

# A Prediction Model for Regional Carbon Emissions Based on GRU Networks

Jun Meng (✉ [84697687@qq.com](mailto:84697687@qq.com))

Beijing Institute of Technology <https://orcid.org/0000-0001-7005-4512>

Gangyi Ding

Beijing Institute of Technology

Laiyang Liu

Beijing Institute of Technology

Zheng Guan

Beijing Institute of Technology

---

## Research Article

**Keywords:** Carbon Emissions, Wireless Carbon Sensor Network, Prediction Model, GRU Networks

**Posted Date:** August 6th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-767805/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# A Prediction Model for Regional Carbon Emissions Based on GRU Networks

Jun Meng, Gangyi Ding, Laiyang Liu, Zheng Guan

(School of Computer Science & Technology, Beijing Institute of Technology, Beijing 100081, China)

**Abstract:** In this study, a data-driven regional carbon emissions prediction model is proposed. The Grubbs criterion is used to eliminate the gross error data in carbon emissions sensor data. Then, according to the nearby valid data, the exponential smoothing method is used to interpolate the missing values to generate the continuous sequence for model training. Finally, the GRU network, which is a deep learning method, is used to process these sequential standardized data to obtain the prediction model. In this paper, the wireless carbon sensor network monitoring data set from August 2012 to April 2014 trained and evaluated the prediction model, and compared with the prediction model based on BP network. The experimental results prove the feasibility of the research method and related technical approaches, and the accuracy of the prediction model, which provides a method basis for the nowcasting of carbon emissions and other greenhouse gas environmental data.

**Keywords:** Carbon Emissions; Wireless Carbon Sensor Network; Prediction Model; GRU Networks

## 1. Introduction

Global climate change affects the survival and development of mankind, and brings serious challenges to the sustainable development of the economy and society. In this context, the "low-carbon economy" based on low energy consumption and low pollution has become a global hot spot. Carbon dioxide (CO<sub>2</sub>) is one of the main ingredients of the atmosphere, it through to the absorption of long wave radiation and stop the surface heat dissipation, also is one of the main greenhouse gases. Therefore, monitoring carbon emissions and designing accurate prediction methods based on data-driven thinking will be of great significance to the development of a low-carbon economy and the analysis of future climate change trends.

At present, the conventional carbon emissions accounting methods adopted in the world mainly include IPCC inventory method (Martin et al. 2006), measured method, material balance algorithm and model decomposition method (Hao et al. 2011). These methods have been applied to varying degrees in practical studies (Wang et al. 2012; Cheng et al. 2013; Wang et al. 2015; Zhang et al. 2016). But the calculation results of these methods will be affected by such as the production process, residents' lifestyle (Wang et al. 2016), calorific value, carbon content, carbon oxidation rate (Sun et al. 2015), calculation methods (Wang et al. 2015) and other factors, resulting in greater

uncertainty. It can be seen that there is still a widespread problem of weak data statistics and prediction methods in the existing field of carbon emissions, and it is urgent to establish scientific, transparent, accurate and reliable data statistics and prediction methods that meet international standards.

In recent years, the maturity of sensor network technology provides convenience for real-time and continuous acquisition of research data (Wang 2021), and also provides a strong technical support for data-driven carbon emissions environmental data prediction methods. At the same time, deep learning has made significant achievements in the field of artificial intelligence. This method can improve the accuracy of classification and prediction by training big data and by mining and capturing deep connections among big data. Deep learning is an effective big data processing method. The typical deep learning model, the convolutional neural network (CNN), has been widely used, but CNNs are not completely suitable for learning time series. To adapt to the processing of time series data, a recurrent neural network (RNN) emerged (Huang et al. 2019). An ordinary RNN is equivalent to a multi-layer deep neural network (DNN) expanded on a time series (Fan et al. 2017). This model is prone to gradient explosion or gradient disappearance. The long short-term memory (LSTM) neural network was first proposed by Hochreiter & Schmidhuber in 1997, and it has been improved and popularized by many experts and scholars. Now, it is widely used due to its excellent performance (Duan et al. 2019). The Gated Recurrent Unit (GRU) network is proposed by Cho, et al. (2014) and can be regarded as a variation of LSTM. However, it has a simpler structure and fewer parameters than LSTM, so it is slightly faster to train or requires less data for generalization and guarantees excellent performance (Chen et al. 2021). This makes GRU more suitable for dynamic process modeling (Wang et al. 2020). The variation trend of CO<sub>2</sub> concentration is related to environmental factors such as temperature and air humidity. These data exhibit a natural continuity in time and achieve strong correlation and causality before and after the time series. Using the GRU network to achieve CO<sub>2</sub> concentration prediction can not only use the correlation of the data in the time dimension but also automatically mine the potential correlations between the data and improve the accuracy of carbon emissions data prediction.

In this study, a data-driven regional carbon emissions prediction model is proposed. This model is based on the comprehensive measurement and perception data of complex environmental systems, connecting with relevant research results in the field of deep learning, to collect information and data on various aspects of the natural environment, industrial production and social life to form a regional meso-social simulation Model, and then design a regional carbon emissions environmental data prediction method.

## 2. Proposed method

### 2.1 Model framework

Based on the above related theories and algorithms, a prediction model for regional carbon emissions is established.

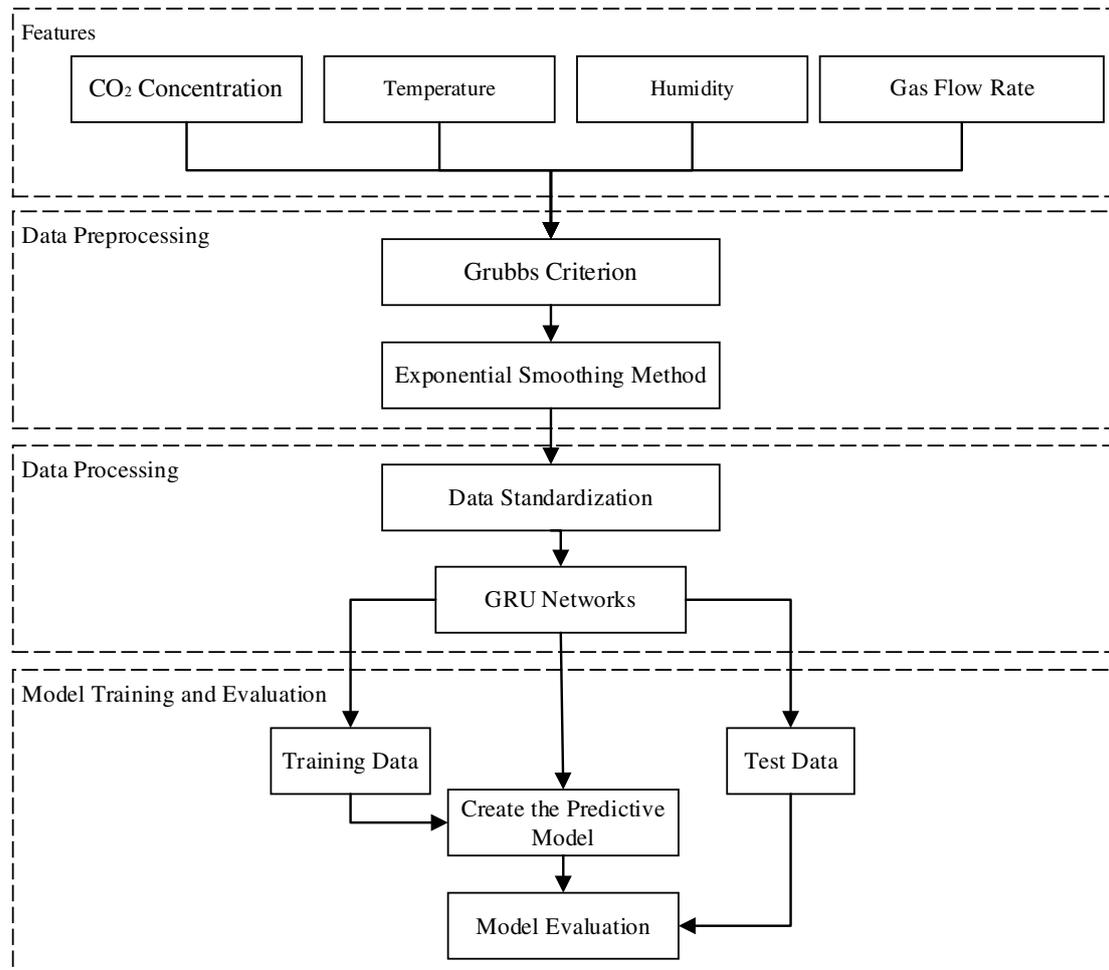


Figure 1. Flowchart of data processing algorithm

The algorithmic flow of the model is shown in figure 1 and the details of the algorithm involved will be introduced.

### 2.2 Eliminate gross error data

In the data preprocessing stage, the feature data are processed according to Grubbs criterion: set the sample as  $s_i$  ( $i=1,2,3,\dots,n$ ),  $n$  is the total number of samples. Then for the  $i$ -th sample value:

- 1) Arrange  $s_i$  in ascending order.
- 2) Calculate the mean  $\bar{s}$  and variance  $\sigma$ .

$$\bar{s} = \frac{\sum_{i=1}^n s_i}{n} \quad (1)$$

$$\sigma = \sqrt{\frac{\sum (s_i - \bar{s})^2}{n - 1}} \quad (2)$$

3) Calculate the lower Grubbs number  $g_{(1)}$  and the upper Grubbs number  $g_{(n)}$ .

$$g_{(1)} = \frac{\bar{s} - s_{(1)}}{\sigma} \quad (3)$$

$$g_{(n)} = \frac{s_{(n)} - \bar{s}}{\sigma} \quad (4)$$

4) Select the larger residual error of the two and select the significance  $\alpha$  to obtain the critical value  $g_0(n, \alpha)$  through the critical value test table shown in table 1. If  $g(i) \geq g_0(n, \alpha)$ , there is gross error in the measured value, which should be eliminated.

Table 1. Grubbs critical value test table

$n \backslash \alpha$	0.1	0.05	0.025	0.01	0.005
3	1.148	1.153	1.155	1.155	1.155
4	1.425	1.463	1.481	1.492	1.496
5	1.602	1.672	1.715	1.749	1.764
6	1.729	1.822	1.887	1.944	1.973
7	1.828	1.938	2.02	2.097	2.139
8	1.909	2.032	2.126	2.221	2.274
9	1.977	2.11	2.215	2.323	2.387
10	2.036	2.176	2.29	2.41	2.482
20	2.385	2.557	2.709	2.884	3.001
30	2.563	2.745	2.908	3.103	3.236
40	2.682	2.866	3.036	3.24	3.381
50	2.768	2.956	3.128	3.336	3.483
60	2.837	3.025	3.199	3.411	3.56
70	2.893	3.082	3.257	3.471	3.622
80	2.94	3.13	3.305	3.521	3.673
90	2.981	3.171	3.347	3.563	3.716
100	3.017	3.207	3.383	3.6	3.754

### 2.3 Data sequence interpolation

The cubic exponential smoothing method is used to obtain missing data values within a time series. Firstly, the length of the missing data sequence is determined, and then data points and smoothing steps are inserted according to the time series value of the previous section of CO<sub>2</sub> concentration of the missing data. Related processing formula is as follows:

$$s_{i+1}^{(1)} = \alpha x_i + (1 - \alpha)s_i^{(1)} \quad (5)$$

$$s_{i+1}^{(2)} = \beta s_i^{(1)} + (1 - \beta)s_i^{(2)} \quad (6)$$

$$s_{i+1}^{(3)} = \gamma s_i^{(2)} + (1 - \gamma)s_i^{(3)} \quad (7)$$

$$\begin{cases} a_i = 3s_i^{(1)} - 3s_i^{(2)} + s_i^{(3)} \\ b_i = \frac{\alpha}{2(1-\alpha)^2} [(6-5\alpha)s_i^{(1)} - (10-8\alpha)s_i^{(2)} + (4-3\alpha)s_i^{(3)}] \\ c_i = \frac{\alpha^2}{2(1-\alpha)^2} (s_i^{(1)} - 2s_i^{(2)} + s_i^{(3)}) \end{cases} \quad (8)$$

$$I_{i+m} = a_i + b_i m + c_i m^2 \quad (9)$$

$\alpha$ ,  $\beta$  and  $\gamma$  are all smoothing constants, the value range is  $[0,1]$ , and its value is selected subjectively, and  $I_{i+m}$  is the predicted value of period  $i+m$ , that is, the missing value of the CO<sub>2</sub> concentration.

## 2.4 Model training method

Since different features have different dimensions and units, in order to reduce the dimensional influence between the features, the data needs to be standardized:

$$s'_i = \frac{s_i - \min(s)}{\max(s) - \min(s)} \quad (10)$$

GRU network is an effective variant of the LSTM network. The LSTM network uses three gate functions: input gate, forgetting gate and output gate to control the input value, memory value and output value. In a GRU network, there are only two gates: the update gate and the reset gate. The structure is shown in figure 2.

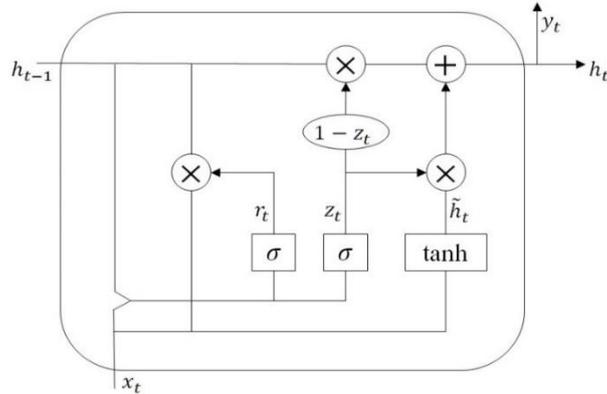


Figure 2. Gated recurrent unit

In the figure,  $z_t$  and  $r_t$  represent update gate and reset gate respectively, and their forward propagation formula is:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (11)$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (12)$$

$$\tilde{h}_t = \tanh(W_{\tilde{h}} \cdot [r_t * h_{t-1}, x_t]) \quad (13)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (14)$$

$$y_t = \sigma(W_o \cdot h_t) \quad (15)$$

It can be seen from the formula in the previous propagation process that the parameters to be learned are  $W_r$ ,  $W_z$ ,  $W_{\tilde{h}}$  and  $W_o$ . The first three parameters are all spliced, so they need to be divided during the training:

$$W_r = W_{rx} + W_{rh} \quad (16)$$

$$W_z = W_{zx} + W_{zh} \quad (17)$$

$$W_{\tilde{h}} = W_{\tilde{h}x} + W_{\tilde{h}h} \quad (18)$$

The input of output layer:

$$y_t^i = W_o h \quad (19)$$

The output of output layer:

$$y_t^o = \sigma(y_t^i) \quad (20)$$

The loss of a single sample at a certain moment is:

$$E_t = \frac{1}{2} (y_d - y_t^o)^2 \quad (21)$$

The loss of a single sample at all times is:

$$E = \sum_{t=1}^T E_t \quad (22)$$

To learn the network by using the backward error propagation algorithm, the partial derivative of the loss function with respect to each parameter must be obtained first:

$$\frac{\partial E}{\partial W_o} = \delta_{y,t} h_t \quad (23)$$

$$\frac{\partial E}{\partial W_{zx}} = \delta_{z,t} x_t \quad (24)$$

$$\frac{\partial E}{\partial W_{zh}} = \delta_{z,t} h_{t-1} \quad (25)$$

$$\frac{\partial E}{\partial W_{\tilde{h}x}} = \delta_t x_t \quad (26)$$

$$\frac{\partial E}{\partial W_{\tilde{h}h}} = \delta_t (r_t \cdot h_{t-1}) \quad (27)$$

$$\frac{\partial E}{\partial W_{rx}} = \delta_{r,t} x_t \quad (28)$$

$$\frac{\partial E}{\partial W_{rh}} = \delta_{r,t} h_{t-1} \quad (29)$$

Among them, the intermediate parameters are:

$$\delta_{y,t} = (y_d - y_t^o) \cdot \sigma' \quad (30)$$

$$\delta_{h,t} = \delta_{y,t} W_o + \delta_{z,t+1} W_{zh} + \delta_{t+1} W_{\tilde{h}h} \cdot r_{t+1} + \delta_{h,t+1} W_{rh} + \delta_{h,t+1} \cdot (1 - z_{t+1}) \quad (31)$$

$$\delta_{z,t} = \delta_{t,h} \cdot (\tilde{h}_t - h_{t-1}) \cdot \sigma' \quad (32)$$

$$\delta_t = \delta_{h,t} \cdot z_t \cdot \phi' \quad (33)$$

$$\delta_{r,t} = h_{t-1} \cdot [(\delta_{h,t} \cdot z_t \cdot \phi') W_{\tilde{h}h}] \cdot \sigma' \quad (34)$$

After calculating the partial derivatives for each parameter, the parameters can be updated, and iteratively until the loss converges. The training process of the prediction model is shown in table 2.

Table 2. Model training procedure.

---

**Input:** A set of features  $X = \{ X_{co2}, X_t, X_h, X_f \}$ ,  $input\_length = 360$ ,  $input\_dim = 4$ ,  $learning\ rate = 0.01$ , the number of hidden layers is 2, the number of hidden units is 20 and the temporal granularity is day.

**Output:** Learned prediction model.

---

1. The array constructor is used to convert the dataset  $X$  and get the matrix  $X_s$ .
2. Data normalization.
3. Divide and create data sets, 67% for the train dataset, and 33% for the test dataset.
4.  $look\_back = 3$
5. Reshape input to be  $[samples, time\_steps=10, features]$
6. Create and fit the GRU networks.
7.  $loss = 'mean\_squared\_error'$
8. The Adam optimizer is used,
9.  $epochs = 10$
10.  $batch\_size = 50$
11. Make predictions
12. Reverse normalization the prediction results.
13. Calculate the mean squared error of the train dataset and test dataset.

**Return**

---

### 3. Experiment and model evaluation

Based on the requirements of the project, the research group designed a wireless carbon emissions monitoring sensor device. It mainly integrates ARM control module,

carbon dioxide data acquisition module, temperature data acquisition module, humidity data acquisition module, gas flow rate acquisition module, GPS positioning module and GPRS data transmission modules. In addition, considering its working environment, the device is designed to be a low power consumption, portable, professional embedded system with industrial standards, with good stability and reliability. The prototype equipment is shown in figure 3:

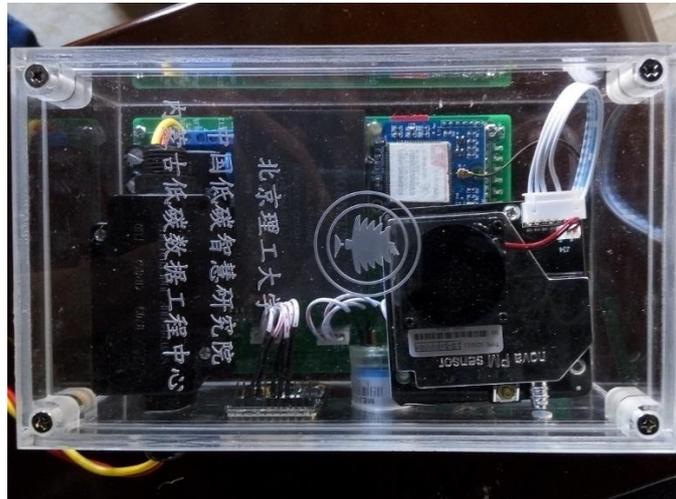


Figure 3. Wireless carbon emissions monitoring sensor device

The experiment selects the CO<sub>2</sub> concentration, temperature, humidity and other data obtained by the wireless carbon sensor network from August 2012 to April 2014. These data are collected by the above-mentioned sensors, and according to the set time period, the data is regularly transmitted to the server of the carbon data processing center through the 4G network. The processing center stores these data into the database and processes them to produce relevant data products. The data collection and processing process is shown in figure 4.

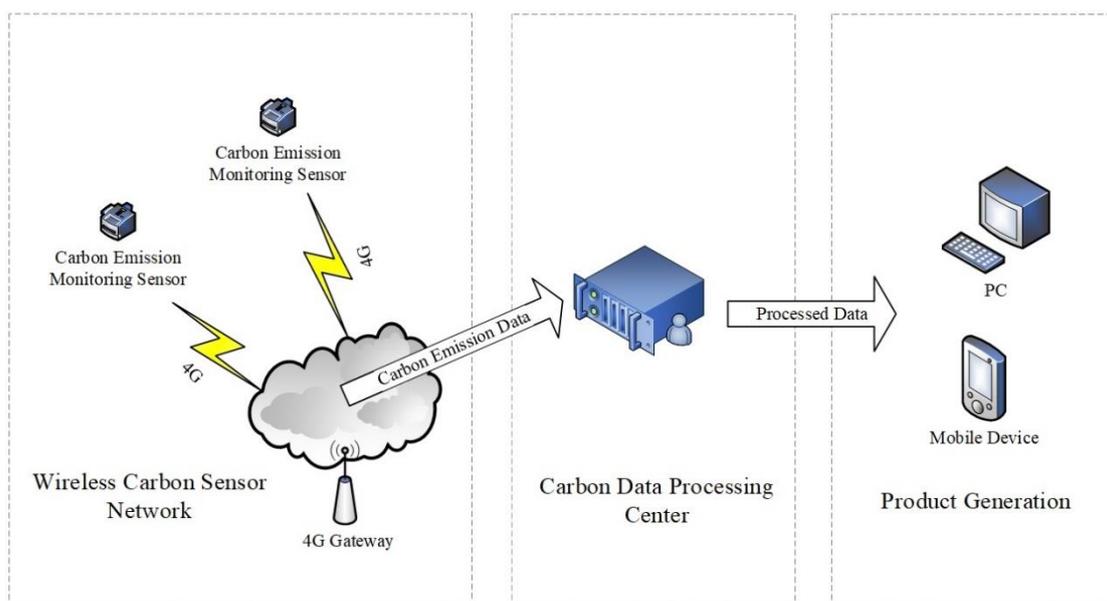


Figure 4. Data collection and processing process

Since August 2010, the research group has set up 14 environmental monitoring sites in Genhe City, Hulunbeier City, Inner Mongolia Autonomous Region, and since April 2014, five monitoring sites have been set up in the new urban area of Huhahaote City, Inner Mongolia Autonomous Region. This number will increase in the future in order to obtain more complete data sets. The monitoring system interface and data of some sites are shown in figure 5.

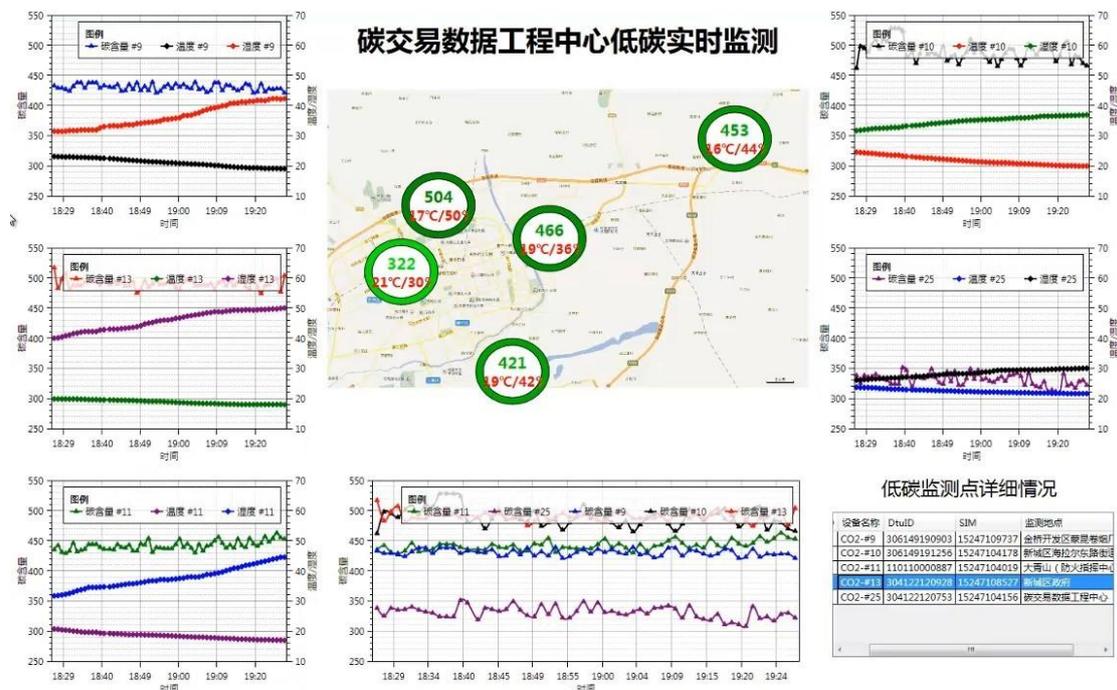


Figure 5. Carbon emissions monitoring system

For the prediction model, the smaller the error of the result, the higher the prediction accuracy. For the same prediction model, the setting of parameter value is the most influential factor. In order to obtain higher accuracy, the parameters need to be adjusted repeatedly to get a satisfactory model. As can be seen from the partial test results in table 3, with the increase of lookback value, the prediction ability of the model is constantly improved, and the prediction effect is the best when its value is 3. However, as the value continues to increase, the accuracy of the model decreases. This reflects that when the lookback value is set too small, the model fitting effect will be poor due to insufficient correlation information between data, while when the lookback value is set too big, the correlation between data will be decreased, which leads to the accuracy of the model and the generalization ability of the model. In addition, theoretically the larger the training set size, the better the learning ability, but if the training set size is too large, the training will also be more time-consuming. For this model, when the size of the training set is 800, its accuracy can be guaranteed, and if the training set size continues to increase, the accuracy does not improve significantly, which indicates that the prediction accuracy becomes insensitive to the increase of the sample number of the training set.

Table 3. Comparison of some experimental results caused by different parameters

Lookback	MAE	RMSE	Training set size	MAE	RMSE
----------	-----	------	-------------------	-----	------

1	0.287543	0.288654	100	0.124598	0.126873
2	0.116542	0.126878	200	0.105687	0.108429
3	0.006852	0.012927	300	0.069654	0.071542
4	0.235643	0.237835	400	0.045641	0.049861
5	0.235644	0.237836	500	0.006487	0.016752
6	0.235646	0.237836	600	0.013466	0.018657
7	0.235647	0.237839	700	0.015471	0.017951
8	0.235649	0.237841	800	0.009745	0.016263
9	0.235651	0.237841	900	0.009612	0.015872
10	0.235655	0.237843	1000	0.008249	0.015269

In order to prove the performance of the prediction model based on the GRU network, it is not only necessary to adjust the parameters to compare with itself, but also to compare with different models. The model is compared with the model based on Back Propagation (BP) network on the same training set, and the results are shown in the figure 6 and table 4.

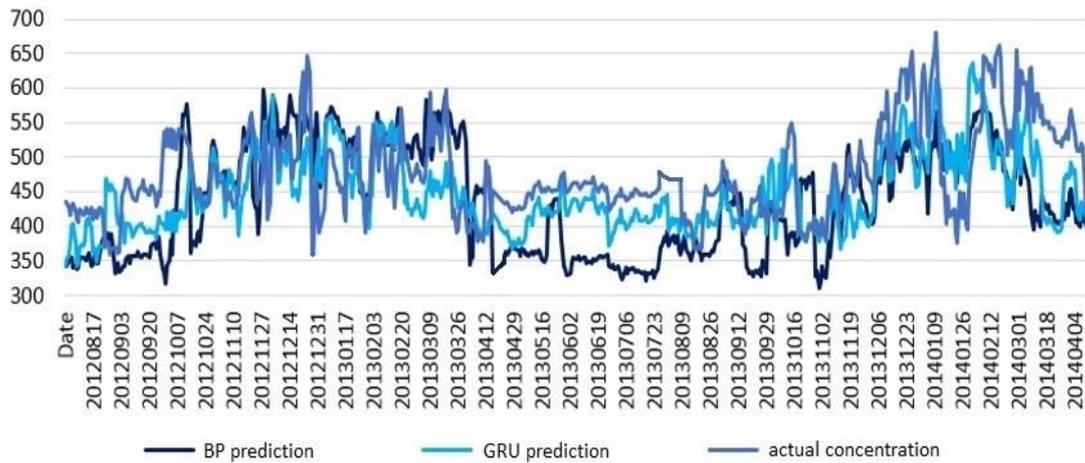


Figure 6. Comparison chart of predicted CO<sub>2</sub> concentration

Table 4. Prediction error comparison

Prediction Model	MAE	RMSE
BP	31.65741	52.14798
GRU	18.65874	21.57742

It can be seen from the above chart and table that this prediction model is more accurate than BP model. This shows that the model can extract deep features from high-dimensional data through various nonlinear operations, and give full play to the unique advantages of time series data, so that the prediction model shows better prediction and fitting effect.

## 4. Conclusion

From the experimental results, it can be seen that the prediction model based on the GRU network has high accuracy and can well fit the variation trend of CO<sub>2</sub> concentration. This also shows that the model gives full play to the unique advantages of deep learning of the GRU network for time series data interpolation. It can extract abstract and deep features from high-dimensional data through a variety of nonlinear operations, making the prediction model perform better prediction and fitting effect. It also proves the feasibility of this research method and related technical approaches, which provides a method basis for the nowcasting of carbon emissions and other greenhouse gas environmental data.

**Author Contributions** Conception and design of study: J. Meng, G.Y. Ding, L.Y. Liu, Z. Guan; acquisition of data: J. Meng, L.Y. Liu; analysis and interpretation of data: J. Meng, G.Y. Ding; drafting the manuscript: J. Meng, Z. Guan.

## Declarations

**Ethical approval** This article does not contain any studies with human participants or animals performed by any of the authors.

**Conflict of interest** The authors declare that they have no conflict of interest.

**Informed consent** Informed consent was obtained from all individual participants included in the study. This statement is to certify that all authors have seen and approved the manuscript being submitted in Soft Computing journal.

## References

1. Martin RM, Dai XS (2006) The Treatment of Uncertainties in the Fourth IPCC Assessment Report. *Advances in climate change research* 2006(05):233-237.
2. Hao QT, Huang MX, Bao G (2011) Study on Carbon Emission Calculation Methods Overview and Its Comparison. *Chinese Journal of Environmental Management* 2011(4):51-55.
3. Huang MX, Wei B, Hao QT, Li S, Zhang FH (2015) A review on research of PM<sub>2.5</sub> retrieval by remote sensing technology. *Environmental Pollution & Control* 37(10):70-76+85.
4. Wang JM, Zhu YP (2012) Research on Carbon Emissions and Its Trend Forecast of Hebei Province, *Journal of Anhui Agri. Sci.* 40(10): 6086-6088+6092.
5. Wang JM, Liu XD (2012) Study on agricultural carbon emissions statistics and monitoring system. *Guangdong Agricultural Sciences* 39(14):233-236.
6. Wang JM, Zhu YP (2012) Some Improvements for DSM Cost-effectiveness Mode in the Environment of Carbon Trading Market. *Journal of Shandong Electric Power College* 15(05):73-77.

7. Wang JM, Xue YT (2013) Efficiency of Power Industry Carbon Emission Quotas Based on Data Envelopment. *Electric Power* 46(10):146-150.
8. Wang JM, Wang C (2013) Development Planning of Nuclear Power Generation in Eastern China. *Electric Power* 46(12):154-159.
9. Cheng YQ, Wang ZY, Zhang SZ, Ye XY, Jiang HM (2013) Spatial econometric analysis of carbon emission intensity and its driving factors from energy consumption in China. *ACTA Geographica sinica* 68(10):1418-1431.
10. Wang YJ, He Y (2015) Methods for Low-Carbon City Planning Based on Carbon Emission Inventory. *CHINA POPULATION, RESOURCES AND ENVIRONMENT* 25(6):72-80.
11. Wei HJ, Zhang YF, Dong XB, Hu XH, Wang XC, Lu NC, Zhao YN (2016) Changes of Vegetation Cover and Carbon Fixation in Wuding River Basin During 2000-2013. *Bulletin of Soil and Water Conservation* 36(1):44-50.
12. Wang JJ, Zhang KS (2016) Uncertainty Analysis of Carbon Emissions Resulting from Daily Activities of Residents of Chengdu. *Environmental Engineering* 34(S1):926-930.
13. Sun ZQ, Chen YN, Wang GJ (2015) Discussion on China's Carbon Market Construction Experience. *Environmental Protection* 43(06):42-44.
14. Sun ZQ, Wang GJ, Chen YN (2015) Analysis on the Uncertainty of Carbon Emission Accounting Based on Energy Balance Sheet. *Ecological Economy* 31(7):33-38.
15. Wang XH, Chen BZ, Zhang HF (2015) Analysis of the provincial carbon list uncertainty. *Journal of Arid Land Resources and Environment* 29(7):7-11.
16. Wang HB (2021) Research Progress of Low-power Carbon Monoxide sensors. *Industry and Mine Automation* 47(7):72-78.
17. Huang J, Zhang F, Du ZH, Liu RY, Cao XP (2019) Hourly concentration prediction of PM<sub>2.5</sub> based on RNN-CNN ensemble deep learning model. *Journal of Zhejiang University* 46(3):370-379.
18. Fan JX, Li Q, Zhu YJ, Hou JX, Feng X (2017) Aspatio-temporal Prediction Framework for Air Pollution Based on Deep RNN. *Science of Surveying and Mapping* 42(7):76-83.
19. Duan DG, Zhao ZD, Liang SH, Yang WJ, Han ZM (2019) Research on PM<sub>2.5</sub> Concentration Prediction Based on LSTM. *Computer Measurement & Control* 27(3):215-219.
20. Chen L, Chen XF (2021) Prediction method of toxic gas diffusion based on gated recycle unit (GRU). *Information Technology and Network Security* 40(04):42-45.
21. Wang W, Zhao W (2020) NO<sub>x</sub> Emission Prediction Model Based on GRU Neural Network in Coal-fired Power Station. *Journal of North China Electric Power University* 47(1):96-103.