

## Research Article

# Short-Term Coalmine Gas Concentration Prediction Based on Wavelet Transform and Extreme Learning Machine

Wu Xiang,<sup>1,2</sup> Qian Jian-sheng,<sup>1</sup> Huang Cheng-hua,<sup>3</sup> and Zhang Li<sup>4</sup>

<sup>1</sup> School of Information and Electrical Engineering, China University Mining & Technology, Xuzhou, Jiangsu 221116, China

<sup>2</sup> School of Medical Informatics, Xuzhou Medical College, Science and Technology Building E206, Dong Dian Zi Campus, Xuzhou, Jiangsu 220009, China

<sup>3</sup> Northern Nenghua Company of Wanbei Coal-Electricity Group, Huaibei, Anhui 235000, China

<sup>4</sup> School of Medical Imaging, Xuzhou Medical College, Xuzhou, Jiangsu 221009, China

Correspondence should be addressed to Wu Xiang; [dmslxzmc@163.com](mailto:dmslxzmc@163.com)

Received 18 February 2014; Accepted 5 July 2014; Published 22 July 2014

Academic Editor: Marcelo J. Colaço

Copyright © 2014 Wu Xiang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

It is well known that coalmine gas concentration forecasting is very significant to ensure the safety of mining. Owing to the high-frequency, nonstationary fluctuations and chaotic properties of the gas concentration time series, a gas concentration forecasting model utilizing the original raw data often leads to an inability to provide satisfying forecast results. A hybrid forecasting model that integrates wavelet transform and extreme learning machine (ELM) termed as WELM (wavelet based ELM) for coalmine gas concentration is proposed. Firstly, the proposed model employs Mallat algorithm to decompose and reconstruct the gas concentration time series to isolate the low-frequency and high-frequency information. Then, ELM model is built for the prediction of each component. At last, these predicted values are superimposed to obtain the predicted values of the original sequence. This method makes an effective separation of the feature information of gas concentration time series and takes full advantage of multi-ELM prediction models with different parameters to achieve divide and rule. Comparative studies with existing prediction models indicate that the proposed model is very promising and can be implemented in one-step or multistep ahead prediction.

## 1. Introduction

It is well known that coalmine gas is one of the most important factors affecting coalmine security in production [1]. The accurate forecasting of coalmine gas concentration is the basis of gas outburst prediction, gas explosion prevention, ventilation design, and so on [2]. Therefore, enhancing research on reliable methods for coalmine gas prediction has positive significance on coalmine security [3]. However, the coalmine gas is influenced by geological conditions, occurrence of coal seam, gas content of coal and rock, permeability coefficient of coal and rock, the depth of coal, mining process, and so on. There are dynamic nonlinear relationships among these factors [4–6]. In addition, these factors are difficult to obtain in the coalmine, which bring great difficulties to the forecast of coalmine gas.

With regard to this, many researchers have turned their attention to gas time series prediction and many methods

have been proposed in the gas forecasting field. Models of gas concentration forecasts are largely based on chaos time series [7–9], grey theory [10, 11], fuzzy mathematics [12, 13], neural networks [14, 15], intelligent algorithm [16–18], support vector machine [19], gaussian process regression [20], and other mathematical or statistical methods [21, 22]. These methods have the same characteristics: the observed original values which are collected by gas sensors are usually directly used for building gas concentration forecasting models [23–26].

However, owing to the high-frequency, nonstationary fluctuations and chaotic properties of the gas concentration time series, a gas concentration forecasting model utilizing the original raw data often leads to an inability to provide satisfying forecast results. To solve this problem, before constructing a forecasting model, many studies would initially utilize information extraction techniques to extract features contained in data and use these extracted characteristics to construct independent forecasting model [27–30].

The useful or interesting information may not be observed directly from the observed original data but can be revealed in the extracted series through suitable signal processing methods. The wavelet decomposition and reconstruction can decompose the multicomponent signal information into a low-frequency approximate signal and a set of high-frequency detail signals. The low-frequency signal reacts to the inherent variation trend of the information while the high-frequency signal reacts to the stochastic disturbance influence of it. In view of the different rules of these two types of signals, different models and parameters can be utilized to independently predict these signals [31]. In this study, the improvement in the accuracy of a forecasting model is achieved by wavelet-based transform. First, we decompose the sample data sequence of gas concentration time series into several components of various time-frequency domains according to wavelet analysis; then we use the ELM particularly established to make forecasts for all domains based on these components; finally, we arrive at the algebraic sum of the forecasts. Thereby, a relatively accurate forecast of mine gas concentration could be achieved. Thus, by means of a combination of ELM with wavelet analysis, we arrive at a model to forecast gas concentration. Based on the research and application in the II826 Coal Face of Luling Coal Mine of Huaibei Mining Group Company in Anhui Province, China, it shows that this method can take advantage of different features contained in data and effectively predict the gas concentration.

The rest of this paper is organized as follows. Section 2 analyzes corresponding basic theories and methods. The proposed hybrid method based on wavelet transform and ELM is described in Section 3. The numerical results and discussions are presented in Section 4. Section 5 includes the conclusions of this paper.

## 2. Methodologies

*2.1. Wavelet Decomposition and Reconstruction.* The essence of the wavelet decomposition and reconstruction is to divide a set of primitive sequences containing comprehensive information into several groups of sequences with different characteristics by a group of band pass filters [32]. In this paper, the Mallat algorithm for discrete wavelet transform (DWT) is adopted as the wavelet decomposition and reconstruction method. Let  $Y = \{y_1, y_2, \dots, y_N\}$  be the original sequence, where  $N$  is the sequence length. The algorithm can be described as follows:

$$\begin{aligned} a_{j+1} &= H(a_j), \\ d_{j+1} &= G(a_j), \\ j &= 0, 1, \dots, J, \end{aligned} \quad (1)$$

where  $H(\cdot)$  and  $G(\cdot)$  represent the low-pass filter and high-pass filter, and  $a_{j+1}$  and  $d_{j+1}$  are the components of the original signal in adjacent frequency band under the resolution of  $2^{-(j+1)}$ , while  $a_{j+1}$  represent the low-frequency approximate component and  $d_{j+1}$  represent the high-frequency detail component. Let  $J$  be the decomposition level. We can get  $J$

detail components  $d_1, d_2, \dots, d_J$  and an approximate component  $a_J$ . For the length of the decomposed sequence is the half of that of the original one, binary interpolation method was adopted in the reconstruction sequence reconstructing [33],

$$\begin{aligned} A_j &= (H^*)^j a_j, \\ D_j &= (G^*)^{j-1} G^* d_j, \\ j &= 0, 1, \dots, J, \end{aligned} \quad (2)$$

where  $H^*$  and  $G^*$  are the dual operators of  $H$  and  $G$ . Detail sequences  $D_1, D_2, \dots, D_J$  and approximate sequence  $A_J$  are the reconstruction sequences of  $d_1, d_2, \dots, d_J$  and  $a_j$ ; they have the same length with original sequence, and the original sequence can be represented as the sum of reconstruction sequences,

$$Y = D_1 + D_2 + \dots + D_J + A_J. \quad (3)$$

It should be emphasized that the stationary wavelet transform (SWT) can also be used for frequency division. However, since the SWT is a nonorthogonal decomposition, there will be cross correlations among the resulted components. By contrast, the components will be independent when using DWT, which is convenient for obtaining the distribution of the original time series based on the forecasted distributions of the components [34].

*2.2. Extreme Learning Machine.* ELM learning algorithm is a kind of the feed forward neural network with a single hidden layer. And the algorithm solves the problems including the slow convergence speed, easily falling into local minimum, and so forth, which exist in the most neural network learning algorithms [35]. Both the theoretical analysis and the numerous experimental results have indicated that the ELM in most cases has better performance than that of the general back propagation neural networks (BPNN) learning algorithm [36]. Besides, with far less learning time than the support vector machine (SVM) algorithm [37], the ELM learning algorithm can achieve almost the same effect as SVM [38]. Therefore, the ELM learning algorithm is suitable for the practical application. And in view of this, this paper chooses the ELM as the base predictor.

Let  $N$  be training samples as  $\{(\mathbf{x}_k, y_k)\}_{k=1}^N$ , where  $\mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{kn}]^T \in R^n$  is the  $k$ th vector of the input sample and  $y_k \in R$  is the output variable corresponding to  $\mathbf{x}_k$ . Besides, the standard single layer feed forward network of the mathematical model with  $L$  hidden layer nodes can be described as follows:

$$o_k = \sum_{i=1}^L \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_k + b_i), \quad k = 1, \dots, N, \quad (4)$$

where  $o_k$  is the output vector of the  $k$ th sample,  $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$  is the input weight vector of the  $i$ th hidden layer node,  $\beta = [\beta_1, \beta_2, \dots, \beta_L]^T$  is the output weight vector of the hidden layer node,  $b_i$  is the bias of the  $i$ th hidden neuron,  $g(\cdot)$  is the activation function of the hidden layer, and  $\mathbf{w}_i \cdot \mathbf{x}_k$  is the inner product of  $\mathbf{w}_i$  and  $\mathbf{x}_k$ .

For  $N$  training samples, to achieve zero error learning, we need to meet  $\sum_{k=1}^N \|o_k - y_k\| = 0$ , and the condition is that (5) must be correct:

$$\sum_{i=1}^L \beta g(\mathbf{w}_i \cdot \mathbf{x}_k + b_i) = y_k, \quad k = 1, \dots, N. \quad (5)$$

Equation (5) can be described as follows in the form of matrix:

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T}. \quad (6)$$

In (6),

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_1 + b_L) \\ \vdots & \vdots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_N + b_L) \end{bmatrix}_{N \times L}. \quad (7)$$

The character of ELM is that the value of input weight  $\mathbf{w}_i$  and the value of hidden bias  $b_i$  are randomly assigned, and we can directly calculate the hidden layer output matrix  $\mathbf{H}$ . Therefore, training ELM is equivalent to obtaining the least-squares solution  $\hat{\boldsymbol{\beta}}$  of the linear equation  $\mathbf{H}\boldsymbol{\beta} = \mathbf{T}$ , and  $\hat{\boldsymbol{\beta}}$  can be described as follows:

$$\hat{\boldsymbol{\beta}} = \mathbf{H}^+ \mathbf{T}, \quad (8)$$

where  $\mathbf{H}^+$  is the Moore-Penrose generalized inverse of the matrix  $\mathbf{H}$  [38–40]. Since the output layer weight can be directly obtained, the ELM has the fast learning speed. At the same time, it also avoids the problem of easily falling into local minimum values due to the repeated iterations which is used by general neural network learning algorithms.

In summary, the ELM algorithm can be divided into the following steps [41].

*Step 1.* Assign the random value of input weight vector  $\mathbf{w}_i$  and threshold value  $b_i$  in the hidden layer. According to Bartlett's theory [42], small weights will get better generalization performance, so we use the random value between 0 and 1 in practice.

*Step 2.* Calculate output matrix  $\mathbf{H}$  in hidden layer.

*Step 3.* Calculate output weight vector  $\boldsymbol{\beta}$  based on (8) and establish ELM model.

*Step 4.* Obtain the predicted values based on input variables.

### 3. The Proposed Wavelet-ELM Gas Concentration Forecasting Method

Firstly, we use the Mallat algorithm to decompose and reconstruct the original gas time series. Then, the different prediction models are established for the low-frequency approximate sequence and high-frequency detail sequences. At last, the final predicted value was calculated by the sum of the results of every prediction model. The flowchart of the multistep ahead prediction framework is depicted in Figure 1.

The forecasting procedure is described as follows.

*Step 1.* Decompose and reconstruct the gas concentration time series into different component series (some detail

sequences  $D_1, D_2, \dots, D_J$  and an approximate sequence  $A_J$ ). In this process, there are two parameters that should be determined: the basic wavelet and decomposition level. Daubechies wavelet families are most appropriate for treating a nonstationary series and have been chosen as the basic wavelet in this paper [43], and the selection of Daubechies wavelet order is discussed in Section 4. The selection of decomposition level has a significant effect on the results obtained and this is also discussed in Section 4.

*Step 2.* C-C method is applied to get the optimum time delay  $\tau$  and embedding dimension  $m$  for each individual-decomposed component ( $D_1, D_2, \dots, D_J$ , and  $A_J$ ) [44].

*Step 3.* Input and output vectors for the ELM model are obtained through phase space reconstruction with the time delay  $\tau$  and embedding dimension  $m$ . The training process of each component is described in the previous section and the only parameter that should be ensured is the count of the hidden layer nodes.

*Step 4.* One-step ahead predicted value of each component series is obtained by trained ELM model.

*Step 5.* Predicted value of the gas concentration time series is obtained by superimposing the predicted values of all components.

*Step 6.* Determine whether the current reaches the need of look-ahead steps. If the condition is met, current predicted value is the last multistep forecasting value. Otherwise, append predicted value to the time series and go to Step 1.

## 4. Experimental Results

In this section, the effectiveness of the proposed method is evaluated by some experiments. The dataset used in our experiment is collected from the Coalmine Security Monitoring System named KJ98 in the II826 Coal Face of Luling Coal Mine of Huaibei Mining Group Company in Anhui Province, China. We test the proposed model with 1000 gas concentration samples; the first 800 data points are used as the training sample, and the remaining 200 data points are used as testing sample, and every data points scale is 10 seconds.

The prediction performance is evaluated by the mean absolute error (MAE), the mean absolute percentage error (MAPE), the root mean square error (RMSE), and the normalized root mean square error (NRMSE). The definitions of these criteria are as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

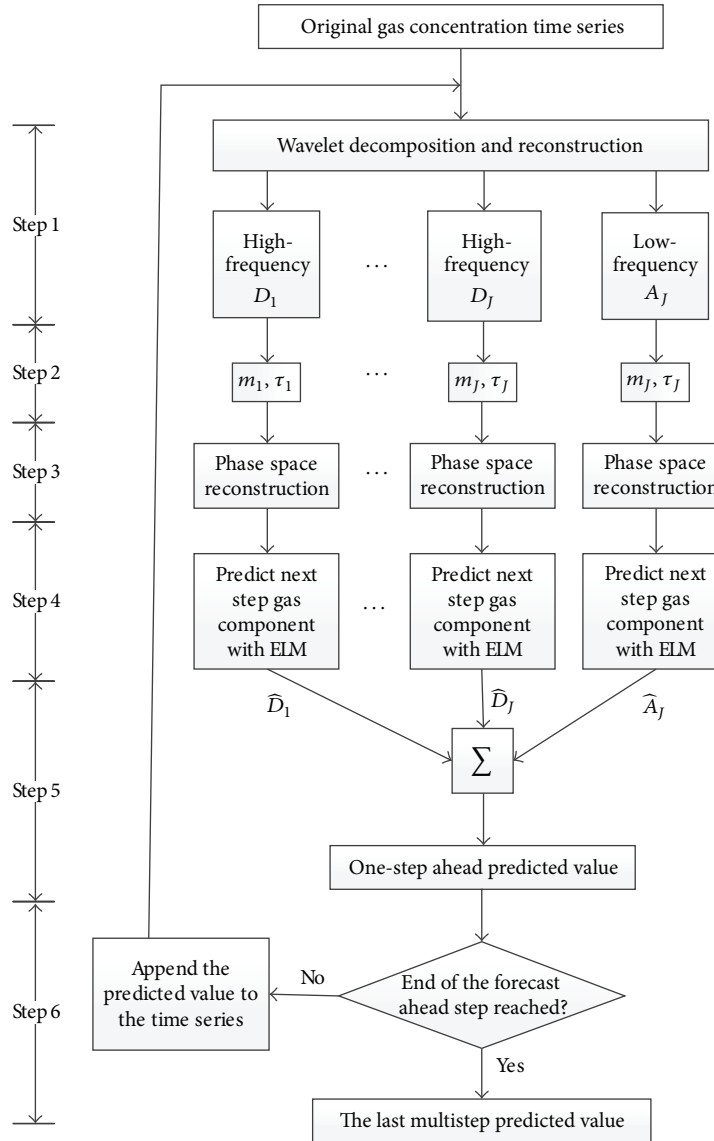


FIGURE 1: Gas concentration multistep ahead prediction framework based on WELM.

$$\text{NRMSE} = \sqrt{\frac{1}{n} \times \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}}, \quad (9)$$

where  $y$  and  $\hat{y}$  represent the actual and predicted value, respectively;  $\bar{y}$  is the mean value;  $n$  is the total number of data points in the test set. A smaller indicator means higher accuracy of the forecast.

The original data is firstly decomposed and reconstructed into 3 levels by wavelet transform based on db3 (db is the abbreviation of Daubechies, and db  $n$  means the Daubechies wavelets of order  $n$ ). The original gas concentration time series and time spectra of the subbands at the 3rd layer are shown in Figure 2. It can be seen from Figure 2 that the low-frequency signals embody the overall trend of the original gas concentration and several other subsignals represent

the uncertainty inference. The wavelet decomposition can well identify the different characteristics from the original data and benefit the gas concentration prediction through different ELM models.

Then, the ELM models are established to get the one-step look-ahead prediction component of each subsignal and their sum indicates the final short time gas concentration prediction value. In the ELM models, the count of hidden layer is firstly set to 20; we get the average value of ten time-independent predictions as the final predicted value. Table 1 gives the prediction performance of each component.

From Table 1, we will find that the order of the prediction performance is A3, D3, D2, and D1. That is because A3 is the smoothed low-frequency approximate component of the original gas concentration series which react to the inherent variation trend of the information and it can be easy to get a high fitness, while D3, D2, and D1 are the

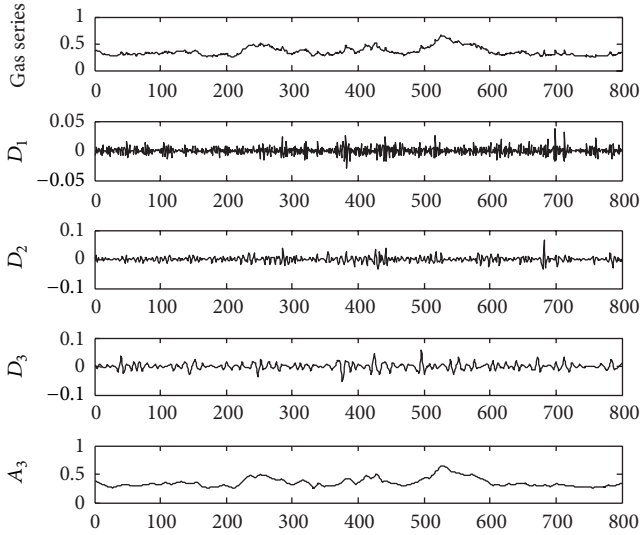


FIGURE 2: Wavelet decomposition and reconstruction of the original gas concentration time series.

TABLE 1: Prediction error of wavelet decomposition sequence.

Reconstructed sequence	MAE	MAPE	RMSE	NRMSE
$D_1$	0.0024	9.0188	0.0032	0.0378
$D_2$	0.0036	4.0222	0.0045	0.0333
$D_3$	0.0024	0.7098	0.0032	0.0233
$A_3$	0.0015	0.0037	0.0022	0.0035

high-frequency component of the original gas concentration series which react to the stochastic disturbance influence of the information; the higher the wavelet decomposition and reconstruction level, the stronger the detail component's randomness and the lower the prediction accuracy. This result coincides with the theory analysis. Figure 3 shows the chart of the final predicted value superimposed upon every subsequence and the actual data. From Figure 3, it can be seen that the final predicted data of the proposed method can fit the actual gas concentration data well.

To verify the effectiveness of the proposed method, routine methods are used to predict the gas concentration samples for comparison. These methods include classification and regression trees prediction model (CART) [45], back propagation neural network prediction model (BPNN) [36], support vector machine prediction model (SVM) [37], and extreme learning machine prediction model (ELM) [35]. In the BPNN prediction model, we get the average value of ten time-independent predictions as the final predicted value, the network hidden layer transfer function is Sigmoid function, the transport layer transfer function is Purelin function, the training algorithm is gradient descent algorithm with variable learning rate momentum, and the learning rate is set to 0.1. In the SVM model, we choose radial basis function as the kernel function, particle swarm optimization (PSO) algorithm is used to optimize the parameters of SVM [46], optimized parameters include the penalty parameter  $C$ , insensitive loss

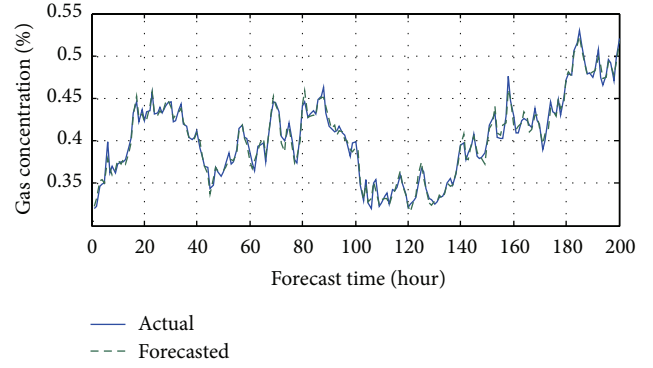


FIGURE 3: Actual value (the solid line) and one-step ahead WELM model output (the dashed line).

parameter  $\epsilon$ , and kernel parameter  $\sigma$ , the number of particles is initialized to 30,  $c_1 = c_2 = 2$ ,  $\omega = 0.7$ , iteration number is set to 1000, and initialization range of  $C$ ,  $\epsilon$ , and  $\sigma$  is set to  $[1, 1000]$ ,  $[0.001, 0.1]$ , and  $[0, 10]$ , respectively. The comparison of predicted results is shown in Table 2.

From Table 2, the forecasting accuracy of the WELM model is more promising than the results of previous works. Improvement in the MAE of the proposed approach with respect to the four previous approaches (CART, BPNN, SVM, and ELM) is 80.32%, 74.74%, 74.61%, and 74.07%, respectively. Improvement in the MAPE of the proposed approach with respect to the four previous approaches is 80.23%, 74.43%, 74.32%, and 73.72%, respectively. Improvement in the RMSE of the proposed approach with respect to the four previous approaches is 80.57%, 74.69%, 74.48%, and 74.26%, respectively. Improvement in the MAE of the proposed approach with respect to the four previous approaches is 80.64%, 74.79%, 74.58%, and 74.29%, respectively. From the column of training time and testing time, it can be seen that WELM method spent only 0.0155 s CPU time for training and 0.0073 s CPU time for testing, it is much less than CART, BPNN, and SVM algorithm, it is slightly more than ELM method because of the extra processing time of wavelet transform, this time is far less than the sampling interval, and it can be trained every time when the new data arrived that means the WELM model is suitable for automatic adjustment according to the time, while other models are not suitable for doing so due to long training time. Supplementary note: SVM algorithm requires a parameter optimization process which is too time-consuming, and the calculated training time is out of the statistical significance, so we did not list the corresponding calculation time.

Figure 4 shows the multistep (from 1 to 24 steps ahead) ahead forecast accuracy of the expectation value measured by the MAPE, the CART, BPNN, SVM, and ELM models used here for comparison.

For multistep ahead forecast, the gas concentration forecast is carried out by recursively taking the previous forecast values which is described in Section 3. That means the error will be recursively along with the increased steps, so the error will be increased according to the look-ahead step. According

TABLE 2: Comparison of different prediction methods.

Method	MAE	MAPE	RMSE	NRMSE	Training time (s)	Testing time (s)
CART	0.0249	0.0622	0.0314	0.0470	0.0280	0.0084
BPNN	0.0194	0.0481	0.0241	0.0361	0.5531	0.0650
SVM	0.0193	0.0479	0.0239	0.0358	N/A	N/A
ELM	0.0189	0.0468	0.0237	0.0354	<b>0.0063</b>	<b>0.0054</b>
WELM	<b>0.0049</b>	<b>0.0123</b>	<b>0.0061</b>	<b>0.0091</b>	0.0155	0.0073

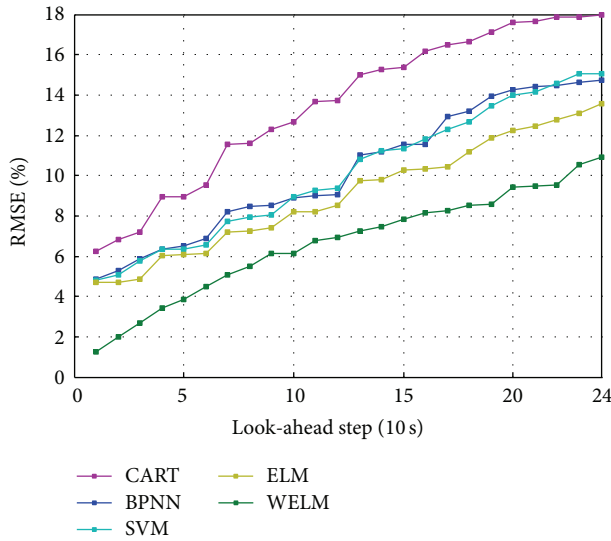


FIGURE 4: Multistep ahead forecast accuracy comparison with different methods.

to Figure 4, BPNN, SVM, and ELM have obviously better forecast accuracy than the CART model. Compared with the other four models, the proposed WELM model can improve the forecast accuracy significantly.

To illustrate the influence of the hidden layer nodes of ELM, the MAPE of WELM using different hidden layer nodes from 1 to 50 is shown in Figure 5. In the figure, we will find that the MAPE values are comparatively higher for less number of hidden nodes (from 1 to 5), while the forecast accuracy is flat which shows that the model performs equally well for different hidden nodes if they have high values, and this fact is equal to [35] that means ELM generalization performance is independent of the number of hidden nodes if the number of hidden nodes is considerably large, so, in the practice, we must choose hidden nodes higher than 5.

To illustrate the influence of the orders of Daubechies mother wavelets, the MAPE of WELM using Daubechies wavelets of different orders from 1 to 45 is shown in Figure 6. In the figure, db is the abbreviation of Daubechies, and db  $n$  means the Daubechies wavelets of order  $n$ . It can be seen from the figure that MAPE according to  $n$  from 1 to 25 is decreased sharply, and the other mother wavelets have almost the same performance, particularly in the tail, so, in the practice, it is better to choose the Daubechies orders from 25 to 45. It should be noted that the MAPE of db1 which has the worst forecasting accuracy than others is only 2.03%, compared in

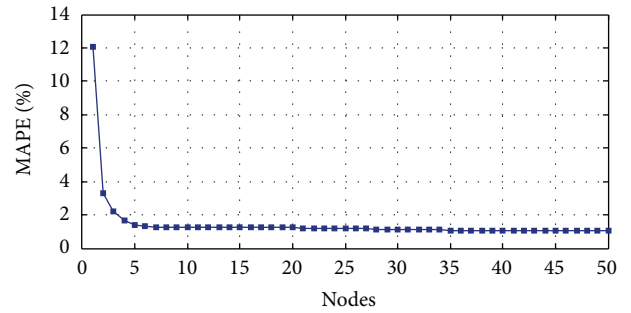


FIGURE 5: Performance comparison with different ELM hidden layer nodes.

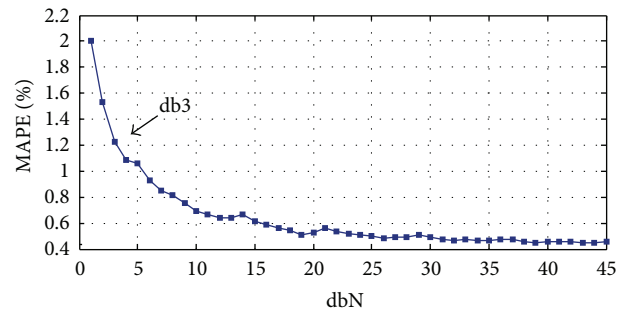


FIGURE 6: Performance comparison with different orders of the Daubechies wavelets.

Table 2, and this is also more promising than other methods (CART, BP, SVM, and ELM) without wavelet transforms; the paper discussed above in Table 2 is using db3, which is comparably higher than Daubechies orders from 4 to 45, so there is still a lot of promising space when using other Daubechies orders.

Furthermore, wavelet decomposition level has influence on the prediction results. Specifically, the higher the wavelet decomposition level is, the smoother and more stable the approximation signal is, and the prediction accuracy is higher as well. However, with the increase of decomposition layers, the number of detail signals will also increase, and the errors will be superimposed because the number of detail signals has increased. More decomposition layers will bring more prediction errors, so the forecast accuracy will not increase with the increase of decomposition level. As a result, the prediction accuracy will fluctuate in a certain range.

Figure 7 shows the prediction errors of WELM using db3 wavelet decomposition when decomposition level is

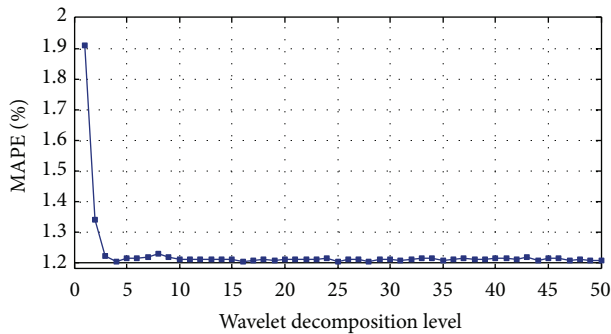


FIGURE 7: Performance comparison with different wavelet decomposition levels.

from 1 to 50. It can be seen from the figure that MAPE according to wavelet decomposition level from 1 to 3 is decreased sharply. The reason is that the forecast performance promising is source by the extraction of the stochastic disturbance influence, but the small wavelet decomposition level cannot extract significantly the random component which still remains in the approximate component. The MAPE according to wavelet decomposition level higher than 3 is fluctuating around 1.21%, according to the calculated performance; we can use 3 to 5 as the decomposition layers in the practical application.

## 5. Conclusions

The focus of this paper is to combine wavelet transform and extreme learning machine for predicting coalmine gas concentration. The coalmine gas time series is influenced by geological conditions, occurrence of coal seam, gas content of coal and rock, mining process, and many other factors. As a result, it shows strong nonstationary and stochastic characteristic. Using a single model to forecast gas concentration is equal to forecasting a mixed signal by unified methods and parameters. Meanwhile, the random factors of gas concentration sequence will have an impact on determination of model parameters and final prediction results. The wavelet decomposition and reconstruction can decompose the multicomponent signal information into a low-frequency approximate signal which reacts to the inherent variation trend and a set of high-frequency detail signals which react to the stochastic disturbance influence. Different ELM models with different parameters can be utilized to predict these new signals independently. The proposed model is compared with CART, BPNN, SVM, and ELM for one-step and multistep prediction. Simulation results show that the ELM model with wavelet-based preprocessing greatly outperforms the other four models. Furthermore, we still discuss the selection principles of ELM hidden layer nodes, the orders of Daubechies mother wavelets, and the wavelet decomposition level. For coalmine gas concentration time series, we must choose hidden nodes higher than 5, it is better to choose the Daubechies orders from 25 to 45, and we can use 3 to 5 as the decomposition layers in the practical application for good performance and accuracy.

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## Acknowledgment

This research was partially supported by the National Natural Science Foundation of China under Grant no. 61379100.

## References

- [1] S. N. Zhou and B. Q. Lin, *Coalbed Gas Occurrence and Flow Theory*, China Coal Industry Publishing House, Beijing, China, 1999.
- [2] J. Q. Zhang, "The application of analytic hierarchy process in mine gas prevention system," *Procedia Engineering*, vol. 26, pp. 1576–1584, 2011.
- [3] W. Yang, B. Lin, C. Zhai, X. Li, and S. An, "How in situ stresses and the driving cycle footage affect the gas outburst risk of driving coal mine roadway," *Tunnelling and Underground Space Technology*, vol. 31, pp. 139–148, 2012.
- [4] K. X. Wang, X. H. Fu, Y. A. Zhou, Y. HE, and H. Wu, "Dynamic development characteristics of amounts of gas and levels of pressure in the Pan-1 coal mine of Huainan," *Mining Science and Technology*, vol. 19, no. 6, pp. 740–744, 2009.
- [5] W. Wang and J. Yan, "The geological factor analysis of influenced Tianchi coal mine gas occurrence," *Procedia Engineering*, vol. 45, pp. 317–321, 2012.
- [6] L. Wang, Y. Cheng, L. Wang, P. Guo, and W. Li, "Safety line method for the prediction of deep coal-seam gas pressure and its application in coal mines," *Safety Science*, vol. 50, no. 3, pp. 523–529, 2012.
- [7] J. Cheng, J. Y. Bai, and J. S. Qian, "Short-term forecasting method of coal mine gas concentration based on chaotic time series," *Journal of China University of Mining and Technology*, vol. 2, no. 37, pp. 231–235, 2008.
- [8] X. Y. Cui and G. Q. Zhao, "Application of chaos forecast of mine gas density monitoring," *Journal of Liaoning Technical University*, vol. 27, no. 1, pp. 184–186, 2008.
- [9] X. J. Ni, "Forecasting of gas emissions based on chaotic time series," *Science & Technology Information*, vol. 31, no. 1, pp. 20–34, 2008.
- [10] J. N. Pan, Z. P. Meng, and Y. C. Liu, "Grey smoothing model for predicting mine gas emission," *Journal of China University of Mining & Technology*, vol. 13, no. 1, pp. 76–78, 2003.
- [11] P. Lu, Y. Ma, and X. Zhou, "Research and application on dynamic forecasting model of gas consistence in top corner," *Journal of the China Coal Society*, vol. 31, no. 4, pp. 461–465, 2006.
- [12] J. Y. Zhang, J. Cheng, and Y. H. Hou, "Forecasting coalmine gas concentration based on adaptive neuro-fuzzy inference system," *Journal of China University of Mining and Technology*, vol. 36, no. 4, pp. 494–498, 2007.
- [13] X. Sun, "Coalmine gas prediction based on fuzzy neural network," *Journal of Anhui University of Technology*, vol. 29, no. 3, pp. 229–232, 2012.
- [14] J. H. Zhao and Q. Xu, "Short-term prediction of coalmine gas concentration based on chaotic series and wavelet neural network," in *Proceedings of the International Conference on*

- Artificial Intelligence and Computational Intelligence (AICI '10)*, vol. 3, pp. 240–244, IEEE, Sanya, China, October 2010.
- [15] L. Gao, Y. Hu, and H. Yu, "Prediction of gas emission time series based on W-RBF," *Journal of the China Coal Society*, vol. 33, no. 1, pp. 67–70, 2008.
  - [16] X. Wang, J. Liu, and J. Lu, "Gas concentration forecasting approach based on wavelet transform and optimized predictor," *Journal of Basic Science and Engineering*, vol. 19, no. 3, pp. 499–508, 2011.
  - [17] W. T. Ma, "Gas emission forecast based on wavelet transform and genetic-least square support vector machine," *Journal of Mining and Safety Engineering*, vol. 26, no. 4, pp. 524–528, 2009.
  - [18] Q. Wang and J. Cheng, "Forecast of coalmine gas concentration based on the immune neural network model," *Journal of the China Coal Society*, vol. 33, no. 6, pp. 665–669, 2008.
  - [19] M. Y. Qiao, X. P. Ma, and J. Y. Lan, "Time series short-Term gas prediction based on weighted LS-SVM," *Journal of Mining and Safety Engineering*, vol. 28, no. 2, pp. 310–314, 2011.
  - [20] D. W. Dong, S. G. Li, X. T. Chang, and H. F. Lin, "Prediction model of gas concentration around working face using multivariate time series," *Journal of Mining and Safety Engineering*, vol. 29, no. 1, pp. 135–139, 2012.
  - [21] L. Gao and H. Yu, "Prediction of gas emission based on information fusion and chaotic time series," *Journal of China University of Mining and Technology*, vol. 16, no. 1, pp. 94–96, 2006.
  - [22] L. Shao and G. Fu, "Dynamic prediction technology for gas based on data fusion theory," *Journal of the China Coal Society*, vol. 33, no. 5, pp. 551–555, 2008.
  - [23] D. Li, Y. Cheng, L. Wang, H. Wang, L. Wang, and H. Zhou, "Prediction method for risks of coal and gas outbursts based on spatial chaos theory using gas desorption index of drill cuttings," *Mining Science and Technology*, vol. 21, no. 3, pp. 439–443, 2011.
  - [24] Q. Meng, X. P. Ma, and Y. Zhou, "Prediction of mine gas emission rate using support vector regression and chaotic particle," *Journal of Computers*, vol. 8, no. 11, pp. 2908–2915, 2013.
  - [25] F. Z. Wang and W. Z. Liu, "Prediction strategy of coal and gas outburst based on artificial neural network," *Journal of Computers*, vol. 8, no. 1, pp. 240–247, 2013.
  - [26] W. Ting, C. Lin-Qin, F. Yao, and Z. Tingcheng, "A wavelet-based robust relevance vector machine based on sensor data scheduling control for modeling mine gas gushing forecasting on virtual environment," *Mathematical Problems in Engineering*, vol. 2013, Article ID 579693, 4 pages, 2013.
  - [27] N. A. Shrivastava and B. K. Panigrahi, "A hybrid wavelet-ELM based short term price forecasting for electricity markets," *Electrical Power and Energy Systems*, vol. 55, no. 7, pp. 41–50, 2014.
  - [28] Y. Q. Liu, J. Shi, Y. P. Yang, and W. J. Lee, "Short-term wind-power prediction based on wavelet transform-support vector machine and statistic-characteristics analysis," *IEEE Transactions on Industry Applications*, vol. 48, no. 4, pp. 1136–1141, 2012.
  - [29] G. Capizzi, C. Napoli, and F. Bonanno, "Innovative second-generation wavelets construction with recurrent neural networks for solar radiation forecasting," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 11, pp. 1805–1815, 2012.
  - [30] A. U. Haque, P. Mandal, J. L. Meng, A. K. Srivastava, T. L. Tseng, and T. Senjyu, "A novel hybrid approach based on wavelet transform and fuzzy ARTMAP networks for predicting wind farm power production," *IEEE Transactions on Industry Applications*, vol. 49, no. 5, pp. 2253–2261, 2013.
  - [31] G. G. He, S. F. Ma, and Y. Li, "A study on forecasting for time series based on wavelet analysis," *Acta Automatica Sinica*, vol. 29, no. 1, pp. 1012–1014, 2012.
  - [32] A. S. Pandey, D. Singh, and S. K. Sinha, "Intelligent hybrid wavelet models for short-term load forecasting," *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1266–1273, 2010.
  - [33] S. G. Mallat, "Theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.
  - [34] M. Yang, S. Fan, and W. J. Lee, "Probabilistic short-term wind power forecast using componential sparse Bayesian learning," *IEEE Transactions on Industry Applications*, vol. 49, no. 6, pp. 2783–2791, 2013.
  - [35] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.
  - [36] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning Internal Representations by Error Propagation*, MIT Press, Cambridge, Mass, USA, 1986.
  - [37] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
  - [38] Q. Wang and S. Chen, "Ensemble learning of ELM regressors based on  $l_1$ -regularization," *Journal of Computer Research and Development*, vol. 49, no. 12, pp. 2631–2637, 2012.
  - [39] C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and Its Applications*, John Wiley & Sons, New York, NY, USA, 1971.
  - [40] D. Serre, *Matrices: Theory and Applications*, vol. 216 of *Graduate Texts in Mathematics*, Springer, New York, NY, USA, 2002.
  - [41] L. Yang and R. Zhang, "Online sequential ELM algorithm and its improvement," *Journal of Northwest University (Natural Science Edition)*, vol. 42, no. 6, pp. 885–896, 2012.
  - [42] P. L. Bartlett, "The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 525–536, 1998.
  - [43] A. Cohen, I. Daubechies, and J. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. 45, no. 5, pp. 485–560, 1992.
  - [44] H. S. Kim, R. Eykholt, and J. D. Salas, "Nonlinear dynamics, delay times, and embedding windows," *Physica D*, vol. 127, no. 1–2, pp. 48–60, 1999.
  - [45] L. Breiman, J. Friedman, C. Stone, and R. A. Olshen, *Classification and Regression Trees*, Chapman & Hall, 1984.
  - [46] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, vol. 4, pp. 1942–1948, Perth, Australia, November–December 1995.





# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

