Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 21, No. 1, 2026

A weighted average temperature model for the Yunnan-Guizhou region considering multiple factors

Xue Wang

School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo, Henan 454003, China

Abstract: Addressing the issue of complex terrain in the Yunnan-Guizhou region and the insufficient applicability of existing weighted mean temperature (Tm) models, this paper constructs a multi-factor, multivariate nonlinear Tm empirical model (YGTm) based on data from six sounding stations spanning 2010 to 2019. Internal consistency, external consistency, and leave-one-out cross-validation were employed to compare the model with Bevis, GPT2w, and HGPT2 models. The results indicate that the YGTm model exhibits a root mean square error (RMS) of 2.52 K, representing a 15%–39% improvement in accuracy compared to the comparative models; the relative error of atmospheric precipitable water vapor (PWV) retrieved based on YGTm is only 0.85% on average. This model can provide high-precision Tm support for GNSS water vapor retrieval in the Yunnan-Guizhou region.

Keywords: Weighted average temperature; Yunnan-Guizhou region; GNSS water vapor retrieval; Multivariate nonlinear regression.

1. Introduction

Although water vapor is present in small amounts in the atmosphere, it plays an extremely important role in meteorological monitoring. It is a key influencing factor for various severe weather events and also has a significant impact on climate regulation. With the development of GPS technology, remote sensing water vapor technology has continuously advanced. Compared to conventional observation methods, GPS remote sensing water vapor technology has advantages such as high spatial coverage, all-weather observation, high accuracy, and fewer limited conditions, especially for the observation of atmospheric water vapor, which is notably effective[1]. To convert zenith wet delay (ZWD) into atmospheric precipitable water vapor (PWV), it is necessary to apply the atmospheric weighted mean temperature of water vapor (Tm). Therefore, accurately determining the value of Tm becomes very important. The accuracy of Tm directly affects the inversion accuracy of PWV, and improving the accuracy of Tm is the core of enhancing the inversion accuracy of precipitable water vapor.

In recent years, many scholars have conducted research on atmospheric weighted mean temperature models in different regions[2]. Among them, the most widely used is the univariate linear expression between Tm and ground temperature Ts established by Bevis et al. using linear regression. Li Haojie et al. [2]established a relationship model between Tm and ground temperature, vapor pressure, elevation, and latitude through the analysis of data from 84 sounding stations in China. Wang Hanhong et al. [3]applied linear regression to sounding data from the Lanzhou region and established single-factor and dual-factor Tm refinement models suitable for the region. Mo Zhixiang et al.[4] constructed a Tm model for the Qinghai-Tibet Plateau based on data from 13 sounding stations, considering ground temperature, altitude, latitude, and seasonal variations. Wei Haifu et al. [5]modeled the Chinese region, taking into account the impact of relative elevation on atmospheric weighted mean temperature. Xu Mingze et al. [6]established atmospheric weighted mean temperature models in the Jinan region, Fan Shijie et al.[7]in the Qingdao region, Zheng Lei[8] in the southwest region, Shi Lingfan et al.[9]in the Qinghai-Tibet Plateau region, Zhu Hai et al.[10]in Shaanxi, Ke Liping et al.[11]in Guizhou Province, Cui Jinye et al.[12]in Hong Kong, and He Shiwei et al.[13]in Anhui.

However, the applicability of existing models in complex terrain areas remains inadequate. Located in the southwest of China, the Yunnan-Guizhou region features steep topographical variations, diverse climatic types, and frequent extreme weather events. The accuracy and stability of existing models in this region urgently need to be improved. To this end, based on the measured data from six sounding stations in the Yunnan-Guizhou region from 2010 to 2019, this paper systematically analyzes the spatiotemporal evolution of Tm and its key influencing factors, constructs a multivariate nonlinear Tm empirical model (YGTm model) suitable for this region, and evaluates its accuracy and generalization ability through various verification methods, aiming to provide more reliable Tm input for regional GNSS water vapor inversion.

2. Data sources and processing

2.1. Data source

This article selects sounding data from six sounding stations in the Yunnan-Guizhou region spanning from 2010 to 2018 for Tm modeling, and utilizes sounding data from 2019 for accuracy verification. The sounding data can be freely downloaded from the website of the University of Wyoming (http://www.weather.uwyo.edu/ upperair/ sounding.html), with a temporal resolution of 12 hours. It primarily includes information such as the station's altitude, longitude and latitude, pressure, vapor pressure, absolute temperature, etc. The distribution and relevant information of each station are shown in Figure 1

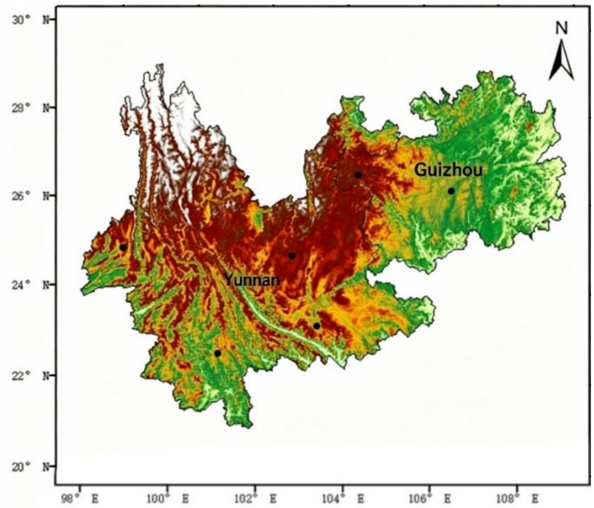

Figure 1. Schematic diagram of the distribution of sounding stations in Yunnan-Guizhou region

2.2. Data Processing

The atmospheric water vapor PWV retrieved from GNSS can be calculated by multiplying the water vapor conversion factor by the zenith wet delay (ZWD):

$$PWV = \Pi \cdot ZWD \tag{1}$$

$$\Pi = \frac{10^6}{\rho_\omega R_v[(k_3/T_m + k_2^{'})]} \tag{2}$$

In the formula: ρw=1×103 kg/m3 represents the density of liquid water; Rv=461.495J·kg-1·k-1 denotes the gas constant of water vapor; k2' and k3 are atmospheric physical parameters, with typical empirical values of k2'=22.13±2.20K/hPa and k3=(3.739±0.012)×105 K/hPa, respectively.

The atmospheric weighted average temperature Tm can be obtained by numerical integration of the water vapor pressure and atmospheric temperature above the measurement station, and its calculation formula is as follows:

$$T_m = \frac{\int \frac{e}{T} dz}{\int \frac{e}{T^2} dz} \tag{3}$$

In the formula, e represents the water vapor pressure (hPa) above the measurement station, T denotes the surface temperature (K), and Z stands for the geopotential height. In practical applications, the above formula is usually discretized to obtain:

$$T_m = \frac{\sum_{i=1}^{n} \left(\frac{e_i}{T_i}\right) \Delta h_i}{\sum_{i=1}^{n} \left(\frac{e_i}{T_i^2}\right) \Delta h_i} \tag{4}$$

In the formula, ei represents the average water vapor pressure of the ith layer of the atmosphere, and Ti represents the average temperature of the ith layer of the atmosphere. The water vapor pressure e in the formula can be obtained through the following formula:

$$e = \frac{RH \cdot e_s}{100} \tag{5}$$

$$e_s = 6.11 \times 10 \left(\frac{7.5 \times T_d}{T_d + 237.7}\right) \tag{6}$$

In equations (5) and (6), es represents the saturated vapor pressure, RH stands for relative humidity, and Td denotes the dew point temperature.

3. Correlation analysis and model establishment

3.1. Correlation analysis

As shown in Figure 2, the Tm time series from six sounding stations in the Yunnan-Guizhou region from 2010 to 2019 exhibit significant annual and semi-annual cycles, generally exhibiting higher values in summer and lower values in winter. This is closely related to its subtropical monsoon climate characteristics: in summer, influenced by warm and humid air currents, rising temperatures drive Tm to peak; in winter, under the control of dry and cold air, decreasing temperatures cause Tm to fall to low values.

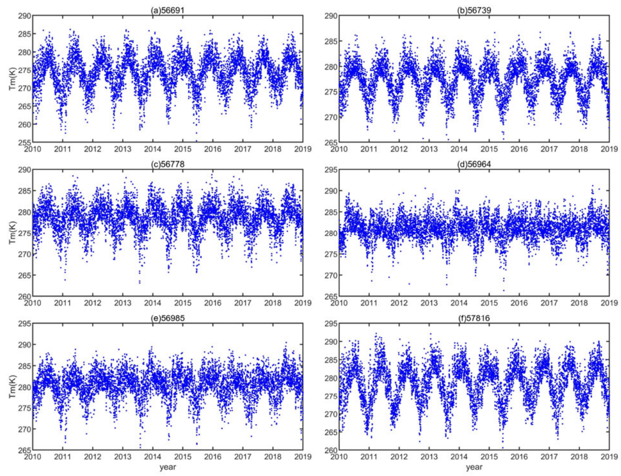

Figure 2. Tm time series plot

Research indicates that using the natural logarithm of vapor pressure, ln(e), yields higher correlation and accuracy than directly using vapor pressure e to construct a weighted average temperature Tm model. The analysis of data from six sounding stations in the Yunnan-Guizhou region spanning 2010 to 2019 reveals a significant positive correlation between Tm and ln(e) at each station, with an average correlation coefficient of 0.62, indicating a strong correlation between the two in the Yunnan-Guizhou region.

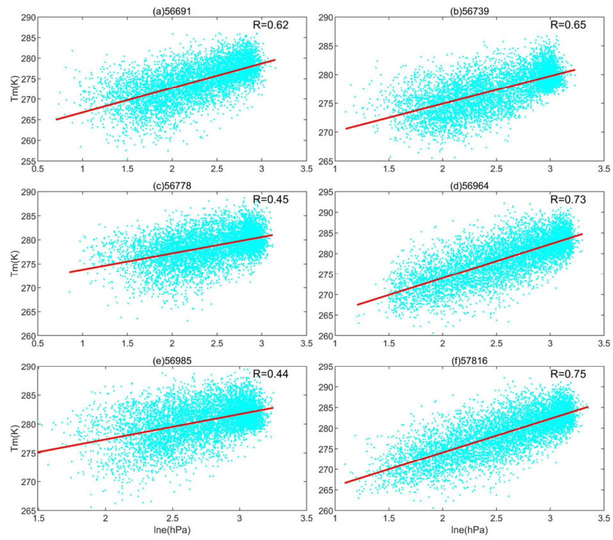

Figure 3.: Correlation between Tm and the natural logarithm of water vapor pressure at the measurement station, lne

To further compare the impact of different meteorological parameters on Tm, this paper analyzes the correlation between Tm and ground temperature Ts in the Yunnan-Guizhou region (Figure 4). The results show that there is a strong linear positive correlation between Tm and Ts at all six stations, with a mean correlation coefficient of 0.81 and a maximum of 0.90. This indicates that Ts can serve as a key parameter for constructing an empirical model of Tm in the region.

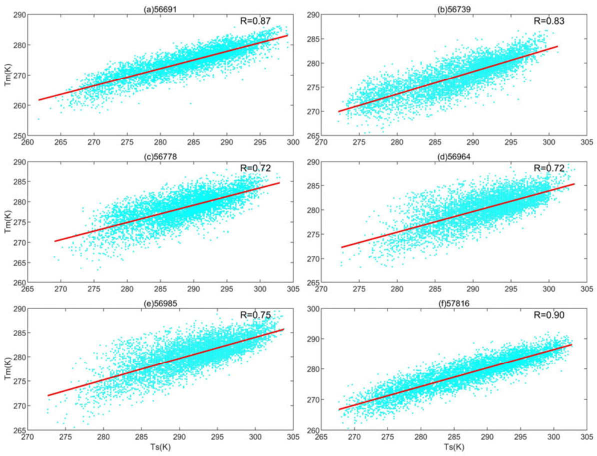

Figure 4. Correlation between Tm and temperature at the measurement station Ts

3.2. Model establishment

Based on the above analysis, it can be observed that the weighted average temperature Tm exhibits characteristics of annual and semi-annual cycles, and has a high correlation with surface temperature Ts and the natural logarithm of vapor pressure ln(es). Therefore, to construct a localized Tm model that adapts to the atmospheric characteristics of the Yunnan-Guizhou region, this paper comprehensively considers the aforementioned key influencing factors and establishes the following multivariate nonlinear empirical model:

$$YGTm = A_0 + A_1 T_s + A_2 \ln es + A_3 \Phi + A_4 H + A_5 \cos\left(2\pi \frac{DOY}{365.25}\right) + A_6 \sin\left(2\pi \frac{DOY}{365.25}\right) + A_7 \cos\left(4\pi \frac{DOY}{365.25}\right) + A_8 \sin\left(4\pi \frac{DOY}{365.25}\right) \tag{7}$$

In the formula: A0, A1, A2, A3, A4, A5, A6, A7, A8 are model coefficients, Ts is ground temperature, es is water vapor pressure, Φ is latitude, H represents the altitude of the station, and DOY is the accumulated day of the year. The coefficients in the model can be solved using the least squares method, and by fitting the sounding data from five modeling stations in Yunnan-Guizhou region from 2010 to 2017, the localized YGTm model coefficients for Yunnan-Guizhou region can be obtained. The specific results are shown in Table 1. When applying this model in practice, it is only necessary to input the ground temperature, water vapor pressure, altitude, latitude, and accumulated day of the year of the target station to quickly estimate the weighted average temperature Tm at the corresponding location.

Table 1. YGTm model coefficients calculated using sounding data from Yunnan-Guizhou region from 2010 to 2017

Model coefficient	A0	A1	A2	A3	A4	A5	A6	A7	A8
fitted value	131.244	0.493	2.020	0.092	-0.002	1.108	0.100	0.122	-0.135

4. Model accuracy verification

To verify the accuracy of the YGTm model, Tm obtained from sounding station data in 2019 was selected as the reference value for precision testing of the YGTm model. Bias, mean absolute error (MAE), and root mean square error (RMS) were selected as evaluation metrics, with the specific formulas as follows:

$$Bias = \frac{1}{N}\sum_{i=1}^{N}\left(X_m^{M_i} - X_m^{R_i}\right) \tag{8}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|X_m^{M_i} - X_m^{R_i}\right| \tag{9}$$

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(T_m^{M_i} - T_m^{R_i}\right)^2} \tag{10}$$

In the formula, $X_m^{M_i}$ represents the calculated value of the Tm model, $X_m^{R_i}$ represents the reference value of Tm, and N represents the number of samples.

4.1. Internal consistency accuracy verification

Internal consistency verification is used to test the prediction accuracy and temporal stability of the model at the modeling sites. In this paper, data from 5 modeling sites spanning 2010-2017 are used for modeling, and data from 2018-2019 are used for verification. Taking the radiosonde integral Tm as the true value, the Bevis, GPT2w, HGPT2, and YGTm models are employed to predict Tm for each site. By comparing the predicted values with the true values, the annual average bias (Bias) and root mean square error (RMS) are calculated. The results are presented in Table 2.

Table 2. Analysis of conformity accuracy within the regional model

model	Bias/K			RMS/K		
	Max	Min	Avg	Max	Min	Avg
Bevis	1.22	-2.77	-0.94	4.42	2.28	3.43
GPT2w-1	2.31	-1.28	0.82	3.84	2.52	3.38
GPT2w-5	4.33	-1.94	1.80	5.52	3.64	4.16
HGPT2	1.96	-1.03	0.59	3.85	2.14	2.94
YGTm	0.83	-0.92	0.28	2.76	2.07	2.34

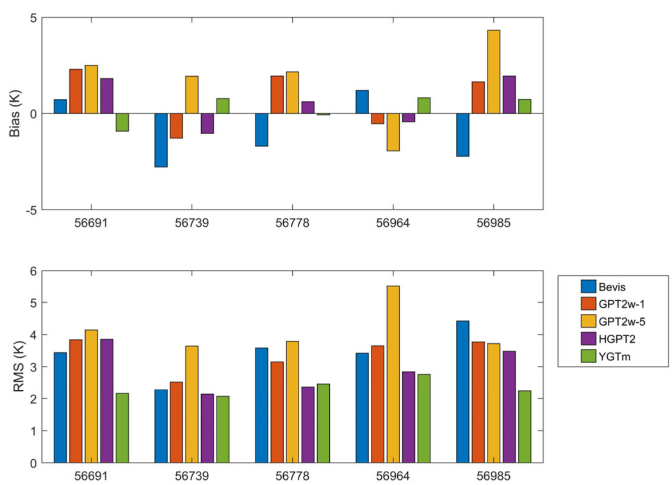

Figure 5. Bar chart of bias and RMS error for various models at the modeling station from 2018 to 2019

From Table 5 and Figure 5, it can be seen that the mean deviation of the Bevis model is -0.94 K, with significant fluctuations; the mean deviations of GPT2w-1 and GPT2w-5 are 0.82 K and 1.80 K, respectively; HGPT2 has a mean deviation of 0.59 K; the YGTm model has a mean deviation of only 0.28 K, with the smallest range (-0.92~0.83 K), indicating the best systematic deviation. In terms of root mean square error (RMS), the RMS values for Bevis, GPT2w-1, GPT2w-5, and HGPT2 are 3.43 K, 3.38 K, 4.16 K, and 2.94 K, respectively, while YGTm has the smallest RMS value of 2.34 K. Compared to the other four models, YGTm improves accuracy by 32%, 31%, 44%, and 20%, respectively, demonstrating higher stability and regional applicability.

4.2. External conformity accuracy verification

Verifying solely with modeling site data is insufficient to comprehensively assess the regional adaptability of the model. To this end, this paper selects independent sounding stations that did not participate in the modeling process, and utilizes their complete data from 2010 to 2019 for external consistency verification, in order to test the model's spatial generalization ability. The results are presented in Table 3.

Table 3. Analysis of external consistency accuracy of regional models

Model	Bias/K			RMS/K		
	Max	Min	Avg	Max	Min	Avg
Bevis	2.52	-3.41	-1.28	4.59	2.36	3.86
GPT2w-1	3.93	-1.03	1.32	3.97	2.94	3.58
GPT2w-5	3.87	-2.68	1.52	4.91	3.58	4.13
HGPT2	2.08	-1.15	0.64	3.94	2.15	2.96
YGTm	1.77	-0.47	0.36	2.83	2.12	2.57

According to Table 3, the YGTm model exhibits the highest accuracy in external validation. Its average bias is merely 0.36 K, significantly lower than that of Bevis (-1.28 K), GPT2w-1 (1.32 K), GPT2w-5 (1.52 K), and HGPT2 (0.64 K). In terms of root mean square error, YGTm stands at 2.57 K, whereas Bevis, GPT2w-1, GPT2w-5, and HGPT2 are 3.86 K, 3.58 K, 4.13 K, and 2.96 K, respectively. Compared to the aforementioned models, YGTm's accuracy is improved by 33%, 28%, 38%, and 13%, respectively. This indicates that YGTm maintains high accuracy and stability across independent sites, effectively adapts to the complex terrain of Yunnan-Guizhou region, overcomes the limited regional applicability of traditional models, possesses good spatial generalization capabilities, and is suitable for high-precision estimation of Tm in this region.

4.3. Leave-one-out cross-validation

To further avoid the contingency of a single validation site and fully utilize the limited sounding data, this paper employs Leave-One-Out Cross-Validation (LOOCV) to comprehensively evaluate the YGTm model, adhering to the small-sample modeling standards in atmospheric science. The specific process involves taking six sounding stations in turn as independent validation stations, with the remaining five stations serving as modeling stations. The modeling and validation periods remain unchanged, and the model is constructed and tested sequentially, recording the Bias and RMS mean values for each time. After six rounds, the arithmetic mean of the six results is calculated, and the average error from cross-validation is used as a comprehensive indicator of model accuracy and generalization performance. The results are presented in Table 4.

Table 4. Analysis of Leave-One-Out Cross-Validation Accuracy

Site number	Bias					RMS				
	Bevis	GPT2w-1	GPT2w-5	HGPT2	YGTm	Bevis	GPT2w-1	GPT2w-5	HGPT2	YGTm
56691	-1.82	1.76	1.94	0.57	0.41	3.58	3.39	4.17	2.94	2.66
56739	-0.71	0.91	1.12	0.38	0.35	3.51	3.27	3.83	3.14	2.41
56778	-1.15	1.58	1.73	0.61	0.49	4.55	4.01	4.72	2.97	2.87
56964	-1.06	1.44	1.85	0.86	0.53	3.81	3.80	4.05	3.05	2.53
56985	-0.94	1.27	1.47	0.48	0.26	4.15	3.62	4.25	2.74	2.84
57816	-1.28	1.32	1.52	0.64	0.36	3.86	3.58	4.13	2.96	2.57
Average	-1.16	1.38	1.61	0.59	0.40	3.91	3.61	4.19	2.97	2.65

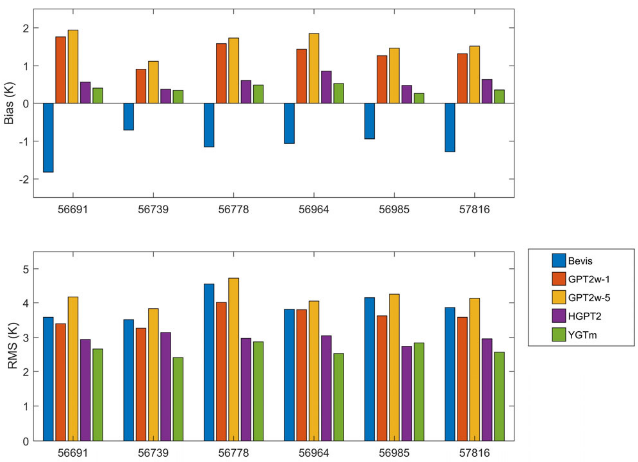

Figure 6. Bar chart of Bias and RMS error of each model in leave-one-out cross-validation

As can be seen from Table 4 and Figure 6, after six rounds of cyclic verification and averaging, the YGTm model exhibits the smallest average Bias and RMS, significantly outperforming other models. The Bevis model shows a pronounced negative bias; both GPT2w-1 and GPT2w-5 exhibit positive biases; the HGPT2 model has a bias of 0.59 K and an RMS of 2.97 K, which is superior to traditional models but still higher than the YGTm. The results indicate that the YGTm model is stable and reliable in accuracy, adaptable to different climates and terrains in Yunnan and Guizhou, and possesses good regional applicability and spatial generalization capabilities.

5. The impact of Tm on GNSS-PWV estimation

The core purpose of constructing the YGTm model is to enhance the accuracy of Tm calculation, thereby improving the precision of GNSS-derived PWV inversion. However, due to the inconsistent locations of GNSS reference stations and radiosonde stations, and the fact that most GNSS stations are not equipped with meteorological sensors, it is difficult to directly assess the impact of Tm on PWV inversion. Therefore, based on the research conducted by Huang Liangke et al., this paper theoretically analyzes the impact of Tm on the accuracy of PWV inversion:

$$\frac{RMS_{PWV}}{PWV} = \frac{RMS_{\Pi}}{\Pi} = \frac{k_3 \cdot RMS_{T_m}}{\left(k_2^{'}+\frac{k_3}{T_m}\right)T_m^2} = \frac{k_3}{\left(k_2^{'}+\frac{k_3}{T_m}\right)T_m} \cdot \frac{RMS_{T_m}}{T_m} \quad (11)$$

In the formula, RMSPWV represents the RMS error of PWV; RMSΠ denotes the RMS error of the conversion factor Π; RMSTm signifies the RMS error of Tm; and RMSPWV/PWV stands for the relative error of PWV. Tm and PWV utilize the annual mean values derived from sounding data in 2019. The calculation results for RMSPWV/PWV and RMSPWV of each model are presented in Table 5

Table 5. Theoretical RMS error and relative error of PWV calculated by different models

Model	RMSPWV/mm			RMSPWV/PWV/%		
	Max	Min	Avg	Max	Min	Avg
Bevis	0.36	0.16	0.27	2.06	1.12	1.52
GPT2w-1	0.47	0.08	0.31	1.53	0.89	1.26
GPT2w-5	0.43	0.12	0.29	1.95	0.92	1.43
HGPT2	0.39	0.15	0.25	1.68	0.96	1.03
YGTm	0.28	0.13	0.21	0.95	0.73	0.85

As can be seen from Table 5, the Bevis model, GPT2w model, and HGPT2 model all exhibit relatively large RMSPWV and RMSPWV/PWV values, whereas the YGTm model has the smallest RMSPWV, with an average value of 0.21 mm, and an average RMSPWV/PWV value of 0.85%, ranging from 0.73% to 0.95%. Compared to other models, it has a smaller fluctuation range and more stable performance. Therefore, the error caused by Tm calculated using the YGTm model has a smaller impact on GNSS-PWV inversion compared to other models, providing more accurate Tm for GNSS-PWV inversion in the Yunnan-Guizhou region of China.

6. Conclusion

Based on the measured data from six sounding stations in the Yunnan-Guizhou region spanning from 2010 to 2019, this paper constructs a weighted average temperature empirical model, YGTm, that takes into account multiple factors. The results indicate that Tm exhibits a significant seasonal variation in the Yunnan-Guizhou region, being higher in summer and lower in winter, and demonstrates good correlation with surface temperature and the natural logarithm of vapor pressure. The YGTm model performs optimally in internal consistency, external consistency, and leave-one-out cross-validation, with a comprehensive average RMS of 2.52 K. Compared to the Bevis, GPT2w-1, GPT2w-5, and HGPT2 models, its accuracy is improved by 32%, 28%, 39%, and 15%, respectively. The theoretical relative error of PWV inversion based on YGTm is only 0.85% on average, significantly outperforming the comparative models. This model can provide reliable Tm support for applications such as high-precision water vapor inversion and extreme weather monitoring in the Yunnan-Guizhou region using GNSS.

References

[1] Cao Yujing. Remote Sensing of Atmospheric Water Vapor Content Using Ground-Based GPS Technology and Its Application in Meteorology [D]. Nanchang University, 2008.

[2] Li Haojie, Liu Lilong, Huang Liangke, et al. Research on the refinement of China's regional Tm model considering multi-factor impacts [J]. Journal of Geodesy and Geodynamics, 2022, 42(04): 393-397.

[3] Wang Hanhong, Wei Guanjun, Zhang Xing, et al. Research on the construction method of regional atmospheric weighted average temperature model in Lanzhou area [J]. Journal of Navigation and Positioning, 2022, 10(01): 90-96.

[4] Mo Zhixiang, Li Jiahao, Zhou Lu, et al. A Spatio-Temporal Weighted Mean Temperature Model for the Tibetan Plateau Region Taking into Account Temporal and Spatial Factors [J]. Journal of Nanjing University of Information Science & Technology (Natural Science Edition): 1-11.

[5] Wei Haifu, Chen Tianwei, Chen Ming. Regional Atmospheric Weighted Mean Temperature Model Considering Regional Relative Elevation in China [J]. Journal of Geodesy and Geodynamics, 2021, 41(10): 1057-1062.

[6] Xu Mingze, Guo Qiuying, Hou Jianhui, et al. Establishment and Accuracy Analysis of Atmospheric Weighted Average Temperature Model in Jinan Area [J]. Journal of Navigation and Positioning, 2021, 9(05):142-151.

[7] Fan Shijie, Liu Zhaojian, Chen Yan, et al. Optimization of the Atmospheric Weighted Average Temperature Model in Qingdao Region [J]. Geospatial Information, 2021, 19(09): 136-138.

[8] Zheng Lei. Atmospheric Weighted Average Temperature Model Considering Seasonal Variations in the Southwest Region [J]. Journal of Navigation and Positioning, 2021, 9(04): 98-103.

[9] Shi Lingfan, Wang Ping, Huang Liangke. Construction of Atmospheric Weighted Average Temperature Model in the Qinghai-Tibet Plateau Region [J]. Bulletin of Surveying and Mapping, 2021(04):52-59.

[10] Zhu Hai, Huang Guanwen, Zhang Juqing. Regional Atmospheric Weighted Mean Temperature Model Taking into Account Climate Differences - A Case Study of Shaanxi Province, China [J]. Journal of Surveying and Mapping, 2021, 50(03): 356-367.

[11] Ke Liping, Liao Liufeng, Yang Yang, et al. Research on the Atmospheric Weighted Average Temperature Model in Guizhou Province [J]. Mid-Low Latitude Mountain Meteorology, 2021, 45(01): 58-62.

[12] Cui Jinye, Ma Xiaping, Liu Xiaopeng. Atmospheric Weighted Mean Temperature Model Taking into Account Altitude Correction in Hong Kong Region [J]. Geodesy and Geodynamics, 2020, 40(10): 1022-1026.

[13] He Shiwei, Kong Xiaoyu, Guo Linyun. Construction of Atmospheric Weighted Average Temperature Model in Anhui Region [J]. Beijing Surveying and Mapping, 2020, 34(11): 1644-1648.