Prediction of groundwater levels using evidence of chaos and support vector machine

Faming Huang, Jinsong Huang, Shui-Hua Jiang and Chuangbing Zhou

ABSTRACT

Many nonlinear models have been proposed to forecast groundwater level. However, the evidence of chaos in groundwater levels in landslide has not been explored. In addition, linear correlation analyses are used to determine the input and output variables for the nonlinear models. Linear correlation analyses are unable to capture the nonlinear relationships between the input and output variables. This paper proposes to use chaos theory to select the input and output variables for nonlinear models. The nonlinear model is constructed based on support vector machine (SVM). The parameters of SVM are obtained by particle swarm optimization (PSO). The proposed PSO-SVM model based on chaos theory (chaotic PSO-SVM) is applied to predict the daily groundwater levels in Huayuan landslide and the weekly, monthly groundwater levels in Baijiabao landslide in the Three Gorges Reservoir Area in China. The results show that there are chaos characteristics in the groundwater levels. The linear correlation analysis based PSO-SVM (linear PSO-SVM) and chaos theory-based back-propagation neural network (chaotic BPNN) are also applied for the purpose of comparison. The results show that the chaotic PSO-SVM model has higher prediction accuracy than the linear PSO-SVM and chaotic BPNN models for the test data considered.

Key words | chaos theory, groundwater level prediction, particle swarm optimization, phase space reconstruction, reservoir landslide, support vector machine

Faming Huang
Geological Survey Institute of China University of Geosciences,
Wuhan 430074,
China

Jinsong Huang (corresponding author)
ARC Centre of Excellence for Geotechnical Science and Engineering,
University of Newcastle,
Newcastle, NSW,
Australia
E-mail: *jinsong.huang@newcastle.edu.au*

Jinsong Huang
Shui-Hua Jiang
Chuangbing Zhou
School of Civil Engineering and Architecture,
Nanchang University,
Nanchang 330031,
China

INTRODUCTION

There are many landslides in the Three Gorges Reservoir Area especially after the impoundment of the Three Gorges Reservoir. The instability of many reservoir landslides has been related to the dramatic changes in the groundwater seepage field (Asch *et al.* 2009; Zhang *et al.* 2012). Therefore, the prediction of groundwater levels is of critical importance for landslide prevention (Keqiang *et al.* 2015).

Generally speaking, groundwater level prediction models include physically based and data-based models (Adamowski & Chan 2011). Some physically based models such as the ARX model (Knotters & Bierkens 2000), process-based spatio-temporal model (Schmidt & Dikau 2004) or the water-table fluctuation method (Park & Parker 2008) have been used to simulate and forecast the processes of groundwater level fluctuation. However, physically based models have practical limitations (Knotters & Bierkens 2000, 2002; Nourani *et al.* 2007). For example, spatial variations and the uncertainty of hydrological investigation also have negative effects on the accuracy of physically based models. Knotters & Bierkens (2001) provided a regionalized ARX model to determine parameters for physically based models considering spatial variability.

It is easy to build data-based models using only an input-output variable approach (Trichakis *et al.* 2011; Nourani & Komasi 2013). In recent years, many data-based models have been adopted to predict groundwater levels, such as regression models (Adamowski & Feluch 1991; Sahoo &

doi: 10.2166/hydro.2017.102

Jha 2013; Raghavendra & Deka 2016), artificial neural networks (ANN) (Nourani *et al.* 2008; Tsanis *et al.* 2008; Dash *et al.* 2010; Mohanty *et al.* 2010, 2015; Adamowski & Chan 2011; Jalalkamali *et al.* 2011; Sreekanth *et al.* 2011; Wang *et al.* 2012; Maheswaran & Khosa 2013; Atiquzzaman & Kandasamy 2015; Behnia & Rezaeian 2015; Chang *et al.* 2015), and genetic programming (Fallah-Mehdipour *et al.* 2013). Moreover, a literature review indicates that linear correlation analysis methods are widely used to select the input and output variables of these data-based models (Daliakopoulos *et al.* 2005; Nayak *et al.* 2006; Wong *et al.* 2007; Chen *et al.* 2009a; Yang *et al.* 2009; Chen *et al.* 2010; Sahoo & Jha 2013; Maiti & Tiwari 2014). The principal component analysis method can be used to reduce the redundant information in the input variables (Jolliffe 2002). Since groundwater level processes are non-linear in unconfined aquifers, a nonlinear model that can capture not only the overall appearance but also the underlying dynamic behavior of all of the nonlinear processes is required.

Both chaos theory (Sivakumar *et al.* 2001) and fractal theory (Zhang & Yang 2010) can be used to explore the nonlinear dynamic behavior of groundwater level time series. Both of these are sensitive to the initial conditions of the dynamic system, and there are self-similar characteristics in the chaos attractor and the fractal structure (Baas 2002). However, these methods usually explore the nonlinear system from different perspectives. The fractal theory mainly explores the structure of the attractor in geometrical space while the chaos theory mainly explores the evolution characteristic of the nonlinear system from the perspectives of time series (Peitgen *et al.* 2006).

Recently, chaos theory has been widely used in the nonlinear analysis of hydrological time series (Gutiérrez *et al.* 2006). Evidence of chaos has been demonstrated in many hydrological phenomena such as water level (Liong *et al.* 2005), precipitation (Jayawardena & Lai 1994), stream flows (Salas *et al.* 2005) and rainfall-runoff processes (Sivakumar *et al.* 2001). However, research on the evidence of chaos in groundwater levels in landslides has been very limited. In this study, based on the finding of evidence of chaos, embedding theory and phase space reconstruction (PSR) method are used to build the chaotic time series model. According to embedding theory (Takens 1981), in the long-term evolution of a chaotic groundwater level time series, information about the hidden states of the whole dynamic system can be preserved through a univariable groundwater level output. It is significant to effectively predict a nonlinear time series using a univariable model because sometimes it is difficult to obtain other correlated variables. The chaotic model is able to do nonlinear prediction using a univariable time series. In the PSR method from the chaotic model (King & Stewart 1992), a univariable groundwater level can be constructed into a multi-dimensional phase-space. As a result, the inputs and output of the nonlinear model can be obtained from the reconstructed multi-dimensional phase spaces.

It is necessary to choose a nonlinear model for chaotic groundwater level model building. As noted previously, ANN models (without the incorporation of chaos theory) have been extensively applied to forecast groundwater levels. ANN has limitations, however, including locally optimal values and the requirement of extensive data. Recently, support vector machines (SVM) were developed for time series prediction (Cortes & Vapnik 1995). SVM models have many advantages, including excellent generalization performance and global optimum. They have gained special attention in many areas such as electronic power prediction (Niu *et al.* 2010), traffic flow forecasting (Chen *et al.* 2009b), rainfall and runoff prediction (Tripathi *et al.* 2006), and landslide prediction (Feng *et al.* 2004; Yao *et al.* 2008). Meanwhile, the SVM model without chaos theory has also been used for groundwater level prediction (Behzad *et al.* 2009; Guzman *et al.* 2015; Gong *et al.* 2016). The main problem with SVM is the determination of its parameters. The particle swarm optimization (PSO) algorithm is widely used for SVM parameter selection because of its excellent global search ability (Lin *et al.* 2008; Fei *et al.* 2009). A novel PSO-SVM model based on chaos theory (chaotic PSO-SVM) is proposed in this study. It is then used to forecast the daily groundwater levels of the Huayuan landslide area, as well as the weekly and monthly groundwater levels of the Baijiabao landslide area.

To show the excellent generalization capability of PSO-SVM model, the back-propagation neural network model based on chaos theory (chaotic BPNN) is also used to predict the groundwater levels. The results show that the chaotic PSO-SVM model has higher prediction accuracy than the chaotic BPNN model for the test data. To show

the better guideline for determining the input and output variables, the PSO-SVM model based on linear correlation analysis (linear PSO-SVM) is also used. The results show that the PSR method is more appropriate to select the optimal input and output variables.

METHOD

The proposed method includes data pre-processing, PSR, evidence of chaos identification, the PSO-SVM model and accuracy assessment. All the programs are implemented in MATLAB R2015b. The PSR and the false nearest neighbor (FNN) algorithms in ChaosToolbox2p9_trial are used. The SVM model is tested and trained using the program libsvm-3.1-[FarutoUltimate 3.111code].

Data pre-processing

To prevent large values from overriding small values, the original groundwater levels are transformed into the desired range [0, 1] as:

$$x_i = \frac{x_{\text{old},i} - x_{\text{old,min}}}{x_{\text{old,max}} - x_{\text{old,min}}} \quad (1)$$

where $x_{\text{old},i}\,(i = 1, 2, \cdots\cdots, N)$ are the original groundwater levels, N is the number of groundwater levels, $x_{\text{old,min}}$ and $x_{\text{old,max}}$ are the lower and upper bounds of the original groundwater levels. x_i are used to predict groundwater level and the results are back-transformed to obtain the final predicted groundwater levels:

$$\hat{y}_i = y_i \times (x_{\text{old,max}} - x_{\text{old,min}}) + x_{\text{old,min}} \quad (2)$$

where y_i is the predicted groundwater level.

Phase space reconstruction

The fundamental properties of chaotic time series are the sensitivity to initial values, an evolution trace that becomes exponentially further apart as time increases, and the amplification of small disturbances in the nonlinear dynamic. Because geological conditions, rainfall and reservoir water level can affect the groundwater levels, the groundwater level time series is treated as a chaotic evolutionary system in this study. The PSR method (Kennel *et al.* 1992) is used to model the deterministic regular of strange attractors of groundwater level. The PSR method provides a simplified, multi-dimensional representation of a univariable nonlinear time series. In this approach, the nonlinear dynamics of a groundwater level can be fully embedded in a multi-dimensional phase space as:

$$X_i = (x_i, x_{i-\tau}, \cdots\cdots, x_{i-(m-1)\tau}) \quad (3)$$

where $X_i(i = 1 + (m-1)\tau, \cdots\cdots, N)$ is the reconstructed phase space, x_i is the normalized groundwater level, m is the embedding dimension ($m \geq d$), d is the dimension of the strange attractor of phase space, and τ is the delay time. τ represents the average length of memory of the system. The phase space of a univariable chaotic time series can be reconstructed by selecting appropriate τ and m (Sauer *et al.* 1991).

Determination of delay time

Typically, τ is determined using either the correlation analysis method (Aguirre 1995) or the mutual information method (Fraser & Swinney 1986). Both methods are generally suitable for noiseless long chaotic time series, and there are no commonly accepted guidelines for selection of parameters. Meanwhile, if a large value is selected for τ, the difficulty of computing the nonlinear model will increase greatly (Wolf *et al.* 1985). Therefore, a relatively small τ is usually chosen for the evidence of chaos identification and nonlinear prediction of the finite hydrological and environment time series (Sivakumar *et al.* 2002; Han & Wang 2009; Huang *et al.* 2016). In this study, the τ of daily, weekly and monthly groundwater levels are all set to 1, because the measurements of groundwater levels are finite and noisy. Huang *et al.* (2015) show that reasonably good groundwater level forecasts were obtained when τ was set to 1.

Determination of embedding dimension

The m is the minimum number of state variables required to describe the chaotic system. If m is too large, longer

groundwater level time series and more complex computations are needed. As a result, the efficiency of the nonlinear model will be reduced because of data redundancy. If m is too small, the strange attractors cannot be reconstructed (Cao 1997). The FNN method (Kennel *et al.* 1992) is one of the most popular methods for estimating optimal m because it is insensitive to the finite and noisy data points. It is used in this study.

The determination of the m using FNN method starts with an embedding space named R_m. Suppose $X_i = (x_i, x_{i-\tau}, \cdots\cdots, x_{i-(m-1)\tau})$ is a data point in the R_m and $X_i^{\text{Near}} = (x_i^{\text{Near}}, x_{i-\tau}^{\text{Near}}, \cdots\cdots, x_{i-(m-1)\tau}^{\text{Near}})$ is its nearest neighbor. The Euclidean distance between these two elements is calculated as:

$$E_m^2 = \sum_{\lambda=0}^{m-1} (x_{i-\lambda\tau} - x_{i-\lambda\tau}^{\text{Near}})^2 \tag{4}$$

The Euclidean distance between the projections of these two points into E_{m+1} is given by:

$$E_{m+1}^2 = E_m^2 + (x_{i-m\tau} - x_{i-m\tau}^{\text{Near}})^2 \tag{5}$$

The parameter S is defined as a measure of the distance between $x_{i-m\tau}$ and $x_{i-m\tau}^{\text{Near}}$ in R_{m+1}, normalized against their distance in R_m as:

$$S = \frac{(x_{i-m\tau} - x_{i-m\tau}^{\text{Near}})^2}{E_m} \tag{6}$$

The FNN method is aimed to search for all the data points which are neighbors in a particular embedding dimension m and which do not remain so when increasing the m to $m+1$. In this study, the ratio S of the distances between a particular data point X_i and its nearest neighbor X_i^{Near} in the $m+1$th and mth dimensions is computed. If the S is larger than a particular threshold S_{tol}, the neighbor is false. The S_{tol} is determined according to the number of reconstructed phase spaces. The S_{tol} can be set to 10 when the distribution of the reconstructed phase spaces in the dynamic system is sparse; and the S_{tol} can be set to a greater value when the distribution of the reconstructed phase spaces in the dynamic system is intensive (Kennel & Abarbanel 2002; Han 2007). There are just tens of phase spaces in the weekly and monthly groundwater levels, and hundreds of phase spaces in the daily groundwater levels in this study. Hence, the S_{tol} is set to 10. A greater S_{tol} can be used if there are thousands of phase spaces in the groundwater levels. When the percentage of FNN falls to 5%, the corresponding embedding dimension is considered high enough to represent the dynamics of the groundwater level.

Evidence of chaos identification

If there are chaos characteristics in the groundwater levels, chaotic models can be used to forecast the groundwater levels. Otherwise, if the groundwater levels are random or periodic time series, chaotic models cannot be used. The chaos characteristics of nonlinear time series are primarily identified by qualitative or quantitative methods (Welch 1967). Qualitative methods identify the chaotic time series through revealing the special spatial structures or frequency features shown in the spatial or frequency domains (Yeragani *et al.* 1993). Qualitative methods are less specific and have lower accuracy. Thus, it is preferred to identify the chaos characteristics through quantitative methods.

Quantitative methods include the largest Lyapunov exponent (LLE) method (Wolf *et al.* 1985), correlation dimension method (Broock *et al.* 1996) and Kolmogorov entropy method (Kosloff & Rice 1981). It is difficult to calculate the Kolmogorov entropy values of groundwater levels because the groundwater level time series are finite and noisy. Meanwhile, the LLE and correlation dimension methods have less stringent requirement than the Kolmogorov entropy method in the length of groundwater level time series. Hence, the LLE and correlation dimension methods are applied to characterize the chaotic system of groundwater levels in this study. The LLE describes the divergence rate of trajectories that starts close but diverges over time. In addition, the correlation dimension is one of the infinite numbers of dimensions in the dimension spectrum which characterizes the multi-fractal structure of the strange attractor (Procaccia *et al.* 1983). If the evidence of chaos in the groundwater levels can be revealed from the results of both of the LLE and correlation dimension methods, there are chaos characteristics in the groundwater levels. Then the chaotic models can be used to predict the groundwater levels.

Calculation of LLE

For a dynamic system, its sensitivity to initial conditions is quantified by the Lyapunov exponents. The small data sets method (Rosenstein *et al.* 1993) is both simple and appropriate for a finite time series. It is used to calculate the LLE in this study. If the LLE is denoted as L, the average divergence at time t can be defined as:

$$d(t) = ke^{Lt} \tag{7}$$

where k is a constant that normalizes the initial separation. The first step is to reconstruct the attractor dynamics from a univariable time series. The reconstructed trajectory, X, can be expressed as a matrix where each row is a phase-space vector. That is:

$$X = (X_1, X_2, \cdots\cdots, X_M)^T \tag{8}$$

where X_i is the ith data point of the dynamic system, M is the number of data points on the reconstructed attractor, For a N-point groundwater level time series, $\{x_1, x_2, \cdots\cdots, x_N\}$, each X_i is given by:

$$X_i = \{x_i, x_{i+\tau}, \cdots\cdots, x_{i+(m-1)\tau}\} \tag{9}$$

Thus X is a $M \times m$ matrix. The constants m, M, τ and N are related as:

$$M = N - (m-1)\tau \tag{10}$$

After reconstructing the dynamics, the nearest neighbor of each point can be located on the trajectory. The nearest neighbor X_j^{Near} is found by searching for the point that minimizes the distance to the particular reference point X_j as:

$$d_j(0) = \min_{X_j^{\text{Near}}} ||X_j - X_j^{\text{Near}}|| \tag{11}$$

where $d_j(0)$ is the initial distance from the jth point to its nearest neighbor, and $||\cdot||$ denotes the Euclidean norm. The L is then estimated as the mean rate of separation of the nearest neighbors. Based on the definition of L given in Equation (7), the jth pair of nearest neighbors diverge approximately at a rate given by:

$$d_j(t_i) \approx k_j e^{L(i\Delta t)} \tag{12}$$

where $t_i = i\Delta t$, Δt is the sampling period of the time series, k_j is the initial separation of the jth pair of nearest neighbors. Taking the logarithm of both sides of Equation (12),

$$\ln d_j(t_i) \approx \ln k_j + L(i\Delta t) \tag{13}$$

Equation (13) represents a set of approximately parallel lines (for $j = 1, 2, \cdots\cdots, M$), each with a slope roughly proportional to L. The L can be calculated using a least-squares fit to the 'average' line defined by

$$y(t_i) = \frac{1}{\Delta t}\langle \ln d_j(t_i)\rangle \tag{14}$$

where $\langle \cdot \rangle$ denotes the average over all j. This process of averaging is the key step to calculate accurate L using finite and noisy time series. If the dynamic system of groundwater levels contains chaos characteristics, the LLE must be greater than zero. This is because that, for a two trajectories with nearby initial conditions on the dynamic system, if the LLE is smaller than zero, the trajectories cannot diverge exponentially when time increases (Eckmann & Ruelle 1985).

Calculation of correlation dimension

Correlation dimension is one of the most efficient methods to identify the evidence of chaos. The method uses a fractal dimension, which is non-integer for chaotic systems. The Grassberger-Procaccia (G-P) approach (Grassberger & Procaccia 1983) is suitable for finite time series and easy to implement. It is used in this study to calculate the correlation dimension of groundwater levels.

Suppose X_i and X_j are the two points in the phase space, the correlation function is given by Theiler (1986):

$$C(r) = \lim_{N\to\infty} \frac{2}{M(M-1)} \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j\neq i}}^{M} H(r - ||X_i - X_j||) \tag{15}$$

where H is the Heaviside step function, with $H(u) = 1$ for $u \geq 0$, and $H(u) = 0$ for $u < 0$, r is the radius of the sphere centered on X_i or X_j; if the time series is characterized by an attractor, $C(r)$ can be related to radius r as:

$$C(r) \propto r^{D(m)} \tag{16}$$

where $D(m)$ is the correlation dimension. Take the logarithm of Equation (16) and rearrange it as:

$$D(m) = \lim_{r \to 0} \frac{\ln C(r)}{\ln r} \tag{17}$$

A series of $D(m)$ can be obtained by increasing the m. For a chaotic time series, $D(m)$ continuously increases and then converges to a constant when m increases. For a random or periodic time series, $D(m)$ increases without converging when m increases (Lai & Lerner 1998).

PSO-SVM model

After the evidence of chaos in the groundwater levels is identified, the evolution of groundwater levels can be predicted as:

$$y_i = f(X_i) \tag{18}$$

where y_i represents the one day, one week and one month ahead groundwater levels. $f(X_i)$ is a deterministic function. In this study, SVM is used to construct $f(X_i)$.

Support vector machine

SVM (Cortes & Vapnik 1995) is a non-linear kernel-based regression method. It is developed based on statistical learning theory. SVM maps the input data into a higher-dimensional feature space by nonlinear mapping and then solves a linear regression problem in the higher-dimensional feature space. It is aimed to find the best regression hyperplane with smallest structural risk in the feature space. One of the most popular SVMs is the ε-SVM which locates the hyperplane with an ε-insensitive loss value. The ϵ-SVM is formulated as:

$$f(X_i) = w^T \cdot \varphi(X_i) + b \tag{19}$$

where $\varphi(X_i)$ is a nonlinear mapping from the input space to the feature space, w is a vector of weight coefficients and b is a bias constant. w and b are estimated by the following optimization problem:

$$\text{minimize} \quad \frac{1}{2}\|w\|^2 \tag{20}$$

$$\text{subjected to} \quad \begin{cases} y_i - (\langle w, \varphi(X_i)\rangle + b) \leq \varepsilon \\ (\langle w, \varphi(X_i)\rangle + b) - y_i \leq \varepsilon \end{cases}$$

To cope with feasibility issues and to make the method more robust, points from the ε-insensitive band are not eliminated. Instead, these points are penalized by introducing slack variables ξ_i, ξ_i*:

$$\text{Minimize} \quad \frac{1}{2}\|w\|^2 + C_0 \sum_{i=1}^{M} (\xi_i + \xi_i*) \tag{21}$$

$$\text{subjected to} \quad \begin{cases} y_i - (\langle w, \varphi(X_i)\rangle + b) \leq \varepsilon + \xi_i \\ (\langle w, \varphi(X_i)\rangle + b) - y_i \leq \varepsilon + \xi_i* \\ \xi_i, \xi_i* \geq 0 \end{cases}$$

where the cost constant $C_0 > 0$ determines the trade-off between model complexity. After taking the Lagrangian and conditions for optimality, the solution in dual form is

$$f(X_i) = \sum_{i=1}^{M} (\alpha_i - \alpha_i*) K(X_i, X) + b \tag{22}$$

where α_i, α_i* are non-zero Lagrangian multipliers and the solution for the dual problem. $K(X_i, X)$ is the kernel function which represents the inner product $\langle \varphi(X_i), \varphi(X)\rangle$. In this study, the radial basis function (RBF) is used:

$$K(X_i, X) = \exp(-\gamma ||X_i - X||^2) \tag{23}$$

where γ is the width parameter of RBF kernel. In this study, the cost constant C_0, the insensitive loss ε and the kernel function parameter γ are determined by the PSO algorithm.

PSO-SVM

PSO (Eberhart & Kennedy 1995) is an adaptive algorithm for parameter selection. The general procedure of a PSO-SVM model (Wang *et al.* 2013) is shown in Figure 1. In the first step, the initial parameters such as the number of particles, learning factors and maximum iterations of the PSO algorithm are determined. In the second step, the SVM is trained and tested. In the third step, the root mean square error (*RMSE*) of SVM predicted values were selected as the fitness function, and the fitness function is regarded as the objective function for PSO algorithm. In the fourth and fifth steps, the position and velocity of all particles are updated by comparing the particle fitness value with the local and global best fitness values. Finally, the process is repeated until the maximum number of iterations is reached.

Accuracy assessment

Three assessment methods are used to evaluate the prediction effectiveness and precision of different models. The *RMSE* is calculated as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N_0} (x_{\text{old},i} - \hat{y}_i)^2}{N_0}} \tag{24}$$

where $x_{\text{old},i}$ is the original groundwater level, $\hat{y}_i$ is the final predicted values, and N_0 is the length of predicted data. The *RMSE* indicates the discrepancy between the monitoring and predicted values. The lower the *RMSE* is, the more accurate the prediction. In addition, the goodness of fit (R^2) is also used to assess the accuracy in this study. The R^2 represents the percentage of the initial uncertainty that is explained by the prediction models:

$$R^2 = 1 - \frac{N_0 \sum (x_{\text{old},i} - \hat{y}_i)^2}{N_0 \sum x_{\text{old},i}^2 - \sum \hat{y}_i^2} \tag{25}$$

Meanwhile, the Nash-Sutcliffe model efficiency coefficient (*NSE*) (Pulido-Calvo & Gutierrez-Estrada 2009) is also used to assess the performance of the three models as:

$$NSE = 1 - \frac{\sum_{i=1}^{N_0} (x_{\text{old},i} - \hat{y}_i)^2}{\sum_{i=1}^{N_0} (x_{\text{old},i} - \bar{x}_{\text{old},i})^2} \tag{26}$$

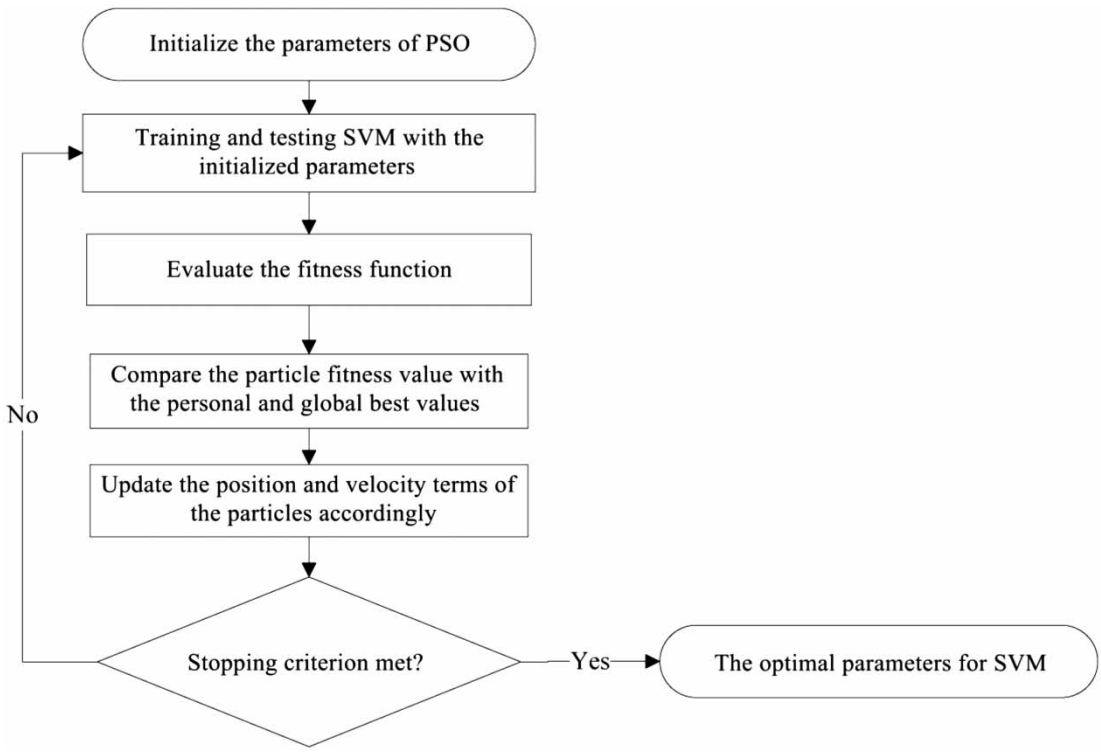

Figure 1 | Flow chart of PSO-SVM model.

The *NSE* falls in the range from 0 to 1. A *NSE* of 1 corresponds to a perfect match of the predictive values to the monitored data. The closer the *NSE* is to 1, the more accurate the model is.

Development of the proposed model

In summary, the proposed chaotic PSO-SVM model includes five main steps. In the first step, the original groundwater levels are pre-processed. In the second step, the phase spaces of the pre-processed groundwater levels are reconstructed, the reconstructed phase spaces of the daily, weekly and monthly groundwater levels are used as the input-output variables for the chaotic PSO-SVM. In the third step, The LLE and correlation dimension are used to identify the evidence of chaos. In the fourth step, the PSO-SVM model is used to train and test the input and output variables which are obtained from the PSR. Finally, the accuracy of the forecast results is assessed. The flow chart is shown in Figure 2.

CASE STUDIES

Research area and materials

The Huayuan and Baijiabao landslides in the Three Gorges Reservoir Area of China are used as the regions of interest. The locations of the Huayuan and Baijiabao landslides are shown in Figure 3. The Huayuan landslide occurs on the left bank of the Yangtze River, located in the Wanzhou district. The Baijiabao landslide is on the western side of the Xiangxi River, in Zigui county.

Huayuan landslide

The Huayuan landslide covers an area of 13.6×10^4 m^2 with a maximum longitudinal length of 380 m and a width of 360 m. The elevation of its leading edge is approximately 125 m and the elevation of the trailing edge is approximately 270 m, while the left and right boundaries are both defined by seasonal gullies. The topographic map of the landslide area is shown in Figure 4.

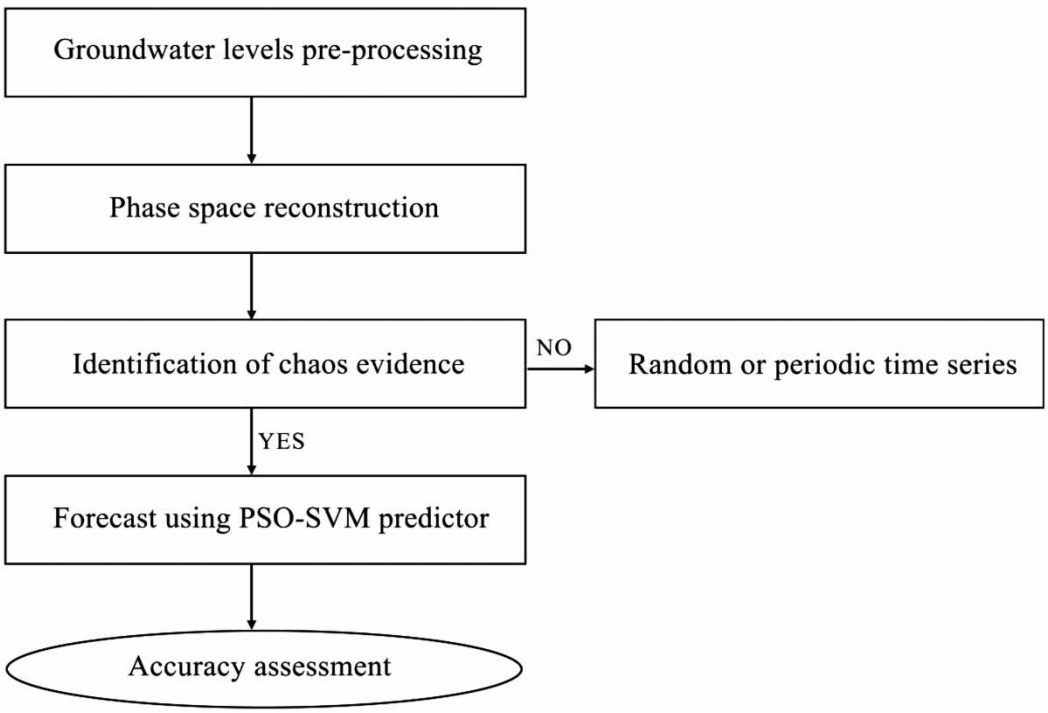

Figure 2 | Flow chart of the proposed chaotic PSO-SVM model.

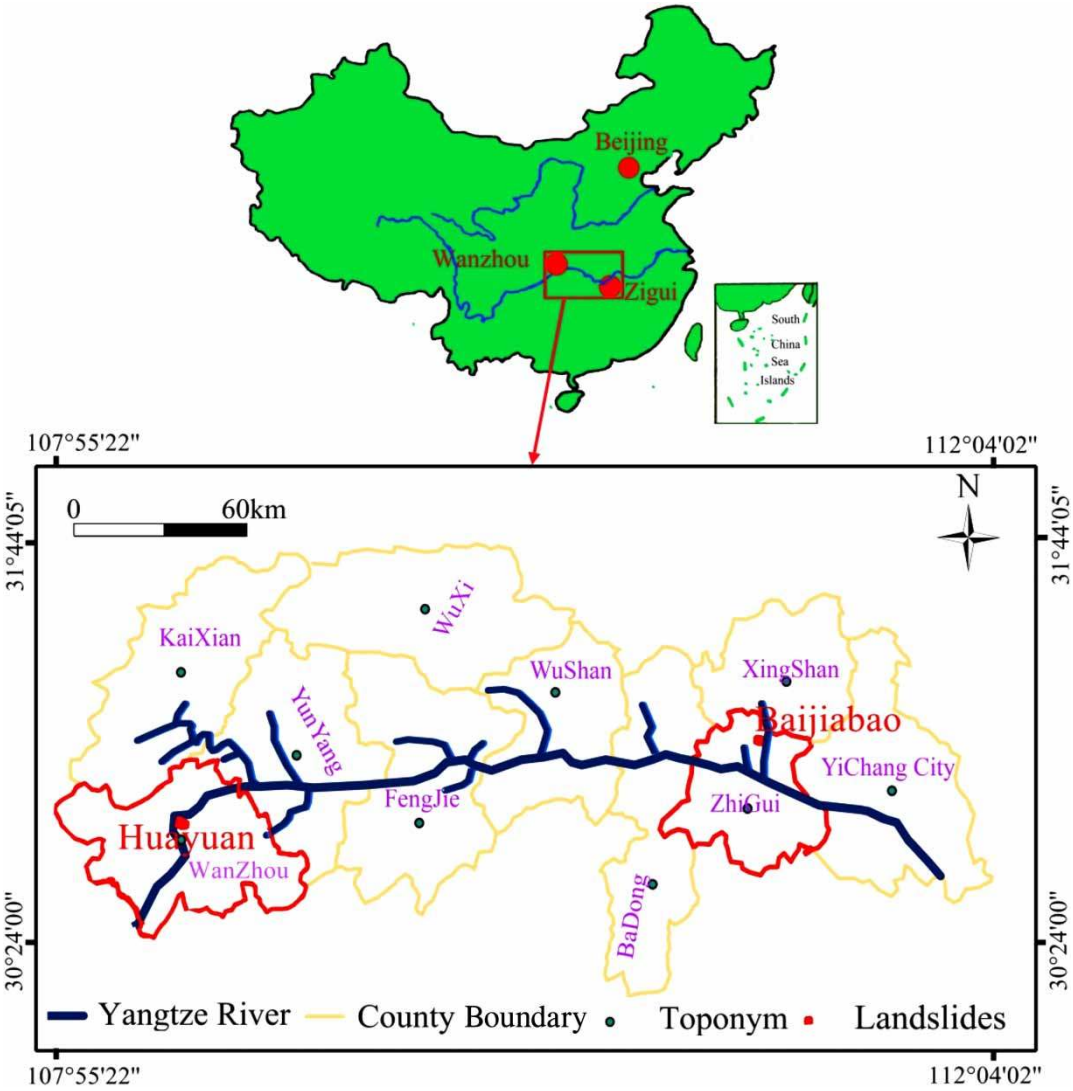

Figure 3 | Map showing the locations of the Huayuan and Baijiabao landslides.

The I-I′ geological section of the Huayuan landslide area is shown in Figure 5. The area consists of silty clay and fragmented rubble, with a loose structure and a medium permeability coefficient. The bedrock of the area is composed of lower Jurassic feldspar quartz sandstone and mudstone. The bedrock is overlain by silty clay fragment stone that is a soil slope with a loose structure. The groundwater types of the Huayuan landslide area are primarily loose debris pore water and bedrock fissure water.

Because of the complex geological conditions and external affecting factors such as reservoir water level, the groundwater level time series shows a complex nonlinear characteristic. As shown in Figure 6, the actual daily groundwater levels from 27 January 2013 to 3 January 2014 in the STK-1 hydrological hole were used as sample data in this study.

Baijiabao landslide

The Baijiabao landslide covers an area of 22×10^4 m^2 with a maximum longitudinal length of 550 m and a width of 400 m. The landslide extends between 135 m and 275 m in elevation and the mean depth of the sliding surface is approximately 45 m. The main sliding direction of Baijiabao landslide is oriented at N 85° E, and the left and right

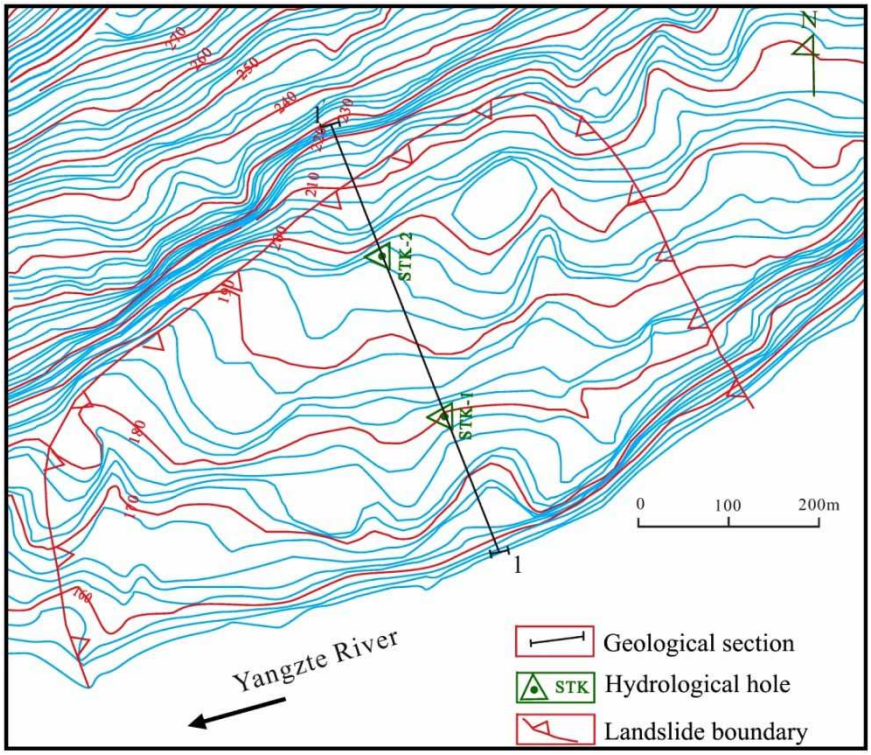

Figure 4 | Topographical map of the Huayuan landslide area and the location of hydrological hole.

boundaries of the landslide are defined by bedrock and a gully, respectively. The topographic map is shown in Figure 7.

The materials of Baijiabao landslide area are composed of silty clay and fragmented rubble with medium permeability. The landslide structure is loose with fragmented rubble content of approximately 15%. The lithology of the bedrock is mainly silty mudstones, muddy siltstones and sandstone. The dip direction and dip angle of the Baijiabao landslide are approximately 153° and 4°, respectively. The groundwater types of the landslide area are mainly loose debris pore water and bedrock fissure water. Its geological section is shown in Figure 8. The monitoring data of the weekly groundwater levels from 21 January 2007 to 4 July

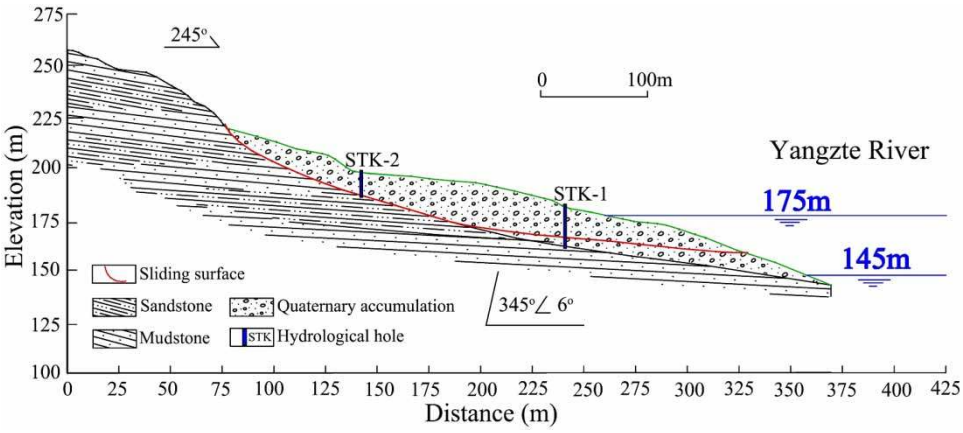

Figure 5 | Geological section of the Huayuan landslide area.

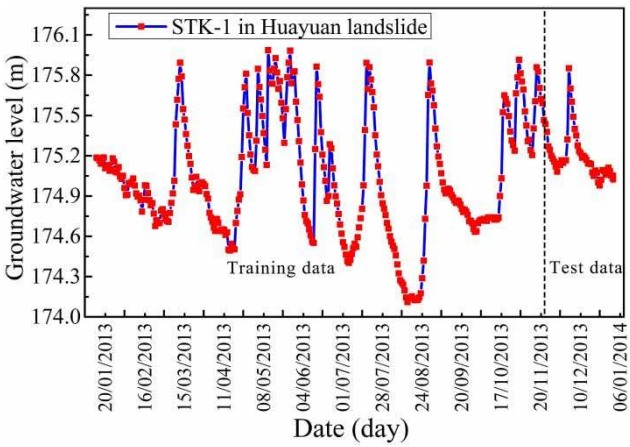

Figure 6 | Daily groundwater levels of the STK-1 hydrological hole.

2010 in the STK-2 hole and monthly groundwater levels from January 2007 to November 2010 in the STK-1 hole are used in this study, as shown in Figure 9.

Reconstruct the phase spaces of groundwater levels

The phase spaces of normalized daily ($x_{D(i)}$), weekly ($x_{W(i)}$) and monthly ($x_{M(i)}$) groundwater levels are reconstructed. The τ of all groundwater levels are set to 1. The optimal m of daily, weekly and monthly groundwater levels are 3, 3 and 2, respectively, as shown in Figure 10. The reconstructed phase spaces are shown in Table 1.

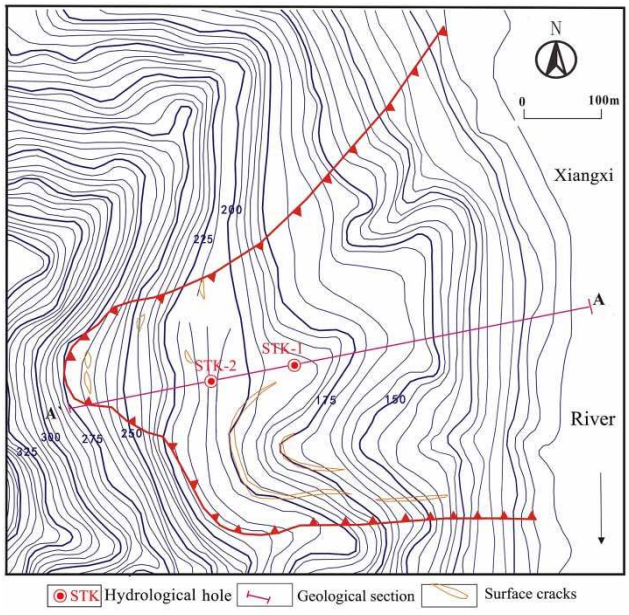

Figure 7 | Topographical map of the Baijiabao landslide area and the location of hydrological hole.

Evidence of chaos in groundwater levels

Figure 11 shows a curve of $\langle \ln d_j(t_i) \rangle$ versus $i\Delta t$. Also shown in Figure 11 is the L of groundwater levels. For the groundwater level series, the initial point is chosen near the attractor and the transient points are discarded. In Figure 11, the y axis is the average distance between all nearest neighbors after $i\Delta t$ discrete time steps. The red line is the fitted average line of the coordinate points $i\Delta t = 2, 3, \cdots, 8$. The slope of the red line is taken as L. The final calculated L of daily, weekly and monthly groundwater levels are 0.2291, 0.1533 and 0.0591, respectively. Although the L of monthly groundwater levels is close to zero, it just means that the orbits in the state space of monthly groundwater levels are very close (An *et al.* 2011; Zhou & Yin 2014). It is clear that there are chaos characteristics in all the groundwater levels because L is greater than zero.

Correlation dimension is also used to identify the evidence of chaos. For the daily, weekly and monthly groundwater levels, the corresponding $D(m)$ are calculated when m is increased from 1 to 20, 15 and 20, respectively. The relationships of $\ln C(r)$ and $\ln r$ in daily, weekly and monthly groundwater levels are shown in Figure 12. It can be seen that when m is increased to 11, 10 and 15, the slopes of the lines converge to a constant. Therefore, there are chaos characteristics in the daily, weekly and monthly groundwater levels.

Parameters of PSO-SVM model

Based on the reconstructed phase spaces and evidence of chaos, groundwater levels are predicted by PSO-SVM model. The position vector of a particle represents a parameter combination $(C_0, \varepsilon, \gamma)$ of the SVM. The final position vector is regarded as the optimal parameter combination of SVM, as shown in Table 2. The final predictive values are shown later in Figure 14.

In this study, the daily groundwater levels are relatively long time series with small fluctuations. Therefore, the C_0 of SVM for the daily groundwater levels should be high enough

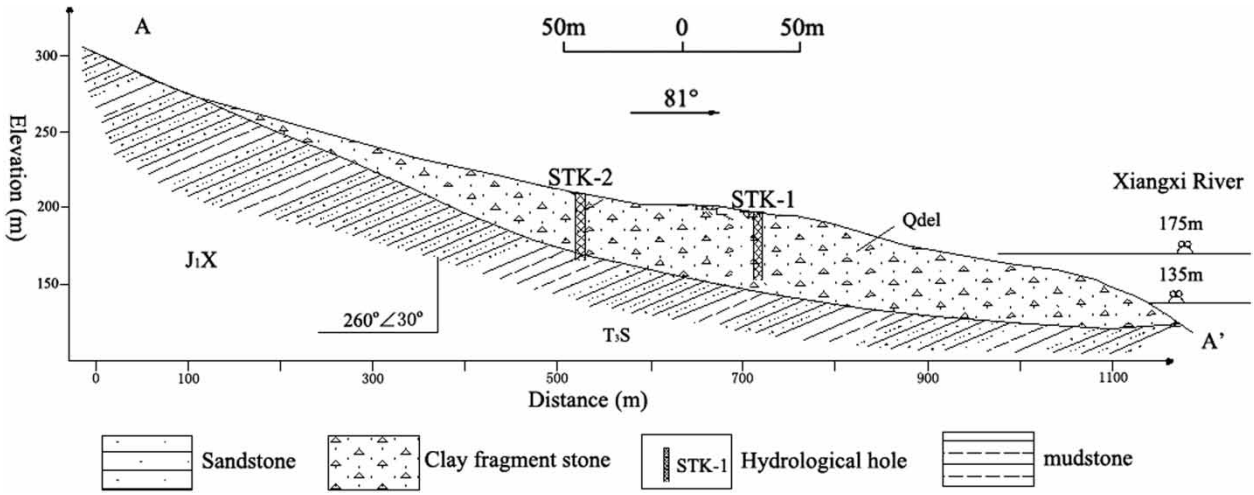

Figure 8 | Geological section of the Baijiabao landslide.

to fit the training data well, so as to predict the test data well. However, the weekly and monthly groundwater levels are very limited time series with large fluctuations. A high C_0 will result in the problem of over-fitting. Therefore, the C_0 of weekly and monthly groundwater levels is low to allow appropriate errors in the training process.

Comparisons with other models

Linear PSO-SVM model

The input and output variables for the linear PSO-SVM model are selected by linear correlation analysis. The input variables are regarded as the number of groundwater levels that have remarkable relationships with the output variable. Because groundwater levels are finite and belong to a univariable time series, the linear autoregressive method is commonly used (Daliakopoulos *et al.* 2005; Wong *et al.* 2007; Yang *et al.* 2009).

The results of the linear autoregressive analysis are shown in Figure 13. For daily groundwater levels, it can be seen from Figure 13 (top) that the autoregressive coefficient of $x_{D(i-6)}$ is 0.216 under a 95% level of significance. This suggests that $x_{D(i-6)}$, $x_{D(i-1)}$ and $x_{D(i)}$ have strong effects on the groundwater level $x_{D(i+1)}$. Therefore, the input variables for daily groundwater levels are $x_{D(i-6)}$, $x_{D(i-1)}$ and $x_{D(i)}$, and the output variable is $x_{D(i+1)}$, $i = 7, 8, \cdots\cdots, 336$. Similarly, the input variable of weekly groundwater levels is $x_{W(i)}$ and

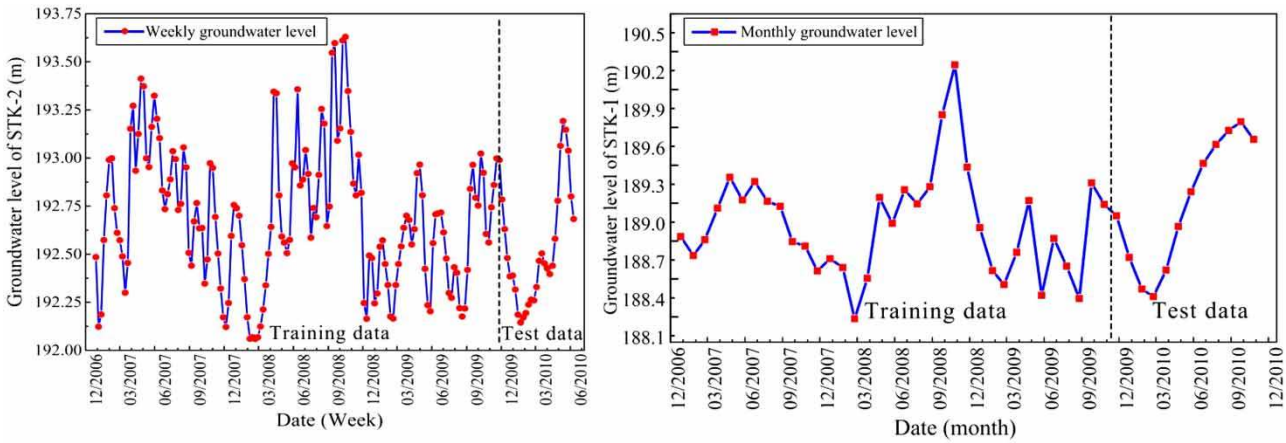

Figure 9 | Weekly groundwater level of STK-2 hole and monthly groundwater level of STK-1 hole in the Baijiabao landslide.

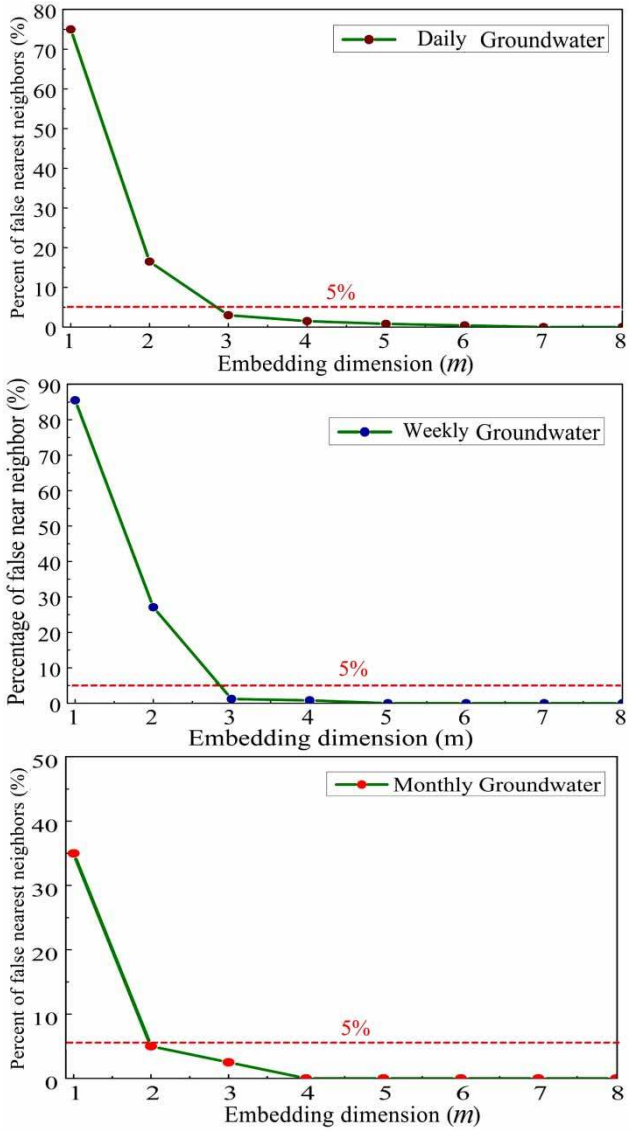

Figure 10 | The optimal m of daily, weekly, and monthly groundwater levels.

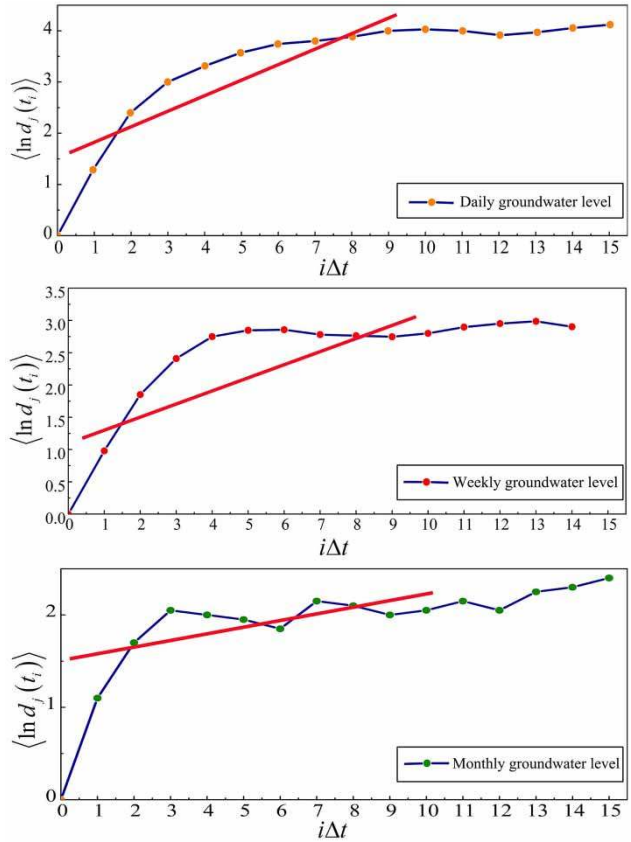

Figure 11 | The L values of daily, weekly and monthly groundwater levels.

the output variable is $x_{W(i+1)}$, $i = 1, 2, \cdots\cdots, 180$ as shown in Figure 13 (middle). The input variable of monthly groundwater levels is $x_{M(i)}$ and the output variable is $x_{M(i+1)}$, $i = 1, 2, \cdots\cdots, 46$ as shown in Figure 13 (bottom). In addition, the parameters of the SVM model are optimized by PSO. The inputs and outputs of the linear PSO-SVM model are shown in Table 3. The obtained parameters are shown in Table 4. The final predictive results are shown later in Figure 14.

Chaotic BPNN model

In order to compare the PSO-SVM model with the BPNN model, the same input and output variables that were used

Table 1 | Reconstructed phase spaces of daily, weekly and monthly groundwater levels

Groundwater levels	Reconstructed phase spaces (input variables)	Output variables
Daily	$X_{D(i)} = (x_{D(i)}, x_{D(i-1)}, x_{D(i-2)}), i = 3, 4, \cdots, 341$	$x_{D(i+1)}$
Weekly	$X_{W(i)} = (x_{W(i)}, x_{W(i-1)}, x_{W(i-2)}), i = 3, 4, \cdots, 180$	$x_{W(i+1)}$
Monthly	$X_{M(i)} = (x_{M(i)}, x_{M(i-1)}), i = 2, 3, \cdots, 46$	$x_{M(i+1)}$

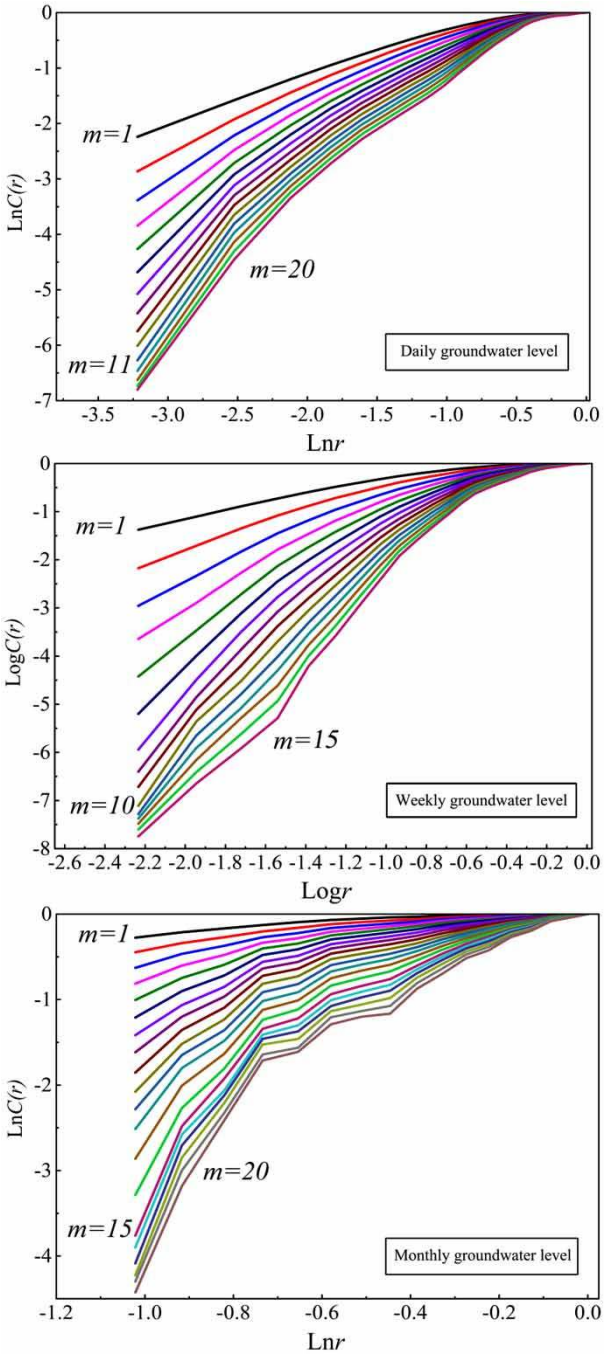

Figure 12 | The correlation dimension curves of daily, weekly and monthly groundwater levels.

in the chaotic PSO-SVM model are used again in the chaotic BPNN model. BPNN (Zhang & Wu 2009) is a commonly used ANN method. Recent research has demonstrated that the BPNN model with three-layer networks can fit nonlinear

Table 2 | Parameters of the chaotic PSO-SVM model

	Groundwater levels	$(C_0, \varepsilon, \gamma)$
Chaotic PSO-SVM	Daily	(6,562.5, 0.011, 0.344)
	Weekly	(67.4, 0.031, 0.287)
	Monthly	(30.5, 0.162, 0.498)

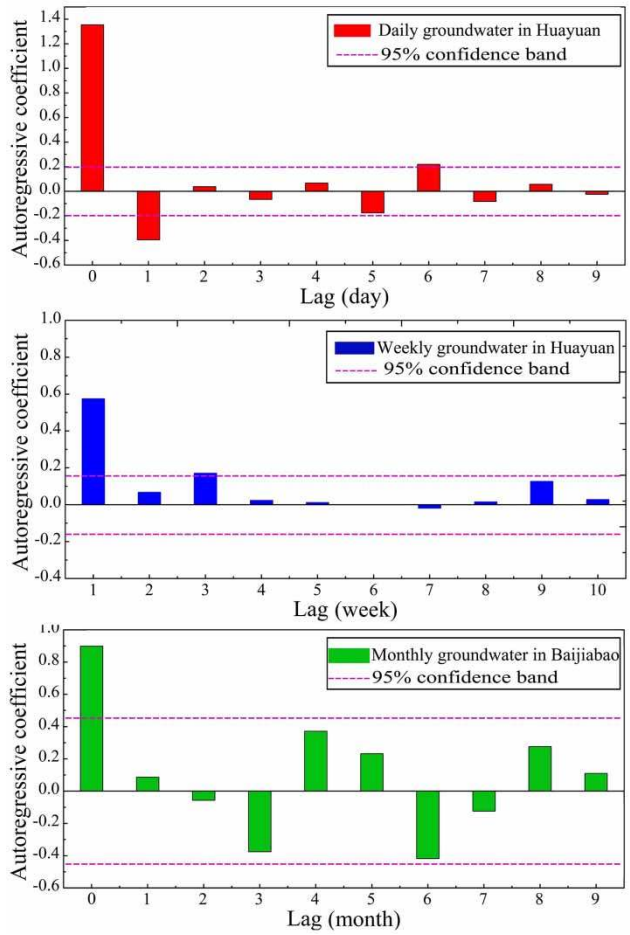

Figure 13 | Autoregressive coefficient of daily, weekly, and monthly groundwater levels.

mapping relationships accurately (Jiang *et al.* 2008). In this study, a three-layered BPNN model is used. The input variables of the chaotic BPNN model are shown in Table 1. Three nodes of daily, two nodes of weekly, and two nodes of monthly groundwater levels are defined as input layers, and one node of all of the groundwater levels is defined as the output layer. The number of hidden layer nodes is determined using the trial-and error method (Basheer & Hajmeer 2000; Yang *et al.* 2009). The final selected hidden layer nodes

Table 3 | Inputs and outputs of daily, weekly and monthly groundwater levels of linear PSO-SVM model

Groundwater levels	Input variables	Output variables
Daily	$(x_{D(i)}, x_{D(i-1)}, x_{D(i-6)}), i = 7, 8, \cdots, 336$	$x_{D(i+1)}$
Weekly	$x_{W(i)}, i = 1, 2, \cdots\cdots, 180$	$x_{W(i+1)}$
Monthly	$x_{M(i)}, i = 1, 2, \cdots\cdots, 46$	$x_{M(i+1)}$

Table 4 | Parameters of the linear PSO-SVM model

	Groundwater levels	$(C_0, \varepsilon, \gamma)$
Linear PSO-SVM	Daily	(3,562.5, 0.014, 0.352)
	Weekly	(72.8, 0.063, 0.291)
	Monthly	(179.2, 0.097, 0.878)

of daily, weekly and monthly groundwater levels are 7, 5 and 5, respectively.

The built-in BPNN model in the MATLAB R2015b is used. The initial weights and thresholds for all connection links are set randomly within the range from 0 to 1. The transferring functions of the neural networks of the hidden layer and the output layer are *tansig* and *purelin* functions. The weight values of connection links are trained by gradient descent algorithm. The maximum iteration number is set to 1,000. The learning rate is set to 0.01. The training error is 0.001. The predictive results of the chaotic BPNN model are shown later in Figure 14.

Comparison of the three models

Training accuracies of the three models

The training accuracies of the three models for the daily, weekly and monthly groundwater levels are shown in Table 5. It can be seen from Table 5 that the daily, weekly and monthly groundwater levels are trained well by the three models. It can also be seen from Table 6 that the testing accuracies of the three models are reasonably good. Hence, there are no overtraining signs in the three models.

Comparison of the prediction results

The final prediction results of one day, one week and one month ahead groundwater levels are compared in Table 6 and Figure 14. It can be seen from Figure 14 that although some prediction values deviate from the monitoring data, the chaotic PSO-SVM model predicts well the groundwater level especially for the daily and weekly groundwater levels. However, as highlighted by the green ellipse in Figure 14, the fluctuation of the daily, weekly and monthly groundwater levels are not well predicted by the chaotic BPNN and linear PSO-SVM models.

The prediction performances of the three models are also compared in Table 6, with the three indices *RMSE*, R^2 and *NSE*. It can be seen from Table 4 that the daily data has the highest prediction accuracy, and the monthly data has the lowest prediction accuracy. It can also be seen from Table 4 that the *NSE* metric indicates that chaotic PSO-SVM model is superior to the chaotic BPNN model and the linear PSO-SVM model for the test data. The other two indices also show that the performance of the chaotic PSO-SVM model is better than the chaotic BPNN model and the linear PSO-SVM model for the test data of all daily, weekly and monthly groundwater levels.

The bias metric is also used to estimate the prediction performances of the three models for the daily, weekly and monthly groundwater levels. It can be seen from Figure 15(a) that the predicted daily groundwater levels on 15 and 16 December 2013 are overestimated, and the other daily groundwater levels are predicted well. It can be seen from Figure 15(b) that the weekly groundwater levels from 27 December 2009 to 21 March 2010 are underestimated while the remaining weekly groundwater levels are overestimated in general. The average bias values of the weekly groundwater level predictions are larger than those of daily groundwater levels. It can be seen from Figure 15(c) that the monthly groundwater levels from December 2009 to April 2010 are underestimated while the remaining monthly groundwater levels are overestimated. The average bias values of the monthly groundwater level predictions are also larger than the ones of daily groundwater levels, but are almost the same as the ones of weekly groundwater levels. The bias metrics in Figure 15 show that the bias values of the chaotic PSO-SVM model are smaller than the chaotic BPNN and linear PSO-SVM models for the daily, weekly and monthly groundwater level predictions.

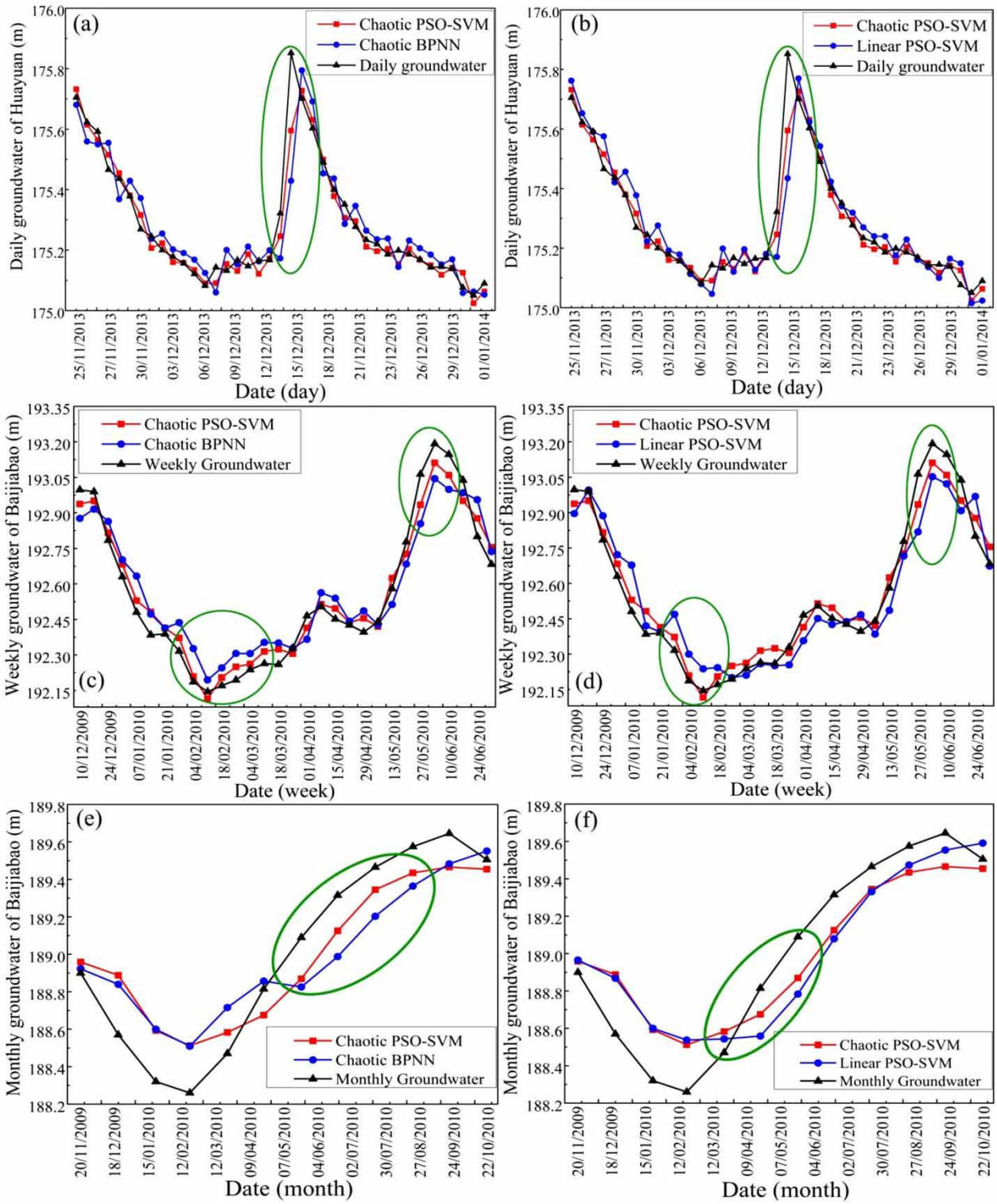

Figure 14 | Final prediction results of groundwater levels: (a) one-day ahead prediction using chaotic PSO-SVM and chaotic BPNN models; (b) one-day ahead prediction using chaotic PSO-SVM and linear PSO-SVM models; (c) one-week ahead prediction using chaotic PSO-SVM and chaotic BPNN models; (d) one-week ahead prediction using chaotic PSO-SVM and linear PSO-SVM models; (e) one-month ahead prediction using chaotic PSO-SVM and linear PSO-SVM models; and (f) one-month ahead prediction using chaotic PSO-SVM and chaotic BPNN models.

Table 5 | Accuracy assessment of the training data

	Chaotic PSO-SVM			Chaotic BPNN			Linear PSO-SVM		
Groundwater	***RMSE* (m)**	**R^2**	***NSE***	***RMSE* (m)**	**R^2**	***NSE***	***RMSE* (m)**	**R^2**	***NSE***
Daily	0.054	0.948	0.952	0.034	0.958	0.964	0.057	0.942	0.943
Weekly	0.097	0.904	0.884	0.072	0.923	0.921	0.106	0.896	0.878
Monthly	0.142	0.923	0.875	0.103	0.941	0.906	0.151	0.914	0.869

Table 6 | Accuracy assessment and performance comparison of the test data

	Chaotic PSO-SVM			Chaotic BPNN			Linear PSO-SVM		
Groundwater	***RMSE* (m)**	**R^2**	***NSE***	***RMSE* (m)**	**R^2**	***NSE***	***RMSE* (m)**	**R^2**	***NSE***
Daily	0.075	0.917	0.938	0.098	0.906	0.912	0.105	0.895	0.904
Weekly	0.127	0.893	0.867	0.148	0.863	0.836	0.157	0.852	0.825
Monthly	0.189	0.896	0.849	0.223	0.857	0.826	0.207	0.831	0.817

Discussion

The forecast results indicate that the chaotic PSO-SVM model is more accurate and credible than the linear PSO-SVM model, and the PSO-SVM model is more appropriate for finite nonlinear data than the BPNN model. The comparisons indicate that the PSR method of chaos theory provides the input and output variables more appropriately than the linear autoregressive model. It can be seen that there are similarities between the PSR method and SVM model. The PSR method can determine the number of input variables by extending one-dimensional groundwater level to high dimensions referred to as 'embedding dimensions' (Sivakumar *et al.* 2002). Similarly, the SVM method provides a nonlinear kernel function to map the input variables into high dimensional feature space. Both methods map the groundwater level from the low-dimensional space to the high-dimensional space (Wang & Shi 2013). Hence, the PSR method is suitable to determine the input and output variables for the SVM model. The nonlinear dynamical characteristics of groundwater level can be reflected in the embedding space.

However, all the three models have relatively low prediction precision for the extreme monitoring values. One important reason is that the length of the groundwater level time series is not long enough and the reconstructed strange attractors are not fully unfolded to reflect the original strange attractors. One year is a very short time period for data-based model training. Another reason is that only groundwater level is used as input. For future studies, a nonlinear model based on multivariable chaos theory (Garcia & Almeida 2005) can be used, which can consider some other input variables such as rainfall, reservoir water level and temperature. Furthermore, the prediction windows of this study are one day, one week and one month, the prediction window can be enlarged by feeding the predicted values to the model again as previously done by Trichakis *et al.* (2009). In addition, all the predictions are conducted in a server with Intel Xeon CPU X5675@3.07 GHz with 256GB RAM. The computational time of the chaotic PSO-SVM, chaotic BPNN and linear PSO-SVM models for the daily groundwater level prediction is 184.5 s, 24.6 s and 179.3 s, respectively.

CONCLUSION

Based on chaos theory, this study proposes the chaotic PSO-SVM model for groundwater level predictions. Two criteria are proposed to ensure that there is evidence of chaos in the groundwater levels. The first one is that the LLE should be greater than zero. The second criterion is that the values

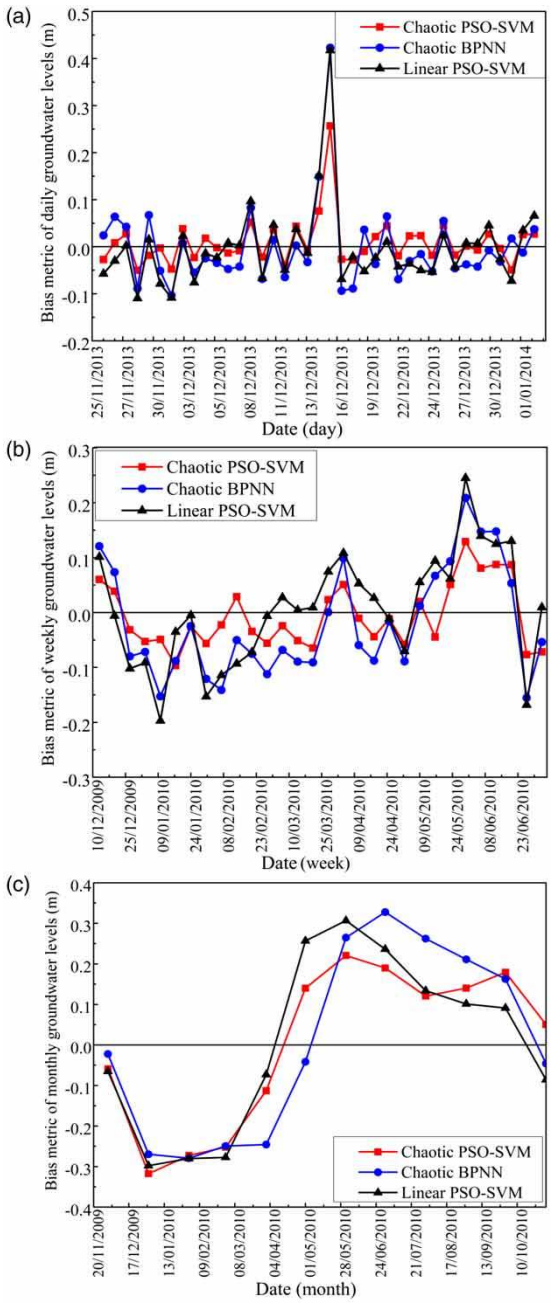

Figure 15 | Bias metrics of the (a) daily, (b) weekly and (c) monthly groundwater levels.

of correlation dimension should converge to a constant when the embedding dimension increases.

The proposed model is used to predict the daily groundwater levels in Huayuan landslide, and weekly and monthly groundwater levels in Baijiabao landslide. Both chaotic BPNN and linear PSO-SVM models are used for comparisons. The results show that the chaotic PSO-SVM model provides the best predictions among the three methods for the test data considered. However, there are still some limitations of the three models. For the chaotic PSO-SVM model, only the groundwater level is used as an input variable without considering the external factors such as rainfall and reservoir water level. Further research is needed to develop models based on multivariable chaos theory. In addition, relatively long groundwater level time series is needed to train and test the chaotic PSO-SVM model well. For the chaotic BPNN model, the prediction accuracy is not ideal although the prediction efficiency is high. For the linear PSO-SVM model, the input variables are selected based on simple linear correlation coefficient, which sometimes cannot identify properly the input variables.

In summary, the chaotic PSO-SVM model has advantages in determining the input-output variables through nonlinear method, and obtaining more accurate predictive values of groundwater levels than the linear PSO-SVM and chaotic BPNN models. The proposed chaotic PSO-SVM model can be used to predict the real world groundwater levels.

ACKNOWLEDGEMENTS

The first author is currently supported by the China Scholarship Council to visit the ARC Centre of Excellence for Geotechnical Science and Engineering, University of Newcastle, NSW, Australia. This work was also supported by the National Natural Science Foundation of China (Project No. 51679117, 51509125).

REFERENCES

Adamowski, J. & Chan, H. F. 2011 A wavelet neural network conjunction model for groundwater level forecasting. *Journal of Hydrology* **407** (1), 28–40.

Adamowski, K. & Feluch, W. 1991 Application of nonparametric regression to groundwater level prediction. *Canadian Journal of Civil Engineering* **18** (4), 600–606.

Aguirre, L. A. 1995 A nonlinear correlation function for selecting the delay time in dynamical reconstructions. *Physics Letters A* **203** (2), 88–94.

An, X., Jiang, D., Liu, C. & Zhao, M. 2011 Wind farm power prediction based on wavelet decomposition and chaotic time

series. *Expert Systems with Applications* **38** (9), 11280–11285.

Asch, T. W. J. V., Malet, J. P. & Bogaard, T. A. 2009 The effect of groundwater fluctuations on the velocity pattern of slow-moving landslides. *Natural Hazards & Earth System Sciences* **9** (3), 739–749.

Atiquzzaman, M. & Kandasamy, J. 2016 Prediction of hydrological time-series using extreme learning machine. *Journal of Hydroinformatics* **18** (2), 345–353.

Baas, A. C. 2002 Chaos, fractals and self-organization in coastal geomorphology: simulating dune landscapes in vegetated environments. *Geomorphology* **48** (1), 309–328.

Basheer, I. & Hajmeer, M. 2000 Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods* **43** (1), 3–31.

Behnia, N. & Rezaeian, F. 2015 Coupling wavelet transform with time series models to estimate groundwater level. *Arabian Journal of Geosciences* **8** (10), 8441–8447.

Behzad, M., Asghari, K. & Coppola Jr, E. A. 2009 Comparative study of SVMs and ANNs in aquifer water level prediction. *Journal of Computing in Civil Engineering* **24** (5), 408–413.

Broock, W., Scheinkman, J. A., Dechert, W. D. & LeBaron, B. 1996 A test for independence based on the correlation dimension. *Econometric Reviews* **15** (3), 197–235.

Cao, L. 1997 Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D: Nonlinear Phenomena* **110** (1), 43–50.

Chang, J., Wang, G. & Mao, T. 2015 Simulation and prediction of suprapermafrost groundwater level variation in response to climate change using a neural network model. *Journal of Hydrology* **529**, 1211–1220.

Chen, L.-H., Chen, C.-T. & Pan, Y.-G. 2009a Groundwater level prediction using SOM-RBFN multisite model. *Journal of Hydrologic Engineering* **15** (8), 624–631.

Chen, S., Wang, W. & Van Zuylen, H. 2009b Construct support vector machine ensemble to detect traffic incident. *Expert Systems with Applications* **36** (8), 10976–10986.

Chen, L.-H., Chen, C.-T. & Lin, D.-W. 2010 Application of integrated back-propagation network and self-organizing map for groundwater level forecasting. *Journal of Water Resources Planning and Management* **137** (4), 352–365.

Cortes, C. & Vapnik, V. 1995 Support-vector networks. *Machine Learning* **20** (3), 273–297.

Daliakopoulos, I. N., Coulibaly, P. & Tsanis, I. K. 2005 Groundwater level forecasting using artificial neural networks. *Journal of Hydrology* **309** (1), 229–240.

Dash, N. B., Panda, S. N., Remesan, R. & Sahoo, N. 2010 Hybrid neural modeling for groundwater level prediction. *Neural Computing and Applications* **19** (8), 1251–1263.

Eberhart, R. C. & Kennedy, J. 1995 A new optimizer using particle swarm theory. In: *Paper Presented at the Proceedings of the Sixth International Symposium on Micro Machine and Human Science*.

Eckmann, J.-P. & Ruelle, D. 1985 Ergodic theory of chaos and strange attractors. *Reviews of Modern Physics* **57** (3), 617.

Fallah-Mehdipour, E., Haddad, O. B. & Mariño, M. 2013 Prediction and simulation of monthly groundwater levels by genetic programming. *Journal of Hydro-Environment Research* **7** (4), 253–260.

Fei, S.-W., Wang, M.-J., Miao, Y.-B., Tu, J. & Liu, C.-L. 2009 Particle swarm optimization-based support vector machine for forecasting dissolved gases content in power transformer oil. *Energy Conversion and Management* **50** (6), 1604–1609.

Feng, X.-T., Zhao, H. & Li, S. 2004 Modeling non-linear displacement time series of geo-materials using evolutionary support vector machines. *International Journal of Rock Mechanics and Mining Sciences* **41** (7), 1087–1107.

Fraser, A. M. & Swinney, H. L. 1986 Independent coordinates for strange attractors from mutual information. *Physical Review A* **33** (2), 1134.

Garcia, S. P. & Almeida, J. S. 2005 Multivariate phase space reconstruction by nearest neighbor embedding with different time delays. *Physical Review E* **72** (2), 027205.

Gong, Y., Zhang, Y., Lan, S. & Wang, H. 2016 A comparative study of artificial neural networks, support vector machines and adaptive neuro fuzzy inference system for forecasting groundwater levels near lake okeechobee, Florida. *Water Resources Management* **30** (1), 375–391.

Grassberger, P. & Procaccia, I. 1983 Characterization of strange attractors. *Physical Review Letters* **50** (5), 346–349.

Gutiérrez, J. M., Galván, A., Cofiño, A. & Primo, C. 2006 Chaos game characterization of temporal precipitation variability: application to regionalization. *Fractals* **14** (02), 87–99.

Guzman, S. M., Paz, J. O., Tagert, M. L. M. & Mercer, A. 2015 Artificial neural networks and support vector machines: Contrast study for groundwater level prediction. In: *Paper Presented at the 2015 ASABE Annual International Meeting*.

Han, M. 2007 *Prediction Theory and Method of Chaotic Time Series*. China Water Power Press, Beijing, pp. 50–92.

Han, M. & Wang, Y. 2009 Analysis and modeling of multivariate chaotic time series based on neural network. *Expert Systems with Applications* **36** (2), 1280–1290.

Huang, F. M., Yin, K. L., Zhang, G.-R., Tang, Z.-Z. & Zhang, J. 2015 Prediction of groundwater level in landslide using multivariable PSO-SVM model. *Journal of Zhejiang University (Engineering Science Edition)* **49** (6), 1193–1200.

Huang, F. M., Wu, P. & Ziggah, Y. Y. 2016 GPS monitoring landslide deformation signal processing using time-series model. *International Journal of Signal Processing, Image Processing and Pattern Recognition* **9** (3), 321–332.

Jalalkamali, A., Sedghi, H. & Manshouri, M. 2011 Monthly groundwater level prediction using ANN and neuro-fuzzy models: a case study on Kerman plain, Iran. *Journal of Hydroinformatics* **13** (4), 867–876.

Jayawardena, A. & Lai, F. 1994 Analysis and prediction of chaos in rainfall and stream flow time series. *Journal of Hydrology* **153** (1), 23–52.

Jiang, S., Ren, Z., Xue, K. & Li, C. 2008 Application of BPANN for prediction of backward ball spinning of thin-walled tubular

part with longitudinal inner ribs. *Journal of Materials Processing Technology* **196** (1), 190–196.

Jolliffe, I. 2002 *Principal Component Analysis*. Wiley Online Library, John Wiley & Sons.

Kennel, M. B. & Abarbanel, H. D. 2002 False neighbors and false strands: a reliable minimum embedding dimension algorithm. *Physical Review E* **66** (2), 026209.

Kennel, M. B., Brown, R. & Abarbanel, H. D. 1992 Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A* **45** (6), 3403.

Keqiang, H., Zhiliang, W., Xiaoyun, M. & Zengtao, L. 2015 Research on the displacement response ratio of groundwater dynamic augment and its application in evaluation of the slope stability. *Environmental Earth Sciences* **74** (7), 5773–5791.

King, G. & Stewart, I. 1992 Phase space reconstruction for symmetric dynamical systems. *Physica D: Nonlinear Phenomena* **58** (1), 216–228.

Knotters, M. & Bierkens, M. F. 2000 Physical basis of time series models for water table depths. *Water Resources Research* **36** (1), 181–188.

Knotters, M. & Bierkens, M. F. 2001 Predicting water table depths in space and time using a regionalised time series model. *Geoderma* **103** (1), 51–77.

Knotters, M. & Bierkens, M. 2002 Accuracy of spatio-temporal RARX model predictions of water table depths. *Stochastic Environmental Research and Risk Assessment* **16** (2), 112–126.

Kosloff, R. & Rice, S. A. 1981 The influence of quantization on the onset of chaos in hamiltonian systems: the Kolmogorov entropy interpretation. *The Journal of Chemical Physics* **74** (2), 1340–1349.

Lai, Y.-C. & Lerner, D. 1998 Effective scaling regime for computing the correlation dimension from chaotic time series. *Physica D: Nonlinear Phenomena* **115** (1), 1–18.

Lin, S.-W., Ying, K.-C., Chen, S.-C. & Lee, Z.-J. 2008 Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Systems with Applications* **35** (4), 1817–1824.

Liong, S., Phoon, K., Pasha, M. & Doan, C. 2005 Efficient implementation of inverse approach for forecasting hydrological time series using micro GA. *Journal of Hydroinformatics* **7** (3), 151–163.

Maheswaran, R. & Khosa, R. 2013 Long term forecasting of groundwater levels with evidence of non-stationary and nonlinear characteristics. *Computers & Geosciences* **52**, 422–436.

Maiti, S. & Tiwari, R. 2014 A comparative study of artificial neural networks, bayesian neural networks and adaptive neuro-fuzzy inference system in groundwater level prediction. *Environmental Earth Sciences* **71** (7), 3147–3160.

Mohanty, S., Jha, M. K., Kumar, A. & Sudheer, K. 2010 Artificial neural network modeling for groundwater level forecasting in a river island of Eastern India. *Water Resources Management* **24** (9), 1845–1865.

Mohanty, S., Jha, M. K., Raul, S., Panda, R. & Sudheer, K. 2015 Using artificial neural network approach for simultaneous forecasting of weekly groundwater levels at multiple sites. *Water Resources Management* **29** (15), 5521–5532.

Nayak, P. C., Rao, Y. S. & Sudheer, K. 2006 Groundwater level forecasting in a shallow aquifer using artificial neural network approach. *Water Resources Management* **20** (1), 77–90.

Niu, D., Wang, Y. & Wu, D. D. 2010 Power load forecasting using support vector machine and ant colony optimization. *Expert Systems with Applications* **37** (3), 2531–2539.

Nourani, V. & Komasi, M. 2013 A geomorphology-based ANFIS model for multi-station modeling of rainfall–runoff process. *Journal of Hydrology* **490**, 41–55.

Nourani, V., Monadjemi, P. & Singh, V. P. 2007 Liquid analog model for laboratory simulation of rainfall–runoff process. *Journal of Hydrologic Engineering* **12** (3), 246–255.

Nourani, V., Mogaddam, A. A. & Nadiri, A. O. 2008 An ANN-based model for spatiotemporal groundwater level forecasting. *Hydrological Processes* **22** (26), 5054–5066.

Park, E. & Parker, J. 2008 A simple model for water table fluctuations in response to precipitation. *Journal of Hydrology* **356** (3), 344–349.

Peitgen, H.-O., Jürgens, H. & Saupe, D. 2006 *Chaos and Fractals: New Frontiers of Science*. Springer Science & Business Media, Berlin.

Procaccia, I., Grassberger, P. & Hentschel, H. G. E. 1983 On the characterization of chaotic motions. In: *Dynamical System and Chaos* (L. Garrido, ed.). Springer, Berlin, pp. 212–222.

Pulido-Calvo, I. & Gutierrez-Estrada, J. C. 2009 Improved irrigation water demand forecasting using a soft-computing hybrid model. *Biosystems Engineering* **102** (2), 202–218.

Raghavendra, N. S. & Deka, P. C. 2016 Multistep ahead groundwater level time-series forecasting using Gaussian process regression and ANFIS. In: *Advanced Computing and Systems for Security* (R. Chaki, A. Cortesi, K. Saeed & N. Chaki, eds). Springer, India, pp. 289–302.

Rosenstein, M. T., Collins, J. J. & De Luca, C. J. 1993 A practical method for calculating largest Lyapunov exponents from small data sets. *Physica D: Nonlinear Phenomena* **65** (1), 117–134.

Sahoo, S. & Jha, M. K. 2013 Groundwater-level prediction using multiple linear regression and artificial neural network techniques: a comparative assessment. *Hydrogeology Journal* **21** (8), 1865–1887.

Salas, J., Kim, H., Eykholt, R., Burlando, P. & Green, T. 2005 Aggregation and sampling in deterministic chaos: implications for chaos identification in hydrological processes. *Nonlinear Processes in Geophysics* **12** (4), 557–567.

Sauer, T., Yorke, J. A. & Casdagli, M. 1991 Embedology. *Journal of Statistical Physics* **65** (3–4), 579–616.

Schmidt, J. & Dikau, R. 2004 Modeling historical climate variability and slope stability. *Geomorphology* **60** (3), 433–447.

Sivakumar, B., Berndtsson, R., Olsson, J. & Jinno, K. 2001 Evidence of chaos in the rainfall-runoff process. *Hydrological Sciences Journal* **46** (1), 131–145.

Sivakumar, B., Jayawardena, A. & Fernando, T. 2002 River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches. *Journal of Hydrology* **265** (1), 225–245.

Sreekanth, P., Sreedevi, P., Ahmed, S. & Geethanjali, N. 2011 Comparison of FFNN and ANFIS models for estimating groundwater level. *Environmental Earth Sciences* **62** (6), 1301–1310.

Takens, F. 1981 Detecting strange attractors in turbulence. In: *Dynamical Systems and Turbulence, Warwick 1980*. Springer, Berlin, pp. 366–381.

Theiler, J. 1986 Spurious dimension from correlation algorithms applied to limited time-series data. *Physical Review A* **34** (3), 2427.

Trichakis, I. C., Nikolos, I. K. & Karatzas, G. P. 2009 Optimal selection of artificial neural network parameters for the prediction of a karstic aquifer's response. *Hydrological Processes* **23** (20), 2956–2969.

Trichakis, I. C., Nikolos, I. K. & Karatzas, G. 2011 Artificial neural network (ANN) based modeling for karstic groundwater level simulation. *Water Resources Management* **25** (4), 1143–1152.

Tripathi, S., Srinivas, V. & Nanjundiah, R. S. 2006 Downscaling of precipitation for climate change scenarios: a support vector machine approach. *Journal of Hydrology* **330** (3), 621–640.

Tsanis, I. K., Coulibaly, P. & Daliakopoulos, I. N. 2008 Improving groundwater level forecasting with a feedforward neural network and linearly regressed projected precipitation. *Journal of Hydroinformatics* **10** (4), 317–330.

Wang, J. & Shi, Q. 2013 Short-term traffic speed forecasting hybrid model based on chaos–wavelet analysis-support vector machine theory. *Transportation Research Part C: Emerging Technologies* **27**, 219–232. doi:10.1016/j.trc.2012.08.004.

Wang, W.-C., Cheng, C.-T., Chau, K.-W. & Xu, D.-M. 2012 Calibration of xinanjiang model parameters using hybrid genetic algorithm based fuzzy optimal model. *Journal of Hydroinformatics* **14** (3), 784–799.

Wang, W.-C., Xu, D.-M., Chau, K.-W. & Chen, S. 2013 Improved annual rainfall-runoff forecasting using PSO–SVM model based on EEMD. *Journal of Hydroinformatics* **15** (4), 1377–1390.

Welch, P. D. 1967 The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics* **15** (2), 70–73.

Wolf, A., Swift, J. B., Swinney, H. L. & Vastano, J. A. 1985 Determining Lyapunov exponents from a time series. *Physica D: Nonlinear Phenomena* **16** (3), 285–317.

Wong, H., Ip, W.-C., Zhang, R. & Xia, J. 2007 Non-parametric time series models for hydrological forecasting. *Journal of Hydrology* **332** (3), 337–347.

Yang, Z., Lu, W., Long, Y. & Li, P. 2009 Application and comparison of two prediction models for groundwater levels: a case study in Western Jilin Province, China. *Journal of Arid Environments* **73** (4), 487–492.

Yao, X., Tham, L. & Dai, F. 2008 Landslide susceptibility mapping based on support vector machine: a case study on natural slopes of Hong Kong, China. *Geomorphology* **101** (4), 572–582.

Yeragani, V. K., Pohl, R., Berger, R., Balon, R., Ramesh, C., Glitz, D., Srinivasan, K. & Weinberg, P. 1993 Decreased heart rate variability in panic disorder patients: a study of power-spectral analysis of heart rate. *Psychiatry Research* **46** (1), 89–103.

Zhang, Y. & Wu, L. 2009 Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. *Expert Systems with Applications* **36** (5), 8849–8854.

Zhang, Y.-K. & Yang, X. 2010 Effects of variations of river stage and hydraulic conductivity on temporal scaling of groundwater levels: numerical simulations. *Stochastic Environmental Research and Risk Assessment* **24** (7), 1043–1052.

Zhang, M., Dong, Y. & Sun, P. 2012 Impact of reservoir impoundment-caused groundwater level changes on regional slope stability: a case study in the Loess plateau of Western China. *Environmental Earth Sciences* **66** (6), 1715–1725.

Zhou, C. & Yin, K. 2014 Landslide displacement prediction of WA-SVM coupling model based on chaotic sequence. *Electr. J. Geol. Eng.* **19**, 2973–2987.

First received 3 September 2016; accepted in revised form 9 February 2017. Available online 10 April 2017