

PAPER • OPEN ACCESS

## Short-term and long-term prediction of coal consumption in Shandong Province- with ARIMA Model and Metabolism GM (1, 1) Model

To cite this article: Minglu Ma *et al* 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **227** 042004

View the [article online](#) for updates and enhancements.



**IOP | ebooks™**

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

# Short-term and long-term prediction of coal consumption in Shandong Province- with ARIMA Model and Metabolism GM (1, 1) Model

Minglu Ma<sup>1</sup>, Hao Wu, Haopeng Wang and Yu Zhang

School of Economics and Management, China University of Petroleum, Qingdao 266580, China.

<sup>1</sup> Email: maminglu-117@hotmail.com

**Abstract.** We have witnessed the rapid growth of coal consumption in past decades. Shandong Province takes 5% of global coal consumption. This paper takes short-term (2016-2020) and long-term (2020-2030) influence into account to predict coal consumption in Shandong Province based on two linear models (ARIMA Model and Metabolism GM (1,1) Model). The MAPE (Mean Absolute Percentage Error) obtained by these two models are respectively 5.6% and 8% and the accuracy obtained by these two prediction models has exceeded 92%, which has reflected the high reliability of the predicted data. The prediction results also show that the coal consumption in Shandong Province will keep increasing with annual growth rate of 1.9% in the next 15 years. The accurate prediction to the coal consumption in Shandong Province will provide scientific basis for the Shandong Province and the country to adjust and formulate the energy supply and demand policies in the future.

## 1. Introduction

Shandong Province only occupies 1.6% jurisdiction area of China (about the size of Nepal), but its coal consumption takes up 10% of the whole country. In other words, its coal consumption accounts for 5% of the whole world (amount to 1/2 of that in Northern America or twice in the whole African Areas). Though governments and experts of various countries advocate the use of new energy, new energy cannot totally replace coal and petroleum as leading fuels at present. According to BP Statistical Review of World Energy, China, as the largest coal consumer country in the world, ranks the first place in the carbon dioxide emissions generated by coal combustion [1]. Under such background, taking short-term and long-term prediction to the coal consumption will better understand the tendency of carbon dioxide emission in the future and the predicted results will provide data support for the future energy structure adjustment in Shandong Province.

At present, the most commonly used models in prediction field are grey model and ARIMA model. In which, grey model, with the advantages of less required samples, simple calculation and high prediction accuracy, is widely applied in the short-term and long-term prediction of fields like administration, engineering and economy [2]. While ARIMA model has also been maturely applied in the prediction of various kinds of energy, for example, energy consumption prediction [3], coal consumption prediction, etc. because of its advantages of simple structure, fast modeling speed and high prediction accuracy [4]. To enhance the prediction accuracy of coal consumption in Shandong Province, the study has simultaneously adopted the two linear models, which can effectively reduce the prediction error. The MAPE (Mean Absolute Percent Error) is used to reflect the prediction



credibility in the study. The prediction results obtained from the combined use of the two models will provide reference basis for clearly mastering the future tendency.

The organization structure of the paper is as follows: Section 2 is the review of the relevant energy literature and present energy prediction methods; Section 3 is the introduction of the method using two linear models (ARIMA Model and Metabolism GM (1,1) Model) in the study; Section 4 is the presentation and discussion of the research results; Section 5 is the summary of the paper.

## 2. Literature review

In recent years, many international experts and scholars have studied from various aspects and used different models to study the coal consumption. Internationally: Challi Nondo, et al. studied the coal consumption in less developed countries. They selected the data of the year 1980-2005 from 19 African countries to verify the relationship between coal consumption and economic growth in these countries and the results showed that: for quite a long time, there was bidirectional causality between these two for these less developed countries; M Shahbaz [5] took the data of the year 1965-2008 as basis, adopted error correction method to study the impact on carbon dioxide emission by financial development, economic development and coal consumption in South Africa and the results showed that coal consumption had caused the environmental degradation in South Africa; O Ocal [6] used asymmetric causality test to test the coal consumption with GDP in Turkey and the results showed that the coal consumption had not influenced the economic growth. SG Chavez [7] used ARIMA to analyze the data in year 1980-1996 and predicted the energy consumption data in 1997-1998 of Asturias (Northern part of Spain); G Aydin [8] used GM (1,1) to predict the primary energy consumption in Turkey in 2013-2020; Yang H Y. A [9] applied Granger's technique to analyze the relationship between energy (coal, petroleum, etc.) and GDP in Taiwan area and discovered that there was bidirectional causality between the total energy consumption and GDP; Wolde-Rufael Y. [10] also took similar study with Shanghai as example; in China, there are also many scholars who have made some research on the regional coal consumption: Zhao et al. researched and analyzed the issue on coal consumption and economic growth in Shandong Province through SPSS and EViews software; Wang Hongjie applied the combined method of department analysis method and coal consumption elastic coefficient method to analyze the mid and long-term coal demand tendency in Fujian Province based on the coal consumption in 1996-2003 and the results showed that the coal demand would present substantial growth in the next 5-10 years; Ma Guoqi used ARIMA model to predict the coal consumption in Jiangsu Province in 2011-2020 with the energy consumption in Jiangsu in 1985-2010 as basis and the results showed that the energy supply rate in Jiangsu Province would be further reduced if the energy policy was not adjusted; Du et al. [11] used grey GM (1,1) to predict the coal consumption in Anhui Province during 2005-2011 and the results showed that GM (1,1) model could reasonably reflect the growth tendency of coal consumption in Anhui Province and the prediction accuracy was very high.

Many experts and scholars have solved some practical problems by using ARIMA Model and GM (1,1), for example: A Montanari [12] used ARIMA Model to predict the water flow rate of Nile River in Aswan; Wang Yaozhong [13] established ARIMA model according to the coal consumption data in Hunan Province during 1982-2010 and predicted the coal consumption in 2013-2015. The prediction effect was good and the MAPE during 2005-2010 was only 4%. It can be seen from above research literature that ARIMA Model fits well the analysis and prediction of various kinds of time sequence with strong randomness in real life. While, GM (1,1) used in the paper to predict the long-term coal consumption in Shandong Province was put forward by Deng Julong in 1982. Later it was developed and has been widely applied in numerous fields like industry, agriculture, society and energy, which has successfully solved the vast practical problems in the industry, production and scientific research.

From above literature review, we can discover that most of the researches focus on the relationship between coal consumption and the influence factor, economic growth and industrial structure. Though the researches on China's coal consumption have also involved different areas and are related with coal consumption prediction, very few are about the prediction of coal consumption in Shandong

Province. In view of the very huge coal consumption in Shandong Province, it is of great significance to predict it. In addition, we can see from above literature that ARIMA Model and GM (1,1) are used frequently in coal prediction, which shows the high stability and reliability of the data predicted by the two models; many people may use one model to predict the coal consumption within one time period, but this paper will use the latest data to respectively take short-term and long-term prediction to the coal consumption in Shandong Province through both ARIMA and Metabolism GM (1,1) so as to obtain more accurate predicted value.

### 3. Methodology

#### 3.1. Data

In this paper, the time series data during 1995-2015 have been selected. To guarantee the data reliability, the data source on the coal consumption in Shandong Province adopted by the paper is from the Statistical Yearbook of Shandong Province issued by Shandong Provincial Bureau of Statistics. These data will be used to take short-term prediction to the coal consumption in Shandong Province during 2016-2020 through ARIMA Model and long-term prediction to the coal consumption in Shandong Province during 2020-2030 through Metabolic GM (1,1) model.

#### 3.2. Model theory

**3.2.1. ARIMA Model.** Many experts and scholars have used ARIMA Model to take prediction to many industries and the reason behind it is that it has many advantages, such as: simple structure, quick modeling speed and high prediction accuracy [14]; in addition, the processing of non-stationary series time series has been solved [15].

The ARIMA Model used in this paper to predict the coal consumption in Shandong Province is based on the generalized ARIMA (p, d, q) equation and it is composed of 3 parts:

Step 1: Make the series pass through d difference stably and turn the non-stationary time series into stationary time series. This stationary time series can be represented as:

$$X_t^* = (1 - B)^d X_t \quad (1)$$

Step 2: Auto-regression (AR) process: the explained variable can be taken as the function of previous data and current data, which is also called their regression. Its model can be written as:

$$X_t^* = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_K X_{t-K} + \varepsilon_1 \quad (2)$$

Where:  $\alpha_i$  is the coefficient;  $X_t^*$  is the explained variable;  $X_{t-K}$  is the previous data;  $\varepsilon_1$  is the error;

Step 3: Moving average (MA) process: during this process, the explained variable is taken as the function of current error and previous error and it can be expressed as follows by mathematic model:

$$X_t^* = u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \dots + \beta_q u_{t-q} \quad (3)$$

Where:  $\beta_1$  is the co-efficient;  $u_{t-q}$  is the error;  $X_t^*$  is the explained variable

Through the regression to previous data, current data and error, we can fit the data during 1995-2015. Through the comparison with original data, we can obtain the MRE and the prediction accuracy, which can be finally applied for prediction.

**3.2.2. Metabolism GM (1,1).** From the prediction aspect, metabolism GM (1,1) is an ideal model because with the continuous update of coal consumption, some old data can no longer reflect the present coal consumption features. Only by supplementing new data at any time and removing old data that the coal consumption can be reflected more reasonably and accurately. The original data series is:

$$X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)) \quad (4)$$

The cumulative data series is  $X^{(1)}$ , and the data series after once accumulation is:

$$X^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)) \quad (5)$$

In addition,  $X^{(1)}$  shall satisfy the first order differential equation:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u \quad (6)$$

In above equation,  $a, u$  are unknown and they are solved by following matrix:

$$A = \begin{bmatrix} a \\ u \end{bmatrix} = (B^T B)^{-1} B^T Y_N \quad (7)$$

$$Y_N = [x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(m)]^T \quad (8)$$

$$B = \begin{bmatrix} -(0.5x^{(1)}(1) + x^{(1)}(2)) & 1 \\ \vdots & \vdots \\ -(0.5x^{(1)}(m-1) + x^{(1)}(m)) & 1 \end{bmatrix} \quad (9)$$

The solution of differential equation (1) is:

$$\hat{x}^{(1)}(k) = \left[ x^{(0)}(1) - \frac{b}{a} \right] e^{-a(k-1)} + \frac{b}{a}, \quad k = 1, 2, \dots, n \quad (10)$$

$$\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1) = \left[ x^{(0)}(1) - \frac{b}{a} \right] (1 - e^a) e^{-a(k-1)}, \quad k = 2, 3, \dots, n \quad (11)$$

**Table 1.** Predicted value, absolute error and relative error during 1995-2015.

| Year | Coal Consumption (Tce) | ARIMA Model            |                |                | Metabolism GM (1,1)    |                |                |
|------|------------------------|------------------------|----------------|----------------|------------------------|----------------|----------------|
|      |                        | Prediction Value (Tce) | Absolute Error | Relative Error | Prediction Value (Tce) | Absolute Error | Relative Error |
| 1995 | 0                      | 7796.9                 | 0              | 0              | 7796.9                 | 0              | 0              |
| 1996 | 7940.3                 | 7940.3                 | 0              | 0              | 7962.017               | 21.717         | 0.002735       |
| 1997 | 7930.4                 | 8425.197               | 494.797        | 0.0623924      | 7908.059               | -22.340        | 0.002817       |
| 1998 | 7875.6                 | 8437.349               | 561.748        | 0.0713277      | 7854.467               | -21.132        | 0.002683       |
| 1999 | 7779.5                 | 8363.864               | 584.363        | 0.0751158      | 7801.238               | 21.738         | 0.002794       |
| 2000 | 7392.0                 | 8263.299               | 871.298        | 0.1178704      | 7748.369               | 356.369        | 0.048210       |
| 2001 | 8955.3                 | 7870.543               | -1084.757      | 0.1211301      | 7329.474               | -1625.826      | 0.181549       |
| 2002 | 10738.9                | 10127.205              | -611.695       | 0.0569606      | 8772.098               | -1966.801      | 0.183147       |
| 2003 | 12694.9                | 12419.214              | -275.686       | 0.0217163      | 11778.684              | -916.215       | 0.072171       |
| 2004 | 14896.7                | 14434.558              | -462.141       | 0.0310230      | 15172.829              | 276.129        | 0.018536       |
| 2005 | 26056.2                | 16631.828              | -9424.371      | 0.3616940      | 17610.527              | -8445.673      | 0.324132       |
| 2006 | 29763.7                | 31417.055              | 1653.355       | 0.0555493      | 33101.552              | 3337.852       | 0.112145       |
| 2007 | 32718.7                | 33597.875              | 879.175        | 0.0268707      | 41146.193              | 8427.493       | 0.257574       |
| 2008 | 34389.6                | 35248.751              | 859.151        | 0.0249828      | 42218.838              | 7829.238       | 0.227662       |
| 2009 | 34795.2                | 35688.970              | 893.770        | 0.0256865      | 38230.833              | 3435.633       | 0.098738       |
| 2010 | 37408.5                | 37623.395              | 214.895        | 0.0057445      | 37241.123              | -167.376       | 0.004474       |
| 2011 | 36470.3                | 38829.496              | 2359.196       | 0.0646881      | 38618.541              | 2148.241       | 0.058903       |
| 2012 | 37077.3                | 38302.730              | 1225.43        | 0.0330507      | 38021.926              | 944.626        | 0.025477       |
| 2013 | 37683.4                | 37630.904              | -52.495        | 0.0013930      | 37913.943              | 230.543        | 0.006117       |
| 2014 | 39561.7                | 39234.947              | -326.752       | 0.0082593      | 37525.206              | -2036.494      | 0.051476       |
| 2015 | 40926.9                | 40800.972              | -125.927       | 0.0030768      | 40252.026              | -674.873       | 0.016489       |

Therefore, based on  $d = 2$ , the  $p$  and  $q$  values can be obtained from the correlation coefficient diagram. We can obtain a maximum  $R$  value from ARIMA (1,2,5). Through the equation of ARIMA (1,2,5) and Metabolism GM (1,1), we can respectively obtain the predicted value, absolute error and relative error during 1995-2015 and the specific data are shown in the table above:

Table 1 shows the predicted value, absolute error and relative error during 1995-2015 calculated through ARIMA (1,2,5) and Metabolism GM (1,1) equation. From the data in the form we can see that the error predicted by ARIMA (1,2,5) was only 5.6%, while the error predicted by Metabolism GM (1, 1) was 8%. The error of the two prediction methods is less than 10%, which shows that they have good fit effect and it means that the method we selected is correct and effective.

#### 4. Results and discussion

In this part, we will demonstrate and discuss the predicted results through ARIMA Model and Metabolism GM (1,1).

**Table 2.** Prediction results (Unit: Tons of Standard Coal).

| Year | ARIMA (1,2,5) | Metabolism GM (1,1) |
|------|---------------|---------------------|
| 2016 | 42204.3048    | 42290.3007          |
| 2017 | 42805.4409    | 43783.5318          |
| 2018 | 43300.3736    | 45329.4876          |
| 2019 | 43247.6136    | 46930.0296          |
| 2020 | 42973.4907    | 48587.0854          |
| 2021 | -             | 50302.6500          |
| 2022 | -             | 52078.7899          |
| 2023 | -             | 53917.6434          |
| 2024 | -             | 55821.4253          |
| 2025 | -             | 57792.4279          |
| 2026 | -             | 59833.0247          |
| 2027 | -             | 61945.6731          |
| 2028 | -             | 64132.9173          |
| 2029 | -             | 66397.3908          |
| 2030 | -             | 68741.8209          |

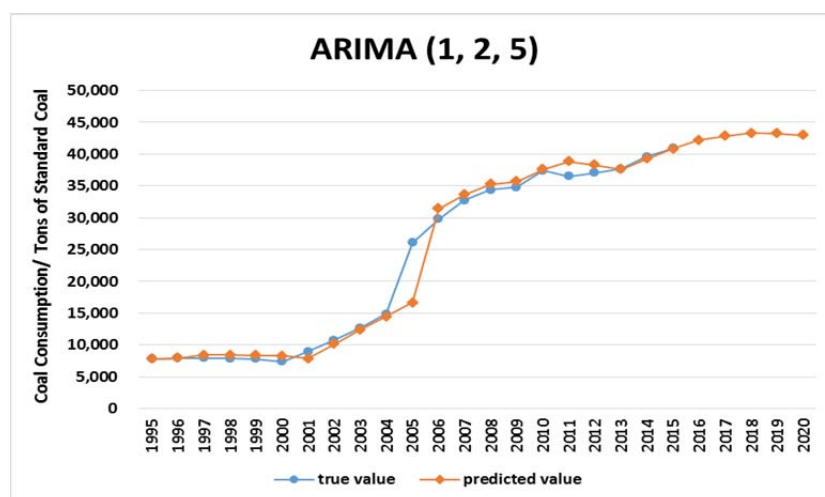


**Figure 1.** Short-term forecast trend.

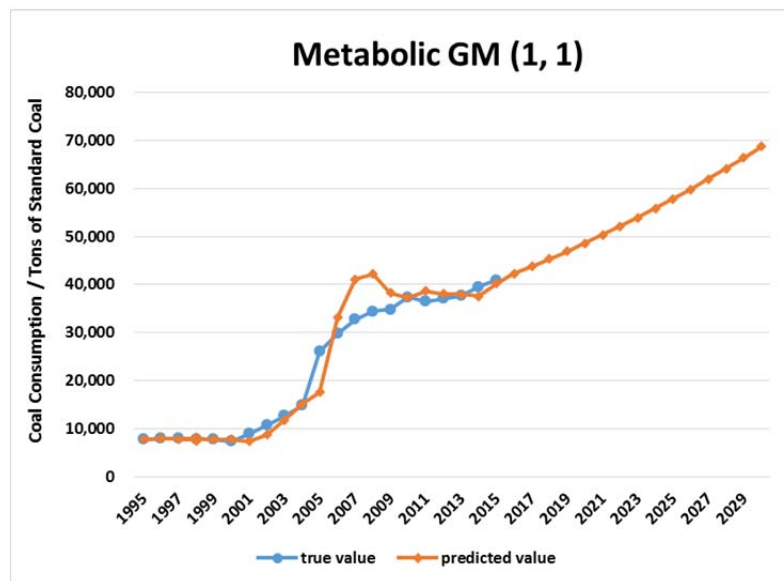


**Figure 2.** Long- term forecast trend.

Through calculation, we can obtain that the error of ARIMA (1,2,5) is 5.6% and that of Metabolism GM (1,1) is 8%. Both of the accuracy obtained by the two models is higher than 92%, so the predicted results are reliable.

This paper has used ARIMA (1,2,5) and Metabolism GM (1,1) to respectively take short-term and long-term prediction to the coal consumption in Shandong Province. From the predicted results of ARIMA (1,2,5) in Table 2, we can tell that the coal consumption in Shandong Province during 2016-2018 is still on the increase; in 2019-2020, though the coal consumption is reduced slightly, the amplitude is small and the coal consumption is in slow growth on the whole, with the annual growth rate of 0.45%. This can be more clearly seen in the short term tendency chart of Figure 1. Moreover, the long term predicted results obtained through Metabolism GM (1,1) shows that the coal consumption is keeping growth trend during 2020-2030 and the coal consumption will reach 68741.8209 Tons of Standard Coal in 2030 and the average annual rate of growth during 2020-2030 is 4.46%. The growth tendency can be more clearly seen through Figure 2.

## 5. Conclusions

This research is to take short-term and long-term prediction to the coal consumption in Shandong Province through ARIMA (1,2,5) and Metabolism GM (1,1). After using ARIMA (1,2,5) to fit the data of the year 2016-2020, it got the error of 5.6%; after using Metabolism GM (1,1) to fit the data of the year 2020-2030, it got the error 8%. Both of the errors are less than 10%, so it means that the short-term and long-term prediction to the coal consumption in Shandong Province is correct and reliable. The prediction results show that though there is small fluctuation for the coal consumption in Shandong Province during 2016-2020, the total consumption will still keep growing with the annual growth rate of 0.453%. During 2020-2030, the annual growth rate of coal consumption in Shandong Province will be 4.46%, which means that coal will still be the main fuel for Shandong Province during 2016-2030. On the other side, these two models have some imperfect in prediction. The time series prediction model is only good at the simulation of the sequence itself. However, the future trend of a sequence is not only influenced by the historical data itself, but also by other variables, this is the part that needs improvement in future predictions.

The accurate prediction of future energy consumption in Shandong Province will enable the energy planners of Shandong Province to take more effective and accurate planning to the future energy policies in Shandong Province and try their best to maintain the sustainable development of energy

and environment in Shandong Province. The prediction results obtained from the study are identical to the present development tendency of Shandong Province, which will also provide reference basis for the carbon emission reduction policymakers.

## References

- [1] BP 1965-2017 *Statistical Review of World Energy*
- [2] Yi Z 2007 *Systems Engineering-Theory & Practice* **27** 141-146
- [3] Wang Q, S. Li, and R. Li 2018 *Energy* **161** 821-831
- [4] Wang Q et al, 2018 *Energy* **160** 378-387
- [5] Shahbaz M., A.K. Tiwari, and M. Nasir 2013 *Energy Policy* **61** 1452-1459
- [6] Ocal, O. and A. Aslan 2013 *Renewable & Sustainable Energy Reviews* **28** 494-499
- [7] Chavez, S.G., J.X. Bernat, and H.L. Coalla 1999 *Energy* **24** 183-198
- [8] Aydin G 2015 *Energy Sources Part B Economics Planning & Policy* **10** 176-182
- [9] Yang H.Y. 2000 *Energy Economics* **22** 309-317
- [10] Menyah K. and Y. *Energy Economics* **32** 1374-1382
- [11] Shuai W, Zhijun L, Shouxu J, 2014 *Dependable, Autonomic and Secure Computing (DASC), 2014 IEEE 12th International Conference on* 344-349
- [12] Montanari, A. R. Rosso, and M.S. Taqqu 2000 *Water Resources Research* **36** 1249-1259
- [13] Shuai W, Kim S, Tian H 2018 *IEEE 38th International Conference on Distributed Computing Systems (ICDCS)* 500-510
- [14] Zhaoxu J, Shuai W 2017 *Technology, Networking, Electronic and Automation Control Conference (ITNEC)* 999-1003
- [15] Kai L, Shuai W 2017 *International Conference on Big Data Analytics and Knowledge Discovery* 471-485