LSTM-based short-term ionospheric TEC forecast model and position accuracy analysis

Xie Ting
Sun Yat-Sen university school of electronics and communication engineering

Zhiqiang Dai (✉ daizhiqiang@mail.sysu.edu.cn)
Sun Yat-Sen university school of electronics and communication engineering

Xiangwei Zhu
Sun Yat-Sen university school of electronics and communication engineering

Biyan Chen
Central South University

Chengxin Ran
Sun Yat-Sen university school of electronics and communication engineering

Research Article

Keywords: GNSS, ionospheric delay, SPP, LSTM forecast model, TEC, RMSE

Posted Date: July 13th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1820577/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at GPS Solutions on February 9th, 2023. See the published version at https://doi.org/10.1007/s10291-023-01406-8.
Abstract

Ionospheric delay is one of the major error sources in global navigation satellite system (GNSS) single point positioning (SPP). Some empirical models have been proposed to correct ionospheric delay, which, however, are limited by the low accuracy. The single-point time-series-based ionospheric total electron content (TEC) forecast method theoretically introduces model error and the accumulation of forecast error increases during a day. Therefore, based on the regular variation characteristics of the ionosphere, to improve the forecast accuracy of the ionosphere, we propose a long short-term memory (LSTM) short-term forecast model using discrete GNSS data and ionospheric space environment data in the same period of multiple days. The LSTM ionospheric forecast model is constructed based on the 2014 single-site GNSS data from different regions (low, mid, and high latitude regions) of the Crustal Movement Observation Network of China (CMONOC). The results are compared with the Klobuchar model, CMONOC regional ionosphere maps (RIM) data, and GNSS derived TEC measurements. The performance of each model in the SPP is also compared and analyzed to further examine the feasibility of the LSTM ionospheric forecast model. The comprehensive statistical analysis shows that: i) LSTM forecast model is consistent with GNSS-TEC observations at high, mid, and low latitudes, and the forecast error is less than 3 TECu, which is much better than that of the Klobuchar model and RIM model, and is robust to anomalous values. The mean absolute error (MAE) and root mean square error (RMSE) of the LSTM forecast model decrease with increasing latitude. ii) When being used in SPP, the LSTM forecast model brings significant improvement to position accuracy, which is overall better than the RIM and Klobuchar models. The RMSE of 3D position error is 2–4 m for the LSTM forecast model, 2–5 m for the RIM model, and 4–5 m for the Klobuchar model. iii) In terms of geographic location, for 3D position accuracy, it is highest at mid-latitudes followed by high latitudes and worst at low latitudes. For the percentage of 3D position corrections relative to the reference, it is highest at high latitudes followed by mid-latitudes, and lowest at low latitudes. The percentage of LSTM forecast model is 85%-90% at low latitudes, which is better than Klobuchar's 40%-80% and RIM's 75%-85%. Such percentage is over 90% at mid-latitudes, which is comparable to the RIM and about 50% better than the Klobuchar model. It is noteworthy that the LSTM performs even better than the reference data at high latitudes.

1 Introduction

Global Navigation Satellite System (GNSS) Single Point Positioning (SPP) technology has been widely used in many applications such as vehicle navigation. In the applications, single-frequency receivers dominate the market share because of their relative low cost. For single-frequency GNSS users, the ionospheric delay is one of the major error sources that affects position accuracy. When GNSS signals travel through the atmosphere, the error introduced by the ionosphere varies from a few meters to more than 20 meters, and can even reach more than 100 meters during periods with high solar activities (L. Chen et al., 2018; Su et al., 2019). Empirical models have been proposed to correct ionospheric delays, and these models are limited by their low accuracy. For example, the Klobuchar model corrects only 50–60% of the ionospheric delay in mid-latitudes during quiet solar activity (Pongracic et al., 2019; Tongkasem et al., 2019). In contrast, ionospheric forecast models based on GNSS measurements are more practical due to
their high accuracy, estimated RMS error for 24 h TEC forecast is 2–5 TECU under geomagnetically quiet conditions (Gulyaeva et al., 2013; Niu et al., 2014; Badeke et al., 2018). However, because the ionosphere is influenced by solar and geomagnetic activities, the spatial variability of the ionosphere is extremely complex, and small-scale irregularities or perturbations often occur in some regions, especially at some low latitudes. Therefore, accurate forecast on ionospheric delay remains a challenge.

So far, various ionospheric forecast models have been proposed and discussed, which are mainly divided into two categories. The first is empirical ionospheric models, such as the Klobuchar model (Cai et al., 2017; Pongracic et al., 2019; F. Wang et al., 2014) and the NeQuick model (N. Wang et al., 2017; Jun Chen et al., 2020). These ionospheric models have been widely used to reduce the influence of the ionosphere on single-frequency signals, contributing to an overall 50–70% reduction of ionospheric delay. The second is the ionospheric parameter reconstruction, which is a statistical model of ionospheric parameters such as total electron content (TEC). The accuracy is high in the field of short-term forecast, for example, time series forecast models such as auto-regression and moving average model (ARMA) (L. Li et al., 2013; Ansari et al., 2019; Lu et al., 2021) and spatial interpolation models like Kriging models (Srinivas et al., 2016; Ghaffari Razin & Moradi, 2021). In recent years, due to their ability to describe complex nonlinear input-output relations, neural networks have been increasingly used for the forecast of ionospheric parameters, especially in the field of ionospheric TEC forecast, mainly including radial basis function (S. Liu et al., 2020), convolutional neural networks (Ruwali et al., 2021), and long short-term memory (LSTM) networks (Kim et al., 2020, 2021; Tang et al., 2020; Wen et al., 2021; Xiong et al., 2021). Because LSTM can take full advantage of the model's input and previous hidden states to generate future TEC values, it has attracted considerable attention and achieved many research results in time series forecast. Sun et al (2017) proposed an LSTM-based model to forecast the vertical TEC (VTEC) of the ionosphere in Beijing. The input of the model is a time series consisting of a vector of daily TEC and other closely related parameters, and the output is the TEC for the next 24 hours. The results indicated that the LSTM model can obtain a more stable convergence trend and a smaller root mean square error (RMSE). Cherrier et al (2017) presented an LSTM-based TEC time series forecast model and verified that LSTM has a good potential for time series forecast. Ruwali et al (2020) introduced LSTM-CNN for ionospheric forecasting at low latitude GNSS stations using data consisting of hourly data points of VTEC time series data from the Bengaluru station. It was shown that LSTM-CNN may be well suited for estimating the ionospheric delay of GNSS signals at low latitudes. Tang (2020) compared the short-term forecast capability of LSTM, autoregressive integrated moving average and sequence-to-sequence for ionospheric TEC under different magnetic storm conditions. It was shown that LSTM can achieve the best forecast accuracy and has strong robustness for accurate trend forecast of strong magnetic storms.

However, on one hand, most of the above ionospheric LSTM forecast models are based on single-point time series to build forecast models. Since the satellite position is dynamically changing, the inversion of ionospheric TEC in the line of sight direction using GNSS pseudorange and carrier phase is usually a discrete set of data. To obtain the time series of a single point, the discrete data must be first extended to a region by a mathematical model to generate the ionospheric TEC values at any point in the region. Then, a forecasting method is used to find the statistical pattern of the ionosphere from the processed single-point
time series and to forecast the ionospheric parameters (Z. Li et al., 2019; P. Chen et al., 2020; Xiong et al., 2021; Jinpei Chen et al., 2022). Thus, such processing methods theoretically introduce more model errors and some original ionospheric information is lost in the forecast stage. Secondly, because of the continuous forecast, the error accumulation phenomenon occurs as the forecast hour increases within a day (Ruwali et al., 2021; Zewdie et al., 2021). Many studies have been done on the variation characteristics of the ionosphere, showing that the ionosphere mainly presents periodic regular variation patterns such as 11-year cycle variation, seasonal variation and diurnal variation. Thus, for a short period such as 15 days, the diurnal variation acts as the primary pattern, meaning that the ionospheric parameter values are stable in the same time epoch for many days under calm space environment. In this paper, based on the regular variation characteristics of the ionosphere, multiple observation data are selected and are split into multi-day data of different time periods. The ionospheric discrete data of fixed time periods of multiple days are used to construct the LSTM forecast model. In addition, the ionosphere is sensitive to the variation of solar activity and near-Earth space. The ionosphere is influenced by various factors, such as geographic location, solar activity level and geomagnetic activity conditions (Kaselimi et al., 2020), these factors should be considered in LSTM modeling. In this paper, a single-station regional ionospheric model is constructed using an LSTM model suitable for multi-feature learning. The TEC, solar activity index and geomagnetic activity index data are used as model inputs to forecast the ionospheric TEC values for the next 24 h. Finally, the performance of the LSTM forecast model is examined in the SPP solution together with Klobuchar model, RIM model and reference ionospheric data. The TEC in the latter refers to VTEC.

The paper is organized as follows. In Section 2, the LSTM forecast model, the influences of ionospheric delay on SPP, and the accuracy evaluation are presented. Then, section 3 describes the data used in this paper and the data processing strategy. Next, the ionospheric delay results obtained from different models and the performance in the SPP solutions are presented in Section 4. The results are briefly discussed in Section 5, and the conclusions are summarized in Section 6.

2. Model And Methodology

This section describes the LSTM prediction model, the influences of ionospheric delay on SPP, and the accuracy evaluation method.

2.1 LSTM forecast model

LSTM is a special type of recurrent neural network that has feedback connections, unlike the standard feedforward neural networks. The recurrent neural network not only can process single data points but also entire data sequences (Hochreiter & Schmidhuber, 1997). The LSTM model has various forms for different types of data inputs, and in this paper, we choose a multi-input model, i.e., multiple sequence inputs and single sequence outputs. The framework and internal structure of the LSTM forecast model are shown in Fig. 1. LSTM works on the concept of gates, where information flows through a mechanism called cell states, which consists of 3 main gates, the forget gate, input gate and output gate. Thereby, LSTM can selectively remember and forget something. In LSTM cells, two main states, the cell state \( C_t \)
and the hidden state \((H^t)\), are transferred to each cell. The information flow process of LSTM is calculated as follows:

\[
\begin{align*}
    & f^t = \sigma(W_f X^t + U_f H^{t-1} + b_f) \\
    & i^t = \sigma(W_i X^t + U_i H^{t-1} + b_i) \\
    & \tilde{C}^t = \tanh(W_c X^t + U_c H^{t-1} + b_c) \\
    & C^t = f^t \odot C^{t-1} + i^t \odot \tilde{C}^t \\
    & o^t = \sigma(W_o X^t + U_o H^{t-1} + b_o) \\
    & H^t = o^t \odot \tanh(C^t)
\end{align*}
\]

where \(X^t, H^t, \tilde{C}^t\) and \(C^t\) are the input, hidden state, cell input activation, and cell state vectors to the LSTM unit at time \(t\), respectively; \(f^t, i^t, \) and \(o^t\) are the activation vectors of the forget, input/update and output gates, respectively; \(W, U, \) and \(b\) are weight matrices and bias vector parameters which need to be learned during training; the symbol \(\odot\) represents the Hadamard product operator; \(\sigma\) is the Sigmoid activation function, and \(\tanh\) denotes the tanh activation function.

The parameters of the LSTM forecast model are shown in Table 1. The input to the model is continuous 30-minute single-station ionospheric data, which are entered through the input layer, including TEC observations and 6 training features of latitude, longitude, Ap, Dst, \(F_{10.7}\), and SSN (Xie et al., 2022). The detailed processing of the input data is described in Section 3. The model consists of a three-layer LSTM with the number of hidden units of 31 in each layer, where the network optimizer is chosen as Adam and the loss function is MSE. In addition, the validation set loss value (val_loss) is monitored by the early stopping method, where patience is set to 3, and the fully connected layer (Dense) uses the LeakyReLU activation function to avoid the dead neuron problem.
### 2.2 Influences of ionospheric delays on single point positioning

Compared with the precise point positioning (PPP) technique, the SPP technique does not need to solve the ambiguity of whole cycles, and has the advantages of easy and fast calculation, which is widely used in the fields of navigation of vehicles, ships, aircraft, and geological exploration and resource exploration (Krasuski et al., 2020). The pseudo-range is generally expressed as:

\[
P = \sqrt{(x^s - x)^2 + (y^s - y)^2 + (z^s - z)^2 + cdt^s - cdt_r + I + T}
\]

in which, \(P\) is the pseudo-range observation; \(x^s, y^s, z^s\) is the satellite orbit coordinate at the time of signal transmission; \(x, y, z\) is the station coordinate at the time of signal reception; \(cdt^s\) is the satellite clock difference; \(cdt_r\) is the receiver clock difference; \(I\) is the ionospheric delay, \(I = \frac{40.28 \times 10^{16}}{f^2} \times VTEC \times M_F\), where \(M_F\) is the projection function; \(T\) is the tropospheric error.

In SPP, the tropospheric delay is usually corrected by empirical models such as the Saastamoinen model, and ionospheric delay is corrected by different models. To evaluate the effect of ionospheric delay on GNSS position performance, ionospheric delay corrections using Klobuchar, RIM and LSTM are implemented and compared with the position results of reference ionospheric data. When performing the
position accuracy analysis, the reference coordinates of each station are determined by the static PPP solution of the last epoch. When performing the pseudo-range SPP solution, the data of satellites with elevation angle below 15° and signal-to-noise ratio below 25 dBHz are excluded, and the specific solution strategy is shown in Table 2.

<table>
<thead>
<tr>
<th>Options</th>
<th>Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constellation</td>
<td>GPS</td>
</tr>
<tr>
<td>Positioning mode</td>
<td>SPP</td>
</tr>
<tr>
<td>Estimator</td>
<td>Least-squares</td>
</tr>
<tr>
<td>Frequencies</td>
<td>L1</td>
</tr>
<tr>
<td>Sampling rate</td>
<td>30s</td>
</tr>
<tr>
<td>Elevation mask</td>
<td>15</td>
</tr>
<tr>
<td>Tropospheric delay correction</td>
<td>Saastamoinen model</td>
</tr>
<tr>
<td>Ionospheric delay correction</td>
<td>Klobuchar/RIM/LSTM/Reference</td>
</tr>
<tr>
<td>Satellite orbit, clock and timing group delay (TGD)</td>
<td>Broadcast ephemeris</td>
</tr>
<tr>
<td>Station reference coordinates</td>
<td>The end position of the static PPP solution</td>
</tr>
</tbody>
</table>

### 2.3 Accuracy evaluation

Forecast results from three models are compared with the reference values of the station. The mean absolute error (MAE) and RMSE are used to evaluate the performance of the model, which are calculated as shown in Eqs. 3 and 4. The MAE and RMSE can indicate the dispersion of the error between the forecast and reference values, and a lower value means a better fit between the forecast and reference values.

\[
\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |(\text{Reference}_i - \text{Forecast}_i)|
\]

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\text{Reference}_i - \text{Forecast}_i)^2}
\]
Taking the regular variation characteristics of the ionosphere into account, the daily measurements are split into 48 datasets with an interval of 30 minutes. The discrete ionospheric data with fixed time intervals of multiple days are used to construct the LSTM model. In addition, the ionosphere is very sensitive to solar and geomagnetic activities, and the TEC depends on various factors such as geographic location, solar activity level, and geomagnetic activity status, such factors will be considered as inputs to the LSTM forecast model. This section summarizes the ionosphere-related data used in the experiment and presents the data processing strategy.

3.1 Retrieval of VTEC

To better represent the forecast result of the model, this paper selects the dataset from 1–17 January 2014, the period in which both geomagnetic activity and solar activity are active (as shown in Fig. 2).

3.1.1 VTEC estimated from GNSS measurements

The experiments use the ground-based GNSS measurements from the Crustal Movement Observation Network of China (CMONOC). Using the GNSS dual-frequency data, the VTEC at different epochs of the selected stations is accurately extracted by the carrier-phase smoothing pseudo-range method, which is used as input for subsequent LSTM models and reference values for evaluation, and the calculation expressions are shown in Eqs. 5. The time interval of VTEC data is 30 s, the satellite cut-off elevation angle is 15°, and the thin layer height of the ionosphere is 350 km.

\[
\text{VTEC} = -\frac{\cos z'}{40.28} \frac{f_1^2 f_2^2}{f_1^2 - f_2^2} \left[ \bar{P}_4(t) + L_4(t) - \bar{L}_4(t) - (\text{DCB}^S + \text{DCB}^r) \right]
\]

where \(P_4 = P_2 - P_1\), \(L_4 = L_2 - L_1\), \(P_i(t)\) is the pseudorange at time \(t\) (\(i\) takes 1 and 2); \(L_i\) is the phase corresponding to this epoch; \(\bar{L}_i\) is the average values of the carrier phase; \(\bar{P}_4\) is the average values of the pseudorange, \(\text{DCB}^S\) and \(\text{DCB}^r\) are the differential code deviations of the satellite and receiver, respectively. \(\text{DCB}^S\) can be corrected directly using the P1-C1 and P1-P2 files of the CODE Center.

3.1.2 VTEC calculated by the Klobuchar model

In this study, the VTEC at the ionospheric pierce point (IPP) is calculated using the Klobuchar model. The GPS Klobuchar model is based on the geomagnetic coordinate system, and the 8 parameters of the Klobuchar model can be used to calculate the VTEC values with the geomagnetic latitude of the IPP. The 8 parameters are obtained by using the global GNSS dual-frequency measurements, which are updated daily and broadcast to the user via a broadcast ephemeris. The advantage of the Klobuchar model is that the model calculates and applies the ionospheric delay value in an easy-to-use way, based on the fact that the ionospheric delay error is inversely proportional to the square of the frequency intensity. The value depends on the frequency of the GNSS satellite, as shown in Eqs. 6.
where,

\[
VTEC_{\text{klobuchar}} = \frac{c \times f^2}{40.28 \times 10^{16}} \left[ 5 \times 10^{-9} + A \cdot \cos \left( \frac{2\pi}{P} (t - 50400) \right) \right]
\]

6

where

\[
A = \begin{cases} 
\sum_{n=0}^{3} \alpha_n |\phi_M/\pi|^n, & A \geq 0 \\
0, & A < 0 
\end{cases}
\]

7

\[
P = \begin{cases} 
172800, & P \geq 172800 \\
\sum_{n=0}^{3} \beta_n |\phi_M/\pi|^n, & 172800 > P \geq 72000 \\
72000, & P < 72000 
\end{cases}
\]

8

where, \(VTEC_{\text{klobuchar}}\) is the VTEC value calculated by the Klobuchar model, \(t\) is the local time in seconds at the IPP; \(c\) is the speed of light; \(f\) is the frequency; \(A\) is the amplitude of the cosine function during the day; \(P\) is the period of the cosine function; \(\alpha_n\) and \(\beta_n\) are the broadcast ionospheric parameters given by the navigation message; \(\phi_M\) is the geomagnetic latitude of the IPP.

### 3.1.3 VTEC calculated from RIM data

In addition, global/regional ionospheric maps constructed from GNSS measurements using mathematical models are also popular (Alizadeh et al., 2011; Krypiak-Gregorczyk et al., 2017; Krypiak-Gregorczyk & Wielgosz, 2018). The International GNSS Service (IGS) regularly provides global ionospheric maps (GIM) in ionospheric exchange (IONEX) format. Due to the sparse distribution of IGS sites in China, the accuracy and resolution of the GIM data inverted from IGS sites are limited in the Chinese region (Luo et al., 2014; Lai et al., 2021). Therefore, the regional ionosphere maps (RIM) data (ftp.cgps.ac.cn/products/ionosphere/data) released by CMONOC are directly used to compare with the LSTM model forecast. The temporal resolution of the RIM data is 2 h. The spatial coverage is from latitude 15.0°N to 55.0°N, longitude 70.0°E to 140.0°E, and the spatial resolution is 1°×1°. Using the RIM model, for single-frequency users, the VTEC values at the IPP can be calculated from the four nearest grid values using a bilinear interpolation method.

### 3.2 Space environment data

Space environment data are also applied in the LSTM model, and the Pearson correlation coefficient method and the Fréchet distance method (Technische Universität Wien et al., 1994) are used to evaluate the data correlation before data selection to ensure a weak correlation between the data (Xie et al., 2022).
The Pearson correlation coefficient method is used when the temporal resolution of the two datasets is the same, otherwise, the Fréchet distance method is used. A correlation coefficient greater than 0.5 or a Fréchet coefficient value less than 0.5 indicates a strong correlation between the data, and only one of the two datasets is selected. After data screening, the sunspot number index (SSN) and 10.7 cm radio flux ($F_{10.7}$) index, representing solar activity, and the geomagnetic activity level index (Ap) and the magnetic storm loop current index (Dst), featuring the Earth's geomagnetic activity, are selected. The specific information is shown in Table 3.

### Table 3
Information on space environment data products

<table>
<thead>
<tr>
<th>Index</th>
<th>Data Sources</th>
<th>Time resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSN/$F_{10.7}$</td>
<td>Prepared by the US Dept. of Commerce, NOAA, Space Weather Prediction Center (Daily Space Weather Indices Product: <a href="http://www.swpc.noaa.gov/wwire.html/">http://www.swpc.noaa.gov/wwire.html/</a>)</td>
<td>1 day</td>
</tr>
<tr>
<td>Ap</td>
<td>NASA/Goddard Space Flight Center (OMNI WEB data: <a href="https://omniweb.gsfc.nasa.gov/form/dx1.html">https://omniweb.gsfc.nasa.gov/form/dx1.html</a>)</td>
<td>3 hours</td>
</tr>
<tr>
<td>Dst</td>
<td>NASA/Goddard Space Flight Center (OMNI WEB data: <a href="https://omniweb.gsfc.nasa.gov/form/dx1.html">https://omniweb.gsfc.nasa.gov/form/dx1.html</a>)</td>
<td>1 hour</td>
</tr>
</tbody>
</table>

The geomagnetic activity index and solar activity index during the study period are shown in Fig. 2. It can be seen that the solar activity was high during this period, reaching high levels on 4 days and moderate levels on 6 days. The monthly average number of SSN in January was 126 and the monthly average $F_{10.7}$ was 159 sfu. The geomagnetic field reached small magnetic storm levels for 3 hours and active levels for 6 hours on January 2. The geomagnetic field reached active levels for 3 hours each on January 12 and 14.

### 3.3 Data processing

For the LSTM model, the data need to be processed in advance. First, the space environment parameters were interpolated using the fast Fourier interpolation technique to match the time resolution of 30 minutes. Integrating the above data, the input datasets are Ap, Dst, $F_{10.7}$, SSN, longitude, and latitude, and the data sets are divided into 30-minute intervals. A total of 48 datasets were used for one day, and each dataset was divided into 3 groups to experiment, where the training set: validation set: forecast set $\approx 15:1:1$. The data were scaled in the range of 0 to 1 using the maximum-minimum normalization method to reduce the effect of the large dynamic range.

In this paper, single-station GNSS data of different latitudes are used for experiments, and the brief forecast process of the LSTM forecast model is as follows:

1. Using the Pearson and Fréchet method to select the weakly correlated geomagnetic and solar activity datasets, and normalizing the data to remove the influence of the magnitude between the data;
2. Pre-processing the single-station GNSS data, i.e., calculating the TEC value and the longitude and latitude of the IPP, and selecting the ionospheric data points with altitude angles greater than 15°;
3. Integrating the above data, the input data sets are Ap, Dst, F₁₀.₇, SSN, longitude and latitude. The datasets were divided into 30-minute intervals for a total of 48 datasets, and each dataset was divided into 3 groups (training set, validation set, and forecast set) for the experiments.
4. Establishing adaptive hyperparameters, and using the smallest validation set loss (var_loss) as the basis for early stopping. Among them, the specific hyperparameter settings are shown in Table 1;
5. Calculating forecast accuracy and comparing the results with RIM data and Klobuchar data.

4. Test And Results

This section presents the ionospheric modeling characteristics of the Klobuchar, RIM and LSTM forecast models, and their performance in the SPP solution.

To evaluate the performance of the above LSTM forecast model in the SPP solution, GNSS measurements collected from CMONOC stations from 1 to 17 January, 2014 were used for statistical analysis. The ionosphere shows complex spatial variations with latitude and longitude. To verify the adaptability and forecast accuracy of the model at different spatial locations, GPS stations were divided into three regions: low (0°~30°), middle (30°~45°), and high latitude (45°~60°), with 12 stations selected in total. The GNSS station locations and detailed information are shown in Fig. 3 and Table 4, including four stations each at low latitude, mid-latitude and high latitude.

4.1 Performance of the LSTM forecast model

First, 48 LSTM forecast models are built for each site (saving the optimal weights) using the split dataset as the parameter input. The ionospheric TEC forecast models are trained to obtain the ionospheric TEC forecast values for different periods, and the results are evaluated using the above criterion. The results of the LSTM forecast model are shown in Table 4.
Table 4 shows the MAE and RMSE results for the LSTM forecast model for 12 different stations. The forecast MAE values range from 0.8 to 3 TECu, with the HISY station showing lower accuracy, while the HLMH station achieved the highest accuracy among the other stations. The forecast MAE and RMSE of the LSTM forecast model are 0.8 to 3 TECu, and the forecast accuracy of the mid-latitude and high latitude stations can be within 1 TECu. Modeling and forecast accuracy is positively correlated with latitude.

For a more detailed representation of the forecast accuracy at each 30-minute interval, the error bars for each period are plotted by combining the forecast MAE and RMSE values, as shown in Fig. 4. It can be seen that the range of error bars decreases with increasing latitude, especially in the period 04:00 UTC-10:00 UTC (12:00 LT-18:00 LT). It can be found that the LSTM forecast model is different from the temporal LSTM-CNN model (Ruwali et al., 2021), in that the RMSE increases as the number of forecasted hours increases. Because it is modeled separately in periods, the RMSE values are independent in different periods and are only related to the ionospheric characteristics in that period, and the errors do not accumulate. As can be seen in Fig. 4 (a), the errors of the low latitude stations are within 5 TECu for most of the periods. The forecast errors are larger in the 06:00 UTC-10:00 UTC (14:00 LT-18:00 LT) period, especially for the KMIN station, and the reason for this can be found in Fig. 5 (c) later, i.e., the presence of anomalous discrete values with large deviations of up to 30 TECu, a phenomenon that may be related to the low latitude equatorial ionospheric anomaly (EIA) (Song et al., 2018). From Fig. 4 (b), it is found that
the GSDX station has a large error in the period 04:00 UTC-6:00 UTC (12:00 LT-14:00 LT), with a maximum of more than 5 TECu, which is also found in Fig. 5 (f) because of the presence of continuous anomalous discrete values.

Figure 5 shows the forecast results for 12 stations at low, mid and high latitudes. Among them, the GNSS measurements are selected to be within 3° difference in longitude and within 1° difference in latitude from the station. As can be seen from Fig. 5, i) the TEC tends to decrease with increasing latitude. Among them, the forecast values of the LSTM forecast model is in the best agreement with the GNSS measurements, which can better forecast the ionospheric TEC at low latitudes; ii) the RIM data are generally in good agreement with the GNSS measurements. However, the TEC values are overestimated in the period of 04:00 UTC-08:00 UTC, the RIM values are closer to the GNSS measurements than the Klobuchar model values; iii) the Klobuchar model only captures the temporal trend of the ionosphere, which is somewhat underestimated at low latitudes and overestimated at mid-latitudes and high latitudes. In the next subsection, the performance of the LSTM forecast model, the Klobuchar model and the RIM model in the SPP solution is comprehensively evaluated and the spatial and temporal characteristics are analyzed.

4.2 Performance in the SPP solution

The ionospheric delay is considered as one of the largest error sources in single-frequency position, and its model accuracy can be indirectly reflected in the accuracy of the SPP solution. The experimental platform used is based on the GiNav program developed by the NASG Key Laboratory of Land Environment and Disaster Monitoring, China University of Mining and Technology (K. Chen et al., 2021). The data required for the experiments are downloaded from the official FTP of the IGS Analysis Center CDDIS (crustal dynamics data information system) (https://cddis.nasa.gov/). The data from CMONOC stations distributed in high, mid and low latitudes are processed separately. Using the position results of ionospheric reference data as a comparison, this paper compares the performance of ionospheric correction by the LSTM forecast model, Klobuchar model and RIM model in SPP solution, and evaluates the feasibility of the LSTM forecast model. For the position accuracy analysis, the SPP error is obtained by subtracting the exact coordinates of each site from the SPP solution. The reference position is determined as the static PPP solution of the last epoch.

Figure 6 shows SPP errors in the east (E), north (N) and up (U) directions generated by the three models and the reference values at the low latitude stations. On the whole, the positioning error of the three ionospheric models in the E direction is the smallest, within ±3 m, the U direction is the largest, within ±20 m. The SPP errors occur mainly in the vertical (U) direction, and the position accuracy is in the meter level, while in the horizontal (N and E) direction, it can reach the sub-meter level. In terms of individual models, the SPP result with LSTM forecast model correction has the best accuracy, especially in the U direction. The difference between the LSTM forecast model position accuracy and the reference data in the E direction is not much, around 0.1 m, and in the N direction, the difference is around 0.2-0.3m, and in the U direction, except for LALB, the difference is 0.516 m. The difference in the U direction is about 0.2m except for LALB, so the overall position performance of the LSTM forecast model is better in the low latitude region. The SPP error using the Klobuchar model is the largest, with RMSE of about 1 m, 1-4 m, and 3-5
m in the E, N, and U directions, respectively. In terms of stations, the LSTM forecast model and the RIM model have better positioning accuracy at KMIN and CQCS stations relative to HISY and LALB stations, especially in the N direction and during 12:00 to 20:00 local time (LT) (corresponding to 4:00 to 14:00 UTC). One reason is that the parameters of GNSS Klobuchar are calculated based on an empirical model, while RIM is generated by processing GNSS measurements from the regional monitoring station network, which has higher accuracy. Another reason is that the RIM is updated every 1 hour, which is more frequent than GPS Klobuchar, and this will help to describe the ionospheric characteristics accurately. The RIM model is generally corrected at the HISY and LALB stations, and even the results are comparable to the Klobuchar model at the HISY station. Since the latitude coverage of RIM data is from 15.0°N to 55.0°N, while the latitude of the HISY station is 18.236°N, the interpolation accuracy of the IPP is limited.

Figure 7 shows the SPP error at the mid-latitude stations. It can be seen that the SPP result in mid-latitudes is similar to that in low latitudes, with the smallest error fluctuations in the E direction and the largest in the U direction. In the E and N directions, the position results based on the RIM model are better than the LSTM forecast model, which is because many CMONOC stations are set up in the mid-latitude region. In the U direction, the LSTM model is better than the RIM, and in the GSDX and XJBC, it is slightly stronger than the reference data, which is because the GNSS measurements have some anomalies as can be seen in Fig. 5 (f). Overall, the position accuracy of the LSTM forecast model and the reference data are comparable. In the E and U directions, the difference between the LSTM forecast model and the reference is not much, about 0.1 m, and in the N direction, about 0.1–0.2 m. The position accuracy of the Klobuchar model in the E and N directions is slightly lower than that of the RIM and LSTM forecast models, which is consistent with the conclusion of the paper that the GPS Klobuchar model is more suitable for mid-latitude regions (Cai et al., 2017).

Figure 8 shows SPP errors at the high latitude stations. It can be seen that the LSTM forecast model has the best position accuracy, even slightly better than the reference data, because the LSTM forecast model results are smoother compared with the GNSS measurement data, which are susceptible to some anomalies thus affecting the position accuracy. In the E and N directions, the RIM position accuracy is poor, especially for HLFY and HLMH, for reasons consistent with the HISY stations at low latitudes, limited by the spatial coverage of the RIM while the Klobuchar model has better position accuracy, but decreases compared to mid-latitudes.

To more visually assess the SPP position accuracy when using different models at different stations, the RMSE of the positioning for each model at 12 stations is shown in Fig. 9. At low latitudes, the RMSE of the Klobuchar model SPP error is the largest, especially in the N direction, which is 2–3 times higher than that of the LSTM forecast model. In the mid-latitude region, the difference in position accuracy of each model is relatively reduced in the E and N directions, and the RMSE is all around 1m, but in the U direction, the Klobuchar model error is twice as high as the other models. In the high latitude region, the LSTM forecast model has much better positioning accuracy than Klobuchar and RIM, and is comparable to the accuracy of the reference data. Overall, the LSTM forecast model has the best positioning accuracy, the RIM is the second best, and the Klobuchar model is the worst. However, for mid and high latitudes, the horizontal
accuracy of the Klobuchar model can reach the sub-meter level. If the user only requires meter level horizontal positioning accuracy and does not care about vertical positioning accuracy, the Klobuchar model can achieve better accuracy.

Table 5
Statistics of 3D position errors of ionospheric models

<table>
<thead>
<tr>
<th>Classification</th>
<th>Station name</th>
<th>3D(m)</th>
<th>Percentage of corrections relative to Reference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reference</td>
<td>Klobuchar</td>
<td>RIM</td>
</tr>
<tr>
<td>Low latitude</td>
<td>HISY</td>
<td>3.02</td>
<td>3.93</td>
</tr>
<tr>
<td></td>
<td>LALB</td>
<td>3.34</td>
<td>4.98</td>
</tr>
<tr>
<td></td>
<td>KMIN</td>
<td>3.23</td>
<td>5.88</td>
</tr>
<tr>
<td></td>
<td>CQCS</td>
<td>1.97</td>
<td>4.56</td>
</tr>
<tr>
<td>Middle latitude</td>
<td>AHBB</td>
<td>1.95</td>
<td>4.58</td>
</tr>
<tr>
<td></td>
<td>GSDX</td>
<td>1.99</td>
<td>4.48</td>
</tr>
<tr>
<td></td>
<td>XJBC</td>
<td>2.85</td>
<td>4.59</td>
</tr>
<tr>
<td></td>
<td>BJGB</td>
<td>1.82</td>
<td>4.45</td>
</tr>
<tr>
<td>High latitude</td>
<td>XJFY</td>
<td>2.91</td>
<td>4.63</td>
</tr>
<tr>
<td></td>
<td>HLFY</td>
<td>2.22</td>
<td>4.46</td>
</tr>
<tr>
<td></td>
<td>NMER</td>
<td>2.42</td>
<td>4.57</td>
</tr>
<tr>
<td></td>
<td>HLMH</td>
<td>2.48</td>
<td>4.59</td>
</tr>
</tbody>
</table>

Note: Percentage of corrections relative to Reference (%) = (RMS_{method})/RMS_{Reference}×100%, RMS_{method} indicates the positioning error RMSE under the adopted ionospheric model, RMS_{Reference} indicates the positioning error RMS under the reference datum data.

To more accurately assess the SPP positioning accuracy when using different models at different stations, the three-dimensional (3D) position errors of different ionospheric models and the percentage of corrections relative to reference are further summarized (Table 5). It is obvious from Table 5 that the 3D position accuracy of the LSTM forecast model in high, mid, and low latitudes is substantially improved, with its 3D error range of 2–4 m, 2–5 m for RIM, and 4–5 m for Klobuchar overall. At mid-latitudes, the correction percentage of the LSTM forecast model relative to the reference data is even more than 90%, which is comparable to the RIM data, while Klobuchar is only about 40%, an improvement of about 50%. It is known from the previous paper (Bi et al., 2017) that the Klobuchar model is too poor in position accuracy mainly in the U direction. Due to the limitation of RIM data coverage, the correction percentage of RIM is only about 60% for high latitude stations except NMER. In terms of 3D position accuracy, the highest position accuracy is achieved at mid-latitude, followed by high latitude and the worst at low
latitude. In terms of the percentage of corrections relative to the reference, it is highest in high latitudes, second highest in mid-latitudes, and lowest in low latitudes.

5. Discussion

So far, SPP is still the most popular positioning method for navigation applications, in which the simple Klobuchar model is used more often. In addition, GIM and RIM are recognized as ionospheric products with higher accuracy (Ren et al., 2019). Therefore, the LSTM forecast model proposed in this paper not only compares with the reference ionospheric data, but also compares the performance of the Klobuchar model and RIM model in the SPP solution, to further evaluate the feasibility of the LSTM forecast ionospheric model. The following recommendations can be provided as a reference through the comprehensive study and further analysis of this paper.

1. In middle and high latitudes, if users only require meter-level horizontal positioning accuracy and do not care about vertical positioning accuracy, the simplest Klobuchar model is preferred. Due to the restricted RIM coverage, the performance of the RIM model in high latitudes is limited, i.e., the application of grid data products is limited.

2. For users near low latitude areas, the Klobuchar model is not the best choice. This is because the RMSE of the Klobuchar model SPP error is at least twice that of the LSTM forecast model and the RIM model in the N direction, with an RMSE of more than 1.7 m. Moreover, the position accuracy in the N direction is extremely unstable, with an error of more than 5 m during 12:00 to 20:00 local time (LT) (corresponding to 4:00 to 14:00 UTC), while the LSTM forecast model is more stable, at around 0.8 m.

3. If vertical position accuracy is considered, the Klobuchar model is less suitable, especially between 12:00 and 20:00 local time (LT) (equivalent to 4:00 to 14:00 UTC), where the error of the Klobuchar model is more than 5 m and can exceed 10 m, and the error of the LSTM forecast model and RIM is around 3 meters.

4. The RIM and Klobuchar models used are post-processed data with higher accuracy than the real-time model, while the LSTM forecast models are forecast models that can reach a level comparable to that of the RIM, which further validates the feasibility of the LSTM forecast model. This study discusses the feasibility of the LSTM forecast model based on single station data, which can be extended to regional or global.

6. Conclusions

This paper shows the proposed LSTM forecast model and presents a more comprehensive statistical analysis of the LSTM forecast model, the Klobuchar model and the RIM model in terms of both modeling accuracy and positioning performance. The data collected at 12 stations from January 1–17, 2014 were processed using different ionospheric schemes. The results are as follows:

1. The forecast results of the LSTM forecast model are consistent with the GNSS measurements at high, middle and low latitudes, and the forecast error is less than 3 TECu. The forecast accuracy is much higher than that of the Klobuchar and RIM models, and is less susceptible to anomalies.
Geographically, the forecast MAE and RMSE of the LSTM forecast model decrease with the increase of latitude.

2. In terms of SPP position accuracy, the ionospheric correction scheme based on the LSTM forecast model is generally better than that using the RIM and Klobuchar models. Compared with the Klobuchar model, the SPP position accuracy in the N, E, U, and 3D directions of the LSTM forecast model and RIM model at high, mid, and low latitudes are greatly improved, with the 3D error range of 2–4 m for the LSTM forecast model, 2–5 m for the RIM, and 4–5 m for the Klobuchar overall.

3. In terms of geographical regions, for 3D position accuracy, the highest position accuracy is achieved at mid-latitude, followed by high latitude and the worst at low latitude. For the percentage of corrections relative to the reference, it is highest in high latitudes, second highest in mid-latitudes, and lowest in low latitudes. For the percentage of 3D position correction relative to the reference, at low latitudes, the LSTM forecast model is 85%-90%, which is much better than Klobuchar's 40%-80% and better than RIM's 75%-85%, with at least about 8% improvement over RIM. At mid-latitudes, the LSTM forecast model is over 90%, which is comparable to the RIM, about 50% better than the Klobuchar model, and even better than the reference data at high latitudes.

Declarations

Data availability

The GNSS datasets can be provided to readers by contacting the corresponding author on reasonable request. The CMONOC RIM can be accessed from the website (ftp.cgps.ac.cn/products/ionosphere/data). Kp, Dst indexes were downloaded from the website (http://isgi.unistra.fr/data_download.php/), and F_{10.7}, SSN were downloaded from the (http://www.swpc.noaa.gov/wwire.html/). The authors would like to thank these organizations for making their data public.

Acknowledgments

This work was supported by the Science and Technology Planning Project of Guangdong Province of China (Grant No.2021A0505030030), Shenzhen Science and Technology Program (Grant No.GXWD20201231165807008, 20200830225317001).

Author information

Sun Yat-Sen university school of electronics and communication engineering, Shenzhen, China

Ting Xie, Zhiqiang Dai, Xiangwei Zhu, Chengxin Ran

School of Geosciences and Info-Physics, Central South University, Changsha, China.

Biyan Chen
Corresponding author: Zhiqiang Dai (daizhiqiang@mail.sysu.edu.cn) and Xiangwei Zhu (zhuxw666@mail.sysu.edu.cn)

Conflict of interest

The authors declare no conflict of interest.

Author contributions

Zhiqiang Dai gave the basic idea of the paper and revised the paper. Xiangwei Zhu, Biyan Chen, and Chengxin Ran participated in the discussion and interpretation of the data and the revision of the paper. Ting Xie performed the experiment and wrote the paper. All authors have read and agree to the published version of the manuscript.

References


Figures
Figure 1

The framework and internal structure of the LSTM forecast model
Figure 2

Space environment data in the selected period. In Figure (b), the green bar represents Kp and the blue represents Ap.

Figure 3

Spatial distribution of stations.
Figure 4

The forecast error bar of LSTM forecast model; (a) low latitude station; (b) mid-latitude station; (c) high latitude station

Figure 5
TEC values of the LSTM forecast model (green solid line) for low latitude (a-d: HISY, LALB, KMIN, CQCS), mid-latitude (e-h: AHBB, GSDX, XJBC, BJGB) and high latitude stations (i-l: XJFY, HLFY, NMER, HLMH) and Klobuchar (yellow solid line), RIM (purple solid line) data, and GNSS measurements (red scattered points) are compared.

Figure 6

The SPP position errors of ionospheric models for low latitudes (a) HISY, (b) LALB, (c) KMIN, and (d) CQCS (LSTM forecast model (green scatter) and Klobuchar (yellow scatter), RIM (purple scatter) data and GNSS reference data (red scatter))
Figure 7

The SPP positioning errors of ionospheric models for mid-latitude (a) AHBB, (b) GSDX, (c) XJBC, (d) BJGB (LSTM forecast model (green scatter) and Klobuchar (yellow scatter), RIM (purple scatter) data and GPS reference data (red scatter))
Figure 8

The SPP positioning errors of ionospheric models for high latitudes (a) XJFY, (b) HLFY, (c) NMER, and (d) HLMH (LSTM forecast model (green scatter) and Klobuchar (yellow scatter), RIM (purple scatter) data, and GNSS reference data (red scatter))
Figure 9

The SPP position accuracy for different stations using different models, the RMSE of positioning error for each model for 12 stations in high, mid and low latitudes