

Article

Zenith Tropospheric Delay Forecasting in the European Region Using the Informer–Long Short-Term Memory Networks Hybrid Prediction Model

Zhengdao Yuan ¹ , Xu Lin ^{1,2,*} , Yashi Xu ¹, Jie Zhao ¹, Nage Du ¹, Xiaolong Cai ¹ and Mengkui Li ³

- ¹ College of Earth and Planetary Science, Chengdu University of Technology, Chengdu 610059, China; yuanzhengdao@stu.cdut.edu.cn (Z.Y.); 2023050071@stu.cdut.edu.cn (Y.X.); zhaoj@stu.cdut.edu.cn (J.Z.); dunage@stu.cdut.edu.cn (N.D.); caixiaolong@stu.cdut.edu.cn (X.C.)
² State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu 610059, China
³ School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China; mkli@sgg.whu.edu.cn
* Correspondence: linux15@cdut.edu.cn

Abstract: Zenith tropospheric delay (ZTD) is a significant atmospheric error that impacts the Global Navigation Satellite System (GNSS). Developing a high-precision, long-term forecasting model for ZTD can provide valuable insights into the overall trends of predicted ZTD, which is essential for improving GNSS positioning and analyzing changes in regional climate and water vapor. To address the challenges of incomplete information extraction and gradient explosion in a single neural network when forecasting ZTD long-term, this study introduces an Informer–LSTM Hybrid Prediction Model. This model employs a parallel ensemble learning strategy that combines the strengths of both the Informer and LSTM networks to extract features from ZTD data. The Informer model is effective at capturing the periodicity and long-term trends within the ZTD data, while the LSTM model excels at understanding short-term dependencies and dynamic changes. By merging the features extracted by both models, the prediction capabilities of each can complement one another, allowing for a more comprehensive analysis of the characteristics present in ZTD data. In our research, we utilized ERA5-derived ZTD data from 11 International GNSS Service (IGS) stations in Europe to interpolate the missing portions of GNSS-derived ZTD. We then employed this interpolated data from 2016 to 2020, along with an Informer–LSTM Hybrid Prediction Model, to develop a long-term prediction model for ZTD with a prediction duration of one year. Our numerical results demonstrate that the proposed model outperforms several comparative models, including the LSTM–Informer based on a serial ensemble learning model, as well as the Informer, Transformer, LSTM, and GPT3 empirical ZTD models. The performance metrics indicate a root mean square error (RMSE) of 1.91 cm, a mean absolute error (MAE) of 1.45 cm, a mean absolute percentage error (MAPE) of 0.60, and a correlation coefficient (R) of 0.916. Spatial distribution analysis of the accuracy metrics showed that predictive accuracy was higher in high-latitude regions compared to low-latitude areas, with inland regions demonstrating better performance than those near the ocean. This study introduced a novel methodology for high-precision ZTD modeling, which is significant for improving accurate GNSS positioning and detecting water vapor content.



Academic Editor: Stephan Havemann

Received: 28 November 2024

Revised: 26 December 2024

Accepted: 28 December 2024

Published: 29 December 2024

Citation: Yuan, Z.; Lin, X.; Xu, Y.; Zhao, J.; Du, N.; Cai, X.; Li, M. Zenith Tropospheric Delay Forecasting in the European Region Using the Informer–Long Short-Term Memory Networks Hybrid Prediction Model. *Atmosphere* **2025**, *16*, 31. <https://doi.org/10.3390/atmos16010031>

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: zenith tropospheric delay; zenith tropospheric delay forecasting; neural network; informer; LSTM; combined model; ensemble learning

1. Introduction

When Global Navigation Satellite System (GNSS) signals travel through the neutral atmosphere to receivers, the propagation path of the signals is altered due to refraction caused by gases and water vapor. This phenomenon is known as tropospheric delay [1,2]. Tropospheric delay is typically expressed as the product of the zenith tropospheric delay (ZTD) and a mapping function corresponding to the direction of the propagation path [3,4]. Previous research has shown that when a satellite's local horizontal elevation angle drops to 10° , the ZTD can reach as high as 20 m [5], significantly limiting the positioning accuracy of GNSS. Therefore, having high-precision a priori ZTD values is crucial for GNSS navigation, particularly in applications like precise point positioning (PPP) [6,7] and real-time kinematic (RTK) [8,9]. In addition, in GNSS meteorology, a priori ZTD values are widely used to detect moisture content and serve as essential data sources for atmospheric analysis and numerical weather prediction [10,11]. Consequently, developing a high-precision long-term model to predict ZTD is vital for enhancing GNSS positioning accuracy and understanding the interactions within the atmosphere. This is of considerable scientific significance for precise GNSS positioning and moisture content detection [12,13].

Currently, commonly used models for predicting ZTD include those based on measured meteorological parameters, empirical models that do not require such parameters, and neural network models [14]. Models that rely on measured meteorological parameters use real-time data from a specific location, incorporating variables such as temperature, pressure and water vapor pressure to calculate ZTD [15–17]. However, the difficulty in obtaining these measured meteorological parameters limits the widespread application of these models. On the other hand, common empirical models without meteorological parameters include the UNB series models [18,19], the European Geostationary Navigation Overlay System (EGNOS) model [20] and the GPT series models [21–24], which primarily utilize global meteorological observation data and geographical observation data to obtain and calculate ZTD-related meteorological parameters through linear fitting, and then calculate the ZTD [25]. While these empirical models do not need direct measurements of meteorological parameters, their accuracy can be affected by the local environment [26].

In recent years, with the continuous development of machine learning (ML) theories, neural network algorithms such as Artificial Neural Networks (ANN) [27], Deep Neural Networks (DNN) [28], Convolutional Neural Networks (CNN) [29], and Long Short-Term Memory Networks (LSTM) [30,31] have provided various new approaches for ZTD prediction modeling. Among them, LSTM effectively addresses the gradient problem that traditional neural networks face when processing long sequences. It achieves this by incorporating a specialized gating structure, which enhances its ability to capture short-term dependencies in the data. As a result, LSTM performs exceptionally well in adapting to the rapid changes in ZTD data over short periods. For instance, Zhang et al. created an effective ZTD prediction model using LSTM in the western Antarctic region [32]. However, LSTM encounters challenges such as information loss and gradient issues when processing long input sequences spanning months or even years, which can limit its effectiveness in long-term ZTD prediction scenarios [33–35]. The attention mechanism intentionally assigns different weights to various elements by calculating the correlation between each element in a sequence and the others. This allows the model to focus on long-distance elements that significantly influence the current prediction. As a result, it captures long-term dependencies in time series data more effectively, offering a new solution for long-term time series prediction scenarios [36,37].

Zhang et al. applied a Transformer model to create a ZTD prediction model that learns complex patterns and dynamics from extensive ZTD time series through a self-attention mechanism. This enables the model to concentrate on more impactful information, thereby

enhancing prediction accuracy [38]. However, the Transformer model also faces issues related to high computational complexity and low processing efficiency for long sequences, which must be considered in practical applications [39].

To address these issues, Hu et al. (2024) developed a novel ZTD forecasting model based on the Informer model, which employs a ProbSparse Self-attention Mechanism to capture the global features of ZTD, effectively reducing the average time of the prediction model while maintaining an accuracy comparable to that of Transformer [40]. The authors utilized the predicted ZTD as an a priori constraint for PPP, which greatly enhances the speed of vertical convergence across all four seasons. The Informer model efficiently extracts key information from large time series datasets, accurately capturing long-term trends while significantly reducing computational complexity. However, it struggles to respond promptly to subtle short-term fluctuations and faces challenges related to incomplete information extraction and the neglect of short-term dependencies in the series [41]. Scholars have achieved promising results by employing ML methods to build ZTD prediction models. However, a single neural network prediction model often has limitations that can restrict the accuracy of ZTD forecasts [42,43].

Due to the limitations of a single prediction model, scholars often turn to ensemble learning to improve the model's prediction accuracy [44–46]. For example, Shi et al. (2024) introduced an integrated prediction model based on Transformer and LSTM. The model first uses LSTM to capture the dependency of input sequences and then utilizes the self-attention mechanism of Transformer to extract sequence information [47]. While this method improves the accuracy, it does face challenges related to high computational complexity and low sequence processing efficiency. Additionally, the assembly does not fully consider the limitations of LSTM when working with long-time series. Yuan et al. (2024) proposed a CNN–Informer combination model. In this model, CNN is utilized to extract the features of the sequences, and then the Informer model is employed to establish the relationship between inputs and outputs, resulting in strong performance [48]. However, this model may encounter issues related to missing information and overlooking short-term dependencies within sequences. Wang et al. proposed an LSTM–Informer model that employs a serial ensemble learning strategy. This approach first utilizes a low-level LSTM model to preprocess the data and then incorporates a top-level Informer model for prediction [49]. The combination method described has some issues, such as gradient explosion, which can occur in LSTM models when handling long time series, and the Informer may overlook short-term data dependencies.

The integrated models are designed to leverage the strengths of different models and enhance overall forecasting performance by utilizing their respective advantages through a serial approach. However, a combined model that utilizes serial ensemble learning often employs a method known as model stacking. This approach creates a strong dependency between individual models, meaning that the limitations of one model can negatively impact the performance of subsequent models. As a result, it does not effectively address the shortcomings of individual models or take full advantage of their strengths, leaving room for improvement in the prediction performance of the combined model. In contrast, some researchers have proposed that combined models based on a parallel ensemble learning strategy can achieve higher prediction accuracy and better generalization abilities [50]. The parallel ensemble learning model reduces the dependencies between individual models by training multiple independent models simultaneously. Each sub-model in this parallel approach can leverage its unique strengths to handle both long and short-term dependencies independently. This synergy among the sub-models ultimately enhances the performance of the overall model [51].

Based on the progress described above, this paper presents a novel approach that combines the LSTM model with the Informer model to create a hybrid prediction model called the Informer–LSTM Hybrid Prediction Model. This model employs a parallel ensemble learning strategy and is applied to long-term predictions of ZTD. Unlike existing models that integrate LSTM and Informer using a serial ensemble learning approach [49], our model allows the Informer and LSTM to function independently. We utilize LSTM to capture the short-term dependencies and dynamic changes in the ZTD time series. This helps to address the limitations of the Informer model, which may not fully extract information or adequately consider the short-term dependencies present in the series. Simultaneously, we employed the Informer model to extract global features from the ZTD time series, which helps improve issues related to information loss and gradient explosion that LSTM may encounter when processing long time series. We input the data into the LSTM and Informer models to generate feature vectors, which were then combined using a fully connected layer. This integration allowed us to utilize Informer’s global information extraction and LSTM’s local temporal modeling, effectively capturing both the long-term and short-term dependencies in the ZTD sequences. As a result, we enhanced the accuracy and robustness of long-term predictions for ZTD.

The major contributions of this paper are summarized as follows: 1. This paper introduces an Informer–LSTM Hybrid Prediction Model that utilizes a parallel ensemble learning strategy based on the LSTM and Informer models. This hybrid model combines the strengths of both the Informer and LSTM in capturing long-range dependencies in lengthy time series tasks, addressing the limitations of each model in prediction tasks. As a result, it enables a more comprehensive capture of data features. 2. We evaluated its performance against existing models, including an LSTM–Informer model based on serial ensemble learning, Informer, Transformer, LSTM, and a GPT3 empirical ZTD model, to showcase its superior effectiveness in ZTD long-time forecasting. 3. This paper offers a new perspective and methodology for ZTD forecasting, which has significant implications for improving GNSS positioning accuracy and enhancing meteorological predictions.

2. Study Area and Data

The study area is defined as the region in Europe that lies between 40° N and 70° N latitude and between 5° E and 35° E longitude. To train the model, we utilized the ZTD time series from 11 International GNSS Service (IGS) stations located within this region. The selected IGS stations are KIRU, METG, POLV, BUCU, GRAZ, GANP, PTBB, ZIM2, TLSG, MAD2 and MORP. Figure 1 illustrates the distribution of these IGS stations, while Table 1 provides specific information about each of them.

Table 1. Basic information of IGS stations used in this study.

IGS Stations	Latitude	Longitude	Elevation (m)	Data Integrity Rate
KIRU	67.857	20.968	390.9	93.05%
METG	60.242	24.384	59.7	88.94%
POLV	49.603	34.543	178.1	88.94%
BUCU	44.464	26.126	143.2	91.57%
GRAZ	47.067	15.493	538.3	91.35%
GANP	49.035	20.323	746	92.22%
PTBB	52.296	10.460	130.2	92.28%
ZIM2	46.877	7.465	956.5	93.10%
TLSG	43.550	1.485	208.7	77.29%
MAD2	40.429	−4.250	829.5	89.22%
MORP	55.213	−1.685	144.5	88.12%

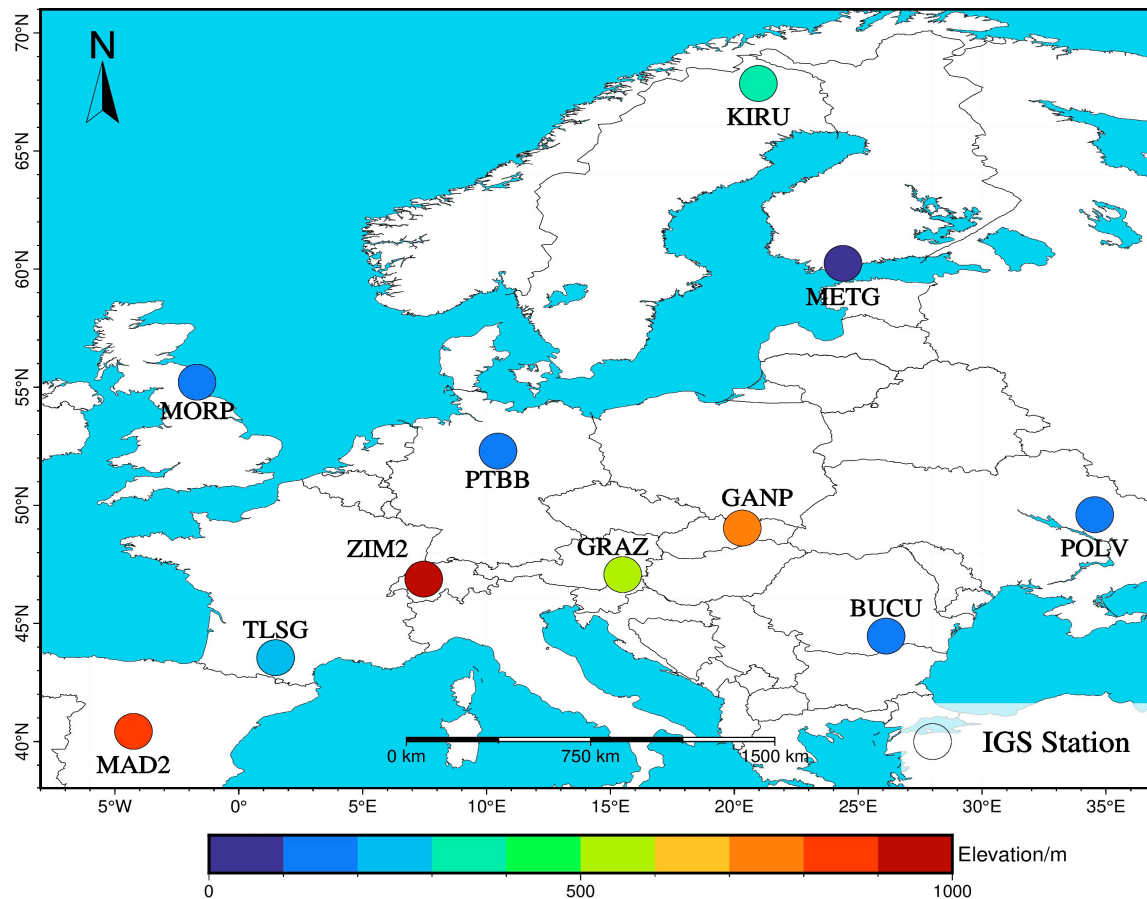


Figure 1. The location of IGS stations used in this study.

2.1. GNSS-Derived ZTD

Based on the observations from global GNSS stations, the IGS center has been releasing products in the form of IGS_ZPD files daily since 1998. Users can access these files on the following website (<https://cddis.gsfc.nasa.gov/archive/gnss/products/troposphere/zpd/>, accessed on 10 October 2024). The IGS_ZPD files provide information on station locations, observation details and GNSS-derived ZTD, along with associated observational errors, gradient north of the total troposphere and gradient north of the total troposphere. In this study, GNSS-derived ZTD was extracted from 11 IGS stations, with the selected observation period ranging from 1 January 2016, to 31 December 2020, and a temporal resolution of 6 h.

2.2. ERA5-Derived ZTD

ERA5 data represents the fifth generation of reanalysis products from the European Centre for Medium-Range Weather Forecasts (ECMWF). This pressure level dataset spans from 1940 to the present and has a spatial resolution of $0.25^\circ \times 0.25^\circ$ with a temporal resolution of 1 h. It includes data at 37 different pressure levels. The dataset used in this study can be accessed at the following website (<https://cds.climate.copernicus.eu/datasets/reanalysis-era5-pressure-levels>, accessed on 13 October 2024). The variables employed for calculating the ERA5-derived ZTD in this study include geopotential height, relative humidity and temperature.

3. Methods

3.1. Data Processing

The ZTD products provided by the IGS are internationally recognized as the highest precision observational data currently available, owing to their relatively low un-

certainty [52–55]. However, continuous observations from GNSS stations can often be disrupted by extreme weather conditions and equipment-related issues. This study interpolated the raw dataset to address the limitations posed by missing GNSS-derived ZTD data. Previous research has shown that the accuracy of ERA5-derived ZTD, computed from ERA5 reanalysis data, is comparable to that of GNSS-derived ZTD [56,57]. Building on this, our study utilized ERA5 data in conjunction with GNSS to calculate the ERA5-derived ZTD and applies interpolation to the GNSS-derived ZTD series. Specifically, first, we obtained the longitude, latitude, and elevation of the IGS stations. Using the ERA5 reanalysis data along with the station information, we then calculated the ERA5-derived ZTD for each station. The details of this calculation process will be described later. After that, we interpolated the GNSS-derived ZTD sequence for each site over a time scale. The goal was to use the calculated ERA5-derived ZTD to fill in the gaps in the GNSS-derived ZTD sequence, thereby creating a complete ZTD dataset. We refer to this interpolated data as GNSS-interpolation ZTD. These data were used to construct the Informer–LSTM hybrid ZTD prediction model and served as the reference for evaluating the model’s prediction accuracy.

The ERA5 reanalysis data are recognized for their high spatiotemporal resolution, offering comprehensive historical ZTD data for GNSS stations located throughout a three-dimensional spatial area. In this study, using an integral method, we utilized the ERA5 layered pressure reanalysis product to compute the zenith tropospheric wet delay (ZWD) for GNSS stations at varying heights. Additionally, we calculated the zenith tropospheric dry delay (ZHD) for the layer above the highest GNSS station using the Saastamoinen model. Combining these two components results in the ERA5-derived ZTD [58].

The specific calculation processes are given by Equations (1) and (2):

$$ZWD = 10^{-6} \int N ds = 10^{-6} \sum_i^{n-1} (N_i + N_{i+1}) \times (h_{i+1} - h_i) / 2 \quad (1)$$

$$ZHD = 0.0022793 \times \frac{[P_1 + (0.05 + \frac{1255}{T})e_1]}{f(\varphi, H)} \quad (2)$$

$$N = k_1(P - e)/T + k_2 \times e/T + k_3 \times e/T^2 \quad (3)$$

$$e = q \times P / 0.622 \quad (4)$$

$$f(\varphi, H) = 1 - 0.00266 \cos(2\varphi) - 2.8 \times 10^{-7} H \quad (5)$$

In these equations, n represents the total number of layers included in the reanalysis data above the GNSS station. The refractive index constants are $k_1 = 77.604$ (K/hPa), $k_2 = 64.79$ (K/hPa), and $k_3 = 3,754,630$ (K²/hPa). P denotes atmospheric pressure (unit: hPa), e represents water vapor pressure (unit: hPa), and T indicates temperature (unit: K). The values of these three variables can be obtained from the ERA5 reanalysis data. φ is the latitude corresponding to the GNSS station (unit: °), while P_1 , e_1 and H correspond to the atmospheric pressure, water vapor pressure, and elevation above sea level (units: m) at the site.

Since GNSS stations typically do not align with the grid points of the ERA5 reanalysis data and their elevations do not match the model layer elevations in the reanalysis, to begin with, we unified the elevation reference system of IGS stations with ERA5 reanalysis data. We then applied the EGM2008 model to convert the geodetic elevations of the IGS stations into positive elevations [59]. Next, we obtained the necessary meteorological parameters for Equations (1) to (4) through interpolation. The specific steps to calculate the ERA5-derived ZTD using the ERA5 reanalysis data are as follows:

1. Based on the coordinates of the IGS station, we identified the four nearest grid points. We performed linear interpolation of the meteorological parameters corresponding to these four grid points, ensuring that the elevation matches that of the GNSS station.
2. We applied the two-dimensional linear interpolation method to estimate the meteorological parameters at the location of the IGS station based on the values obtained from the four grid points in Step 1.
3. We used the two-dimensional linear interpolation method again to compute the meteorological parameters for the layers above the IGS station, ensuring that both the latitude and longitude coordinates aligned with the station coordinates and that the heights corresponded to those of the layers.
4. We calculated the ZWD and ZHD using Equations (1) and (2), respectively, and then summed the two results.

During the same time period and at the same time resolution, we used ERA5-derived ZTD to interpolate the GNSS-derived ZTD. This process allowed us to obtain the GNSS-interpolation ZTD time series for each IGS station in the study area. By comparing the box plots of the ZTD data before and after interpolation (refer to Figure 2), we observed that the two datasets exhibited strong similarities. This included their central tendency, degree of dispersion, and behavior of outliers. These findings suggest that the datasets share similar characteristics overall. For example, the GNSS-interpolation ZTD for the KIRU, GANP, POLV and TLSC stations was analyzed, as shown in Figure 3. The results indicate that using ERA5-derived ZTD to interpolate GNSS-derived ZTD effectively fills the data gaps. This approach achieves high precision in matching with GNSS-derived ZTD while adequately capturing its trends and periodic variations. By analyzing the characteristics of GNSS-interpolation ZTD, we found that values are generally higher in summer and lower in winter [60]. This pattern indicates significant annual periodicity and observable seasonal traits. Additionally, the local dynamic changes within the data series are sharp and irregular, highlighting notable instability and contributing to the challenges associated with ZTD prediction. Therefore, this study utilized an Informer–LSTM Hybrid Prediction Model to address these issues to develop a long-time predictive model for ZTD.

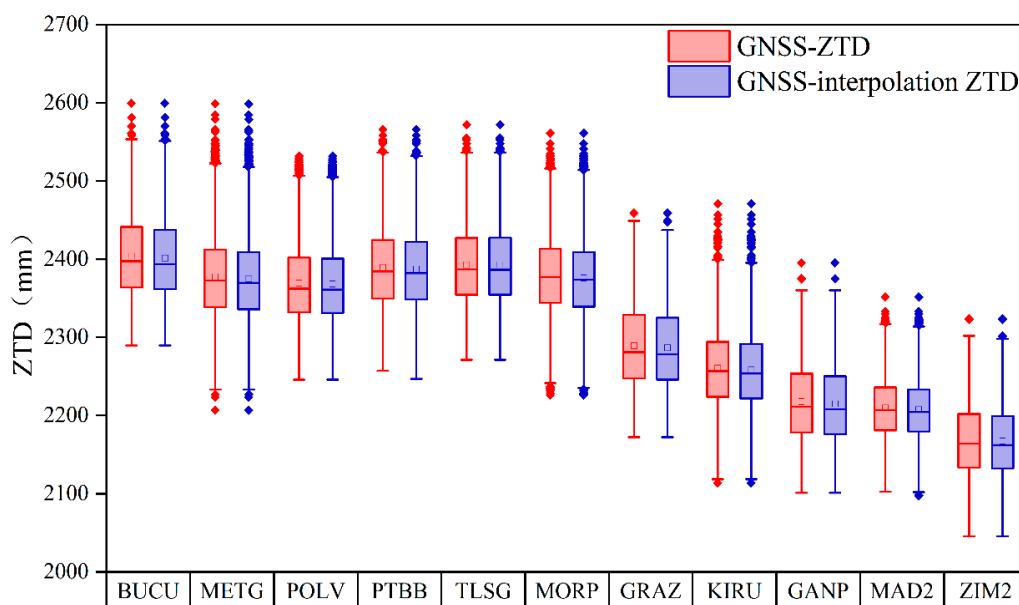


Figure 2. Comparison of the box plots of ZTD before and after interpolation for each IGS station used in this study. The line in the middle of the boxes represents the mean level of the data. In contrast, the upper and lower edges of the boxes correspond to the upper and lower quartiles, respectively, indicating the variability of the data. Points outside the boxes represent outliers in the data.

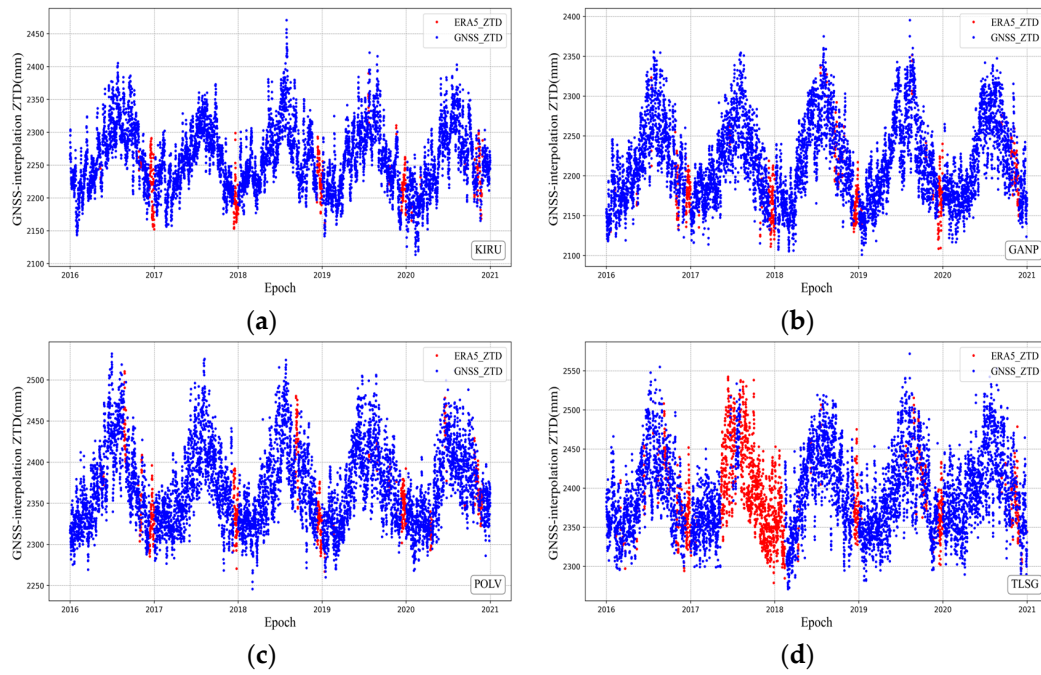


Figure 3. GNSS-interpolation ZTD from 2016 to 2020 for the selected IGS stations in this study. Panels (a), (b), (c), and (d) represent the stations KIRU, GANP, POLV and TLSG, respectively.

During the data preprocessing stage, we addressed the issue of missing data in GNSS-derived ZTD through interpolation. However, this approach may still introduce some uncertainties. For instance, the GNSS-interpolation ZTD may have inherent limitations and may not fully account for all possible atmospheric conditions and geographic changes. Additionally, the use of two-dimensional linear interpolation for meteorological parameters relies on the assumption that these parameters change linearly in space. In reality, variations in meteorological parameters are not always linear, which can lead to discrepancies between the interpolated results and the actual values. This, in turn, can affect the accuracy of predictions made by subsequent models that are based on GNSS-interpolation ZTD. Therefore, we will carefully consider these factors in our future studies to enhance the reliability and generalizability of our model.

3.2. The Forecasting Model of This Study

3.2.1. LSTM

The LSTM model effectively retains important historical information and discards irrelevant information through the control of gating units and the state transmission of memory units [34]. In the Informer–LSTM Hybrid prediction model, we leveraged LSTM to capture the sequential data's short-term dependencies and the dynamic variation present in the sequential data.

The structure of the LSTM model is shown in Figure 4, in which C_t represents the state of the LSTM memory cell, which is transmitted across time steps and serves to preserve the long-term memory of the network. f_t denotes the forget gate, i_t denotes the input gate, and O_t denotes the output gate. The mathematical formulas are as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (6)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (7)$$

$$\hat{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (8)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \hat{C}_t \quad (9)$$

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (10)$$

$$h_t = O_t \cdot \tanh(C_t) \quad (11)$$

First, f_t determines which information to delete from the cell state C_{t-1} . This process is represented by Equation (6), where W_f denotes the weights, σ represents the sigmoid activation function, and $[h_{t-1}, x_t]$ indicates the concatenated column vector of h_{t-1} and x_t , while b_f is the bias term. Next, i_t decides which new information to add to the cell state. After determining what information to update based on h_{t-1} and x_t , a candidate state $\hat{C}_t$ is generated using a tanh activation function. The historical cell state C_{t-1} is then updated to the new cell state C_t . This updating process is represented by Equations (7)–(9), where the different subscripts of W and b correspond to the respective weight matrices and bias terms. Finally, based on the inputs h_{t-1} and x_t , O_t is used to determine which features of the cell state should be output, and the output result h_t is computed. This process is represented by Equations (10) to (11), where W_o and b_o are the weights and biases for O_t and h_t is the output result. When an LSTM processes a sequence of data, it sequentially repeats the steps from the beginning to the end of the sequence at each time step. This allows the model to continuously update the state of the memory cells based on the information from the input sequence. As a result, it produces corresponding hidden states that are ultimately used for making predictions. This process effectively captures and manages both long-term and short-term dependencies in the sequence data.

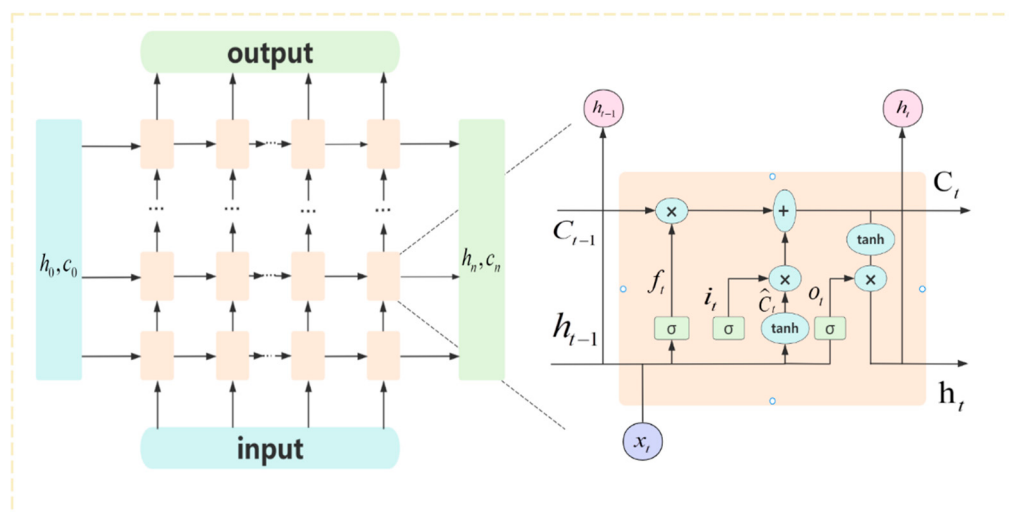


Figure 4. The architecture of the LSTM model.

Although the LSTM model offers several advantages, it encounters challenges when handling long sequence data. These challenges include difficulties in capturing long-term dependencies as well as issues related to vanishing or exploding gradients, which restrict its effectiveness in long-term time series prediction scenarios.

3.2.2. Informer

To address challenges such as the difficulties LSTM models face with gradient solving in long-time series, we propose the Informer–LSTM Hybrid Prediction Model. This model utilizes Informer to capture global features and long-term dependencies of the input data, enabling it to work in tandem with the LSTM model. This collaboration enhances the overall processing capabilities and prediction accuracy of the hybrid model for long-time series data. Firstly, the Informer model significantly reduces time complexity by incorporating the ProbSparse Self-Attention mechanism. In conventional self-attention models, the computational load increases dramatically with longer sequence lengths, leading to decreased efficiency when handling extended data sequences. The ProbSparse Self-attention mechanism adopts a probabilistic approach, actively ignoring those weights that have a relatively small impact on the final result. Its expression is shown in Equation (12):

$$A(Q, K, V) = \text{softmax}\left(\frac{\overline{Q}K^T}{\sqrt{d}}\right)V \quad (12)$$

In this context, Q , K , and V are three matrices of the same size obtained through linear transformations of the input data with hidden features. Here, d represents the feature dimension of the input data and Softmax is the normalization exponential function. $\overline{Q}$ is a matrix of the same size as Q , which includes only the attention weights that significantly impact the predictions compared to Q . The Informer model also uses a self-attention distillation process to downsample the data features. In addition, the information model downsamples data features through a self-attention distillation mechanism. The expression for the distillation mechanism from layer j to layer $j + 1$ in the network is given by Equation (13):

$$X_{j+1}^t = \text{MaxPool}(\text{ELU}(\text{conv1d}([X_j^t]_{AB}))) \quad (13)$$

In this context, conv1d denotes the one-dimensional convolution operation on the sequence, $\text{ELU}(\cdot)$ is the activation function, and MaxPool represents the downsampling process of the max pooling layer. Through distillation operations, the informer reduces the complexity of the data and reduces the cumulative error in the delivery of long series. Finally, the informer understands long-term trends and seasonal patterns in the data by adding timestamp information [39].

The structure of Informer is shown in Figure 5, which is mainly composed of two parts: an encoder and a decoder. The encoder's primary role is to transform the input data into intermediate features, while the decoder generates the forecast sequence based on these features produced by the encoder. Initially, the encoder employs a Prob-Sparse Self-attention mechanism combined with a feed-forward neural network to process the input data and extract key feature information. Following this, the decoder works to reconstruct the output data from the intermediate features. It produces final features by integrating the Multiple Self-attention mechanism, the Mask Self-attention mechanism, and another feed-forward neural network. The resultant features are then mapped to the final prediction outcomes via a fully connected layer. In the decoder, the input comprises both the hidden intermediate features generated by the encoder and the original input vectors. To prevent the model from relying on past information, the values intended for prediction are set to zero. This strategy allows the model to concentrate on forecasting future locations effectively.

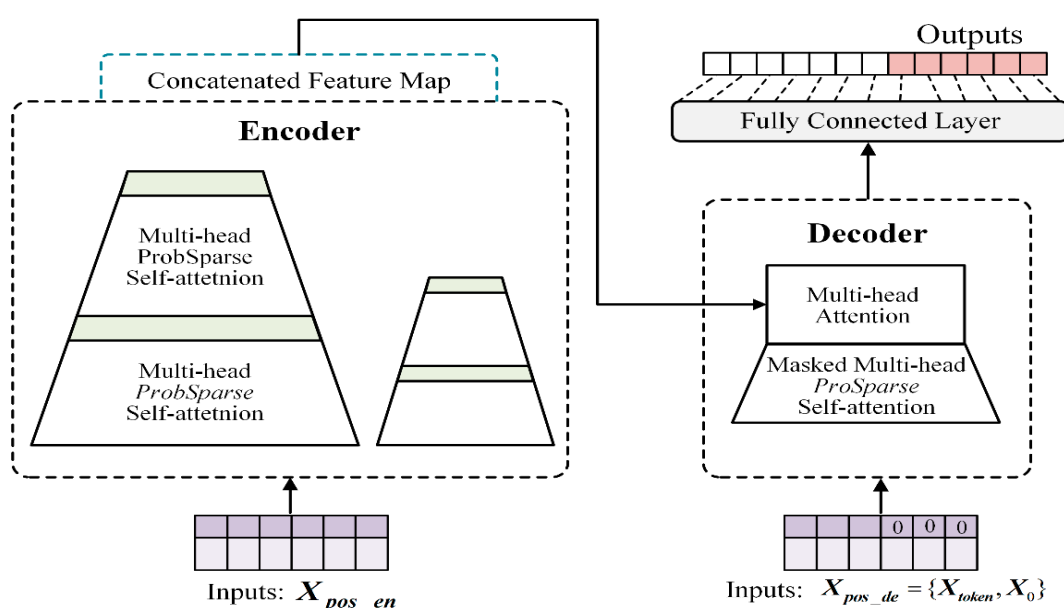


Figure 5. The architecture of the Informer model.

The Informer model excels in long-term time series forecasting but has some limitations. First, the ProbSparse Self-attention Mechanism improves computational efficiency but may not fully capture all significant dependencies in some situations. Second, while the Informer can identify long-term trends, it may struggle to learn dynamic changes in scenarios where the time series patterns shift rapidly.

3.2.3. Informer–LSTM Hybrid Prediction Model

As previously mentioned, both the informer and LSTM models have limitations in processing time series data. However, the Informer model demonstrates strong capabilities in handling long-time series. It excels at extracting global features and capturing long-term dependencies while minimizing computational resource consumption, making it highly applicable to various tasks. In contrast, LSTM models have a significant memory effect that enables them to effectively capture short-term dependencies and dynamic changes, resulting in impressive performance in data regression tasks. To harness the strengths of both the LSTM and Informer models and effectively capture both long-term and short-term dependencies in the data, this paper proposes an Informer–LSTM Hybrid Prediction Model that combines the features of both models. The overall structure of the proposed model is illustrated in Figure 6.

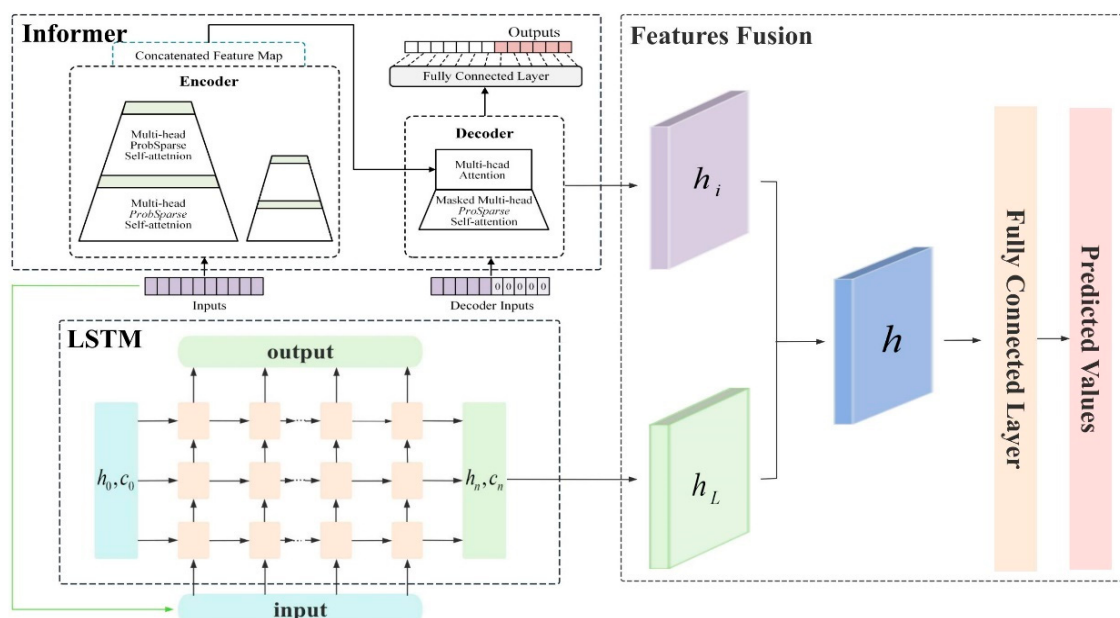


Figure 6. The architecture of the Informer–LSTM Hybrid Prediction Model.

The model is implemented using parallel learning with two branches: LSTM and Informer. Initially, the data is separately input into both the Informer and LSTM models. The Informer utilizes its strong capability in modeling long-term dependencies to extract global features. It employs the ProbSparse Self-attention Mechanism to capture the long-term trends in the sequence, compressing these features through a Self-attention distilling mechanism that ultimately generates the global feature h_i in the decoder. Meanwhile, LSTM leverages its memory effect to extract short-term dependencies and dynamic changes in the data. It retains useful information through three gating units and generating h_L . Subsequently, the model stacks and fuses h_i and h_L to integrate the feature representations from both the Informer and LSTM to create a unified feature representation, denoted as h , for the Informer–LSTM Hybrid Prediction Model. Finally, a fully connected layer maps these features to produce the final prediction results. The process of feature fusion is illustrated in Expressions (14) and (15):

$$h_{combine} = \text{concat}(h_i, h_L) \quad (14)$$

$$h = \text{projection}(h_{combine}) \quad (15)$$

First, the feature representations h_i and h_L are concatenated to obtain $h_{combine}$, where h_i represents the feature representation extracted by the Informer with a dimension of (B, T, D_1) and h_L represents the feature representation extracted by the LSTM with a dimension of (B, T, D_2) . Here, B denotes the batch size, T denotes the time steps, and D_1 and D_2 represent the feature dimensions of the Informer and LSTM, respectively. The concatenated feature representation $h_{combine}$ thus has a dimension of $(B, T, D_1 + D_2)$. This process aims to preserve the data features extracted by both and capture more fully the long-term and short-term dependencies of the time series data. Next, through a linear projection layer, $h_{combine}$ is projected to produce the final feature representation h , with an output dimension of (B, T, D_{out}) . This process aims to combine the feature representations of Informer and LSTM to address periodic trends and dynamic changes in the sequence data. Ultimately, the predicted value is derived from calculations performed by a fully connected layer. The parallel ensemble learning strategy enhances the limitations of both LSTM and Informer in predicting long time series while also improving the accuracy and robustness of the predictions.

Both LSTM and Informer possess unique strengths in time series forecasting. This strategy acknowledges the periodic trends and dynamic changes present in the sequential data, enabling a more thorough capture of both long-term and short-term dependencies. By integrating the characteristics of both models, we can overcome their individual limitations and improve prediction accuracy. This approach also enhances the model's robustness and generalization ability, making it more reliable and practical for real-world applications.

3.3. Construction of ZTD Forecast Model

In this research, we propose a ZTD prediction model by employing the Informer–LSTM Hybrid Prediction Model. The model is built upon GNSS-interpolation ZTD data, which possess a temporal resolution of 6 h and cover a time span from 1 January 2016 to 31 December 2020. These data were collected for each of the 11 IGS stations. For model training, we utilized ZTD data from each IGS station from 2016 to 2019, which constitutes approximately 80% of the dataset, while the remaining data served as the test set. The inputs to our model from each station comprised its historical GNSS-interpolation ZTD data along with the corresponding temporal information, allowing the model to learn the ZTD's temporal patterns. We designated one year as the prediction horizon, enabling us to forecast ZTD data for each station for the year 2020. In addition, for the ZTD prediction process, we used the method of forecasting by sliding window. We used the ZTD data of the past t calendar elements at each IGS station to forecast the ZTD data at the moment of $t + 1$ calendar elements. After completing one prediction, the ZTD value obtained from the new prediction was updated into the ZTD sequence and then continued to carry out the next round of prediction with the updated sequence in accordance with the same sliding-window method, and so on and so forth, continuously updating the ZTD sequence, thus realizing the long-term prediction of ZTD.

The model was constructed based on the PyTorch 1.13.1 framework. The initial learning rate of both the Informer and LSTM was set to 0.0003, the dropout rate was 0.1, the Batch-Size was set to 64, the GELU activation function was adopted, the Loss function was set to Mean Root Square, the number of epochs was set to 80 and the Adaptive Moment Estimation (ADAM) optimizer was used. In terms of the model architecture, the number of Encoder Layers in the Informer part was set to 2, the number of Decoder Layers was set to 1, the number of Multi-head was set to 8, the Encoder Input Length was set to 72, the Decoder Label Length was set to 48, the number of Hidden Layers was set to 2048 and the

number of Hidden features was set to 512. For the LSTM part, the number of layers was set to 3, the hidden size was set to 1000, and the input size was set to 1. To compare the prediction performance of the model in this paper with that of the contrast models, we also implemented the LSTM–Informer model, the Informer model, the Transformer model and the LSTM model in PyTorch. The device used was 1 NVIDIA GeForce RTX 4060 GPU. The device we used was a Lenovo Legion Y9000p laptop, which was manufactured by Lenovo Ltd. in Beijing, China.

Figure 7 illustrates the forecasting workflow of the proposed model. The steps are outlined as follows: First, ERA5-derived ZTD was used to interpolate GNSS-derived ZTD, resulting in GNSS-interpolation ZTD. Second, this GNSS-interpolation ZTD was standardized, and the data were divided into training and test sets at an 80:20 ratio. Finally, the model was trained, and the regression results from the trained model were evaluated against the known data, producing the prediction-fitting results.

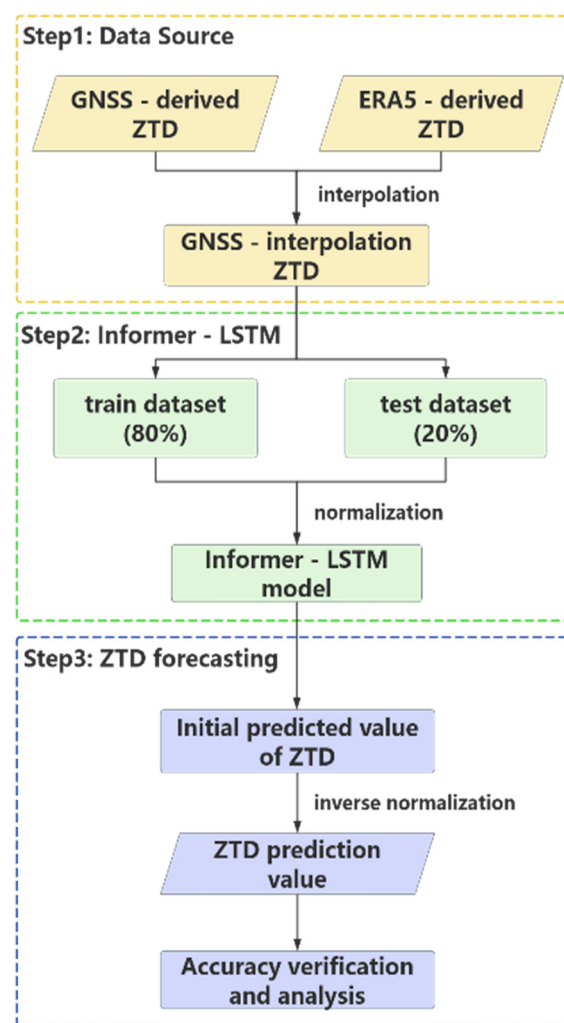


Figure 7. Flowchart of Informer–LSTM Hybrid ZTD prediction model.

3.4. Methods for Comparison

LSTM–Informer [49]: This model integrates LSTM and Informer through a strategy of serial ensemble learning. LSTM is utilized for processing time series data at the lower layer, while Informer is employed for feature extraction at the upper layer. The model demonstrates strong prediction performance.

Informer [39]: This model is a supervised learning framework based on the Transformer architecture. It utilizes a ProbSparse Self-Attention Mechanism for efficient global

feature extraction. It can generate prediction values in a single step, making it well-suited for long-term time series forecasting. It has achieved solid results in ZTD prediction tasks.

Transformer [61]: The Transformer model introduces a self-attention mechanism that has significantly advanced developments across various fields. In time series forecasting, this mechanism allows the model to dynamically adjust attention weights based on different parts of the input sequence, enabling it better to capture important features and patterns within the data.

LSTM [62]: LSTM is a classical algorithm for time series forecasting, employing memory cells to store and retain crucial information for current predictions. This capability allows LSTM to effectively capture short-term dependencies and dynamic characteristics in time series data.

GPT3 [24]: The GPT3 model can estimate temperature and pressure near the Earth's surface by utilizing the coordinates of observation stations and data over the years. This functionality allows for the calculation of ZTD. As an empirical model for ZTD prediction, GPT3 demonstrates good performance.

3.5. Accuracy Evaluation Indicators

In this study, we utilized the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) to compare the accuracy of different models. R^2 indicates how closely the prediction model results approximate the reference data. The closer R^2 is to 1, the better the independent variables explain the dependent variable in the regression analysis, representing strong predictive performance. RMSE measures the total deviation between the predicted values and the reference values. At the same time, MAE represents the average absolute deviation of each measurement, providing insight into the actual situation of ZTD prediction errors. MAPE ranges from $[0, +\infty]$, a MAPE of 0% indicates a perfect model, whereas a MAPE greater than 100% suggests a poor model. Smaller RMSE, MAE, and MAPE values indicate better predictive performance of the model. The calculation methods for each indicator are shown in Equations (16)–(19):

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_i^N (ZTD_i - ZTD_i^{pre})^2}{\sum_i^N (ZTD_i - \overline{ZTD_i})^2} \quad (16)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_i^N (ZTD_i - ZTD_i^{pre})^2} \quad (17)$$

$$MAE = \frac{1}{N} \sum_i^N \text{abs}(ZTD_i - ZTD_i^{pre}) \quad (18)$$

$$MAPE = \frac{100\%}{N} \sum_i^N \text{abs}\left(\frac{ZTD_i - ZTD_i^{pre}}{ZTD_i}\right) \quad (19)$$

where ZTD_i and ZTD_i^{pre} represent the GNSS-interpolation ZTD and the ZTD predictions output by the Informer–LSTM Hybrid Prediction Model, respectively, and $\overline{ZTD_i}$ is the mean value of the GNSS-interpolation ZTD.

4. Results and Discussion

In this study, we conducted ZTD prediction experiments at 11 IGS stations in Europe, using data from 2016 to 2019 as the training set to predict ZTD for 2020. We utilized the GNSS-interpolation ZTD data from 2020 for accuracy validation. We compared our approach with other models, including the LSTM–Informer based on the serial ensemble learning model, popular machine learning models (Informer, Transformer, and LSTM models) and the GPT3 empirical model to verify the superiority of our model in ZTD prediction modeling. In this section, Informer–LSTM refers to the Informer–LSTM Hybrid

Prediction Model proposed in this study, while LSTM-Informer is the comparison model based on a serial ensemble learning strategy. Each comparison model performs ZTD prediction under the same experimental conditions, including identical dataset division. Both models utilize data from 2016 to 2019 for training and validation using 2020 data, ensuring consistency in the data format and preprocessing methods. This setup allows for an objective comparison of accuracy between the method presented in this paper and the comparison models in the context of ZTD long-term series prediction.

Figure 8 and Table 2 present the statistical accuracy metrics of the different models. The statistical results indicate that the Informer–LSTM Hybrid Prediction model has the smallest average RMSE compared to the LSTM–Informer, Informer, Transformer, LSTM, and GPT3 models, with improvements in prediction accuracy of 6.37%, 7.72%, 6.83%, 23.91%, and 44.2%, respectively. It also achieves the lowest average MAE, with corresponding accuracy improvements of 6.45%, 8.23%, 7.05%, 27.86%, and 47.3%. The metrics R and MAPE also show significant improvements.

Table 2. Average Accuracy Comparison Among Different ZTD Prediction Models. In the table, Informer–LSTM refers to the Informer–LSTM Hybrid Prediction Model proposed in this study, while LSTM–Informer is the comparison model.

	RMSE (cm)	MAE (cm)	MAPE	R
Informer–LSTM	1.91	1.45	0.60	0.916
LSTM–Informer	2.04	1.55	0.66	0.905
Informer	2.07	1.58	0.68	0.901
Transformer	2.05	1.56	0.67	0.902
LSTM	2.51	2.01	0.87	0.841
GPT3	3.42	2.75	1.19	0.713

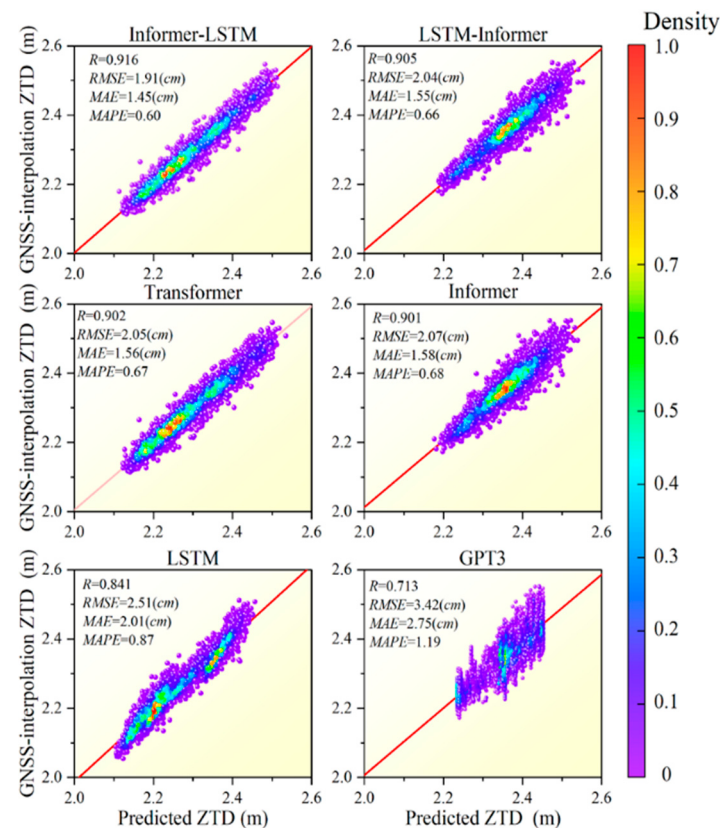


Figure 8. Comparison of ZTD prediction accuracy of different models. The figure uses colors to indicate the density of data points. It mainly shows the distribution of each model's predicted and true values on a two-dimensional plane.

The numerical results from all stations indicate that our Informer–LSTM Hybrid prediction model (shown as Informer–LSTM in the charts) achieved an average RMSE of 1.91 cm, an average MAE of 1.45 cm, an average MAPE of 0.60, and an average R of 0.916, all of which are the best among all models evaluated. This demonstrates our approach's advantage and the parallel ensemble learning strategy's effectiveness in constructing ZTD prediction models. The LSTM–Informer based on the serial ensemble learning model exhibited prediction accuracy comparable to the Informer and Transformer models but did not achieve ideal predictive performance. This may be attributed to the serial ensemble learning strategy's inability to address the Informer model's incomplete feature extraction limitations fully. Additionally, the prediction accuracy of the Transformer model was slightly higher than that of the Informer model, which may be related to the incomplete information extraction of the Informer model. The average RMSE for LSTM was 2.51 cm, higher than that of other comparable machine learning models, demonstrating the superiority of the attention mechanism in long-term sequence prediction. In contrast, the prediction accuracy of GPT3 was relatively low, likely due to the empirical model's difficulty in capturing the nonlinear characteristics of ZTD. This highlights the advantages of machine learning algorithms in constructing ZTD prediction models.

Figure 9 and Table A1 compare the prediction accuracy between this paper's model and the comparison model at 11 IGS stations. It is obvious from the bar charts that this paper's model outperforms the comparison model at all stations, and this result fully reflects the advantages of this paper's model in terms of prediction ability and robustness.

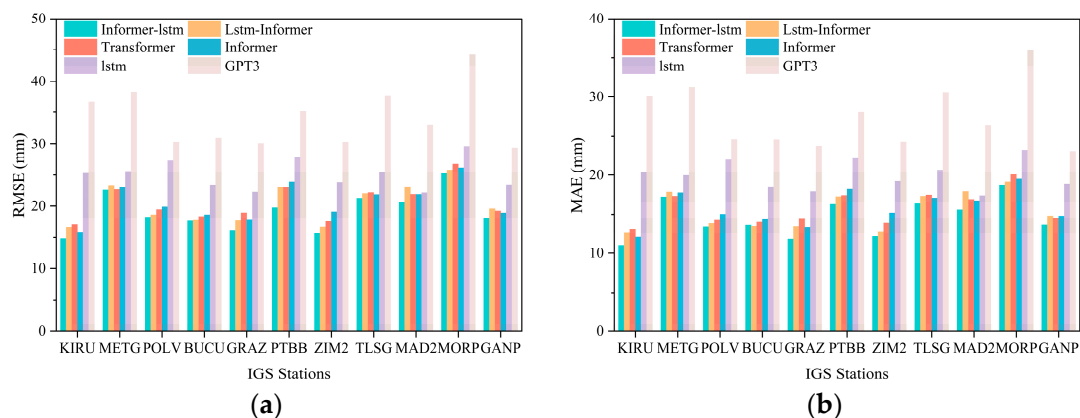


Figure 9. Comparison of prediction accuracy indexes of different ZTD prediction models for each station. (a) RMSE, (b) MAE.

Figure 10 shows the prediction effect of this paper's model at the different locations of the stations. We chose the KIRU and GRAZ stations with higher prediction accuracy and the METG and MORP stations with lower accuracy for demonstration. Through comparison, it can be found that this paper's model fits the ZTD time series better, not only successfully capturing the long-term trend and cyclic changes of ZTD but also performs well in fitting the local dynamics, which further verifies that the parallel ensemble learning strategy of this model gives full play to the respective advantages of the Informer model and the LSTM model.

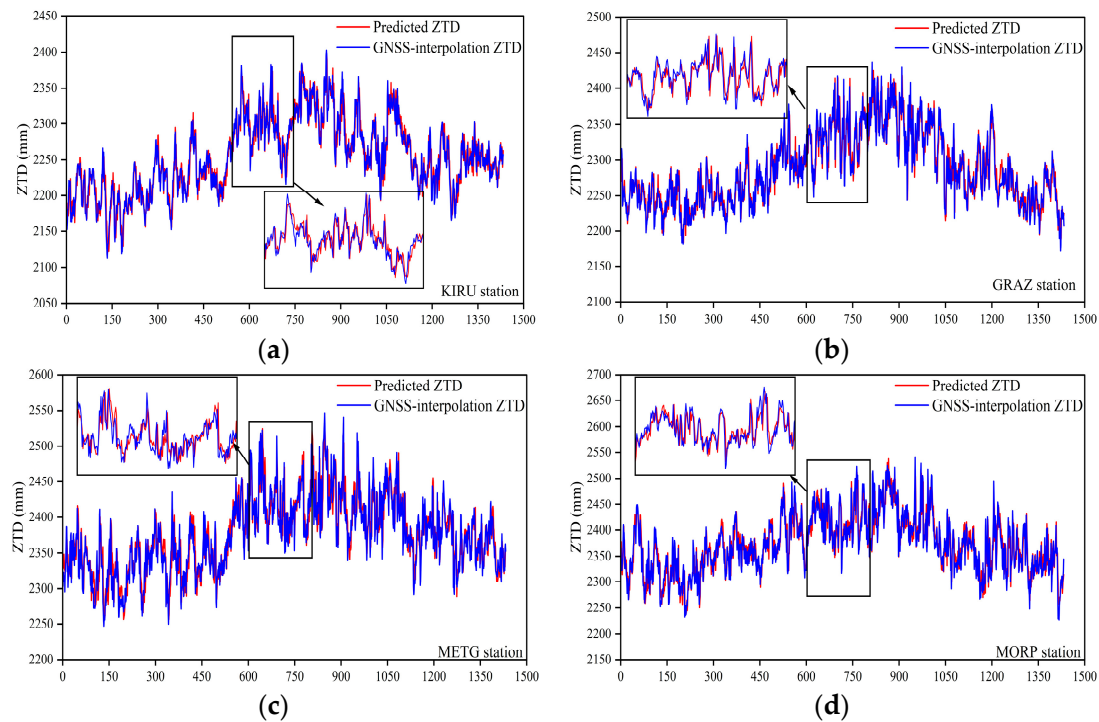


Figure 10. Comparison of the prediction results of the Informer–LSTM Hybrid ZTD Prediction Model with the GNSS interpolation ZTD. The horizontal coordinates represent the predicted four moments of each day in 2020 (0:00, 6:00, 12:00, and 18:00), and the vertical coordinates represent the ZTD values. Panels (a), (b), (c), and (d) represent the stations KIRU, GRAZ, METG and MORP, respectively.

Many studies have been conducted in the field of ZTD prediction, and the results vary from one study to another due to the differences in the data, models and experimental setups used. For example, in the study conducted by Hu et al. in 2024, they used the Informer model to predict ZTD for nine GNSS stations in China. After counting the overall 24 h prediction ZTD accuracies of all the GNSS station models in 2022, the average RMSE is about 2.21 cm and the average MAE is about 1.66 cm. The average RMSEs of the comparative models, namely Transformer and LSTM in this study, are 2.15 cm and 3.65 cm, respectively [39]. In comparison, our proposed Informer–LSTM Hybrid Prediction Model obtains an average RMSE of 1.91 cm and an average MAE of 1.45 cm, which are significantly better than the results of this previous study in terms of accuracy metrics. This is partly due to our innovative parallel integrated learning combination strategy, which gives full play to the respective advantages of the Informer model and the LSTM model; on the other hand, it may also be related to the pre-processing method of the ZTD data we selected and the spatial and temporal coverage of the experimental data.

To further analyze the variations in accuracy of the Informer–LSTM Hybrid Prediction Model proposed in this study, we examined its global spatial distribution. Figure 11 illustrates the overall distribution of accuracy metrics for the Informer–LSTM Hybrid Prediction model. Our spatial distribution analysis of the RMSE reveals that prediction accuracy is relatively low at the MORP, METG, and TLSG stations. In contrast, the IGS stations in inland areas demonstrate significantly higher prediction accuracy than coastal regions. We speculate that this phenomenon may be attributed to the higher water vapor content at coastal stations, which increases the impact of ZWD on ZTD, complicating the model’s ability to learn the characteristics of ZTD. Additionally, we observed that KIRU exhibited the highest prediction accuracy, with a correlation coefficient R approaching 0.957. However, even some inland stations that were less affected by water vapor did not achieve

ideal prediction accuracy. This leads us to believe that latitude may also play a role in influencing the prediction accuracy of the stations.

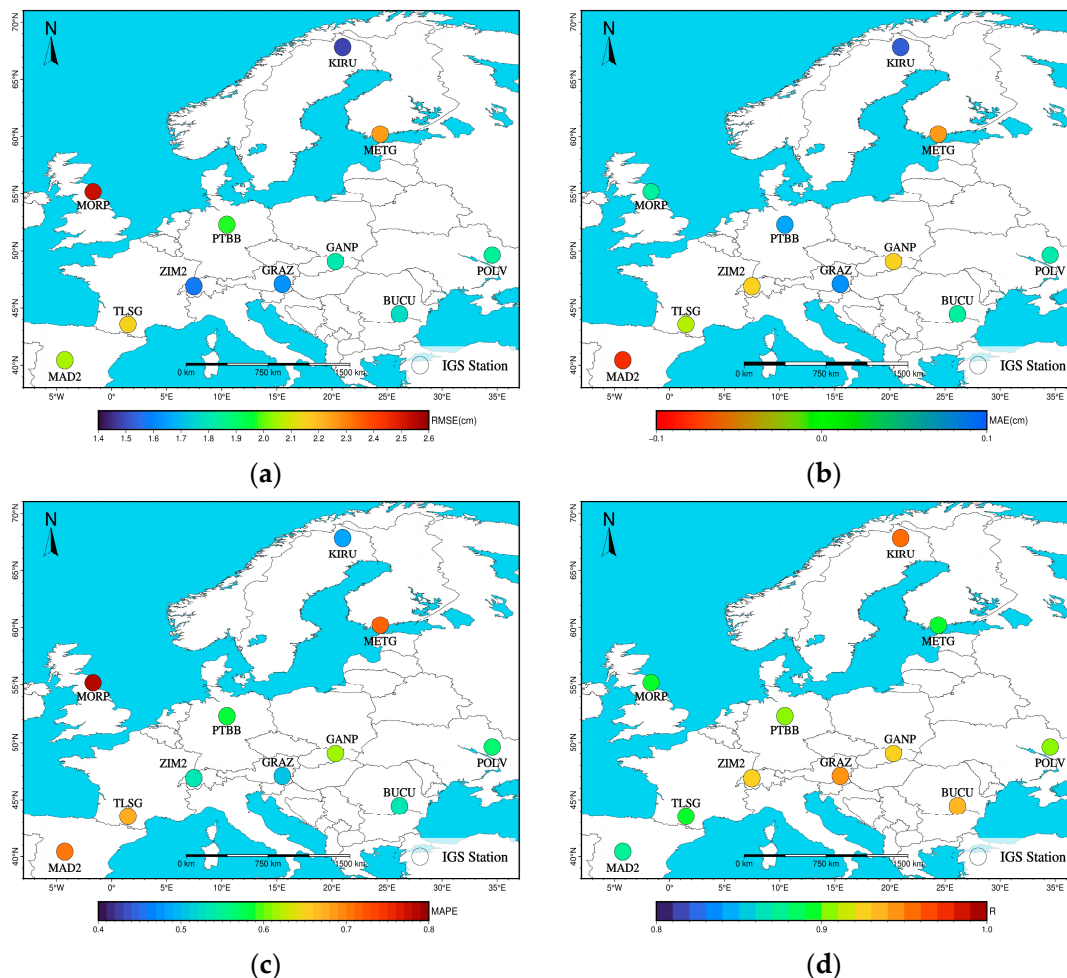


Figure 11. Spatial distribution of forecast accuracy metrics: (a) RMSE, (b) MAE, (c) MAPE, (d) R.

Since accurate prediction of ZTD is of great significance to improving the accuracy of GNSS positioning. Our proposed Informer–LSTM Hybrid Prediction Model, with its high prediction accuracy, is able to provide more reliable atmospheric delay correction information for PPP. This capability significantly enhances the accuracy of positioning results. Additionally, the variation of ZTD can reflect regional meteorological characteristics and water vapor dynamics processes. Our model can accurately capture the long-term trends, cyclic changes, and local dynamics of ZTD. Based on the prediction results, meteorologists can further invert the atmospheric water vapor distribution and trends, which is potentially valuable for short-term weather prediction, climate monitoring, and early warning of meteorological disasters.

In addition, while the Informer–LSTM Hybrid Prediction Model proposed in this study yields satisfactory prediction results for the data from 11 IGS stations in the European region, certain limitations may arise when extending its use to other regions or different datasets. Firstly, the significant variations in atmospheric, topographic, and climatic conditions across different areas can lead to distinct ZTD characteristics. Consequently, when applying the model to regions with considerable climatic differences, it may struggle to accurately capture the local ZTD's changing patterns due to mismatched data characteristics, ultimately affecting the prediction accuracy. Secondly, our model is trained optimally based on a specific structure and parameter settings tailored to European regional data. As

we promote its application, inherent characteristics of the model could limit its generalizability. To enhance the model's performance, several improvements can be made. Firstly, expanding and enriching the training dataset by collecting more samples from various regions and types can facilitate better learning of diverse ZTD change patterns, thereby improving the model's adaptability to different data characteristics. Secondly, conducting model fine-tuning or re-training for particular regions allows for the optimization and adjustment of hyperparameters based on specific climate and terrain information, ensuring a better fit with local data characteristics.

5. Conclusions

Aiming at the limitations of a single neural network in ZTD prediction, an Informer–LSTM Hybrid Prediction Model based on parallel ensemble learning was proposed and applied to ZTD long-time series prediction. We utilized the Informer to extract global features of ZTD data and employed LSTM to capture the dynamic variations of the ZTD data. We obtained a more comprehensive representation of ZTD features by adopting a stacked fusion approach. This model effectively compensates for the shortcomings of both LSTM and Informer in ZTD predictions, enabling us to better capture the long-term and short-term dependencies of ZTD data, thereby enhancing the accuracy of ZTD predictions.

We initially utilized the GNSS source ERA5-derived ZTD to interpolate GNSS-derived ZTD, resulting in what we called GNSS-interpolation ZTD. Building on this, we developed the Informer–LSTM Hybrid ZTD Prediction model. This model was then applied to forecast ZTD variations at the 11 IGS stations in Europe in 2020.

To evaluate the predictive performance of our model, we compared it with several benchmark models, including the LSTM–Informer based on serial ensemble learning, Informer, Transformer, LSTM, and empirical ZTD models. The quantitative analysis results indicate significant improvements in our algorithm's average RMSE, MAE, MAPE, and R^2 . When compared to LSTM–Informer, Informer, Transformer, LSTM, and GPT3, the Informer–LSTM Hybrid Prediction model achieved the lowest average RMSE, with prediction accuracy improvements of 6.37%, 7.72%, 6.83%, 23.91%, and 44.2%, respectively. It also exhibited the lowest average MAE, with prediction accuracy improvements of 6.45%, 8.23%, 7.05%, 27.86%, and 47.3%, respectively. The proposed Informer–LSTM Hybrid Prediction model not only surpasses individual ML models and empirical ZTD models in prediction accuracy but also outperforms LSTM–Informer based on serial ensemble learning, thoroughly validating the effectiveness and superiority of our approach. Additionally, through spatial distribution analysis of accuracy metrics, we concluded that prediction accuracy is higher in high-latitude regions compared to low-latitude regions, and inland areas show greater prediction accuracy than those near the ocean.

The accuracy of ZTD prediction significantly impacts GNSS positioning accuracy. Our proposed Informer–LSTM Hybrid Prediction Model demonstrates high accuracy in ZTD predictions, which can positively influence GNSS positioning outcomes. By leveraging the precise ZTD predictions from this model, we anticipate a reduction in positioning errors, leading to improved reliability and accuracy of the positioning results. This advancement is expected to provide robust technical support for the high-quality development of industries that depend on precise positioning.

In addition, considering the significant differences in atmospheric conditions, water vapor changes, and ZTD characteristics across various climate regions, we plan to apply the Informer–LSTM Hybrid Prediction Model to a broader range of climatic zones for further testing in future research. By conducting experiments in diverse climatic environments, we can more comprehensively evaluate the generalization ability of the model and identify potential challenges it may face in dealing with different atmospheric conditions. This

approach will allow us to optimize and adjust the model, enhancing its adaptability for a wider array of application scenarios.

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

Funding: This research was funded by the National Natural Fund of China grants (grant number 42271461), the Natural Science Foundation of Sichuan Province (grant number 24NSFSC1797) and the State Key Laboratory of Geohazard Prevention and Geoenvironment Protection Independent Research Project (SKLGP2021Z022).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The GNSS-derived ZTD data used in our experiments were obtained from data published on NASA’s CDDIS website (<https://cddis.gsfc.nasa.gov/archive/gnss/products/troposphere/zpd/>) (accessed on 10 October 2024). The ERA5-derived ZTD we used for performing interpolation on the GNSS-derived ZTD data were obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) website (<https://cds.climate.copernicus.eu/datasets/reanalysis-era5-pressure-levels>) (accessed on 13 October 2024). We thank the CDDIS and ECMWF for providing the relevant data products.

Acknowledgments: The authors would like to thank the reviewers for their careful reading of our paper and for their valuable suggestions for revision, which have improved its presentation.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A. Prediction Metrics Among Different Models for IGS Stations

Table A1. Comparison of prediction accuracy indicators of different ZTD prediction models for each station (cm). In the table, Informer–LSTM refers to the Informer–LSTM Hybrid Prediction Model proposed in this study, while LSTM–Informer is the comparison model.

Site	Accuracy	Informer–LSTM	LSTM–Informer	Informer	Transformer	LSTM	GPT3
KIRU	RMSE	1.48	1.67	1.59	1.71	2.53	3.67
	MAE	1.10	1.27	1.21	1.31	2.03	3.01
METG	RMSE	2.25	2.33	2.30	2.26	2.56	3.83
	MAE	1.72	1.78	1.77	1.73	2.00	3.12
POLV	RMSE	1.83	1.87	1.99	1.95	2.73	3.03
	MAE	1.34	1.38	1.50	1.43	2.19	2.46
BUCU	RMSE	1.77	1.78	1.87	1.84	2.34	3.10
	MAE	1.36	1.35	1.44	1.40	1.85	2.46
GRAZ	RMSE	1.62	1.78	1.79	1.90	2.22	3.00
	MAE	1.18	1.35	1.34	1.44	1.78	2.37
PTBB	RMSE	1.97	2.30	2.39	2.30	2.78	3.52
	MAE	1.62	1.72	1.81	1.73	2.21	2.80
ZIM2	RMSE	1.58	1.68	1.91	1.76	2.38	3.04
	MAE	1.22	1.28	1.52	1.39	1.92	2.43

Table A1. Cont.

Site	Accuracy	Informer-LSTM	LSTM-Informer	Informer	Transformer	LSTM	GPT3
TLSG	RMSE	2.13	2.20	2.18	2.21	2.55	3.78
	MAE	1.63	1.73	1.70	1.74	2.07	3.06
MAD2	RMSE	2.06	2.31	2.18	2.19	2.21	3.30
	MAE	1.55	1.78	1.67	1.68	1.73	2.64
MORP	RMSE	2.52	2.58	2.62	2.68	2.96	4.42
	MAE	1.87	1.91	1.95	2.00	2.32	3.59
GANP	RMSE	1.82	1.96	1.90	1.93	2.34	2.94
	MAE	1.37	1.48	1.48	1.45	1.89	2.31

References

- Bevis, M.; Businger, S.; Herring, T.A.; Rocken, C.; Anthes, R.A.; Ware, R.H. GPS Meteorology' Remote Sensing of Atmospheric Water Vapor Using the global positioning system. *J. Geophys. Res. Atmos.* **1992**, *97*, 15787–15801. [\[CrossRef\]](#)
- Xia, P.; Ye, S.; Guo, M.; Jiang, W.; Xu, C. Retrieval of Tropospheric Refractivity profiles using Slant Tropospheric Delays Derived from a single ground-based global Navigation Satellite System Station. *Earth Space Sci.* **2019**, *6*, 1081–1097. [\[CrossRef\]](#)
- Yang, L.; Wang, J.; Li, H.; Balz, T. Global Assessment of the GNSS Single Point Positioning Biases Produced by the Residual Tropospheric Delay. *Remote Sens.* **2021**, *13*, 1202. [\[CrossRef\]](#)
- Sun, P.; Zhang, K.; Wu, S.; Wan, M.; Lin, Y. Retrieving Precipitable Water Vapor from Real-Time Precise Point Positioning Using VMF1/VMF3 Forecasting Products. *Remote Sens.* **2021**, *13*, 3245. [\[CrossRef\]](#)
- Zhao, J.; Song, S.; Chen, Q.; Zhou, W.; Zhu, W. Establishment of a global tropospheric zenith delay model based on the functional form of vertical profiles. *Chin. I. Geophys.* **2014**, *57*, 3140–3153. [\[CrossRef\]](#)
- Lou, Y.; Huang, J.; Zhang, W.; Liang, H.; Zheng, F.; Liu, J. A New Zenith Tropospheric Delay Grid product for real-time PPP applications over China. *Sensors* **2018**, *18*, 65. [\[CrossRef\]](#)
- Xia, P.; Tong, M.; Ye, S.; Qian, J.; Fang, H. Establishing a high precision real-time ZTD model of China with GPS and ERA5 historical data and its application in PPP. *GPS Solut.* **2023**, *27*, 2. [\[CrossRef\]](#)
- Xu, Y.; Wu, C.; Li, L.; Yan, L.; Liu, M.; Wang, S. GPS/BDS Medium/Long-Range RTK Constrained with Tropospheric Delay Parameters from NWP Model. *Remote Sens.* **2018**, *10*, 1113. [\[CrossRef\]](#)
- Dabov, P.; Bagheri, M. Enhancing Atmospheric Monitoring Capabilities: A Comparison of Low- and High-Cost GNSS Networks for Tropospheric Estimations. *Remote Sens.* **2024**, *16*, 2223. [\[CrossRef\]](#)
- Mohammed, J. Adaptive neuro fuzzy inference system for predicting sub-daily Zenith Wet Delay. *J. Geod. Geodyn.* **2022**, *13*, 352–362. [\[CrossRef\]](#)
- Gong, Y.; Liu, Z.; Chan, P.; Hon, K. Assimilating GNSS PWV and radiosonde meteorological profiles to improve the PWV and rainfall forecasting performance from the Weather Research and forecasting (WRF) model over the South China. *Atmos Res.* **2023**, *286*, 106677. [\[CrossRef\]](#)
- Lian, d.; He, Q.; Li, L.; Fu, E.; Zhang, K. Accuracy Assessment of ERA5-ZTD/PWV and Response of Typhoon Events in China. *J. Catal.* **2024**, *39*, 23–28. [\[CrossRef\]](#)
- Zhang, H.; Yao, Y.; Xu, C.; Hu, M.; Tang, F.; Ji, C.; Ma, X. A Novel Multilayer Perceptron-Based Nonmeteorological Parameters PWV Retrieval Model. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–9. [\[CrossRef\]](#)
- Yang, F.; Guo, J.; Zhang, C.; Li, Y.; Li, J. A Regional Zenith Tropospheric Delay (ZTD) Model Based on GPT3 and ANN. *Remote Sens.* **2021**, *13*, 838. [\[CrossRef\]](#)
- Black, H.D. An easily implemented algorithm for the tropospheric range correction. *J. Geophys. Res. Solid Earth* **1978**, *83*, 1825–1828. [\[CrossRef\]](#)
- Hopfield, H.S. Two-quartic tropospheric refractivity profile for correcting satellite data. *J. Geophys. Res.* **1969**, *74*, 4487–4499. [\[CrossRef\]](#)
- Saastamoinen, J. Contributions to the theory of atmospheric refraction. *Bull. Géodésique (1946–1975)*. **1972**, *105*, 279–298. [\[CrossRef\]](#)
- Leandro, R.; Santos, M.; Langley, R. UNB Neutral Atmosphere Models: Development and Performance. In Proceedings of the 2006 National Technical Meeting of the Institute of Navigation, Monterey, CA, USA, 18–20 January 2006; pp. 564–573. [\[CrossRef\]](#)
- Leandro, R.F.; Langley, R.; Santos, M. UNB3m_pack: A neutral atmosphere delay package for radiometric space techniques. *GPS Solut.* **2008**, *12*, 65–70. [\[CrossRef\]](#)
- Penna, N.; Dodson, A.; Chen, W. Assessment of EGNOS Tropospheric Correction Model. *J. Navig.* **2001**, *54*, 37–55. [\[CrossRef\]](#)
- Böhm, J.; Heinkelmann, R.; Schuh, H. Short Note: A global model of pressure and temperature for geodetic applications. *J. Geod.* **2007**, *81*, 679–683. [\[CrossRef\]](#)

22. Lagler, K.; Schindelegger, M.; Böhm, J.; Krásná, H.; Nilsson, T. GPT2: Empirical slant delay model for radio space geodetic techniques. *Geophys. Res. Lett.* **2013**, *40*, 1069–1073. [\[CrossRef\]](#)
23. Böhm, J.; Möller, G.; Schindelegger, M.; Pain, G.; Weber, R. Development of an improved empirical model for slant delays in the troposphere (GPT2w). *GPS Solut.* **2015**, *19*, 433–441. [\[CrossRef\]](#)
24. Landskron, D.; Böhm, J. VMF3/GPT3: Refined discrete and empirical troposphere mapping functions. *J. Geod.* **2018**, *92*, 349–360. [\[CrossRef\]](#)
25. Wang, Y.; Jin, S.; Liu, C.; Zhang, G.; Zhang, Y. A real-time ZTD model in China using ERA5 and GNSS based on empirical orthogonal function method. *GPS Solut.* **2025**, *29*, 11. [\[CrossRef\]](#)
26. Osah, S.; Acheampong, A.A.; Fosu, C.; Dadzie, I. Deep learning model for predicting daily IGS zenith tropospheric delays in West Africa using TensorFlow and Keras. *Adv. Space. Res.* **2021**, *68*, 1243–1262. [\[CrossRef\]](#)
27. Pikridas, C.; Katsougiannopoulos, S.; Ifadis, I. Predicting Zenith Tropospheric Delay using the Artificial neural network technique. *EUREF 2010 Annual Symp.* **2010**.
28. Anantrasirichai, N.; Biggs, J.; Albino, F.; Bull, D. A deep learning approach to detecting volcano deformation from satellite imagery using synthetic datasets. *Remote Sens. Environ.* **2019**, *230*, 111179. [\[CrossRef\]](#)
29. Shi, J.; Li, X.; Li, L.; Ouyang, C.; Xu, C. An efficient deep learning-based troposphere ZTD dataset generation method for massive GNSS CORS stations. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–11. [\[CrossRef\]](#)
30. He, L.; Yao, Y.; Xu, C.; Zhang, H.; Tang, F.; Ji, C.; Liu, Z.; Wu, W.; He, L.; Yao, Y.; et al. A New Global ZTD Forecast Model Based on Improved LSTM Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2024**, *17*, 9606–9614. [\[CrossRef\]](#)
31. Zhang, H.; Yao, Y.; Hu, M.; Xu, C.; Su, X.; Che, D.; Peng, W. A Tropospheric Zenith Delay forecasting Model based on a Long Short Term Memory Neural Network and its impact on Precise Point Positioning. *Remote Sens.* **2022**, *14*, 5921. [\[CrossRef\]](#)
32. Zhang, Q.; Li, F.; Zhang, S.; Li, W. Modeling and forecasting the GPS zenith troposphere delay in west Antarctica based on different blind source separation methods and deep learning. *Sensors* **2020**, *20*, 2343. [\[CrossRef\]](#)
33. Greff, K.; Srivastava, R.K.; Koutník, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A search space odyssey. *IEEE Trans. Neural Networks. Learn. Syst.* **2017**, *28*, 2222–2232. [\[CrossRef\]](#)
34. Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D.* **2020**, *404*, 132306. [\[CrossRef\]](#)
35. Bandara, K.; Bergmeir, C.; Smyl, S. Forecasting across time series databases using long short-term memory networks on groups of similar series. *Mach. Learn.* **2017**, *8*, 805–815. [\[CrossRef\]](#)
36. Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv* **2014**, arXiv:1409.0473.
37. Manghi, R.L.; Bernacchia, D.; Casajus, L.G.; Zannoni, M.; Tortora, P.; Martellucci, A.; De Vicente, J.; Villalvilla, J.; Maschwitz, G.; Cappuccio, P.; et al. Tropospheric delay calibration system performance during the first two BepiColombo solar conjunction. *Radio Sci.* **2023**, *58*, 1–15. [\[CrossRef\]](#)
38. Zhang, H.; Yao, Y.; Xu, C.; Xu, W.; Shi, J. Transformer-based global Zenith Tropospheric Delay forecasting model. *Remote Sens.* **2022**, *14*, 3335. [\[CrossRef\]](#)
39. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtually, 2–9 February 2021; Volume 35, pp. 11106–11115. [\[CrossRef\]](#)
40. Hu, F.; Sha, Z.; Wei, P.; Xia, P.; Ye, S.; Zhu, Y.; Luo, J. Deep learning for GNSS zenith tropospheric delay forecasting based on the informer model using 11-year ERA5 reanalysis data. *GPS Solut.* **2024**, *28*, 182. [\[CrossRef\]](#)
41. Xie, X.; Huang, F.; Peng, Y.; Zhou, W. A New Design of an Optimized Informer Wind Power Prediction Model Utilizing Wind Turbine Health Assessment. *J. Circuits. Syst. Comput.* **2024**, *10*, 1142. [\[CrossRef\]](#)
42. Nowotarski, J.; Liu, B.; Weron, R.; Hong, T. Improving short term load forecast accuracy via combining sister forecasts. *Energy.* **2016**, *98*, 40–49. [\[CrossRef\]](#)
43. Lu, H.; Ma, X.; Huang, K.; Azimi, M. Carbon trading volume and price forecasting in China using multiple machine learning models. *J. Cleaner. Prod.* **2020**, *249*, 119386. [\[CrossRef\]](#)
44. Jeong, S.; Park, I.; Kim, H.S.; Song, C.H.; Kim, H.K. Temperature Prediction Based on Bidirectional Long Short-Term Memory and Convolutional Neural Network Combining Observed and Numerical Forecast Data. *Sensors* **2021**, *21*, 941. [\[CrossRef\]](#)
45. Zhang, J.; Li, S. Air quality index forecast in Beijing based on CNN-LSTM multi-model. *Chemosphere* **2022**, *308*, 136180. [\[CrossRef\]](#)
46. Yan, X.; Yang, W.; Gao, M.; Ding, N.; Zhang, W.; Li, L.; Hou, Y.; Zhang, K. The WOA-CNN-LSTM-Attention Model for Predicting GNSS Water Vapor. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 1–15. [\[CrossRef\]](#)
47. Shi, J.; Wang, S.; Qu, P.; Shao, J. Time series prediction model using LSTM-Transformer neural network for mine water inflow. *Sci. Rep.* **2024**, *14*, 18284. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Hu, Y.; Fan, Z.; Liu, W.; Wu, L. An integrated navigation algorithm assisted by CNN-Informer during short-time GNSS outages. *Meas. Sci. Technol.* **2024**, *35*, 096309. [\[CrossRef\]](#)

49. Wang, K.; Zhang, J.; Li, X.; Zhang, Y. Long-term power load forecasting using lstm-Informer with ensemble learning. *Electronics*. **2023**, *12*, 2175. [[CrossRef](#)]
50. Zheng, G.-Q.; Kong, L.-R.; Su, Z.-E.; Hu, M.-S.; Wang, G.-D. Approach for Short-Term Power Load Prediction Utilizing the ICEEMDAN-LSTM-TCN-Bagging Model. *J. Electr. Eng. Technol.* **2024**. [[CrossRef](#)]
51. Ganaie, M.; Hu, M.; Malik, A.; Tanveer, M.; Suganthan, P. Ensemble deep learning: A review. *Eng. Appl. Artif. Intell.* **2022**, *115*, 0952–1976. [[CrossRef](#)]
52. Shangguan, G.; Cheng, X.; Pan, X.; Dan, M.; Wu, L.; Xie, Z. Assessments of global tropospheric delay retrieval from reanalysis based on GNSS data. *Chin. J. Geophys.* **2023**, *66*, 939–950. [[CrossRef](#)]
53. Braun, J.; Rocken, C.; Ware, R. Validation of line-of-sight water vapor measurements with GPS. *Radio Sci.* **2001**, *36*, 459–472. [[CrossRef](#)]
54. Yao, Y.; Xu, X.; Hu, Y. Precision analysis of GGOS tropospheric delay product and its application in PPP. *Acta Geod. Et Cartogr. Sin.* **2017**, *46*, 278–287. [[CrossRef](#)]
55. Yu, C.; Li, Z.; Blewitt, G. Global comparisons of ERA5 and the operational HRES Tropospheric Delay and Water Vapor products with GPS and MODIS. *Earth Space. Sci.* **2021**, *8*, 1417. [[CrossRef](#)]
56. Chen, Q.; Song, S.; Heise, S.; Liou, Y.-A.; Zhu, W.; Zhao, J. Assessment of ZTD Derived from ECMWF/NCEP Data with GPS ZTD over China. *GPS Solut.* **2011**, *15*, 415–425. [[CrossRef](#)]
57. Zhang, Y.; Cai, C. Method and Accuracy Assessment of Zenith Tropospheric Delay Derived from ERA5 Re-Analysis Data. *Geod. Geodyn.* **2020**, *40*, 62–65. [[CrossRef](#)]
58. Chen, B.; Liu, Z. A Comprehensive Evaluation and Analysis of the Performance of Multiple Tropospheric Models in China Region. *IEEE Trans. Geosci. Remote. Sens.* **2016**, *54*, 663–678. [[CrossRef](#)]
59. Zhang, C.Y.; Guo, C.X.; Chen, J.Y.; Zhang, L.; Wang, B. EGM2008 and Its Application Analysis in Chinese Mainland. *Acta Geod. Cartogr. Sin.* **2009**, *38*, 5–11. [[CrossRef](#)]
60. Zhao, Z.; Feng, J.; Hong, L. Comparison and Analysis of Tropospheric Correction Models in Satellite Positioning. *Bull. Surv. Mapp.* **2016**, *18–22*, 136. [[CrossRef](#)]
61. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural. Inf. Process Syst.* **2017**, *30*, 5998–6008. [[CrossRef](#)]
62. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.