RVM based on PSO for Groundwater Level Forecasting

Weiguo Zhao, Yanfeng Gao
Hebei University of Engineering, Handan 056038, China
Email:zwg770123@163.com

Chunliu Li
College of Urban Construction, Hebei Normal University of Science&Technology, Qinhuangdao 066004, China

***Abstract*—Relevance Vector Machine (RVM) is a novel kernel method based on Sparse Bayesian, which has many advantages such as its kernel functions without the restriction of Mercer's conditions, the relevance vectors automatically determined. In this paper, a new RVM model optimized by Particle Swarm Optimization (PSO) is proposed, and it is applied to groundwater level forecasting. The simulation experiments demonstrate that the proposed method can reduce significantly both relative mean error and root mean squared error of predicted groundwater level. Moreover, the model achieved is much sparser than its counterpart, so the RVM based on PSO is applicable and performs well for groundwater data analysis.**

***Index Terms*—Relevance Vector Machine, Particle Swarm Optimization, Support Vector Machine, groundwater level forecasting**

I. INTRODUCTION

Water is a limited resource and essential for agriculture, industry and creatures existence on earth including human beings. Now, water resource scarcity has become a restricting factor for the economic and social development in the world. In the past decades there has been a tremendous pressure on this precious natural resource mainly due to rapid industrialization and human population, as an increase in the human population will simply result in increasing the demand for irrigation purpose to meet food production requirements [1].

Groundwater provides one-third of the world's drinking water for human beings, animals, and plants. Since surface water is largely allocated, demand on the finite groundwater resources is increasing. However, groundwater is highly susceptible to contamination, various human factors and factitious factors would result to declining of groundwater level and pollution of groundwater For example, groundwater can be contaminated by localized releases from waste disposal sites, landfills, and underground storage tanks. Pesticides, fertilizers, salt water intrusion, and contaminants from other source pollutants are also major sources of groundwater pollution [2]. So, we must understand the groundwater level and the change characteristic, the groundwater resources are reasonably used and protected, then it is important for us to research the change trend of groundwater level using the existing data and materials.

In this direction several studies were carried out for forecasting the groundwater levels using conceptual models that are not only laborious, but also have practical limitations, In the recent past, soft computing tools like Support Vector Machine （SVM）has been used increasingly in various fields of science and technology for prediction purposes. But SVM does exhibit some significant drawbacks. For example, it lacks of necessary probability information and is unable to obtain forecast in uncertainty; it need to entail a cross-validation procedure which is wasteful both of data and time; its kernel function must satisfy Mercer's conditions; the number of kernel functions will increase with the increase of training set; its output is a point estimate rather than the conditional distribution in order to capture uncertainty in the prediction, therefore its predictions are not probabilistic [3].

RVM is a probabilistic sparse kernel model identical in functional form to SVM, where a Bayesian approach to learning is adopted, introducing a prior over the weights governed by a set of hyperparameters, one associated with each weight whose most probable values are iteratively estimated from the data. One of the main advantages of RVM is its capability to obtain a generalization performance comparable to SVM but using dramatically fewer kernel functions. Furthermore, RVM suffers from none of the other limitations of SVM outlined above [4, 5]. In this paper, a novel RVM model based on PSO algorithm is proposed and is applied to groundwater level forecasting, compared with the SVM model, the simulation experiments show it has more accurate predicting results than its counterpart, moreover, it can achieve the greatest levels of sparsity.

II. RELEVANCE VECTOR MACHINE

Tipping [5, 6] proposed the Relevance Vector Machine (RVM) in 2000. Followed by Noslen Hernández's illustration [3], we show a main idea of the RVM. For a regression problem, given a training dataset $\{x_n, t_n\}_{n=1}^{N}$, the following generalized linear regression model can be used to describe the mapping relation between the input pattern vector $\mathbf{x}$ and the scalar target t:

$$t_n = y(x_n, \mathbf{w}) + \varepsilon_n \quad \mathbf{t} = \mathbf{y} + \boldsymbol{\varepsilon} \tag{1}$$

doi:10.4304/jcp.7.5.1073-1079

Where the errors $\boldsymbol{\varepsilon}=(\varepsilon_1,.....,\varepsilon_n)$ are modeled probabilistically as independent zero-mean Gaussian, with variance σ^2, so $p(\varepsilon)=\prod_{n=1}^{N}N(\varepsilon_n|0,\sigma^2)$, $\mathbf{w}=(w_1.....,w_M)$ is the parameter vector and $y(x_n,\mathbf{w})$ can be expressed as a linearly weighted sum of some basis functions $\phi(\mathbf{x})$:

$$y(\mathbf{x},\mathbf{w})=\sum_{m=1}^{M}w_m\phi_m(\mathbf{x})+w_0\ \mathbf{y}=\boldsymbol{\Phi}\mathbf{w} \tag{2}$$

Here $\boldsymbol{\Phi}=[\phi_1,\cdots,\phi_M]$ is the $N \times M$ design matrix whose columns comprise the complete set of M basis vectors. Note that the form of the function (2) is equal to the form of the function for a SVM, where we identify our general basis functions with the kernel as parameterized with the training vectors: $\phi_m(\mathbf{x})=K(\mathbf{x},x_m)$ and $\phi(x_n)=[1,K(x_1,x_n),.....K(x_N,x_n)]$. The error model assumed implied $p(t_n|x_n)=N(t_n|y(x_n),\sigma^2)$, where the notation specifies a Gaussian distribution over t_n with mean $y(x_n)$ and variance σ^2. Due to the assumption of independence of the t_n, the likelihood of the complete dataset can be written as

$$p(\mathbf{t}|\mathbf{w},\sigma^2)=(2\pi\sigma^2)^{-N/2}\exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{t}-\boldsymbol{\Phi}\mathbf{w}\|^2\right\} \tag{3}$$

Maximum-likelihood estimation of $\mathbf{w}$ and σ^2 from the above equation will generally lead to overfitting problem. Here, though, RVM adopts a Bayesian perspective, and constrains the parameter by defining an explicit prior probability distribution over them, encoding a preference for smoother (less complex) functions by making the popular choice of a zero-mean Gaussian prior distribution:

$$p(\mathbf{w}|\boldsymbol{\alpha})=\prod_{i=0}^{N}N(w_i|0,\alpha_i^{-1})$$
$$=(2\pi)^{-M/2}\prod_{m=1}^{M}\alpha_m^{1/2}\exp(-\alpha_m w_m^2/2) \tag{4}$$

Where $\boldsymbol{\alpha}=(\alpha_1.....,\alpha_M)$ is a vector of M hyperparameters. There is an individual hyperparameter associated independently with every weight, moderating the strength of the prior over its associated weight. It is this form of prior that is ultimately responsible for the sparseness properties of the model..

Given $\boldsymbol{\alpha}$, the posterior parameter distribution conditioned on the data is given by combining the likelihood and prior within Bayes's rules:

$$p(\mathbf{w}|\mathbf{t},\boldsymbol{\alpha},\sigma^2)=p(\mathbf{t}|\mathbf{w},\sigma^2)\,p(\mathbf{w}|\boldsymbol{\alpha})/\,p(\mathbf{t}|\boldsymbol{\alpha},\sigma^2) \tag{5}$$

is a Gaussian distribution $N(\mu,\Sigma)$ with

$$\mu=\sigma^{-2}\Sigma\boldsymbol{\Phi}^T\mathbf{t}$$
$$\Sigma=(A+\sigma^{-2}\boldsymbol{\Phi}^T|\boldsymbol{\Phi})^{-1} \tag{6}$$

and A is defined as diag$(\alpha_1.....,\alpha_M)$. Rather than extending the model to include Bayesian inference over these hyperparameters (which is analytically intractable), Sparse Bayesian learning can be formulated as a type–II maximum likelihood procedure; that is, a most probable point estimate α_{MP} may be found throughout the maximization of the marginal likelihood with respect to the hyperparameters a_i (the same can be done to estimate the hyperparameter σ):

$$L(\boldsymbol{\alpha})=\log p(\mathbf{t}|\boldsymbol{\alpha},\sigma^2)$$
$$=\log\int_{-\infty}^{\infty}p(\mathbf{t}|\mathbf{w},\sigma^2)p(\mathbf{w}|\boldsymbol{\alpha})\,d\mathbf{w}$$
$$=-\frac{1}{2}\left[N\log 2\pi+\log|\mathbf{C}|+\mathbf{t}^T\mathbf{C}^{-1}\mathbf{t}\right], \tag{7}$$

With $$\mathbf{C}=\sigma^{-2}I+\boldsymbol{\Phi}A^{-1}\boldsymbol{\Phi}^T \tag{8}$$

Once most probable values $\boldsymbol{\alpha}_{MP}$ (and σ_{MP}) have been found, a point estimate μ_{MP} for the parameters is the obtained by evaluating (6) with $\boldsymbol{\alpha}=\boldsymbol{\alpha}_{MP}$ and $\sigma=\sigma_{MP}$. The crucial information is that typically the optimal values of many hyperparameters are infinite. From (6) this leads to a parameter posterior infinitely peaked at zero for many weights w_m with the consequence that μ_{MP} correspondingly comprises very few non-zero elements. Those training vectors associated with the remaining non-zero weights are called Relevance Vectors, after the principle of automatic relevance determination.

Predictions are made based on the posterior distribution over the weights, conditioned to the maximized values $\boldsymbol{\alpha}_{MP}$ and σ_{MP}^2. The predictive distribution for a new datum $\mathbf{x}_*$ using (5) is defined as follows:

$$p(t_*|\mathbf{t},\boldsymbol{\alpha}_{MP},\sigma_{MP}^2)=$$
$$\int p(t_*|\mathbf{w},\sigma_{MP}^2)p(\mathbf{w}|\mathbf{t},\boldsymbol{\alpha}_{MP},\sigma_{MP}^2)\,d\mathbf{w} \tag{9}$$

which is easily computed due to the fact that both integrated terms are Gaussian, resulting in a Gaussian too

$$p(t_*|\mathbf{t},\boldsymbol{\alpha}_{MP},\sigma_{MP}^2)=N(t_*|y_*,\sigma_*^2)$$

with $y_*=\mu^T\phi(\mathbf{x}_*)\ \sigma_*^2=\sigma_{MP}^2+\phi(\mathbf{x}_*)^T\Sigma\phi(\mathbf{x}_*)$ (10)

It can be seen that the predictive mean is intuitively $y(\mathbf{x}_*,\mu)$ giving a final (posterior mean) approximation $\mathbf{y}=\boldsymbol{\Phi}\mu_{MP}$.

III. PARTICLE SWARM OPTIMIZATION

PSO is a method for performing numerical optimization without explicit knowledge of the gradient of the problem to be optimized [7, 8]. It is simplified and

first intended for simulating social behaviour. The features of the method are as follows [9]:

(1) The method is based on researches on swarms such as fish schooling and bird flocking.

(2) It is based on a simple concept. Therefore, the computation time is short and it requires few memories.

PSO is basically developed through simulation of bird flocking in two-dimension space. The position of each individual (agent) is represented by *XY* axis position and also the velocity is expressed by *vx* (the velocity of *X* axis) and *vy* (the velocity of *Y* axis). Modification of the agent position is realized by the position and velocity information.

An optimization technique based on the above concept can be described as follows: namely, bird flocking optimizes a certain objective function. Each agent knows its best value so far (*pbest*) and its *XY* position. Moreover, each agent knows the best value so far in the group (*gbest*) among *pbests*. Each agent tries to modify its position using the following information: the current positions (*x*,*y*), the current velocities (*vx*,*vy*), the distance between the current position, *pbest* and *gbest*.

This modification can be represented by the concept of velocity. Velocity of each agent can be modified by the following equation:

$$v_i^{k+1} = wv_i^k + c_1 rand \times (pbest_i - x_i^k) + c_2 rand \times (gbest - x_i^k), \quad (11)$$

where,

v_i^k: velocity of agent *i* at iteration *k*,

w : inertia weight,

c_j : weight factor,

rand : random number between 0 and 1,

x_i^k: current position of agent *i* at iteration *k*,

$pbest_i$: *pbest* of agent *i*.

Using the above equation, a certain velocity, which gradually gets close to *pbest* and *gbest* can be calculated. The current position (searching point in the solution space) can be modified by the following equation:

$$x_i^{k+1} = x_i^k + v_i^{k+1}. \quad (12)$$

Fig. 1 shows scheme of of position update of a particle.

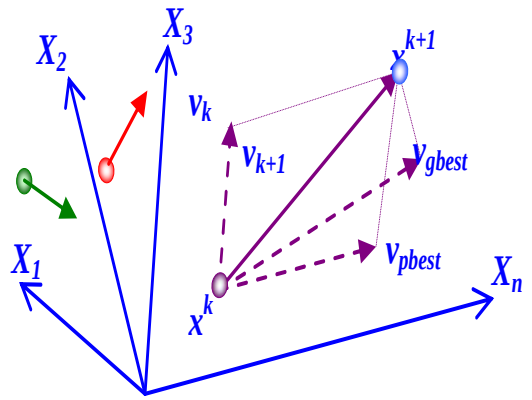

Figure 1 Scheme of position update of a particle 。

IV. THE RVM BASED ON PSO MODEL

In this section, we describe the proposed hybrid PSO-RVM method for groundwater level forecasting, PSO-RVM method combines two machine learning methods by optimizing the parameters of RVM using PSO. PSO starts with *n*-randomly selected particles and searches for the optimal particle iteratively. Each particle is a *m*-dimensional vector and represents a candidate solution. PSO algorithm can gain the optimal parameters of RVM that lead to best prediction accuracy in present circumstances, and it uses the fittest particles to contribute to the next generation of candidate particles. Thus, on the average, each successive population of candidate particles fits better than its predecessor. This process continues until the performance of RVM converges [10]. The procedure describing proposed PSO-RVM approach is as follows.

[Step 1] Parameters *m, w,* c_1, c_2,*T* are initialized, and *k*=0 is set, where

m: number of population,

w: inertia weight,

c_i: weight factor,

T: number of iteration.

[Step 1.1] Velocity and position of each particle are initialized randomly. Each particle vector stand for a RVM model, which is corresponding to kernel function parameter σ.

[Step 1.2] Each of the particle fitness is evaluated, *gbest* is initialized with the index of the particle with the best function value among the population and $pbest_i$ is initialized with a copy of each previous particle.The fitness function is defined as:

$$F_{fitness} = \sum_{i=1}^{N} \left(\frac{y - y_i}{N}\right)^{\frac{1}{2}}, \quad (13)$$

where y is the measured value of a sample, y_i is the predictive value of a sample, and N is the number of samples for i=1,2, ⋯,N.

[Step 2] The following steps are repeated until a stopping criterion is satisfied.

[Step 2.1] Each particle's velocity is updated according to (11) and each particle's position is updated according to (12).

[Step 2.2] Each particle's fitness is evaluated.

[Step 2.3] The personal best position $pbest_i$ and the global best position *gbest* are updated.

[Step 3] Let *k*=*k*+1 and return to Step 2 until the last maximum generation is reached or the best gbest is found.

V. THE SIMULATION EXPERIMENTS

A. Experimental Data

There is 12-year (1983-1994) record of shallow monthly groundwater level in Beijing of China from the literatures [11, 12, 13], that is {*Z*(*t*),t=1,2,…,144}. The first nine years time series are used for training the model, and the remaining three years data are used for verification or

TABLE I.
THE OBSERVED GROUNDWATER LEVEL ON THE REMAINING THREE YEARS

Month	1992	1993	1993
Jan.	42.07	42.29	42.12
Feb.	42.04	42.17	42.09
Mar.	42.02	42.44	41.98
Apr.	42.44	42.49	42.28
May	42.48	42.37	42.24
Jun.	42.53	42.44	42.20
Jul.	42.72	42.73	42.54
Aug.	43.04	42.68	42.75
Sep.	42.8	42.32	42.86
Oct.	42.66	42.37	42.55
Nov.	42.61	42.4	42.55
Dec.	42.47	42.3	42.55

testing purposes. Table I shows the observed groundwater level on the remaining three years.

Parameter Selection

Because RVM is a kernel method, it is important to choice the parameter of its model, including the selection of kernel function and its parameter. Some researches show that when the lack of a prior knowledge about the process. Gaussian kernel function outperforms others in fitting for regression problems. So Gaussian kernel function is used in the RVM model, meanwhile it is vital to determine the parameter of kernel function because of its direct effect on generalization capability of the model.

The PSO algorithm we used is the standard global version with inertia weight, the acceleration factors c_1 and c_2 are both 2.0, a decaying inertia weight w starting at 0.9 and ending at 0.4 is used, and the population size is set at 30 [14]. The parameter setup of PSO-RVM is shown in Table II detail.

TABLE II.
PARAMETER SELECTION FOR THE RVM-PSO

Parameter	m	T	c_1	c_2	σ
Value	30	200	2	2	[0.1,100]

B. Prediction Methodology

Suppose the current time is t, we want to predict $y(t+1)$ for the future time $t+1$ with the knowledge of the value $y(t-m)$, $y(t-n+1),\ldots, y(t)$ for past time $t-m$, $t-m+1,\ldots,t$, respectively. The prediction function is expressed as

TABLE III.
THE INPUT AND OUTPUT SAMPLES

Sample number	Input samples	Output samples
1	$x(1), x(2), \ldots, x(m)$	$x(m+1)$
2	$x(2), x(3), \ldots, x(m+1)$	$x(m+2)$
…	……	…
$(n-m)$	$x(n-m), x(n-m+1), \ldots, x(n-1)$	$x(n)$

[15]: $y(t+1)=f(t, y(t), y(t-1), \ldots, y(t-m))$. So, the input and output samples for the build model are shown in Table III, where $m=5$ [16] is reasonable and the number of training samples is $n-m=103$ in this paper. We examine the groundwater level prediction of different time, Relative Mean Error (*RME*) and Root Mean Squared Error (*RMSE*) are applied as performance indices.

$$RME = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - Y_i^*}{Y_i}\right| \tag{14}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - Y_i^*}{Y_i}\right|^2} \tag{15}$$

C. Results and Analysis

The 103 samples are used as training, the 36 samples are used as test, the convergence process is shown in Fig. 2, and the optimal Gaussian kernel parameter is determined $\sigma^2=4.382$. For the sake of comparison，the SVM based PSO model ($C=92$, $\varepsilon=0.05$, $\sigma^2=3.25$) is also applied to this study [12].

.The RVM and the SVM regression models optimized by the PSO are separately established according to the above parameters for groundwater level forecasting, 36 test samples are tested by the two models, Fig. 3 and Fig. 4 show the foretasted groundwater level and the absolute error for the RVM on the remaining three years, Fig. 5 and Fig. 6 show the foretasted groundwater level and the absolute error for the RVM on the remaining three years, respectively. Table IV shows the *RME* and *RMSE* for the two predicting models.

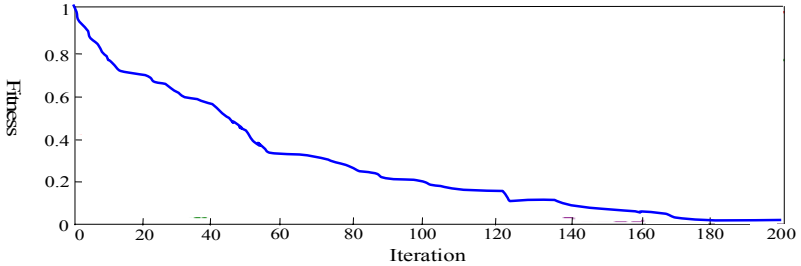

Figure 2 Converge process of PSO.

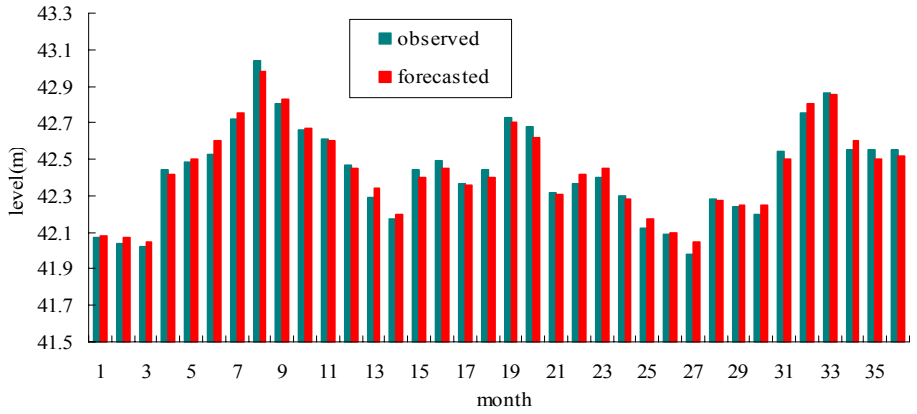

Figure 3. The forecasted groundwater level for the RVM

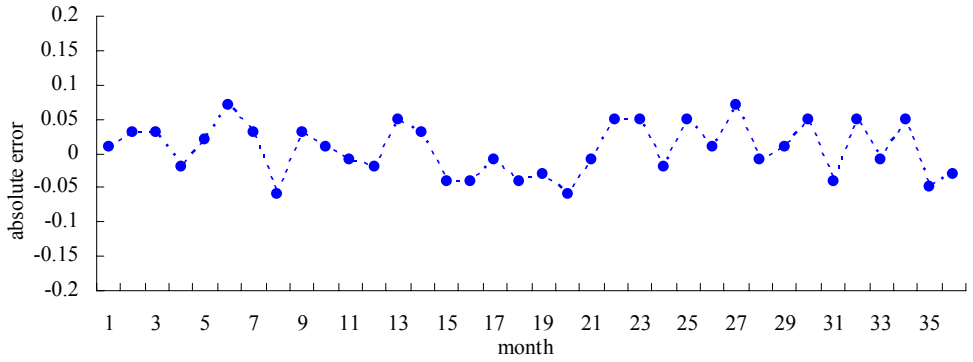

Figure 4. The absolute error for the RVM

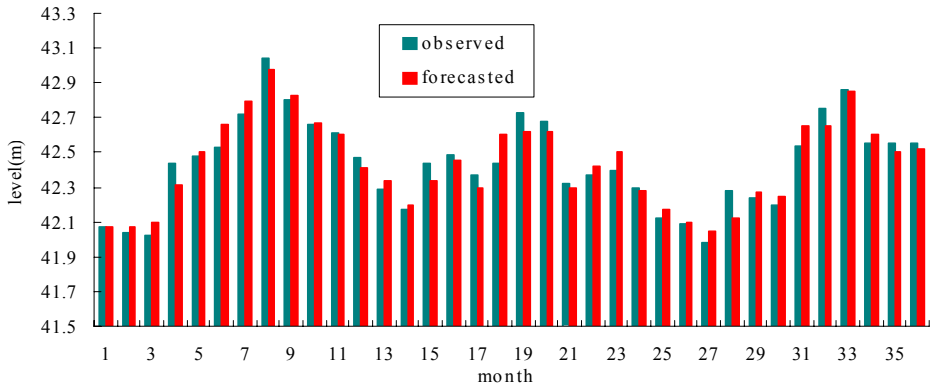

Figure 5 .The forecasted groundwater level for the SVM

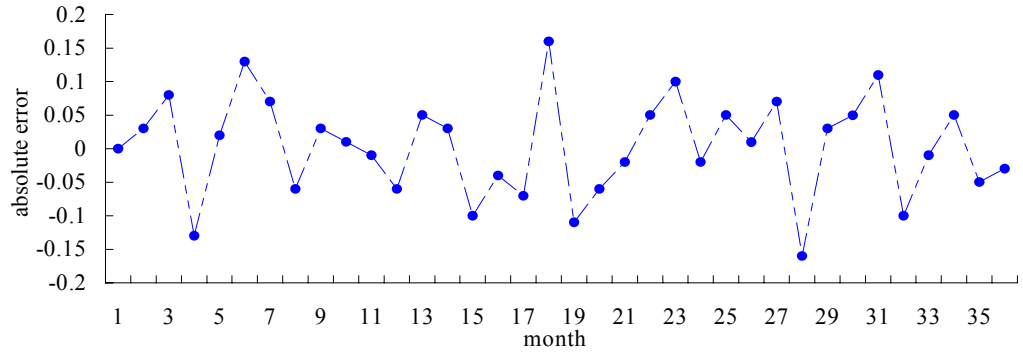

Figure 6. The absolute error for the SVM

TABLE IV.
THE COMPARISON FOR THE TWO PREDICTORS

Predictors	*RME*	*RMSE*
RVM	3.31 %	5.51%
SVM	6.72 %	11.26 %

It can be noticed that the RVM model outperforms the SVM model, whether the *RME* or the *RMSE* of training samples, the former is less than the letter, and with the increase of time, the predicting error increase gradually for the SVM model, however, the forecasted results show good agreement with observed results in last months for the RVM model.

To exhibit the influence of each training sample to the final solution, the estimated variances values of the models are used [3]. For the RVM, the estimation of the inverse variance hyperparameter a_i according to the maximum likelihood II method can be considered as a weighting of the sample importance in the regression, the estimated Lagrange multipliers are taken into account for the SVM. According to this, the highest alpha values (lowest a_i^{-1}) for the RVM are assigned to the least relevance vector samples and the highest Lagrange multipliers a_i for the SVM to the least support vector samples for their model fitting respectively.

Fig.7 and Fig. 8 show the most important training samples for two models as well as those who do not contribute to the solution ($a_i^{-1}\rightarrow 0$ for the RVM and $a_i=0$ for the SVM). There are some samples that are used by the two models as well as others that are not significant for any of them.

The RVM model is comprehensively compared with the SVM model in Table V, it presents not only the numbers of support vectors for the SVM and relevance vectors for the RVM needed to build respective models for the prediction but only the MSE of the training samples and the test samples. It is evident that the RVM exhibits some important advantages over the SVM. The RVM is drastically more sparse, 25 relevant vectors may establish a good model, its sparsity is reduced more than a half with respect to the SVM implying that less than a quarter of the training samples contribute to the fitted model, all the relevant vectors for the RVM are used as the support vectors for the SVM. And then the MSE of training samples for the RVM are smaller than that for the SVM as a result of its conditional distribution instead of a point estimate [3, 17, 18].

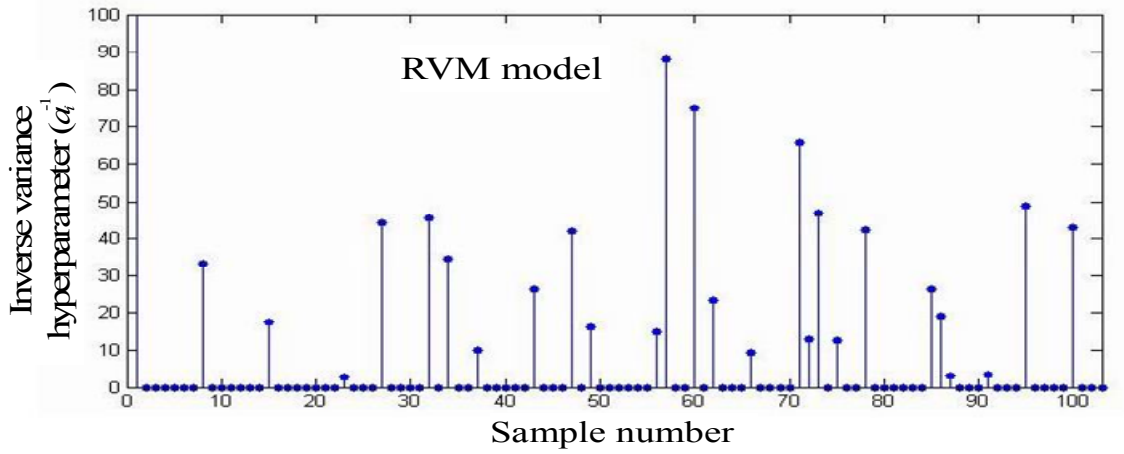

Figure 7. Importance of training samples for the SVM

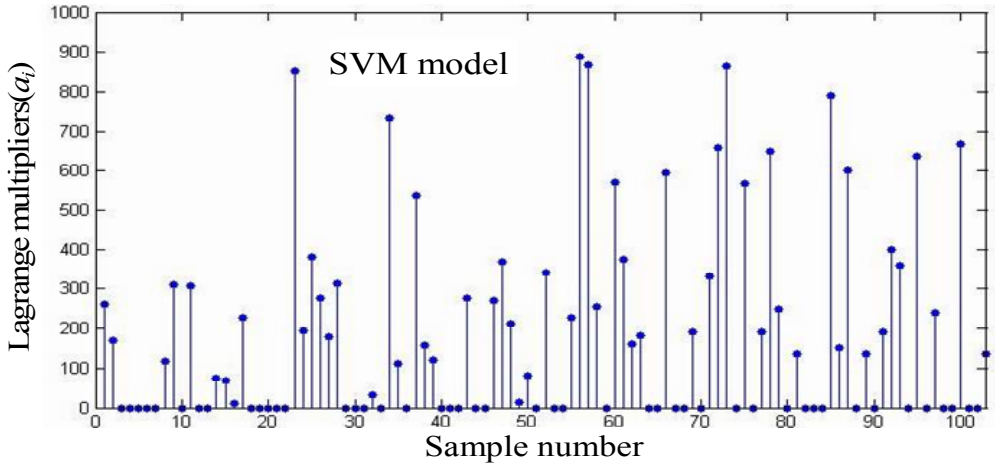

Figure 8. Importance of training samples for the RVM

TABLE V.
PERFORMANCE COMPARISON FOR THE SVM AND THE RVM

vector machine	number of vector	percent of training set	MSE of training set
SVM	57	55.3%	3.871%
RVM	25	24.2%	2.328%

VI. CONCLUSION

The RVM model based on PSO is proposed in this paper, the PSO has good convergence speed and search ability to the optimum, it can well optimize the parameter of RVM model, groundwater level forecasting is applied to the optimized RVM model, and the simulation experiments demonstrated its success. And compared with the SVM model, the RVM model is more accurate than the SVM model; furthermore, it can obtain the best results and achieve the greatest levels of sparsity. So, the RVM based on PSO will provide reliable method for time-series analysis and statistical learning.

ACKNOWLEDGMENTS

This work is supported by the Science Research Foundation of Hebei Education Department of China No. 2009422, and Natural Science Foundation of Hebei Province of China No. E2010001026.

REFERENCES

[1] P. D. Sreekanth1, N. Geethanjali, P. D. Sreedevi3, Shakeel Ahmed, N. Ravi Kumar and P. D. Kamala Jayanthi, "Forecasting groundwater level using artificial neural networks," *CURRENT SCIENCE*, 2009,96(7): 933-939.

[2] Abedalrazq Khalil, Mohammad N. Almasri，Mac McKee, and Jagath J, Kaluarachchi1, "Applicability of statistical learning algorithms in groundwater quality modeling," *Water Resurces Research*, 2005.

[3] Noslen Hernández, Isneri Talavera, Angel Dago, Rolando J. Biscay, Marcia M. Castro Ferreira, and Diana Porro, "Relevance vector machines for multivariate calibration purposes," *Chemometrics*, 2008(22):386-694.

[4] Abedalrazq F Khalil，Mac McKee，Mariush Kemblowski, Tirusew Asefa, and Luis Bastidas, "Multiobjective analysis of chaotic dynamic systems with spame learning machines," Advances *in Water Resources*, 2006(29):72-88.

[5] Tipping ME, "Sparse Bayesian learning and the relevance vector machine," *Journal of machine learning research*, 2001, 1(3):211-244.

[6] Tipping ME, "Sparse Bayesian learning and the relevance vector machine," *Journal of machine learning research*, 2001, 1(3):211-244.

[7] Kennedy, J. and Eberhart, R.C, "Particle swarm optimization," *Proc. IEEE Int'l. Conf. on Neural Networks*, IV, 1995, 1942-1948.

[8] Li Bingyu, Xiao Yunshi, and Wang Lei, "A hybrid Particle swarm optimization algorithm for solving complex functions with high dimension," *Informalion and Coarrol*" 2004, (33): 27-30.

[9] Amany El-Zonkoly, "PARTICLE SWARM OPTIMIZATION FOR SOLVING THE PROBLEM OF TRANSMISSION SYSTEMS AND GENERATION EXPANSION," *Nansoura Engineering* 2005, 30(1):15-20.

[10] Radoslav Goldman, et.al, "Candidate Markers for the Detection of Hepato Cellular Carcinoma in Low-mofraction of Serum," *Carcinogenesis*, 2007, 28(10):2149-2153.

[11] Wensheng Wang and Jing Ding, "Wavelet Network Model and Its Application to the Prediction of Hydrology," *Nature and Science*, 2003, 1(1): 67-71.

[12] Qing G, Ding J, and Liu G, "Self-adapted BP algorithm and its application to flood forecast of river," *Advance in Water Science*, 2002, 13(1):37-41.

[13] WANG Hongrui, YE Letian, and LIU Changming, et al, "Problems in wavelet analysis of hydrologic series and some suggestions on improvement," *Progress In Natural Science*, 2007, 17(1): 80-86, 2007

[14] Hai Shen, Xiaoming Zhou, Xiaoming Zhou, and Xiaoming Zhou, "Bacterial foraging optimization algorithm with particle swarm optimization strategy for global numerical optimization," *Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation,* 2009, 497-504

[15] C.-H. Wu, Jan-Ming Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Trans. Intelligent Transportation Systems*, 2004, 5(4):276-281.

[16] CAO Cheng-tao, and XU Jian-min, "SVM based on PSO and its application in traffic flow predication," *Computer Engineering and Applications*, 2007, 432(15): 12-14.

[17] Shivam, Tripathi, and Rao S, "On selection of kernel parameters in relevance vector machines for hydrologic applications," *Stochastic Environmental Research and Risk Assessment,* 2007, 21(6):747-764.

[18] Kropotov Dmitry, and Ptashko Nikita, "On Kernel Selection in Relevance Vector Machines Using Stability Principle," *Proceedings of the 18th International Conference on Pattern Recognition*, 2007 (4): 233-236.

Weiguo Zhao was born in Xingtai, Hebei Province, China, on October 23, 1977. He received the Master Degree in Computer Technology and Application from Hebei University of Technology, Tianjin, China, in 2008. He is a Engineer of Hebei University of Engineering, teaches and conducts research in the areas of system sciences.and intelligent optimization.

Yanfeng Gao was born in Handan Hebei Province, China, on September 14, 1979. He received the Master Degree in Electrical engineering from Hebei University of Technology, Tianjin, China, in 2007. He is a Lecturer of Hebei University of Engineering, teaches and conducts research in the areas of Electrical engineering and automation.

Chunliu Li received the MS. Degree from Dalian Fishery University in 2005 in China; she has published 5 papers in referred journals.