2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Performance of Combined Artificial Neural Networks for Forecasting Landslide Displacement

Cheng Lian, Zhigang Zeng, Wei Yao and Huiming Tang

***Abstract*— An efficient and accurate method for landslide displacement prediction is very important to reduce the casualties and property losses caused by this type of natural hazard. In recent years, many kinds of artificial neural networks (ANNs) have been widely applied to landslide displacement prediction. But we can't know which type of ANN is the best until we have calculated the prediction error. An improper choice of ANN may result in bad prediction results. In this paper, we use a neural networks combination prediction method based on the discounted MSFE (mean squared forecast error) to reduce the risk of selecting the types of ANNs. Four popular ANNs, radial basis function neural network (RBFNN), support vector regression (SVR), least squares support vector machine (LSSVM) and extreme learning machine (ELM), are selected as candidate neural networks. The performance of our model is verified through two case studies in Baishuihe landslide and Bazimen landslide. Experimental results reveal that the combining neural networks can improve the generalization abilities of ANNs.**

I. Introduction

In the Three Gorges Reservoir of China, frequent landslides often result in significant damage to people and property. The possible time when landslides are likely to occur should be identified in advance and thus people can decrease landslide damage through proper preparation. Landslide is a complex nonlinear dynamical system with uncertainty, a sufficiently precise mechanism model can hardly be obtained. In recent years, ANNs have been widely applied to landslide displacement prediction, as ANNs have ability to learn nonlinear functions from the data [2] [4] [7] [11] [12] [13]. With the fast development of ANNs, there are kinds of candidate ANNs for us to choose. Different ANNs may suit for different data. For a specific case, especially for landslide displacement prediction, we don't know which ANN is the best until we have calculated the prediction error. An improper choice of ANN may result in bad prediction results. In this case, neural network combination prediction may be a useful way to reduce the risk of choosing one single ANN. Combining ANNs may help to integrate the knowledge acquired by each component ANN and thus improve model accuracy [1] [8] [16]. In this paper, four popular ANNs, RBFNN [3], SVR [6], LSSVM [14] and ELM [9] [10], are selected as candidate ANN forecasters.

A two step procedure is proposed to build the landslide displacement prediction model. In the first step, each alternative ANN is supplied by an independent professional forecaster. In the second step, the discounted MSFE is used to combine the ANNs in the first step. Two typical landslides in Three Gorges Reservoir, Baishuihe landslide and Bazimen landslide, are presented to illustrate the capability of our method.

C. Lian and Z.G. Zeng are with the School of Automation, Huazhong University of Science and Technology, and Key Laboratory of Image Processing and Intelligent Control of Education Ministry of China, Wuhan, Hubei, 430074, China (e-mail: zgzeng521@126.com).

W. Yao is with School of Computer Science, South-Central University for Nationalities, Wuhan 430074, China.

H.M. Tang is with Faculty of Engineering, China University of Geosciences, Wuhan 430074, China

The work is supported by the Natural Science Foundation of China under Grant 60974021 and 61203286, the 973 Program of China under Grant 2011CB710606, the Specialized Research Fund for the Doctoral Program of Higher Education of China under Grant 20100142110021.

II. Individual Forecasts

A. NAR Structure

Considering a time series of landslide displacement $\{y_t\}_{t=1}^N$, a nonlinear autoregressive (NAR) model can be expressed as follows:

$$\widehat{y}_t = \widehat{f}(y_{t-1}, \ldots, y_{t-\ell}) \tag{1}$$

where the function $\widehat{f}$ is estimated using ANN, ℓ is the time lag. In this article, we use four popular ANNs, RBFNN, SVR, LSSVM and ELM, as a tool for estimating $\widehat{f}$ in NAR model, respectively.

B. Artificial Neural Network Model

1) RBF Network: The RBF neural network is a feedforward network, which is generally composed of three layers: an input layer, a hidden layer and a linear output layer [3] [17]. External data are presented to the RBF network via the input layer. The hidden layer applies a nonlinear transformation from the input space to the hidden space by using non-linear RBF activation functions. The final outputs of RBF network are produced though the output layer. The output of RBF network can be mathematically modeled as:

$$f(\mathbf{x}) = \sum_{j=1}^{L} w_j h_j(\mathbf{x}) \tag{2}$$

where $\mathbf{x}$ is the input variable, L is the number of neurons in the hidden layer, w_j is the output weights and $\{h_j\}_{j=1}^L$ are the RBF activation functions. In this paper, Gaussian function is used as RBF activation function, which is defined as:

$$h_j(\mathbf{x}) = \exp(-\|\mathbf{x} - \mathbf{c}_j\|/2\sigma^2) \tag{3}$$

where $\mathbf{c}_j$ is the center, σ is the width, $\|\cdot\|$ is the Euclidean norm.

Given a set of training data with N number of samples $(\mathbf{x}_i, \mathbf{t}_i), i = 1, \ldots, N$, a RBF network consisting of m RBFs

with centers $\{\mathbf{c}_j\}_{j=1}^m$ and widths $\{\sigma_j\}_{j=1}^m$. Take one output node as an example, the goal of RBF network is to minimize the sum of squared errors:

$$E = \sum_{i=1}^{N}(f(\mathbf{x}_i) - t_i)^2 \tag{4}$$

This leads to a set of N linear equations in L unknown weights, and the output weights $\mathbf{w}$ can be obtained using least squares method:

$$\mathbf{w} = (\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{T} \tag{5}$$

where $\mathbf{T} = [t_1 \ t_2 \ldots t_N]^{\mathrm{T}}$, $\mathbf{H}$ is the hidden layer output matrix, its elements are $H_{ij} = h_j(\mathbf{x}_i)$.

2) SVR: In the past two decades, Support Vector Machine (SVM) has been widely used in classification applications [5]. A modified version of SVM, called SVR, is developed for regression applications [6]. SVR is capable of solving non-linear time-series problems. Given a set of training data with N number of samples $(\mathbf{x}_i, \mathbf{t}_i), i = 1, \ldots, N$. The regression function can be expressed as:

$$f(\mathbf{x}) = \mathbf{w} \cdot \phi(\mathbf{x}) + b \tag{6}$$

where $\phi : \mathbf{x}_i \to \phi(\mathbf{x}_i)$ is the nonlinear mapping, b is the bias. The regression parameters $\mathbf{w}$ and b are estimated by minimizing the sum of the norm of the weights, $\|\mathbf{w}\|^2$, and the empirical risk. In this paper, Vapnik's ϵ-insensitive loss function is used, which is defined as:

$$L_\epsilon(f(\mathbf{x}_i), t_i) = \begin{cases} 0, & \text{if } |t_i - f(\mathbf{x}_i)| \le \epsilon \\ |t_i - f(\mathbf{x}_i)| - \epsilon, & \text{otherwise} \end{cases} \tag{7}$$

Then, the nonlinear regression problem can be expressed as the following optimization problem.

$$\begin{aligned} \text{Minimize} : & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \\ \text{Subject to} : & f(\mathbf{x}_i) - t_i \le \epsilon + \xi_i \\ & t_i - f(\mathbf{x}_i) \le \epsilon + \xi_i^* \\ & \xi_i, \xi_i^* \ge 0, \quad i = 1, \ldots, N \end{aligned} \tag{8}$$

where ξ_i, ξ_i^* are slack variables, C is a user-specified parameter that determines the degree of penalized loss. Based on the Karush-Kuhn-Tucker(KKT) theorem, this optimization problem can be transformed to the dual problem. The final solution is given by:

$$f(\mathbf{x}) = \sum_{s=1}^{N_s}(a_i^* - a_i)K(\mathbf{x}_s, \mathbf{x}) + b, \qquad 0 \le a_i^*, a_i \le C \tag{9}$$

where a_i^*, a_i are the Lagrange multiplier, N_s is the number of support vectors $\mathbf{x}_s$'s, $K(\mathbf{x}_s, \mathbf{x})$ is the kernel function, which needs to satisfy Mercer's condition. In particular, Gaussian kernel function is selected as the kernel function:

$$K(\mathbf{x}_s, \mathbf{x}) = \exp(-\|\mathbf{x} - \mathbf{x}_s\|^2/\gamma) \tag{10}$$

where γ is a user-specified kernel parameter.

3) LSSVM: LSSVM is a modification of the standard SVM, which has been proven to have excellent generalization performance and low computational cost [14]. A given training data set $(\mathbf{x}_i, \mathbf{t}_i), i = 1, \ldots, N$ is assumed, the regression function can also be expressed as Eq.(6). In LSSVM, the regression problem is formulated as:

$$\begin{aligned} \text{Minimize} : & \frac{1}{2}\mathbf{w} \cdot \mathbf{w} + \frac{1}{2}C\sum_{i=1}^{N} e_i^2 \\ \text{Subject to} : & t_i = \mathbf{w} \cdot \phi(\mathbf{x}_i) + b + e_i, \\ & i = 1, \ldots, N \end{aligned} \tag{11}$$

where e_i is the error between the actual output and the predictive output of the ith sample, C is the regularization parameter that balances model complexity and approximation accuracy. The optimization problem can be transformed to the dual problem based on KKT theorem. The final LS-SVM model is given by:

$$f(\mathbf{x}) = \sum_{i=1}^{N} a_i K(\mathbf{x}_i, \mathbf{x}) + b \tag{12}$$

where $a_i, i = 1, \ldots, N$ are the Lagrange multipliers, $K(\mathbf{x}_i, \mathbf{x})$ is the kernel function, which is selected as Gaussian kernel function in Eq.(10).

4) ELM with kernel function: Recently, a novel learning algorithm for single-hidden-layer feedforward neural networks (SLFNs) called ELM has been proposed [9] [18]. ELM not only learns much faster with higher generalization performance than the traditional gradient-based learning algorithms but also avoids many difficulties faced by gradient-based learning methods such as stoping criteria, learning rate, learning epochs and local minimum. The main characteristic of ELM is that the input weights and biases can be randomly chosen and need not be tuned.

For N distinct samples $(\mathbf{x}_i, \mathbf{t}_i), i = 1, \ldots, N$, where $\mathbf{x}_i \in \mathbf{R}^p, \mathbf{t}_i \in \mathbf{R}^q$. Take one output node as an example, ELM with L hidden nodes can be mathematically modeled as:

$$f_L(\mathbf{x}) = \sum_{i=1}^{L} \beta_i h_i(\mathbf{x}) = \mathbf{h}(\mathbf{x})\boldsymbol{\beta} \tag{13}$$

where $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), \ldots, h_L(\mathbf{x})]$ is the hidden layer output vector with respect to the input $\mathbf{x}$, $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_L]^{\mathrm{T}}$ is the vector of the output weights. The goal of ELM is to minimize the training error and the norm of the output weights:
Minimize: $\|\mathbf{H}\boldsymbol{\beta} - \mathbf{T}\|^2$ and $\|\boldsymbol{\beta}\|$
where $\mathbf{T}$ is the target vector and $\mathbf{H}$ is the hidden layer output matrix:

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}(\mathbf{x}_1) \\ \vdots \\ \mathbf{h}(\mathbf{x}_N) \end{bmatrix} = \begin{bmatrix} h_1(\mathbf{x}_1) & \cdots & h_L(\mathbf{x}_1) \\ \vdots & \vdots & \vdots \\ h_1(\mathbf{x}_N) & \cdots & h_L(\mathbf{x}_N) \end{bmatrix} \tag{14}$$

Based on the minimal norm least square method, the output weights can be obtained by:

$$\boldsymbol{\beta} = \mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{H}^{\mathrm{T}})^{-1}\mathbf{T} \tag{15}$$

According to the ridge regression theory, a positive value can be added to the diagonal of $\mathbf{HH}^{\mathrm{T}}$ in order to have better generalization performance:

$$\boldsymbol{\beta} = \mathbf{H}^{\mathrm{T}}(\frac{\mathbf{I}}{C} + \mathbf{HH}^{\mathrm{T}})^{-1}\mathbf{T} \tag{16}$$

Then, the output function of ELM can be rewritten as:

$$\mathbf{f}(\mathbf{x}) = \mathbf{h}(\mathbf{x})\boldsymbol{\beta} = \mathbf{h}(\mathbf{x})\mathbf{H}^{\mathrm{T}}(\frac{\mathbf{I}}{C} + \mathbf{HH}^{\mathrm{T}})^{-1}\mathbf{T} \tag{17}$$

Huang et al. [10] defined a kernel matrix for ELM based on Mercer's conditions as follows:

$$\boldsymbol{\Omega}_{\mathrm{ELM}} = \mathbf{HH}^{\mathrm{T}} : \Omega_{\mathrm{ELM}} = h(\mathbf{x}_i) \cdot h(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j) \tag{18}$$

Finally, the output function of ELM with kernel function can be written as:

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} K(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ K(\mathbf{x}, \mathbf{x}_N) \end{bmatrix}^{\mathrm{T}} (\frac{\mathbf{I}}{C} + \boldsymbol{\Omega}_{\mathrm{ELM}})^{-1}\mathbf{T} \tag{19}$$

Here, $K(\mathbf{x}, \mathbf{x}_i)$ is select as Gaussian kernel function in (10).

III. Combination Method

Ideally, ANN can approximate any continuous non-linear function with arbitrary precision. However, practically, the generalization ability of ANN is always unstable. ANN with high fitting precision may also obtain a bad prediction error. With the development of several decades, there are kinds of candidate ANNs that we can choose. In a real situation, we don't know which ANN is the best until we have observed the prediction errors. Different ANNs may fit different data sets, even the same ANN with different parameters may lead to very different prediction errors. In general, the best available ANN predictor is always not clear. Combining the multiple neural networks may help to integrate the knowledge acquired by each individual ANN and appears to be a very promising approach in improving the generalization ability of ANN [1] [8] [16].

Given a set of K individual predictors, the prediction value of predictor k for the landslide displacement in time $t+1$ is labeled as $\widehat{y}_{k,t+1}$. In this paper, some most popular ANNs are selected as individual predictors. The linear combination procedure can be described as follows:

$$\widehat{y}_{c,t+1} = \sum_{k=1}^{K} w_{k,t}\widehat{y}_{k,t+1} \tag{20}$$

where c refers to the method of selecting the weights. The goal is to find the optimal weights $w_{k,t}$ such that $\widehat{y}_{c,t+1}$ is a good estimator for the true value y_{t+1}. The total data set is divided into three parts: the training set, the validation set and the prediction set. The training set is used to train ANNs, the validation set is used to determine the parameters of ANNs, and also used to calculate the combination weights, the prediction set is used to evaluate the performance of the prediction model. The discounted MSFE is used as the combination method, where the weights depend inversely on the historical performance of each individual ANN [15]. The combining weights formed at time t are calculated as follows:

$$w_{k,t} = \frac{(\psi_{k,t})^{-1}}{\sum_{j=1}^{K}(\psi_{j,t})^{-1}} \tag{21}$$

$$\psi_{k,t} = \sum_{m}^{m+q0} \theta^{t-1-m}(y_{m+1} - \widehat{y}_{k,m+1})^2 \tag{22}$$

where θ is the discount factor which is less than or equal to 1. θ less than 1 means that more recent observations have more influence in calculating the weights. In this paper, θ is selected as 0.95. $q0$ is the length of holdout out-of-sample data. Particularly, if we want to predict the first data in the prediction set, $q0$ equals the length of the validation set, the starting value for m is selected as the first time step in the validation set. Note that, the weights are constantly updated when new data are continuously added.

IV. Experiments

A. Case study 1: Baishuihe landslide

1) Date Collection: Three Gorges reservoir region where has been badly affected by landslides in recent years, is selected as suitable for this study. Baishuihe landslide is located on the south bank of Yantze River and its 56km away from the Three Gorges Dam of China. There are 11 GPS monitoring points on the landslide surface. The monitoring data at ZG118 point is selected to establish the prediction model. Fig. 1 shows the monitoring data of landslide accumulative displacement at ZG118 monitoring point which is monitored from August 2003 to December 2011.

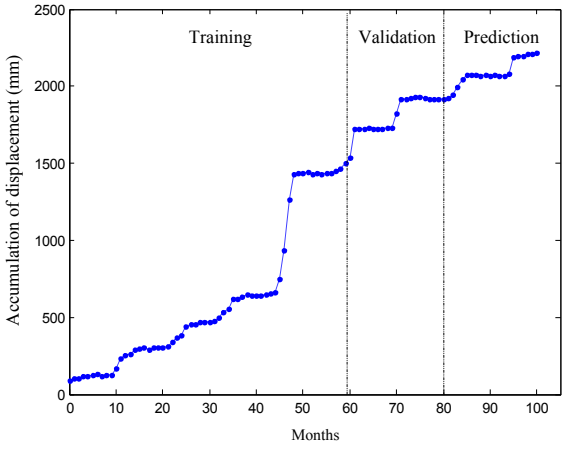

Fig. 1. Monitoring curves of landslide accumulative displacement at ZG118 point

2) Selection of Parameters: There are a total of 101 observations. The 59 observations of data between August 2003 and June 2008 are selected as training set, 21 observations of data from July 2008 to March 2010 are selected as validation set, and the rest of 21 observations of data are the prediction set. Firstly, regressors should to be selected, i.e.,

TABLE I

PARAMETERS USED IN RBFNN, SVR, LSSVM AND ELM.

	RBFNN	SVR	LSSVM	ELM
C	/	2^{19}	2^{13}	2^{11}
γ	/	2^{-13}	2^{12}	2^{9}
σ	/	/	/	/
lag	4	2	9	9

which lags of inputs are going to be included in (1). This selection is done by the training and validation set methods. The optimal number of time lag ℓ is selected as the one which results in the lowest validation error. Secondly, the parameters for ANNs need to be selected. For ANNs, the parameters are also determined by the training and validation set methods. For RBFNN, the Gaussian function is used as RBF activation function, K-means clustering algorithm is used to obtain representative centers for RBFs. The Gaussian kernel function is used in SVR, LSSVM and ELM. The regularization parameter C and kernel parameter γ need to be selected. We test 50 different values of C and 50 different values of γ, resulting in a total of 2500 pairs of (C, γ). The different values of C and γ are $\{2^{-24}, 2^{-23}, \ldots, 2^{24}, 2^{25}\}$. The user-specified parameters finally chosen are given in Table 1. Note that, the number of hidden nodes in RBFNN is 6.

3) Experimental Results: Note that only one-step-ahead prediction is performed in the experiments. And one month ahead prediction is enough to provide early warnings in the landslide prediction. Four individual forecasters, RBFNN, SVR, LSSVM and ELM are used to provide independent forecasts. The discounted MSFE is used to do the ANNs combination forecast. The comparison between monitoring data and prediction results is shown in Fig. 2. The prediction

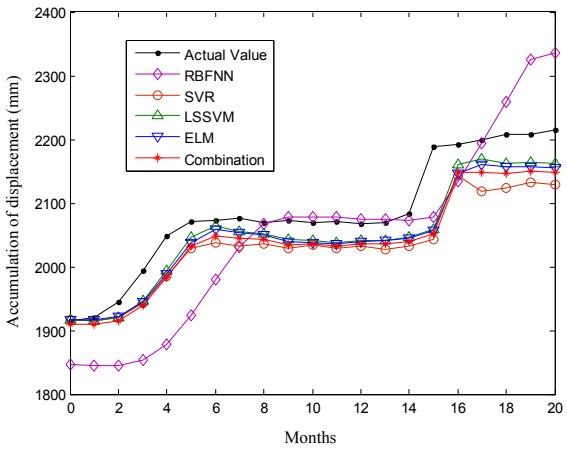

Fig. 2. The comparison between monitoring data and prediction results

results obtained from RBFNN are much lower than the actual values. The prediction results obtained from SVR, LSSVM and ELM, are acceptable. The combination forecast method is not the best one, but it performs similarly to the best individual ANN predictor and can reduce the risk of selecting

TABLE II

COMPARE THE PERFORMANCE FOR THE ABOVE-MENTIONED FIVE MODELS

	RBFNN	SVR	LSSVM	ELM	Combination
RMSE	84.0152	59.3113	42.9812	46.3649	52.2111

the types of ANNs. The comparison between these methods for the landslide displacement prediction via RMSE is shown in Table 2. The results obtained from TABLE 2 indicate that LSSVM is better than the other ANNs in case 1. But we can find that LSSVM is not the best one in case 2, which will be shown in the next subsection. It shows that different ANNs may fit different data sets, there may not be a best one for all cases.

B. Case study 2: Bazimen landslide

1) Date Collection: Bazimen landslide is situated on the right bank of XiangXi river, and it is 38km away from the Three Gorges Dam of China. The estimated volume of Bazimen landslide is about $2 \times 10^6 m^3$, with a maximum longitudinal dimension of $380m$ and width of 100 to $350m$. Three GPS monitoring points were located in the main cross-section of the landslide. Z111 point is selected as a case study. Fig. 3 shows the monitoring data of landslide accumulative displacement at Z111 monitoring point.

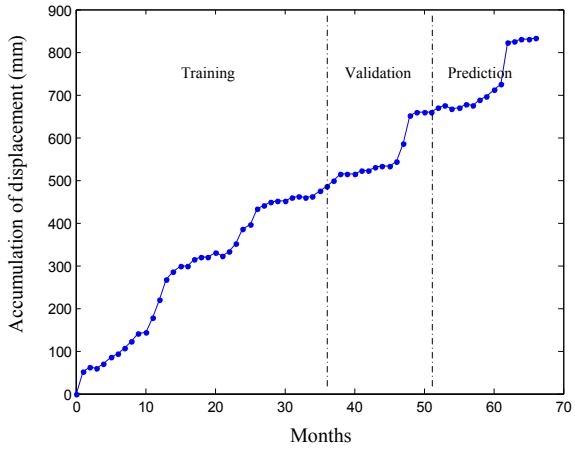

Fig. 3. Monitoring curves of landslide accumulative displacement at Z111 point

2) Selection of Parameters: The total number of monitoring data is 67 from July 2003 to January 2009. The whole data is divided into three data sets: the training set, the validation set and the prediction set. The 37 observations of data between July 2003 and July 2006 are selected as the training set, 15 observations of data from August 2006 to October 2007 are selected as the validation set, and the rest of 15 observations of data are the prediction set. The user-specified parameters used in ANNs are given in Table 3. The way to choose these parameters is the same to the case 1. Note that, the number of hidden nodes in RBFNN is 3.

TABLE III

PARAMETERS USED IN RBFNN, SVR, LSSVM AND ELM.

	RBFNN	SVR	LSSVM	ELM
C	/	2^{10}	2^{25}	2^{25}
γ	/	2^{-2}	2^{8}	2^{16}
σ	/	/	/	/
lag	1	2	5	2

TABLE IV

COMPARE THE PERFORMANCE FOR THE ABOVE-MENTIONED FIVE MODELS

	RBFNN	SVR	LSSVM	ELM	Combination
RMSE	45.7551	62.8610	69.0574	24.8314	24.5102

3) Experimental Results: Note that only one-step-ahead prediction is performed in the experiments. Four individual forecasters, RBFNN, SVR, LSSVM and ELM are used to provide independent forecasts. The discounted MSFE is used to do the ANNs combination forecast. The comparison between monitoring data and prediction results is shown in Fig. 4. The comparison between above methods for the landslide

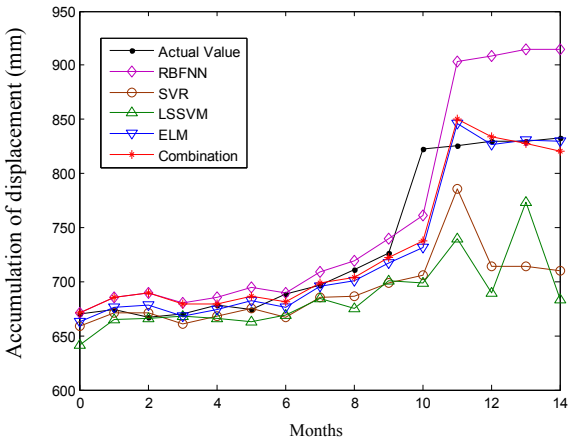

Fig. 4. The comparison between monitoring data and prediction results

displacement prediction via RMSE is shown in Table 4. As is shown in Fig. 4, the prediction result obtained by the combination forecast method is the best. The prediction result obtained by ELM is better than other individual ANNs. But ELM is not the best one in case 1. We can't decide which ANN to use until we have observed the prediction error. That is to say we can't know ELM is more suitable for case 2 in a real situation. But the ANNs combination based on the discounted MSFE can reduce the risk of choosing ANN model. A individual ANN with high prediction error may obtain a small weight in the combination model. From case 1 and case 2, we can find that the combined model is not always the best model, but it can perform better than or similarly to the best individual ANN predictor. The advantage of the proposed combination procedure is that you can use alternative ANN methods without worrying much about whether they are good at specific prediction work.

V. CONCLUSION

Nowadays, the study of landslide displacement prediction is a very important problem. A large number of geological models based on ANNs have been proposed for landslide prediction. At the same time, we will face with a number of ANN choices. In a real situation, we can't know which type of ANN is better until we have observed the prediction errors. In this paper, a ANNs combination method based on the discounted MSFE is proposed, which can reduce the risk of selecting the type of ANNs. The experimental results based on two case studies, Baishuihe landslide and Bazimen landslide, show that ANNs combination method can effectively improve the generalization ability of landslide displacement prediction and help to make future planning of the dangerous area.

REFERENCES

[1] Z. Ahmad and J. Zhang, "Selective combination of multiple neural networks for improving model predictionin nonlinear systems modelling through forward selection and backward elimination," *Neurocomputing*, vol. 72, pp. 1198–1204, 2009.

[2] P. Aleotti and R. Chowdhury, "Landslide hazard assessment: summary review and new perspectives," *Bulletin of Engineering Geology and the Environment*, vol. 58, no. 1, pp. 21–44, 1999.

[3] D.S. Broomhead and D. Lowe, "Multivariate functional interpolation and adaptive networks," *Complex Systems*, vol. 2, pp. 321–355, 1988.

[4] H.Q. Chen and Z.G. Zeng, "Deformation prediction of landslide based on improved back-propagation neural network," *Cognitive Computation*, vol. 5, no. 1, pp. 56–62, 2013.

[5] C. Cortes and V.N. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[6] H. Drucker, C. Burges, L Kaufman L and A. Smola and V.N. Vapnik, "Support Vector Regression Machines," in Advances in Neural Information Processing Systems 9, NIPS 1996, MIT Press, 1997, pp.155–161.

[7] J. Du, K.L. Yin and S. Lacasse, "Displacement prediction in colluvial landslides, Three Gorges Reservoir, China," *Landslides*, vol. 10, no. 2, pp. 203–218, 2013.

[8] S. Hashem, "Optimal Linear combinations of neural networks," *Neural Networks*, vol. 10, no. 4, pp. 599–614, 1997.

[9] 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.

[10] G.B. Huang, H.M. Zhou, X.J. Ding, R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 2, pp. 513–529, 2012.

[11] C. Lian, Z.G. Zeng, W. Yao, H.M. Tang, "Displacement prediction model of landslide based on a modified ensemble empirical mode decomposition and extreme learning machine," *Natural Hazards*, vol. 66, no. 2, pp. 759-771, 2013.

[12] C. Lian, Z.G. Zeng, W. Yao and H.M. Tang, "Ensemble of extreme learning machine for landslide displacement prediction based on time series analysis," *Neural Computing and Applications*, doi:10.1007/s00521-013-1446-3, 2013

[13] B. Pradhan, S. Lee and M.F. Buchroithner, "A GIS-based back-propagation neural network model and its cross-application and validation for landslide susceptibility analyses," *Computers, Environment and Urban Systems*, vol. 34, no. 3, pp. 216–235, 2010.

[14] J.A.K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.

[15] J.H. Stock and M.W. Watson, "Combination forecasts of output growth in a seven-country data set," *Journal of Economic Literature*, vol. 23, pp. 405–430, 2004.

[16] L. Yu, K.K. Lai and S.Y. Wang, "Multistage RBF neural network ensemble learning for exchange rates forecasting," *Neurocomputing*, vol. 71, pp. 3295–3302, 2008.
[17] J. Park and I.W. Sandberg, "Universal approximation using radial-basis-function networks," *Neural Computation*, vol. 3, no. 2, pp. 246–257, 1991.
[18] G.B. Huang, Q.Y. Zhu, and C.K. Siew, "Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks," International Joint Conference on Neural Networks (IJCNN'2004), (Budapest, Hungary), July, pp. 25–29, 2004.