Simulation and uncertainty analysis for groundwater levels using radial basis function neural network and support vector machine models

Siyu Nie, Jianmin Bian, Hanli Wan, Xiaoqing Sun and Bingjing Zhang

ABSTRACT

Accurate measurement of groundwater levels is often difficult and involves great uncertainty. Therefore, simulating and predicting the fluctuating behavior of groundwater levels is necessary for water resource planning and management. In this study, radial basis function (RBF) neural networks and support vector machines (SVM) were employed to simulate groundwater level fluctuations. The time series data of precipitation, evaporation, and temperature were used as model inputs. Groundwater level data from the first 10 years, from 2003 to 2014, were used as the training dataset, while data from the last 2 years were used as the test dataset. Uncertainties caused by errors in the measurements of the variables or in outputs were estimated at 95% confidence intervals. The results showed that the SVM model had a superior simulation and prediction capability according to four statistical standards. The comparisons of the outputs and the confidence intervals of the two models showed that the SVM model was more accurate and had less uncertainty. The conclusions suggest that SVM is an effective method for simulating groundwater levels and analyzing model uncertainties using confidence intervals and can be used to facilitate sustainable groundwater management strategies.

Key words | groundwater level, radial basis function neural networks, support vector machines, uncertainty analysis

Siyu Nie
Jianmin Bian (corresponding author)
Hanli Wan
Xiaoqing Sun
Bingjing Zhang
College of Environment and Resources,
Jilin University,
Changchun 130021,
China
and
Key Laboratory of Groundwater Resources and
   Environment, Ministry of Education,
Jilin University,
Changchun 130021,
China
E-mail: *bianjianmin@126.com*

INTRODUCTION

Groundwater is an important component of the global freshwater supply and a precious natural resource for agricultural, domestic, and industrial purposes in many countries. Water shortages, the over-exploitation of groundwater, and related environmental and geological problems have attracted increasing attention and become one of the most critical global concerns, especially in arid, semi-arid, and the fragile ecological environments (Adamowski & Chan 2011). Da'an, a semi-arid region with the highest salinization rate in western Jilin Province of China, is located in an ecologically and economically fragile area. The variation in groundwater levels is the main indicator of the amount of groundwater resources, and the somewhat unstable changes in groundwater levels are the result of changes in many

complex and interactive factors. Therefore, an accurate and reliable prediction of groundwater levels is essential in determining the resource quantity and allowable exploitation level of groundwater and in avoiding or reducing adverse effects such as the loss of pumpage in water wells, land surface subsidence, and aquifer compaction (Vahid et al. 2013; Verma & Singh 2013).

Mathematical models are generally used to improve our understanding of groundwater systems. There are many prediction models, such as nonlinear empirical models, mathematical groundwater models, and physically-based models, that have been used to simulate and forecast groundwater levels and applied to problems ranging from aquifer safe yield analysis to groundwater remediation and quality

issues (Sun & Xu 2011; Emamgholizadeh *et al.* 2014). Although conceptual and physically-based models are used to depict hydrological variables and characterize the complex structures of aquifers, they have practical limitations (Nourani *et al.* 2008). These modeling techniques are very data- and labor-intensive, such as Darcy's law-based differential equation systems of groundwater dynamics (Bense *et al.* 2009; Ge *et al.* 2011). Since data are typically limited in regions under severe environmental conditions, it is difficult to analyze the geological parameters and predict the results accurately in those regions. Therefore, empirical models, such as artificial neural networks (ANN) and support vector machines (SVM), may serve as attractive alternatives, because they can provide useful results using a smaller amount of data, are less labor-intensive, more cost-effective, and suited to solve the dynamic nonlinear systems (Emamgholizadeh *et al.* 2014; Chang *et al.* 2015; Gong *et al.* 2016).

In various branches of hydrology, ANNs have been well-developed and applied for the prediction of nonlinear problems, such as precipitation (Nastos *et al.* 2014), sediment load (Afan *et al.* 2015), and river flow (He *et al.* 2014). The radial basis function (RBF) neural network is one of the ANN models and has superior performance to the back-propagation ANN model and has a fast impending speed. The applications of the RBF technique in hydrology range from real-time modeling to event-based modeling. It has been used for the prediction of rainfall and groundwater levels as well as for the modeling of stream flows and water quality (Garcia & Shigdi 2006; Ghose Dillip *et al.* 2010). SVM is a relatively new structure in modeling nonlinear systems. It is based on structural risk minimization (SRM) instead of the empirical risk minimization of ANN. SRM minimizes the empirical error and model complexity simultaneously, which can improve the generalization ability of SVM for classification or regression problems in many disciplines. SVM has been used to solve hydrogeological problems, such as estimating evapotranspiration in a semi-arid environment (Tabari *et al.* 2012), and predicting groundwater levels in a coastal aquifer (Yoon *et al.* 2011) and stream flow (Noori *et al.* 2011). These studies showed that two data-driven models could be applied in formal hydrology studies, and models could be improved or combined with other models for higher accuracy in results. However, the results of the numerical models are subject to randomness and

uncertainty whether the models are combined or not, which makes it difficult to calculate the groundwater levels accurately. Thus far, very little research has been conducted on the analysis of the correlation between the results of numerical models with their uncertainties.

Using the extensive field monitoring data collected from 2003 to 2014 in Da'an, in western Jilin Province of China, this study aims to construct a groundwater level model by using RBF and SVM frameworks, examine the validity of the model, and compare the results of two frameworks. We investigate and analyze the impacts of uncertainty on the simulated results at a 95% confidence interval. The results provide an important theoretical basis for improving the accuracy of groundwater level simulations and predictions, and thus serve as a reference for sustainable exploitation, utilization, and protection of groundwater resources.

METHODS

RBF

RBF is a kind of centrosymmetric nonnegative and nonlinear function. It has a strong biological background and the ability to approximate arbitrary nonlinear function (Schilling *et al.* 2001), and it also possesses the advantages of the optimal approximation point. In 1985, multivariate interpolation of the RBF method was proposed by Powell. In 1988, RBF neural network was applied to the design of ANN and this method was successfully applied to identify the nonlinear time series prediction field. Basically, a RBF network is composed of a large number of simple and highly interconnected artificial neurons and can be organized into several layers, i.e. input layer, hidden layer, and output layer as shown in Figure 1.

Input layer: An input pattern enters the input layer and is subjected to direct transfer function. The input layer serves as a distributor to the hidden layer and output from the input layer is also subjected to transfer function. The number of nodes in the input layer is equal to the dimension of input vector L. The output from the input layer with element $I_{i(i=1 \text{ to } L)}$ is I_i.

Hidden layer: The hidden layer does all the important processes and these nodes satisfy a unique property, being
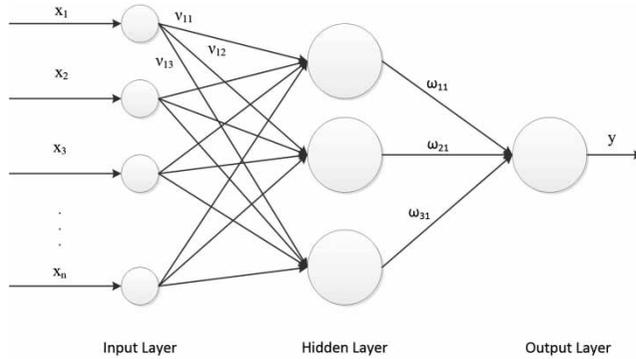
Figure 1 | Architecture of RBFN.

of radially symmetric structure. Being a radially symmetric structure, it must have the following:

(a) A center vector v_j in the input space, made up of a cluster center with the element v_{ji} (j=1 to M). 'M $\leq$ P', where M is the number of center vectors and P is the number of training patterns. The vector typically is stored as weight factors from the input layer to the hidden layer.

(b) A distance measured to determine how far an input pattern with element I_i is from the cluster center v_{ji}. We have used Euclidean distance norm for this purpose:

$$\text{Euclidean distance } ed_j = ||I - v_j||$$
$$= \sqrt{\sum_{i=1}^{L} (I_i - v_{ji})^2} \qquad (1)$$

(c) A transfer function which transfers Euclidean distance to give output for each node. In our case we used the Gaussian function for this purpose:

$$\text{Output}_j = \exp\left(\frac{ed_j^2}{\sigma^2}\right) \qquad (2)$$

where σ is the spread parameter determined from:

$$\sigma = \frac{\max(ed)}{\sqrt{M}} \qquad (3)$$

and $\max(ed)$ is the maximum Euclidean distance between selected centers and M is the number of centers.

Output layer: There are weight factor ω_{kj} (k=1 to N, j=1 to M) between kth nodes of the output layer and jth nodes of the

hidden layer. 'N' is the dimension of the output vector. Output from the output layer transferred through a transfer function like log sigmoid or tan sigmoid (Ghose Dillip *et al.* 2010).

Output from the output layer is given by:

$$\text{Output}_k = f\left(\sum_{k=1}^{N} \omega_{kj} \times \text{Output}_j\right) \qquad (4)$$

SVM

SVM is a relatively new machine-learning approach in data-driven research fields based on statistical learning theory (Vapnik 1995, 1998). The process of an SVM estimator (f) in regression can be expressed as follows:

$$f(x) = \omega \cdot \phi(x) + b \qquad (5)$$

where ω is a weight vector and b is a bias. ϕ denotes a non-linear transfer function that maps the input vectors into a high-dimensional feature space in which theoretically a simple linear regression can cope with the complex non-linear regression of the input space. Vapnik (1995) introduced the following convex optimization problem with an ε-insensitivity loss function to obtain the solution:

$$\underset{\omega,b,\xi,\xi^*}{\text{minimize}} \quad \frac{1}{2}||\omega||^2 + C\sum_{k=1}^{N}(\xi_k + \xi_k^*)$$

$$\text{Subject to } \begin{cases} y_k - \omega^T\varphi(x_k) - b \leq \varepsilon + \xi_k \\ \omega^T\varphi(x_k) + b - y_k \leq \varepsilon + \xi_k^* \\ \xi_k, \xi_k^* \geq 0 \end{cases} \qquad (6)$$
$$k = 1, 2, \ldots, N$$

where ξ_k and ξ_k^* are slack variables that penalize training errors by the loss function over the error tolerance ε, and C is a positive tradeoff parameter that determines the degree of the empirical error in the optimization problem. Equation (6) is usually solved in a dual form using Lagrangian multipliers and imposing the Karush–Kuhn–Tucker (KKT) optimality condition. The input vectors that have non-zero Lagrangian multipliers under the KKT condition support the structure of the estimator and are called support

vectors (Gong *et al.* 2016). The architecture of SVM is shown in Figure 2.

Study area description and data collection

Da'an (123 °08′45″ to 124 °21′56″E, 44 °57 ′00″ to 45 °45′ 51″N) is located in the northwest of Jilin Province, in eastern China. It covers a total area of about 4,924 km², is in an ecologically fragile local environment, and belongs to the Songnen plain hinterland (Figure 3). Da'an experiences one of the most serious soil salinization problems in the western Jilin Province. The severe saline-alkali land area accounts for 60.3% of the total saline-alkali land area. Da'an features semi-arid climatic conditions with dry and windy weather in spring, rainy weather in summer, light precipitation in autumn and moderate snow in winter. The annual average temperature is 4.8 °C, the annual average precipitation is 422 mm, and the average annual evapotranspiration is 1,681 mm. The main types of groundwater aquifers in the region are phreatic and confined aquifers. Groundwater level is shallow and subsurface runoff is slow. With the recent developments in agriculture and urbanization, increased groundwater extraction has altered the natural dynamic equilibrium of groundwater and left the water resources issues unresolved.

In the study area, the locations of the wells were determined using a GARMIN handheld Global Positioning System and are shown in Figure 3. The monthly groundwater level data were collected by recording the levels in a manual drilling well. The monthly precipitation, air temperature, and evaporation data were downloaded from the Da'an hydrological station in Baicheng County for the period from January 2003 to December 2014. In view of the reliability and completeness of the data source, for groundwater level, evaporation, average temperature, and rainfall has a slight correlation. Therefore, temperature, evaporation, and precipitation are chosen as the models input, and the groundwater levels as output. Of the 12 years of observed groundwater level data (2003–2014), the first 10 years were used as the training dataset and the last 2 years were used as the test dataset. The time-series data were normalized by Equation (7) to eliminate the dimensional differences between different influence factors, and the variables in the training dataset were scaled to a limit

between 0 and 1.

$$Y = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{7}$$

where Y is the normalized data, X is the time-series data, $X_{\min}$ is the minimum value of the time-series data and $X_{\max}$ is maximum values of time-series data (Yoon *et al.* 2011).

Performance criteria

The performances of the models developed in this study were assessed using four standard statistical parameters, including the coefficient of correlation (R), root mean squared error (RMSE), mean absolute error (MAE) and Nash–Sutcliffe efficiency coefficient (NS). Coefficient of correlation (R) measures the degree to which two variables are linearly related. RMSE and MAE provide different types of information about the predictive capabilities of the model. The Nash–Sutcliffe efficiency coefficient (NS) evaluates the reliability of model results (Chang *et al.* 2015). The following equations were used for the computation of these parameters:

$$R = \frac{\sum_{i=1}^{n}(O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2 \sum_{i=1}^{n}(P_i - \bar{P})^2}} \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{9}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - O_i| \tag{10}$$

$$NS = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \tag{11}$$

where n is the number of input samples, O_i and P_i are the observed and predicted groundwater level depths at time t, and $\bar{O}$ and $\bar{P}$ are the means of the observed and predicted groundwater level values, respectively. The best fit between the observed and predicted values would have R = 1, RMSE = 0, MAE = 0 and NS = 1.
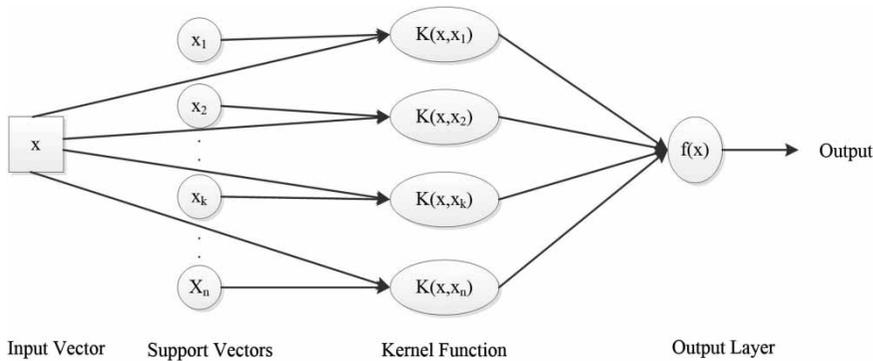
Figure 2 │ Architecture of SVM.

RESULTS AND DISCUSSION

The RBF modeling

The RBF models for monthly groundwater levels simulation from observation wells are developed using the Matlab R2011software program. In the RBF model, the variables temperature, precipitation and evaporation are used as the input data to simulate and predict the groundwater level. To select the best one in number of neurons in the hidden layer, a trial and error method is made and the optimal numbers of hidden neurons are determined to be 8. During the
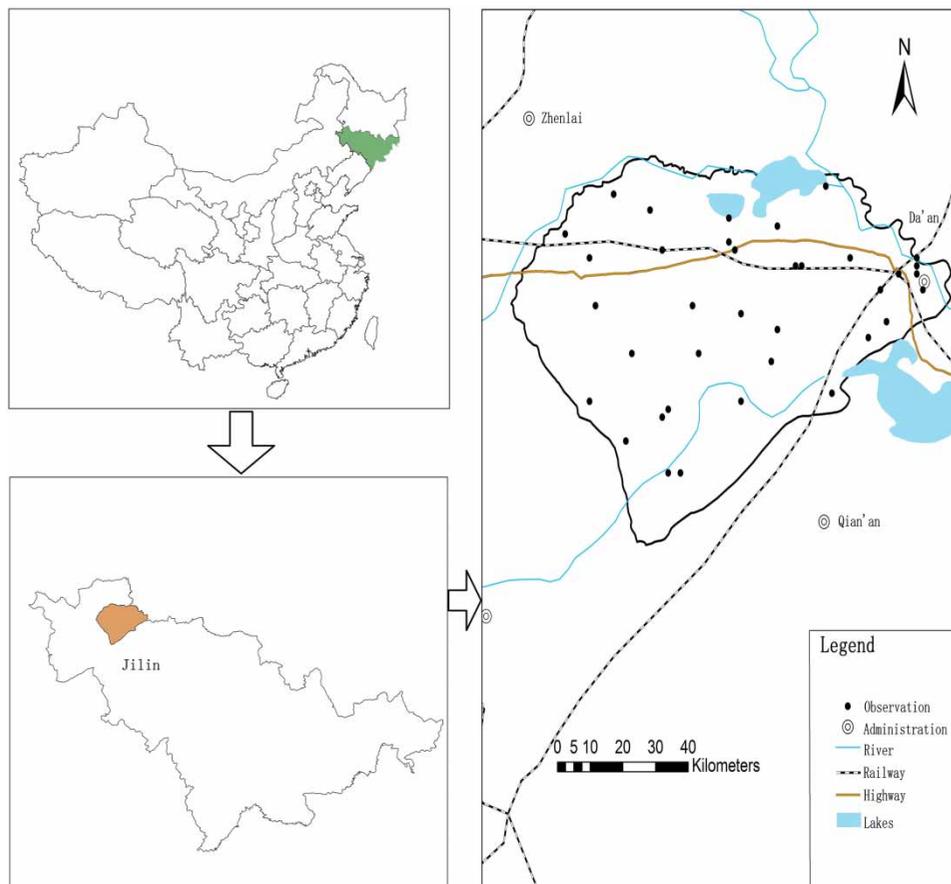
Figure 3 │ Location map of study area and observation wells.

training period, the RBF models are used to compute the monthly groundwater level for observation wells. Figure 4 shows the comparisons of observed and simulated groundwater level values using RBF model for observation wells.

Figure 4 shows that the values simulated by the RBF model reasonably match the observed groundwater levels in the training period. The correlation coefficient R^2 between the RBF models simulated value and observed data was 0.8483, which indicates that the RBF models had good fitting accuracy in the training period.

The SVM modeling

The same input parameters and driven factors are introduced to the SVM model. Based on the theory of SVM, the RBF kernel function (Huang & Wang 2006) is presented and the SVM model is set up by using the Matlab R2011. The performance of SVM model for simulation the groundwater level in study area is shown in Figure 5.

Figure 5 shows that the correlation coefficient R^2 between the SVM model simulated values and observed data was 0.9307 during the training period. Compared
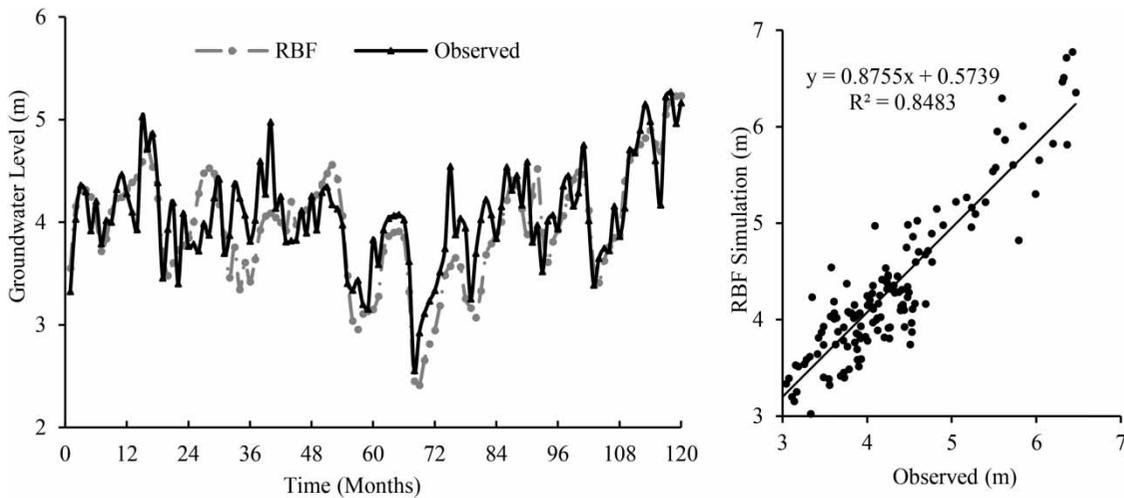

Figure 4 | The observed and simulated groundwater level by the RBF model during the training period.
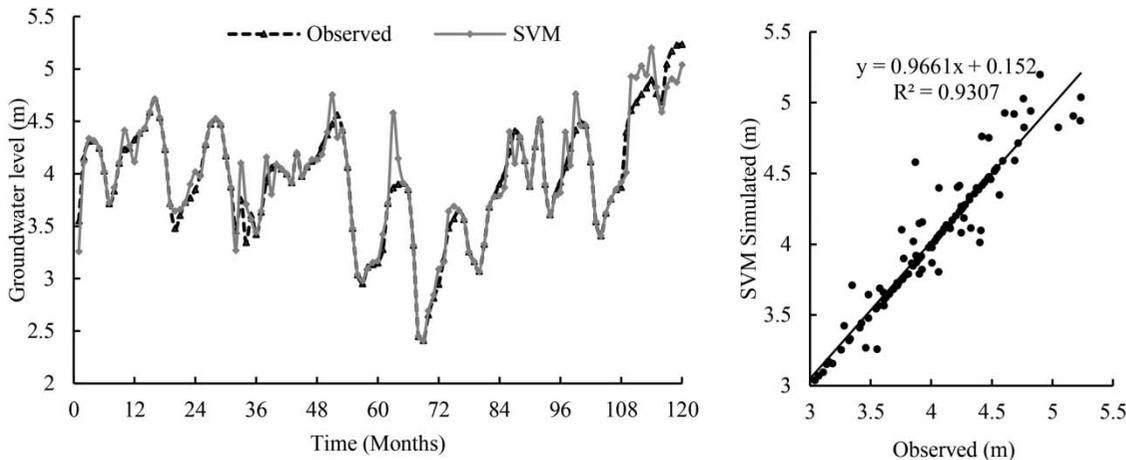

Figure 5 | The observed and simulated groundwater level by the SVM model during the training period.

with the results of RBF, the SVM model had better fitting accuracy in the training period. Thus, the two models can be used to simulate and predict monthly groundwater levels.

Comparison of RBF and SVM models

The performance of the RBF model and SVM model during the training period and validation is summarized in Table 1 in terms of R, RMSE, MAE and NS.

Table 1 and Figure 6 show that in the training stage, the RMSE values for RBF and SVM models are 0.315 and 0.148, respectively. The MAE values for the two models were 0.257 and 0.082, respectively. The mean RMSE values of SVM

Table 1 │ Result of modeling from RBF and SVM models in different period

	RBF		SVM	
	Training period	Validation period	Training period	Validation period
R	0.839	0.905	0.964	0.919
RMSE	0.315	0.413	0.148	0.297
MAE	0.257	0.336	0.082	0.208
NS	0.679	0.676	0.928	0.849

were smaller than those of the RBF model. In the validation stage, the mean RMSE values in RBF and SVM were 0.413 and 0.297; the mean MAE values were 0.336 and 0.208, respectively. The prediction results of SVM were more accurate, which implies that the simulating and predicting capability of the SVM model is better than that of the RBF model for the given data.

In addition, if the NS and R criteria in a model are equal to 1, then that model is capable of producing a perfect estimation. In general, a model can be considered accurate and effective if the NS is higher than 0.8 (Shu & Ouarda 2008). The R values of the two models were over 0.8 and the NS values for the SVM model in the training stage were greater than 0.8 (Table 1). These values suggest that the SVM model achieved acceptable results, but RBF did not; and the SVM model is more capable of capturing the nonlinear relationships with the input data than the RBF model.

Uncertainty analysis

One of the digital drive model hypotheses includes the valuable information (change rule) that understands the changes of the input and output data with the time change and the trend of changes is recorded and simulating by using model.
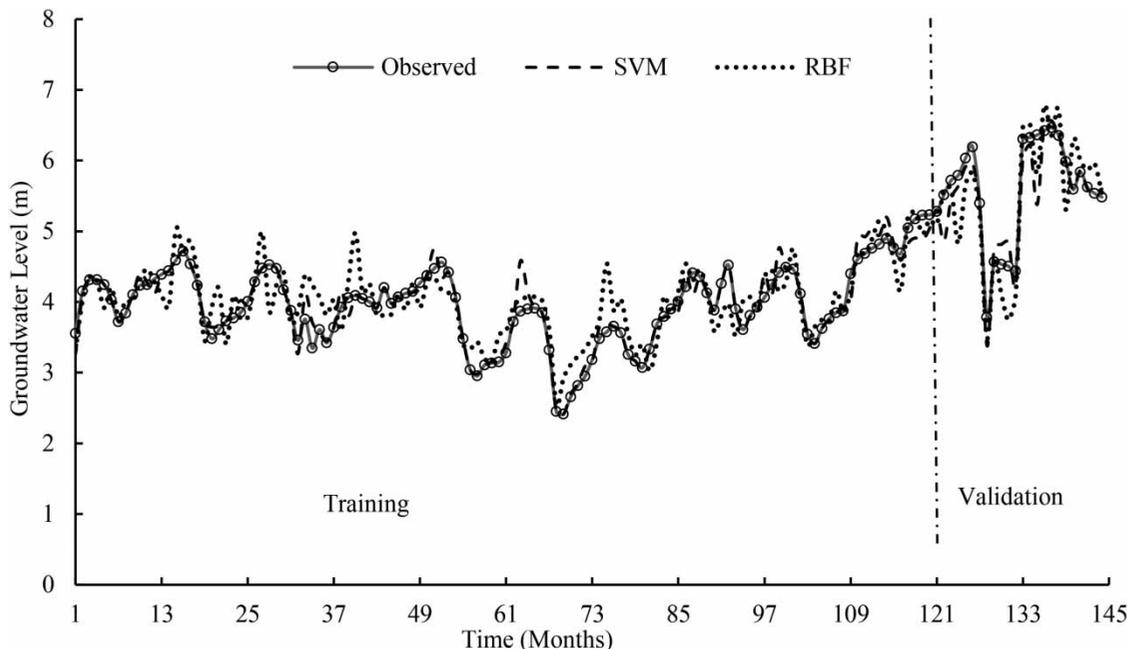

Figure 6 │ SVM and RBF model simulations and predictions versus observed data.

However, the data source and model parameters bring uncertainty problems. Therefore, it is necessary to perform uncertainty analysis for evaluating. In order to compare and measure the uncertainty related to the results of RBF and SVM models, it is necessary to apply objective criteria. Therefore, in this study, we used the d-factor (Talebizadeh & Moridnejad 2011), where the greater the d-factor, the more uncertainty the model has. Calculation of the d-factor can be achieved according to the following:

$$\bar{d}_I = \frac{1}{n} \sum_{i=1}^{n} (I_{Uj} - I_{Lj}) \qquad (12)$$

$$d - \text{factor} = \frac{\bar{d}_I}{\sigma_I} \qquad (13)$$

where $\bar{d}_I$ is the mean distance between the lower I_L and the upper limits I_U of the 95% confidence interval; σ_I is the standard deviation of observed data.

According to the statistical principle (Yang & Wen 2012), the upper and lower limit values of simulation and prediction of RBF and SVM models are calculated with the 95% confidence level. The results are shown in Figures 7 and 8.

According to Equations (12) and (13), the values of the d-factor for the SVM and RBF models were 0.91 and 2.16, respectively, which indicated that the overall uncertainty in SVM model results was lower than that in RBF in this case study. Figures 7 and 8 show the relationship between observed groundwater levels and the predicted values within 95% confidence interval for two models. The 95% confidence interval for RBF predictions was much wider than the interval for SVM predictions. The lower the model uncertainty is, the narrower the confidence interval is, and the more reliable the predicted results are. In addition, the majority of observed groundwater levels fell within the confidence interval, which shows that the confidence level in simulation results reached 95% (Figures 7 and 8).

CONCLUSIONS

The accurate and reliable simulation and prediction of groundwater levels is one of the most important issues in water resources management. In this study, monthly groundwater data were used to assess the ability of SVM and RBF models to simulate and predict groundwater levels in Da'an, in Jilin Province of China. Hydrological variables were used as model inputs and monthly groundwater levels were used as the model output. Four standard statistical criteria, R, MAE, RMSE, and NS, were used for evaluating the performance of these two models.

The overall results showed that RBF and SVM models provided a good fit to the observed data. However, the values of four standard statistical parameters indicated that the SVM model was more reliable in simulating and predicting the groundwater levels compared to the RBF model
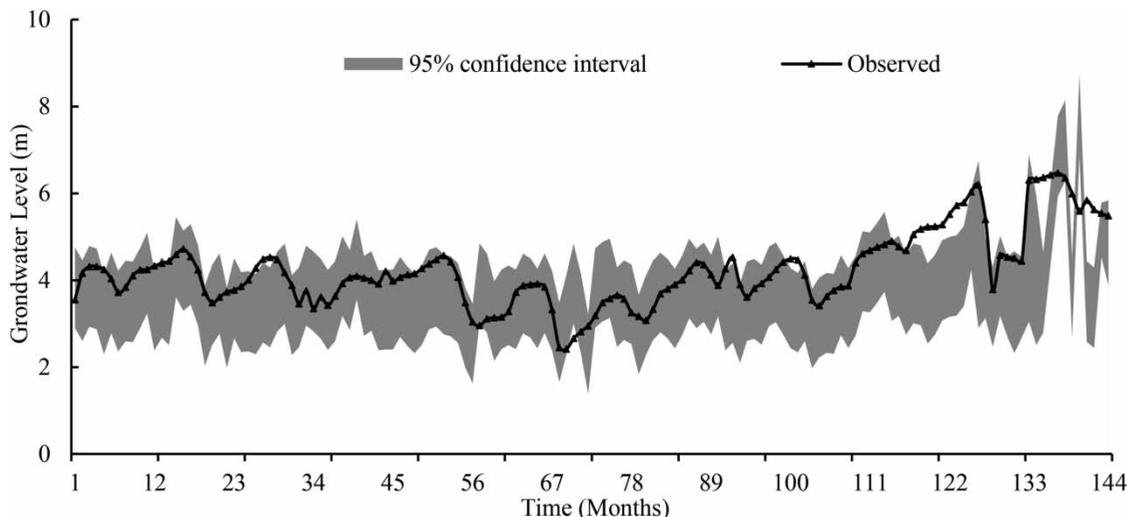

Figure 7 | Predicted and observed groundwater levels for RBF with confidence intervals.
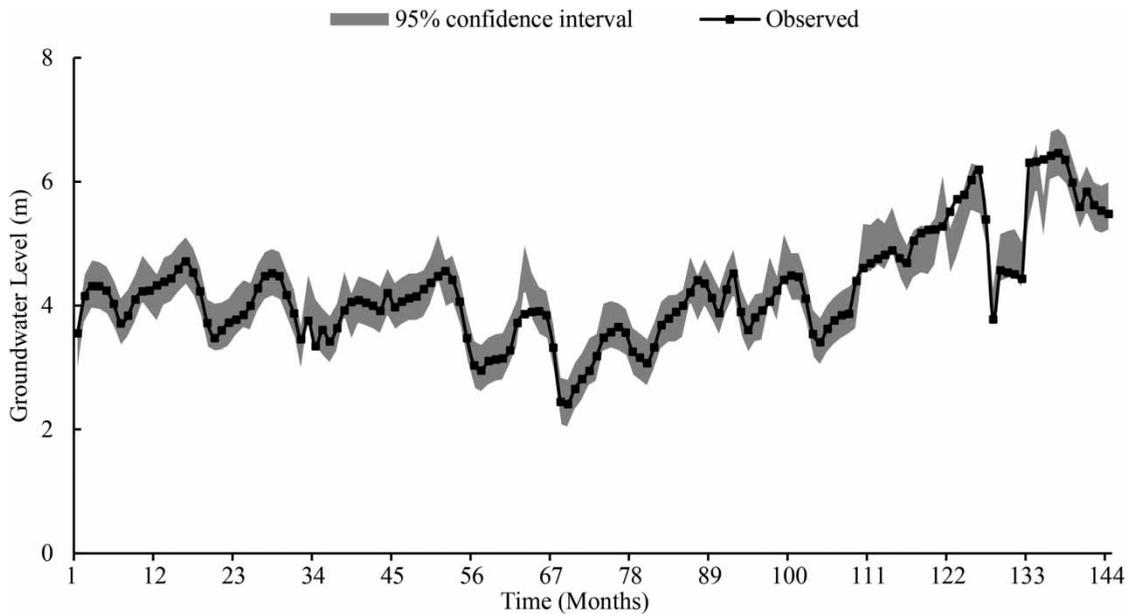
Figure 8 │ Predicted and observed groundwater levels for SVM with confidence intervals.

during the training and validation steps. Another advantage of the SVM model over RBF, based on the objective criterion (d-factor), was the lower uncertainty (narrower confidence interval) in the results. Thus, the SVM model is considered an effective method for predicting the groundwater levels.

The uncertainty quantification is an important aspect of model predictions. Based on the results of deterministic simulation and prediction, the 95% confidence interval is proposed to calculate model uncertainty and predict the results in a probabilistic sense. Additional studies should be conducted to further explore this proposed method, which can improve the accuracy of the predictions under varied environmental conditions and facilitate the development of more effective and sustainable groundwater management strategies.

REFERENCES

Adamowski, J. & Chan, H. F. 2011 A wavelet neural network conjunction model for groundwater level forecasting. *J. Hydrol.* **407**, 28–40.

Afan, H. A., El-Shafie, A., Yaseen, Z. M., Hameed, M. M., Wan Mohtar, W. H. M. & Hussain, A. 2015 ANN based sediment prediction model utilizing different input scenarios. *Water Resour. Manage.* **29**, 1231–1245.

Bense, V. F., Ferguson, G. & Kooi, H. 2009 Evolution of shallow groundwater flow systems in areas of degrading permafrost. *Geophys. Res. Lett.* **36**, L23401. http://dx.doi.org/10.1029/2009GL039225.

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

Emamgholizadeh, S., Moslemi, K. & Karami, G. 2014 Prediction of the groundwater level of bastam plain (Iran) by artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS). *Water Resour. Manage.* **28**, 5433–5446.

Garcia, L. A. & Shigdi, A. 2006 Using neural networks for parameter estimation in ground water. *J. Hydrol.* **318** (1–4), 215–231.

Ge, S., McKenzie, J., Voss, C. & Wu, Q. 2011 Exchange of groundwater and surface water mediated by permafrost

response to seasonal and long term air temperature variation. *Geophys. Res. Lett.* **38**, L14402.

Ghose Dillip, K., Panda Sudhansu, S. & Swain Prakash, C. 2010 Prediction of water table depth in western region, Orissa using BPNN and RBFN neural networks. *J. Hydrol.* **394**, 296–304.

Gong, Y. C., Zhang, Y. X., Lan, S. 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 Resour. Manage.* **30**, 375–391.

He, Z., Wen, X., Liu, H. & Du, J. 2014 A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *J. Hydrol.* **509**, 379–386.

Huang, C. L. & Wang, C. G. 2006 A GA-based feature selection and parameters optimization for support vector machines. *Expert Syst. Appl.* **31**, 231–240.

Nastos, P. T., Paliatsos, A. G., Koukouletsos, K. V., Larissi, I. K. & Moustris, K. P. 2014 Artificial neural networks modeling for forecasting the maximum daily total precipitation at Athens, Greece. *Atmos. Res.* **144**, 141–150.

Noori, R., Karbassi, A. R., Moghaddamnia, A., Han, D., Zokaei-Ashtiani, M. H., Farokhnia, A. & Gousheh, M. G. 2011 Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction. *J. Hydrol.* **401**, 177–189.

Nourani, V., Moghaddam, A. A., Nadiri, A. O. & Singh, V. P. 2008 Forecasting spatiotemporal water levels of Tabriz aquifer. *Trends Appl. Sci. Res.* **3**, 319–329.

Schilling, R. J., Carroll, J. J. & Al-Ajlouni, A. F. 2001 Approximation of nonlinear systems with radial basis function neural networks. *IEEE Trans. Neural Networks* **12** (1), 21–28.

Shu, C. & Ouarda, T. B. M. J. 2008 Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. *J. Hydrol.* **349**, 31–43.

Sun, S. & Xu, X. 2011 Variational inference for infinite mixtures of Gaussian processes with applications to traffic flow prediction. *IEEE Trans. Intell. Transp. Syst.* **12**, 466–475.

Tabari, H., Kisi, O., Ezani, A. & Hosseinzadeh Talaee, P. 2012 SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. *J. Hydrol.* **445**, 78–89.

Talebizadeh, M. & Moridnejad, A. 2011 Uncertainty analysis for the forecast of lake level fluctuations using ensembles of ANN and ANFIS models. *Expert Syst. Appl.* **38** (4), 4126–4135.

Vahid, M., Mehdi, V., Bagher, S. & Negin, B. 2013 A wavelet-ANFIS hybrid model for groundwater level forecasting for different prediction periods. *Water Resour. Manage.* **27**, 1301–1321.

Vapnik, V. N. 1995 *The Nature of Statistical Learning Theory.* Springer-Verlag, New York, p. 314.

Vapnik, V. N. 1998 *Statistical Learning Theory.* Wiley, New York, p. 736.

Verma, A. K. & Singh, T. N. 2013 Prediction of water quality from simple field parameters. *Environ. Earth Sci.* **69**, 821–829.

Yang, H. S. & Wen, Y. D. 2012 Forecasting of wind speed and estimation of confidence interval based on wavelet neural network. *J. Anhui Polytech. Uni.* **27** (3), 65–68.

Yoon, H., Jun, S. C., Hyun, Y., Bae, G. O. & Lee, K. K. 2011 A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J. Hydrol.* **396**, 128–138.