
# Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 13 (15): 3055-3060, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.3055.3060


# Prediction of China's Coal Price During Twelfth Five-year-plan Period Based on ANN and RNM-BCC-LS-SVM Method

Zeng Ming, Wang Yi, Xue Song and Wang Zhijie
School of Economics and Management, North China Electric Power University,
Beijing, 102206, China

**Abstract:** Coal resource plays a significant role in primary energy production and consumption in C hina. Thus, coal price has a great influence on national economy, it is very meaningful to predict coal price. However, because of limited data availability, further study is requied to investigate the precision and reasonability of those methods such as multiple regression method, dynamic analysis model, neural network method and so on. In this study, we use Artificial Neural Network (ANN) to select key influencing factors of coal price. And then, we introduce the Bacterial Colony Chemotaxis (BCC) algorithm based on random Nelder Mead (RNM) to determine the extra-parameters used in least squares-support vector machine (LS-SVM) for coal price prediction rapidly and reasonably. At last, a case study of D atong premium blend coal at Qinhuangdao port is presented predicting its coal price during the twelfth Five-year-planning. Compared with the prediction results of ANN and BCC, the suitability and novelty of ANN and RNM-BCC-LS-SVM is fully demonstrated.

**Key words:** Coal price prediction, ANN, RNM-BCC-LS-SVM, datong premium blend coal at Q inhuangdao port

## INTRODUCTION

As we all know, coal price basically depends on its supply and demand and is also influenced by many other factors. From the perspective of realistic value, it is very essencial to predict Chinese coal price and understand the related dynamic changes.

Many investigations on coal forecast are associated with the supply and demand of coal (Cattaneo *et al*., 2011), the economic analysis (Zaklan *et al*., 2012) and envrionmental protection analysis (Bloch *et al*., 2012) and so on. Various methods have been employed to the prediction of coal price, such as multiple regression method (Dahl, 2012), dynamic analysis model, neural network method and the like. But further study is requied to investigate the precision and reasonability of those methods due to limited data availablity. Therefore, the ANN and RNM-BCC-LS-SVM method is introduced to predict the coal price in this study.

In this study, we first get the influencing factors of coal price using artificial neural network and then establish the prediction model of coal price based on random nelder mead-bacterial colony chemotaxis-least squares-support vector machine. It is should be noted that bacterial colony chemotaxis algorithm is optimized using random nelder mead algorithm avoiding local optimal value. Finally, a case study is presented getting the price of Datong premium blend at Qinhuangdao port during the twelfth Five-year-plan in China. The suitability and novelty of the method used in this study is also demonstrated.

## COAL PRICE INFLUENCING FACTORS SELECTION BASED ON ANN

With relevant government's encouragement in China, a series of market reforms have been implemented in coal market. In 2007, the model that coal ordering fair organized by the government which had lasted for over 50 years was eliminated. Consequently, coal demand and supply became the key factors affecting coal price which indicated that market liberation had been successfully introduced into coal price in China Yuan *et al*. (2008). With new coal ordering reforms adopted previously, market reforms has been explored further to achieve the aim of independent consultant pricing between providers and consumers and to form the coal price mechanism that reflects the coal supply and demand, shortage of energy resources and environmental costs as soon as possible ever since 2008. Nowadays, the mechanisms of coal production, distribution and market-oriented pricing have been established virtually. Therefore, main factors that affect coal price include coal demand and supply, international energy price, transportation costs and production costs (Fig. 1).

**Corresponding Author:** Wang Zhijie, School of Economics and Management,North China Electric Power University, 102206, Beijing, China Tel: 010-51963851 Fax: 010-51963851

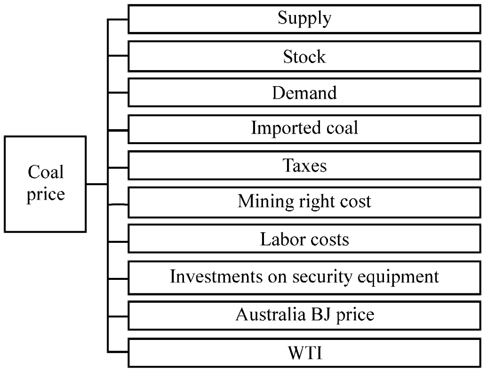


Fig. 1: Influencing factors of coal price

Artificial Neural Network (ANN) is a number of artificial neurons connecting with each other simply applied to simulate the biology neural network and then obtain information from outer environment or other neural. After simple calculations, outputs are transported back to outer environment or other neural. Back Propagation (BP) error back propagation learning algorithm with two processes: forward propagation of information and backward propagation of error, which is the most representative and widely used in ANN. So here we choose BP to find out the factors affecting the coal price.

## PREDICTION MODEL OF COAL PRICE BASED ON RNM-BCC-LS-SVM OPTIMIZATION METHOD

**Model of least squares-support vector machine:** Support Vector Machine (SVM) is specialized in solving actual problems characterized by small samples, non-linear features, high dimensions and local minimum values and has been successfully applied to deal with categories, function approaching, time series forecasting and so on. Least Squares-Support Vector Machine (LS-SVM) proposed by Suykens is viewed as the improvement of standard SVM with a lot of advantages: the non-equation constraints in standard SVM are replaced by equation ones and quadratic programming problem is converted to solving linear equations directly (Li *et al*., 2011). Given a set:

$$\{x_i, y_i\}_{i=1}^{N}$$

where, $x_i \epsilon F^N$, is the inputs and $y_i \epsilon F^N$ is the corresponding outputs. The non-linear function f(.) is aimed at converting the samples to the characteristics space. The model of LS-SVM can be described as (Zhong *et al*., 2005):

$$h(a_i) = G^T f(x_i) + m \quad (1)$$

where, G and m are parameters to be examined. G and m can be obtained by minimizing the following objective function.

$$F = 0.5\|G\|^2 + \xi.F_1 \quad (2)$$

where, $\xi$ is regulation factor; $F_1$ is loss function. The optimization problem can be described as:

$$\begin{cases} \min Z(G, d_i) = 0.5\|G\|^2 + 0.5\zeta\sum_{i=1}^{N} d_i^2 \\ s.t. y_i = G^T f(x_i) + m + d_i, i = 1, 2, ..., N \end{cases} \quad (3)$$

The Lagrange function of 12 is expressed as:

$$L(G, \lambda_i, m, d_i) = Z + \sum_{i=1}^{N} \lambda_i [y_i - G^T f(a_i) - m - d_i] \quad (4)$$

where, $\lambda_i = 0$, is the Lagrange mutiplier; $d_i$ is the error. According to Karush-Kuhn-Tucker conditions, the following equations can be obtained.

$$\partial L/\partial G = 0, \partial L/\partial \lambda_i = 0, \partial L/\partial m = 0, \partial L/\partial d_i = 0 \quad (5)$$

Then:

$$G = \sum_{i=1}^{N} \lambda_i f(x_i), \quad \sum_{i=1}^{N} \lambda_i = 0, \lambda = \zeta d_i \quad (6)$$

$$y_i - G^T f(x_i) - m - d_i = 0 \quad (7)$$

Linear equations are obtained after eliminating G and $d_i$:

$$\begin{bmatrix} 0 & 1 & L & 1 \\ 1 & k(x_1, x_1) + \frac{1}{r} & L & k(x_1, x_N) \\ M & M & M & M \\ 1 & k(x_N, x_1) & L & k(x_N, x_N) + \frac{1}{r} \end{bmatrix} \begin{bmatrix} m \\ \lambda_1 \\ M \\ \lambda_N \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ M \\ y_N \end{bmatrix} \quad (8)$$

$\lambda_i$ and m can be obtained by calculating these linear equations. The model associated with inputs x is described as:

$$\hat{y}(x) = \sum_{i=1}^{N} \lambda_i \varphi(x_i, x) + m \quad (9)$$

where, $\varphi(x_i, x) = f(x_i)^T f(x)$ is a kernel function subject to Phil Mercer theorem. In this study, RBF kernel function is adopted: $\varphi(x, x_k) = \exp(-\|x - x_k\|^2/2\mu^2)$.

It can be seen that kernel parameter ì and regulation factor $\zeta$ in Eq. 2 have a large impact on the accuracy of LS-SVM model. These two parameters can be obtained in the training period of RNM-BCC optimization algorithm.

**Parameters optimization with RNM-BCC:** Bacterial Colony Chemotaxis (BCC) algorithm which is developed from the Bacterial Chemotaxis (BC) algorithm, is a swarm intelligent algorithm that can realize the information communication among the individuals. Compared with genetic algorithm, BCC algorithm is characterized by its local search after each iteration and then a final overall optimal solution after a sufficient number of iteration. However, algorithms for local search of BCC seem simple and random. Therefore, it is wise to replace the algorithms with better local search algorithms to strengthen the performance of algorithms. Random Nelder-Mead simpex (RNM) is an improved algorithm of Nelder-Mead simplex proposed by Nelder and Mead. It can make the search direction more precise and locate the potential solutions around the search direction. RNM adopts direct search strategy which is insensitive to the initial value and does not require the objective function to be continuous or differentiable. Hence, RNM rather than the usual algorithms in BCC algorithm can be utilized in local search. Compared with BCC algorithm, RNM-BCC is effective and promotes the overall accuracy. The calculation process of RNM-BCC is presented as Fig. 2.

**Prediction model of coal price based on ANN and RNM-BCC-LS-SVM method:** Steps of ANN and RNM-BCC-LS-SVM method are presented in Fig. 5.

Prediction model of coal price based on ANN and RNM-BCC-LS-SVM is established as follows:

**Step 1:** Establish regression function. Optimal regression Eq. 1 is established according to non-linear function. Inputs of the SVM are coal price key influencing factors obtained by ANN while the output is the coal price

**Step 2:** Determine the kernel function and parameters. RBF kernel function is adopted in this study and regulation factor and kernel parameters are obtained using RNM-BCC-LS-SVM method

**Step 3:** Establish the prediction model of coal price. The regression optimization problem and corresponding constraints Eq. 3 are firstly created. Optimal coefficients of the regression function are obtained using Lagrange function. Then get the coal price prediction model shown as Eq. 9

**Step 4:** Predict the coal price. The model is actually a black box model. Coal price influencing factors are the inputs and the future coal price is obtained by simulation learning

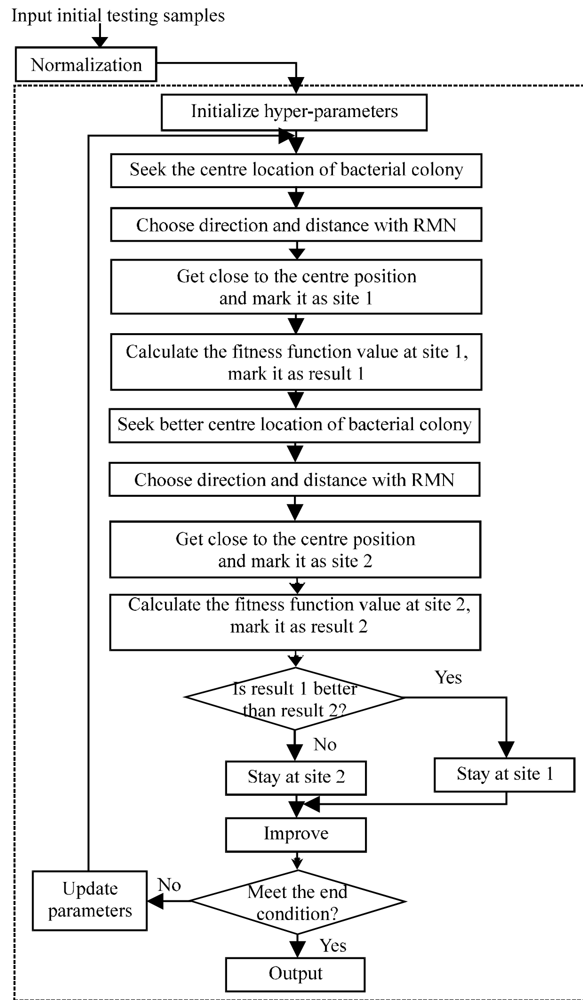


Fig. 2: The process of BCC algorithm optimized by RMN method

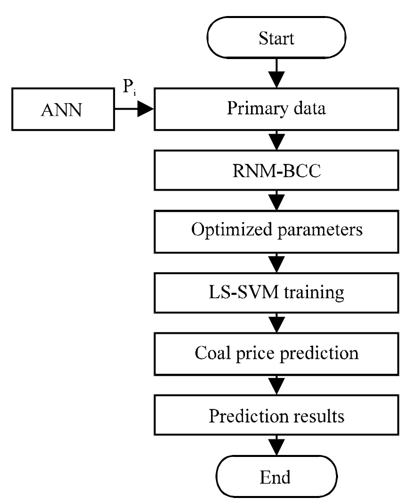


Fig. 3: Steps of ANN and RNM-BCC-LS-SVM method

## CASE STUDY

**Initial data:** It should be noted that initial data used in this study mainly come from energy market in China such as Coal market in China (http://www.cctd.com.cn), Energy Prices and Taxes Quarterly Statistics 2nd quarter 2011(http://www.iea.org) by the International Energy Agency, International oil economy as well as data from State Statistics Bureau and China Energy Statistic Yearbook. According to experts' suggestions and data availability, inputs of the prediction model include eight influencing factors: coal supply, coal demand, coal storage, coal imports and exports, taxes, transportation costs of coal, BJ price in Australia and WTI between 1995 and 2010; and the output involves coal price.

**Coal price influencing factors selection based on ANN:** Based on MATLAB 7.0 platform and ANN simulation analysis, the impact of each influencing factor on the error of training results under different smoothing factors is analyzed and compared. Combined with the weights generated in the pattern layer of ANN, influencing factors whose impact on the error of training results is great are selected as the inputs of the model. Four influencing factors are selected as inputs: coal supply, coal demand, coal imports and exports and WTI. Data from 2000-2010 are regarded as training sample and two continuous samples are chosen randomly as testing sample. Sample data are calculated under different smoothing factors. Error between predicted value and actual value are shown in Fig. 4, in which the horizontal axis denotes the sequence number of sample while the vertical axis denotes error. And σ is the smoothing factor. It can be seen from Fig. 4 that the error between predicted coal price and actual price is lowest when σ = 0.25.

**Data pretreatment and parameters determination:** Both inputs and outputs should be standardized to make all variables non-dimensional as follows:

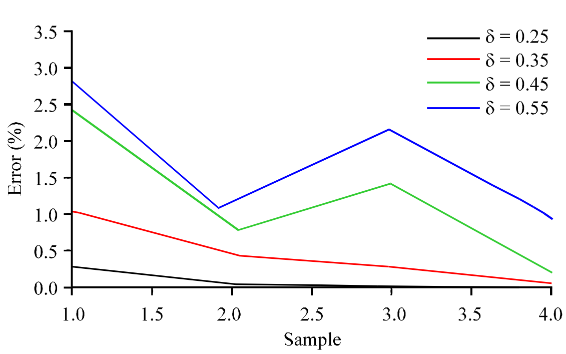


Fig. 4: Errors under different smoothing factors

$$x'_{ij} = \frac{x_{ij} - m_j}{M_j - m_j} \tag{10}$$

where, $x_{ij}$ is the primary data, $m_j$ is the minimum $M_j$ is the maximum, $x'_{ij}$ is non-dimensional and $x'_{ij}\in[0, 1]$.

Both inputs and outputs can be converted into numbers between 0 and 1 with Eq. 20. It should be noted that inputs are the influencing factors obtained by ANN method while the output is coal price.

Parameters are optimized with RNM-BCC algorithm. The population size is 20, the maximum iteration number is 500, the dimension is 2 and the trace back step number is 5. Fitness function is designed to calculate the error of every generation. Coal price prediction has been optimized and fitness of every iteration is shown in Fig. 5.

From Fig. 5, it is clear that the error of coal price prediction is lowest when the number of iteration exceeds 350. And at this time, ζ is 639.07683 and μ is 8.36746.

### Forecasting results

**Test on the prediction error based on sample learning:** Due to limited data, data between 1995 and 2000 are used to train while data from 2000-2010 are employed to test in this study and the forecasting results are shown in Fig. 6. Error is described as Eq. 11 and Mean Square Error (MSE) is selected as the evaluation index of the error (Eq. 12):

$$\varepsilon_i = \frac{(x'_i - x_i)}{x_i} \quad (i = 0, 1, 2, \cdots) \tag{11}$$

$$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})^2} \quad (n = 0, 1, 2, \cdots) \tag{12}$$

where, $\varepsilon_i$ is the error, $x_i$ the actual value, $x'_i$ is the corresponding predicted value, S is the MSE and $\bar{\varepsilon}$ is the average value of the error.

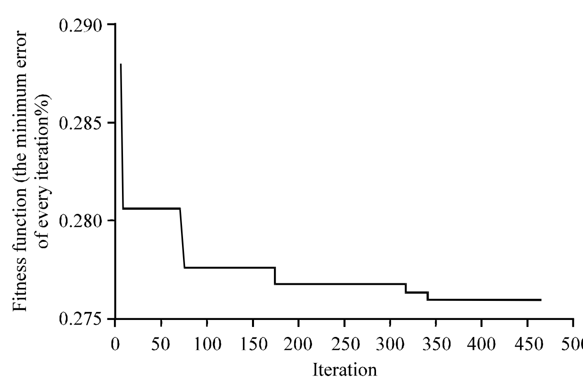


Fig. 5: Fitness function of coal price prediction with RNM-BCC algorithm

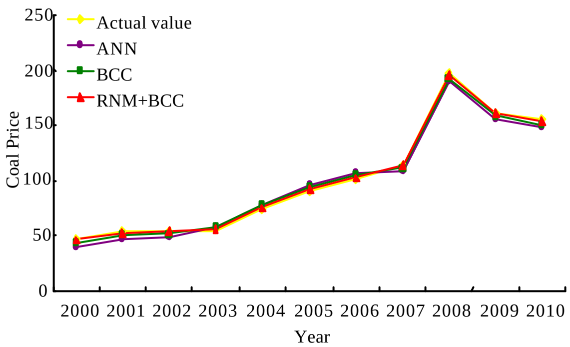


Fig. 6: Comparision of the predicted and actual values of various optimization methods between 2000 and 2009

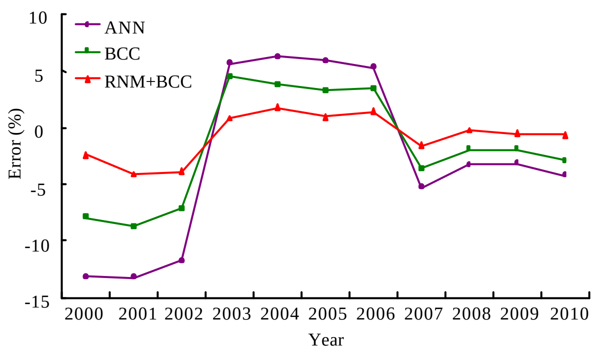


Fig. 7: The prediction errors of coal price of various optimization methods

The predicted and actual values under different optimized parameters methods between 2000 and 2010 are shown in Fig. 6. Obviously, RNM-BCC algorithm is better and more effective than BCC since the former generates a lower error (Fig. 6 and 7).

**Comparision of the error under different prediction methods:** To prove the effectiveness of the prediction method proposed in this study, MSE is designed as the evaluation index of the error, as shown in Eq. 30. Table 1 shows the MSE and errors of various methods. From Table 1, it can be seen that prediction MSE of ANN is 7.35% while the combination of ANN and RNM-BCC-LS-SVM enables the prediction MSE to be 1.02% which is much lower than others. Obviously, the prediction model of ANN and RNM-BCC-LS-SVM perfporms very well. Actually in this study, the influencing factors of coal price are compared with ANN before the LS-SVM is adopted. Then, key influencing factors are selected as the inputs and outputs of LS-SVM

Table 1: Comparision of the prediction errors of coal price of various optimization methods

| Methods | Average value of errors | MSE |
|---|---|---|
| ANN | -2.812 | 7.35 |
| BCC | -1.736 | 4.69 |
| ANN and RNM-BCC | -0.787 | 1.94 |

Table 2: The prediction of Datong premium blend coal price at Qinhuangdao port based on RNM-BCC-LS-SVM

| Time | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|
| Predicted values (US\$ toe$^{-1}$) | 180.70 | 190.70 | 210.40 | 230.00 | 242.30 |
| Predicted values (US\$ ton$^{-1}$) | 899.42 | 926.98 | 966.46 | 980.66 | 1028.19 |

method. Parameters are also optimized with RNM-BCC algorithm to reduce the error. The combination of ANN and RNM-BCC-LS-SVM has an advantage of accuracy over traditional methods. By comparing ANN with the combination of ANN and RNM-BCC-LS-SVM, errors from data itself can be effectively avoided. Therefore, the prediction model of ANN and RNM-BCC-LS-SVM perfporms better than other prediction methods.

**Forecast on coal price during the twelfth five-year-plan in China:** Prediction model of coal price established in this study is used to predict the Datong premium blend coal price at Qinhuangdao port. Forecasting results are shown in Table 2. It is found out that in November 2011, the price of Datong premium blend at Qinhuangdao with calorific value of 5800 Kcal/kg is 900-910¥/ton, indicating that prediction method proposed in this study is precise enough.

## CONCLUSION

Accurate prediction of coal price plays a significant role in promoting the sustainable development of national economy and contributes to decision-making for relevant industries. In this study, a novel prediction method based on ANN and RNM-BCC-LS-SVM was proposed to predict coal price accurately. With the novel prediction method, the average prediction error of coal price between 2000 and 2010 is much lower than other methods used in this study. In addition, factors which affect coal price are determined, contributing to master the regularity of coal price to some extent. Indeed, the suitability and novelty of the ANN and RNM-BCC-LS-SVM is fully demonstrated by predicting the price of Datong premium blend coal at Qinhuangdao port between 2010 and 2015 in this study. Results indicate that the coal price in 2015 will be 242.3US\$/toe, increasing by 31% compared with 184.3US\$/toe in 2010. During the twelfth Five-year-plan in China, the energy price would have an evident tendency to rise. Therefore, enterprises, especially those who depend largely on coal, should make unremitting efforts to promote production technologies, boost energy efficiency and thus reduce production costs.

## ACKNOWLEDGMENT

The study described in this study was supported by National Science Foundation of China (NSFC)(70671041, 70771039) and Energy Foundation of U.S (G-1006-12630).

## REFERENCES

Bloch, H., S. Rafiq and R. Salim, 2012. Coal consumption, $CO_2$ emission and economic growth in China: Empirical evidence and policy responses. Energy Econ., 34: 518-528.

Cattaneo, C., M. Manera and E. Scarpa, 2011. Industrial coal demand in China: A provincial analysis. Resour. Energy Econ., 33: 12-35.

Dahl, C.A., 2012. Measuring global gasoline and diesel price and income elasticities. Energy Policy, 41: 2-13.

Li, R., H. Su, Z. Wang, G. Deng and T. Chen, 2011. Medium-and long-term load forecasting based on heuristic least square support vector machine. Power Syst. Technol., 35: 195-199.

Yuan, J.H., J.G. Kang, C.H. Zhao and Z.G. Hu, 2008. Energy consumption and economic growth: Evidence from China at both aggregated and disaggregated levels. Energy Econ., 30: 3077-3094.

Zaklan, A., A. Cullmann, A. Neumann and C. von Hirschhausen, 2012. The globalization of steam coal markets and the role of logistics: An empirical analysis. Energy Econ., 34: 105-116.

Zhong, W.M., D.Y. Pi and Y.X. Sun, 2005. Support vector machine based nonlinear model multi-step-ahead optimizing predictive control. J. Cent. South Univ. Technol., 12: 591-595.