ATLANTIS PRESS

Advances in Engineering Research, volume 184

2nd International Conference on Sustainable Energy, Environment and Information Engineering (SEEIE 2019)

Prediction of Gas Concentration Based on ARIMA and GARCH Model

Lizhi Zhang and Zaihua Yang

Xi'an International University, Shaanxi Xi'an, 710077, China

***Abstract*—In order to accurately predict the dynamic gushing process of gas in fully mechanized mining face, based on the historical monitoring data of underground gas concentration, with the help of R language, the ARIMA model is first established and fitted to determine the prediction equation of the ARIMA (*p, d, q*). The results of data fitting show that the model has a high degree of fitting to the gas concentration time series. Then the GARCH *(u, v)* is applied to the residual sequence of ARIMA (*p, d, q*), and the predicted value of the noise term in the ARIMA model is simulated, and the prediction result of the gas emission concentration is optimized. Finally, the 1001 fully mechanized mining face of Huangling No. 1 Mine in Shaanxi Province is taken as an application example. The results show that the combined model of ARIMA (*p, d, q*) and GARCH (*u, v*) can not only reflect the change trend of gas emission concentration but also has a high fitting effect and prediction accuracy.**

***Keywords*—gas emission concentration; time series; ARIMA model; GARCH model; prediction; fitting; R language**

I. INTRODUCTION

Relevant data show that 57.3% of the major accidents in China's coal mines are gas accidents, and 91.7% of the 24 major accidents since the founding of the People's Republic of China were gas dust accidents. It shows that gas disasters have become the first major disasters in coal mines[1]. Study the change law of gas emission concentration and predict and alarm it, and prevent the accident before it will ensure the safe production of coal mines.

At present, the methods applied to mine gas time series management mainly include ARIMA neural network[2], wavelet transform[3], Lyapunov exponent algorithm[4], gray system method[5] support vector machine [6], least squares support vector machine[7], etc., including various improved models based on the above algorithm. In addition, some scholars combine two research methods to predict time series, such as IABC-RBF algorithm and wavelet analysis[8], ARIMA-BP based model[9] , ARIMA-GM model[10]. Each of the above prediction models has its own advantages and disadvantages, but generally the algorithm is complex and the prediction step size is short, which is highly likely to result in inaccurate prediction results.

II. RESEARCH IDEAS

Based on the analysis and research of historical monitoring data of gas emission concentration, according to its sequence and randomness, the working face gas concentration is regarded as a non-stationary random time series. The ARIMA model is established for the random time series, and the reliability of the ARIMA model prediction is tested. Aiming at the mean regression problem of ARIMA model in the prediction process, the regression heteroscedastic model (GARCH) is used to simulate the fitting residual of ARIMA model, and the simulated result is used as the noise term predicted in ARIMA model. And use it to optimize the prediction of the ARIMA model.

III. CONSTRUCTION OF THE ARIMA MODEL

A. Model Identification

The stabilized gas concentration monitoring data is a first-order differential stationary stochastic process, so the ARIMA (p, d, q) model can be established. In practical applications: d generally takes 1 and does not exceed 2.

B. Parameter Estimation

The prediction model of the gas concentration random time series can be expressed as:

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \tag{1}$$

Where: W_t represents the value at time t in the random process, e_t represents a new information item that is independent of the random process.

For the values of ϕ_1, ϕ_2, …, ϕ_p and θ_1, θ_2,…, θ_q, the least squares method is used to estimate, and the formula of the prediction model is transformed into such a form, which is specifically expressed as follows:

$$e_t = W_t - \phi_1 W_{t-1} - \phi_2 W_{t-2} - \ldots - \phi_i W_{t-i} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q} \tag{2}$$

Where: e_t represents a white noise process.

Use a numerical algorithm to minimize $S_c(\phi, \theta)$ value to get the conditional least squares of all parameters:

$$S_c(\phi_1,\phi_2,\cdots,\phi_p,\overline{Y}) = \sum_{p+1}^{n}[(Y_t-\overline{Y})-\phi_1(Y_{t-1}-\overline{Y})-\phi_2(Y_{t-2}-\overline{Y})-\cdots-\phi_p(Y_{t-p}-\overline{Y})]^2 \tag{3}$$

C. Evaluation of the Model

It mainly judges whether the residual sequence of the ARIMA (p, d, q) model is independent of the random process, whether the model order and parameters are reasonable, and whether it needs to be re-estimated. The expression of the autoregressive moving average summation model is:

$$\hat{Y}_{t+1}(\ell) = \pi_1\hat{Y}_t(\ell) - \pi_2\hat{Y}_{t-1}(\ell) + \pi_3\hat{Y}_{t-2}(\ell) + \cdots \tag{4}$$

Where: $\hat{Y}_t$ represents the predicted value of gas concentration.

1) Method of establishing GARCH model

Since the residual sequence of the ARIMA (p, d, q) process constitutes a stochastic process, the sequence is uncorrelated but there are high-order correlation structures (ie, undulating clusters) and thick-tailed distributions, so it is necessary to the residual sequence of ARIMA (p, d, q) fitted to the GARCH(u, v) process, and then the ARIMA (p, d, q)+ GARCH(u, v) model of the gas concentration in the fully mechanized mining face is established.

The general form of the GARCH(u, v) model is:

$$\begin{cases} \sigma^2_{t|t-1} = \omega + \beta_1\sigma^2_{t\text{-}1|t-2} + \cdots + \beta_p\sigma^2_{t\text{-}u|t-u-1} + \alpha_1 r^2_{t-1} + \alpha_2 r^2_{t-2} + \cdots + \alpha_v r^2_{t-v} \\ r_t = \sigma_{t\,|\,t-1}\varepsilon_t \end{cases} \tag{5}$$

Where: u and v represent the order of the ARCH model and the GARCH model respectively; ω, α and β are unknown parameters respectively; $\sigma^2_{t|t-1}$ is the estimated value of the conditional variance at time t; $\{r_t\}$ is the time series fitted by the GARCH model.

2) Judging the pros and cons of ARIMA and GARCH combined forecasting model

There are generally four judgment indicators: Mean Absolute Deviation (MAD), Mean Absolute Percent Error (MAPE), Mean Squared Error (MSE), and Standard Deviation Error (SDE), which are expressed as follows:

$$MAD = \frac{1}{n}\sum_{1}^{n}\left|\frac{Y_i-\hat{Y}_i}{Y_i}\right| \tag{6}$$

$$MAPE = \frac{1}{n}\sum_{1}^{n}\left|\frac{Y_i-\hat{Y}_i}{Y_i}\right| \tag{7}$$

$$MSE = \frac{1}{n}\sum_{1}^{n}(Y_i-\hat{Y}_i)^2 \tag{8}$$

$$SDE = \sqrt{\frac{1}{n}\sum_{1}^{n}(Y_i-\hat{Y}_i)^2} \tag{9}$$

IV. Case Analysis

A. ARIMA Model Construction Process

1) Data source and processing

The data sample of this paper is from the 1001 fully mechanized mining face of Huangling No. 1 Mine in Shaanxi Province. The gas concentration monitoring data of 4 time points per hour on April 11, 12, 13 and 14 of April 2018 is selected as sample data. The time series prediction model is fitted by adjusting the gas concentration monitoring data from 0:00 on April 11 to 23:50 on April 13. The forecast interval is from 0:00 on April 14 to 11:50 on April 14. The data of this interval is used to verify the feasibility of the gas concentration prediction model.

Firstly, the plot of the gas concentration history monitoring data is drawn by using the plot instruction in the R language, and the first-order difference is made to the gas concentration monitoring data. The result of the difference is shown in Fig. 1:

It is observed that the process is roughly stable, the values of the sequence are evenly distributed around the value of 0, and the value of the sequence lags up and down with time, and there is no tendency to increase or decrease. Therefore, the random process after the first-order difference is smooth, in line with the basic requirements of the ARIMA model.

2) Order of gas concentration prediction model

The extended autocorrelation function (EACF) is used to determine the order of the model. The sample extended ACF (EACF) function values for the random time series are shown in Table 1.

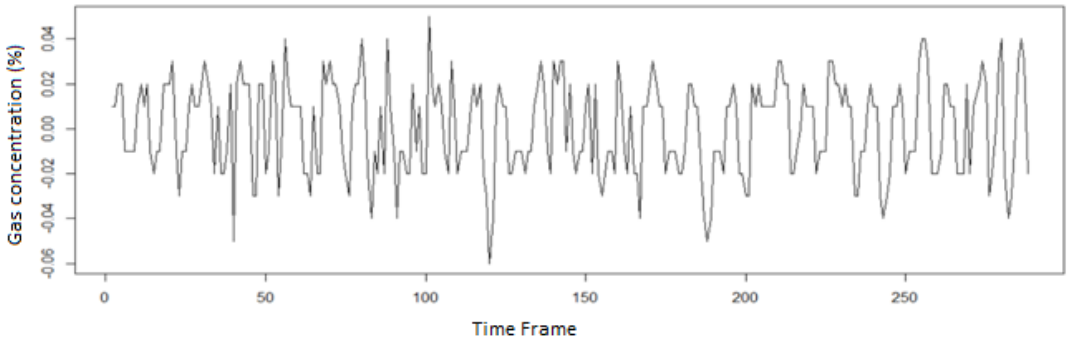


FIGURE I. FIRST-ORDER DIFFERENCE TIME SERIES DIAGRAM OF GAS CONCENTRATION IN FULLY MECHANIZED MINING FACE

TABLE I. EXTENDED ACF (EACF) OF THE FIRST-ORDER DIFFERENTIAL TIME SERIES W_t

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	0	0	x	x	x	x	0	x	X	0	0	0	0	0
1	x	x	0	x	x	x	0	x	0	0	0	0	0	0
2	x	0	x	0	0	0	0	x	0	0	0	0	0	0
3	x	x	x	0	0	0	x	x	X	0	0	x	0	x
4	x	x	x	0	0	x	0	0	0	0	0	0	0	0
5	0	x	x	0	0	0	0	0	0	0	0	0	0	0
6	0	x	0	0	x	0	0	x	0	0	0	0	0	0
7	x	x	x	x	0	0	0	0	0	0	0	0	0	0

In Table 1: the upper left corner of the zero triangle is at the p=4th row and the q=6th column, whereby the prediction model of the gas concentration can be determined to be ARIMA (4, 1, 6).

3) Parameter estimation of ARIMA model

The statistical software R language is used to solve the minimum 1 and 2 values in Eq. 4, which is the weighted estimate of the AR term and the MA term in the ARMA model. The calculation results are shown in Table 2.

TABLE II. PARAMETER ESTIMATION RESULTS OF THE ARMA (4, 6) MODEL

ARIMA(4,1,6)	AR1	AR2	AR3	AR4	MA1
CoefficientErr	0.3242	0.2512	-0.2954	-0.1210	0.4298
or estimate	0.0980	0.0742	0.0790	0.0874	0.1053
MA2	MA3	MA4	MA5	MA6	Intercept
-0.0484	-0.1080	0.0475	-0.2568	-0.0324	0.0004
0.0968	0.1339	0.1241	0.0721	0.0730	0.0012

Where: The estimate for σ^2 is 0.0002958; AIC = -1490.52; the log likelihood is 756.26.

Process mean can be expressed as:

$$W_t = 0.3242W_{t-1} + 0.2512W_{t-2} - 0.2954W_{t-3} - 0.1210W_{t-4} + e_t + 0.4298e_{t-1} - 0.0484e_{t-2} - 0.1080e_{t-3} + 0.0475e_{t-4} - 0.2568e_{t-5} - 0.0324e_{t-6}$$

Replace W_t in the prediction model of the first-order differential time series with the form of $W_t = \nabla Y_t = Y_t - Y_{t-1}$, and perform a simple shift operation to obtain an expression of Y_t, which is expressed as follows:

$$Y_t = 1.3242Y_{t-1} + 0.0730Y_{t-2} - 0.5466Y_{t-3} + 0.1216Y_{t-4} - 0.1210Y_{t-5} + e_t + 0.4298e_{t-1} - 0.0484e_{t-2} - 0.1080e_{t-3} + 0.0475e_{t-4} - 0.2568e_{t-5} - 0.0324e_{t-6}$$

4) Model evaluation of ARIMA model

Initial value of weight π is $\pi_0 = -1$, π can be obtained after iterative calculations. The operation result is expressed as follows:

$$\hat{Y}_t(\ell) = 0.9214\hat{Y}_{t-1}(\ell) + 0.1214\hat{Y}_{t-2}(\ell) - 0.4376\hat{Y}_{t-3}(\ell) + 0.0741\hat{Y}_{t-4}(\ell) + 0.2568\hat{Y}_{t-5}(\ell) + 0.0324\hat{Y}_{t-6}(\ell) + 0.4298Y_{t-1}(\ell) - 0.0484Y_{t-2}(\ell) - 0.1080Y_{t-3}(\ell) + 0.0475Y_{t-4}(\ell) - 0.2568Y_{t-5}(\ell) - 0.0324Y_{t-6}(\ell),$$

Using the above formulas to fit the historical monitoring data of gas concentration in the fully mechanized mining face, the fitting effect is shown in Fig. 2 (the actual value is indicated by the solid line and the fitted value is indicated by the dotted line):

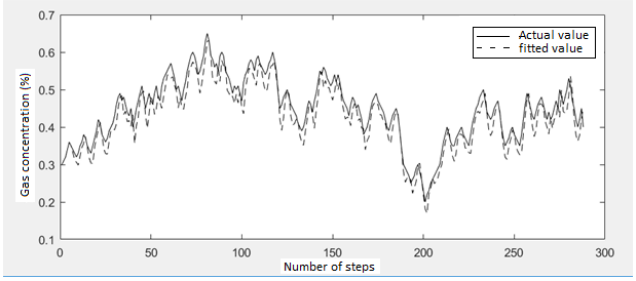


FIGURE II. COMPARISON OF ACTUAL AND FITTED VALUES OF GAS CONCENTRATION

B. GARCH Processing of Residual Sequences

1) Identification of GARCH models

The EACF method is used to identify the residual sequence with GARCH model, and the sample EACF of the squared value of the residual sequence fitted by the ARIMA (4, 1, 6) model is obtained. The results are shown in Table 3:

TABLE III. THE SAMPLE EACF OF THE SQUARED VALUE OF THE RESIDUAL SEQUENCE

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	0	0	x	0	0	0	x	x	0	0	0	0	x	0
1	x	0	x	0	0	0	0	0	0	0	0	0	x	0
2	x	x	0	x	0	0	0	0	0	0	0	0	x	0
3	x	x	x	0	0	0	0	0	0	0	0	0	x	0
4	x	x	0	0	0	0	0	0	0	0	0	0	0	0
5	x	0	x	0	0	0	0	0	0	0	0	0	0	0
6	x	0	x	0	x	0	0	0	0	0	0	0	0	0
7	x	x	x	0	x	x	0	0	0	0	0	0	0	0

The ARMA of the square of the residual sequence is ARMA(4,3). Therefore, it can be determined that the order u and v of the GARCH process are 3 and 4, respectively, so the residual random process of ARIMA(4,1,6) model fitting is recognized as GARCH(3,4).

2) Determination of GARCH model parameters

The parameters of the GARCH(3,4) model need to be estimated using the maximum likelihood method. The estimation results are shown in Table 4:

TABLE IV. ESTIMATES OF GARCH(3,4) MODEL PARAMETERS

parameter	estimate	standard error	t	Pr（>|t|）
a0	2.084e-06	<0.0001	<0.0001	<0.0001
a1	2.633e-02	<0.0001	<0.0001	<0.0001
a2	8.416e-04	<0.0001	<0.0001	<0.0001
a3	1.979e-16	<0.0001	<0.0001	<0.0001
a4	4.100e-02	<0.0001	<0.0001	<0.0001
b1	7.658e-03	<0.0001	<0.0001	<0.0001
b2	1.061e-02	<0.0001	<0.0001	<0.0001
b3	1.920e-02	<0.0001	<0.0001	<0.0001

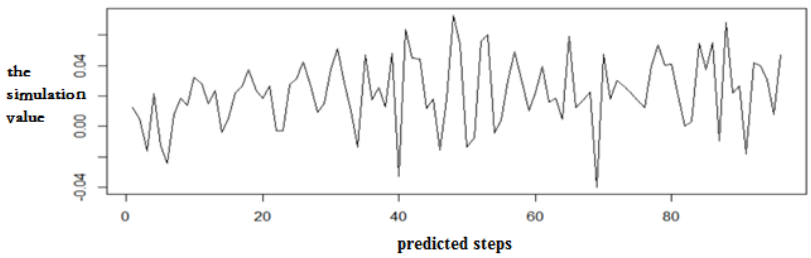
```
> m1=garch(r.date,order=c(3,4))；  summary(m1)
```

The expression for the GARCH model is as follows:

$$\sigma_{t|t-1}^2 = 0.000002084 + 0.007658\sigma_{t-1|t-2}^2 + 0.01061\sigma_{t-2|t-3}^2 + 0.0192\sigma_{t-3|t-4}^2 + 0.02633r_{t-1}^2 + 0.0008416r_{t-2}^2 + (1.979e\text{-}16)r_{t-3}^2 + 0.041r_{t-4}^2$$

The ARIMA and GARCH models were used to predict the gas concentration (sample size 96) of the fully mechanized mining face from 0:00 on April 14 to 11:50 on April 14.

Use the R language to simulate the stochastic process of the GARCH model, as shown in Fig. 3:




```
>set.seed(12345678)
>garch11.sim=garch.sim(alpha=c(2.084e-06,2.633e-02,8.416e-04,1.979e-16,4.100e-02)),beta=7.658e-03,1.061e-021.92,0e-02,n=96)
>plot(garch11.sim,ylab='GARCH(3,4) simulation value', xlab=' predicted steps')
```

FIGURE III. GARCH MODEL SIMULATION RESULTS

The simulation results in Table 3 are derived, and the derived series is the noise term of forward 1 to 96 in $\hat{Y}_t(\ell)$. After the iterative operation, the predicted gas concentration of the fully mechanized mining face can be obtained. The predicted result is shown in Fig. 4 (the dotted line indicates the predicted value; the solid line indicates the actual gas concentration):

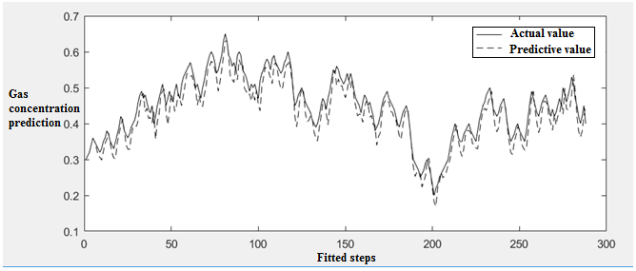


FIGURE IV. COMPARISON OF PREDICTIVE AND ACTUAL VALUES OF GAS CONCENTRATION

Analysis of Fig. 4 shows that the trend of the predictive value of the combined model of ARIMA and GARCH is basically consistent with the trend of the actual value of the gas concentration.

According to the calculation, the MAD (Mean Absolute Deviation) is 0.0268, the MAPE (Mean Absolute Percent Error) is 5.56%, the MSE (Mean Squared Error) is 0.0019, and the SDE (Standard Deviation Error) is 0.0436. The calculation results of the above four errors are small, indicating that ARIMA and GARCH model has high prediction accuracy for gas concentration in fully mechanized mining face, and can be used to recursively calculate the future value of gas concentration.

V. SUMMARY

This paper mainly uses R language as an auxiliary tool to carry out empirical research on the feasibility of ARIMA model and GARCH model in gas concentration prediction in time series analysis method. Firstly, the gas concentration prediction model ARIMA (4,1,6) of the fully mechanized mining face was established. Then the GARCH(3,4) model was established for the fitting residual sequence of the ARIMA model. Finally, the gas concentration was predicted and the prediction effect was analyzed. Conclusion as below:

1) ARIMA has a high degree of fitting between the simulated and actual values when fitting the gas concentration historical monitoring data. However, the ARIMA model has certain defects in predicting future gas concentration, that is, the error term when predicting the future value of gas concentration cannot be obtained. This leads to a large deviation between the predicted value of the ARIMA model and the actual value of the gas concentration. The GARCH model is built for the residual sequence of the ARIMA model. The stochastic process simulated by the GARCH model can be used as the residual sequence of the future ARIMA model fitting to solve the problem that the error term cannot be obtained in the ARIMA model prediction process.

2) The MAPE with ARIMA and GARCH combined model for gas concentration prediction is only 5.56%, and the deviation is small, which indicates that the combined model can reflect the change trend of the true value of gas concentration in the fully mechanized mining face.

ACKNOWLEDGEMENT

This research was financially supported by the Shaanxi Provincial Department of Education Research Project (18JK1131).

REFERENCES

[1] Lan Hang;Chen Dongke;Mao Debing. Current status of deep mining and disaster prevention in China, Coal Science and Technology [J]. Coal Science and Technology, 2016, 44(01):39-46.

[2] Wang Xinying. Research on Multivariate Time Series Prediction Based on Random Project Neural Networks [D].2015

[3] Cheng Xie. Methods of Risk Measurement Estimation And Empirical Study on Financial Time Series Based on Wavelet Analysis [D]. 2016

[4] ZHANG Xin-yu;LI Ru;WANG Jin-yu. Study of Improved Algorithm of Gas Concentration Based on the Largest Lyapunov Exponent [J]. Computer and Modernization, 2014(10):119-122.

[5] Liu Haiyan. Study on Monitoring Data Analysis and Deformation Prediction of Deep Foundation Pit [D]. 2012

[6] N V V. Statistical learning theory [M]. New York: John Wiley, 1998.

[7] Yin Xiaoqin. Research And Application of Hybrid Time Series Model Based on Support Vector Machine [D]. 2016

[8] Liu Liqiang. Gas Time Series Optimization Prediction Based on IABC-RBF Algorithm and Wavelet Analysis [D]. 2015

[9] Xing Haoran; Yang Yingdi. Prediction of coal mine gas emission based on ARIMA and BP combination model [J]. Inner Mongolia Coal Economy, 2017(17):3-6.

[10] WU Bing;GUO Zhiguo;WANG Ziwei. Gas Emission Forecasting at Heading Face Based on ARIMA-GM Model [J]. Safety in Coal Mines, 2015, 46(11):152-155.