Forecasting of gas supply in self-provided power plant of iron and steel enterprises based on time series

Hua Meng 1,2,4 and Guofeng Fan 3

1 Chongqing Chemical Industry Vocational College, Chongqing 400020, China;
2 University Electronic and Technology of China, Chengdu, Sichuan 611731, China;
3 Pingdingshan University, Pingdingshan, Henan 467000, China.

4 Email: mhua08@163.com

Abstract. The self-provided power plants are the main buffer customers for iron and steel industry. They have great effect on reducing gas emission and improving enterprise efficiency. The ARMA and ARMA–ARCH models are used to short-term forecast respectively based on sampling data, in order to master the varying trend of gas supply in iron and steel enterprises of self-provided power plants. The actual comparison and application results indicate that the ARMA-ARCH model has a well-pleasing forecast performance with a Mean Abs. Percent Error (MAPE) of 3.32%, plays an important role in adjusting boiler switches frequently and optimal scheduling in actual production.

1. Introduction

By-product gas is an important secondary energy in iron and steel industry. The supply gas as the main buffer user, which can improve the reliability of the power supply, reduce the discharge of the gas, reduce the cost of electricity consumption [1-3]. Because the working conditions are different in iron and steel industry, the whole by-product gas system is very complicated. The amount of residual gas fluctuates with time frequently, so it is difficult to figure out the amount of residual gas flow. Therefore, in order to reasonably and accurately forecast the varying trend of by-product gas supply is necessary, which can use to controlling load level according to the production plan and ensure the maximization of by-product gas utilization. Besides, it will reduce boiler supply instability and the fluctuation caused by boiler frequent adjustment. This forecast can provide a guide for scheduling decision to keep production balance and ensure that the pipeline network pressure in a safe range.

Time series is applied to forecast varying tendency for the by-product gas supply of self-provided power plant because of the complex of the gas system. Time series analysis is a kind of data processing method to observe and analyse the random data, which is an effective tool to forecast in the lack of necessary information for the research object [4-5]. In recent years, many scholars have focus on the model establishment by using time series analysis method [6-12], and much effort which is given to the forecasting in generation and consumption of by-product gas [13-16]. However, the forecast varying trend of gas supply in self-provided power plant has less research. Therefore, ARMA and ARMA-ARCH model are established here and applied to the varying trend of gas supply in self-provided power plant. Furthermore, the analysis of residual error sequence is presented.
2. Built the model

Although there are many factors affecting the fluctuation of gas supply, the external performance is mainly reflected by the fluctuation of gas flow. The fluctuation of gas supply is a time-delay and random, so the randomness will weaken the relevance of the gas supply sequence. It is determined that the prediction of gas supply is difficult and can’t be predicted accurately, but it can only make the optimal prediction from the statistical significance, and finally make the mean square of the prediction error satisfy a certain precision requirement. Therefore, based on the change characteristics and the difficulty of predicting the supply of gas supply in power plant, combined with the applicability of various prediction methods, the time series prediction method is selected for short-term prediction.

The characteristic of the model is to study the smoothing error term, and can recognize the structure and characteristics of time series more essentially, and make it flat. The smoothing error reaches the best prediction under the minimum variance.

2.1. Model of ARMA

Time series model is a practical method based on linear model, which handle dynamic random data for parametric model [17]. A system, if the response is $X_t$ at time $t$, which is not only related to the previous value, but also exist a certain relation with disturbance, then the system is auto-regressive moving average model as ARMA (p, q), the general form as Eqn. (1):

$$X_t - \varphi_1 X_{t-1} - \cdots - \varphi_p X_{t-p} = \theta_1 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q}$$

(1)

Where $X_t$ is time series, $\{ \varphi_i \}$ $(1 \leq i \leq m)$ is autoregressive coefficients, $\{ \theta_j \}$ $(1 \leq j \leq n)$ is moving average coefficient, $\{ \epsilon_i \}$ is white noise sequence, $B$ is the backward shift operator. $\varphi(B) = 1 - \varphi_1 B - \cdots - \varphi_p B^p, \theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$.

It is assumed that $\varphi(B)$ and $\theta(B)$ have no common factor. Eqn. (1) can be indicated as Eqn. (2).

$$\phi(B)X_t = \theta(B)\epsilon_t$$

(2)

The best forecast value of first step at time $t$ in the forward for ARMA (p, q) model as Eqn. (3).

$$\hat{X}_t(1) = E[X_{t+1}] = E[\varphi_1 x_t + \varphi_2 x_{t-1} + \cdots + \varphi_p x_{t-p} + \epsilon_{t+1} - \theta_1 \epsilon_t - \theta_2 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q}]$$

(3)

$x_t, x_{t-1}, \cdots, x_{t-p}, \epsilon_{t+1}$ have determined at time $t$, and $\epsilon_{t+1}$ has not occurred, then $E[\epsilon_{t+1}] = 0$. So Eqn. (3) can be represented as Eqn. (4).

$$\hat{X}_t(1) = \varphi_1 x_t + \varphi_2 x_{t-1} + \cdots + \varphi_p x_{t-p} - \theta_1 \epsilon_t - \theta_2 \epsilon_{t-1} - \cdots - \theta_q \epsilon_{t-q}$$

(4)

Similarly, the best forecast value of second step at time $t$ in the forward for ARMA (p, q) model as Eqn. (5).

$$\hat{X}_t(2) = \varphi_1 \hat{X}_t(1) + \varphi_2 x_{t-1} + \cdots + \varphi_p x_{t-2-p} - \theta_1 \hat{X}_t(1) - \theta_2 \epsilon_t - \cdots - \theta_q \epsilon_{t+q-q}$$

(5)

The best forecast value of ARMA (p, q) model forward l-step can be obtained by analogy as Eqn. (6).

$$\hat{X}_t(l) = \begin{cases} \sum_{i=1}^{p} \varphi_i x_{t+l-i} - \sum_{j=1}^{q} \theta_j x_{t+1-j} & l = 1 \\ \sum_{i=1}^{p} \varphi_i \hat{X}_t(l-i) + \sum_{j=1}^{q} \varphi_j x_{t+l-i} - \sum_{j=1}^{q} \theta_j \hat{X}_t(l-j) - \sum_{j=1}^{q} \theta_j x_{t+1-j} & -1 < l \leq p, q \\ \sum_{i=1}^{p} \varphi_i \hat{X}_t(l-i) - \sum_{j=1}^{q} \theta_j x_{t+1-j} & l > q \end{cases}$$

(6)
2.2. Model of ARCH

Some time series usually show fluctuations phenomenon in a period of time, then maintain stable state relatively in the next period of time, and variance changes in time series is considered.

In a certain period of time, when the variance of time series changes with time, ARCH model can be used for modeling to random disturbance of time-series. The model equation is shown in Eqn. (7).

\[ y_t = X_t \beta + \epsilon_t \quad (7) \]

Where \( \epsilon_t = \sqrt{h_t} \nu_t \), \( \nu_t \) obey the normal independent distribution, \( h_t \) is the conditional variance of \( \epsilon_t \). \( h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 \), \( \alpha(B) \epsilon_t^2 \), \( \alpha(B) \) is lag operator polynomial, and nonnegative constraints \( \alpha_0 > 0, \alpha_i > 0, i = 1, 2, \cdots, q \). It shows that \( \{ \epsilon_t \} \) obey ARCH (q) process if satisfy the above conditions.

Judging the existence of ARCH effect in a time series by LM method, Auxiliary regression equation is established as Eqn. (8):

\[ h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 \quad (8) \]

ARCH effect is judged according to all regression coefficient of Eqn. (8) are zero or not at the same time. Test statistic is shown in Eqn. (9):

\[ \chi^2 = nR^2 \sim \chi^2(q) \quad (9) \]

Where \( n \) is the sample capacity, \( R^2 \) is the determination coefficient, the statistic LM is converged with \( \chi^2 \) distribution of \( q \) [18]. At last, the model parameters are estimated by the method MLE maximum likelihood estimation (MLE).

3. Case Study

There are 120t/h boiler, 180t/h boiler and 200t/h boilers for the iron and steel plant. The mixed gas for BOF and COD is 0-500000 m³/h. Taking the measured data of the supply gas as the historical data, the supply gas are defined that the production of the by-product of iron and steel enterprises in the statistical period (a certain time) reduce the consumption of the gas users in the main process.

3.1. Data Preprocessing

The actual data of mixed gas are selected for modeling from Aug. 5, 2016 to Oct. 12, 2016 in this paper. There are 1400 data after excluding some invalid data. The sample data are firstly pretreated by using the method of taking logarithm, and then treated by first order difference. The processed data are proved to be stable by ADF test [19].

3.2. ARMA Modeling

The varying trend of the autocorrelation function and partial autocorrelation function are shown in Figure 1. We can see that the autocorrelation function and partial autocorrelation function are not need difference, so we can’t establish the ARIMA model. The model of ARMA (p, q) is used to forecast because of tailing of the both parameters, where p is 1 or 2, q is 2 to 5. The parameters are estimated by using the method of nonlinear least squares, and the best model ARMA (2,3) is determined by considering the AIC (Akaiake Info Criterion), SC (Sehwarz Criterion), Adjusted R-squared and inverted AR roots in the unit circle or not. Those results of coefficient estimation are shown in Table 1 [18]. Compared to the other order model, the effect of fitting is better at the adjusted R-squared of 0.873, and the accuracy of this model is also explicated with the help of AIC and SC criteria. The model of ARMA is shown in Eqn. (10) and Eqn. (11) based on Eqn. (2), respectively.

\[ \varphi(B) = 1 - 0.3311B^{-1} - 0.9321B^{-2} \quad (10) \]
\[
\theta(B) = 1 + 0.2315B^{-1} - 0.5267B^{-2} + 0.6723B^{-3}
\]  
(11)

Table 1. Coefficient estimation of ARMA (2,3) model.

| Model   | Adjusted R-squared | AIC    | SC    |
|---------|--------------------|--------|-------|
| ARMA(1,4) | 0.8138             | 0.7613 | 0.8432 |
| ARMA(2,3) | 0.8703             | 0.7089 | 0.8023 |
| ARMA(2,4) | 0.8308             | 0.7332 | 0.8310 |
| ARMA(2,5) | 0.8126             | 0.7851 | 0.8735 |

Figure 1. Summary of autocorrelation and partial correlation.

3.3. ARCH Effect and Modelling
ARCH effect is judged by the LM test for the residuals error of ARMA (2, 4) model. The LM test results of ARCH (1) and ARCH (2) are shown in Table 2.

Table 2. LM test results of ARCH effect.

| Model   | Statistic | Result  | Prob.     |
|---------|-----------|---------|-----------|
| ARCH(1) | F-statistic | 22.0452 | 0.000005 |
|         | Obs*R-squared | **20.3193** | **0.000007** |
| ARCH(2) | F-statistic | 11.3746 | 0.000019 |
|         | Obs*R-squared | 20.9856 | 0.000028 |

*a denotes that the LM test results of ARCH (1) and ARCH (2) model are 20.3193 and 20.9856, respectively. The probability P of \( \chi^2 \) test is less than the significance level \( \alpha = 0.05 \) obviously.

Furthermore, it is obvious that the residual error series of ARCH (1) model is significant, which is presented as the best model. Parameters estimation and significance test results of ARCH model are shown in Table 3.
ARMA (2,3)-ARCH (1) is selected as the final forecast model, which can be indicated as Eqn. (12) and Eqn. (13).

Part of ARMA:
\[ y_t = -0.3311y_{t-1} - 0.9321y_{t-2} + 0.2315\epsilon_{t-1} - 0.5267\epsilon_{t-2} + 0.6723\epsilon_{t-3} \] (12)

Part of ARCH:
\[ h_t = 0.0683 + 0.2689\epsilon_{t-1}^2 \] (13)

**Table 3. Parameters estimation and significance test results of ARCH model**

| Model   | Variable | Coefficient | Prob.  |
|---------|----------|-------------|--------|
| ARCH(1) | C        | 0.083593    | 0.0000 |
|         | Resid^2(-1) | 0.293523    | 0.0000 |
|         | C        | 0.078967    | 0.0000 |
| ARCH(2) | Resid^2(-1) | 0.275423    | 0.0000 |
|         | Resid^2(-2) | 0.060562    | 0.3564a |

*a denotes that it is reasonable to refuse the original assumptions that the regression equation coefficient is zero by LM test, but the coefficients of ARCH (2) model is not pass significant test for the confidence of 95%.

### 3.4. Trend forecast

ARMA model and ARMA-ARCH model are used for forecasting the mixed gas from Oct.13, 2016 to Oct.18, 2016 according to the optimal order and parameter estimation calculated above. In comparison of the actual and forecast result, the applicability and forecast ability of the model can be confirmed as shown in Figure 2.

![Figure 2. The forecast results for variation trend.](image)

From Figure 1, it can be seen that the maximum prediction error of the model is 6.3%, and the prediction accuracy is good, but there is a certain gap. It is mainly due to the complexity of the gas
system in the iron and steel enterprises, and it is difficult for the gas supply to occur in various working conditions and events.

**Table 4.** Comparison of the forecasting abilities.

| Model               | MAPE/% | Bias proportion | Variance proportion |
|---------------------|--------|-----------------|---------------------|
| ARMA(2,3)           | 8.87   | 0.00012         | 0.1839              |
| ARMA(2,3)-ARCH(1)   | 3.32   | 0.000089        | 0.1428              |

Comparison of the forecast capability based on ARMA (2, 3) and ARMA (2, 3)-ARCH (1) is shown in Table 4 and Figure 1. The conclusions are as follows:

1. The MAPE indexes of two models are in the range of 3% to 9%, ARMA-ARCH model is better than ARMA model for the system of complex gas in iron and steel industry.
2. The Bias proportion of the ARMA-ARCH model is much smaller than the ARMA model, indicating that the difference between the forecasting mean value and the actual mean value are small.
3. Variance proportion of the ARMA-ARCH model is smaller than the ARMA model, mainly because the variance of the gas supply is changed, and the ARCH model is put forward to solve this kind of problem, which shows that the modeling results are in agreement with the reality.

**Table 5.** Wilcoxon signed-rank test.

| Compared models          | Wilcoxon signed-rank test |
|--------------------------|---------------------------|
|                          | \( \alpha = 0.025; W = 4 \) | \( \alpha = 0.05; W = 6 \) |
| ARMA(2,3)-ARCH(1) vs. AR(3) | 8                         | 3\(^a\)                         |
| ARMA(2,3)-ARCH(1) vs. ARMA(2,3) | 8                         | 0\(^a\)                         |
| ARMA(2,3)-ARCH(1) vs. ARMA(2,3)-ARCH(2) | 6                         | 2\(^a\)                         |

\(^a\) denotes that the ARMA(2,3)-ARCH(1) model significantly outperforms other alternative models.

Furthermore, to verify the significance of the accuracy improvement of the ARMA(2,3)-ARCH(1) model, the forecasting accuracy comparisons in both Cases among AR(3), ARMA(2,3) and ARMA(2,3)-ARCH(2) models are conducted by a statistical test, namely a Wilcoxon signed-rank test, at the 0.025 and 0.05 significance levels in one-tail-tests. The test results are shown in Table 5. Clearly, the proposed ARMA(2,3)-ARCH(1) model is significant (under a significant level 0.05) superior to other alternative models.

**4. Conclusions**

Forecasting of gas supply in self-provided power plant based on ARMA-ARCH Time series model is researched in this paper, which aimed at the characteristics of the gas system in iron and steel enterprises. Through the model application, it shows that The MAPE of this forecasting model is 3.32%, the mathematical model is effective for the gas system with complex working conditions.

**Acknowledgement**

Item Sponsored by humanities and social sciences research of the Chongqing Municipal Education Committee (16SKGH266); Item on Chongqing foundation and frontier plan (cstc2016jcyjA0341).

**References.**

[1] Wu hong-liang, LIU Kun 2007 *J. Gas and Heat*.27 35
[2] WANG Hua-feng, GUO Ming-zhou, BAI Hong-bin 2008 *J. Gas and Heat*.28 A01
[3] YANG Hua-feng 2012 *J. Gas and Heat*.32 A39
[4] YANG Wei-qin, GU Lan 1988 *M. Beijing: Beijing Institute of Technology Press.*
[5] YANG Shu-z, WU Ya, XUAN Jian-ping 2007 *M.Wuhan: Huazhong of Science and Technology Press.*
[6] LI Gang, WU Hui-xin, XUE Hui-feng 2008 *J. Computer Simulation. 25* 122
[7] Andrés Ubiernaa, Santiago Velilla 2007 *J. Journal of Statistical Planning and Inference. 137* 2903
[8] A. Aghababaii Mobarakhe, F. R. Rofooei, G. Ahmadi 2003 *J. Probabilistic Engineering.* 17 15
[9] Hemerly E M, Davis M H A 1989 *J. Annals of Statistics.* 17 941
[10] Aksasse B, Radouane L 1999 *J. IEEE Transactions on Signal Processing.* 47 2072
[11] Michael Nwogugu 2006 *J. Applied Mathematics and Computation.* 18 1735
[12] QI Er-shi, KONG Hai-ning, HE Shu-guang 2010 *J. Systems Engineering-Theory & Practice.* 30 2071
[13] ZHANG Qi, TI Wei, CAI Jiu-ju 2011 *J. Journal of Iron and Steel Research, International.* 18 37
[14] Li lei, Li hong-Juan 2015 *J. Journal of central south university.* 22 1437
[15] WU Cheng-zhong, JIAO De-mei 1988 *J. Energy for Metallurgical Industry.* 7 48
[16] Ives A, Abbott K C, Ziebarth N L 2010 *J. Ecology.* 91 858
[17] Shiqing Ling 2007 *J. Journal of Econometrics.* 14 849
[18] He Shu-yuan 2003 *M. Beijing: Peking University press.*
[19] Mohammadi K, Eslami H R, Kahawita R 2006 *J. Journal of Hydrology.* 331 293