

## Article

# Ionospheric TEC Prediction in China Based on the Multiple-Attention LSTM Model

Haijun Liu <sup>1</sup>, Dongxing Lei <sup>1</sup>, Jing Yuan <sup>2,\*</sup>, Guoming Yuan <sup>1,\*</sup>, Chunjie Cui <sup>3</sup>, Yali Wang <sup>2</sup> and Wei Xue <sup>2</sup>

<sup>1</sup> Institute of Disaster Prevention, Institute of Intelligent Emergency Information Processing, Langfang 065201, China

<sup>2</sup> School of Information Engineering, Institute of Disaster Prevention, Langfang 065201, China

<sup>3</sup> Beijing Jingwei Textile Machinery New Technology Co., Ltd., Beijing 100176, China

\* Correspondence: yuanjing@cidp.edu.cn (J.Y.); yuanguoming@cidp.edu.cn (G.Y.); Tel.: +86-18612068986 (J.Y.); +86-17301397759 (G.Y.)

**Abstract:** The prediction of the total electron content (TEC) in the ionosphere is of great significance for satellite communication, navigation and positioning. This paper presents a multiple-attention mechanism-based LSTM (multiple-attention Long Short-Term Memory, MA-LSTM) TEC prediction model. The main achievements of this paper are as follows: (1) adding an L1 constraint to the LSTM-based TEC prediction model—an L1 constraint prevents excessive attention to the input sequence during modelling and prevents overfitting; (2) adding multiple-attention mechanism modules to the TEC prediction model. By adding three parallel attention modules, respectively, we calculated the attention value of the output vector from the LSTM layer, and calculated its attention distribution through the softmax function. Then, the vector output by each LSTM layer was weighted and summed with the corresponding attention distribution so as to highlight and focus on important features. To verify our model's performance, eight regions located in China were selected in the European Orbit Determination Center (CODE) TEC grid dataset. In these selected areas, comparative experiments were carried out with LSTM, GRU and Att-BiGRU. The results show that our proposed MA-LSTM model is obviously superior to the comparison models. This paper also discusses the prediction effect of the model in different months. The results show that the prediction effect of the model is best in July, August and September, with the R-square reaching above 0.99. In March, April and May, the R-square is slightly low, but even at the worst time, the fitting degree between the predicted value and the real value still reaches 0.965. We also discussed the influence of a magnetic quiet period and a magnetic storm period on the prediction performance. The results show that in the magnetic quiet period, our model fit very well. In the magnetic storm period, the R-square is lower than that of the magnetic quiet period, but it can also reach 0.989. The research in this paper provides a reliable method for the short-term prediction of ionospheric TEC.

**Keywords:** attention mechanism; multiple-attention LSTM (MA-LSTM); ionosphere; total electron content



**Citation:** Liu, H.; Lei, D.; Yuan, J.; Yuan, G.; Cui, C.; Wang, Y.; Xue, W. Ionospheric TEC Prediction in China Based on the Multiple-Attention LSTM Model. *Atmosphere* **2022**, *13*, 1939. <https://doi.org/10.3390/atmos13111939>

Academic Editor: Yuichi Otsuka

Received: 28 August 2022

Accepted: 15 November 2022

Published: 21 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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/>).

## 1. Introduction

The ionosphere is an important ionization region of the earth's upper atmosphere, with an altitude of approximately 60–1000 km. The ionosphere is affected by solar storms, geomagnetic activities and other factors, which will generate ionospheric disturbances and have an important impact on radio and other communication signals transmitted in the ionosphere [1,2]. The total electron content (TEC) is among the important parameters for studying ionospheric changes [3,4]. Accurate prediction of TEC is very important for communication systems such as satellite positioning and remote sensing systems [5,6]. With TEC prediction, corresponding protection measures can be started in satellite navigation, satellite positioning, and radio communication in the case of poor space weather conditions. Therefore, it is very important to study the TEC prediction model and predict it as early as possible.

There are two main types of ionospheric short-term prediction methods: the first type are the methods combining observation data with the ionospheric theoretical model [7], and the other type is based on the artificial neural network (ANN) [8,9]. Among them, the artificial neural network has become a popular tool in ionospheric TEC modeling and prediction due to its strong non-linear representation ability [10]. Yakubu et al. [11] used an ANN model to predict the daily and seasonal effects of TEC on the Indian equatorial station Changanacherry. Watthanasangmechai et al. [12] proposed a neural network model to predict TEC in Thailand. However, TEC data are typical time series data, which contain a strong time correlation. The TEC prediction methods based on the artificial neural network only consider the spatial position of the data and cannot characterize the characteristics of the time series, which will lead to a large prediction error. Inyurt et al. [13] showed that the ANN model could not reflect the time series characteristics of data, resulting in large prediction errors and low prediction accuracy in different seasons; Huang et al. [14] showed that a RBF neural network is not sensitive to the daily changes of TEC, resulting in a large prediction error at night. Habarulema et al. [15] showed that the ANN model is susceptible to the interference of solar activity, and the TEC prediction error varies greatly in the high and low years of solar activity. The model is insensitive to the seasonal changes of TEC, resulting in low prediction accuracy. The recurrent neural network (RNN) is a chain-connected neural network that takes sequence data as input and performs recursion in the evolution direction of the sequence. It is a deep learning model that can characterize both the spatial characteristics and the temporal characteristics of the data. It is the mainstream algorithm for time series modeling [16]. Yuan et al. [17] showed that the RNN can predict TEC; however, a unit in the RNN is mainly affected by the units near it, which will make it difficult for the gradient to be transferred back from the back layer during long-time series prediction, causing the gradient to disappear, which makes it unable to represent the non-linear relationship of long-time span, that is, it cannot solve the long-term dependence of data. Long Short-Term Memory (LSTM) can solve the gradient disappearance problem by its unique gate structure [18]. Chimsuwan et al. [19] used the LSTM model to predict TEC. However, since the LSTM model treats historical time series equally, it cannot adaptively focus on important features, which results in limited prediction accuracy. To solve this problem, this paper adds multiple attention modules and an L1 constraint to the traditional LSTM model. The attention modules can redistribute weights according to the importance of multiple historical data input to the model to improve the attention of the model to important features [20–22]. L1 constraints can prevent the model from paying too much attention to certain features. The proposed LSTM model with multiple attention modules is called multiple-attention LSTM (MA-LSTM). This paper uses MA-LSTM to conduct TEC prediction experiments in Harbin (45° N, 125° E), Beijing (40° N, 115° E), Guangzhou (22.5° N, 115° E), Tibet (30° N, 90° E), Yushu (35° N, 90° E), Ziyang (30° N, 105° E), Hangzhou (30° N, 120° E), and Heze (35° N, 115° E) in the next 2 h, compared with other time series prediction models, such as LSTM [23], GRU [24], and Att-BiGRU [25]. We also tested the performance of the MA-LSTM model for predicting TEC in a magnetic quiet period and a medium magnetic storm period, and for different months.

The contributions of this paper are as follows:

- (1). The LSTM model with multiple attention modules is applied to TEC prediction to make TEC modelling more adaptive and obtain higher prediction accuracy.
- (2). An L1 constraint is added to TEC prediction, which can avoid overfitting caused by excessive attention to a historical observation value in TEC modeling.

## 2. Data and Proposed Model

### 2.1. Data Description

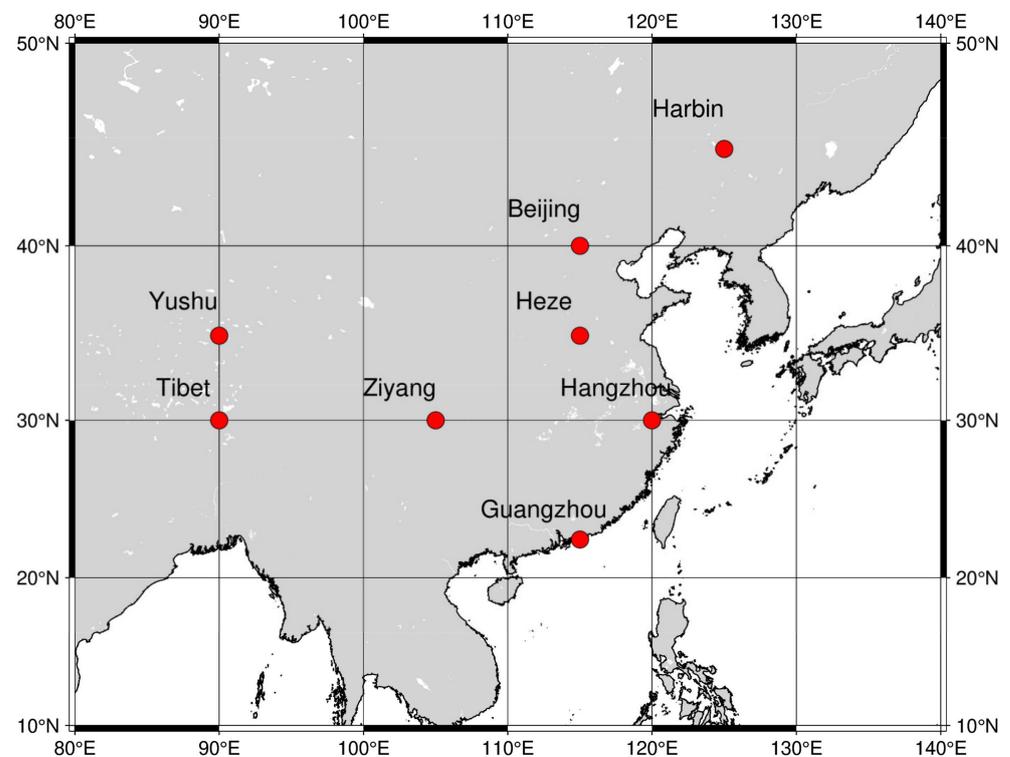
The ionospheric data used in this paper are from the Center for Orbit Determination of Europe (CODE) TEC grid data, and the data have a temporal resolution of 2 h [26,27]. Eight regions in China are selected as the research objects, namely Beijing (40° N, 115° E), Guangzhou (22.5° N, 115° E), Harbin (45° N, 125° E), Tibet (30° N, 90° E), Yushu (35° N,

90° E), Ziyang (30° N, 105° E), Hangzhou (30° N, 120° E), and Heze (35° N, 115° E). The time range for selecting experimental data is from 0:00 on 1 January 2002 to 24:00 on 31 December 2011.

The information on the geographical location of the eight areas selected is shown in Table 1 and Figure 1, and the values of TEC are shown in Figure 2.

**Table 1.** The coordinates of the research areas in this paper.

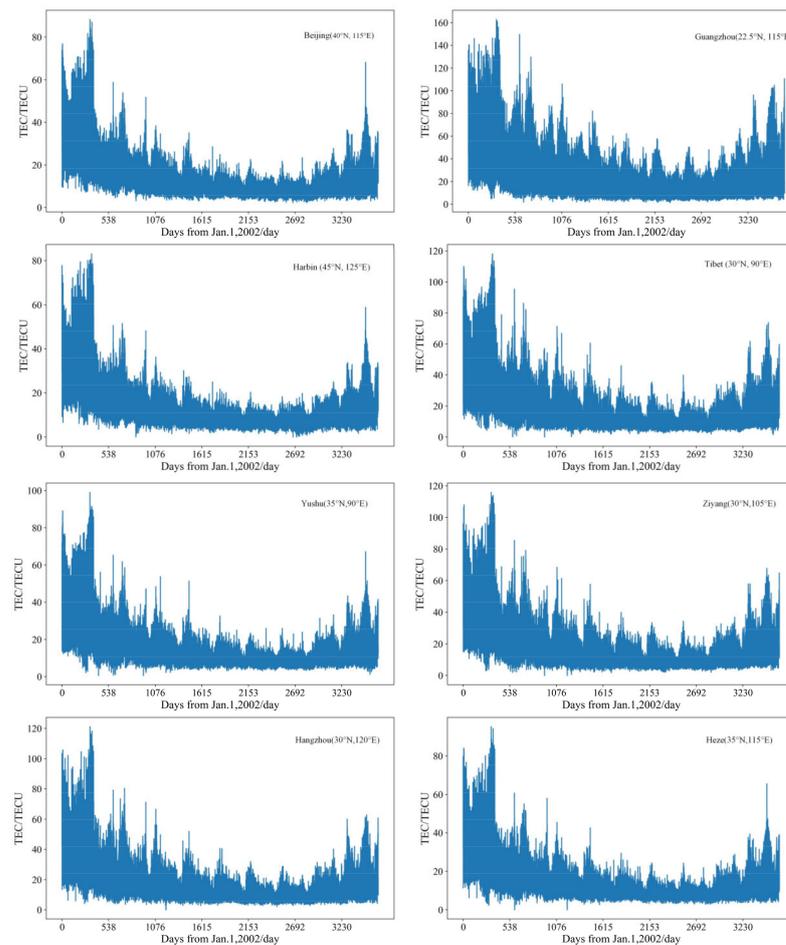
Region	Lat	Lon
Beijing	40° N	115° E
Guangzhou	22.5° N	115° E
Harbin	45° N	125° E
Tibet	30° N	90° E
Yushu	35° N	95° E
Ziyang	30° N	105° E
Hangzhou	30° N	120° E
Heze	35° N	115° E



**Figure 1.** The information on the location of the eight regions on the map.

**2.2. Data Preprocessing**

TEC data are typical time series data, and only stable non-random time series can be predicted. Therefore, a series of preprocessing is required before TEC prediction. The preprocessing of the selected ionospheric TEC data in this paper includes a TEC data stationarity test, difference processing, a pure random stationarity test, and TEC data normalization.



**Figure 2.** TEC data of eight regions selected in this paper.

### 2.2.1. The TEC Data Stationary Test and Difference Processing

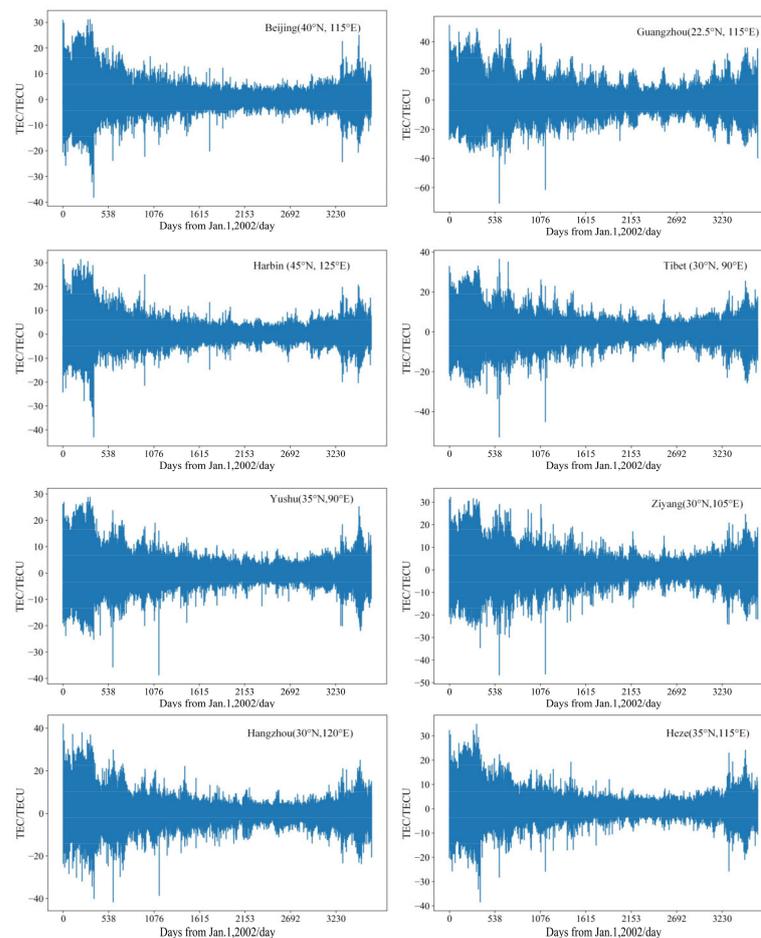
The stationarity of time series is the basic assumption of time series analysis. Therefore, the stationarity of time series needs to be tested first before TEC prediction. In this paper, we use the ADF test to determine the stationarity of TEC sequences. The stationarity test results of the eight regions are all non-stationary time series. In order to transform the data into stationary time series for prediction, first-order difference processing is required. The calculation formula of the first-order difference is as follows:

$$\Delta x_t = x_t - x_{t-1} \quad (1)$$

where  $\Delta$  is the first-order difference operator and  $x_t$  is the observation data at the time  $t$ . Figure 3 shows the results of the first-order difference of the TEC data of the eight regions in Table 1. After the first-order difference processing, the ADF (Augmented Dickey–Fuller) test is performed again. This time, all eight regions pass the ADF test, that is, the first-order difference data of the eight regions are stable time series.

### 2.2.2. The Pure Randomness Stationarity Test

Stationary time series are not necessarily predictable, and pure random stationary time series are unpredictable. Therefore, it is also necessary to detect the pure randomness of the TEC differential sequence. In this paper, the LB (Ljung–Box) method is used to test the pure randomness of time series. The LB method is used to test whether a series of observations in a given time period is purely random. LB test results show that TEC data after first-order difference processing is not purely random data and can be predicted.



**Figure 3.** TEC values after first-order difference.

### 2.2.3. Normalization of TEC Data

After the first-order difference processing, the original TEC data become a stationary non-random time series and can be predicted. However, there are still great changes in the data of the whole data space, which will affect the prediction results, so the data needs to be normalized. In this paper, min–max normalization is used to map the first-order differential TEC data between 0 and 1. The calculation formula of min–max normalization is as follows:

$$x_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (2)$$

where  $x_i$  is the TEC observation value at time  $i$ , and  $x$  is a vector composed of all TEC observations in a selected area.

### 2.2.4. Sample Making

After the previous stationarity test, difference processing, the pure randomness test and TEC data normalization processing, the experimental samples are prepared next. This paper selects the TEC data from 0:00 on 1 January 2002 to 24:00 on 31 December 2011 in eight Chinese regions, and the total number of observation points at each region is 47476. After the first-order differential processing, it becomes 47475. A segmentation method with a sliding window of  $13 + p$  ( $p$  indicates the number of samples to be predicted in the future) is used to make the normalized data into samples. The sample making process is shown in Figure 4.  $x_i$  is the input of sample  $i$ , including 13 observations, as shown by the orange data in Figure 4.  $y_i$  is the output of the sample  $i$ . When predicting the data within  $t$  hours in the future,  $y_i$  contains  $t / 2$  points, as shown by the blue points in Figure 3 (this paper is

predicting TEC data for the next two hours, so  $t = 2$  and  $p = 1$ ). The whole experiment is shown in Figure 5.

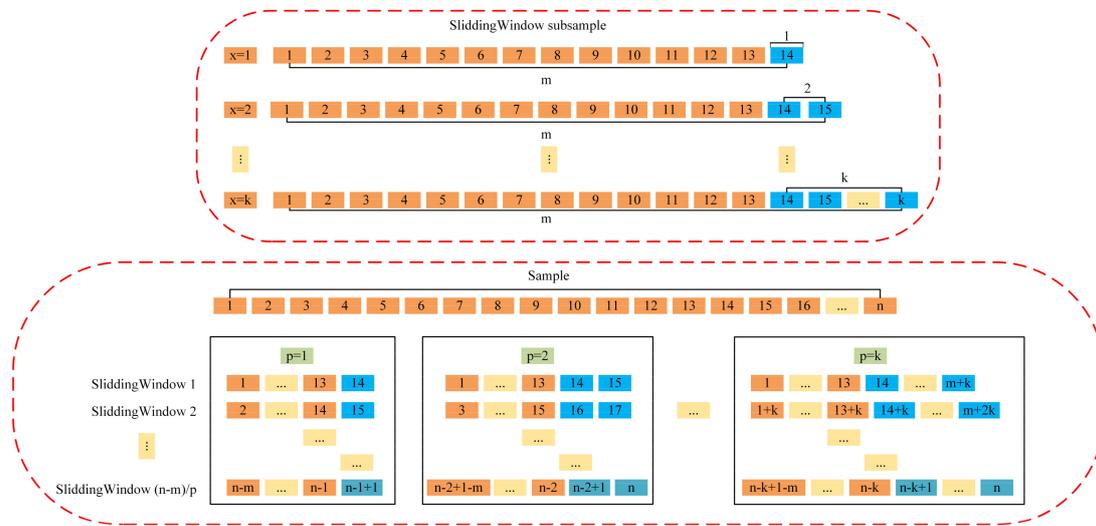


Figure 4. Schematic diagram of sample making process.

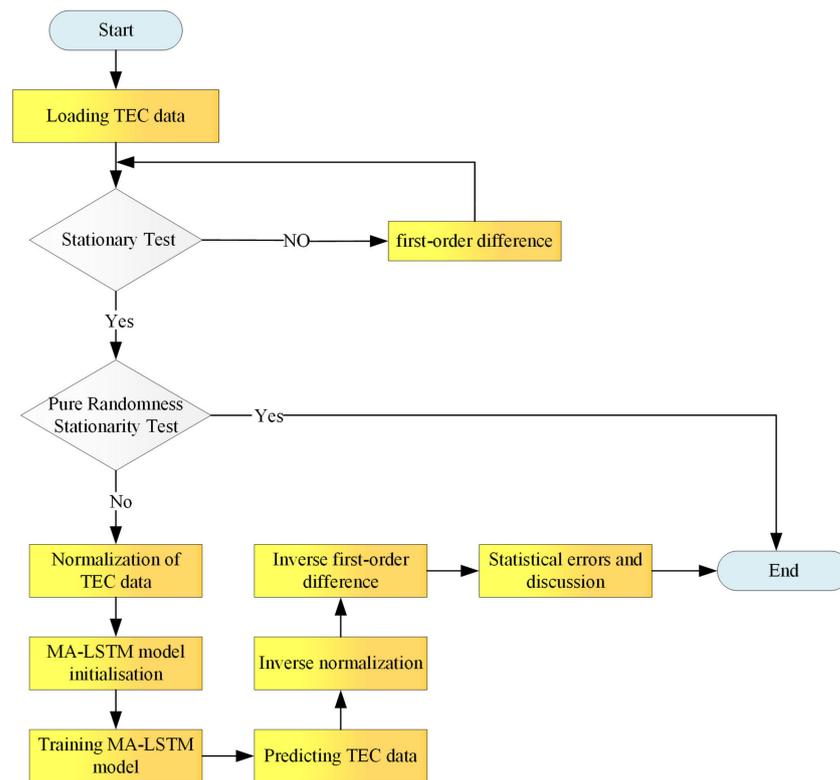


Figure 5. The entire experimental process.

### 2.3. Experimental Environment

This paper builds the MA-LSTM model based on Python 3.6 and the keras deep learning library. The experimental equipment is configured with Intel i5-7200u CPU, 8G memory, 500G solid-state hard disk, and the GPU is NVIDIA Geforce 940MX.

### 2.4. Evaluation Indexes

In order to test the performance of various models in TEC prediction, this paper uses two evaluation indexes to evaluate the model: the root mean square error (RMSE), the

R-square and the mean absolute percentage error (MAPE). The calculation formulas are shown in Equations (3)–(5):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{true_i} - y_{pre_i})^2} \quad (3)$$

$$R - square = 1 - \frac{\sum_{i=1}^n (y_{true_i} - y_{pre_i})^2}{\sum_{i=1}^n (y_{true_i} - y_{mean})^2} \quad (4)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_{pre_i} - y_{true_i}}{y_{true_i}} \right| \times 100\% \quad (5)$$

where  $n$  is the number of test samples,  $y_{true_i}$  is the true value of test sample  $i$ ,  $y_{pre_i}$  is the predicted value of test sample  $i$ , and  $y_{mean}$  is the average value of all the test samples. RMSE is used to describe the prediction error. The smaller the RMSE, the better the prediction performance of the model; the R-square is used to describe the fitting degree between the predicted value and the real value. The closer it is to 1, the better the fitting ability of the model to the TEC observation data; MAPE represents the error accuracy between the predicted and true values, the closer MAPE is to zero, the higher the prediction accuracy of the model.

## 2.5. Our Proposed Model

### 2.5.1. The Long Short-Term Memory (LSTM) Network

LSTM [28] is a recursive neural network, which is composed of several LSTM units. One LSTM unit includes three gate structures, as shown in Figure 6, namely, input gate, forgetting gate and output gate. These three gate structures are connected by memory cell units to realize purposeful selection of features in the network. The calculation formula of each module of LSTM unit type are as follows:

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \quad (6)$$

$$f_t = o_t(W_f \times [h_{t-1}, x_t] + b_f) \quad (7)$$

$$\tilde{c}_t = \tanh(W_c \times [h_{t-1}, x_t] + b_c) \quad (8)$$

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \quad (9)$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \quad (10)$$

$$h_t = o_t * \tanh(c_t) \quad (11)$$

where  $[ ]$  represents vector connection,  $\times$  represents matrix dot product,  $*$  represents Hadamard product,  $\sigma$  represents sigmoid activation function, and  $\tanh$  represents hyperbolic tangent activation function.  $W_i, W_f, W_c, W_o$  are weight matrix and  $b_i, b_f, b_c, b_o$  indicate bias matrix. The weight matrix and bias matrix can be learned during the training process.  $x_t$  represents the input of the network at time  $t$ ,  $\tilde{c}_t$  represents the updated value of memory cell state at time  $t$ .  $h_t$  represents the state of the hidden layer at the time  $t$ .  $i_t$  determines how much information in  $\tilde{c}_t$  is used to update  $c_t$ .  $f_t$  determines how much information is retained in the memory cell unit at time  $t-1$ . The state of the hidden layer at time  $t$  is determined by  $h_t, o_t$  and  $c_t$ .

### 2.5.2. LSTM Based on Multiple-Attention Modules (MA-LSTM) Proposed in This Paper

When the traditional LSTM model is used to model TEC data, the prediction weights of the data at any position in the historical sequence to the future data are equal, so it is impossible to adaptively and accurately model TEC. To solve this problem, this paper adds three attention modules [29], which can adaptively assign weights to each input sequence, so that the model can selectively focus on the historical sequence and reduce the prediction error. The proposed MA-LSTM TEC prediction model is shown in Figure 7.

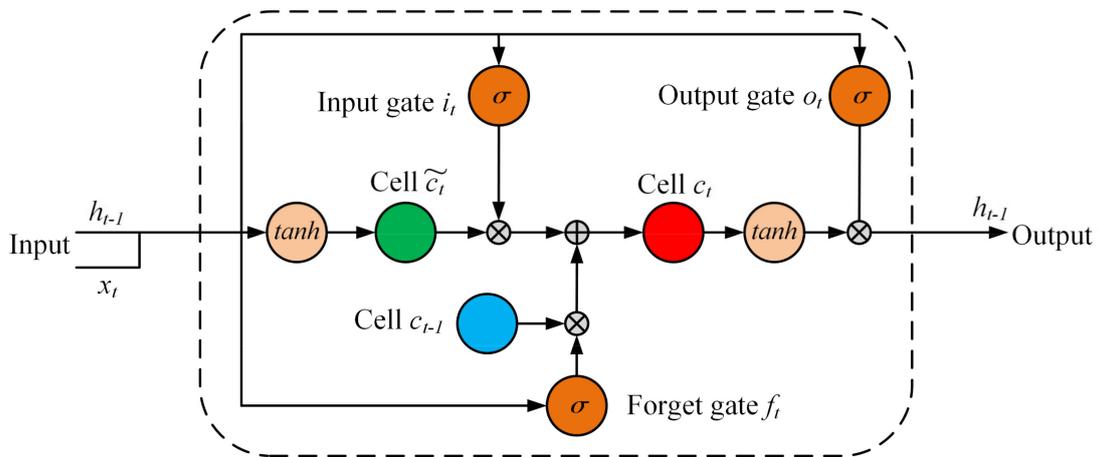


Figure 6. Structure of an LSTM unit.

The proposed MA-LSTM model includes four parts: the input layer, the LSTM layer, the multiple-attention layer and the output layer.

**Input layer:** Each sample includes two parts: feature vector  $x_n$  and regression value  $y_n$  (i.e., prediction value), which form an input sample in pairs as  $[x_n, y_n]$ .

**LSTM Layer:** This layer includes two independent LSTM neuron layers, one of which is provided with an L1 constraint and the other is not. These two layers, respectively, receives a pair of feature and regression value samples of the input layer  $[x_n, y_n]$  and process them separately. The processed results are spliced by concatenation function and transmitted to the next layer. The calculation formula of this layer is as follows:

$$k_n = LSTM([x_n, y_n]) \tag{12}$$

$$m_n = LSTM([x_n, y_n]) \tag{13}$$

$$r_n = concat([k_n, m_n]) \tag{14}$$

where  $k_n$  represents the output of the first LSTM layer with an L1 constraint.  $m_n$  represents the output of the unconstrained LSTM layer.  $r_n$  represents the vector after connecting  $k_n$  and  $m_n$ . All  $r_n$  are connected to form the output of this layer, which is  $h_n$ .

**Multiple-attention layer:** This layer contains three parallel attention branches, which, respectively, receive the data from the LSTM layer. After receiving the data transmitted from the LSTM layer, each branch first calculates the similarity between each feature in the received data and the regression value through the attention function, and each feature obtains an attention score. The calculation formula of attention function selected in this paper is as follows:

$$score(h_n, y_n) = V^T tanh[Wh_n + Uy_n] \tag{15}$$

where  $W, V, U$  are the parameters that can be learned in the training process. After obtaining the attention score, then normalize it with the softmax function to obtain the probability distribution of attention. The specific calculation formula is as follows:

$$a_{1n} = softmax(score(h_n, y_n)) \tag{16}$$

$$a_{2n} = softmax(score(h_n, y_n)) \tag{17}$$

$$a_{3n} = softmax(score(h_n, y_n)) \tag{18}$$

$a_{1n}, a_{2n}, a_{3n}$  represents the respective attention distribution values of the three branches. The softmax function is as follows:

$$\text{Softmax}(\text{score}(h_i, y_i)) = \frac{e^{\text{score}(h_i, y_i)}}{\sum_{n=1}^n e^{\text{score}(h_n, y_n)}} \quad (19)$$

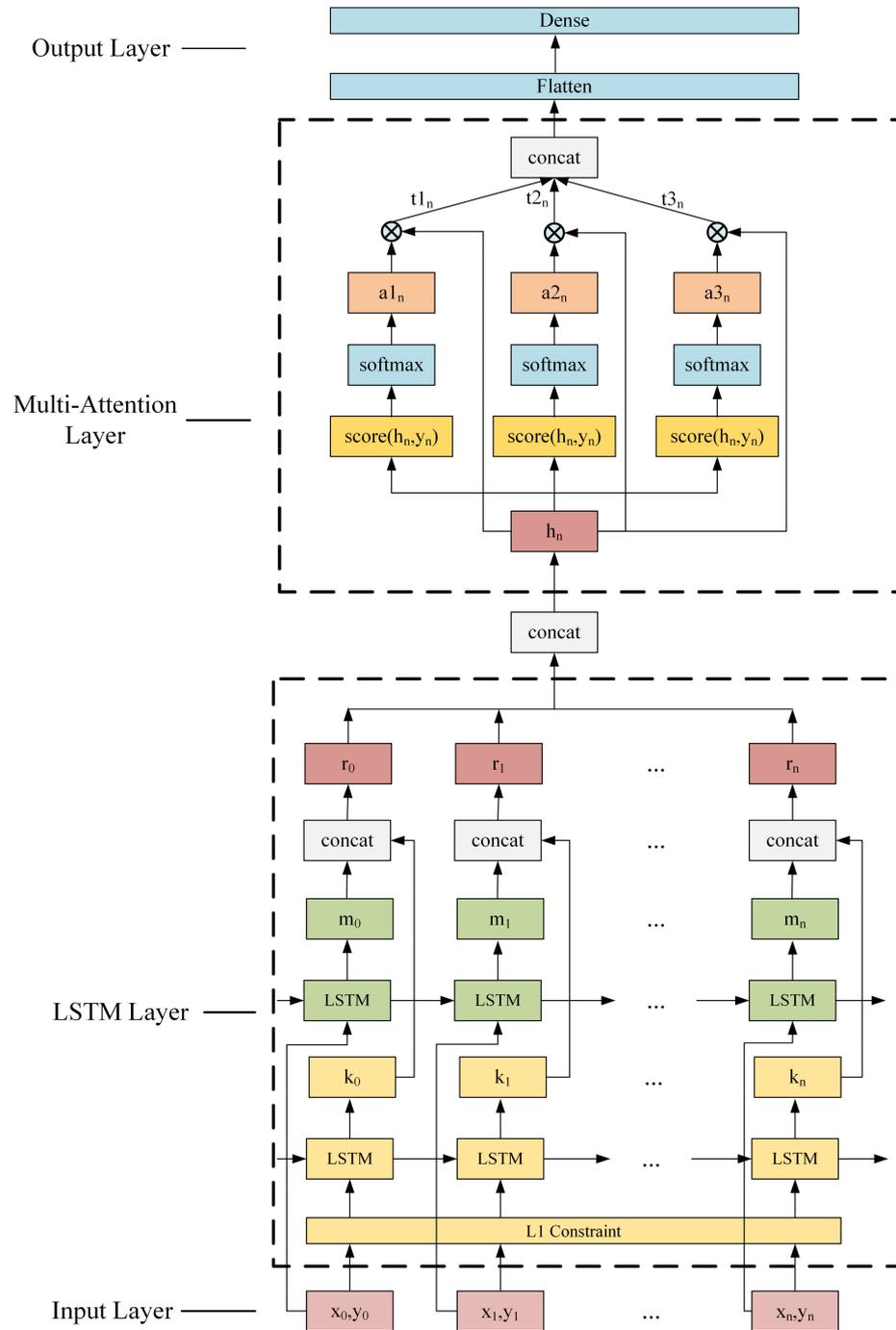


Figure 7. MA-LSTM model structure for TEC prediction.

Then,  $a_{1n}, a_{2n}, a_{3n}$  are, respectively, multiplied  $h_n$  to obtain the output value of each attention branch. The calculation formula is as follows:

$$t_{1n} = a_{1n} \times h_n \quad (20)$$

$$t_{2n} = a_{2n} \times h_n \quad (21)$$

$$t_{3n} = a_{3n} \times h_n \quad (22)$$

Finally, connect the attention values of the three branches as the output of the attention layer.

**Output layer:** This layer includes a flattened layer and a fully connected layer. It is used to map the previously results into predicted values and then output them.

### 3. Experimental Results and Discussion

#### 3.1. Model Parameter Selection

When MA-LSTM is used for TEC modeling, the optimal parameters of the model need to be determined first. In this paper, the grid search method is adopted to find the optimal super parameters. Result of grid search is as shown in Table 2.

**Table 2.** Optimal parameters of the MA-LSTM model gained by grid search.

Parameter Name	Optimal Value
input_dim	13
optimizer	Adagrad
loss	“RMSE” and “R-Square”
activation	“Relu”and “Softmax”
activity_regularizer	L1(0.01)
learning rate	0.01
batch_size	64
num of hidden units	64

#### 3.2. Prediction Comparison of Different Stations

This paper first discusses the situation of TEC prediction at the second hour. The MA-LSTM model proposed in this paper is compared with the classical time series models LSTM, GRU and Att-BiGRU (in terms of structure, the LSTM is a three-layer gate structure, the GRU is a two-layer gate structure, and the Att-BiGRU is a two-layer gate structure in both directions). In the selected eight regions, samples were made according to the method described in Section 2.2. A total of 90% of the samples were used for training models and 10% for testing. The training data are 42715 TEC observations from 0:00 on 1 January 2002 to 18:00 on 30 December 2010, and the test data are 4747 TEC observations from 20:00 on 30 December 2010 to 24:00 on 31 December 2011. The predictive performance of each model is shown in Figure 8. The specific values are shown in Table 3.

**Table 3.** Comparison of prediction performance of different models.

Algorithm	Indicator	Mean	Beijing	Guangzhou	Harbin	Tibet	Yushu	Ziyang	Hangzhou	Heze
LSTM	RMSE (TECU)	4.137	2.962	4.683	6.834	2.745	3.880	4.561	4.225	3.208
GRU		3.846	3.221	4.613	6.676	2.683	3.031	4.392	3.487	2.664
Att-BiGRU		1.581	0.801	1.867	4.244	0.650	0.957	1.354	1.445	1.331
MA-LSTM		<b>1.171</b>	<b>0.718</b>	<b>1.550</b>	<b>3.377</b>	<b>0.602</b>	<b>0.934</b>	<b>1.283</b>	<b>0.543</b>	<b>0.358</b>
LSTM	R-square	0.857	0.860	0.856	0.874	0.877	0.831	0.850	0.850	0.855
GRU		0.874	0.839	0.860	0.880	0.880	0.890	0.859	0.893	0.895
Att-BiGRU		0.979	0.988	0.974	0.953	0.992	0.987	0.985	0.980	0.972
MA-LSTM		<b>0.988</b>	<b>0.991</b>	<b>0.982</b>	<b>0.970</b>	<b>0.993</b>	<b>0.988</b>	<b>0.986</b>	<b>0.997</b>	<b>0.998</b>
LSTM	MAPE	0.188	0.153	0.217	0.268	0.149	0.183	0.205	0.184	0.150
GRU		0.179	0.166	0.214	0.267	0.153	0.143	0.200	0.154	0.135
Att-BiGRU		0.100	0.139	0.191	0.166	0.047	0.051	0.065	0.069	0.072
MA-LSTM		<b>0.054</b>	<b>0.037</b>	<b>0.075</b>	<b>0.136</b>	<b>0.034</b>	<b>0.047</b>	<b>0.061</b>	<b>0.024</b>	<b>0.018</b>

As can be seen from Figure 8 and Table 3, the performance of the models with attention mechanism (MA-LSTM and Att-BiGRU) is significantly better than those of the models without attention mechanism (LSTM and GRU). The multiple-attention mechanism model proposed in this paper (MA-LSTM) is superior to the single attention mechanism model (Att-BiGRU). In the eight selected regions, the average RMSE of the MA-LSTM model proposed in this paper is 1.171TECU, which is reduced by 253.29, 228.44 and 35.01%,

respectively, compared with LSTM, GRU and Att-BiGRU. The average R-square is 0.988, increased by 15.29, 13.04 and 0.01%, respectively. The average MAPE of our model is 0.054, which is increased by 248.15, 231.48 and 85.19%, respectively.

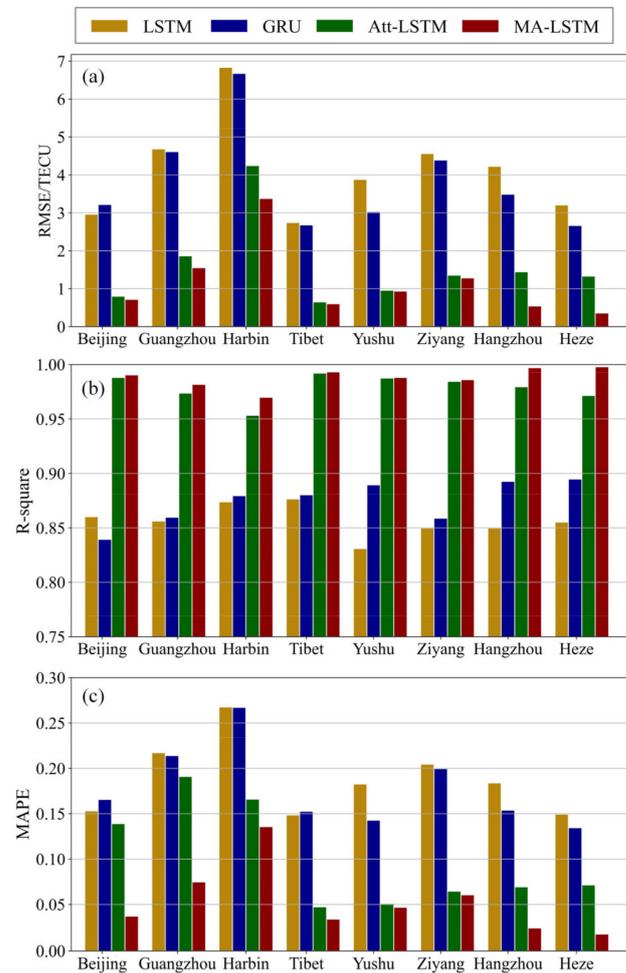


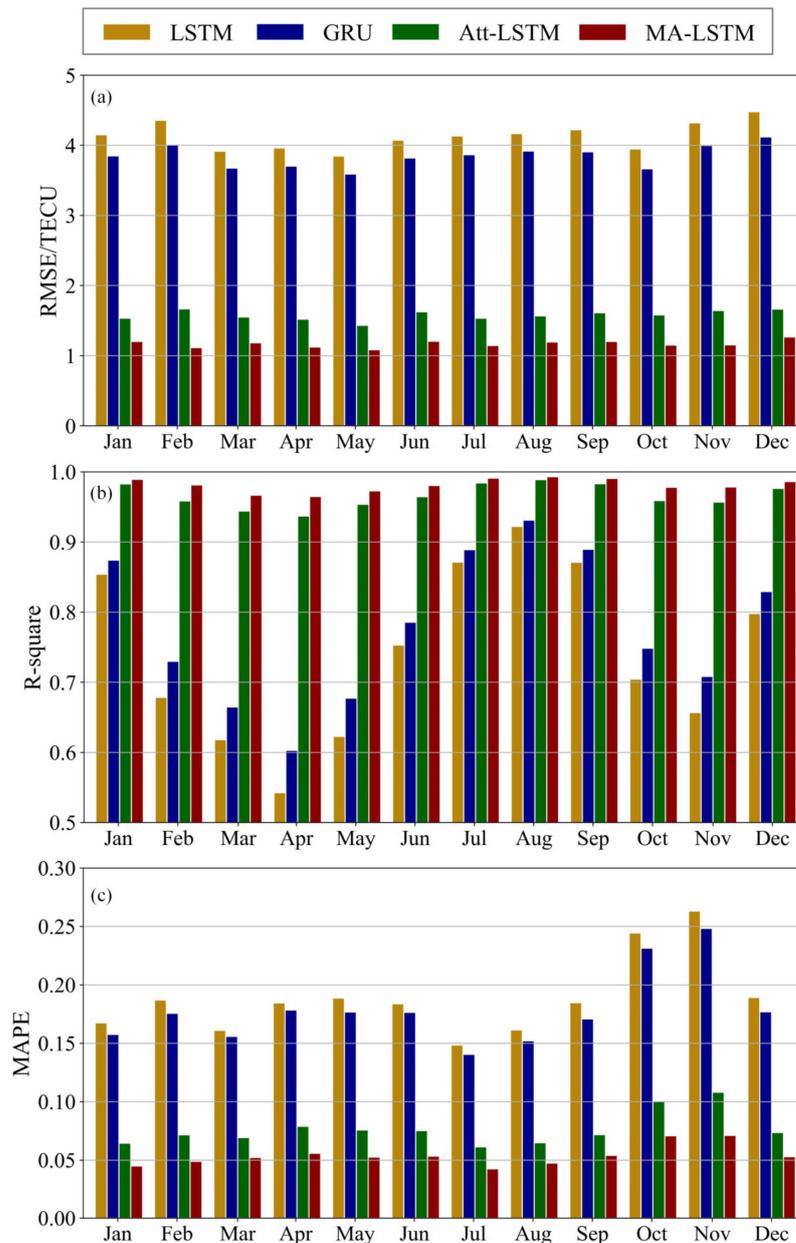
Figure 8. Comparison of (a) RMSE, (b) R-square, (c) MAPE values of different models in all regions.

### 3.3. Prediction Comparison of Different Months

This section mainly discusses the prediction performance of the LSTM, GRU, Att Bi-GRU and MA-LSTM models in different months of the eight regions in Table 1. The training data are 42715 TEC sequences from 0:00 on 1 January 2002 to 18:00 on 30 December 2010, and the test data are TEC sequences from 12 different months in 2011. The mean prediction performance of different models in different months are shown in Table 4 and Figure 9. It can be seen that the fitting index R-square of our model exceeded 0.99 in July, August and September, and the maximum reaches 0.993, which appears in August. The lowest and highest RMSE values of our model are 1.086TECU and 1.267TECU, which are lower than those of the comparison models. The minimum MAPE of a single month of our model is 0.042, and the maximum is 0.071, which are significantly lower than those of the comparison model. That is to say, in all selected regions, in all months, in all comparison indicators, our MA-LSTM mode outperforms other comparison models. From Figure 9 we also can see that the R-square of our model is high in July, August and September, which indicates that the predicted value has a good fit with the real value in these three months. In March, April and May, the R-square is slightly low, and the prediction performance of the model is slightly poor, but even at the worst time, the fitting degree between the predicted value and the real value still reaches 0.965.

**Table 4.** Comparison of prediction performance by month.

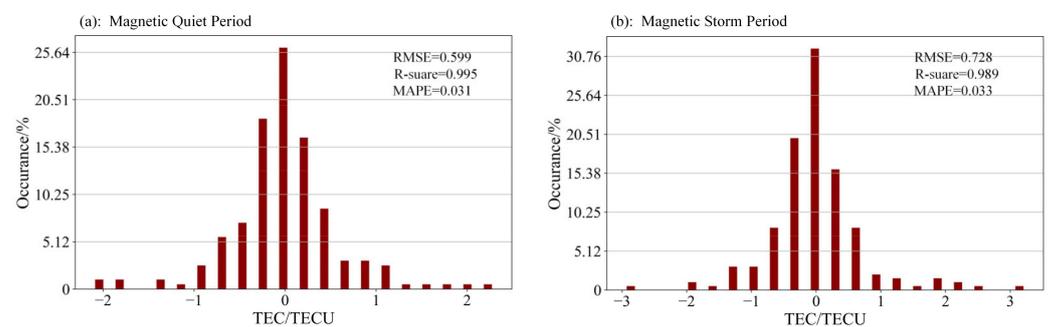
Algorithm	Indicator	January	February	March	April	May	June	July	August	September	October	November	December
LSTM	RMSE (TECU)	4.150	4.357	3.915	3.964	3.848	4.072	4.132	4.168	4.222	3.950	4.321	4.479
GRU		3.848	4.006	3.675	3.703	3.589	3.818	3.866	3.917	3.907	3.664	3.998	4.118
Att-BiGRU		1.533	1.666	1.552	1.522	1.431	1.625	1.534	1.568	1.613	1.581	1.645	1.664
MA-LSTM		<b>1.202</b>	<b>1.115</b>	<b>1.184</b>	<b>1.122</b>	<b>1.086</b>	<b>1.206</b>	<b>1.145</b>	<b>1.195</b>	<b>1.202</b>	<b>1.150</b>	<b>1.154</b>	<b>1.267</b>
LSTM	R-square	0.854	0.678	0.618	0.542	0.623	0.753	0.871	0.922	0.871	0.704	0.657	0.798
GRU		0.874	0.730	0.665	0.603	0.678	0.786	0.889	0.931	0.890	0.749	0.708	0.829
Att-BiGRU		0.983	0.959	0.944	0.937	0.954	0.965	0.984	0.989	0.983	0.959	0.957	0.977
MA-LSTM		<b>0.989</b>	<b>0.981</b>	<b>0.967</b>	<b>0.965</b>	<b>0.973</b>	<b>0.981</b>	<b>0.991</b>	<b>0.993</b>	<b>0.991</b>	<b>0.978</b>	<b>0.979</b>	<b>0.986</b>
LSTM	MAPE	0.168	0.187	0.161	0.185	0.189	0.184	0.148	0.161	0.185	0.244	0.263	0.189
GRU		0.158	0.176	0.156	0.178	0.177	0.177	0.141	0.152	0.171	0.231	0.248	0.177
Att-BiGRU		0.064	0.072	0.069	0.079	0.076	0.075	0.061	0.065	0.072	0.101	0.108	0.074
MA-LSTM		<b>0.045</b>	<b>0.049</b>	<b>0.052</b>	<b>0.056</b>	<b>0.053</b>	<b>0.053</b>	<b>0.042</b>	<b>0.048</b>	<b>0.054</b>	<b>0.071</b>	<b>0.071</b>	<b>0.053</b>



**Figure 9.** Comparison of (a) RMSE, (b) R-square, and (c) MAPE values for various models in different months.

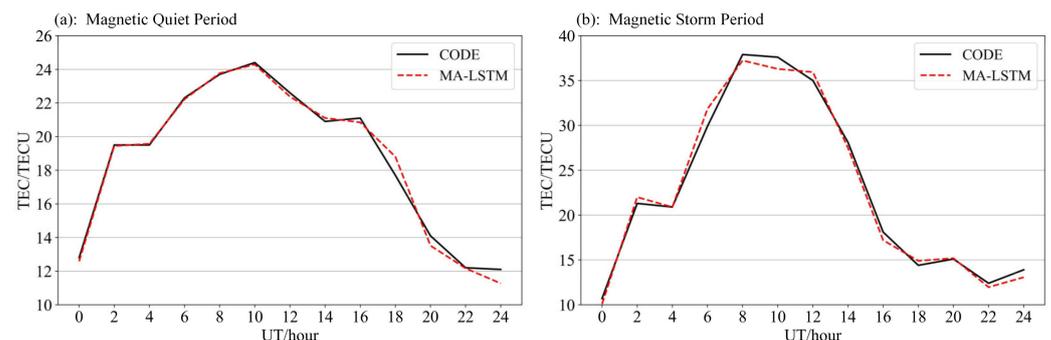
### 3.4. Prediction Comparison of Different Geomagnetic Conditions

In order to further study the prediction ability of the MA-LSTM model under different geomagnetic activities, we divided the test data into a magnetic quiet period and a magnetic storm period for comparative analysis (in this paper, samples with  $K_p < 3$  are taken as magnetic quiet days, and samples with  $K_p > 3$  and  $-100 < Dst \leq -50$  are taken as magnetic storm periods). Histograms of absolute error distribution during magnetic quiet periods and magnetic storm periods are shown in Figure 10, from which we can see that during the magnetic quiet period, the absolute error distribution of the prediction is more concentrated, basically between  $[-2, 2]$ , while during the magnetic storm, the absolute error distribution of the model is wider, reaching  $[-3, 3]$ , that is, the prediction effect of the MA-LSTM model during the magnetic quiet period is better than that during the magnetic storm. However, even during the magnetic storm period, the R-square of our model can still reach 0.989 in the selected area.



**Figure 10.** Histograms of absolute error distribution during (a) magnetic quiet periods and (b) magnetic storm periods.

Figure 11 shows the comparison between the predicted value and the true value of a magnetic quiet day and a magnetic storm day, from which we can see that on the magnetostatic day, the predicted value matches the real value perfectly, while on a magnetic storm day, the error between the predicted value and the true value is relatively large, and the fitting is not very good.



**Figure 11.** Comparison of the predictive performance of a (a) magnetic quiet day and a (b) magnetic storm day.

## 4. Conclusions

In this paper, we proposed a multiple-attention LSTM (MA-LSTM)-based TEC prediction model, and applied it to the TEC prediction of eight locations—Beijing, Guangzhou, Harbin, Tibet, Yushu, Ziyang, Hangzhou and Heze in China. We compared our model in this paper with LSTM, GRU and Att-BiGRU. The experimental results show that the MA-LSTM model in this paper is obviously superior to the comparison models. We also discussed the prediction performance of our model in different month. Results show that our MA-LSTM model has the highest prediction performance in July, August and September, with the R-square reaching above 0.99. In March, April and May, the prediction

performance is slightly poor, but even at the worst time, the R-square still reaches 0.965. We further discussed the prediction effect during a magnetic quiet period and a magnetic storm period. The results show that the prediction effect of a magnetic quiet period is slightly better than that of magnetic storm.

The author will study the grid prediction of TEC at multiple time points in the future.

**Author Contributions:** Writing—review and editing, H.L.; writing—original draft and methodology, D.L.; validation and resources, J.Y.; data curation, G.Y.; investigation, C.C.; software, Y.W.; project administration, W.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Self funded scientific research and development program of Langfang Science and Technology Bureau, China (2022011046).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. Ionospheric TEC data can be found here: [https://pan.baidu.com/s/1yojZd9MKWc\\_COcbK5Xk\\_XA](https://pan.baidu.com/s/1yojZd9MKWc_COcbK5Xk_XA), accessed on 27 August 2022.

**Acknowledgments:** Thanks to the TEC data provided by the European orbital determination Center (CODE), Thank Le Huijun of the Chinese Academy of Sciences for his help and suggestions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Kaselimi, M.; Voulodimos, A.; Doulamis, N.; Doulamis, A.; Delikaraoglou, D. A causal long short-term memory sequence to sequence model for TEC prediction using GNSS observations. *Remote Sens.* **2020**, *12*, 1354. [CrossRef]
2. Tang, J.; Li, Y.; Yang, D.; Ding, M. An Approach for Predicting Global Ionospheric TEC Using Machine Learning. *Remote Sens.* **2022**, *14*, 1585. [CrossRef]
3. Xiong, P.; Zhai, D.; Long, C.; Zhou, H.; Zhang, X.; Shen, X. Long short-term memory neural network for ionospheric total electron content forecasting over China. *Space Weather* **2021**, *19*, e2020SW002706. [CrossRef]
4. Sharma, G.; Mohanty, S.; Kannaujia, S. Ionospheric TEC modelling for earthquakes precursors from GNSS data. *Quat. Int.* **2017**, *462*, 65–74. [CrossRef]
5. Belehaki, A.; Stanislawski, I.; Liliensten, J. An overview of ionosphere—Thermosphere models available for space weather purposes. *Space Sci. Rev.* **2009**, *147*, 271–313. [CrossRef]
6. Samardjiev, T.; Bradley, P.A.; Cander, L.R.; Dick, M.I. Ionospheric mapping computer contouring techniques. *Electron. Lett.* **1993**, *29*, 1794–1795. [CrossRef]
7. Tang, J.; Zhang, S.; Huo, X.; Wu, X. Ionospheric Assimilation of GNSS TEC into IRI Model Using a Local Ensemble Kalman Filter. *Remote Sens.* **2022**, *14*, 3267. [CrossRef]
8. Qiao, J.; Liu, Y.; Fan, Z.; Tang, Q.; Li, X.; Zhang, F.; Song, Y.; He, F.; Zhou, C.; Qing, H.; et al. Ionospheric TEC data assimilation based on Gauss–Markov Kalman filter. *Adv. Space Res.* **2021**, *68*, 4189–4204. [CrossRef]
9. Yue, X.A.; Wan, W.X.; Liu, L.B.; Le, H.J.; Chen, Y.D.; Yu, T. Development of a middle and low latitude theoretical ionospheric model and an observation system data assimilation experiment. *Chin. Sci. Bull.* **2008**, *53*, 94–101. [CrossRef]
10. Akhoondzadeh, M. A MLP neural network as an investigator of TEC time series to detect seismo-ionospheric anomalies. *Adv. Space Res.* **2013**, *51*, 2048–2057. [CrossRef]
11. Unnikrishnan, K.; Haridas, S.; Choudhary, R.K.; Dinil Bose, P. Neural Network Model for the prediction of TEC variabilities over Indian equatorial sector. *Indian J. Sci. Res.* **2018**, *18*, 56–58.
12. Watthanasangmechai, K.; Supnithi, P.; Lerkvaranyu, S.; Tsugawa, T.; Nagatsuma, T.; Maruyama, T. TEC prediction with neural network for equatorial latitude station in Thailand. *Earth Planets Space* **2012**, *64*, 473–483. [CrossRef]
13. Inyurt, S.; Sekertekin, A. Modeling and predicting seasonal ionospheric variations in Turkey using artificial neural network (ANN). *Astrophys. Space Sci.* **2019**, *364*, 62. [CrossRef]
14. Huang, Z.; Yuan, H. Ionospheric single-station TEC short-term forecast using RBF neural network. *Radio Sci.* **2014**, *49*, 283–292. [CrossRef]
15. Habarulema, J.B.; McKinnell, L.A.; Cilliers, P.J. Prediction of global positioning system total electron content using neural networks over South Africa. *J. Atmos. Sol.-Terr. Phys.* **2007**, *69*, 1842–1850. [CrossRef]
16. Ruwali, A.; Kumar, A.J.S.; Prakash, K.B.; Sivavaraprassad, G.; Ratnam, D.V. Implementation of hybrid deep learning model (LSTM-CNN) for ionospheric TEC forecasting using GPS data. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 1004–1008. [CrossRef]
17. Yuan, T.; Chen, Y.; Liu, S.; Gong, J. Prediction model for ionospheric total electron content based on deep learning recurrent neural network. *Chin. J. Space Sci.* **2018**, *38*, 48–57. [CrossRef]

18. Tang, R.; Zeng, F.; Chen, Z.; Wang, J.-S.; Huang, C.-M.; Wu, Z. The comparison of predicting storm-time ionospheric TEC by three methods: ARIMA, LSTM, and Seq2Seq. *Atmosphere* **2020**, *11*, 316. [[CrossRef](#)]
19. Chimsuwan, P.; Supnithi, P.; Phakphisut, W.; Myint, L.M.M. Construction of LSTM model for total electron content (TEC) prediction in Thailand. In Proceedings of the 2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Chiang Mai, Thailand, 19–22 May 2021; pp. 276–279.
20. Ren, Q.; Li, M.; Li, H.; Shen, Y. A novel deep learning prediction model for concrete dam displacements using interpretable mixed attention mechanism. *Adv. Eng. Inform.* **2021**, *50*, 101407. [[CrossRef](#)]
21. Li, X.; Yuan, A.; Lu, X. Vision-to-language tasks based on attributes and attention mechanism. *IEEE Trans. Cybern.* **2019**, *51*, 913–926. [[CrossRef](#)]
22. Liu, F.; Zhou, X.; Cao, J.; Wang, Z.; Wang, T.; Wang, H.; Zhang, Y. Anomaly detection in quasi-periodic time series based on automatic data segmentation and attentional LSTM-CNN. *IEEE Trans. Knowl. Data Eng.* **2020**, *34*, 2626–2640. [[CrossRef](#)]
23. Liu, L.; Zou, S.; Yao, Y.; Wang, Z. Forecasting global ionospheric TEC using deep learning approach. *Space Weather* **2020**, *18*, e2020SW002501. [[CrossRef](#)]
24. Iluore, K.; Lu, J. Long short-term memory and gated recurrent neural networks to predict the ionospheric vertical total electron content. *Adv. Space Res.* **2022**, *70*, 652–665. [[CrossRef](#)]
25. Lei, D.; Liu, H.; Le, H.; Huang, J.; Yuan, J.; Li, L.; Wang, Y. Ionospheric TEC Prediction Base on Attentional BiGRU. *Atmosphere* **2022**, *13*, 1039. [[CrossRef](#)]
26. Li, Z.; Wang, N.; Liu, A.; Yuan, Y.; Wang, L.; Hernández-Pajares, M.; Krankowski, A.; Yuan, H. Status of CAS global ionospheric maps after the maximum of solar cycle 24. *Satell. Navig.* **2021**, *2*, 19. [[CrossRef](#)]
27. Li, Z.; Wang, N.; Hernández-Pajares, M.; Yuan, Y.; Krankowski, A.; Liu, A.; Zha, J.; García-Rigo, A.; Roma-Dollase, D.; Yang, H.; et al. IGS real-time service for global ionospheric total electron content modeling. *J. Geod.* **2020**, *94*, 32. [[CrossRef](#)]
28. Graves, A. Supervised sequence labelling. In *Supervised Sequence Labelling with Recurrent Neural Networks*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 37–45.
29. Galassi, A.; Lippi, M.; Torrioni, P. Attention in natural language processing. *IEEE Trans. Neural Netw. Learn. Syst.* **2020**, *32*, 4291–4308. [[CrossRef](#)]