2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)

Research on the application of a combined model in carbon emission prediction

Liu Rui
School of Accounting
Wuxi Taihu University
Wuxi, China
e-mail: aprillr@sina.com

Cai Feijun
School of Accounting
Wuxi Taihu University
Wuxi, China
e-mail: hxlian@126.com

***Abstract*—Nowadays, the prediction accuracy of carbon emissions is required to be improved, a combination model for prediction is proposed. First, calculate the carbon emissions according to the carbon emission conversion formula of petrochemical energy consumption, then use the trend moving average method to pre-process the calculated carbon emissions, and finally combine the pre-processed data with the grey linear regression model to realize the prediction of future carbon emissions. The experimental results show that the prediction accuracy of using traditional linear regression model and GM (1,1) is low, while using the grey linear regression model is good, but it is still lower than using the combined model proposed.**

Keywords-linear regression model; GM (1,1); trend moving average method; combined model; carbon emissions

I. INTRODUCTION

For the prediction of carbon emissions, many scholars have proposed many better theoretical analysis and feasible verification methods for the calculation and prediction of carbon emissions. For example, classic carbon emission calculation models include LMDI[1] decomposition method, Kaya model [2], CGE model [3], production function theory [4], etc. But they still have problems such as insufficient theory and low prediction accuracy [4]. To this end, taking the historical data from 2005 to 2018 published in China and the carbon emissions data for 2019 and 2020 predicted in the paper [12] as the object, on the basis of analyzing the historical data from 2005 to 2018, a method for predicting carbon emissions by a combined model is proposed. Results show that the prediction accuracy of carbon emissions using the combined model exceeds three traditional models.

II. DATA ANALYSIS

In general, the energy consumption structure is mainly composed of coal, oil, natural gas and other non-petrochemical energy sources. The combustion of petrochemical energy mainly causes CO2 emission, and the conversion formula of its carbon emissions [6]:

$$C_i = E_i \times K_i \times \frac{44}{12} \qquad (1)$$

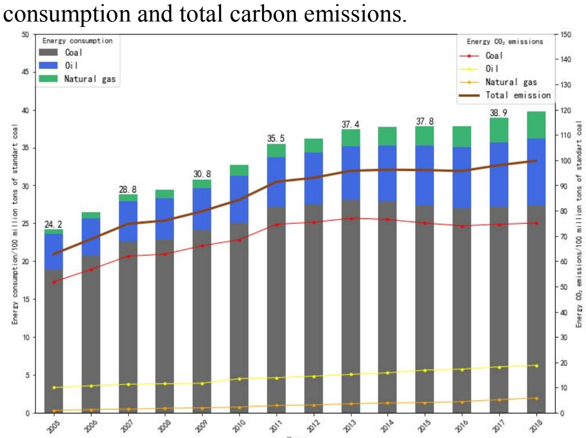
In the formula, C_i is the carbon emission of the ith energy; K_i is the CO_2 emission coefficient of the ith energy, and its value can be based on the carbon emission coefficient recommended by the country. The CO_2 emission coefficients of coal, oil, natural gas and non-petrochemical energy sources are taken as: 0.7476, 0.5825, 0.4435 and 0 respectively. Because the emission coefficient of non-petrochemical energy is 0, i refers to three petrochemical energies respectively..

Therefore, from the data published in China, the energy consumption from 2005 to 2018 was selected, and the formula (1) was used to convert the calculation to obtain the CO_2 emissions generated by each energy. It is shown in Table I.

TABLE I Total energy consumption and total CO_2 emissions of Sinopec from 2005 to 2018

	CC	CCE	OC	OCE	NGC	NGCE	TPEC	TCE
2005	18.9	51.8	4.7	10.0	0.6	0.98	24.2	62.78
2006	20.7	56.7	5.0	10.7	0.8	1.30	26.5	68.7
2007	22.6	62.0	5.3	11.3	0.9	1.46	28.8	74.76
2008	22.9	62.8	5.4	11.5	1.1	1.79	29.4	76.09
2009	24.1	66.1	5.5	11.7	1.2	1.95	30.8	79.75
2010	25.0	68.5	6.3	13.5	1.4	2.28	32.7	84.28
2011	27.2	74.6	6.5	13.9	1.8	2.93	35.5	91.43
2012	27.5	75.4	6.8	14.5	1.9	3.09	36.2	92.99
2013	28.1	77.0	7.1	15.2	2.2	3.58	37.4	95.78
2014	27.9	76.5	7.4	15.8	2.4	3.90	37.7	96.2
2015	27.4	75.1	7.9	16.9	2.5	4.07	37.8	96.07
2016	27.0	74.0	8.1	17.3	2.7	4.39	37.8	95.69
2017	27.2	74.6	8.5	18.2	3.2	5.20	38.9	98
2018	27.4	75.1	8.8	18.8	3.6	5.85	39.8	99.75

The relevant abbreviations with the same unit in Table I are used for explanation. CC, OC and NGC represent each energy consumption / 100 million tons of standard coal for coal, oil and natural gas respectively. Then, CCE, OCE and NGCE represent CO_2 emissions of each energy respectively with the same unit while TPEC and TCE stand for total petrochemical energy consumption and total carbon emissions.



DOI 10.1109/DCABES50732.2020.00081

Figure 1 Schematic diagram of petrochemical energy consumption and total CO_2 emissions from 2005 to 2018

The bar in Figure 1 indicates the total petrochemical energy consumption and includes the proportion of various types of petrochemical energy consumption. The above data is the total consumption of petrochemical energy; the curve shows the CO_2 emissions and total CO_2 emissions from various types of petrochemical energy consumption. From the distribution and changes of data in the above figure, it can be shown that China's total CO_2 emissions have increased significantly along with the growth of total petrochemical energy consumption, and the trend of change is linear and exponential. Therefore, using a linear regression model combined with GM(1,1) to predict carbon emissions can reflect the general law of its changes.

III. COMBINED MODEL

A. Grey linear regression model [7] [8]

Supposing $\widetilde{X}^{(0)} = (\widetilde{x}^{(0)}(1), \widetilde{x}^{(0)}(2), \cdots, \widetilde{x}^{(0)}(n))$ is the original sequence as a basis , and from that, $\widetilde{X}^{(1)}$ = $(\widetilde{x}^{(1)}(1), \widetilde{x}^{(1)}(2), \cdots, \widetilde{x}^{(1)}(n))$ is the generated sequence getted from the 1-AGO of $\widetilde{X}^{(0)}$ (in which $\widetilde{x}^{(1)}(i) = \sum_{i=1}^{i} \widetilde{x}^{(0)}(i)$. Then, it can be obtained from the GM (1,1) model:

$$\hat{\widetilde{x}}^{(1)}(t+1) = (\widetilde{x}^{(0)}(1) - \frac{b}{a})\exp(-at) + \frac{b}{a} \quad (2)$$

The above formula obviously has the law of exponential growth, which can be recorded as：

$$\hat{\widetilde{x}}^{(1)}(t+1) = C_1 \exp(\alpha t) + C_2 \quad (3)$$

Therefore, the sum of $\widetilde{Y} = a\widetilde{X} + b$ and $\tilde{Y} = a * \exp(\tilde{X})$ is used to fit the cumulative generation sequence $\widetilde{x}^{(1)}(t)$. And then the following form of sequence can be getted:

$$\hat{\widetilde{x}}(t) = C_1 \exp(\alpha t) + C_2 t + C_3 \quad (4)$$

Obviously, in formula (4), if C_1 =0, the generated sequence obtained by 1-AGO is a linear regression model; if C_2 =0, the generated sequence is a GM(1,1) model. Therefore, the combined model can improve the deficiencies in the linear regression model and the GM(1,1) model. So the combined new model can improve the situation that the original linear regression model doesn't contain exponential growth trends and the GM (1,1) model does not contain linear factors. Among them, the parameter α is：

$$\alpha = \ln[\frac{\tilde{y}_k(t+1)}{\tilde{y}_k(t)}] \quad (5)$$

For different m , we can get different α . And their average value can be used as the estimated value of $\hat{\alpha}$. If recording $m(t) = \exp(\hat{\alpha} t)$, formula (4) is as follows:

$$\hat{x}^{(1)}(t) = C_1 m(t) + C_2 t + C_3 \quad (6)$$

If recording

$$\widetilde{X}^{(1)} = \left[\widetilde{x}^{(1)}(1) \quad \widetilde{x}^{(1)}(2) \quad \cdots \quad \widetilde{x}^{(1)}(n)\right]^T,$$

$$C = \left[C_1 \quad C_2 \quad C_3\right]^T,$$

$$A = \begin{bmatrix} m(1) & 1 & 1 \\ m(2) & 2 & 1 \\ \cdots & \cdots & \cdots \\ m(n) & n & 1 \end{bmatrix} \quad (7)$$

Then the available parameters C_1, C_2, C_3 source from the least square method.

$$C = (A^T A)^{-1} A^T \widetilde{X}^{(1)} \quad (8)$$

Then, $z(t) = \hat{\widetilde{x}}^{(1)}(t+1) - \hat{\widetilde{x}}^{(1)}(t)$, $y_k(t) = z(t+k) - z(t)$.

B. Trend moving average method

This method is to predict the average value based on the gradual passage of time series and moving average over a certain period. It can eliminate the impact of periodic fluctuations and irregular changes in the time series on the prediction results, and can correct the lag deviation caused by only one moving average[9][10].The total CO_2 emissions from petrochemical energy consumption are often affected by many factors, such as population, GDP, economic policies, industrial structure adjustments, economic crises and natural disasters, which will cause large fluctuations in energy consumption. The projections of the emissions are disturbing. Therefore, using historical data of total CO_2 emissions with trend moving average to eliminate or weaken the impact of fluctuations can highlight the trend and periodic variability of carbon emissions and further improve the accuracy of model prediction.

Suppose the time series $\{y_t\}$ starts to change in a straight-line trend from a certain moment, and then the trend moving average method is based on the starting point of a moving average of the latest actual value. Using the quadratic moving average to estimate the slope of the trend change, a prediction model [9] is established as:

$$y_{t+k} = a_t + b_t \times k \quad (9)$$

In the formula (9), k is the number of trend forecast periods and y_{t+k} is the forecast value of the $t+k$. Among them:

$$a_t = 2m_t^{(1)} - m_t^{(2)}$$

$$b_t = \frac{2}{n-1}(m_t^{(1)} - m_t^{(2)})$$

In the above formulas, a_t is the intercept of the predicted straight line, b_t is the slope of the predicted straight line, n is the length of each moving average, t is the number of periods, $m_t^{(1)}$ is a moving average and $m_t^{(2)}$ is a second moving average.

C. prediction model validation

Taking the total carbon emission data for 2005-2016 published [5] as the object, after applying the formula (1) carbon emission conversion, three traditional models and the combined model in this paper are used to predict the total carbon emissions from 2017 to 2020. However, since the data in 2019 and 2020 have not be released, the calculation results in Table 7 of [12] are cited as the actual data for comparison. Among them, for the trend moving average method of carbon emission data preprocessing, the sliding window length is 5. The verification results are as follow.

TABLE II Total carbon emissions predicted by different models

Year / Algorithm	Result	2017	2018	2019	2020
	Actual	98	99.75	99.04	100.3
LR	Predict	108.02	111.5	115.0	118.5
	Error	10.02	11.77	15.99	18.19
GM(1,1)	Predict	108.67	112.9	117.2	121.6
	Error	10.67	13.08	18.13	21.31
GLR	Predict	96.72	97.06	97.31	97.50
	Error	-1.27	-2.68	-1.73	-2.83
CM	Predict	99.06	99.71	100.2	100.7
	Error	1.06	-0.04	1.20	0.32

It can be seen from the above table that the prediction accuracy of the improved grey linear regression model exceeds that of the traditional linear regression model, GM (1,1) model and grey linear regression model. At the same time, the aboved indicates that the prediction results for 2019 and 2020 are also close to the calculation results in [11], indicating that the combined model in this paper is effective and feasible for the prediction of carbon emissions.

In addition, the visualization results of the total carbon emissions predicted for 2017-2020 under different models are shown in Figure 2. It can also be seen from Figure 2 that the combined models proposed are superior to the prediction accuracy of traditional linear regression models, GM (1,1) models, grey linear regression models, etc.

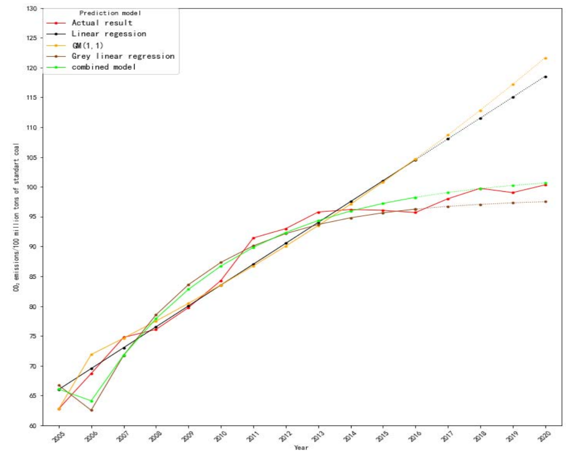


Figure 2 Different models predicting the total carbon emissions from 2017 to 2020

IV. CONCLUSION

With the economic development of various countries, resulting in a substantial increase in carbon emissions, the greenhouse effect caused has become one of the hot issues of global concern. In order to achieve a low-carbon development transformation and reduce CO_2 emissions, countries around the world are jointly formulating relevant emission reduction targets. Therefore, for improving the prediction accuracy of carbon emissions, a combined model method using trend moving average method to improve grey linear regression is established, which can assist countries to formulate relevant carbon emission legislation, adjust and optimize industry and energy structure, and master carbon emissions. The volume growth trend provides an effective theoretical analysis method.

REFERENCES

[1]ANG B W. LMDI decomposition approach: a guide for implementation[J]. Energy Policy,2015,86:233-238.

[2] LIMA F,NUNES M L,CUNHA J,et al. A cross-country assessment of energy-related co2 emissions: an extended Kaya index decomposition approach[J]. Energy,2016,115:1361-1374.

[3] FREIRE-GONZALEZ J. Environmental taxation and the double dividend hypothesis in CGE modeling literature: a critical review[J]. Journal of Policy Modeling,2018,40(1):194-223.

[4] Wu Jian, Xu Jiayu. Carbon emission model and its application based on production function theory [J]. Journal of Jiangsu University, 2019, 40 (3): 320 -324.

[5] China Statistical Yearbook 2019 of the National Bureau of Statistics. Http://www.stats.gov.cn/tjsj/ndsj/2019/indexch.htm.

[6] Zeng Xiaoli. Energy consumption structure adjustment and intensity effect of carbon emission[J]. Journal of Zhengzhou Institute of Aeronautical Industry Management, 2016, 34 (4): 38-43.

[7] Deng Julong. Grey System (Social&Economic) [M]. Beijing: National Defense Industry Press, 1985.

[8] Liu Sifeng, Dang Yaoguo, Fang Zhigeng, etc. Grey System Theory and Application [M]. Beijing: Science Press, 1991.

[9] Liang Shichun, Zhang Xiaodong, Lin Peifeng et al. A power prediction algorithm for hybrid energy storage photovoltaic power generation system [J]. China Electric Power, 2014, 47 (3): 24-27.

[10] Liu Jiheng, Xu Yong, Yu Fengping, et al. Forecast of hepatitis A virus by seasonal trend model using moving average method [J]. China Health Education, 2015, 31 (11): 1069-1072.

[11] Chen Wei. Application of moving average in monitoring and early warning of tuberculosis epidemic situation in Hefei City [J]. Chinese Journal of Antituberculosis, 2011, 33 (10): 689-691.

[12] Fang Debin, Dong Wei, Yu Qian. Optimization of China's energy consumption structure under the low-carbon transition trend [J]. Technology and Economy, 2015, 35 (7): 71-79, 128.