

## Prediction of Coal Face Gas Concentration by Multi-Scale Selective Ensemble Hybrid Modeling

WU Xiang<sup>1,3\*</sup>, YANG Zhong-ru<sup>2</sup>, Zhang Li<sup>4</sup> and Pilati Silvia<sup>5</sup>

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

<sup>2</sup> China Huajin Coal Energy Co., Ltd. Hejin, Shanxi 043300, China

<sup>3</sup> School of Medical Informatics, Xuzhou Medical College, Xuzhou, Jiangsu 221116, China

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

<sup>5</sup> Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003, United States

Received 15 December 2013; Accepted 23 June 2014

### Abstract

A selective ensemble hybrid modeling prediction method based on wavelet transformation is proposed to improve the fitting and generalization capability of the existing prediction models of the coal face gas concentration, which has a strong stochastic volatility. Mallat algorithm was employed for the multi-scale decomposition and single-scale reconstruction of the gas concentration time series. Then, it predicted every subsequence by sparsely weighted multi unstable ELM (extreme learning machine) predictor within method SERELM (sparse ensemble regressors of ELM). At last, it superimposed the predicted values of these models to obtain the predicted values of the original sequence. The proposed method takes advantage of characteristics of multi scale analysis of wavelet transformation, accuracy and fast characteristics of ELM prediction and the generalization ability of L1 regularized selective ensemble learning method. The results show that the forecast accuracy has large increase by using the proposed method. The average relative error is 0.65%, the maximum relative error is 4.16% and the probability of relative error less than 1% reaches 0.785.

*Keywords:* Gas Concentration, Multi-scale, Selective Ensemble Learning, Hybrid Modeling

### 1. Introduction

Accurately predicting real-time trends of gas concentration is the prerequisite to control coal mine gas accident. At the same time, accurate prediction, real-time monitoring and the control of coalmine gas concentration are important measures to prevent gas explosion, gas outburst and ensure miners' safety [1]. As known to all, the installed safety monitoring system has helped collect vast amounts of data concerning coal mine safety in most coal mines. Moreover, it is significant for making an accurate and effective trends prediction of the mine gas concentration, if the collected data can be fully used to know the rule and trend of gas movement.

It is a fact that the change of gas concentration is affected by gas content of coal seam, gas drainage volume of coal seam, air volume in the work area and many other factors. Also, there are complicated nonlinear relations between the factors. It is a feasible method to predict real-time trends of gas concentration by unified modeling of combining all related factors. However, in practical applications, most coal mines can only provide the historical data of gas concentration and can hardly provide the data of all related factors.

In recent years, on the basis of the mathematical methods

such as chaos theory, grey theory, fuzzy mathematics, neural network, intelligent algorithm and under the help of the computer, the prediction method about gas concentration time series has achieved some application effect. Papers [2-4] show that the gas concentration sequence is chaotic time series. In [2], a prediction model was constructed by an adding-weight one-rank local-region method in the reconstruction phase space. In [3], a prediction model was built using time series and adaptive fuzzy reasoning neural system. In [4], it proposed a max Lyapunov index model. In LS-SVM case, the sparseness and robustness may lose, in [5], it proposed the weighted LS-SVM to overcome these draw backs. In [6], the chaotic phase space reconstructive method was used to reconstruct the sample space of gas concentration in multivariate time series and the Gaussian process regression model was used to predict the gas concentration around the work face. These methods provide a good guidance for gas data prediction.

However, gas concentration in coalface is affected by many nonlinear factors. As a result, it shows strong non-stationary and stochastic volatility. In addition, mining areas have complex environmental factors. Thus, using a single model to predict gas concentration is equal to using a group of gas time series including comprehensive information to make a general prediction for gas concentration by unified methods and parameters [7]. Meanwhile, the random factors of gas concentration sequence will have an impact on

\* E-mail address: dmshxzm@163.com

determination of model parameters and final prediction results. Therefore, the prediction method using a single model brings a problem that the prediction result has a good fitting degree but a poor generalization ability. In paper[8], the method combining wavelet transform and particle swarm optimization SVM is shown. The modeling method based on combining prediction has achieved good predictions. Yet, the method with the SVM will bring lots of problems, for example, hardly setting up parameters, long computing time. Moreover, the decline of the prediction accuracy appears in the high-frequency sequence with small differences in the sample. Besides, when the multi-scale method with the SVM is used to make a prediction, the final prediction error will be enlarged owing to the superposition of each sequence's prediction error. Because of these problems, this method will make overall prediction accuracy affected and is hardly spread in the practical application. The selective ensemble learning is the machine learning method that can effectively improve the generalization ability and reduce the prediction error. It can also overcome the drawbacks including the slow prediction speed and occupying too much storage space that bring by the integration learning. However, the selective ensemble learning like GASEN and SVM-PRE have their own problems, for instance, parameter selection, long learning time. Therefore, it is not conducive to real-time prediction of gas concentration. Sparse ELM regression integration learning algorithm makes full use of characters of ELM's fast prediction and ELM's strong generalization ability. By using the L1 regularization constraint method, the algorithm avoid the difficult problem in selecting diversity predictors. In general, the algorithm is suitable for real-time prediction of gas concentration. In view of the above analysis, this article proposes a new prediction method called prediction of coal face gas concentration by multi-scale selective hybrid modeling. By wavelet transform, the new method makes the multi-scale decomposition for original gas time series and restructures the results from the multi-scale decomposition of original gas time series to get subsequences under different resolutions. Then, on basis of using the ELM of L1 regularization, the new method improves the prediction accuracy and the generalization ability of each subsequence[9]. Finally, the new method will improve the overall prediction result.

## 2. Multi-scale Method Analysis

The gas concentration in coalface is affected by many factors. Further, these geological factors including gas content of coal seam, the permeability of coal seam and surrounding rock and so on compose the trend component of gas concentration time series. Mining factors including mining scale, mining methods and technology, air volume and so on compose the volatility component of gas concentration time series. In fact, gas concentration time series reflect the overall performance caused by a variety of internal and external factors. By the method called decomposition and reconstruction of wavelet, a group of the original time series including comprehensive information can be decomposed into several groups of time series having different characters. Specifically, one of the groups reflects the trend of internal changes about original time series and this is low-frequency approximation signal. The other groups reflect the influence caused by random disturbance and they are high-frequency detail signals[10]. The Mallat

algorithm is adopt 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 follow.

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

$H(\cdot)$  and  $G(\cdot)$  represent the low-pass filter and high-pass filter.  $a_{j+1}$  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  $d_1, d_1 \dots d_J$  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 adapted in the reconstruction sequence reconstructing [11].

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

$H^*$  and  $G^*$  are the dual operators of  $H$  and  $G$ . Detail sequences  $D_1, D_1 \dots D_J$  and approximate sequence  $A_J$  are the reconstruction sequences of  $d_1, d_1 \dots d_J$  and  $a_J$ . They have the same length with original sequence. And the original sequence can be represented as the sum of reconstruction sequences.

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

## 3. ELM Selective Ensemble Learning based on L1 Regularization Constraint

### 3.1 Extreme learning machine ELM

ELM[12] 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, etc, which is exist in most neural network learning algorithms. 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 BP learning algorithm. Besides, with far less learning time than the SVM algorithm[9], the ELM learning algorithm can achieve almost the same effect as it. 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 in the selective ensemble learning.

Let  $N$  training samples as  $\{(\mathbf{x}_k, y_k)\}_{k=1}^N$ , where  $\mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{km}]^T \in \mathbb{R}^n$  is the  $k$  th vector of the input sample and  $y_k \in \mathbb{R}$  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 follow:

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

where  $O_k$  is the output vector of the  $k$  th sample,  $w_i = [w_{i1}, w_{i2}, \dots, w_{im}]^T$  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  $w_i \cdot x_k$  is the inner product of  $w_i$  and  $x_k$ .

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

$$\sum_{i=1}^L \beta g(w_i \cdot x_k + b_i) = y_k, k = 1, 2, \dots, N. \quad (5)$$

The equation (5) can be described as follow in form of matrix.

$$H\beta = T \quad (6)$$

In equation (6)

$$H = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & L & g(w_L \cdot x_1 + b_L) \\ M & M & M \\ g(w_1 \cdot x_N + b_1) & L & g(w_L \cdot x_N + b_L) \end{bmatrix}_{N \times L}.$$

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

$$\hat{\beta} = H^+ T \quad (7)$$

where  $H^+$  is the Moore-Penrose generalized inverse of the matrix  $H$ . 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 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[13]:

Step 1) Assign the value of input weight vector  $w_i$  in the hidden layer and randomly assign threshold value  $b_i$ .

Step 2) Calculate output matrix  $H$  in hidden layer.

Step 3) Calculate weight vector  $\beta$  based on the equation(7) and establish ELM model.

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

### 3.2 Sparse ELM regression integration learning algorithm[9]

With the given training samples  $\{(x_k, y_k)\}_{k=1}^N$ , where  $x_k \in \mathbb{R}^n$  and  $y_k \in \mathbb{R}$ . Determine the number of primary based predictor  $m$ , we can get a group of predictors  $\{ELM_i\}_{i=1}^m$  by randomly selecting parameters, where  $w = [w_1, w_2, \dots, w_m]^T$  is the weight vector of these predictors in the group. Therefore, the output from the integration learning for the input  $x_k$  can be described as follows,

$$H(x_k) = \sum_{i=1}^m w_i ELM_i(x_k), k = 1, 2, \dots, N \quad (8)$$

where  $ELM_i(x_k)$  is the result of  $i$  th ELM predictor for the input  $x_k$ .

The aim of regularization theory is to minimize the sum of empirical risk and regularization term,

$$\begin{aligned} \min R_{emp}[f] + \lambda \|w\|_1 \\ = \min \frac{1}{N} \sum_{k=1}^N l(x_k, y_k, f(x_k)) + \lambda \|w\|_1 \end{aligned} \quad (9)$$

where  $\|w\|_1$  is the  $l_1$  norm of the weight vector  $w$ ,  $\lambda$  is the penalty factor which is used to make a balance between  $R_{emp}[f]$  and the sparseness,  $l(x_k, y_k, f(x_k))$  is the loss function. In this paper, we use the quadratic loss function and the equation (9) can be converted as follows.

$$\begin{aligned} \min \frac{1}{N} \sum_{k=1}^N (f(x_k) - y_k)^2 + \lambda \|w\|_1, \\ s.t. w_i \geq 0, i = 1, 2, \dots, m. \end{aligned} \quad (10)$$

The weight vector  $w$  which is obtained by calculating is sparse. These ELM predictors whose vector elements are unequal to zero ( $w_i \neq 0$ ) will be selected as predictors participating in the integration. Finally, the final predictive result of the selective learning can be calculated by the equation (8).

### 4. Prediction Method of Multi-scale Selective Ensemble Hybrid Model

First, with the Mallat algorithm, make the multi-scale decomposition for gas time series and reconstruct the result of the decomposition. Then, make a mathematic model and predict for low-frequency approximation signals. Besides, do the same thing on high-frequency detail signals. Moreover, the final prediction result can be obtained by calculating the sum of the predicted data of each subsequence. The prediction framework can be described as Fig.1.

The specific steps of the multi-scale selective ensemble hybrid modeling prediction are shown as follows.

Step 1) With Mallat algorithm, gas concentration time series are decomposed into high-frequency components including  $d_1, d_2, \dots, d_J$  and the low-frequency component  $a_J$ , according to equation(1). Then, the component in each group is reconstructed alone to obtain the detail sequences including  $D_1, D_2, \dots, D_J$  and the approximation sequence  $A_J$  according to equation (2).

Step 2) Do delay mapping  $D: I \rightarrow I^u$  for each sequence. Because of this, the one-dimensional gas concentration subsequence can be mapped into  $u$  dimensional sequences with delay from 1 to  $u$  to construct the input and output variables of the ELM model.

Step 3) For each subsequence, Train  $m$  ELM predictors by randomly giving input weight vector  $w_i$ , threshold value  $b_i$  and  $L$  the number of hidden layers.

Step 4) For each subsequence, the weight of each predictor can be obtained by the method of L1

regularization. Besides, the prediction result of each subsequence can be obtained as well.

Step 5) The final prediction result of gas concentration time series can be obtained by superimposing the predictive result of each subsequence.

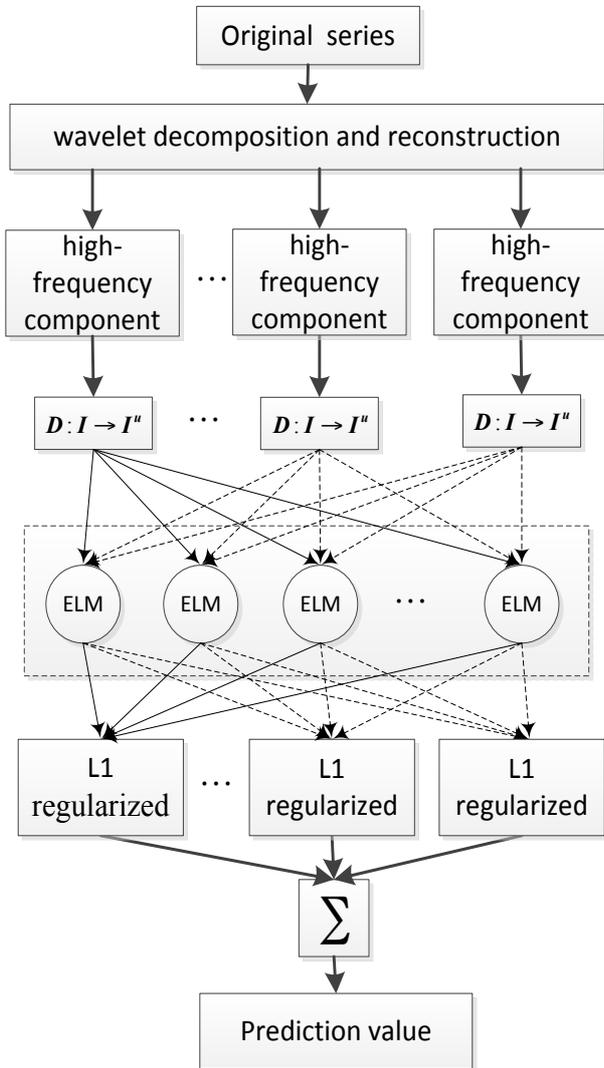


Fig.1. Hybrid modeling framework for multi-scale selective ensemble forecasting

### 5. Experiment Results

We test the collected models including 1400 gas concentration samples coming from the II826 Coal Face of Luling coal mine of Huaibei Mining Group Company in Anhui Province. We select the first 1000 samples as training samples, data from 1001 to 1200 as validation samples, and the last 200 data are selected as test samples. Firstly, the original sequence is decomposed and reconstructed by wavelet, wavelet using db4, decomposition

level is set 3. The reconstructed sequences are shown in Fig.2.

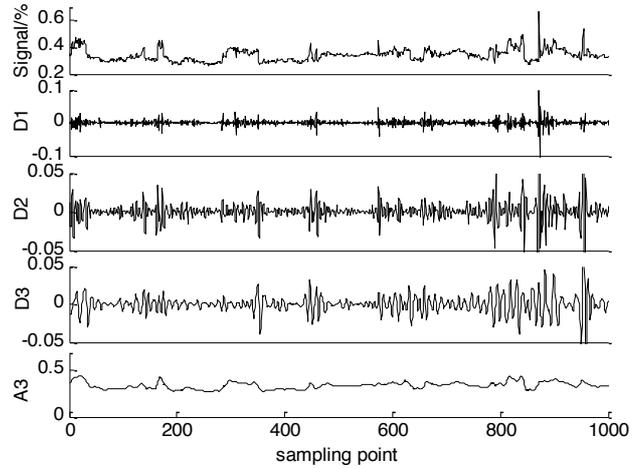


Fig.2. Wavelet decomposition and refactoring of the original gas concentration

### 5.1 Parameter settings

The input dimension of each ELM base predictor is randomly selected from (4, 5, 6, 7, 8, 9), the number of nodes in the hidden layer is randomly selected from (10, 20, 30, 40, 50, 60). The sigmoid function is used as the activation function. And, initial number of the ELM used as the integrated prediction is 500. Next, in order to avoid the effect of randomly setting parameters, we do fifty experiments and use the average value of the fifty prediction values as the final prediction result. For comparison, this paper uses MRE(Mean Relative Error) and NRMSE(Normalized Root Mean Square Error) as the evaluation criteria. The equations are shown as follows.

$$MRE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (11)$$

$$NRMSE = \sqrt{\frac{1}{n} \times \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})}} \quad (12)$$

### 5.2 Result analysis

Fig.3 shows the prediction results and the actual value of wavelet decomposition subsequences. And Tab.1 shows the prediction error of each subsequence.

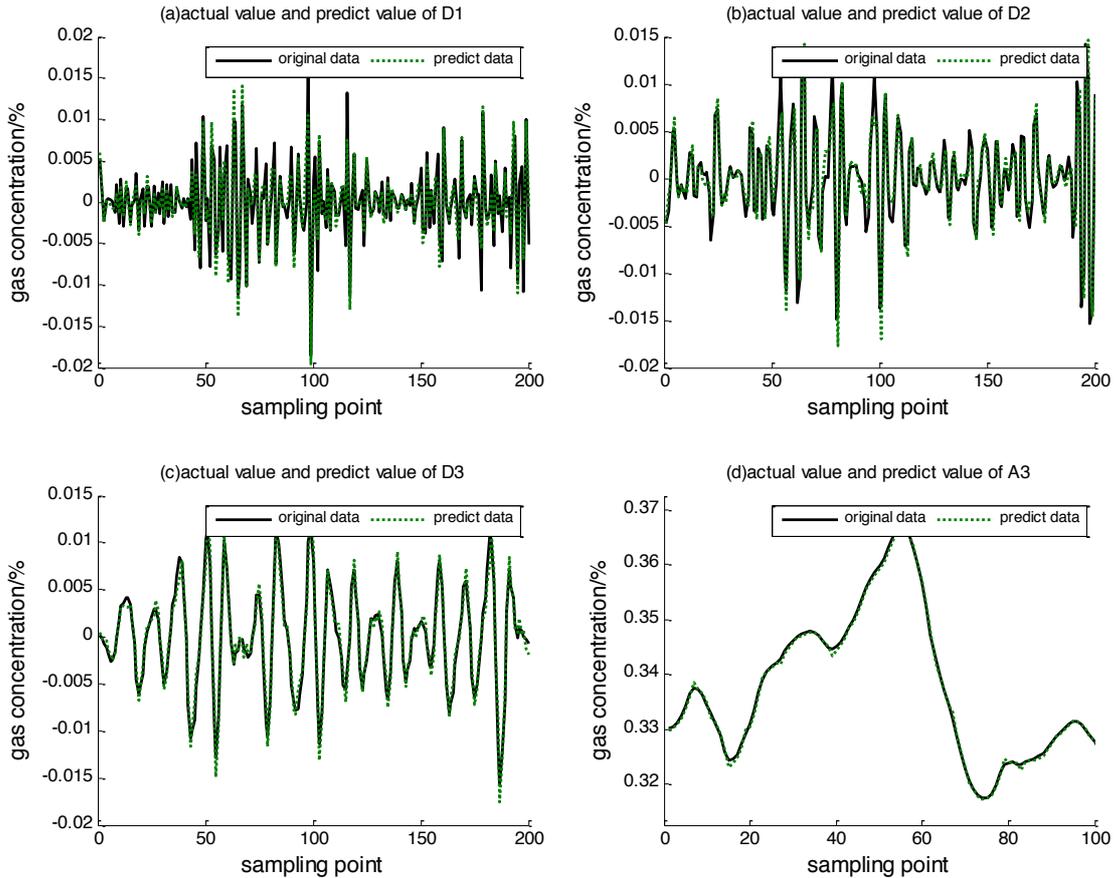


Fig.3. Comparison chart of each sub-series forecasting results

Fig.3 (a) is the high-frequency detail sequence  $D_1$ . Fig.3 (b) is high-frequency detail sequence  $D_2$ . Fig.3 (c) is the high-frequency detail sequence  $D_3$ . Fig.3 (d) is the prediction curve of the low-frequency trend component  $A_3$ . As Fig.3 (d) shows, the curve  $A_3$  which retains gas concentration time series internal trend information is smoothed, so it gets a high prediction accuracy. In addition, Tab.1 shows that the MRE(Mean Relative Error) of the curve  $A_3$  is only 0.13%, these indicate that fitting and generalization capability of ELM selective ensemble learning regularized by L1 is strong.  $D_1$ ,  $D_2$  and  $D_3$  is different high-frequency detail components in three scales separately. And, they have small values, carry many random and noise components. Because their actual values are close to zero,  $D_1$ ,  $D_2$  and  $D_3$  have large MRE(Mean Relative Error) respectively being 79.72%, 40.62% and 24.18%. However, they have a small proportion in gas time series and have little effect on the overall prediction accuracy. The MRE(Mean Relative Error) of the overall prediction is 0.65% and the NRMSE(Normalized Root Mean Square Error) of the overall prediction is 0.0087. Also, both of them indicate that the method proposed by this paper has good prediction results.

Table.1 Prediction error of wavelet decomposition sub-series

Reconstructi on sequence	D1	D2	D3	A3	Actual sequence
MRE/%	79.72	40.62	24.18	0.13	0.65
NRMSE	0.0302	0.0216	0.0136	0.0017	0.0087

### 5.3 Comparison experiments

In order to show advantages of the method proposed in this paper, according to the same gas concentration samples, we respectively use the BP neural network prediction model, the SVM prediction model, the sparse ELM regression integration learning algorithm and the W-PSO-SVM method which is proposed in paper[8] to make predictions. Then, we compare the prediction results obtained by using different methods with the prediction result obtained by the method proposed in this paper. More, the BP neural network prediction model, the SVM prediction model and the W-PSO-SVM methods use the first 1200 samples as the training samples, the last 200 samples as the test samples. BP parameter settings are: the hidden layer transfer function using Sigmoid function, the transfer function of the transport layer uses Purelin function, training algorithm using variable learning rate momentum gradient descent algorithm, learning rate is set to 0.1. SVM parameter settings are: SVM kernel function using radial basis function, PSO algorithm is used to optimize the parameters of SVM, optimized parameters include the penalty parameter  $C$ , insensitive loss 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 [1,100], [0.001,0.1] and [0,10], respectively. The SERELM method uses the samples consistent with the method proposed in this paper. The prediction result of the BP and SERELM method comes from the average value by doing fifty experiments. The results of all experiments are

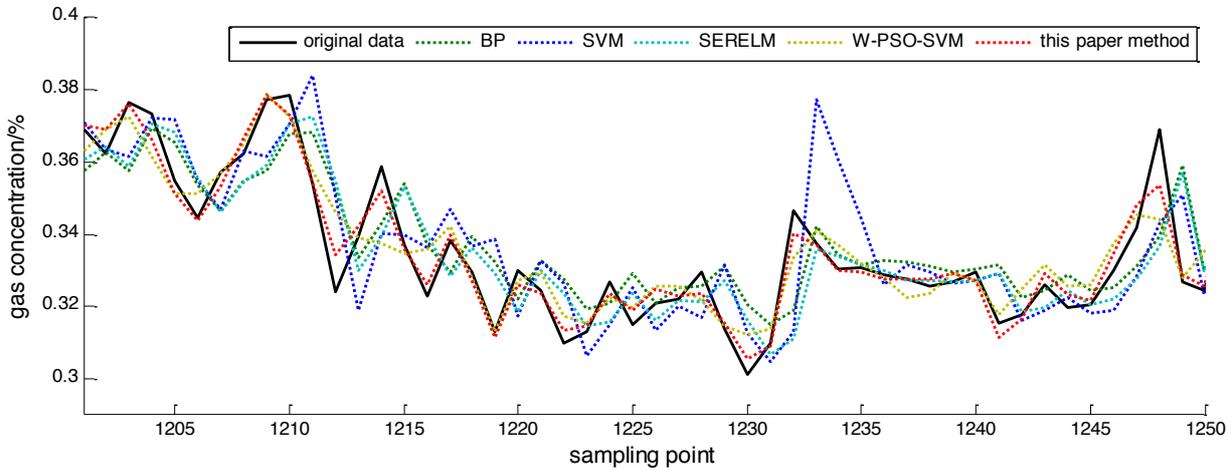
shown in Tab.2. At the same time, the maximum relative error, the number of the relative error under 1% and the percentage of the relative error under 1% are shown as well in Tab.2. In Fig.4, the comparison charts of these prediction methods are shown. For the display of the comparison of these prediction methods, we select the prediction results from the 1201th data point to the 1250th data point.

As the Fig.4 shows, the BP method, the SVM method, the SERELM method and the W-PSO-SVM method all showed hysteresis in different degree and the mutation point error. And these phenomena are caused by great randomness of the gas data, these are consistent with the theoretical analysis. As Tab.2 shows, these methods except the new method proposed by this paper have almost same prediction performance. Specifically, their maximum relative errors are more than 10%. Besides, the percentage of the precise prediction points(the relative error is less than 1%) is under 40%. However, when we use the new method proposed by this paper, the number of the points having

the relative error under 1% can be up to 157, And, these points account for 78.5% of all points. Moreover, their maximum relative errors are under 5%. In general, the prediction values obtained by the new modeling method proposed in this paper fit the actual values better. As a result, they can meet practical requirements well.

**Table.2** Prediction errors of five methods

Prediction method	BP	SVM	SERELM	W-PSO-SVM	Suggested method
MRE/%	3.23	2.66	2.35	1.68	0.65
NRMSE	0.0376	0.034	0.0303	0.0267	0.0087
Maximum relative error /%	15.62	14.34	13.24	10.73	4.16
Number of RE under 1%	39	56	58	72	157
Percentage of RE under 1%	19.50	28.00	29.00	36.00	78.50



**Fig.3.** Prediction comparison chart of five methods

**Tab.3** Prediction error of wavelet decomposition sequence by db4

Wavelet decomposition level	1	2	3	4	5	6	7	8	9	10
MRE/%	1.11	0.72	0.65	0.65	0.65	0.65	0.77	0.76	0.76	1.37
NRMSE	0.0137	0.0093	0.0087	0.0087	0.0087	0.0088	0.102	0.0101	0.0102	0.0154

**5.4 Wavelet decomposition level analysis**

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, the prediction accuracy is higher as well. On the other hand, with the increase of decomposition layers, the number of detail signals will also increase, the errors will be superimposed because the number of detail signals increase, more decomposition layers will bring more prediction errors, as a result, the prediction accuracy will decline.

Table 3 shows the prediction errors using db4 wavelet decomposition when decomposition level is from 1 to 10. It can be seen from table 3,, we can obtain the minimum prediction error when we use the decomposition layers from 3 to 5. Therefore, we can use 3 to 5 as the decomposition layers in the practical application.

**6. Conclusions**

1) A new prediction method of coal face gas concentration by using multi-scale selective ensemble hybrid modeling has been proposed. Firstly, by wavelet transform, the original sequence is decomposed into low-frequency signals reflecting the trend of data and high-frequency signals reflecting the stochastic volatility of data. Then, with the method of L1 regularization, make a sparsely weighted combination of the multiple unstable ELM predictors to form a strong predictor. Finally, the final prediction result can be obtained by superposing the predictive result of each subsequence.

2) Do weight optimization for lots of primary selected ELM predictors by using the method of L1 regularization. Because the sparse weights coming from the method overcome the difficulties of hardly constructing difference predictors in selective ensemble learning, the prediction accuracy and generalization ability of wavelet

decomposition subsequences can be improved. Finally, the prediction result can be improved largely.

3)The maximum prediction error of the method proposed by this paper is 4.16%. The average error of the method proposed by this paper is 0.64%. Both of them are far below the errors coming from the BP method, the SVM

method, the SERELM method and the W-PSO-SVM method. The experimental results show that the method proposed by this paper has the high prediction accuracy and the great generalization ability. Particularly, the method is suitable for the gas concentration time series with strong stochastic volatility.

---

## References

1. Zhou S N, Lin B Q. Coalbed gas occurrence and flow theory. Beijing: China Coal Industry Publishing House, 1999. (in Chinese)
2. Cheng J, Bai J Y, Qian J S, et al. "Short-term forecasting method of coalmine gas concentration based on chaotic time series", *Journal of China University of Mining & Technology*, 37(2), 2008, pp. 231-235. (in Chinese)
3. Zhang J Y, Cheng J, Hou Y H, et al. "Forecasting coalmine gas concentration based on adaptive neuro-fuzzy Inference system", *Journal of China University of Mining & Technology*, 36(4), 2007, pp.494-498. (in Chinese)
4. Cui X Y, Zhao G Q. "Application of chaos forecast of mine gas density monitoring", *Journal of Liaoning Technical University*, 27, 2008, pp.184-186. (in Chinese)
5. Qiao M Y, Ma X P, Lan J Y, et al. "Time series short-Term gas prediction based on weighted LS-SVM", *Journal of Mining and Safety Engineering*, 28(2), 2011, pp. 310-314. (in Chinese)
6. Dong D W, Li S G, Chang X T, et al. "Prediction model of gas concentration around working face using multivariate time series", *Journal of Mining and Safety Engineering*, 29(1), 2012, pp. 135-139. (in Chinese)
7. He G G, Ma S F, Li Y. "A study on forecasting for time series based on wavelet analysis", *Acta Automatica Sinica*, 29(1), 2012, pp. 1012-1014. (in Chinese)
8. Wang Xiaolu, LIU Jian, LU Jianjun. "Gas concentration forecasting approach based on wavelet transform and optimized predictor", *Journal of Basic Science and Engineering*, 19(3), 2011, pp. 499-508. (in Chinese)
9. Wang Q, Chen S C. "Ensemble learning of LEM regressors based on  $l_1$ -regularization", *Journal of computer Research and Developmen*, 49(12), 2012, pp. 2631-2637. (in Chinese)
10. Zhang X D, Bao Z. Non-stationary signal analysis and processing. Beijing: National Defense Publishing House, 1998. (in Chinese)
11. Mallat S G. "A theory for multiresolution signal decomposition: The wavelet representation", *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 11(7), 1989, pp. 674-693.
12. HUANG G B, ZHU Q Y, SIEW C K. "Extreme learning machine: Theory and applications", *Neural Computation*, 70, 2006, pp.489-501.
13. Yang L, Zhang R. "Online sequential ELM algorithm and its improvement", *Journal of Northwest University*, 42(6), 2012, pp. 885-896. (in Chinese)