Original Research

Influential Factor Analysis and Projection of Industrial CO₂ Emissions in China Based on Extreme Learning Machine Improved by Genetic Algorithm

Yanmei Li, Hongdan Hu*

North China Electric Power University, Baoding, China

Abstract

In China, CO_2 emissions from industrial sectors on a larger scale than other end-use sectors. In order to reduce CO_2 emissions, it is necessary to study the influencing factors and projection of industrial CO_2 emissions. Based on accounting for CO_2 emissions from the industrial sectors, this paper carries out bivariate correlation analysis and linear regression analysis on 15 preselected influencing factors and industrial CO_2 emissions, removing two factors that have failed the significance test. In order to obtain some potential commonalities among the influencing factors, the remaining 13 influencing factors are divided into four categories, and then factor analysis is performed on each category in order to obtain five latent factors. An extreme learning machine algorithm that uses genetic algorithms to optimize the input weights and bias thresholds – the genetic algorithm extreme learning machine (GA-ELM) algorithm – to predict industrial CO_2 emissions, the empirical results show that the GA-ELM algorithm using five factors as inputs has a higher prediction accuracy and performance for industrial CO_2 emissions than the extreme learning machine, back propagation neural network, and back propagation neural network optimized by the genetic algorithm. It also shows that the five influencing factors have a significant impact on industrial CO_2 emissions. Finally, based on the analysis of five influencing factors, some policy recommendations are proposed for the CO_2 emissions reduction path in the industrial sectors.

Keywords: industry, CO_2 emissions, genetic algorithm-extreme learning machine

Introduction

Global warming caused by massive greenhouse gas emissions and its consequences have been serious environmental issues for every country in the world.

China is the world's top CO_2 emitter [1-3], accounting for 30% of global emissions [4]. Consequently, China is playing an important role in global emissions reduction and climate change mitigation. The Chinese government has promised that its CO_2 emissions will achieve its maximum CO_2 emissions in 2030 [5] and that it will achieve a 40-45% reduction in its CO_2 intensity per GDP by 2020 compared with its 2005 levels [6, 7].

*e-mail: 569373581@qq.com

At the 2015 Paris Climate Conference [8], China continued to firmly reiterate its promised actions regarding climate change [9]. According to the fifth assessment report (AR5), climate change 2014 by IPCC, energy consumption in industrial sectors accounts for 19% of the terminal energy use, and the greenhouse gas contribution rate is 30% in 2010 in the world [10]. In China, the situation is even more serious [11], even if the proportion of total industrial energy consumption in China decreases year by year. As of 2014, the total industrial energy consumption still accounted for 69.4%. The current status of China's carbon flow indicated that 75.12% of total CO_2 emissions flow mainly into several sectors, such as ferrous sectors, chemical industry and heat and the power production sector. These sectors are all basically industrial sectors [12].

In this background, it is necessary to analyze the influencing factors of industrial CO_2 emissions and to forecast industrial CO_2 emissions and then accordingly propose feasible policy advice. Studies about industrial CO_2 emissions have been developed over many years, and these studies can be classified into two aspects. The first aspect of studies is the influencing factors of industrial CO_2 emissions. Osamu Akashi et al. developed a three-part simulation system to study industrial CO_2 emissions depending on changes in both technology and industrial activity [13]. Jinlong MA et al. used the adaptive weighting divisia index approach to examine contributions to the changes in CO_2 emissions from industrial energy consumption in China over the period 1980-1997, and concluded that the increases in industrial production and energy efficiency had great impact on CO_2 emissions, and fuel mix and fuel quality were found to make only a small impact on CO_2 emissions [14]. Through other researches, the factors that have an impact on carbon emissions also include industrial output [15, 16], population size, per capita industrial output value, industrial technology level [17], industrial structure and energy structure et al. [18-20]. The second aspect of these studies is the CO_2 emissions from industry-specific industries. Daniel Summerbell et al. studied CO_2 emissions from cement plants and found that there exists significant opportunity to reduce the emissions from cement plants by operational means [21]. Wenqiang Sun et al. applied the logarithmic mean divisia index (LMDI) technique in the CO_2 emissions analysis from the iron and steel industry in China, and put forward policy implications regarding the reduction of CO_2 emissions [22]. Since the power generation industry is the largest sector of industrial CO_2 emissions, there is more research on the CO_2 emissions of the power industry. M. Karmellos measured the CO_2 emissions of European Union countries' power industries and analyzed the influencing factors [23]. Ming Meng et al. applied the scenario analysis method to CO_2 emissions from China's electric power [24, 25]. CO_2 emissions from other sectors, including mining, manufacturing and so on, are all the study hotspots.

For predicting CO_2 emissions, there are many prediction models. Tudor [26] attempted to forecast the evolution of carbon dioxide emissions in Bahrain during 2012-2021 by employing seven automated forecasting methods, including the exponential smoothing state space model (ETS), the Holt-Winters model, the BATS/TBATS model, ARIMA, the structural time series model (STS), the naive model, and the neural network time series forecasting method (NNAR), which arrived at the conclusion that Bahrain cannot meet its assumed target. In addition to the above, the STRIPAT model [27] and partial least squares [28-30] are also often used for carbon emissions predictions. However, in recent years, more and more people are using intelligent algorithms to predict carbon emissions. Their commonly used intelligent algorithms are a genetic algorithm and a back propagation neural network (GA-BPNN), back propagation neural network (BPNN) [31, 32], and least squares support vector machine (LSSVM) [33] et al.

Although there is a lot of literature on the factors affecting industrial carbon dioxide emissions and carbon dioxide emissions in specific industrial sectors, only a handful of studies have focused on Chinese industry-level carbon emissions performance analysis and lack a comprehensive comparison of CO_2 emissions forecasts for all industrial sectors in China. Although some commonly used intelligent algorithms have achieved good performance in practical applications, they often have certain defects such as overfitting, falling into local optimum, and poor generalization ability et al. The purpose of this study is to fill these gaps by predicting the CO_2 emissions from Chinese industry using an extreme learning machine based on genetic algorithm optimization.

This paper first calculates the CO_2 emissions of the industrial sectors and preselects 15 factors related to industrial CO_2 emissions. Then, a bivariate correlation analysis and a linear regression analysis of the preselected factors and industrial CO_2 emissions are carried out to remove two factors that have no significant test coefficient. In order to reduce the correlation between influencing factors and to obtain some potential commonalities among the influencing factors, the remaining influencing factors are divided into four categories, performing factor analysis for each category, and finally obtaining five factors. The values of these five factors in 1995-2010 are used as the input values of the algorithm, and the industrial CO_2 emissions from 1995 to 2010 are used as output values to train the optimized extreme learning machine algorithm by the genetic algorithm (GA-ELM). The values of five factors in 2011-2015 are input into the trained GA-ELM algorithm to predict the industrial CO_2 emissions from 2011 to 2015. Finally, the forecast results and the actual values are compared and analyzed to verify the validity of the algorithm and the significant impact of the five factors on the industrial CO_2 emissions.

The remainder of the paper is structured as follows. The material and methods section introduces

the principle of GA-ELM algorithm. The empirical application section calculates the industrial CO_2 emissions and analyzes the preselect factors with SPSS, and uses the extracted five factors as the input of the GA-ELM algorithm to predict industrial CO_2 emissions. The results and discussion sections analyzes the forecast results. The conclusions section concludes the paper and proposes some policy recommendations for reducing industrial CO_2 emissions based on influencing factors.

Material and Methods

Extreme Learning Machine

The extreme learning machine (ELM) was proposed by Professor Huang Guangbin of Nanyang Institute of Technology in 2004. Its motivation was to overcome the problems of low learning efficiency and the complicated parameter settings of traditional neural network algorithms [34, 35]. ELM is a kind of machine-learning algorithm designed for a feed-forward neural network. Its main features are that the hidden layer node parameters do not need to be adjusted and that the learning process only needs to calculate the output weights. The ELM algorithm is usually applied to a single-layer feed-forward neuron network (SLFN). Its basic principle is as follows:

For a single hidden-layer neural network, its structure is shown in Fig. 1. Supposing there are n arbitrary samples (X_j, t_j), where

$$X_j = \left[x_{j1}, x_{j2}, ..., x_{jn} \right]^T \in R^n, t_j = \left[t_{j1}, t_{j2}, ..., t_{jm} \right] \in R^m$$

For a single hidden layer neural network with L, hidden nodes can be expressed as:

$$\sum_{i=1}^{L} \beta_i g(W_i \cdot X_j + b_i) = o_j, j = 1, 2, ..., N$$

…where g(x) is the activation function, $W_i = [w_{i1}, w_{i2}, ..., w_{im}]^T$ is the input weight of the ith hidden layer unit, b_j is the bias threshold of the ith hidden layer unit, and β_j is the output weight of the ith hidden layer unit. The goal of single hidden layer neural network learning is to minimize the output error and can be expressed as $\sum_{j=1}^{N} \left\| o_j - t_j \right\| = 0$, that is, W_i, X_j, and b_i exist

so that $\sum_{i=1}^{L} \beta_i g(W_i \cdot X_j + b_i) = t_j, j = 1, 2, ..., N$ can be

expressed as a matrix: $H \cdot \beta = T$, where H is the output of hidden nodes, β is output weight, and T is expected output.

$$H(W_1, ..., W_L, b_1, ..., b_L, X_1, ..., X_N) =$$

$$\begin{bmatrix} g(W_1 \cdot X_1 + b_1) & ... & g(W_L \cdot X_1 + b_L) \\ ... & ... & ... \\ g(W_1 \cdot X_N + b_i) & ... & g(W_L \cdot X_N + b_L) \end{bmatrix}_{N \times L}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ ... \\ \beta_L^T \end{bmatrix}_{L \times m} \quad T = \begin{bmatrix} t_1^T \\ ... \\ t_N^T \end{bmatrix}_{N \times m} \tag{1}$$

In order to be able to train a single hidden layer neural network, we hope to obtain $\hat{w}_i, \hat{b}_i$ and $\hat{\beta}_i$ so that $\left\| H(\hat{W}_i, \hat{b}_i) \cdot \hat{\beta} - T \right\| = \min_{w, b, \beta} \left\| H(W_i, b_i) \cdot \beta - T \right\|$, where $i = 1, 2, ..., L$ is equivalent to minimizing the loss function $E = \sum_{j=1}^{N} \left(\sum_{i=1}^{L} \beta_i \cdot g(W_i \cdot X_j + b_i) - t_j \right)^2$. Traditional gradient descent-based algorithms can be used to solve such problems, but the basic gradient-based learning algorithm needs to adjust all parameters during the iteration. In the ELM algorithm, once the input weight W and the hidden layer bias threshold b_i are randomly determined, the output matrix H of the hidden layer is uniquely determined. Training the hidden-layer neural network can be transformed into solving a linear system: $H \cdot \beta = T$, and the output weight can be determined $\hat{\beta} = H^\dagger \cdot T$, where $H^\dagger$ is the Moore-Penrose generalized inverse matrix of matrix H.

ELM is essentially a linear-in-the-parameter model, and its learning process tends to converge at a global minimum [36]. The advantages of ELM in learning efficiency and generalization performance have been confirmed in many fields, but in practical applications the randomly determined initial weights and biased thresholds will have a greater impact on the performance of ELM, leading to instability in the prediction results.
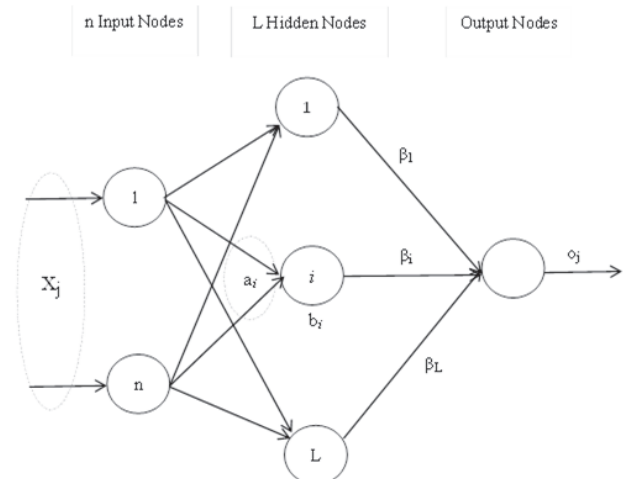

Fig. 1. Single-layer feedforward neuron network.

Genetic Algorithm

Genetic algorithm is a computational model that simulates the natural selection and genetic mechanism of Darwinian biological evolution. It is a method of searching for optimal solutions by simulating natural evolutionary processes. A genetic algorithm starts with a population that represents the potential solution set of a problem, while population consists of a certain number of coded individuals. After generating the initial population, according to the principle of survival of the fittest, generational evolution produces better and better approximate solutions. In each generation, individuals are selected according to the degree of fitness of the individual in the problem domain, and combined with the genetic operators of natural genetics to perform crossovers and mutations to produce a population representing the new solution set. This process will result in a population of natural evolution like the descendant population, and is more adaptive to the environment than the previous generation, the best individual in the last generation population is decoded and can be used as a problem to approximate the optimal solution.

Genetic algorithm, as a new global optimization search algorithm, has been widely used in various fields. It has achieved good results with simple features such as simplicity, versatility, robustness, suitability for parallel processing, high efficiency and practicability, and has gradually become one of the most important intelligent algorithms.

An Extreme Learning Machine Based on Genetic Algorithm Optimization

Since the genetic algorithm has good performance in search, in order to overcome the deficiency of the ELM algorithm, the genetic algorithm is used to optimize the initial values of input weights and hidden layer bias thresholds of the ELM algorithm to form a new algorithm, which is the genetic algorithm-extreme learning machine (GA-ELM) algorithm. The algorithm is used to predict carbon dioxide emissions from China's industrial sector. The flow chart of the research idea in this paper is shown in Fig. 2. The first part is to screen the influencing factors of industrial CO_2 emissions; the second part is to use the GA algorithm to optimize the
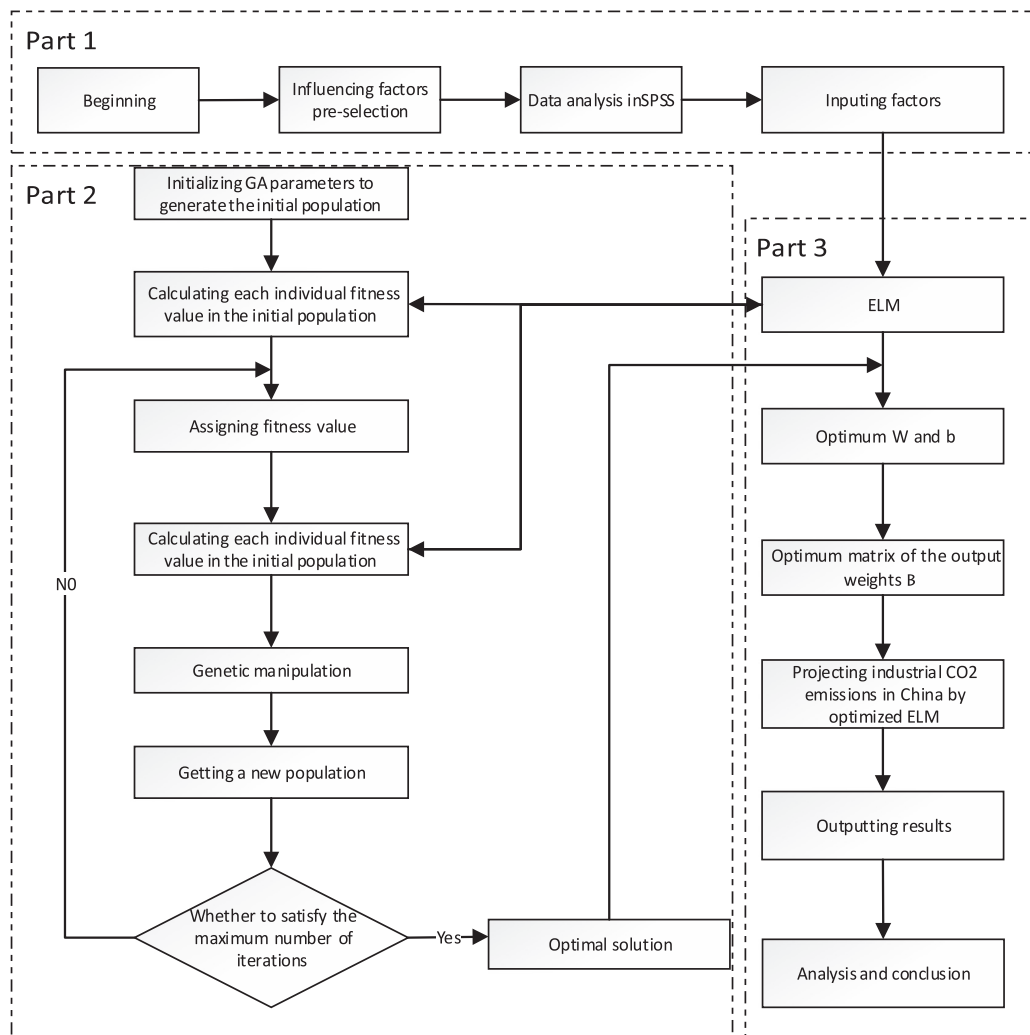

Fig. 2. Flowchart of the research idea in this paper.

initial values of input weights and hidden layer bias thresholds of the ELM algorithm; the third part uses the GA-ELM algorithm to predict industrial CO_2 emissions in China.

Prediction Application of ELM and GA-ELM Algorithms

The application of ELM and GA-ELM algorithms in prediction is more extensive. For example, in the prediction of the viscosity of ionic liquids [37], through the error analysis predicted by the ELM algorithm, the ELM model is more suitable for predicting the viscosity of ionic liquids than other algorithms. The prediction of the monthly effective drought index [38], through the ELM model, improved the prediction of drought duration and severity, and found that ELM is a faster tool for predicting drought and its related characteristics. The prediction of water solubility of carbon dioxide [39], comparing the estimation and prediction results of the ELM model with genetic programming (GP) and artificial neural network (ANN) models, the ELM model can be used safely to develop new water-soluble carbon dioxide. For the refractive index of ionic liquids [40], the average absolute relative deviation of the results predicted by ELM is only 0.295%. This reliable and accurate result highlights the potential of ELM algorithms in this field. The prediction of gas emission quantity [41], using the correlation data of a mine's gas emission quantity, the example analysis of the model shows that the GA-ELM model has higher prediction accuracy and can predict the gas emission in the working face relatively accurately and efficiently. Solar radiation prediction [42], using the optimized GA-ELM model, predicting the prediction of time-lapse solar radiation shows that compared with ELM and BP neural networks, the new method has higher prediction accuracy and can adapt to the needs of irradiation prediction under the condition of mutation of external meteorological conditions. For the prediction of wind power fluctuation range [43] we found that the GA-ELM prediction model can effectively track wind power variation and predict its fluctuation range. For the critical flow velocity prediction of slurry pipeline transportation [44], the average relative error value predicted by the GA-ELM model is 1.58%, while the BP neural network error is 12.95% and the SVM error is 3.19%, which indicates that the ELM model is more accurate and efficient.

From the above practical application of ELM or GA-ELM algorithms in different fields, it is not difficult to find that the ELM or GA-ELM algorithm has higher prediction accuracy than other intelligent algorithms. The ELM algorithm itself has the advantage of high learning efficiency, but its input weight matrix and hidden layer deviation are random. A GA algorithm has good global search ability, which can optimize ELM input weight matrix and hidden layer deviation randomness. The effect of the fitting is better, and the

predicted error value is smaller. Therefore, this paper uses the GA-ELM algorithm to predict industrial carbon emissions.

Empirical Application

Pre-selection of Industrial CO_2 Emissions Accounting and Influencing Factors

Since there is no an indicator of CO_2 emissions in the various statistical indicators of the Chinese government, we cannot directly obtain the CO_2 emissions in the Chinese industrial sectors. Therefore, CO_2 emissions from the industrial sectors can only be calculated through the consumption of various energies and related conversion factors. According to existing research [45], when accounting for CO_2 emissions, there are two situations: one is to use three sources of energy consumption for accounting, and another is to use nine sources of energy consumption for accounting. The use of three energy sources to account for CO_2 emissions often has large errors, because this method ignores secondary energy consumption, and using nine energy sources for accounting will greatly reduce these errors. Therefore, in order to objectively and truthfully reflect and accurately predict the CO_2 emissions of the industrial sectors, this paper uses the consumption of nine energy sources to calculate the CO_2 emissions of the industrial sectors. The formula for calculating CO_2 emissions is as follows:

$$C_t = \sum_{i=1}^{9} \alpha_i \beta_i E_i \qquad (2)$$

...where C_t is CO_2 emissions; i represents energy type; α is the conversion coefficient of different kinds ofenergy into standard coal, which are shown in Table 1; β is the CO_2 emissions conversion coefficient for different kinds

Table 1. Conversion coefficients of different kinds of energy into standard coal.

Energy	Unit	Conversion coefficient
Coal	million ton	0.7143 kgce/kg
Coke	million ton	0.9714 kgce/kg
Crude oil	million ton	1.4286 kgce/kg
Gasoline	million ton	1.4714 kgce/kg
Kerosene	million ton	1.4714 kgce/kg
Diesel oil	million ton	1.4571 kgce/kg
Fuel	million ton	1.4286 kgce/kg
Natural gas	billion cubic meters	1.2721 kgce/m^3
Power	billion kWh	0.1229 kgce/kWh

Table 2. CO_2 emissions conversion coefficients for various energies.

Energy	Conversion coefficient (C(t/t))
Coal	0.747
Coke	0.855
Crude oil	0.585
Gasoline	0.553
Kerosene	0.571
Diesel oil	0.592
Fuel	0.618
Natural gas	0.448
Power	1.814

of energy, which are shown in Table 2; and *E* is the energy consumption for different kinds of energy.

Through the China Energy Statistical Yearbook and the China Statistical Yearbook of 1983-2017, the consumption of nine types of energy in the industrial sectors can be obtained, and the CO_2 emissions of the industrial sectors from 1985 to 2015 can be calculated according to Formula (1). The result is shown in Fig. 3.

As can be seen from Fig. 3, overall, the CO_2 emissions from the industrial sectors have increased year by year. From the perspective of the growth rate and trends, the industrial CO_2 emissions from 1985 to 2015 can be divided into two phases. The first stage was 1985 to 2000. While industrial CO_2 emissions at this stage are also increasing year by year, the emissions and growth rates are relatively small. This was mainly due to the state of economic development in China at that time. From 1985 to 2000, China's industry was still in the process of the reform of the socialist economic system. The vitality of the development of the industrial sector has not yet been fully released. However, at this stage, as China's reforms in the economic and industrial sectors

continue to increase, China's industrial development under the planned economic system has gradually shifted to the development of industries under the socialist market economy. The industrial development model from semi-closed shifted to the open. Industrial enterprises from only state-owned and collectively-owned ownership forms changed to the pattern of public ownership as the main body and multiple ownership common development. Therefore, the first phase of the CO_2 emissions growth trend is relatively slow.

It is worth mentioning that the Asian financial turmoil occurred in 1998, which had a certain impact on young Chinese industries, which led to the second phase of the decline in 1998-2000. After entering the new century, the vitality of China's industrial reforms began to appear. At the same time, the Chinese government has further deepened the industrial economic system by reforming state-owned industries and encouraging non-public ownership industries. In opening up to the outside world, China joined the World Trade Organization (WTO) in 2001, and in 2005 it initiated the reform of the RMB exchange rate formation mechanism, which has had a profound impact on China's industrial development. Therefore, in the second phase, Chinese industry achieved rapid and steady development, and the corresponding CO_2 emissions also showed the same growth trend. However, in recent years, as environmental pressures have increased, the use ratio of renewable clean energy has begun to increase rapidly. The use of fossil fuels has been limited to some extent. Moreover, China is engaged in a new round of deepening reform. As a result, CO_2 emissions from the industrial sectors showed a downward trend at the end of the second phase.

There are many factors that affect the CO_2 emissions in China's industrial sectors. This paper pre-selects 15 influencing factors for industrial CO_2 emissions based on a comprehensive consideration of macroeconomics, population, industrial sector structure, energy consumption, energy consumption structure, and the availability of relevant data, including: industry GDP,
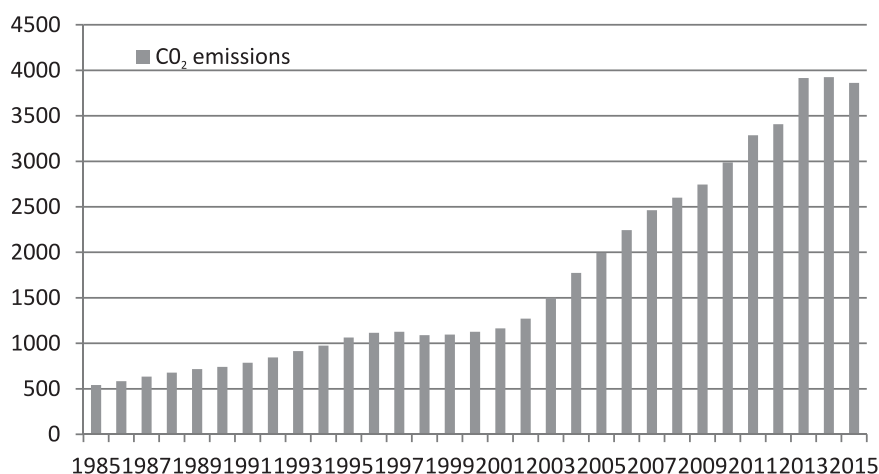

Fig. 3. CO_2 emissions in Chinese industrial sectors (unit: megatons).

total value of exports, energy industry investment in fixed assets, annual average employees of above-designated size, industrial added value, the number of industry enterprises, primary energy production, the production of pig iron, the production of steel, the production of cement, coal consumption in industry, coal consumption in mining and quarrying, coal consumption in manufacturing, coal consumption in electric power, gas and water production and supply, and fossil energy consumption ratio in industry.

Analysis of Influencing Factors

Bivariate Correlation Analysis and Linear Regression Analysis

Since the above 15 influencing factors are subjectively selected, there is a certain degree of randomness and subjectivity. In order to determine that the selected factors are indeed factors that affect industrial CO_2 emissions, this paper will conduct a bivariate analysis of preselected factors and industrial CO_2 emissions. The goal of bivariate analysis is to determine the correlation between two variables and measure their ability to predict or explain. The binary statistical analysis technique includes correlation analysis and regression analysis.

In this paper, we will first conduct the bivariate correlation analysis in SPSS; the results of the operation are shown in Table 3. According to the person coefficient and bilateral significance in Table 3, removing the factor of number of industrial enterprises, the other 14 preselected factors all have a strong correlation with industrial CO_2 emissions. The main reason that the correlation between number of industrial enterprises and industrial CO_2 emissions is not significant is that the data on the number of industrial enterprises used is from the Chinese Statistical Yearbook, and the data after 1997 from the statistical yearbook is only the number of enterprises above the designated size. That is, the statistical range before and after 1997 is inconsistent. In view of this, the factor of the number of industrial enterprises is removed from the pre-selected influencing factors.

Then, the remaining 14 influencing factors and industrial CO_2 emissions are conducted in a linear regressed in SPSS, and the regression method selected is "Enter". According to Table 4 and Table 5 of the operating results, the adjusted R^2 of the regression equation is 0.995, which indicates that the equation fits well. The F-test is significant, indicating that the linear relationship between industrial CO_2 emissions and the influencing factors of the entry equation are quite significant. However, at the same time, the factor of industrial coal consumption did not enter the regression equation. It can be seen from Table 6 that this factor and other factors have very strong multicollinearity. In fact, the sum of the coal consumption in mining and quarrying and the coal consumption in manufacturing and the coal consumption in electric power, gas and water production and supply is the consumption of coal in industry. Therefore, the factor of coal consumption in industry can be completely replaced by coal consumption in mining and quarrying and coal

Table 3. Bivariate correlation analysis.

Factor	Person coefficient	Significance (bilateral)
Industry GDP	0.992**	0.000
Total Value of Exports	0.992**	0.000
Energy Industry Investment in Fixed Assets	0.986**	0.000
Annual Average Employees of above Designated Size	0.963**	0.000
Industrial added value	0.991**	0.000
Number of Industry Enterprises	-0.221	0.231
Primary Energy Production	0.992**	0.000
Production of Pig Iron	0.992**	0.000
Production of Steel	0.994**	0.000
Production of Cement	0.996**	0.000
Coal Consumption in Industry	0.997**	0.000
Coal Consumption in Mining and Quarrying	0.961**	0.000
Coal Consumption in Manufacturing	0.977**	0.000
Coal Consumption in Electric Power, Gas and Water Production and Supply	0.987**	0.000
Fossil Energy Consumption Ratio in Industry	-0.955**	0.000

Note: **indicates a significant correlation at the bilateral significance level of 0.01.

Table 4. Model summary[b].

Model	R	R-squared	Adjusted R-squared	Standard estimated error
1	0.999[a]	0.997	0.995	78.50863

a. Predictive variables (constants): Fossil Energy Consumption Ratio in Industry, Coal Consumption in Mining and Quarrying, Annual Average Employees of above Designated Size, Coal Consumption in Manufacturing, Production of Pig Iron, Energy Industry Investment in Fixed Assets, Total Value of Exports, Primary Energy Production, Coal Consumption in Electric Power, Gas and Water Production and Supply, Industrial added value, Production of Cement, Production of Steel, Industry GDP

b. Dependent variable: Carbon dioxide emissions

Table 5. ANOVA[a]

Model		Quadratic sum	df	Mean square	F	Sig.
	Regression	37307215.29	13	2869785.792	465.602	0.000[b]
1	Residual	104781.274	17	6163.604		
	Total	37411996.57	30			

a. Dependent variable: Carbon dioxide emissions

b. Predictive variables (constants): Fossil Energy Consumption Ratio in Industry, Coal Consumption in Mining and Quarrying, Annual Average Employees of above Designated Size, Coal Consumption in Manufacturing, Production of Pig Iron, Energy Industry Investment in Fixed Assets, Total Value of Exports, Primary Energy Production, Coal Consumption in Electric Power, Gas and Water Production and Supply, Industrial added value, Production of Cement, Production of Steel, Industry GDP.

Table 6. Excluded variables[a]

Model		Beta in	t	Sig.	Partial correlation	Collinearity statistics		
						Tolerance	VIF	Minimum tolerance
1	Coal Consumptionin Industry	.[b]	.	.	.	0.000	.	0.000

a. Dependent variable: Carbon dioxide emissions

b. Predictive variables (constants): Fossil Energy Consumption Ratio in Industry, Coal Consumption in Mining and Quarrying, Annual Average Employees of above Designated Size, Coal Consumption in Manufacturing, Production of Pig Iron, Energy Industry Investment in Fixed Assets, Total Value of Exports, Primary Energy Production, Coal Consumption in Electric Power, Gas and Water Production and Supply, Industrial added value, Production of Cement, Production of Steel, Industry GDP.

consumption in manufacturing and coal consumption in electric power, gas and water production and supply. Therefore, this paper removes the factors of industrial coal consumption from the remaining 14 influencing factors.

Factor Analysis

There are many factors that affect carbon dioxide emissions in the industrial sector, but there may be correlations between many of the influencing factors. The presence of correlation will cause instability of the prediction model and some unpredictable results. Therefore, in order to reduce the correlation between factors and to obtain some potential common factors in the influencing factors, this paper uses factor analysis to analyze the remaining 13 factors. Factor analysis refers to statistical techniques for extracting commonalities from variable groups. It was originally proposed by British psychologist C.E. Spearman. Factor analysis can

find hidden and representative factors in many variables. Combining variables of the same nature into one factor reduces the number of variables and also tests the hypothesis of relationships between variables.

In order to avoid factor analysis to unduly reduce the dimensions of selected factors, on the basis of the research and comprehensive consideration of the existing literature on the influencing factors of industrial carbon dioxide emissions, 13 influencing factors are divided into four categories, namely social development (including total value of exports, energy industry investment in fixed assets, annual average employees of above designated size, industrial added value); production structure (including primary energy production, production of pig iron, production of steel, production of cement); energy efficiency (including industrial GDP, coal consumption in mining and quarrying, coal consumption in manufacturing, coal consumption in electric power, gas and water production and supply); and energy consumption structure

Table 7. Factor analysis of social development.

Indicator	Loading of F_{11}	F_{12}	Scoring coefficient of F_{11}	F_{12}	Index	Value
Total Value of Exports	0.763	0.634	0.423	-0.098	KMO	0.763
Energy Industry Investment in Fixed Assets	0.811	0.577	1.137	-0.915	Bartlett's test	294.552
Annual Average Employees of above Designated Size	0.587	0.809	-1.928	2.586	Sig.	0.000
Industrial added value	0.811	0.583	1.092	-0.862	Contribution rate	99.315%

Table 8. Factor analysis of production structure.

Indicator	Loading of F_2	Scoring coefficient of F_2	Index	Value
Primary Energy Production	0.997	0.25	KMO	0.752
Production of Pig Iron	0.999	0.251	Bartlett's test	417.076
Production of Steel	0.999	0.251	Sig.	0.000
Production of Cement	0.997	0.25	Contribution rate	99.566

Table 9. Factor analysis of energy efficiency.

Indicator	Loading of F_3	Scoring coefficient of F_3	Index	Value
Industry GDP	0.994	0.256	KMO	0.803
Coal Consumption in Mining and Quarrying	0.983	0.253	Bartlett's test	265.25
Coal Consumption in Manufacturing	0.989	0.255	Sig.	0.000
Coal Consumption in Electric Power, Gas and Water Production and Supply	0.974	0.251	Contribution rate	97.08%

(including fossil energy consumption ratio in industry). Then factor analysis in these four categories. Since the energy consumption structure contains only one influencing factor, factor analysis cannot be performed, and therefore it is considered to be a factor in itself. For factor analysis in the other three types of influencing factors with SPSS, the results are shown in Tables 7-9. It is generally believed that the closer the KMO value is to 1, the more suitable for factor analysis, and the KMO value greater than 0.7 can be considered suitable for factor analysis. The KMO values in Table 7-9 are all greater than 0.7, so the three types of influencing factors are all suitable for factor analysis, moreover, the Bartlett tests in each table are significant.

Table 7 shows that two special factors are extracted in the social development category. According to the rotated component matrix, the first component has a rotating load squared of 56.077%, but the cumulative rotational load can reach 99.315% together with the second component. From the specific data in Table 7 we can see that the first factor is the common factor of the total value of exports, energy industry investment in fixed assets and industrial added value. The second factor can reflect the main information of the influencing factors of annual average employees of above designated size. Therefore, the two factors in this category of social development are named economic scale and population scale respectively. There is only one factor for Table 8 and Table 9, the principal component contribution rates are 99.315% and 97.08%, respectively, so the two special factors are named as production structure and energy efficiency in this paper.

In summary, the four types of influencing factors can be further extracted into five factors, namely economic scale, population size, production structure, energy efficiency, and energy consumption structure. The data for the first four factors can be calculated from the component matrix. These five factors will be used to predict China's industrial carbon dioxide emissions.

Application of Projection Algorithms

In this paper, the extreme learning machine optimized by genetic algorithm are used to predict industrial CO_2 emissions. Data for five factors are used as inputs, and the CO_2 emissions of the industrial sectors are taken as output. The algorithm uses 1995-2010 data for training and predicts carbon dioxide emissions from the industrial sector in 2011-2015. In order to measure the predictive effect of the GA-ELM algorithm,

Table 10. Parameter settings of projection algorithms.

Algorithm	Nodes in hidden layer	Population size	Maximal genetic algebra	Training epochs	Training goal	Learning rate
GA-ELM	16	40	50			
ELM	16					
BPNN	18			2000	0.001	0.1
GA-BPNN	18	40	50	2000	0.001	0.1

the extreme learning machine (ELM), back propagation neural network (BPNN), and backpropagation neural network optimized by the genetic algorithm (GA-BPNN) are selected as comparisons. Table 10 shows some of the parameter settings of each algorithm. At the same time, the performance of each prediction algorithm is evaluated using mean absolute percentage error (MAPE), maximum absolute percentage error (MaxAPE), median absolute percentage error (MdAPE) and root mean square error (RMSE):

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100\% \tag{2}$$

$$MaxAPE = \max\left(\left|\frac{y_t - \hat{y}_t}{y_t}\right|\right) \times 100\% \tag{3}$$

$$MdAPE = \underset{i}{median}\left(\left|\frac{y_t - \hat{y}_t}{y_t}\right|\right) \times 100\% \tag{4}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|^2} \tag{5}$$

…where n represents the number of the year, y_i is the actual data, and $\hat{y}_i$ is the forecasted data.

Results and Discussion

The projection procedures of all the algorithms in this paper are all run on the Matalab R2016a version under Windows 10. The forecast results of the industrial CO_2 emissions of each algorithm from 2011 to 2015 are shown in Fig. 4, which shows that the best fitting curve to the real-value curve is the prediction curve of the GA-ELM algorithm. This model's predictions of industrial CO_2 emissions for 2011-2014 are almost the same as the real values; only for 2015 was the difference between the predicted value and the actual value slightly greater. There is a large difference between the forecast results of the ELM, BPNN and the GA-BPNN algorithms and the real values. Only the trend of the ELM algorithm's predicted results curve and the real value curve are similar.

According to the data in Table 11, we can see that the values of the four indicators corresponding to the GA-ELM algorithm are the smallest among the four prediction algorithm indicators. The MAxAPE, MAPE, and RMSE values of GA-ELM are less than one-sixth,
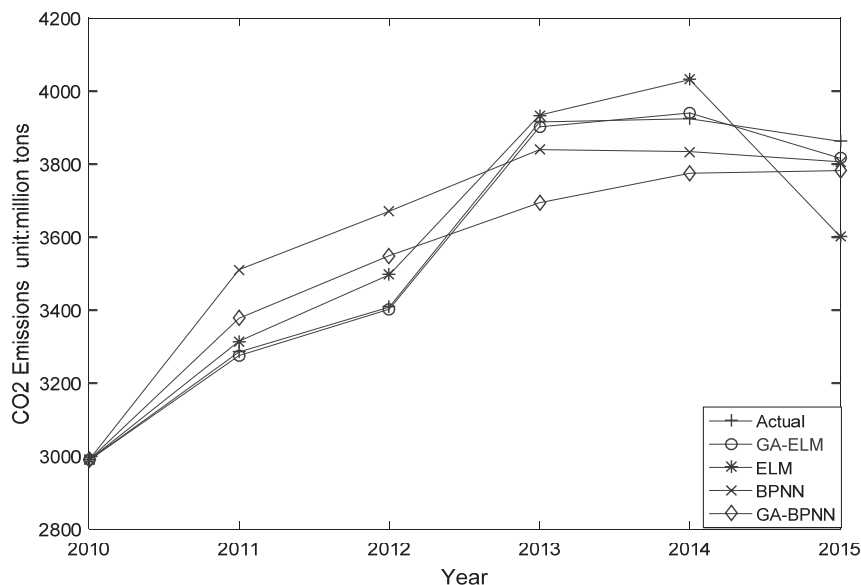

Fig. 4. Projected results of industrial CO_2 emissions of various predictive algorithms.

Table 11. Errors of the fitting results of different algorithms.

Index	GA-ELM	ELM	BPNN	GA-BPNN
MaxAPE(%)	1.195	6.7345	7.7099	5.6481
MAPE(%)	0.47964	2.6899	4.0504	3.6943
MdAPE(%)	0.33369	2.5996	2.2948	3.8031
RMSE(megatons)	23.155	132.85	165.47	145.62

one-eighth, and one-seventh of BPNN, respectively. By comparing the corresponding indicators of BPNN and GA-BPNN, we can see that – except for the value of MdAPE of GA-BPNN being higher than that of BPNN – the other indicators of GA-BPNN are all smaller than these indicators of BPNN, which shows that GA-BPNN has better prediction accuracy than BPNN, and that GA is effective for BPNN optimization.

It can be obtained by analyzing the above prediction results. The GA-ELM algorithm has high accuracy and performance in China's industrial CO_2 emissions forecasting. This not only indicates that the algorithm is suitable for such research, but also shows that the processing method of the influencing factors is reasonable. The five influencing factors have significant influence on China's industrial CO_2 emissions. These influencing factors are suitable for the input value of the algorithm selected in this paper.

In today's countries, we focus on carbon dioxide emissions and promote low-carbon energy. Influencing factors in the study of CO_2 emissions are extremely important for governments and relevant agencies around the world to develop targeted strategies. In order to realize the energy-saving potential, it is necessary to formulate industrial policies and guidelines to implement higher environmental protection. This paper can reduce the carbon emissions generated by the industrial sector from the perspective of economic scale, population size, production structure, energy efficiency and energy consumption. Formulate corresponding policies and guidelines for various factors, such as adjusting economic structure, enhancing humanistic quality, optimizing production structure, improving energy efficiency, improving energy consumption structure, actively developing renewable energy and advanced nuclear technology, adjusting national economic structure, and formulating legislation and regulations, and establish energy-saving institutions to achieve maximum energy conservation and emission reduction.

Conclusions

This paper first selects the factors that affect the carbon emissions of the industrial sector, and selects the most critical factors affecting industrial carbon emissions. Secondly, using 1995-2010 as sample data,

the GA-ELM algorithm is used to predict the carbon dioxide emissions of the industrial sector from 2011 to 2015. The GA-ELM algorithm and other intelligent algorithms are used to predict the carbon emissions in 2011-2015 and compare them with the actual carbon emissions. Finally, the following conclusions are drawn:

(1) Economic scale, population size, production structure, energy efficiency and energy consumption are important factors affecting industrial carbon emissions.

(2) The GA-ELM algorithm has the habit of self-organization, self-adaptation and self-learning. It can quickly search for optimal weights and thresholds, make the ELM network model more compact, overcome the randomness defects of ELM algorithm input weight and deviation threshold. Comparing the prediction results of the GA-ELM algorithm with the actual values in 2011-2015, and then using the ELM algorithm, the BPNN algorithm and the GA-BPNN algorithm also make a comparison between the predicted value and the actual value in 2011-2015. We found that the GA-ELM algorithm has a higher degree of fit and the error is smaller.

(3) The ELM and GA-ELM algorithms are widely used in forecasting. For example, for predicting ionic liquid viscosity [37], the monthly effective drought index [38], carbon dioxide water solubility [39], the ionic liquid refractive index [40], gas emissions (41), solar radiation [42], wind power fluctuation range [43], the critical flow rate of mud pipeline transportation [44] and so on. The GA-ELM algorithm can be widely used in different fields for prediction, which shows that this prediction model has obvious advantages.

(4) Scholars around the world attach great importance to the issue of carbon dioxide emissions, and the introduction of this algorithm into the prediction of carbon emissions is an innovative attempt. In this paper, the prediction of carbon dioxide by this method has also achieved a good predictive effect, which has played an active role in the application of this algorithm in the forecasting field.

(5) Considering that the most critical factors affecting CO_2 emissions in the industrial sector are economic size, population size, production structure, energy efficiency and energy consumption, it is necessary to reduce the carbon emissions generated by the industry from these aspects.

Concrete measures are:
- Guide funds to strengthen investment in low-carbon and high-efficiency departments to accelerate the withdrawal of overcapacity-type heavy industry products.
- Follow the principle of product supply and demand balance, strictly review new reporting items, and refuse to repeat construction.
- Break the price monopoly, clarify the mechanism of energy price formation, and promote price marketization.
- Accelerate the construction of enterprise automation and improve the mechanization and intelligence of industrial processes.
- Optimize the industrial structure and effectively control the further expansion of high-energy-consuming industries through taxation, credit, legal and other means.
- Encourage mergers and restructuring of similar companies and eliminate backward production capacity.
- Implement structural reforms in supply.
- Capture capacity reductions, destock, deleverage, reduce costs and reduce shortages.
- Improve energy efficiency, transform traditional industries with advanced technologies.
- Actively introduce foreign capital and advanced technologies.
- Optimize business processes and manufacturing processes, and improve the scientific management of level managers.
- Improve the energy consumption structure, increase the proportion of clean energy use.
- Prioritize the use of clean solar energy, wind energy and nuclear energy.
- Increase the electrification of various industrial sectors.
- Develop and use clean energy in the industrial sectors to provide policy support and tax incentives.

Acknowledgements

Conflict of Interest

The authors declare no conflict of interest.

References

1. GUAN D., PETERS G.P., WEBER C.L., HUBACEK K. Journey to world top emitter: an analysis of the driving forces of China's recent CO_2 emissions surge.Geophysical Research Letters, **36** (4), **2009**.

2. WANG T., WATSON J. Scenario analysis of China's emissions pathways in the 21st century for low carbon transition.Energy Policy, **38** (7), 3537, **2010**.

3. HE J., DENG J., SU M. CO_2 emission from China's energy sector and strategy for its control. Energy, **35** (11), 4494, **2010**.

4. European Commission.Emission Database for Global Atmospheric Research (EDGAR v4.3.2).Joint Research Centre (JRC)/PBLNetherlands Environmental Assessment Agency http://edgar.jrc.ec.europe.eu(accessed on 22 September **2018**).

5. U.S-China.U.S-China.joint announcement on climate change: http://www.whitehouse.gov/the-press-office/2014/11/11/us-china-joint-announcement-climate-change.htm(accessed on 22 September 2018).

6. The Chinese Government. Enhanced Actions on Climate Change: China's Intended Nationally Determined Contributions, http://www.scio.gov.cn/xwfbh/xwbfbh/wqfbh/33978/35364/xgzc35370/Document/1514539/1514539.htm(accessed on 23 September 2018).

7. ZHOU N., FRIDLEY D., KHANNA N.Z., KE J., MCNEIL M., LEVINE M. China's energy and emissions outlook to 2050: Perspectives from bottom-up energy end-use model. Energy Policy, **53** (53), 51, **2013**.

8. United Nations Framework Convention on Climate Change. In: 2015 United Nations climate change conference; 2015. Available from: https://unfccc.int/process/conferences/pastconferences/paris-climate-change-conference-november-2015/overview(accessed on 24 September 2018).

9. YANG T., PAN Y.K., YANG Y.K., LIN M.S., QIN B.Y., XU P., HUANG Z.Z. CO_2 emissions in China's building sector through 2050: A scenario analysis based on a bottom-up model.Energy, **128** (2017), 208, **2017**.

10. IPCC. Fifth assessment report (AR5) climate change 2014: mitigation of climate change. Intergovernmental Panel on Climate Change,**2014**.

11. ZHANG N., WANG B., LIU Z. Carbon emissions dynamics, efficiency gains, and technological innovation in China's industrial sectors. Energy, **99**, 10, **2016**.

12. LI H., WEI Y M., MI Z. China's carbon flow: 2008–2012. Energy Policy, **80**, 45, **2015**.

13. AKASHI O., HANAOKA T., MATSUOKA Y., KAINUMA M. A projection for global CO_2, emissions from the industrial sector through 2030 based on activity level and technology changes. Energy, **36** (4), 1855, **2011**.

14. JINLONG M.A., EVANS D.E., FULLER R.J., STEWART D.S. Energy-related CO_2 emissions in china's industrial sector.sustainable Energy And Environmental Technologies, **2015**.

15. CAO Z., PENG H., FANG L.U. Industrial Carbon Emissions Status and Analysis on Emission Reduction Methods. Ecological Economy, **2011**.

16. ZHAO M., TAN L., ZHANG W., JI M.,LIU Y., YU L. Decomposing the influencing factors of industrial carbon emissions in Shanghai using the LMDI method. Energy, **35** (6), 2505, **2010**.

17. REN X.S., ZHAO G.H. Gray Forecasting of China's Industrial Carbon Emissions and its Influencing Factors – A study based on the STIRPAT model. Journal of Beijing Jiaotong University, **2014**.

18. XU X., YANG G., TAN Y., ZHUANG Q., TANG X., ZHAO K., WANG S. Factors influencing industrial carbon emissions and strategies for carbon mitigation in the Yangtze River Delta of China. Journal of Cleaner Production, **142**, **2016**.

19. LIU L., WANG S., WANG K., ZHANG R.,TANG X. LMDI decomposition analysis of industry carbon emissions in Henan Province, China: comparison between different 5-year plans. Natural Hazards, **80** (2), 997, **2016**.

20. LIU L.C., FAN Y., WU G., WEI Y.M. Using LMDI method to analyze the change of China's industrial CO_2 emissions from final fuel use: An empirical analysis. Energy Policy, **35** (11), 5892, **2007**.

21. SUMMERBELL D.L., BARLOW C.Y., CULLEN J.M. Potential reduction of carbon emissions by performance improvement: A cement industry case study. Journal of Cleaner Production, **135**, 1327, **2016**.

22. SUN W., CAI J., DAI L. Decomposition analysis of energy-related carbon dioxide emissions in the iron and steel industry in China. Frontiers of Environmental Science & Engineering, **6** (2), 265, **2012**.

23. KARMELLOS M., KOPIDOU D., DIAKOULAKI D. A decomposition analysis of the driving factors of CO_2, (Carbon dioxide) emissions from the power sector in the European Union countries. Energy, **94** (9), 680, **2016**.

24. SUN W., MENG M., HE Y., CHANG K. CO_2 Emissions from China's Power Industry: Scenarios and Policies for 13th Five-Year Plan. Energies, **9** (10), 825, **2016**.

25. MENG M., JING K., MANDER S. Scenario analysis of CO_2, emissions from China's electric power industry. Journal of Cleaner Production, **142**, 3101, **2017**.

26. TUDOR C. Predicting the Evolution of CO2 Emissions in Bahrain with Automated Forecasting Methods. Sustainability, **8** (9), 923, **2016**.

27. MING M., NIU D.,WEI S. CO_2, emissions and economic development: China's 12th five-year plan. Energy Policy, **42**, 468, **2012**.

28. CAOAO C., CHUNLAN L., HAO W., JING G., LONG C., HAIHUA W., JIPING Z., ZHENG L., XIAOJIE L. Examining the impact factors of energy consumption related carbon footprints using the STIRPAT model and PLS model in Beijing.China Environmental Science, **34** (6), 1622, **2014**.

29. MENG M., SHANG W., NIU D., GAO Q. A New Algorithm of Parameter Estimation for the Logistic Equation in Modeling CO_2 Emissions from Fossil Fuel Combustion. Mathematical Problems in Engineering, **2014** (3), 1, **2014**.

30. MENG M., NIU D., WEI S. A small-sample hybrid model for forecasting energy-related CO_2 emissions. Energy, **64** (1), 673, **2014**.

31. SUN W., XU Y. Using a back propagation neural network based on improved particle swarm optimization to study the influential factors of carbon dioxide emissions in Hebei Province, China. Journal of Cleaner Production, **112**, 1282, **2016**.

32. SUN W., YEM., XU Y. Study of carbon dioxide emissions prediction in Hebei province, China using a BPNN based on GA. Journal of Renewable & Sustainable Energy, **8** (4), 043101, **2016**.

33. SUN W., LIU M. Prediction and analysis of the three major industries and residential consumption CO_2, emissions based on least squares support vector machine in China. Journal of Cleaner Production, **122**, 144, **2016**.

34. HUANG G.B., ZHU Q.Y., SIEW C.K. Extreme learning machine: theory and applications. Neurocomputing, **70** (1), 489, **2006**.

35. HUANG G.B., ZHU Q.Y., SIEW C.K. Extreme learning machine: a new learning scheme of feed forward neural networks. Neural Networks, Proceedings. IEEE International Joint Conference on, IEEE, **2004**.

36. HUANG G., HUANG G.B., SONG S.,YOU K. Trends in extreme learning machines: A review. Neural Networks, **61**, 32, **2015**.

37. KANG X., ZHAO Z., QIAN J., RAJA M.Predicting the Viscosity of Ionic Liquids by the ELM Intelligence Algorithm. Industrial & Engineering Chemistry Research, **56** (39), 11344, **2017**.

38. DEO R.C., ŞAHIN MEHMET Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. Atmospheric Research, **153**, 512, **2015**.

39. MOHAMMADIAN E., MOTAMEDI S., SHAMSHIRBAND S., HASHIM R., JUNIN R., ROY C., AZDARPOUR A. Application of extreme learning machine for prediction of aqueous solubility of carbon dioxide. Environmental Earth Sciences, **75** (3), 215, **2016**.

40. KANG X., ZHAO Y., LI J. Predicting refractive index of ionic liquids based on the extreme learning machine (ELM) intelligence algorithm. Journal of Molecular Liquids, **250**, 44, **2017**.

41. YIBO H. Gas Emission Prediction Based on GA-ELM. Safety in Coal Mines, **46**, 4, **2015**.

42. XU JING, HUANG NANTIAN, WANG WENTING, QI JIAJIN, XU SHIBING, YU ZHIYONG Hourly Solar Radiation Forecasting Based on GA-ELM Neural Network. Power System and Clean Energy, **32**, 8, **2016**.

43. WANG HAO, WANG YAN, JI ZHICHENG Simulation of Wind Power Prediction Based on Improved ELM. Journal of System Simulation, **30**, 11, **2018**.

44. XINMIN W., TIANZHENG L.I., QINLI Z. Study on prediction model of critical flow velocity in slurry pipeline transportation based on GA-ELM. Journal of Safety Science and Technology, **11**, 8, **2015**.

45. SHAN Y., GUAN D., ZHENG H., OU J., LI Y., MENG J., MI Z., LIU Z., ZHANG Q. China CO_2 emission accounts 1997-2015. Scientific Data, **5** (2018), 170201, **2018**.