

## Research Article

# TEC Data Forecasting Using a Novel Nonlinear Model

Jun Wang,<sup>1</sup> Bihua Zhou,<sup>1</sup> and Shudao Zhou<sup>2</sup>

<sup>1</sup>National Key Laboratory on Electromagnetic Environmental Effects and Electro-Optical Engineering, PLA University of Science and Technology, Nanjing 210007, China

<sup>2</sup>College of Meteorology and Oceanography, PLA University of Science and Technology, Nanjing 211101, China

Correspondence should be addressed to Jun Wang; wangjun19007@163.com

Received 3 April 2015; Revised 14 July 2015; Accepted 16 July 2015

Academic Editor: Elmetwally Elabbasy

Copyright © 2015 Jun Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A novel nonlinear TEC forecasting model is proposed in the paper; the main produces of the model are as follows: first the EOF decomposition of TEC data is made, then the genetic algorithm is used to establish the nonlinear time field model, and finally the decomposed space field and the predicted time field are reconstructed to achieve the purpose of forecasting the TEC data. Experiments indicate that the performance of the novel forecasting model is effective and superior to the direct forecasting and linear forecasting models.

## 1. Introduction

The ionosphere is the main area of spaceflight activities, and during these activities the processes of energy transfer and severe disturbance will occur [1, 2]. These ionospheric activities have considerable influence on human military activities and daily life. TEC (Total Electron Content) is the key parameter in describing ionospheric characteristics, so that measuring and forecasting TEC is significant [3–6], since now there have been some TEC forecasting methods proposed in many literatures [7–10]. However, the forecasting accuracies of those methods are not high enough and are possible to be improved further. Hence, it is essential to propose other novel methods that further improve the accuracy of TEC forecasting.

In the experiment, the global TEC data of 2007 are derived from the Shanghai observatory of Chinese Academy of Sciences. The data are observed every two hours, and there are 4380 observations in a year. The observatory area is from 180W to 180E and 87.5S to 87.5N, the data pixel size is  $5183 \times 31 \times 12$ , and the resolution in longitude is  $5^\circ$  and  $2.5^\circ$  in latitude.

In the actual practice, the global TEC data are derived from the Shanghai observatory of Chinese Academy of Sciences, and the TEC of the area of 180W to 180E and 87.5S to 87.5N are observed every two hours; the resolution in longitude is  $5^\circ$  and  $2.5^\circ$  in latitude, so there are total 5183 (73

$\times 71$ ) grid points. It is feasible in theory to forecast the TEC of the 5183 grid points; however, the forecasting error is very large without consideration of changing physics features of the TEC of grid points. Therefore, in order to get over the above defects and improve the forecasting accuracy, a new forecasting model is proposed, and the main processes are described as follows: firstly the TEC values of the 5183 points on the Earth-fixed coordinate are converted to the values on the sun-fixed coordinate to reflect the physical features of TEC data distribution, and then the data on the sun-fixed coordinate are decomposed by the method of EOF [11, 12]. Taking the time field as a series of nonlinear time series, then the genetic algorithm [13–15] is used to establish the best functional relation of time series to forecast the time field. The results confirm that the method is better than traditional forecasting models.

The paper is organized as follows. The principle and detailed processes of the proposed forecasting model are described in Section 2. The results and analytics are elaborated in Section 3. A conclusion of the work is given in Section 4.

## 2. Principle of Nonlinear Forecasting Model

The principle of the proposed forecasting model contains three parts: the first step is transforming the original

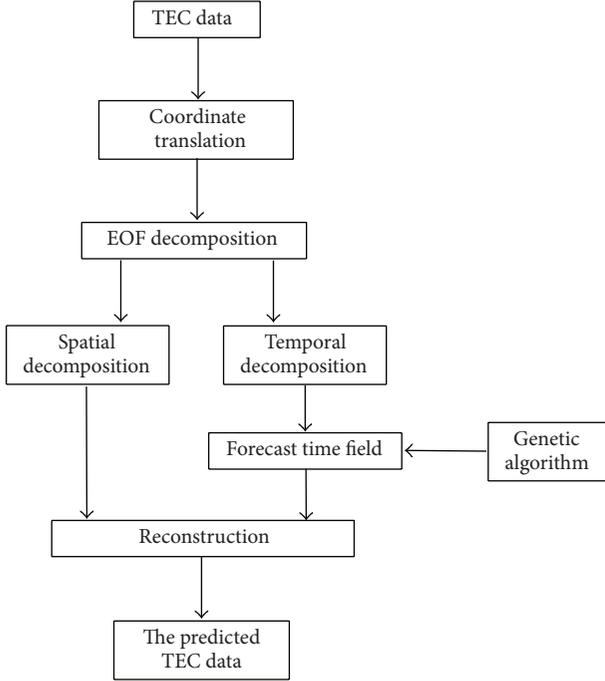


FIGURE 1: General graphical procedures of forecasting model.

Earth-fixed coordinate to sun-fixed coordinate, the second step is making EOF decomposition of the TEC data, the third step is using the genetic algorithm to establish the nonlinear forecasting equation to forecast the time field, and the final step is reconstructing the decomposed space field and the predicted time field; after finishing the above steps the purpose of forecasting TEC and reducing noise can be achieved.

The general graphical procedures of forecasting model are illustrated in Figure 1, and the detailed processes of the above steps are described in Figure 1.

**2.1. Coordinate Transform.** The TEC data are recorded on Earth-fixed coordinate and, in order to reflect physical features of the TEC distribution, the original Earth-fixed coordinate needs to be converted to sun-fixed coordinate.

The way of transforming the Earth-fixed coordinate system to the sun-fixed coordinate system can be described in (1), where  $TEC(Lat, Lon, UT)$  denotes the three-dimension distribution of TEC data on the Earth-fixed coordinate and  $TEC(Lat, LT)$  denotes the two-dimension distribution of TEC data on the sun-fixed coordinate system,  $Lat$  and  $Lon$  are geographical latitude and longitude, respectively (in degrees for unit), and  $UT$  and  $LT$  are the world and local time (with the hour for the unit), respectively. The main advantages of the sun-fixed coordinate system are as follows: first the calculation is simplified and its stability is improved by reducing the dimension, and then the precision of estimating is improved by making full use of the high temporal resolution of GPS observation:

$$\begin{aligned} TEC(Lat, LT) &= TEC(Lat, Lon, UT), \\ LT &= UT + \frac{Lon}{15}. \end{aligned} \quad (1)$$

**2.2. EOF Decomposition and Reconstruction.** Assume matrix  $X$  is the original observation data and the dimension of  $X$  is  $M \times N$  ( $M > N$ ), where  $M$  is the spatial dimension and  $N$  is the temporal dimension. Each column of  $X$  corresponds to a dataset of size  $M$ , and there are  $N$  observation data. The EOF decomposition of  $X$  can be written as  $X = UDV^T$ , where  $U$  represents the spatial decomposition field and its dimension is  $M \times r$ ,  $V$  represents temporal decomposition field and its dimension is  $N \times r$ , and  $D$  is a  $r \times r$  ( $r \leq \min(M, N)$ ) diagonal matrix filled with eigenvalues. When the values of matrix  $U$ ,  $D$ , and  $V$  are confirmed, the matrix  $X$  can be resolved via reconstruction formula  $X = UDV^T = \sum_{i=1}^k \rho_i U_i V_i^T$ , where  $\rho_i$  is the  $i$ -order eigenvalue and  $k$  ( $k \leq r$ ) is the order. The optimal  $k$  is a key point during the EOF data reconstruction. On one hand, if the value of  $k$  is too small, the reconstructed data cannot reflect the inner physical distribution. On the other hand, if the value of  $k$  is too large, the errors are large and the calculation time is long. The general criterion of determining the optimal  $k$  is as follows: the variance contribution rate of the  $k$  orders eigenvalues should be 90% of the variance contribution of the whole eigenvalues established by the general approach of sensitivity analysis [16]. The criterion can make the reconstructed data reflect the inner physical distribution and avoid the large error.

**2.3. Nonlinear Forecasting of Time Field by Genetic Algorithm.** The time field can be taken as a series of nonlinear time series which are described in (2), where  $t$  is a scalar index for the time series,  $\tau$  is the time delay, and  $m$  is the embedding dimension. Consider

$$\vec{X}(t) = [x(t), x(t-\tau), \dots, x(t-(m-1)\tau)]. \quad (2)$$

We need to determine the dependence of the state value  $x(t)$  on its previous state values. Takens Embedding Theorem guarantees that the system's state information can be recovered from a sufficiently long observation of the output time series. According to the theorem, the system state follows the existence of a smooth map  $T: R^m \rightarrow R$  satisfying

$$x(t) = T(x(t-\tau), x(t-2\tau), \dots, x(t-m\tau)). \quad (3)$$

Thus, the time series can be forecasted by establishing the functional relation  $T(\cdot)$  with the genetic algorithm. The genetic algorithm is an evolutionary method that mimics the process of natural selection. GA generates solutions to optimization problems using techniques inspired by natural evolution, such as selection, crossover, and mutation. During each generation, a proportion of the existing population is selected to breed a new generation through a fitness-based process. Crossover is a process of taking more than one parent solutions and producing a child solution from them. It is analogous to reproduction and biological crossover. Mutation is used to maintain genetic diversity from one generation of a population of genetic algorithm chromosomes to the next. It is analogous to biological mutation. Mutation alters one or more gene values in a chromosome from its initial state. The purpose of mutation is preserving and introducing diversity. Hence GA can come to better solution by using mutation.

TABLE 1: Encoding rule of equation strings.

	Second coordinate	First coordinate	Pair
Operator term	1	1	+
		2	-
		3	×
		4	÷
Constant term	2	$k \in [-\lambda, \lambda]$	$k$
Variable term	3	1	$x(t - 1 \cdot \tau)$
		$\vdots$	$\vdots$
		$i$	$x(t - i \cdot \tau)$

The detailed procedures of using genetic algorithm to establish the best functional relation are described as follows.

*Step 1* (set the encoding rules). In order to encode the symbolic form of the equation strings into a numerical structure, the encoding utilizes coordinate pairs, and the detailed rules are listed in Table 1. The second coordinate

indicates an argument or operator in the equation. If the second coordinate is 1, then the pair represents operator. If the second coordinate is 2, then the pair represents constant term of the equation. If the second coordinate is 3, then the pair represents the variable of the equation.

*Step 2* (initialize). The precursor process is the generation of initial population of individuals as a basis for future generations. As the encoding rule of the equation is special, in order to make the decoded individual form the equation, there are several rules when building the initial individuals.

The first two elements of the individual must be constant or variable terms; the last one must be an operator.

Given one location in the individual the number of nonoperators (constant or variable terms) on the left must be greater than the number of operators.

The total number of nonoperators in the individual must be the total number of operators plus one.

*Step 3*. Calculate the fitness value. The fitness value of the  $i$ -individual is computed as

$$\text{Fitness}(i) = 1 - \frac{\sum_{t=m-\tau+1}^Q [x(t) - T_i(x(t-\tau), x(t-2\cdot\tau), \dots, x(t-(m-1)\cdot\tau))]^2}{\sum_{t=m-\tau+1}^Q [x(t) - 1/(N-m\cdot\tau) \cdot \sum_{t=m-\tau+1}^Q x(t)]^2}, \quad (4)$$

where  $Q$  is the total length of the training set.

*Step 4*. Make selection, crossover, and mutation genetic operation. The selection operator chooses individuals that are used for crossover and mutation based on the fitness of the individuals. Proportional selection operator is used to select the individuals, and the probability of selecting  $i$ -individual is proportional to its fitness value. The probability is computed as

$$\text{prob}(i) = \frac{\text{fitness}(i)}{\sum_{k=1}^Q \text{fitness}(k)}. \quad (5)$$

Once the mates are selected according to their strength, the crossover operation between the two parent individuals is carried out to generate two new offspring. The crossover procedure starts determining randomly one of the arguments, constant terms, or variables terms of the time series in the first individual. A mutation operation is taken to yield solutions with new information. In order to preserve the information of the top ranked individual, the mutation is not applied to the best individual with the minimal fitness value of the current iteration. Each element of a determined string is changed by a mutation process with some probability. An element is randomly selected and, depending on whether it is a number, a variable, or an operator, if the element is an operator, then the element is self-mutated; if the element is a number or variable, then the element is self-mutated or mutual mutated.

*Step 5*. Determine the best individual of the current iteration and check whether the stopping criterion is met; if the criterion is met then decode the optimum individual and obtain the final equation expression; otherwise, continue the iteration processes until the criterion is met or the maximum iteration number is achieved.

### 3. Results and Analysis

In the numerical experiment 365 data samples of the whole year 2007 are selected and the front 280 samples are used to establish the forecasting model, the left 85 samples are used to test and verify. The distribution of forecasting errors is shown in Figure 2, where Figures 2(a), 2(b), 2(c), and 2(d) denote that the forecast span is one day ahead, two days ahead, four days ahead, and seven days ahead, respectively. The detailed forecasting precisions of different time scales are listed in Table 2, where  $\Delta$  is the error of the real and the predicted value; the percentages of the values of  $\Delta$  belonging to different intervals ( $\Delta < 1$ ,  $1 \leq \Delta < 2$ ,  $2 \leq \Delta < 3$ , and  $\Delta \geq 3$ ) are also listed in the table. It can be seen from the results that the forecasting performance is good overall, and the forecasting average error of one day ahead is the lowest; with the forecasting span increasing the errors increase overall. Besides, it can be also found from Figure 2 that the areas with large forecasting errors are concentrated in the equatorial region, and the time is concentrated from 10LT to 17LT; the main reason is that the asymmetry structures of ionospheric always exist in the equatorial region and the

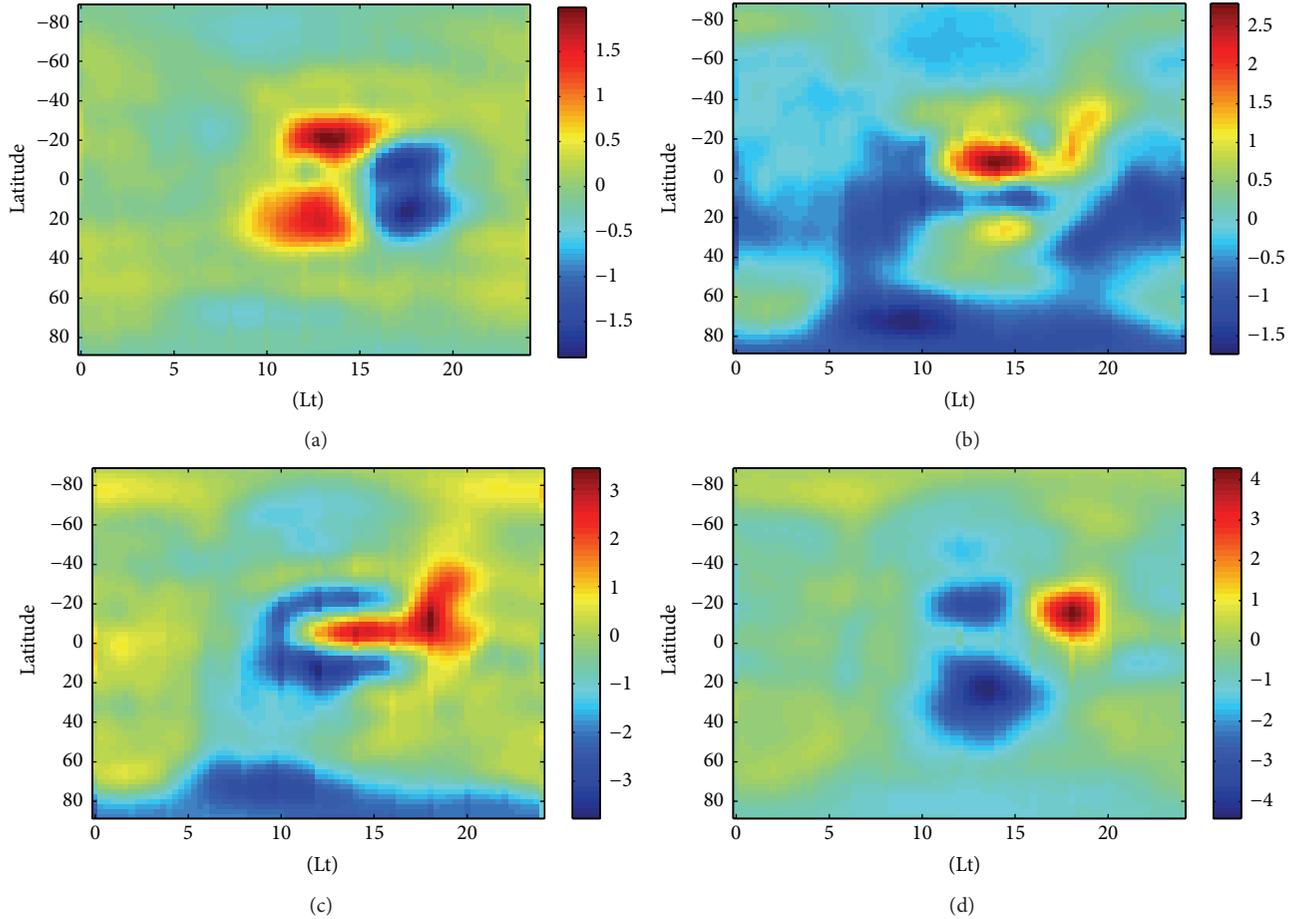


FIGURE 2: Error distributions by nonlinear forecasting based on genetic algorithm.

TABLE 2: Statistics of forecasting precisions of different time scales by nonlinear forecasting.

Forecast span ( $d$ )	Averaged error $\Delta$ (TECU)	$\Delta < 1$	$1 \leq \Delta < 2$	$2 \leq \Delta < 3$	$\Delta \geq 3$
1	0.2175	96.48%	3.52%	0.00%	0.00%
2	0.4268	92.14%	7.45%	0.41%	0.00%
3	0.5137	85.95%	11.52%	2.32%	0.21%
4	0.6824	74.36%	18.69%	6.64%	0.31%
5	0.7032	73.65%	19.30%	6.60%	0.45%
6	0.7703	71.25%	21.05%	6.27%	1.43%
7	0.6887	80.71%	13.24%	4.15%	1.91%

TABLE 3: Statistics of forecasting precisions of different time scales by direct forecasting.

Forecast span ( $d$ )	Averaged error $\Delta$ (TECU)	$\Delta < 1$	$1 \leq \Delta < 2$	$2 \leq \Delta < 3$	$\Delta \geq 3$
1	0.2895	95.37%	3.94%	0.69%	0.00%
2	0.5257	85.72%	13.81%	0.46%	0.00%
3	0.6246	76.77%	17.33%	5.05%	0.85%
4	0.9561	62.43%	25.56%	9.80%	2.20%
5	1.0849	56.38%	27.76%	12.98%	2.87%
6	1.2431	44.59%	41.62%	7.41%	6.39%
7	1.2411	48.77%	37.87%	6.69%	6.66%

maximum density of ionospheric mainly appears at 10LT to 17LT.

In order to indicate the advantage of the nonlinear forecasting model based on genetic algorithm, the traditional forecasting models such as direct forecasting and linear forecasting are also used to forecast under the same condition. The detailed results and analysis are also elaborated in the following.

The forecasting model of the direct forecasting is ARIMA (Auto Regressive Integrated Moving Average) model [8], and it is adequate for nonstationary time series. The distribution of forecasting errors is shown in Figure 3, where Figures 3(a), 3(b), 3(c), and 3(d) denote that the forecast span is one day ahead, two days ahead, four days ahead, and seven days ahead, respectively. The detailed forecasting precisions of different time scales are listed in Table 3. Compared with the above proposed nonlinear forecasting model it can be

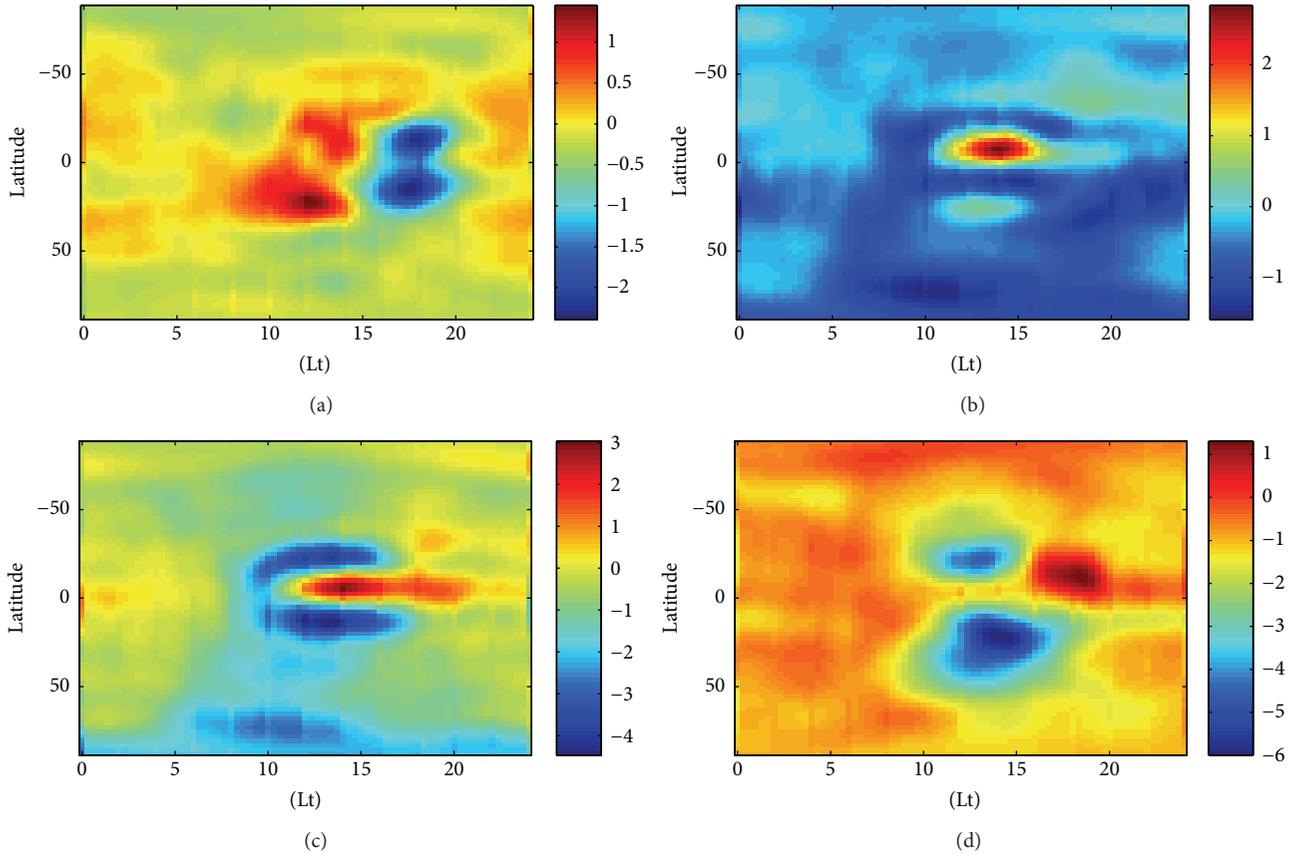


FIGURE 3: Error distributions by direct forecasting.

found that the forecasting errors are much larger and the forecasting performance of long forecasting span is obviously poor. The main reason is that the grid TEC data are simply taken as independent time series and the inner correlations between different grid TEC data are not considered.

The linear time series forecasting model is also ARIMA model; the coefficients of ARIMA model are evaluated by least square method. The distribution of forecasting errors is shown in Figure 4, where Figures 4(a), 4(b), 4(c), and 4(d) denote that the forecast span is one day ahead, two days ahead, four days ahead, and seven days ahead, respectively. The detailed forecasting precisions of different time scales are listed in Table 4. It can be seen from the results that the forecasting performance is better than the direct forecasting model, and the main reason is that the EOF decomposition and reconstruction are used in the linear forecasting model, and then the inner correlations between different grid TEC data are considered. However, compared with the nonlinear forecasting model the performance is worse; the main reason is that the linear ARIMA model is used to express time field in the linear forecasting model and the nonlinear model which is established by genetic algorithm is used to express time field; in fact, the distribution characteristics of time field are nonlinear and chaotic, so the nonlinear model estimated by genetic algorithm is better than the linear ARIMA model.

TABLE 4: Statistics of forecasting precisions of different time scales by linear forecasting.

Forecast span ( $d$ )	Averaged error $\Delta$ (TECU)	$\Delta < 1$	$1 \leq \Delta < 2$	$2 \leq \Delta < 3$	$\Delta \geq 3$
1	0.2606	93.92%	6.08%	0.00%	0.00%
2	0.4770	86.98%	12.33%	0.69%	0.00%
3	0.5721	83.43%	11.60%	4.57%	0.41%
4	0.7667	71.04%	18.60%	9.78%	0.58%
5	0.7514	71.37%	19.02%	9.32%	0.29%
6	0.8387	66.66%	24.73%	6.10%	2.51%
7	0.7692	75.19%	17.73%	4.32%	2.76%

#### 4. Conclusion

In the paper, a novel forecasting model is proposed. There are three main features in the model: first the Earth-fixed coordinate is converted to sun-fixed coordinate to reflect the physical features of TEC data distribution, then the data on the sun-fixed coordinate are decomposed by the method of EOF to reflect inner correlations between different grid TEC data, lastly the time field is taken as a series of nonlinear time series, and then the genetic algorithm is used to establish

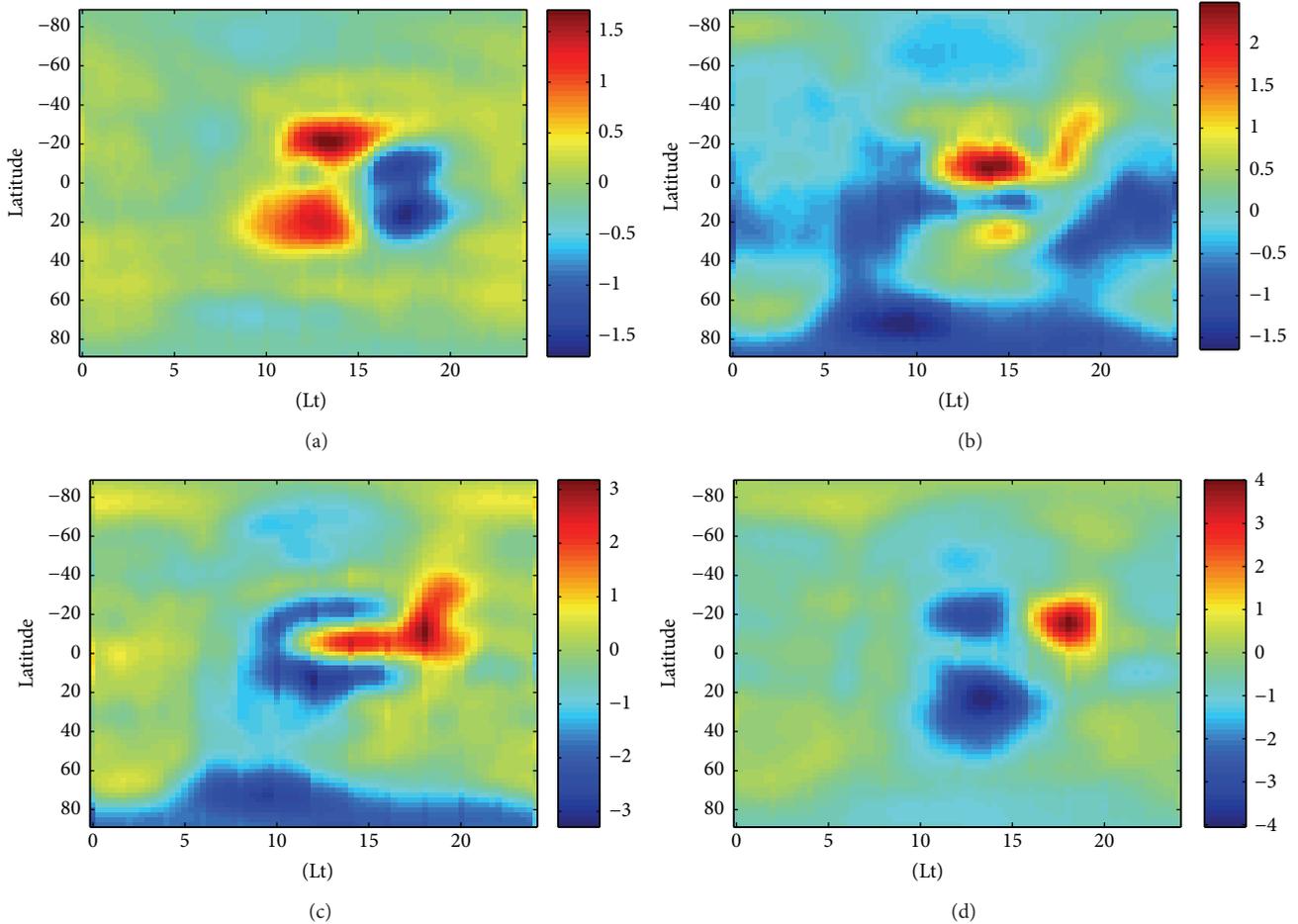


FIGURE 4: Error distributions by linear forecasting.

the nonlinear equation to forecast the time field. The results demonstrate the effectiveness performance of the model. In order to compare the model with other existing forecasting models and demonstrate the superior performance of the novel model, the traditional forecasting models such as direct forecast model and linear forecasting model are also tested; the results indicate that the errors of the novel model are the smallest, and the errors of the direct model are the largest. So it is proved that the thought of using nonlinear chaotic time series forecasting combined with genetic algorithm and EOF decomposition to forecast TEC data is feasible, energy-efficient, and promising.

### Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

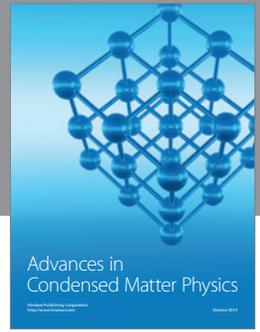
### Acknowledgments

This work is supported by National Natural Science Foundation of China for Young Scholars under Grants nos. 41305017 and 41105013 and National Natural Science Foundation of China under Grant 61271106.

### References

- [1] D. Bilitza, "International reference ionosphere," *Radio Science*, vol. 36, no. 2, pp. 261–275, 2000.
- [2] Z. Huang and H. Yuan, "Ionospheric refraction correction based on spatial statistic for China region," *Chinese Journal of Space Science*, vol. 32, no. 2, pp. 209–215, 2012.
- [3] J. B. Habarulema, L.-A. McKinnell, P. J. Cilliers, and B. D. L. Opperman, "Application of neural networks to South African GPS TEC modelling," *Advances in Space Research*, vol. 43, no. 11, pp. 1711–1720, 2009.
- [4] L. B. Weng, H. X. Fang, Z. Q. Miao, and S. G. Yang, "Forecasting of ionospheric TEC one hour in advance by artificial neural network," *Chinese Journal of Space Science*, vol. 32, no. 2, pp. 204–208, 2012.
- [5] Z. X. Deng, R. Y. Liu, W. M. Zhen, D. Liu, and Y. W. Wu, "Study on the ionospheric TEC storms over China," *Chinese Journal of Geophysics-Chinese Edition*, vol. 55, no. 7, pp. 2177–2184, 2012.
- [6] R. F. Leandro and M. C. Santos, "A neural network approach for regional vertical total electron content modelling," *Studia Geophysica et Geodaetica*, vol. 51, no. 2, pp. 279–292, 2007.
- [7] Z. Huang and H. Yuan, "Ionospheric single-station TEC short-term forecast using RBF neural network," *Radio Science*, vol. 49, no. 4, pp. 283–292, 2014.

- [8] D. Sur and A. Paul, "Comparison of standard TEC models with a Neural Network based TEC model using multistation GPS TEC around the northern crest of Equatorial Ionization Anomaly in the Indian longitude sector during the low and moderate solar activity levels of the 24th solar cycle," *Advances in Space Research*, vol. 52, no. 5, pp. 810–820, 2013.
- [9] C. Manoj and S. Maus, "A real-time forecast service for the ionospheric equatorial zonal electric field," *Space Weather*, vol. 10, no. 9, pp. 18–22, 2012.
- [10] J. Wang, Z. Sheng, Y. Jiang, and H. Q. Shi, "New reconstruction and forecasting algorithm for TEC data," *Chinese Physics B*, vol. 23, no. 9, Article ID 099401, 2014.
- [11] A. Alvarez, "Performance of satellite-based ocean forecasting (SOFT) systems: a study in the Adriatic Sea," *Journal of Atmospheric and Oceanic Technology*, vol. 20, no. 5, pp. 717–729, 2003.
- [12] J. M. Beckers and M. Rixen, "EOF calculations and data filling from incomplete oceanographic datasets," *Journal of Atmospheric and Oceanic Technology*, vol. 20, no. 12, pp. 1839–1856, 2003.
- [13] J. Wang, Z. Sheng, B. H. Zhou, and S. D. Zhou, "Lightning potential forecast over Nanjing with denoised sounding-derived indices based on SSA and CS-BP neural network," *Atmospheric Research*, vol. 137, pp. 245–256, 2014.
- [14] J. Wang, B.-H. Zhou, S.-D. Zhou, and Z. Sheng, "Forecasting nonlinear chaotic time series with function expression method based on an improved genetic-simulated annealing algorithm," *Computational Intelligence and Neuroscience*, vol. 2015, Article ID 341031, 10 pages, 2015.
- [15] J.-H. Ma, Y.-S. Chen, and B.-G. Xin, "Study on prediction methods for dynamic systems of nonlinear chaotic time series," *Applied Mathematics and Mechanics*, vol. 25, no. 6, pp. 605–611, 2004.
- [16] G. Volpe, B. B. Nardelli, P. Cipollini, R. Santoleri, and I. S. Robinson, "Seasonal to interannual phytoplankton response to physical processes in the Mediterranean Sea from satellite observations," *Remote Sensing of Environment*, vol. 117, pp. 223–235, 2012.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

