GUO ZHIGUO
WU BING
MA XIAOLIN
and
CHEN JUAN

Forecasting gas emission concentration using wavelet decomposition and GM-ARIMA model

In order to improve the prediction accuracy of the dynamic gas emission concentration in coal mining and heading working face, an approach based on wavelet decomposition and GM-ARIMA model prediction method was proposed. Firstly, this paper utilized the Grey Model GM (1, 1); the Autoregressive Integrated Moving Average Model ARIMA and the combination prediction model, GM-ARIMA, built by the method of variance reciprocal weighting to forecast the gas emission concentration respectively based on the gas concentration time series, and then employed wavelet theory to decompose the gas concentration time series into approximation sequence and detail sequence, after analyzing the characteristics of decomposition sequences respectively, matching methods-GM (1, 1) and ARIMA model, were employed to forecast the decomposed sub sequences. Then the final prediction values of the gas concentration were obtained through the reconstruction of sub sequences. Finally, it was verified by the engineering application and show that the average error rate of wavelet decomposition and GM-ARIMA prediction method was 3.64%, which provided a better fitting effect and a higher prediction accuracy compared with single model GM (1, 1); ARIMA and GM-ARIMA combination prediction model. It follows that this method is feasible and effective.

Keywords: *Gas concentration, Wavelet theory, decomposition series, GM-ARIMA model, forecasting.*

1. Introduction

Mine gas has always been the most important factor that threatens the safety of coal mine production, especially, the gas concentration limit brings a serious security risk to mining safety in the process of mining [1-2]. Although mine gas concentration is one of the important indicators of mine safety monitoring, now coal mine safety monitoring system generally do not have the ability to predict and forecast. Therefore, the timely, accurate and effective prediction of the gas concentration change in the key position of the coal mining face and to grasp the trend and the law of the real time variation of the gas concentration in the mining space not only can make full use of the gas concentration data obtained by the coal mine safety monitoring system, but also can get the precursor of the gas concentration exceeding limit in advance, so as to ensure the safety of mine excavation.

In recent years, the technology of mine gas concentration prediction has been developed rapidly. Many experts and scholars at home and abroad have made a research on the prediction method and have put forward one after another many artificial intelligence prediction methods and their combination prediction methods which all based on nonlinear theory. Such as Li Runqiu [3] predicted gas emission based on self-organizing data mining; Ai Li [4] predicted gas emission based on particle swarm optimization algorithm; Qiao Meiying [5] utilized least squares support vector machine regression (LS-SVR) to forecast coal mine gas emissions; Fan Baolong [6] used the combination of local mean decomposition (LMD) method and support vector machine (SVM) to predict gas concentration. Although a large number of empirical research show that [7-9]: above the artificial intelligent prediction methods have been widely used, and can use historical data effectively .At the same time, some prediction methods are considered highly nonlinear relationship and the dynamic emission process among the various influencing factors, but are still exist some limitations; Because most of the prediction methods can not contain the characteristics of the dynamic change of gas emission, and the prediction process is static; Even if the gas emission and its influencing factors have nonlinear correlation in some prediction methods which contain the dynamic changes of gas emission, but general mine only can provide historical data about gas emission concentration, and the historical data of the related influencing factors can't give, such as coal seam depth, seam thickness etc., which makes the prediction accuracy and reliability can't be convincing.

In fact, the historical data of gas emission concentration itself has already contained the main information and related factors, so that the time sequence of mine gas concentration often presents a complex dynamic trend and fluctuation random. For the prediction of gas concentration which has larger dynamic trend complexity and fluctuation randomness, if the GM (1, 1) and ARIMA model prediction methods are

Messrs. Guo Zhiguo, Wu Bing, Ma Xiaolin and Chen Juan, Faculty of Resources and Safety Engineering, China University of Mining and Technology (Beijing), Beijing 100 083, China. Email: guozhiguo2012@126.com

used separately, the prediction of some mutation data may be over fitting. And if the correlation method can be used to effectively decompose the time series of gas concentration with strong dynamic trend complexity and volatility, and the complexity and volatility of the different components are reduced, then using phase matched GM (1, 1) and ARIMA model to predict, so as to avoid over fitting.

Wavelet analysis is a kind of time frequency localization signal analysis method after Flourier transform, which can do a multi scale refinement analysis for function or signal by stretching and translation and other computing functions. The time series can be decomposed into different scales by using the transform of the scale parameter, which can be decomposed into the low frequency approximate component and the high frequency detail component [10]. In view of this, this paper presents a new method of using wavelet decomposition and GM-ARIMA model to predict gas concentration. By using wavelet transform, the time series of mine gas concentration is decomposed into low frequency approximate sequence which reflects the complexity of dynamic trend of gas concentration and high frequency detail sequence which reflects the fluctuation of gas concentration, and the GM (1, 1) and ARIMA model were established respectively to obtain the prediction results by analyzing the characteristics of each sub sequence. Empirical results show that the method of wavelet decomposition and GM-ARIMA model has high precision and adaptability, and it is an effective method to predict the gas concentration in coal mining and heading working face.

2. The establishment of GM-ARIMA model

2.1 The establishment of the grey model GM (1, 1)

Grey theory is a kind of dynamic tendency prediction theory, which takes small samples that partial information is known and some information is unknown as the research object. By identifying the different degrees of development trend of the system factors, it looks for the rule of system changing by generating and processing the original data to generate the data sequence which has stronger regularity, then the corresponding mathematical model is established to predict the future development trend of things. It is more suitable for the prediction of dynamic sequence with deterministic factors [11]. Meanwhile, the time series of mine gas concentration has a strong complex dynamic trend, so it is suitable to use the grey theory model GM (1, 1) to forecast.

In the process of establishing the grey model GM (1, 1), firstly, it will turn the irregular raw data into the regular generated data by a certain numerical processing method, and then using these regular generated data to build a model. The modeling steps are given as follows:

(1) When we obtain the sample in the original data, if the sample contains N data, it can form series $X^{(0)}$, where:

$$X^{(0)}=\left\{X^{(0)}(1),X^{(0)}(2),\ldots,X^{(0)}(N)\right\}.$$

(2) We make a one-time cumulative for series $X^{(0)}$and get a new series:

$X^{(1)}=\left\{X^{(1)}(1),X^{(1)}(2),\ldots,X^{(1)}(N)\right\}$, where the values of $X^{(1)}(t)$ is shown as Eq(1) below:

$$X^{(1)}(t)=\sum_{k=1}^{t}X^{(0)}(k),\ t=1,2,\ldots,N \quad \ldots \quad (1)$$

(3) And then we construct the constant term vector Y_N and the accumulation matrix B, and the constant term vector Y_N and the accumulation matrix B are shown as Eq(2) and Eq(3) respectively below:

$$Y_N=\left[X_1^{(0)}(2),X_1^{(0)}(3),\ldots,X_1^{(0)}(N)\right]^T \quad \ldots \quad (2)$$

$$B=\begin{bmatrix}-\frac{1}{2}\left(X^{(1)}(1)+X^{(2)}(2)\right) & 1\\ -\frac{1}{2}\left(X^{(1)}(2)+X^{(2)}(3)\right) & 1\\ \vdots & \vdots\\ -\frac{1}{2}\left(X^{(1)}(N-1)+X^{(2)}(N)\right) & 1\end{bmatrix} \quad \ldots \quad (3)$$

(4) We use the method of least squares to solve grey parameter α, and the grey parameter α is given as Eq(4) below:

$$\alpha=\begin{bmatrix}\alpha\\ u\end{bmatrix}=\left(B^TB\right)^{-1}B^TY_N \quad \ldots \quad (4)$$

where u is the grey parameter of endogenous control.

(5) Finally, we bring grey parameter into the function of time, and then obtain GM (1, 1) model as follows:

$$X^{(1)}(t)=\left(X^{(0)}(1)-\frac{u}{\alpha}\right)e^{-\alpha(t-1)}+\frac{u}{\alpha} \quad \ldots \quad (5)$$

Then make the series of estimated values $X^{(1)}(t)$ obtained from Eq(5) subtract consecutively and reductive generation, and it can be calculate to get the estimated values of the original data $X^{(0)}(t)$:

$$X^{(0)}(t)=X^{(1)}(t)-X^{(1)}(t-1) \quad \ldots \quad (6)$$

So we can employ Eq(6) to calculate the prediction values of gas emission concentration at the t time.

2.2 The establishment of ARIMA (p, d, q) model

Autoregressive Integrated Moving Average Models, called ARIMA, is a time-series prediction methods, where AR of ARIMA (p, d, q) is the autoregressive process, P is the auto regressive terms, MA is moving average processes, d is differential times conducted by time series becoming stationary time series, q is the moving average term. The basic idea of which is to view the data series formed by forecasting objects with the time passing as a random series, and then established certain mathematical model to describe this series.

Once the model is identified can we employ the past values and present values of the time series to forecast the future values, which is more suitable for the prediction of non-stationary random series [12]. At the same time, the time series of mine gas concentration has a strong volatility, so it is suitable to use ARIMA model to predict.

2.2.1 The structure of ARIMA (p, d, q) model

ARIMA (p, d, q) model has the following structure shown as Eq(7) below:

$$\Phi(B)(1-B)^d Y_{(t)} = C + \Theta(B)\alpha_{(t)} \quad \dots \quad (7)$$

where B is backward operator, $BY_{(t)} = Y_{(t-1)}$; $\Phi(B)$ is autoregressive operator, the coefficient of regression polynomial is: $\Phi(B) = 1-\phi_1 B - \phi_2 B^2 \dots \phi_p B^p$; $\Theta(B)$ is the moving average operator, and its coefficient polynomial is: $\Theta(B) = 1-\theta_1 B - \theta_2 B^2 \dots \theta_p B^p$; α_t is the random perturbation terms, that is, random error term; C is constant.

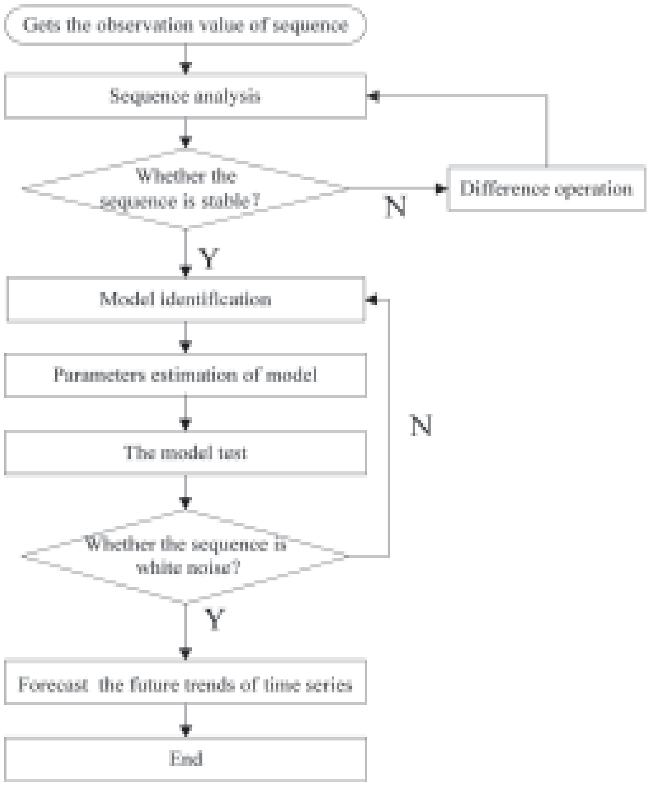

Fig.1 Schematic diagram of the ARIMA modelling

2.2.2 The modeling process of ARIMA

The modeling process of ARIMA is shown in Fig.1:

2.3 The establishment of GM-ARIMA model

Combination prediction method is firstly to use two or more different single model to forecast the target [13, 14], and use some kind of criteria for each single model to form a combined model, and then use the combination model to forecast. The model of the combination can be expressed as Eq(8) below:

$$f(f_1, \dots, f_k) = \sum_{i}^{k} w_i f_i \quad \dots \quad (8)$$

where f_i is the prediction vales of each model, w_i is the weight of each model.

The next step is to determine the weight w_i by using the variance reciprocal method, and the weight w_i can be expressed as Eq(9) below:

$$w_i = D_i^{-1} \Big/ \sum_{i=1}^{n} D_i^{-1}, \; i = 1,2,\dots,k \quad \dots \quad (9)$$

where D_i is the sum of error square of model i, n is the fitting value number that used to forecast.

GM-ARIMA, modeling steps of prediction model are as follows [13]:

(1) Use GM (1, 1) model to forecast the gas emission, the value of prediction is f_1, and calculating the sum of error square D_1.

(2) Use the ARIMA (p, d, q) model to forecast the gas emission, the value of forecasting is f_2, and calculating the sum of error square D_2.

(3) Calculate their weight by the method of variance reciprocal:

$$w_1 = D_1^{-1} \big/ \left(D_1^{-1} + D_2^{-1}\right), w_2 = D_2^{-1} \big/ \left(D_1^{-1} + D_2^{-1}\right).$$

(4) The calculation results: $f = w_1 f_1 + w_2 f_2$.

3. The prediction model of wavelet analysis theory

Wavelet analysis is a kind of time and frequency domain analysis method, it has good localization properties both in time domain and frequency domain and can automatically adjust the sampling signal according to the frequency density, it is easy to capture and analyze weak signals, especially for singular signals, which can be very good to deal with weak or abrupt signals. The goal is to transform the information of a signal into the wavelet coefficients, it can be easily processed, stored, transferred or used to reconstruct the original signal. These advantages determine the wavelet analysis can be effectively applied to the prediction of gas concentration in coal mine [9]. By using wavelet transform to the gas concentration time series, the original sequence is projected onto different scales and a number of sub sequences are obtained, and different models for different sub sequences were built to predict, and then the whole prediction results were obtained through the sequence reconstruction. This method not only improves the prediction accuracy, but also can improve the modeling efficiency.

Let a function $\Psi(t)\in L^2(R)$, in which Fourier transform to $\widehat{\Psi}(\omega)$, supposing that $\widehat{\Psi}(\omega)$ satisfy the admissibility condition below:

$$C_\Psi=\int_0^{\infty}\frac{\left|\widehat{\Psi}(\omega)\right|}{\omega}d\omega<\infty \quad \text{... (10)}$$

The function $\Psi(t)$ is called the mother wavelet. Generating a series of wavelet sequences by using the wavelet transform of the mother wavelet $\{\Psi_{a,b}(t)\}$:

$$\Psi_{a,b}(t)=|a|\frac{1}{2}\Psi\left(\frac{t-b}{a}\right)(a,b\in R;\ a\neq 0) \quad \text{... (11)}$$

The function $\{\Psi_{a,b}(t)\}$ is called baby wavelet, in which *a* is the scale factor for the mother wavelet to stretch out and draw back, and *b* is a translation factor to translate mother wavelet.

The data process of this paper is divided into 4 steps below:

(1) According to the characteristics of historical data series, the suitable wavelet decomposition function and decomposition scale are used to decompose the gas concentration time series of the original working face, and then obtaining the wave form and sequences of the decomposed sub sequences.

(2) Analyzing the characteristics of each sub sequence, and using GM (1, 1) and ARIMA model to predict the approximate sequence and the detail sequence after decomposition.

(3) The predicted sequences in step two are reconstructed to produce the final predicted sequence.

(4) Comparing the predicted gas concentration sequence and the original gas concentration sequence and calculating the error rate. At the same time, comparing the prediction results of single prediction model GM (1, 1) and ARIMA with the combined forecasting model GM-ARIMA to verify the prediction effect of the model.

4. The engineering application analysis

4.1 The original data of gas emission

In order to test the feasibility and validity of the established prediction model above, this paper choose a total of 27 gas concentration data as an example which from Xin Shun mining 15101 heading working face return air trough on December 15, 2015 from 11:03 a.m. to noon 12:21. The first 20 of the 27 gas concentration data were selected as the original data of the experimental samples, and the remaining 7 data were used to verify the performance of the proposed prediction method. The raw data from 11:00 a.m. to 12:21 at noon is shown in Fig.2; And which describe the timing diagram of the 27 gas concentrations (where a gas overrun occurs at the 12:18 time).

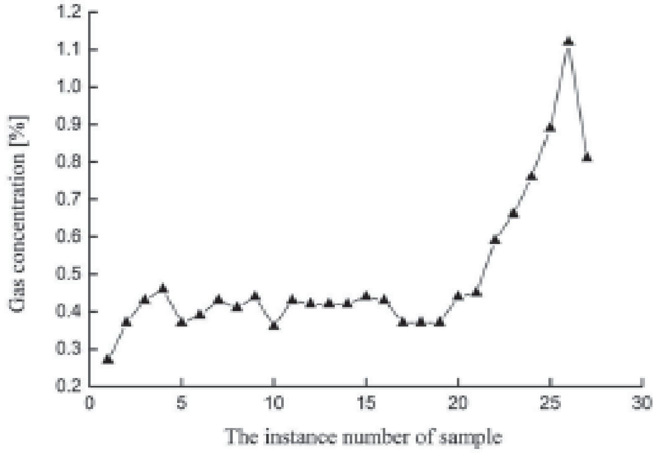

Fig.2 The time series of gas emission concentration

4.2 GM-ARIMA combination model prediction

4.2.1 GM (1, 1) model prediction

Using the first 20 original gas concentrations data in Fig.2 and the established steps of grey GM (1, 1) prediction model in the previous chapters to predict the value of the gas emission concentration from the time 12:03 to 12:21, prediction results and the average error rate (not include the prediction error rate that the measures taken after the gas overrun, the same below) are shown in Table 2.

4.2.2 ARIMA (p, d, q) model prediction

According to the first 20 original data of the gas concentration in Fig.2 and the established steps of ARIMA model in the previous chapters, this paper firstly uses the unit root test of Eviews software to determine the sequence of stationarity, then, the values of Q and P are determined which based on stationary sequence auto correlation and partial correlation diagram. In this paper, the Q, P value is 1, 1, and the difference is 2 times, so choosing ARIMA (1, 2, 1) model to predict. The least square method is used to get the parameter estimation, and the final prediction value of the model is obtained as the Eq(12). ARIMA (1, 2, 1) model prediction results are shown in Table 2.

$$(1+0.924463B)(1-B)^2Y_{(t)}=0.003911+(1+0.117875B)\alpha_{(t)} \quad \text{... (12)}$$

4.2.3 GM-ARIMA combination model prediction

According to the prediction results of above GM (1, 1) model and ARIMA model and the established steps of GM-ARIMA combination prediction model, we can get the final prediction value of GM-ARIMA combination prediction model, and is shown in Eq(13) below. The prediction results of GM-ARIMA model are shown in Table 2.

$$f=w_1f_1+w_2f_2=0.9089X_{(t)}+0.0911Y_{(t)} \quad \text{... (13)}$$

4.3 Prediction of wavelet analysis and GM-ARIMA model

4.3.1 The wavelet function and decomposition level

The selection of mother wavelet should be based on the characteristics of gas emission concentration time series. At the same time, the gas concentration time series can be decomposed by constructing different wavelet bases and comparing the results of transformation, so as to select the mother wavelet which can reflect mostly the sub series. In this paper, by analyzing the original sequence waveform, we choose the four order Daubechies function db4as the mother wavelet. In a certain prediction, supposing that the decomposition level is too small, it is not effective to separate the components of the original signal with different frequency characteristics; supposing that the decomposition level is too large, we need to use more models to forecast decomposition of each component, each model will introduce a certain error, which leads to the larger error about final prediction; this paper select the 3 layer db4 wavelet to decompose gas concentration time series after a lot of data analysis.

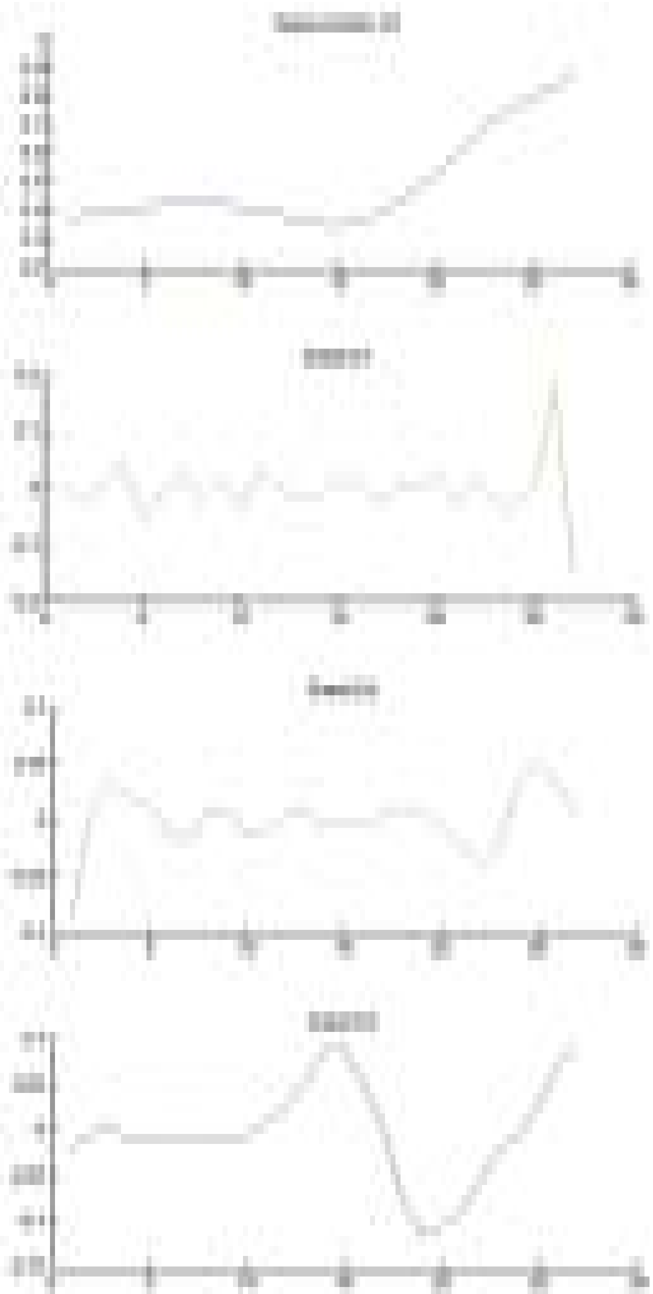

Fig.3 Wavelet decomposition series

4.3.2 Wavelet decomposition

In this paper, the MATLAB software is used to decompose the gas concentration time series data in Figure 3 and using Daubechies function db4 wavelet function, and the decomposition level is 3. MATLAB software output the approximate sequence and the detail sequence after decomposition as shown in Fig.3 below, where the A3 Approximation is the approximate sequence, Detail D1, D2 and D3 are respectively detail sequence corresponding to each decomposition scale.

4.3.3 Prediction of decomposition sub sequence

(1) GM (1, 1) prediction of approximate sequence: It can be seen that the sequence reflects the dynamic trend of the whole gas concentration by the variation trend of A3 Approximation in Figure 3, and the sample data of this paper is less, the use of BP neural network to predict approximate sequence will lead to poor learning effect, therefore, this paper uses GM (1, 1) model to predict the approximate sequence so that the prediction about dynamic change trend of sequence is reasonable. The predicted results which from GM (1, 1) model are shown in Table 1.

Table 1: The prediction results of each decomposed sequence

Time	A3	D1	D2	D3
12:03	0.598646	-0.022867	-0.016441	-0.081105
12:06	0.641246	0.017913	-0.042170	-0.047541
12:09	0.732815	-0.014122	-0.016113	-0.017105
12:12	0.740128	-0.039838	0.035195	-0.002592
12:15	0.831992	0.005661	0.054676	0.035029
12:18	0.842211	0.174105	0.028441	0.070147
12:21	0.907918	0.269857	0.045296	0.091681

(2) ARIMA model prediction of detail sequence: Each scale detail sequence is a non-stationary sequence after wavelet decomposition shown in Figure 3, therefore, this paper uses ARIMA model to predict. According to the above ARIMA modeling steps (will not repeat them here), the wavelet decomposition detail sequence D1, D2 and D3 were respectively constructed ARIMA (3, 1, 1) model, ARIMA (2, 0, 1) and ARIMA (2, 0, 2) to predict. The predicted results are shown in Table 1 below.

4.4 Prediction of wavelet decomposition and GM-ARIMA model

On the basis of prediction of each decomposition sequence, reconstructing the predictive value about approximate and detail sequences to get final gas concentration prediction results of this heading working face from the time 12:03 to 12:21, and comparing with realistic gas concentration in order to calculate the average error rate. The predictive results and average error rate of wavelet decomposition and GM-ARIMA model are shown in Table 2. In order to directly reflect the effect of various prediction methods, the comparison of the predicted value and the actual value is drawn, as shown in Fig.4.

Table 2 prediction results of various prediction models show that: gas concentration will be out of limit at 12:18 (concentration >1%) and we should alarm at this time. The real situation is that the gas concentration is out of limit at 12:18, in this way, we can take relevant safety measures in advance to avoid the influence of gas overrun on safety production. The prediction value of 12:21 time appeared a big error which is due to taking measures to gas overrun. At the same time, from the average error rate comparison results of all kinds of prediction models from the Table 2 and the comparison results of various prediction models from Fig. 4 the prediction model of Figure 4 we can draw a conclusion that: wavelet decomposition and GM-ARIMA prediction model proposed in this paper has higher accuracy than other prediction models.

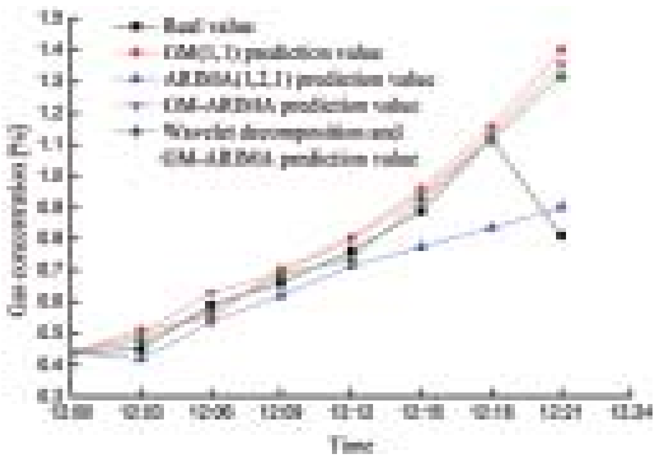

Fig.4 The comparison between prediction values and actual values of gas emission concentration

Table 2: The prediction results of various prediction methods

Time	Real value	GM (1,1)	ARIMA (1,2,1)	GM-ARIMA	Wavelet decomposition and GM-ARIMA
12:03	0.45	0.509703	0.418952	0.495771	0.478233
12:06	0.59	0.550328	0.536918	0.628893	0.569448
12:09	0.66	0.702841	0.620483	0.690145	0.685475
12:12	0.76	0.802573	0.713126	0.794424	0.732893
12:15	0.89	0.959217	0.772614	0.942217	0.927358
12:18	1.12	1.154775	0.835724	1.125709	1.114904
12:21	0.81	1.400729	0.902780	1.355366	1.314752
Average error rate / %	-	7.16	11.13	5.37	3.64

5. Conclusions

In this paper, the new prediction model combined wavelet decomposition with GM-ARIMA predicted gas emission concentration in coal mining and heading working face, example analysis results show that: the original gas concentration sequence is decomposed for the low frequency approximation sequence by wavelet to reflect the complexity of the dynamic trend of gas concentration, and high frequency detail sequence to reflect the fluctuation of gas concentration. Through the analysis of the characteristics of different sub sequences, the GM (1, 1) and ARIMA model were established respectively to predict, and the predicted results are in good agreement with the actual test results, so then it can effectively improve the prediction accuracy and the fitting effect, and overcome the shortcomings of the single forecasting method. Finally, compared GM (1, 1) and ARIMA prediction method with GM-ARIMA combination prediction method, the average error rate is only 3.64%, so it is proved that the new method is an effective method for gas concentration prediction, which can provide strong support and technical basis for the short-term prediction of gas concentration in coal mining and heading working face.

Acknowledgment

The authors gratefully acknowledge the support of National Natural Science Foundation of China (51274205).

References

1. Wang, L., Cheng, Y. P., Wang, L., Guo, P. K. and Li, W. (2012): "Safety line method for the prediction of deep coal-seam gas pressure and its application in coal mines." *Safety Science*, 50(3)5: 523-529.
2. Perrone, D. and Amelio, M. (2016): "Numerical simulation of MILD (moderate or intense low-oxygen dilution) combustion of coal in a furnace with different coal gun positions." *International Journal of Heat and Technology*, 34(S2): S242-S248.
3. Li, R. Q., Shi, S. L., Wu, A. Y., Luo, W. K. and Zhu, H. P. (2014): "Research on prediction of gas emission based on self-organizing data mining in coal

mines." *Procedia Engineering*, 84(4): 779-785.

4. Ai, L., Cheng, J. T. and Xu, S. K. (2012): "Coal mine gas prediction model based on particle swarm optimization algorithm." *Advanced Materials Research*, 546-547: 8-12.
5. Qiao, M. Y., Ma, X. P., Lan, J. Y. and Wang, Y. (2011): "Time-series gas prediction model using LS-SVR within a Bayesian framework." *Mining Science and Technology*, 21(1): 153-157.
6. Fan, B. L., Bai, C. H. and Li, J. P. (2013): "Forecasting model of coalface gas emission based on LMD-SVM method." *Journal of Mining and Safety Engineering*, 30(6): 946-952.
7. Liu, S. W., Qu, S. J., Li, J. L. and Dai, L. C. (2013): "Analysis on influence factors of top corner gas concentration and trend prediction." *Advanced Materials Research*, 634-638(1): 3650-3654.
8. Tang, J., Jiang, C., Chen, Y., Li, X., Wang, G. and Yang, D. (2016): "Line prediction technology for forecasting coal and gas outbursts during coal roadway tunneling." *Journal of Natural Gas Science and Engineering*, 34: 412-418.
9. Xu, S., Ba, J., Chen, X., Zheng, T., Yang, Y. and Guo, L. (2016): "Predicting strata temperature distribution from drilling fluid temperature." *International Journal of Heat and Technology*, 34(2): 345-350.
10. Zhang, D. F. (2011): MATLAB wavelet analysis. Beijing, China: Machinery Industry Press, 87-99.
11. Cheng, H., Wei, F., Yang, T. and Zhao, Y. (2016): "Relation degree analysis of controllable factors in the bitumen foaming process," *International Journal of Heat and Technology*, 34(3): 364-370.
12. Yuan, C., Liu, S. and Fang, Z. (2016): "Comparison of china's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM(1, 1) model." *Energy*, 100(1): 384-390.
13. Zhu, X. and Shen, M. (2012): Based on the ARIMA model with grey theory for short term load forecasting model. International Conference on Systems & Informatics, 564-567.
14. Lin, W. R., Xu, B. and Wei, J. L. (2016): "Forecasting VaR with Combination of Factors and Variables Using High-Frequency Information." *Revista de la Facultad de Ingeniería*, 31(5): 268-277.