Research in Astronomy and Astrophysics, 25:104002 (10pp), 2025 October
https://doi.org/10.1088/1674-4527/adf70f
CSTR: 32081.14.RAA.adf70f
© 2025. National Astronomical Observatories, CAS and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

Enhanced Zenith Tropospheric Delay Forecasting Using a Hybrid GRU-LSTM Deep Learning Model

Ming-Shuai Li[1,2,3], Yu Li[4], Na Wang[1,2,3], Lang Cui[1,2,3], Ming Zhang[1,3], Jian Li[1,5], and Xue-Feng Duan[1,5]
[1] Xinjiang Astronomical Observatory of the Chinese Academy of Sciences, Urumqi 830011, China; limingsh@xao.ac.cn
[2] Key Laboratory of Radio Astronomy and Technology, CAS, Beijing 100101, China
[3] Xinjiang Key Laboratory of Radio Astrophysics, Urumqi 830011, China
[4] China Earthquake Networks Center, Beijing 100045, China
[5] Xinjiang Key Laboratory of Microwave Technology, Urumqi 830011, China
Received 2025 April 23; accepted 2025 May 18; published 2025 September 12

Abstract

Accurate estimation of Zenith Tropospheric Delay (ZTD) is essential for mitigating atmospheric effects in radio astronomical observations and improving the retrieval of precipitable water vapor (PWV). In this study, we first analyze the periodic characteristics of ZTD at the NanShan Radio Telescope site using Fourier transform, revealing its dominant seasonal variations, and then investigate the correlation between ZTD and local meteorological parameters, to better understand atmospheric influences on tropospheric delay. Based on these analyses, we propose a hybrid deep learning Gated Recurrent Units-Long Short-Term Memory model, incorporating meteorological parameters as external inputs to enhance ZTD forecasting accuracy. Experimental results demonstrate that the proposed approach achieves a Root Mean Squared Error of 7.97 mm and a correlation coefficient R of 96%, significantly outperforming traditional empirical models and standalone deep learning architectures. These findings indicate that the model effectively captures both short-term dynamics and long-term dependencies in ZTD variations. The improved ZTD predictions not only contribute to reducing atmospheric errors in radio astronomical observations but also provide a more reliable basis for PWV retrieval and forecasting. This study highlights the potential of deep learning in tropospheric delay modeling, offering advancements in both atmospheric science and geodetic applications.

Key words: atmospheric effects – methods: data analysis – site testing

1. Introduction

Zenith Tropospheric Delay (ZTD) is a crucial factor affecting the accuracy of Global Navigation Satellite System (GNSS) positioning and radio astronomical observations, particularly in Very Long Baseline Interferometry (VLBI) and high-frequency radio measurements. In VLBI, precise tropospheric corrections are essential to minimize phase errors and improve the accuracy of celestial source localization (Reid et al. 2009). Variations in tropospheric delay directly impact interferometric baseline solutions and astrometric measurements, making accurate ZTD estimation vital for high-precision astronomical observations (Teke et al. 2013; Zhou et al. 2022). Additionally, ZTD plays a crucial role in atmospheric science, as it comprises two components: Zenith Hydrostatic Delay (ZHD) and Zenith Wet Delay (ZWD). Among them, ZWD is strongly correlated with atmospheric water vapor content, making ZTD an important indicator for retrieving precipitable water vapor (PWV). This capability renders ZTD data highly valuable for weather forecasting, climate monitoring, and astronomical site assessment (Bevis et al. 1992; Li et al. 2020).

To enhance tropospheric corrections, GNSS-derived ZTD has been widely integrated into VLBI data processing (Heinkelmann et al. 2007; Puente 2021). The International VLBI Service incorporates high-precision GNSS observations to refine atmospheric delay modeling for geodetic and astrophysical applications (Schuh & Böhm 2013). Honma et al. (2008) demonstrated that incorporating GNSS ZTD data into VLBI processing significantly improves celestial reference frame solutions. Moreover, short-term ZTD forecasting is crucial for real-time GNSS positioning, improving VLBI tropospheric delay modeling, and enhancing numerical weather prediction (NWP) models (Tregoning & Watson 2009). However, the dynamic nature of atmospheric conditions introduces significant temporal variations in ZTD, which cannot be fully captured by static corrections alone. This necessitates the development of accurate ZTD prediction models that can adapt to atmospheric changes in real time.

Extensive research has been conducted on ZTD estimation and correction (Liwosz 2013; Li et al. 2015), including empirical models, mapping functions, and data-driven correction techniques. Traditional empirical models, such as the Hopfield and Saastamoinen models, estimate ZTD based on

meteorological parameters. While computationally efficient, these models suffer from regional adaptability issues and cannot fully capture short-term atmospheric fluctuations (Saastamoinen 1972). The mapping functions, such as the Vienna Mapping Function (VMF; Rocken et al. 2001; Boehm & Schuh 2004) and the Global Mapping Function (Boehm et al. 2006), provide site-specific tropospheric delay estimates but require external meteorological inputs. NWP models, such as the ERA-Interim and JRA-55 reanalysis data sets, offer improved accuracy but are computationally expensive and depend on initial atmospheric conditions (Dee et al. 2011; Kobayashi et al. 2015).

In contrast, data-driven correction techniques have shown great potential in capturing complex atmospheric dynamics by learning patterns from historical observations. Among them, statistical time series models (Qu et al. 2008; Liu et al. 2017) (e.g., SARIMAX, Prophet) excel in capturing periodic variations and long-term trends, offering high interpretability and efficient computation. Meanwhile, machine learning sequence models (Chen & Guestrin 2016; Li et al. 2023) (e.g., XGBoost, Random Forest) leverage ensemble learning to model nonlinear relationships in ZTD data, significantly improving short-term forecasting accuracy.

Despite these advancements, both models above are limited by their reliance on manual feature engineering and their inability to model long-term dependencies effectively. Deep learning models, particularly recurrent neural networks (RNN), have demonstrated superior performance in capturing both short- and long-term dependencies in sequential data. Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) mitigate the vanishing gradient problem and enhance sequential modeling capabilities (Hochreiter & Schmidhuber 1997; Cho et al. 2014; Greff et al. 2016). Chen et al. (2023) applied LSTM networks to GNSS time series and achieved better prediction accuracy than traditional autoregressive models. Lu et al. (2023) incorporated external meteorological features into GRU-based ZTD models, further enhancing forecast performance. In addition, another model with transformer-based architectures (Vaswani et al. 2017) has gained prominence in time series modeling due to its self-attention mechanism, which allows for long-range dependency modeling. Lim et al. (2021), Zhang et al. (2022), Haji-Aghajany et al. (2025) explored the application of the Temporal Fusion Transformer for atmospheric delay prediction, demonstrating superior performance in capturing seasonal and meteorological influences on ZTD. However, while Transformer-based models excel in capturing long-range dependencies in time series, their high computational cost remains a major limitation for real-time applications.

The NanShan Radio Telescope (NSRT) is a 26 m radio telescope located at the Nanshan Station of the Xinjiang Astronomical Observatory, Chinese Academy of Sciences, with an altitude of 2080 m (43°28′ N, 87°10′ E). It is an important VLBI observation station and a key component of China's deep-space tracking network. Beyond its role in radio astronomy, the Nanshan Station is equipped with three sets of high-precision GNSS equipment: an IGS reference station, a lunar exploration network GPS station, and a land network reference station. These GNSS receivers provide continuous and high-quality ZTD observations, making NSRT an ideal site for studying atmospheric effects on GNSS signals. Due to its high-altitude location and minimal interference from urban environments, the station provides stable and reliable long-term ZTD measurements. The goal of this study is to improve the accuracy and stability of ZTD forecasting, providing a more reliable foundation for VLBI atmospheric corrections, GNSS positioning error reduction, and PWV retrieval.

2. Materials and Methods

2.1. Data Sources

The GNSS observation data used in this study were obtained from the Crustal Movement Observation Network of China (CMONOC; Wang & Shen 2020; Wang et al. 2023). This network consists of 260 GNSS reference stations and 2000 regional stations, covering mainland China and surrounding areas, providing high-precision, high spatiotemporal resolution GNSS observations. Based on these observations, a continuous time series of ZTD can be derived. The meteorological data at the Nanshan station are obtained from a PC-4 automatic meteorological station located within the research facility. This station is capable of collecting various atmospheric parameters, including temperature, pressure, relative humidity (RH), wind speed, and wind direction. The measurement resolutions for each parameter are as follows: temperature at 0.1°C, pressure at ±0.1 hPa, RH at 0.1%, wind speed at 0.1 m s^{-1}, and wind direction at 3°. The temperature, humidity, and pressure sensors are installed at a height of 1.5 m above the ground. All data are recorded at a time interval of 1 minute.

2.2. GNSS Data Processing and ZTD Estimation

In this study, GAMIT/GLOBK version 10.7 was used to estimate ZTD from GNSS observations. The data were processed using the Relax mode, which allows partial parameter adjustments during the solution process, thereby improving flexibility and accuracy.

The GNSS observation data used in this study were obtained from the XJNS station, located at the NSRT site. To extract the ZTD time series for XJNS, precise GNSS data processing was performed using three surrounding International GNSS Service (IGS) stations–XJSS, XJSH, and XJBY Beutler et al. (1999). These reference stations were selected based on data quality, spatial distribution, and geometric relationships with XJNS to ensure the stability and accuracy of the ZTD estimation. The satellite orbits were computed using the high-

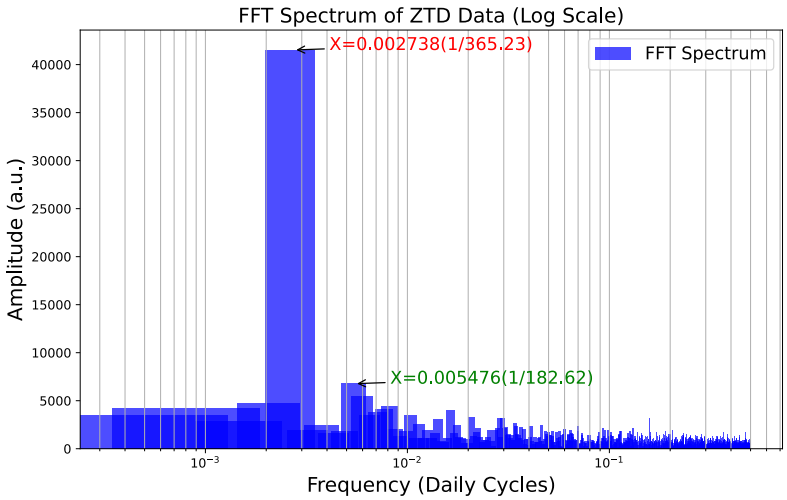


Figure 1. Time series of ZTD at the XJNS station (2016–2021). The periodic fluctuations indicate significant annual and semi-annual variations.

precision final ephemeris provided by IGS, ensuring centimeter-level orbit accuracy.

To mitigate multipath effects and atmospheric delay errors in low-elevation signals, a 10° elevation cutoff angle was applied. The VMF1 mapping function was employed to project the tropospheric delay into the zenith direction, improving tropospheric modeling accuracy. ZTD was estimated hourly and stored as a time series for subsequent analysis using a piecewise linear model.

2.3. Deep Learning Model for ZTD Prediction

GRU and LSTM are widely used variants of RNN, both capable of capturing temporal dependencies. GRU simplifies the memory mechanism of LSTM, reducing computational complexity while retaining its ability to model long-term dependencies. In contrast, LSTM utilizes dedicated gating mechanisms to mitigate the vanishing gradient problem, making it more effective for learning long-range patterns in sequential data.

The proposed GRU-LSTM hybrid model is designed to exploit the strengths of both architectures. The GRU layer extracts local temporal features with lower computational overhead, while the LSTM layer refines long-term dependencies, ensuring stability and accuracy in ZTD prediction. By integrating these two units, the model effectively captures both short-term variations and long-term atmospheric trends, addressing periodic fluctuations and irregular perturbations in ZTD dynamics.

2.4. Characteristics and Influencing Factors of ZTD

Understanding the temporal characteristics and influencing factors of ZTD is crucial for improving prediction accuracy and model robustness. In this section, we first analyze the periodic variations of ZTD using Fourier transform, identifying dominant seasonal patterns. Then, we explore the correlation between ZTD and local meteorological parameters, such as temperature and pressure, to gain insights into the atmospheric influences on tropospheric delay.

2.4.1. Temporal Variations of ZTD

To analyze the temporal variation of ZTD at the NSRT station, we first applied Fast Fourier Transform to the ZTD time series to identify dominant periodic components (see Figure 1). The results revealed two significant frequencies at 0.0027 Hz and 0.0055 Hz, corresponding to annual (365.23 days) and semi-annual (182.62 days) cycles, respectively. These findings indicate that the ZTD variations at the NSRT station exhibit strong periodicity, consistent with general ZTD temporal patterns.

2.4.2. Correlation with Meteorological Parameters

To further investigate the driving factors of ZTD variations and its relationship with atmospheric pressure, temperature, and RH were analyzed from 2016 to 2021, as shown in Figure 2(a)–(d).

As shown in Figure 2(a), it illustrates the temporal variations of ZTD at the XJNS station from 2016 to 2021. The time series analysis reveals a distinct seasonal pattern, with higher values during summer and lower values in winter. This periodic fluctuation suggests a strong correlation between ZTD and atmospheric water vapor content, as the ZWD is the primary variable component of ZTD.

Atmospheric pressure, temperature, and RH all influence ZTD to varying degrees. As shown in Figure 2(b), the atmospheric pressure exhibits an inverse relationship with ZTD. Since the ZHD is directly proportional to pressure, an

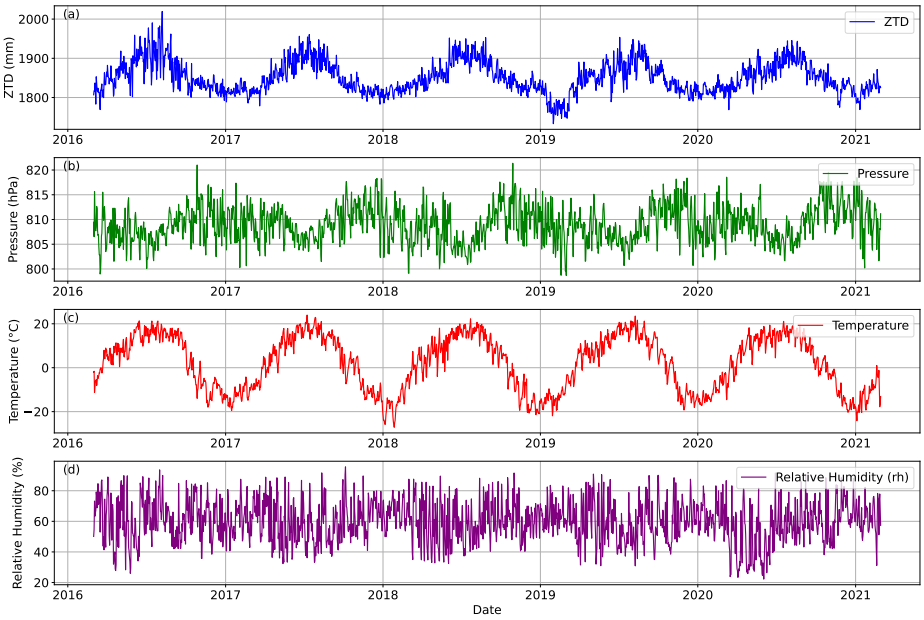


Figure 2. ZTD variations and its relationship with pressure, temperature and relative humidity.

increase in pressure leads to a higher ZHD. However, because ZTD is the sum of ZHD and ZWD, higher pressure often coincides with reduced atmospheric water vapor, resulting in a decrease in total ZTD.

Although the overall trend of pressure appears negatively correlated with ZTD, its time series exhibits relatively small seasonal amplitude but noticeable short-term fluctuations. These high-frequency variations introduce noise into the correlation, as ZHD is relatively stable and less sensitive to rapid atmospheric changes. Additionally, short-term pressure changes are not necessarily accompanied by proportional variations in water vapor. This explains the relatively weak correlation between pressure and ZTD, with a Pearson correlation coefficient of -0.192.

Temperature, shown in Figure 2(c), follows a well-defined annual cycle, with higher values in summer months. Increased temperature is associated with greater atmospheric water vapor, leading to greater ZWD and, consequently, higher ZTD values. Correlation analysis confirms this strong dependency, with a Pearson coefficient of 0.767, indicating a significant positive relationship between temperature and ZTD. This finding emphasizes the dominant role of temperature-driven water vapor variation in ZTD dynamics.

Relative humidity, displayed in Figure 2(d), is expected to influence ZWD; however, our analysis reveals a negligible correlation between RH and ZTD, with a Pearson coefficient of only -0.011. This weak relationship suggests that at the NSRT site–characterized by a relatively dry continental climate–RH does not serve as a reliable proxy for atmospheric water vapor variation relevant to ZWD estimation. Similar findings have been reported in previous studies, indicating that the impact of RH on tropospheric delay depends on regional climate conditions.

These relationships highlight the strong dependence of ZTD on atmospheric moisture dynamics, particularly through temperature and pressure variations. Understanding these correlations provides a foundation for improving ZTD prediction models by incorporating relevant meteorological parameters. The insights gained from this analysis will guide the selection of input features in the subsequent modeling process, ensuring that the most influential factors are considered for accurate ZTD forecasting.

2.5. Model Construction and Optimization

To build the GRU-LSTM model, we first preprocess the ZTD and temperature data. The `Date` column is converted into a datetime format and set as the index to maintain a daily frequency, as shown in Figure 3. Then, the StandardScaler is applied to standardize the ZTD, pressure and temperature data, ensuring numerical stability during training. A sliding window approach with a window length of 30 days is adopted to construct the input dataset for model training. This window size was selected based on empirical testing, which showed that 30 days provide an optimal balance between model accuracy and training stability. A detailed comparison of different window lengths is presented in the Appendix. This method transforms the original time series into supervised learning samples by taking a fixed-length window of historical observations as an input and the subsequent value as the prediction target. Specifically, for each sample, the data are structured as follows:

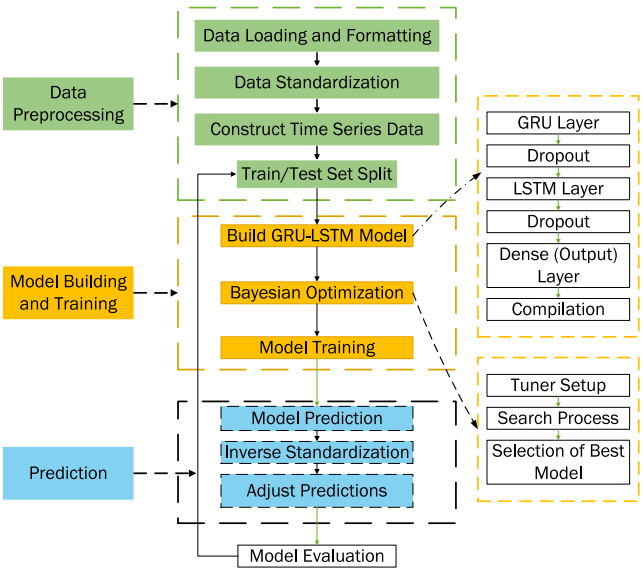


Figure 3. GRU-LSTM forecasting model structure.

1. X: A sequence of 30 consecutive days of ZTD data, used to capture local time series patterns;
2. X_exog: The corresponding pressure and temperature data for the same period, serving as an exogenous variable;
3. y: The ZTD value for the day immediately following the window, serving as the prediction target.

The data set is split into training (90%) and testing (10%) sets, and reshaped into a three-dimensional array (samples $\times$ 30 $\times$ 1) to match the GRU and LSTM input requirements.

Using the `Sequential` API in Keras Chollet (2021), a hybrid GRU-LSTM model is constructed with the following architecture:

1. *GRU Layer*. Extracts local temporal features, with the number of units controlled by the hyperparameter `gru_units`. The `return_sequences` parameter is set to `True` to pass sequence information to the next layer.
2. *Dropout Layer*. Added after the GRU layer to mitigate overfitting, with the dropout rate determined by `gru_dropout`.
3. *LSTM Layer*. Captures long-term dependencies, with the number of units defined by `lstm_units`.
4. *Second Dropout Layer*. Applied after the LSTM layer, with the dropout rate set by `lstm_dropout`.
5. *Fully Connected Layer (Dense)*. Outputs the final ZTD prediction.

The model is compiled using the Adam optimizer with an initial learning rate of 0.001 and trained with a batch size of 64. The loss function is set as mean squared error (MSE), minimizing the difference between predicted and observed ZTD values.

To enhance model performance, hyperparameters such as the number of hidden units, dropout rates, and recurrent layers are fine-tuned through grid search. Early stopping with a patience of 20 epochs is applied to prevent overfitting, ensuring training halts when no further improvement in validation loss is observed.

Once training is completed, the model is evaluated using statistical metrics, which will be discussed in Section 3.1.

3. Results and Discussion

With the trained GRU-LSTM model, we now examine its predictive performance using various statistical metrics and compare it with traditional and machine learning-based models. we employ five commonly used evaluation metrics: Mean Absolute Error (MAE), Mean Bias Error (MBE), Root Mean Squared Error (RMSE), Mean Squared Error (MSE), and the correlation coefficient (R). The mathematical definitions of these metrics are as follows:

$$\mathrm{MAE}=\frac{1}{N}\sum_{i=1}^{N}|y_i-\hat{y}_i| \tag{1}$$

$$\mathrm{MBE}=\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i-y_i) \tag{2}$$

$$\mathrm{MSE}=\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2 \tag{3}$$

$$\mathrm{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2} \tag{4}$$

$$R=\frac{\sum_{i=1}^{N}(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i-\bar{y})^2}\sqrt{\sum_{i=1}^{N}(\hat{y}_i-\bar{\hat{y}})^2}} \tag{5}$$

where:

1. y_i represents the observed ZTD value at time step i.
2. $\hat{y}_i$ represents the predicted ZTD value at time step i.
3. $\bar{y}$ and $\bar{\hat{y}}$ denote the mean values of the observed and predicted ZTD over all N samples.
4. N is the total number of data points.

To further assess the model's performance and characterize the prediction errors, we conduct a comprehensive residual analysis through the following visualizations:

1. A time series plot of the observed and predicted ZTD values, including a zoomed-in view of the test set, to evaluate the model's ability to follow the temporal variation;

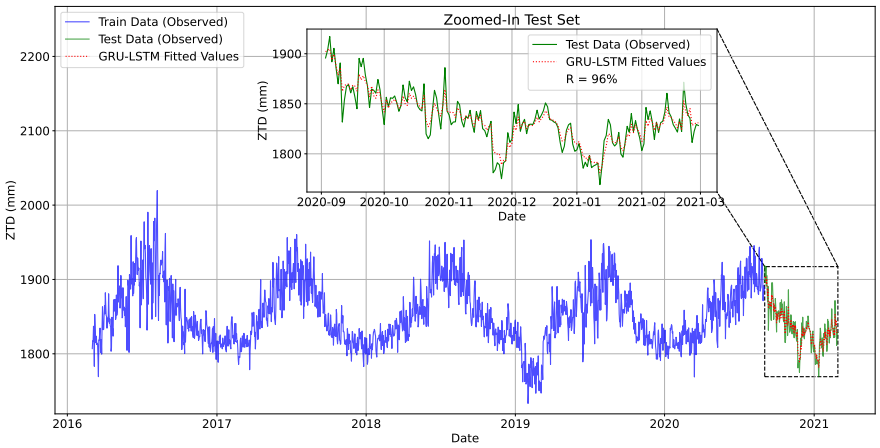


Figure 4. Comparison between fitted and observed ZTD values. (The main panel displays the full time series, where the first 90% of data is used for training and the last 10% for testing. The blue and green curves represent observed ZTD values in the training and test sets, respectively, while the red dotted line shows the GRU-LSTM model's fitted results. The inset provides a zoomed-in view of the test set.)

2. A scatter density plot comparing predicted and actual values, accompanied by key evaluation metrics such as *R*, MAE, RMSE, MBE, and MSE;
3. A histogram and a quantile–quantile (Q–Q) plot of the residuals, used to inspect the distribution's symmetry and normality. Skewness and kurtosis are also computed to quantify deviations from Gaussianity.

3.1. Model Performance Evaluation

Figure 4 presents a comparison between the fitted ZTD values and observed data, including a zoomed-in view focusing on the test set (10% of the data). The fitted values align well with the observed values across the entire time span, confirming the model's strong capability in tracking ZTD fluctuations. The computed *R* value of 0.96 further supports the model's high accuracy.

While the predicted ZTD values closely follow the trend of the observed data, especially during rapid changes, the zoomed-in panel reveals a slight reduction in the amplitude of the predicted values compared to the observed ones. This "amplitude compression" effect is commonly seen in sequential learning models and may be attributed to the model's regularization behavior or the scarcity of extreme events in the training data. Nevertheless, the GRU-LSTM model demonstrates robust performance in capturing both the direction and magnitude of ZTD variations.

To assess how training data volume affects model performance, we analyze the variation of RMSE and correlation coefficient (*R*) as a function of training dataset size, as shown in Figure 5. Several trends are observed:

1. *Performance Improvement with More Data*. With an increasing number of training samples, RMSE steadily decreases, indicating a reduction in prediction error. Simultaneously, the correlation coefficient *R* consistently rises toward 1, demonstrating improved agreement between predicted and observed ZTD values.
2. *Reduction in Seasonal Influence*. Initially, RMSE and R exhibit seasonal fluctuations, likely due to variations in temperature, pressure, and atmospheric water vapor. However, as training data accumulates, these seasonal variations diminish, suggesting that the model progressively learns long-term temporal patterns in ZTD.
3. *Model Convergence Trend*. When trained on a small dataset, RMSE drops rapidly and R increases significantly, indicating that additional data greatly enhance predictive performance in early stages. However, once a sufficient amount of data is used, the performance improvements become less pronounced, suggesting convergence toward the model's optimal capability.

These observations indicate that the GRU-LSTM model gradually learns long-term atmospheric patterns and benefits from expanded historical data.

To further investigate whether prediction accuracy varies with seasonal conditions, we conducted separate forecasting experiments across four seasons. Each test set consisted of 90 consecutive days of data corresponding to winter, spring, summer, and fall, respectively.

As shown in Table 1, the model achieves the highest accuracy in winter, with an RMSE of 6.67 mm and a correlation coefficient of 0.95. Conversely, summer forecasts show the largest errors (RMSE = 16.28 mm) and the lowest correlation (R = 0.77), highlighting the increased complexity of atmospheric conditions in this season. This performance degradation is primarily attributed to the higher water vapor content and more active convective processes during summer, which introduce greater nonlinear and transient fluctuations in the ZTD signal–particularly in the ZWD component. These

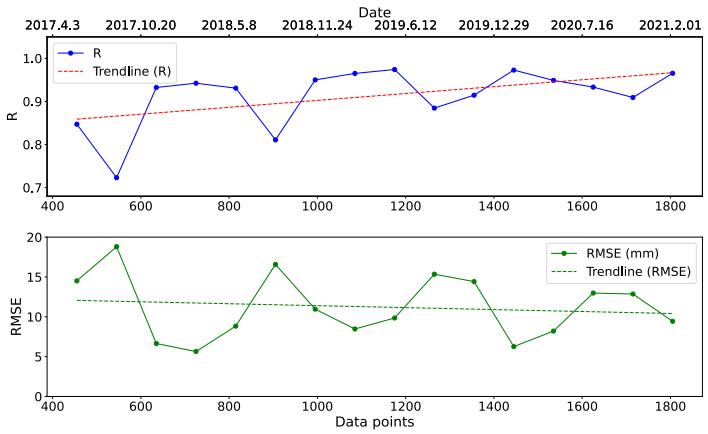


Figure 5. Evaluation of model performance as a function of the training dataset size. The RMSE and correlation coefficient (R) show a steady improvement as more data are used for training. The seasonal influence on performance metrics weakens over time, indicating the model's enhanced stability.

dynamics pose challenges for sequential models to generalize. In contrast, the more stable atmospheric behavior in winter facilitates more accurate and consistent forecasts. Despite the reduced performance in summer, the model maintains reasonable accuracy across all seasons, demonstrating strong generalization ability under varying meteorological conditions.

In summary, the GRU-LSTM model demonstrates high accuracy in forecasting ZTD, with performance consistently improving as the amount of training data increases. In particular, long-term data availability enables the model to better capture seasonal patterns and atmospheric variability, thereby reducing the influence of seasonal fluctuations on prediction accuracy. These results also underscore the importance of continuous, long-term GNSS and meteorological observations in supporting stable and reliable tropospheric delay forecasting.

3.2. Performance Comparison of Different Models

In the task of ZTD prediction, different models exhibit varying levels of accuracy, stability, and computational efficiency. Table 2 presents the key evaluation metrics for each model, which are all computed based on the actual observed and predicted ZTD values derived from our unified test data set. while Figure 6 visually compares the distribution of predicted and observed ZTD values. Overall, the GRU-LSTM hybrid model outperforms all other approaches, demonstrating superior generalization capability and predictive accuracy.

Traditional time-series models, such as SARIMAX and Prophet, rely on linear assumptions and struggle to capture the complex nonlinear variations in ZTD. As a result, they exhibit relatively high RMSE and lower correlation coefficients. While these models offer computational efficiency, their predictive performance is significantly limited when dealing with strongly seasonal and nonlinear time series.

Machine learning models, including XGBoost and Random Forest, show improvements in predictive accuracy, particularly

Table 1
GRU-LSTM Model Prediction Performance Across Different Seasons

Season	RMSE (mm)	MSE	MAE (mm)	MBE (mm)	Correlation
Winter (Dec–Feb)	6.67	44.51	5.20	−0.31	0.9461
Autumn (Sep–Nov)	8.92	79.57	6.73	0.55	0.9567
Summer (Jun–Aug)	**16.28**	**264.88**	**13.18**	**−3.30**	**0.7736**
Spring (Mar–May)	12.00	143.92	9.19	−1.26	0.9188

Note. Bold values highlight the poorest seasonal performance (largest RMSE, MSE, and MAE; largest absolute MBE; and lowest correlation coefficient R), which occurs in summer.

in capturing short-term fluctuations. However, these methods still face limitations in modeling long-term dependencies, as evidenced by their relatively dispersed error distributions in Figure 6. This is particularly pronounced in cases where ZTD exhibits significant seasonal variations.

Deep learning models, such as LSTM and GRU, significantly enhance forecasting accuracy by effectively learning the long-term dependencies in ZTD time series. Compared to traditional methods, these models exhibit tighter clustering around the ideal prediction line in Figure 6, indicating improved predictive performance. However, while LSTM captures long-term dependencies well, its higher computational complexity may limit efficiency. GRU, on the other hand, offers a computationally efficient alternative but may be slightly less effective in modeling extended temporal patterns.

The GRU-LSTM hybrid model combines the strengths of both architectures, allowing it to efficiently extract short-term dynamics while accurately modeling long-term variations. As shown in Table 2, this model achieves the lowest RMSE (7.97 mm) and the highest correlation coefficient (0.96), indicating a high degree of agreement between predictions and observations. Furthermore, Figure 6 demonstrates that the GRU-LSTM model produces the most concentrated density

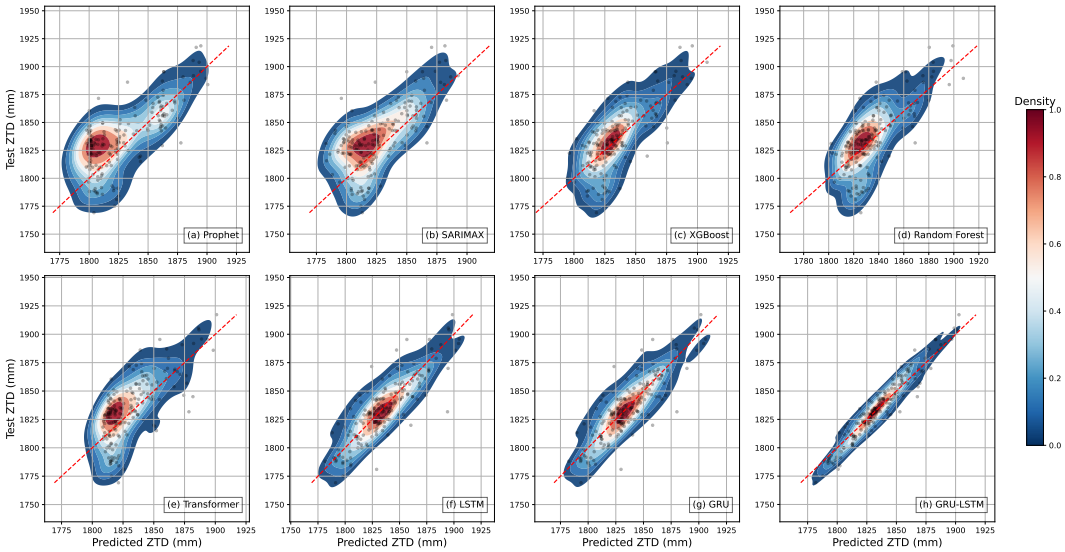


Figure 6. Model performance comparison for ZTD prediction.

Table 2
Performance Comparison of Different Models

Category	Model Name	MAE	MBE	R	RMSE	MSE
Time Series	Prophet	18.29	−8.40	0.72	22.78	519.10
	Sarimax	18.75	−10.20	0.69	23.18	537.24
Machine Learning	XGBoost	14.35	−1.08	0.75	18.77	352.40
	Random Forest	14.90	−1.64	0.74	19.12	365.49
Deep Learning	LSTM	11.51	2.02	0.84	15.17	230.22
	Transformer	14.81	−0.34	0.72	18.73	350.91
	GRU	10.79	0.55	0.85	14.33	205.31
Model Fusion	GRU+LSTM	6.05	0.04	0.96	7.97	63.47

Note. All evaluation metrics shown in Table 1 are computed using the same test set and based on actual predicted and observed ZTD values obtained in this study.

distribution around the ideal prediction line, signifying minimal error dispersion and strong robustness.

3.3. Residual Analysis

To further examine the prediction errors and assess the robustness of the GRU-LSTM model, we conduct a residual analysis based on two statistical visualizations, as shown in Figure 7. These plots help reveal the underlying distribution characteristics of the residuals and provide insight into the model's error structure. A well-behaved residual distribution is crucial for validating model assumptions and identifying potential areas for refinement.

Figure 7(a) presents a histogram of residuals with kernel density estimation (KDE). The distribution is approximately symmetric and centered around zero, indicating that the model does not exhibit significant systematic bias. The majority of residuals are concentrated near zero, suggesting consistent predictive performance. The calculated skewness of −0.255 and kurtosis of 3.90 indicate a slightly left-skewed and moderately leptokurtic distribution, which reflects a small deviation from ideal normality.

Figure 7(b) shows the corresponding quantile–quantile (Q–Q) plot, rotated to align with the histogram's orientation. Most residuals lie close to the diagonal reference line, supporting the assumption of near-normal distribution. Minor deviations in the extreme quantiles suggest the presence of a few outliers or slight heteroskedasticity, potentially due to complex nonlinear variations not fully captured by the current model structure.

Overall, the residual analysis confirms that the GRU-LSTM model yields largely unbiased and near-normally distributed prediction errors, validating its suitability for time series forecasting. The observed deviations, though small, highlight opportunities for further enhancement, such as incorporating additional features or optimizing the model architecture.

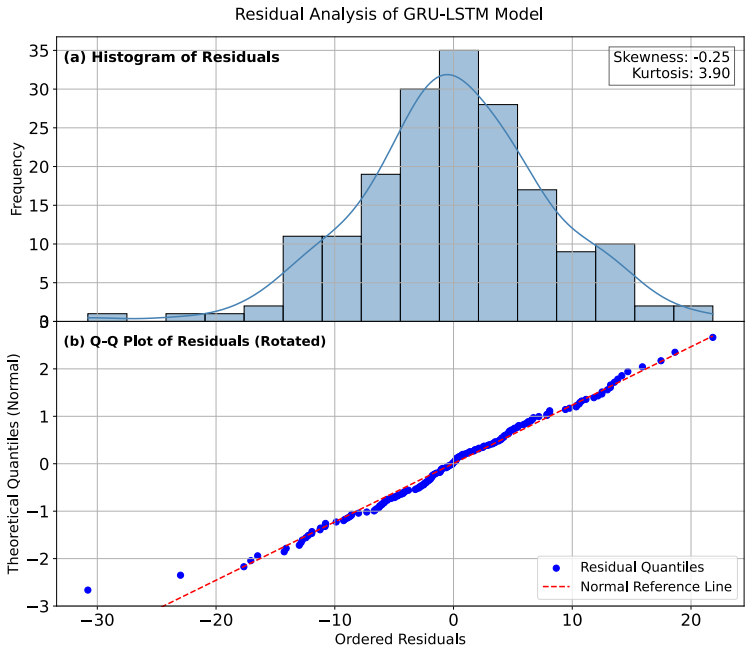


Figure 7. Residual analysis of the GRU-LSTM model. (a) Histogram of residuals with kernel density estimation (KDE). (b) Quantile–Quantile (Q–Q) plot, rotated to align with the histogram.

3.4. Model Limitations and Future Improvements

While the proposed GRU-LSTM model significantly improves the accuracy of ZTD forecasting, several limitations remain and offer directions for future enhancement.

Data Dependency: The model relies on long-term GNSS observations to effectively learn temporal patterns. However, in scenarios where historical GNSS data is sparse or unavailable, its predictive capability may be compromised. This highlights the need for strategies to enhance model robustness under limited data conditions.

Limited Meteorological Inputs and Seasonal Sensitivity: The current model primarily utilizes temperature and pressure as external inputs, which have demonstrated strong correlation with ZTD variations in general. However, the model exhibits significantly reduced accuracy during summer, likely due to intensified atmospheric water vapor and convective activity, which introduce greater nonlinearity and rapid fluctuations–especially in the wet delay component. These dynamics are difficult to fully capture using only standard meteorological variables. Future work should consider incorporating additional seasonally sensitive atmospheric parameters such as precipitation, PWV, or convective instability indices to improve the model's responsiveness under highly variable conditions. This enhancement is particularly critical for improving ZTD forecasting accuracy in humid, high-variability periods.

Computational Efficiency: Although the GRU-LSTM hybrid model achieves superior accuracy compared to traditional empirical models, its computational cost remains a consideration. While it is more efficient than Transformer-based architectures, further optimization techniques, such as model pruning, quantization, or reduced-parameter architectures, could enhance its real-time applicability.

Given these limitations, future research should aim to improve the model's generalizability, responsiveness to dynamic atmospheric conditions, and computational efficiency. One promising direction is to refine the GRU-LSTM architecture by integrating attention mechanisms or adaptive learning rate strategies tailored to atmospheric time series. Additionally, expanding validation to multiple GNSS stations across diverse climatic regions will further test the model's robustness and enhance its applicability for regional tropospheric delay forecasting and atmospheric studies.

4. Conclusion

This study developed a hybrid GRU-LSTM deep learning model for short-term ZTD forecasting, integrating local meteorological parameters as exogenous inputs. The proposed model was trained and validated using multi-year GNSS-derived ZTD data from the NSRT station.

Experimental results demonstrated that the GRU-LSTM model outperformed traditional empirical models, statistical time series methods (e.g., SARIMAX, Prophet), and machine learning models (e.g., XGBoost, Random Forest), highlighting its ability to effectively capture both short-term fluctuations and long-term trends in ZTD variations.

Compared to standalone GRU and LSTM architectures, the hybrid model leveraged the strengths of both recurrent units, resulting in improved sequence modeling capability and reduced overfitting. The integration of local temperature and pressure further enhanced predictive accuracy, reducing seasonal influences over time, and ensuring higher accuracy and robustness in different meteorological conditions. These findings significantly improve the short-term ZTD forecasting accuracy at the NSRT station, demonstrating the effectiveness of integrating deep learning with local meteorological parameters. Furthermore, this approach provides valuable support for further enhancing the accuracy of GNSS positioning error corrections and atmospheric water vapor monitoring.

Acknowledgments

This research was funded by the CAS "Light of West China" Program (grant Nos. 2021-XBQNXZ-030 and 2021-XBQNXZ-005), the Xinjiang Key Laboratory of Radio Astrophysics (grant No. 2023D04064), and the National Key R&D Program of China (grant No. 2024YFA1611503).

The data used in this study were provided by the Crustal Movement Observation Network of China (CMONOC).

Appendix
Sensitivity Analysis of Input Window Length

To investigate how input sequence length affects the performance of the GRU-LSTM model, we tested five different window sizes: 7, 15, 30, 45, and 60 days. All other model parameters were held constant. The resulting performance metrics are summarized in Table A1.

Table A1
Performance Metrics of the GRU-LSTM Model Under Different Time Window Lengths

Days	RMSE (mm)	MSE	*R*	MAE	MBE	Skewness	Kurtosis
7	9.15	83.80	0.96	7.04	2.06	0.12	4.20
15	8.56	73.26	0.95	6.35	0.12	−0.33	4.78
30	**7.97**	**63.47**	**0.96**	**6.05**	**0.04**	**−0.25**	**3.90**
45	7.96	63.32	0.96	6.04	0.09	−0.26	3.85
60	7.95	63.19	0.95	6.03	0.19	−0.25	3.81

Note. Bold values indicate the 30-day input window length selected in this study as the optimal trade-off between prediction accuracy and stability. While longer windows increase computational cost, they do not provide significant accuracy gains compared with the 30-day configuration.

These results confirm that the 30 days window achieves the optimal trade-off between prediction accuracy and stability. Shorter windows suffer from higher bias and error, while longer windows do not yield significant improvements and may introduce unnecessary complexity.

ORCID iDs

Ming-Shuai Li https://orcid.org/0000-0002-8951-4053

References

Beutler, G., Rothacher, M., Schaer, S., et al. 1999, AdSpR, 23, 631
Bevis, M., Businger, S., Herring, T. A., et al. 1992, JGRD, 97, 15787
Boehm, J., Niell, A., Tregoning, P., & Schuh, H. 2006, GeoRL, 33, L07304
Boehm, J., & Schuh, H. 2004, GeoRL, 31, L01603
Chen, H., Lu, T., Huang, J., et al. 2023, RemS, 15, 3694
Chen, T., & Guestrin, C. 2016, in Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining *(Long Beach, CA, USA.)* (New York: ACM), 785
Cho, K., van Merriënnboer, B., Gulcehre, C., et al. 2014, in Proc. 2014 Conf. on Empirical Methods in Natural Language Processing (EMNLP) *(Doha, Qatar)* (Stroudsburg, PA, USA: Association for Computational Linguistics), 1724
Chollet, F. 2021, Deep Learning with Python (2nd ed.; Shelter Island, NY: Simon and Schuster), 478
Dee, D. P., Uppala, S., Simmons, A. J., & Berrisford 2011, QJRMS, 137, 553
Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. 2016, IEEE Transactions on Neural Networks and Learning Systems, 28, 2222
Haji-Aghajany, S., Rohm, W., Hadas, T., & Bosy, J. 2025, GPSS, 29, 1
Heinkelmann, R., Böhm, J., Schuh, H., & Tesmer, V. 2007, JGeod, 81, 483
Hochreiter, S., & Schmidhuber, J. 1997, Neural Computation, 9, 1735
Honma, M., Tamura, Y., & Reid, M. J. 2008, PASJ, 60, 951
Kobayashi, S., Ota, Y., Harada, Y., et al. 2015, JMeSJ, 93, 5
Li, J., Zhang, Q., Liu, L., et al. 2023, GPSS, 27, 172
Li, M. S., Li, R., Wang, N., & Zheng, X. W. 2020, RAA, 20, 200
Li, W., Yuan, Y., Ou, J., et al. 2015, JGeod, 89, 73
Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. 2021, International Journal of Forecasting, 37, 1748
Liu, J., Chen, X., Sun, J., & Liu, Q. 2017, AdSpR, 59, 824
Liwosz, T. 2013, ArtSa, 48, 191
Lu, C., Zheng, Y., Wu, Z., et al. 2023, JGeod, 97, 34
Puente, V. 2021, AdSpR, 67, 2359
Qu, W.-J., Zhu, W.-Y., Song, S.-L., & Ping, J.-S. 2008, ChA&A, 32, 429
Reid, M. J., Menten, K. M., Zheng, X. W., et al. 2009, ApJ, 700, 137
Rocken, C., Sokolovskiy, S., Johnson, J. M., & Hunt, D. 2001, JAtOT, 18, 1205
Saastamoinen, J. 1972, in The Use of Artificial Satellites for Geodesy, ed. S. W. Henriksen, A. Mancini, & B. H. Chovitz, Vol. 15 (Washington, DC: American Geophysical Union), 247
Schuh, H., & Böhm, J. 2013, in Sciences of Geodesy - II: Innovations and Future Developments, ed. G. Xu (Berlin: Springer), 339
Teke, K., Nilsson, T., Böhm, J., et al. 2013, JGeod, 87, 981
Tregoning, P., & Watson, C. 2009, JGRB, 114, B09403
Vaswani, A., Shazeer, N., Parmar, N., et al. 2017, Advances in Neural Information Processing Systems, 30, 6000
Wang, M., & Shen, Z. 2020, Earthquake Research in China, 36, 660
Wang, Y. B., Shi, H. B., Li, Y., et al. 2023, Earthquake Research in China, 39, 516
Zhang, H., Yao, Y., Xu, C., Xu, W., & Shi, J. 2022, RemS, 14, 3335
Zhou, W. L., Song, S. L., Hu, X., et al. 2022, RaSc, 57, 1