A Nonlinear Grade Estimation Method Based on Wavelet Neural Network

LI Xiao-li [#1], Xie Yu-ling [*1], Li Li-hong [#2]， Guo Qin-jin[#3]

[1] *Civil & Environmental Engineering School, University of Science and Technology Beijing 100083, China*
[2] *College of Vehicle and Motive Power Engineering, Henna University of Science and Technology 471003, China*
[3] *Shenyang Inst of Automation, Chinese Academy of Sciences, Shenyang 110016, China*

***Abstract*—Grade estimation is one of the most complicated aspects in mining. Its complexity originates from scientific uncertainty. This paper introduces a nonlinear Wavelet Neural Network (WNN) approach to the problem of ore grade estimation. The nonlinear WNN method combing the properties of the wavelet transform and the advantages of Artificial Neural Networks (ANN) provide fast and reliable ore grade estimation, with minimum assumptions and minimum requirements for modeling skills. The WNN grade estimation method has been tested on a number of real deposits. The result shows that the WNN has advantages of rapid training, generality and accuracy grade estimation approach. It can provide with a very fast and robust alternative to the existing time-consuming methodologies for ore grade estimation.**

I. INTRODUCTION

Grade estimation is probably one of the most important stages in reserve calculations and mineral deposit[1].The problem of grade estimation is quite complicated. Over the past 30 years, geostatistics became the most established methodology for grade estimation[2]. However, Geostatistics have been based on certain assumptions about the spatial distribution of ore grades within the orebody. Effects of these assumptions have led to far more complicated methods requiring a large amount of knowledge in order to be effectively applied [3]. And it is proved to be very difficult to learn and apply efficiently and also very time-consuming[4]. The need for a new method of ore grade estimation comes from the difficulties in applying conventional methods such as geostatistics.

An alternative approach that is considered particularly in the last decade is the application of Artificial Neural Networks (ANNs) systems to grade estimation. ANNs systems typically approach grade variance and distribution as complex functions in space, approached by their various components[5][6]. Wavelet neural networks (WNN) combing the properties of the wavelet transform and the advantages of ANNs have attracted great interest and become a popular tool for various fields of mathematics and engineering. It shows surprising effectiveness in solving the grade estimation [7][8].

This paper is organized as follows: In Section 2, the structure of a wavelet neural network and its advantage in developing the intelligent grade estimation system are presented. In Section 3, the proposed optimization method of neural networks and a training algorithm of grade estimation are detailed described. In Section 4, the structure of an intelligent grade estimation module and an example of grade estimation in a real ore deposit are discussed. Finally, in Section 5, the summary and conclusions drawn from this study are presented.

II. WAVELET NEURAL NETWORKS FOR GRADE ESTIMATION

The WNN employed in this study for grade estimation are designed as a three-layer structure with an input layer, wavelet layer (hidden layer) and output layer. The topological structure of the WNN is illustrated in Fig.1, where w_{jk} denotes the weights connecting the input layer and the hidden layer, v_{ij} denote the weights connecting the hidden layer and the output layer. In this WNN models, the hidden neurons have wavelet activation functions of different resolutions, the output neurons have sigmoid activation functions.

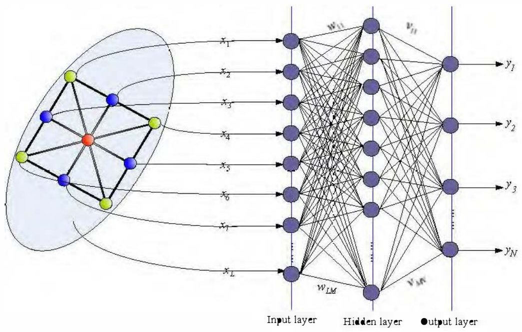

Fig.1. WNN topology structure for grade estimation

The activation functions of the wavelet nodes in the wavelet layer are derived from a mother wavelet $\psi(x)$. It should satisfy the admissibility condition [9]:

$$\int_{-\infty}^{+\infty} \frac{|\hat{\psi}(x)|^2}{x} dx < \infty \tag{1}$$

where $\psi(x) \in L^2(R)$, which represents the collection of all measurable functions in the real space; $\hat{\psi}(x)$ indicates the Fourier transform of $\psi(x)$.

The output of the wavelet neural network Y is represented by the following equation:

$$y_i(t) = \sigma(x_n) = \sigma\left(\sum_{j=0}^{M} v_{ij}\psi_{a,b}\left(\sum_{k=0}^{L} w_{jk}x_k(t)\right)\right)(i = 1,2,\dots,N) \tag{2}$$

$$\sigma(x_n) = 1/(1+e^{-x_n}) \tag{3}$$

[1] Fund item: Key Projects in the National Science & Technology Pillar Program during the Eleventh Five-Year Plan Period（2006BAB01A04）; Supported by Doctoral Fund of Liaoning Province(20081005)

978-1-4244-3867-9/09/$25.00 ©2009 IEEE

Where y_i denotes the ith component of the output vector; v_{ij} the connection weight between the output unit i and the hidden unit j; x_k denotes the kth component of the input vector; w_{jk} the weight between the hidden unit j and input unit k; a_j is dilation coefficient of wavelons in hidden layer; b_j is translation coefficient of wavelons in hidden layer;L, M, N is the sum of input, hidden and output nodes respectively.

III. TRAINING ALGORITHM

The network is trained with BP algorithm in batch way [10], During the training phase, wavelet node parameters, a, b and WNN weights, w_{jk}, v_{ij}, are adjusted to minimize the least-square error, the cost function can be written as:

$$E = \frac{1}{2}\sum_{p=1}^{P}\sum_{i=1}^{N}(d_i^p - y_i^p)^2 \tag{4}$$

where the d_i^p denotes the ith desired target output of pth input pattern,

The selection of the mother wavelet is very important and depends on the particular application. There are a number of well-defined mother wavelets such as Morlet, Harr, Mexican Hat, and Meyer. Groups of them are called families, such as Daubechies, Biorthogonals, Coifets, and Symmlets [11]. For this wavelet neural network, Morlet wavelet has been chosen to serve as an adoption basis function to the network's hidden layer, which has been the preferred choice in most work dealing with WNN, due to its simple explicit expression.

$$\psi_{a,b}(t) = \cos(1.75t_z)e^{-\frac{t_z^2}{2}} \tag{5}$$

where $t_z = (\frac{t-b}{a})$

$$\psi_{a,b}'(t) = -\left(1.75\sin(1.75t_z)\exp(-\frac{t_z^2}{2}) + \cos(1.75t_z)\exp(-\frac{t_z^2}{2})t_z\right) \tag{6}$$

Denote that $net_j = \sum_{k=0}^{L} w_{jk}x_k$

Thus

$$\psi_{a,b}(net_j) = \psi\left(\frac{net_j - b_j}{a_j}\right) \tag{7}$$

$$y_i(t) = \sigma\left(\sum_{j=0}^{M} v_{ij}\psi_{a,b}(net_j)\right)(i = 1,2,...N) \tag{8}$$

$$\sigma'(u) = \frac{\partial\sigma(u)}{\partial u} = \sigma(u)[1-\sigma(u)] \tag{9}$$

Such that

$$\delta_{v_{ij}} = \frac{\partial E}{\partial v_{ij}} = -\sum_{p=1}^{P}(d_i^p - y_i^p)y_i^p(1-y_i^p)\psi_{a,b}(net_j^p) \tag{10}$$

$$\delta_{w_{jk}} = \frac{\partial E}{\partial w_{jk}} = -\sum_{p=1}^{P}\sum_{i=1}^{N}(d_i^p - y_i^p)y_i^p(1-y_i^p)v_{ij}\psi_{a,b}'(net_j^p)x_k^p/a_j \tag{11}$$

$$\delta_{b_j} = \frac{\partial E}{\partial b_j} = \sum_{p=1}^{P}\sum_{i=1}^{N}(d_i^p - y_i^p)y_i^p(1-y_i^p)v_{ij}\psi_{a,b}'(net_j^p)/a_j \tag{12}$$

$$\delta_{a_j} = \frac{\partial E}{\partial a_j} = \sum_{p=1}^{P}\sum_{i=1}^{N}(d_i^p - y_i^p)y_i^p(1-y_i^p)v_{ij}\psi_{a,b}'(net_j^p)(\frac{net_j^p - b_j}{a_j^2}) \tag{13}$$

The learning rate and momentum are set as η and μ in the experiments respectively.

$$\Delta w_{jk}(t+1) = -\eta\frac{\partial E}{\partial w_{jk}} + \mu\Delta w_{jk}(t) \tag{14}$$

$$\Delta v_{ij}(t+1) = -\eta\frac{\partial E}{\partial w_{ij}} + \mu\Delta v_{ij}(t) \tag{15}$$

$$\Delta a_j(t+1) = -\eta\frac{\partial E}{\partial a_j} + \mu\Delta a_j(t) \tag{16}$$

$$\Delta b_j(t+1) = -\eta\frac{\partial E}{\partial b_j} + \mu\Delta b_j(t) \tag{17}$$

Then the parameters are updated as follows:

$$w_{jk}(t+1) = w_{jk}(t) + \Delta w_{jk}(t+1) \tag{18}$$

$$v_{ij}(t+1) = w_{ij}(t) + \Delta v_{ij}(t+1) \tag{19}$$

$$a_j(t+1) = a_j(t) + \Delta a_j(t+1) \tag{20}$$

$$b_j(t+1) = b_j(t) + \Delta b_j(t+1) \tag{21}$$

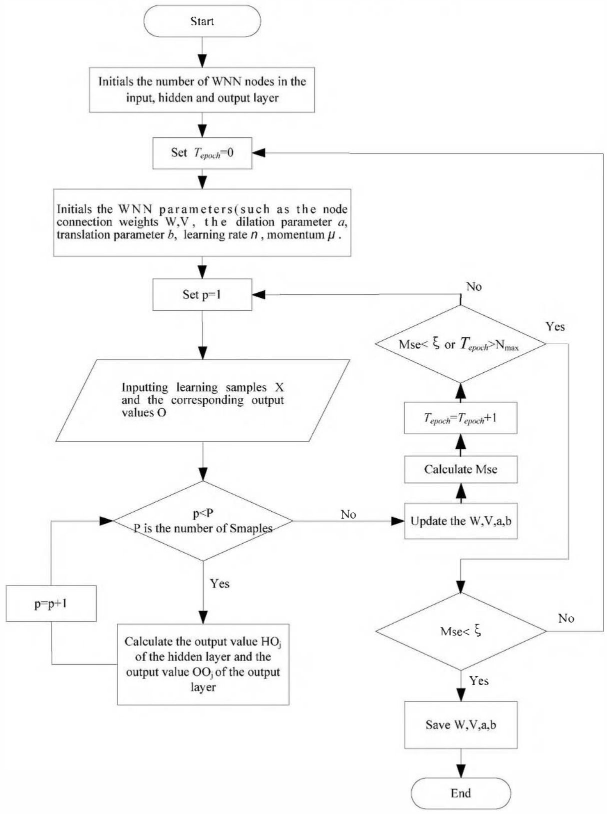

Fig.2. Training algorithm diagram structure

According to the above-mentioned reasoning based on the gradient descend algorithm, the training algorithm can be summarized in Fig.2. As a result, the network model in this paper is constructed by using error back propagation as training algorithm and Morlet wavelet function as node activation function.

IV. CASE STUTY FOR GRADE ESTIMATION

Here the work uses a nonlinear grade estimation model that we have developed to test WNN method for grade estimating. This model mainly contains the sample data acquisition, sampling neighborhood scheme extraction and data normalization, network training, and grade estimation etc.

A. Sample Data Acquisition

In order to assess the performance of WNNs as prediction techniques for grade estimation, a sufficient amount of data is needed to train, validate, and test the different topologies. This section selected actual grade data of a porphyry copper deposit, which was taken from David's 'Geostatistical Ore Reserve Estimation' book[13]. According to David, the deposit is homogeneous. The values used are supposed from one level of the simulated deposit, divided into 100'×100'×50' blocks. The grades given are the actual grades of the blocks. But now the grades are assigned 20×20 blocks when the deposit will be used to train and validation WNNs. Fig.3 is a contour map of the Copper example deposit and Fig.3 is the data posting. Grade information from Fig.4 is given to WNN for training and testing.

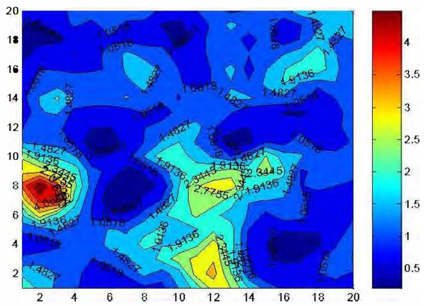

Fig.3. Contour map of the Copper example deposit

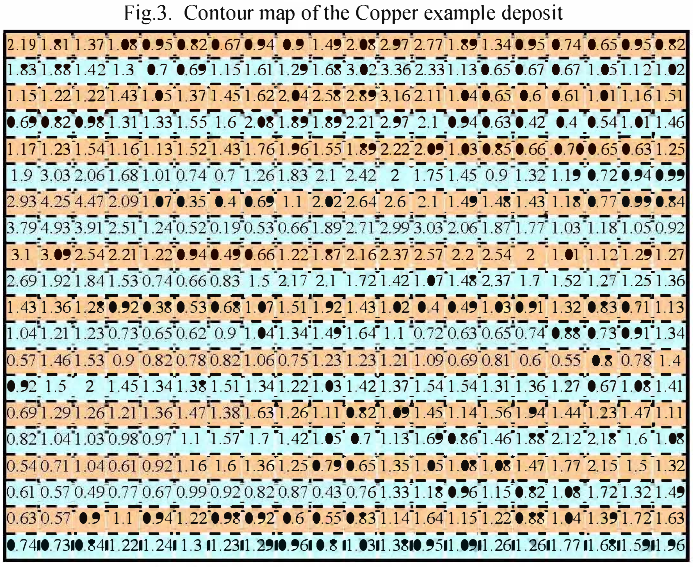

Fig.4. Data allocation for Copper A example

B. Sampling Neighborhood Scheme And Data Normalization

After the selection of source data deposits, the sampling scheme that used to train and validate the WNN topologies will be discussed. Quite reasonably, some researchers tried to take advantage of the information hidden in the relationship between neighboring samples. This approach is followed in general terms by the most advanced existing methods for grade estimation like inverse distance weighting and Kriging. Accordingly, the sampling scheme in this work follows the similar rules that choose the samples closest to the estimation point as neighbors and treat the problem of ore grade estimation as a mapping between the surrounding grades and the grade at the estimation point. Grade values from surrounding sample points are given as input to the network while the grade of the central point is given an output. The mapping scheme is one estimation point with eight surrounding neighbor's points, according to this regulation, the sample dataset are acquired from copper deposit (see in Fig.3 and Fig.4). The sampling scheme topology is shown in Fig.5 and the subsets of sample data shown in Table I. It is supposed that the deposit is homogeneous, and geometric distant is not considered at present.

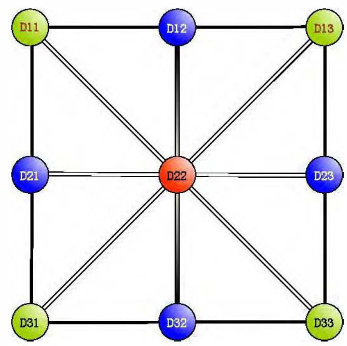

Fig.5. I/O configurations (red indicate estimation)

Table I The subsets of training sample

Inputs No.	D11	D12	D13	D21	D23	D31	D32	D33	Targets (D22)
1	1.22	1.22	1.43	0.82	1.31	1.23	1.54	1.16	0.98
2	1.05	1.37	1.45	1.33	1.60	1.13	1.52	1.43	1.55
3	1.62	2.04	2.58	2.08	1.89	1.76	1.96	1.55	1.89
4	2.89	3.16	2.11	2.21	2.10	1.89	2.22	2.09	2.97
5	1.04	0.65	0.60	0.94	0.42	1.03	0.85	0.66	0.63
6	0.61	1.01	1.16	0.40	1.01	0.70	0.65	0.63	0.54
7	3.03	2.60	1.68	4.25	2.09	4.93	3.91	2.51	4.47
8	1.01	0.74	0.70	1.07	0.40	1.24	0.52	0.19	0.35
9	1.26	1.83	2.10	0.69	2.02	0.53	0.66	1.89	1.10
10	2.42	2.00	1.75	2.64	2.10	2.71	2.99	3.03	2.60
11	1.45	0.90	1.32	1.49	1.43	2.06	1.87	1.77	1.48
12	1.19	0.72	0.94	1.18	0.99	1.03	1.18	1.05	0.77
13	3.09	2.54	2.21	1.92	1.53	1.36	1.28	0.92	1.84
14	1.22	0.94	0.49	0.74	0.83	0.38	0.53	0.68	0.66
15	0.66	1.22	1.87	1.50	2.10	1.07	1.51	1.92	2.17
16	2.16	2.37	2.57	1.72	1.07	1.43	1.02	0.40	1.42

Before training, it is often useful to scale the inputs and targets so that they always fall within a specified range. Neural network training can be made more efficient if you perform certain preprocessing steps on the network inputs and targets. The algorithm to normalize a given data set X so that the inputs and targets will fall in the range [-1,1] as follows.

$$Y = \frac{X - MinValue}{MaxValue - MinValue} \tag{22}$$

Where X denotes the given data; *MinVale* denotes the min data of the input data; *MaxVale* denotes the max data of the input data; Y denotes the normalization data.

According to this method, the sets of sample in Table I can be normalized between 0 and 1 for training (see in Table II). These normalized vectors are used as input and output of WNN.

Table II The normalized sets of training sample

Number	D11	D12	D13	D21	D23	D31	D32	D33	Targets (D22)
1	0.24	0.24	0.29	0.16	0.26	0.25	0.31	0.23	0.20
2	0.21	0.27	0.29	0.27	0.32	0.23	0.30	0.29	0.31
3	0.32	0.41	0.52	0.42	0.38	0.35	0.39	0.31	0.38
4	0.58	0.63	0.42	0.44	0.42	0.38	0.44	0.42	0.59
5	0.21	0.13	0.12	0.19	0.08	0.21	0.17	0.13	0.13
6	0.12	0.20	0.23	0.08	0.20	0.14	0.13	0.13	0.11
7	0.61	0.52	0.34	0.85	0.42	0.99	0.78	0.50	0.89
8	0.20	0.15	0.14	0.21	0.08	0.25	0.10	0.04	0.[illegible]7
9	0.25	0.37	0.42	0.14	0.40	0.11	0.13	0.38	0.22
10	0.48	0.40	0.35	0.53	0.42	0.54	0.60	0.61	0.52
11	0.29	0.18	0.26	0.30	0.29	0.41	0.37	0.35	0.30
12	0.24	0.14	0.19	0.24	0.20	0.21	0.24	0.21	0.15
13	0.62	0.51	0.44	0.38	0.31	0.27	0.26	0.18	0.37
14	0.24	0.19	0.10	0.15	0.17	0.08	0.11	0.14	0.13
15	0.13	0.24	0.37	0.30	0.42	0.21	0.30	0.38	0.43
16	0.43	0.47	0.51	0.34	0.21	0.29	0.20	0.08	0.28

C. Network Training And Determinations

1) Network initialization

Initializing the wavelet network parameters is an important issue. A random initialization of all the parameters to small randomly values (as usually done with neural networks) is not desirable since this may make some wavelets too local (small dilations) and make the components of the gradient of the cost function very small in areas of interest.

In [12], a method for dilation and translation initialization is proposed. This work uses this method to initialize the dilation a_j and translation b_j. In general, one wants to take advantage of the input space domains where the wavelets are not zero. According to the wavelet theory, given the t^* as the center of the time domain, $\Delta\psi$ as the radius of the mother wavelet, the time domain so can be written as:

$$\left[b + at^* - a\Delta\psi, b + at^* + a\Delta\psi\right] \tag{23}$$

In order to guarantee that the wavelets extend initially over the whole input domain, the initialization of the dilation and translation parameters must agree to the following equation:

$$\begin{cases} b_j + a_j t^* - a_j \Delta\psi = \sum_{i=1}^{I} w_{ji} x_{i\min} \\ b_j + a_j t^* - a_j \Delta\psi = \sum_{i=1}^{I} w_{ji} x_{i\max} \end{cases} \tag{24}$$

Thus

$$\begin{cases} a_j = \dfrac{\sum_{i=1}^{I} w_{ji} x_{i\max} - \sum_{i=1}^{I} w_{ji} x_{i\min}}{2\Delta\psi} \\ b_j = \dfrac{\sum_{i=1}^{I} w_{ji} x_{i\max}(\Delta\psi - t^*) + \sum_{i=1}^{I} w_{ji} x_{i\min}(\Delta\psi + t^*)}{2\Delta\psi} \end{cases} \tag{25}$$

The choice of the learning rate factor η has an important effect on stability and convergence rate. If it is too small, too many steps are needed to reach an acceptable solution. On the contrary, a large learning rate may possibly lead to skip the optimal solution and cause oscillation. The momentum factor μ scales the influence of the previous step on the current. It renders the learning procedure stable and accelerates convergence in shallow regions of error function. In the experiment, η and μ are initialized to 0.08 and 0.5, respectively.

The choice of the weights is less critical: these parameters are initialized to small random values.

2) Network training and determination

In the experiment, 200 groups of sample data are acquired on Fig. 4; 50 groups of the data are used as the training sets, the remaining 150 groups are used as the validate sets. Expecting output error threshold is 0.001. Sample vectors are used as input and output of WNN. After all possible normal operating modes of the grade are learned, the system enters the grade estimation stage.

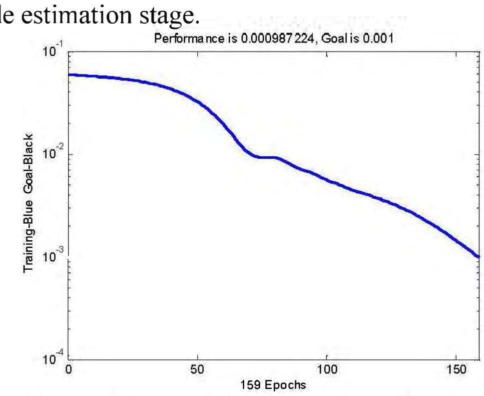

Fig.6. WNN training curve with 15 hidden nodes

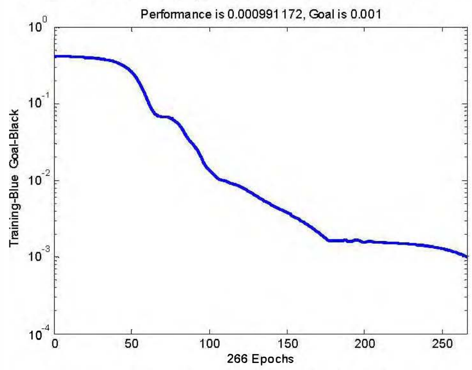

Fig.7. WNN training curve with 8 hidden nodes

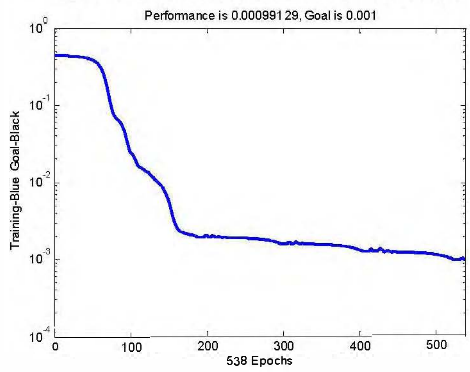

Fig.8. WNN training curve with 20 hidden nodes

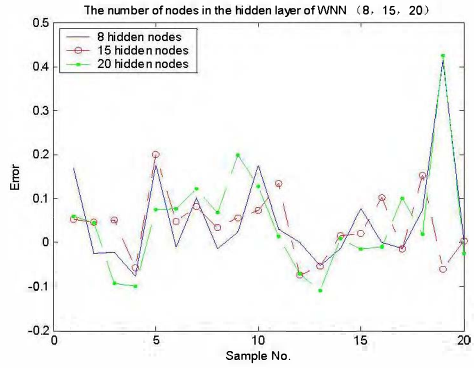

Fig.9. Error in WNN with 8,15, 20 hidden nodes

WNN start as a network of nodes arranged in three layers--the input, hidden, and output layers. The input and output layers serve as nodes to buffer input and output for the model, respectively, and the hidden layer serves to provide a means for input relations to be represented in the output. Here, the network architecture used for grade estimation consists of 8 inputs corresponding to the 8 neighboring samples that grade have been given (listed in table I), and 1 outputs corresponding to 1 unknown grade.

This determines the number of nodes there will be in the hidden layer. Unfortunately, there is no concrete rule of thumb, so you must find the optimal number on your own. From there, you can adjust up or down to try to create a better-trained network. If the number of hidden layer nodes is too small, WNN may not reflect the complex function relationship between input data and output value. On the country, a large number may create such a complex network that might lead to a very large output error caused by over-fitting of the training sample. Thus it can be seen; selecting proper hidden nodes play a very important role in building WNN model. For solving this problem, a new method was developed by using advanced computer technology. The determination of hidden nodes of WNN model becomes very easy by this method.

The network was trained in different hidden nodes (see Fig.6. Fig.7.and Fig.8.), and percentage of error in WNN model were calculated. Then the number of hidden nodes as horizontal scale, and percentage of error as vertical scale, we can draw the 2-dimension plot (see Fig.9.). The optimal number of hidden nodes and the training times can be determined easily using the 2-dimension plot.

V. CONCLUSION

A wavelet neural network based model for grade estimation was presented in this work. Due to its multiscale, multiresolution, and localization, the WNN can accurately capture the local nonlinearity of the dynamic systems. The performance of the WNN approach were tested on the rotating machinery and compared with BP methods. The test results show that the proposed WNNs approach improves the diagnosis accuracy and learning speed of the conventional approaches.

The ongoing activities are oriented to: i) the optimization of the developed software architecture, ii) the performance comparison with other neural network architecture, iii) and the increment of the measurement quality on parameter number.

REFERENCES

[1] Bulent Tutmez. An uncertainty oriented fuzzy methodology for grade estimation. Computers &Geosciences, Vol.33, Issue 2, February 2007, Pages 280-288.

[2] Kapageridis I., Denby B., Neural Network Modelling of Ore Grade Spatial Variability. In: Proceedings of the International Conference for Artificial Neural Networks (ICANN 98), Vol. 1, pp 209 – 214, Springer-Verlag, Skovde, 1998.

[3] Kapageridis I., Denby B., and Hunter G., Integration of a Neural Ore Grade Estimation Tool In a 3D Resource Modeling Package. In: Proceedings of the International Joint Conference on Neural Networks (IJCNN '99), International Neural Network Society and The Neural Networks Council of IEEE, Washington D.C., 1999.

[4] Kapageridis I., Denby B., Neural Network Modelling of Ore Grade Spatial Variability. In: Proceedings of the International Conference for Artificial Neural Networks (ICANN 98), Vol. 1, pp 209 – 214, Springer-Verlag, Skovde, 1998.

[5] Wu X, Zhou Y. Reserve estimation using neural network techniques. Computers &Geosciences, Vol. 19, No. 4, pp. 567-575, Pergamon Press, 1993

[6] Clarici E, Owen D, Durucan S, Ravencroft P. Recoverable reserve estimation using a neural network. In: Elbrond J, Tang X (ed) 24th International Symposium on the Application of Computers and Operations Research in the Minerals Industries (APCOM). Montreal, Quebec, 1993

[7] Q Zhang and A Benveniste, Wavelet Networks, IEEE Trans. On Neural Networks, 1992, 3(11):889-898.

[8] D. S. Chen and Ramesh C. Jain, A Robust Back Propagation Learning Algorithm for Function Approximation, IEEE Trans. Neural Networks, 1994, 5(3):467-479.

[9] I.Daubechies, The wavelet transform, time-frequency localization, and signal analysis, IEEE Trans. Inform Theory, 1990,36(10): 961~1005.

[10] D.E.Rumelhart, G.E.Hinton et al, Learning Internal Representations by Error Propagation, Nature, 1986, 323: 533~536.

[11] Misiti M, Misiti Y et al. Wavelet Toolbox User's Guide, The Math Works Inc, 1996.

[12] Zhao xue-zhi, Zou Chun-hua et al. A research on the initialization of Parameters of Wavelet Neural Networks. Journal of South China University of Technology, 2003, 31(2):77~84.

[13] David, M., Geostatistical Ore Reserve Estimation. Elsevier, Amsterdam, 1977.