac{1-\alpha}{2} \|\beta\|_2^2 \right) \quad (6)$$

where $\|\beta\|_1 = \sum_{j=1}^p |\beta_j|$ is the L_1 penalty and $\|\beta\|_2^2 = \sum_{j=1}^p \beta_j^2$ is the L_2 penalty. The tuning parameters λ (the regularization strength) and mixing parameter $\alpha \in [0, 1]$ determines the relative weight of the L_1 and L_2 penalties (degenerating to Lasso when $\alpha = 1$, and to Ridge when $\alpha = 0$) [Zou and Hastie, 2005].

(2) Multi-Hidden-Layer Neural Network (MHL)

A multi-hidden-layer neural network provides a fundamental deep-learning framework with strong theoretical capacity to approximate complex functional relationships, enabling exploration of deep-model potential in T_m prediction. However, due to the large parameter space and network complexity, sufficient model training can be time-consuming and sensitive to hyperparameter settings with a risk of becoming trapped in local minima. Consequently, an extensive hyperparameter tuning and a model validation are required to ensure robustness and generalization.

3.3 Construction of the MHL_ T_m Model

3.3.1 Data Processing

To the raw data, a data-cleaning pipeline was systematically applied, including missing-value handling, outlier detection and removal, and data normalization. Given the large sample size, missing values were removed outright. For outlier detection, a multivariate approach based on the residuals of a Ridge regression model was used, classifying samples with $|z\text{-score}| > 5$ as outliers. Employing a $|z\text{-score}|$ threshold of 5 is reasonable when the dataset is sufficiently large: a more lenient threshold ensures that only the most extreme samples are excluded, thereby preserving the original data distribution while removing anomalies that could disproportionately affect the model. Afterwards, the data were normalized using Z-score normalization:

$$z = (x - \hat{\mu}) / \hat{\sigma} \quad (7)$$

where x is the value of a given sample associated with a particular feature; $\hat{\mu}$ is its mean value computed from the dataset; $\hat{\sigma}$ is its standard deviation computed from the dataset; z is the standardized value (Z-score). So, all features have their samples standardized with their means of 0 and standard deviations of unity, preserving their original distribution characteristics. The standardization facilitates the model's ability to capture nonlinear relationships.

3.3.2 Feature Selection

To identify the model inputs, the correlations

between T_m and various meteorological parameters were first examined at the 65 radiosonde stations from 2014 to 2018. According to both statistical significance and physical rationale, surface temperature (T_s), water-vapor pressure (e), latitude (Lat), elevation (H), and day of year (DOY) were ultimately selected as the predictor variables for T_m .

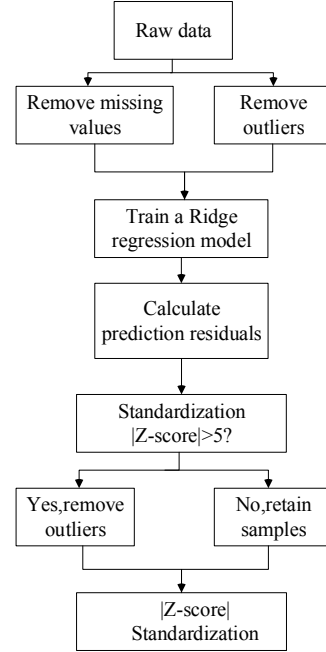


Fig. 2 Data Preprocessing Structure Diagram

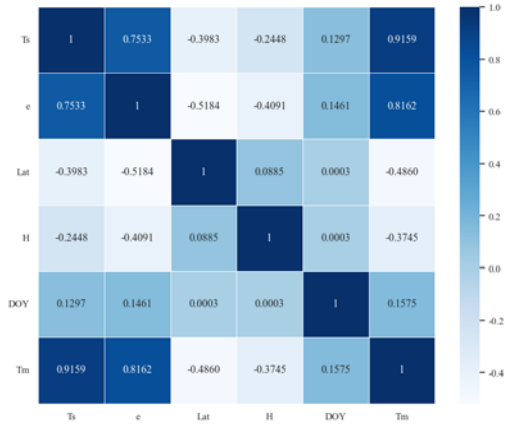


Fig. 3 Correlation Coefficient Plots of T_m with Various Features

Specifically, T_s ($r=0.9159$) and e ($r=0.8162$) exhibited a strong positive correlation with T_m , reflecting the direct driving roles of the surface thermal radiation and atmospheric moisture in near-surface thermodynamic processes, in accordance

with thermodynamic energy-transfer theory. Lat ($r=-0.486$) and H ($r=-0.3745$) show their negative correlations, revealing geographic spatial differentiation whereby T_m decreases significantly at high latitudes and elevations due to the reduced solar insolation and lower atmospheric pressure, and consistent with the vertical climatic zonation. Although DOY ($r=0.1575$) has a weak linear correlation with T_m, previous studies [Sun et al., 2019; Wang et al., 2016] have shown that T_m exhibited long-term diurnal variations, justifying its inclusion. Feature selection was based not only on Pearson correlation screening ($|r|>0.3$) but also on meteorological significance, to avoid loss of physical rationale inherent in the purely data-driven approaches.

3.3.3 Model Parameters

Hyperparameter optimization is used to systematically search the combinations of parameters that must be set prior to training in order to achieve the best generalization on validation data. Five-fold cross-validation (5-fold CV) splits the training set into five non-overlapping subsets and iteratively uses one fold as the validation set and the remaining four as the training set. Each hyperparameter configuration is thus evaluated multiple times under different splits, and its performance is estimated by the mean (and standard deviation) of the validation metric, which reduces sensitivity to a particular split. Based on this procedure, we performed hyperparameter optimization using 5-fold CV and chose the hyperparameter set with the smallest validation mean (and relatively small variance) as the optimal configuration. The final hyperparameters are briefed below. The network consisted of four hidden layers with 256, 128, 64 and 32 neurons, respectively, which was designed to progressively extract and integrate features so that the model could learn the complex patterns in the data and thus improve predictive performance. Hidden layers here used the Sigmoid activation function. The loss function was the mean squared error (MSE) to directly measure the discrepancy between predictions and observations. The used optimizer was the adaptive optimizer Adam

(Adaptive Moment Estimation), which could adaptively adjust each parameter's update step using the first- and second-moment estimates. The maximum number of the training epochs was set to 1000, but in practice training was stopped adaptively by early stopping based on validation performance to avoid overfitting and improve computational efficiency. To further guard against overfitting, L2 regularization was applied to enhance the generalization. A schematic diagram of the model architecture is shown in Figure 4.

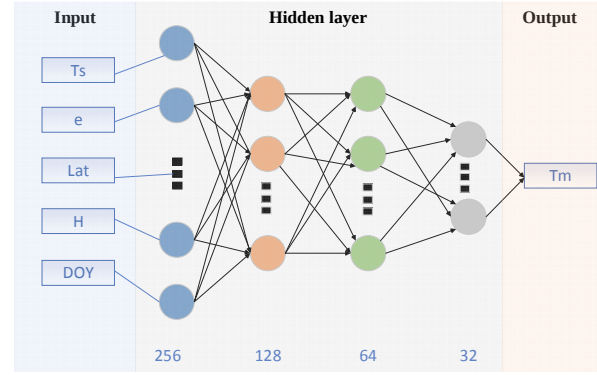


Fig.4 MHL_Tm Model Structure Diagram

4 Experimental Analysis

To assess the accuracy of the MHL_Tm model, the bias and RMSE were taken as evaluation metrics as follows:

$$bias = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i) \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (9)$$

wherein n is the number of the independent test samples, $\hat{y}_i$ is the T_m predicted by the model and y_i is the reference value.

4.1 MHL_Tm Model Accuracy Analysis

To assess the accuracy of MHL_Tm in computing T_m over China, the radiosonde-derived T_m values for the year 2019 obtained by integration were used as the reference, and the bias and RMSE of the Bevis, GPT3, Elastic Net, and MHL_Tm models were compared (Table 1).

Tab. 1 The Accuracy Comparison of T_m from

Different Models		Bevis	GPT3	Elastic Net	MHL_Tm
Bias /K	Max. value	6.46	1.77	1.85	0.21
	Min. value	-3.93	-5.45	-2.06	-2.16
	Average value	0.87	-1.46	-0.11	-0.61
RMSE /K	Max. value	7.33	12.82	4.52	3.85
	Mini. value	2.28	3.03	2.23	1.86
	Average value	4.27	7.25	3.37	2.77

(1) Bias Analysis

From the bias performance perspective, the differences between the maximum and minimum biases with individual models were: 10.39 K (Bevis), 7.22 K (GPT3), 3.91 K (Elastic Net), and 2.37 K (MHL_Tm). Apparently, MHL_Tm had the smallest fluctuation in the bias prediction. The bias fluctuation

can serve as an indicator of model robustness. The relatively narrow bias fluctuation of MHL_Tm suggested that its error distribution was more uniform across different samples, which is particularly advantageous for applications requiring stable predictive performance. In terms of the mean bias, the GPT3 model exhibited a significant negative bias of -1.46 K, the Bevis model showed a positive bias tendency of 0.87 K, as MHL_Tm had a mean bias of -0.61 K, demonstrating strong performance. Compared with the GPT3 model, which had the largest bias, the bias magnitude of MHL_Tm was reduced by 58.22 %.

To visually illustrate the distribution of prediction biases across models, a representative station (28° N, 102° E, elevation $\approx$ 1600 m, sample size $\geq$ 700) was chosen for the bias distribution analysis (Fig. 5).

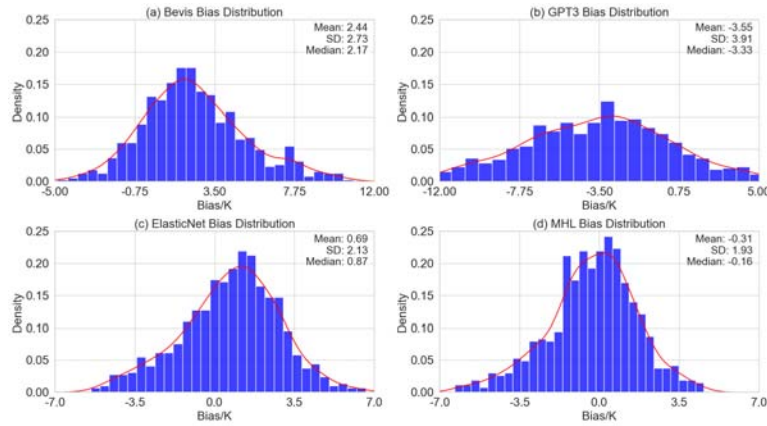


Fig. 5 Bias Distributions of Different Models at a station (28°N, 102°E, elevation $\sim$ 1600m, sample size $\geq$ 700)

Unlike the overall mean-bias metric, the single-station case study can reveal the distribution of prediction biases, facilitating an in-depth assessment of model performance under local climatic conditions. As shown in Fig. 5, the MHL_Tm model has the lowest mean bias, -0.31 K and a moderate dispersion with the standard deviation of 1.93 K, indicating that its predictions are both stable and close to the expectation. Specifically, the distribution curve from MHL_Tm was relatively symmetric, with its peak located near zero, demonstrating that the majority of predictions closely matched the observations.

(2) RMSE Analysis

According to the root-mean-square error (RMSE), the ranking of models by their average prediction error is: GPT3 (7.25 K) > Bevis (4.27 K) > Elastic Net (3.37 K) > MHL_Tm (2.77 K). MHL_Tm achieved the best predictive accuracy, representing a 35.13 % reduction in error relative to the traditional Bevis model and a 61.79 % reduction compared to the GPT3 model. Notably, MHL_Tm also excelled in controlling extreme error magnitude: its maximum RMSE (3.85 K) is lower than those of Bevis (7.33 K) and Elastic Net (4.52 K) whilst its minimum RMSE (2.77 K) is near the optimal value, indicating that the

model maintains robust performance even under extreme scenarios. Box plots of the RMSE for each model are shown in Fig. 6.

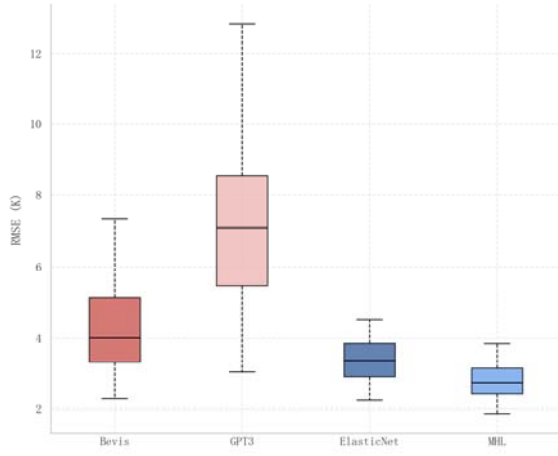


Fig. 6 Boxplot of RMSE for Different Models

To further analyze the accuracy of MHL_Tm across China, the annual mean bias and RMSE at each radiosonde station were compared among the four models (Fig. 7).

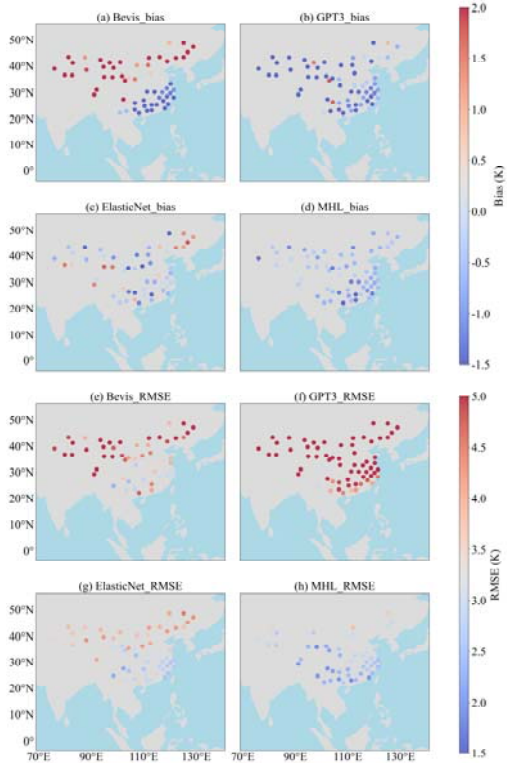


Fig. 7 Distribution of Bias and RMSE of Different Models across China

From Fig. (a)–(d), the Bevis model exhibits positive bias in northern regions and smaller or even

negative bias in southern regions, indicating pronounced regional differences. The GPT3 model shows an overall large negative bias, particularly in the north and west, revealing a tendency to systematically underpredict. In contrast, the Elastic Net and MHL_Tm models have biases close to zero with more uniform spatial distributions, demonstrating higher predictive accuracy and stability across the country.

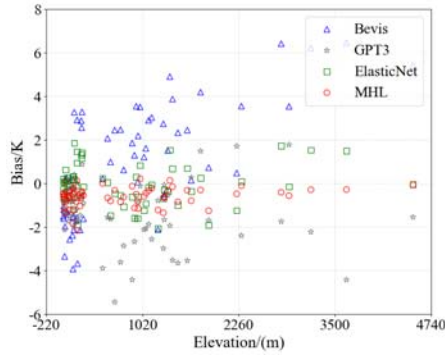
From Fig. 7e – 7h, the Bevis model had its RMSE generally high in northern China (especially North China and the Northeast) and relatively low in the south, indicating poorer adaptability at high latitudes. The GPT3 model exhibited a high RMSE almost nationwide, reflecting a suboptimal performance in geographically complex regions. The Elastic Net model showed the elevated RMSE in eastern and southern areas, particularly along the middle and lower Yangtze River region, while having performed better in the Southwest and Northwest, suggesting its limitations in capturing regional climate variability. The MHL model, however, had its RMSE distribution more balanced with low errors in central and most southern regions and only slightly larger in some northern areas, indicating an overall stable performance and strong generalization ability in the face of geographic complexity.

4.2 Effects of Elevation and Latitude on the MHL_Tm Model

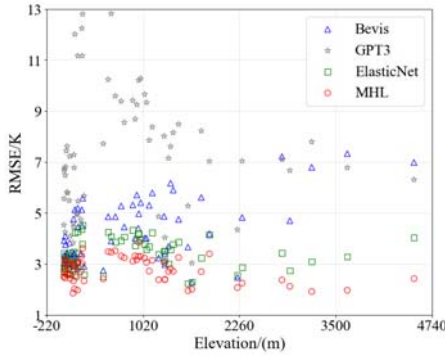
Significant differences in predictive accuracy and error distribution are evident among models. The traditional empirical models (e.g., Bevis) and large-scale models (e.g., GPT3) exhibited biases under complex conditions, whereas, however, the deep-learning models (e.g., MHL_Tm) performed more stably nationwide. To further explore the sources of these model errors, it is necessary to analyze the trends in error variation subject to geographic conditions vs. latitude and elevation.

(1) Effect of Elevation on the MHL_Tm Model

The scatter plots of bias and RMSE values versus the elevation associated with different models are shown in Fig. 8:



(a) Bias distribution



(b) RMSE distribution

Fig. 8 Distribution of Bias and RMSE with respect to Altitude

In the low-altitude regions (< 1000 m), the MHL model exhibited a higher prediction-error variability (mean RMSE of 2.87 K) and a slight underestimation (mean bias of -0.54 K). This is mainly due to the extreme environmental complexity in these areas (such as plains, coastal zones, and urban centers), dense human activities (e.g., urban heat islands), highly variable local circulations, heterogeneous surface types (water bodies, vegetation, buildings), and the strong spatial inhomogeneity of water vapor, all of which substantially increase the difficulty of modeling atmospheric weighted mean temperature (T_m), resulting in a higher uncertainty and volatility in predictions. By contrast, in high-altitude regions (≥ 1000 m), especially in very high-altitude areas (> 3000 m), the prediction errors decreased markedly (mean RMSE drops to 2.59 K, and to 2.11 K in the very high-altitude sites) with an improved consistency. This improvement stems from the relative simplicity and stability of mountain and

plateau environments, minimal anthropogenic interference, clearer atmospheric boundary-layer structures, low and more uniformly distributed water-vapor content, and meteorological processes (such as lapse rates) less affected by local complexities, making it easier for modeling the governing physical laws and thus yielding the more stable and accurate predictions, although the scarcity of very high-altitude stations may introduce a slight systematic estimation (0.13 K). The detailed results are shown in Table 2.

(2) Effect of Latitude on the MHL_ T_m Model

The scatter plots of bias and RMSE values versus latitudes associated with different models are shown in Fig. 9.

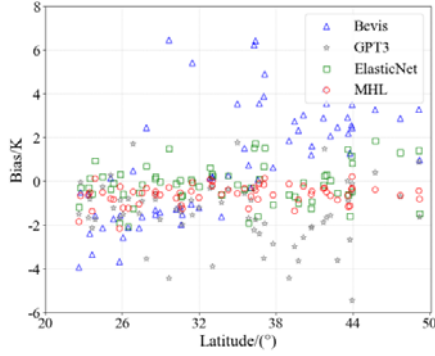
The latitude-induced differences in model performance stem primarily from the complexity of the climatic systems and the underlying data availability. In the low-latitude regions ($< 30^\circ$), the model achieved its lowest prediction error (mean RMSE 2.31 K), benefitting from the relatively uniform and stable monsoonal climate characteristics of the southern subtropical/tropical zone—small annual temperature range, abundant moisture, and strongly regular seasonal variability. This “mild” climate regime allowed models trained on large-scale climatic features to obtain accurate and stable predictions more easily. However, as the latitude increases, the climatic complexity intensifies, and the model’s prediction error rises markedly (mean RMSE 3.20 K) particularly in high-latitude regions ($> 40^\circ$). This is mainly attributable to the pronounced temperate continental monsoon climate, large annual temperature swings, the stark contrast between extremely cold, dry winters and hot, humid summers, and highly variable weather systems (e.g., cold waves, heavy rainfall), all of which greatly complicated the modeling process. Additionally, compared with the lower latitudes, the meteorological stations at the high latitudes are typically sparser, leading to an insufficient training-data coverage and hindering the model’s ability to learn and accurately represent these complex, extreme, and highly localized atmospheric processes, thereby significantly

increasing the prediction uncertainty and error. The

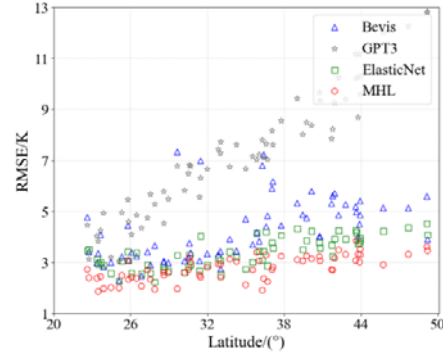
detailed results are shown in Table 3.

Tab. 2 Statistical summary of the effect of elevation on different models

Elevation/m	Bias/K				RMSE/K			
	MHL	ElasticNet	GPT3	Bevis	MHL	ElasticNet	GPT3	Bevis
<1000	-0.54	-0.11	-1.39	-0.11	2.87	3.39	7.477	3.97
1000~3000	-0.32	-0.40	-1.40	2.12	2.59	3.33	6.77	4.45
>3000	0.13	0.77	-2.74	6.03	2.11	3.40	6.96	7.03



(a) Bias distribution



(b) RMSE distribution

Fig. 9 Distribution of Bias and RMSE of Different Models with respect to Latitude

Tab. 3 Statistical summary of the effect of altitude on different models

Latitude/°	Bias/K				RMSE/K			
	MHL	ElasticNet	GPT3	Bevis	MHL	ElasticNet	GPT3	Bevis
<30	-0.62	-0.18	-1.23	-1.17	2.31	2.91	4.64	3.47
30~40	-0.33	-0.19	-1.69	1.32	2.79	3.33	7.27	4.41
>40	-0.39	-0.09	-1.38	2.57	3.20	3.97	10.18	4.96

5 Conclusions

This study introduced a multilayer-perceptron neural-network algorithm into modeling of Tm to address the inadequate representation of its spatial nonuniformity by traditional empirical models under complex geographical and climatic conditions. A multiparameter cooperative Tm-modeling framework was constructed. Accordingly, a regionally adaptive MHL_Tm model was proposed, which were compared with the Bevis, GPT3, and Elastic Net models. The results are summarized here below:

(1) The annual mean bias of MHL_Tm was -0.61 K, representing the reductions of 29.89 % and 58.22 % relative to Bevis and GPT3, respectively, and a slight increase compared to Elastic Net (-0.11 K). The annual mean RMSE was 2.77 K,

corresponding to the improvements of 35.13 %, 61.79 %, and 17.80 % over Bevis, GPT3, and Elastic Net, respectively.

(2) From an elevation perspective, MHL_Tm's performance varies significantly with respect to altitude. In the low-altitude regions (< 1000 m), the model's mean bias was -0.54 K and mean RMSE was 2.87 K, indicating a systematic underestimation and a larger error variability. By contrast, at mid-altitudes (1000–3000 m), the mean bias decreased to -0.32 K whilst the mean RMSE dropped to 2.59 K, showing improved prediction accuracy and reduced error range. In the high-altitude regions (> 3000 m), the mean bias further decreased to 0.13 K and the mean RMSE down to 2.11 K. This likely reflects the simpler, more consistent climatic conditions and stronger regularity in data distribution at high

elevations, which facilitate more accurate modeling, although the relatively small sample size at these altitudes may affect the statistics.

(3) From a latitudinal perspective, MHL_Tm's performance also shows the marked differences. In the low-latitude regions ($< 30^\circ$), the mean bias was -0.62 K with the mean RMSE of 2.31 K, indicating a moderate negative bias (predictions generally below observations) but relatively small error range. In the mid-latitude regions ($30^\circ\text{--}40^\circ$), the mean bias approached zero (-0.33 K) with the mean RMSE slightly increased to 2.79 K, demonstrating a stable predictive performance and good data-fit within this band. However, in the high-latitude regions ($> 40^\circ$), the mean bias raised to -0.39 K with the mean RMSE up to 3.20 K, signifying the larger prediction errors and the poorer adaptability, likely due to the more complex and variable climates, frequent extreme weather events, and uneven sample distribution. Overall, as the latitude increased, MHL_Tm's prediction errors first rose modestly and then increased significantly, reflecting varying adaptability across climate zones: strong performance at low latitudes versus great challenges at high latitudes.

References

- Bevis M, Businger S, Herring T A, et al (1992): GPS meteorology: Remote sensing of atmospheric water vapor using the global positioning system, *Journal of Geophysical Research: Atmospheres*, 1992, 97(D14): 15787-15801.
- Böhm J, Heinkelmann R, and Schuh H.(2007): Short note: a global model of pressure and temperature for geodetic applications, *Journal of Geodesy*, 2007, 81(10): 679-683.
- Cai M, Liu L, Huang L, et al. (2022): Accuracy evaluation of precipitable water vapor retrieved by GPT3 model from GNSS, *Journal of Geodesy and Geodynamics*, 2022, 42(5): 451-456.
- Ding M. (2018): A neural network model for predicting weighted mean temperature, *Journal of Geodesy*, 2018, 92(10): 1187-1198.
- Hagemann S, Bengtsson L, Gendt G. (2003): On the determination of atmospheric water vapor from GPS measurements, *Journal of Geophysical Research: Atmospheres*, 2003, 108(D21).
- Ji D, Shi J. (2014): Water vapor retrieval over cloud cover area on land using AMSR-E and MODIS[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, 7(7): 3105-3116.
- Jiang C, Chen S, Gao X, et al. (2015): A direct conversion model from ZTD to PWV based on the random forest, *Measurement Science and Technology*, 2025, 36(2): 025802.
- Jiang C, Chen S, Wang S, et al. (2014): A grid model of direct conversion between zenith tropospheric delay and precipitable water vapor in tropical regions[J]. *GPS Solutions*, 2024, 28(3): 127.
- Jiang C, Chen S, Xu T, et al. (2014): An improved global Tm stratification model for GNSS-PWV retrieval, *Advances in Space Research*, 2024, 74(3): 1225-1237.
- Jiang C, Gao X, Wang S, et al. (2023): Comparison of ZTD derived from CARRA, ERA5 and ERA5-Land over the Greenland based on GNSS, *Advances in Space Research*, 2023, 72(11): 4692-4706.
- Kishore P, Ratnam M V, Namboothiri S P, et al. (2011): Global (50 S–50 N) distribution of water vapor observed by COSMIC GPS RO: Comparison with GPS radiosonde, NCEP, ERA-Interim, and JRA-25 reanalysis data sets, *Journal of Atmospheric and Solar-Terrestrial Physics*, 2011, 73(13): 1849-1860.
- Kraus H. (2004): *Die Atmosphäre der Erde: eine Einführung in die Meteorologie*, Springer, 2004.
- Landskron D, Böhm J. (2018): VMF3/GPT3: refined discrete and empirical troposphere mapping functions, *Journal of geodesy*, 2018, 92(4): 349-360.
- Li J, Wang Y, Liu L, et al. (2024): A grid model for vertical correction of precipitable water vapor over the Chinese mainland and surrounding areas using random forest[J]. *Geoscientific Model Development*, 2024, 17(7): 2569-2581.
- Liu Y, Teng Y, Van Hove T, et al. (2001): Comparison of precipitable water observations in the near tropics by GPS, microwave radiometer, and radiosondes, *Journal of Applied Meteorology*, 2001, 40(1): 5-15.

Ma X, Wang X, Yao Y, et al. (2025): Enhanced multi-GNSS precise point positioning based on ERA5 precipitation water vapor information, *Journal of Geodesy*, 2025, 99(3): 1-16.

Rocken C, Anthes R, Exner M, et al. (1997): Analysis and validation of GPS/MET data in the neutral atmosphere, *Journal of Geophysical Research: Atmospheres*, 1997, 102(D25): 29849-29866.

Shi Y, Liu L, Lan S, et al. (2023): A PSO-BP neural network-based model for weighted mean temperature in China, *Journal of Geodesy and Geodynamics*, 2023, 43(12): 1300-1306.

Sobrino J A, Romaguera M. (2008): Water-vapour retrieval from Meteosat 8/SEVIRI observations, *International journal of remote sensing*, 2008, 29(3): 741-754.

Sun Z, Zhang B, Yao Y. (2019): An ERA5-based model for estimating tropospheric delay and weighted mean temperature over China with improved spatiotemporal resolutions[J]. *Earth and Space Science*, 2019, 6(10): 1926-1941.

Sun Z, Zhang B, Yao Y. (2021): Improving the estimation of weighted mean temperature in China using machine learning methods, *Remote Sensing*, 2021, 13(5): 1016.

Vaquero-Martínez J, Antón M, de Galisteo J P O, et al. (2017): Validation of MODIS integrated water vapor product against reference GPS data at the Iberian Peninsula, *International journal of applied earth observation and geoinformation*, 2017, 63: 214-221.

Wang X, Zhang K, Wu S, et al. (2017): Determination of zenith hydrostatic delay and its impact on GNSS-derived integrated water vapor[J]. *Atmospheric measurement techniques*, 2017, 10(8): 2807-2820.

Yao Y, Zhang B, Xu C, et al. (2014): Analysis of Tm-Ts Correlation and Construction of a Global Latitude-Related Linear Model, *Chinese Science Bulletin*, 2014, 59(09): 816-824.

Yao Y, Zhao Q. (2022): Research progress and prospect of monitoring tropospheric water vapor by GNSS technique, *Acta Geodaetica et Cartographica Sinica*, 2022, 51(6): 935.

Zhang Q, Ye J, Zhang S, et al. (2018): Precipitable

water vapor retrieval and analysis by multiple data sources: Ground-based GNSS, radio occultation, radiosonde, microwave satellite, and NWP reanalysis data, *Journal of Sensors*, 2018, 2018(1): 3428303.

Zhang W, Zhang H, Liang H, et al. (2019): On the suitability of ERA5 in hourly GPS precipitable water vapor retrieval over China, *Journal of Geodesy*, 2019, 93(10): 1897-1909.

Zou H, Hastie T. (2005): Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 2005, 67(2): 301-320.

Authors



Ruizhao Jiang is a candidate for Master's degree at School of Surveying and Geoinformatics, Liaoning Technical University. His main research interests include GNSS high-precision positioning and GNSS meteorology.



Dr. **Bo Li** is a Lecturer at School of Surveying and Geoinformatics, Liaoning Technical University. His main research is focused on high-precision BeiDou/GNSS positioning inclusive of server-side multi-frequency UPD and ionospheric/tropospheric products, RTK, PPP, PPP-AR, PPP-RTK technologies, and trustworthy evaluation of high-precision positioning algorithms.



Huizhong Zhu is a Professor and Dean, School of Surveying and Geoinformatics, Liaoning Technical University. His main research interests include high-precision BDS/GPS positioning algorithms and data processing; research and applications of BDS/GPS network RTK algorithms.