Periodica Polytechnica
Civil Engineering

60(4), pp. 511–515, 2016
DOI: 10.3311/PPci.8206

RESEARCH ARTICLE

Determination of Electrical Resistivity of Soil Based on Thermal Resistivity Using RVM and MPMR

Pijush Samui, *Dookie* Kim

Received 05-05-2015, accepted 17-07-2015

Abstract

This article adopts Relevance Vector Machine (RVM) and Minimax Probability Machine Regression (MPMR) for prediction Soil Electrical Resistivity(R_E) of soil. RVM uses an improper hierarchical prior. It optimizes over hyperparameters. MPMR is a probabilistic model. Two models (MODEL I and MODEL II) have been adopted. Percentage sum of the gravel and sand size fractions (F) and Soil Thermal Resistivity(R_T) has been takes as inputs in MODEL I. MODEL II uses F,R_T and saturation of soils(S) as input variables. The results of RVM and MPMR have been compared with the Artificial Neural Network (ANN). The developed RVM and MPMR proves his ability for prediction of R_E of soil.

Keywords

Soil Electrical Resistivity · Soil Thermal Resistivity · Relevance Vector Machine · Minimax Probability Machine Regression · Artificial Neural Network

Pijush Samui
Department of Civil Engineering, NIT Patna, Patna, Bihar, India
e-mail: pijushsamui@gmail.com

Dookie Kim
Department of Civil Engineering, Kunsan National University, Kunsan, Jeonbuk, South Korea
e-mail: kim2kie@chol.com

1 Introduction

Soil Electrical Resistivity (R_E) is an important parameter for constructing high voltage buried power cables [1, 2]. The value of R_E depends on different parameters such as water content, degree of saturation, organic content, pore water composition, geologic formation, temperature, compaction, specific surface area, etc. Osman and Harith [3] showed that an increase in electrical resistivity with the increase of angle of shearing resistance, bulk density, and Standard Penetration Test value. Magnesium, sulfate content, calcium and sodium have significant effect on R_E of soil. So, the determination of R_E of soil is a complicated task [4]. Geotechnical engineers use different methods for determination of R_E based on soil thermal resistivity (RT) [5]. R_T is influenced by moisture content, dry density, mineral composition and temperature. So, a strong correlation exists between R_T and R_E [6]. Recently, Erzin et al., [7] successfully adopted Artificial Neural Network (ANN) for prediction of R_E of soil. However, ANN has some limitations such as black box approach, low generalization capability, arriving at local minima, etc. [8, 9]. This article adopts Relevance Vector Machine (RVM) and Minimax Probability Machine Regression (MPMR) for determination of R_E of soil. RVM was developed by Tipping [10]. It is a sparse bayesian nonlinear regression technique Tipping [11]. It uses improper hierarchical prior and optimizing over hyper parameters. There are lots of applications of RVM in literatures [12–14]. Li [12] successfully applied fuzzy progressive transductive relevance vector machine classifier for network attack detection. RVM has been also used by Wang [13] for intrusion detection of internet of things. Batt and Stevens [14] successfully applied RVM for modelling of suspended fine sediment transport in a shallow lake [15] identified soil line by using RVM. Wang et al., [16] used RVM for machine fault diagnosis. MPMR is developed by Lanckriet et al., [17]. It maximizes the minimum probability that future predicted output of the regression model will be within some bound of the true regression function [18]. Researchers have successfully used MPMR for solving different problems in engineering [19–21]. Sun et al., [19] used MPMR for modelling of a chaotic time series. Yang et al., [20] successfully applied MPM for feature classification.

Zhou et al., [21] examined the capability of MPM for face recognition. This article adopts the database collected from the work of Erzin et al., [7]. Table 1 shows the statistical parameter of the dataset.

The datasets contain information about R_E, R_T, percentage sum of the gravel and sand size fractions (F) and saturation of soil (S_r). For obtaining the dataset, soil samples were collected from the different offshore locations in India. Two models (MODEL I and MODEL II) have been developed for prediction of R_E of soil. In MODEL I, input variables are R_T and F. MODEL II adopts R_T, F and S_r as input variables. The developed RVM and MPMR have been compared with the ANN model.

2 Details of RVM

RVM is trained in Bayesian framework [10]. In RVM, the relation between input(x) and output(y) is given below:

$$y = \Phi w + \varepsilon \tag{1}$$

where w is weight, ε is noise, $\Phi = [\varphi(x_1), \ldots, \varphi(x_n)]\, \varphi(x_n) = [K(x_n, x_1), K(x_n, x_2), \ldots, K(x_n, x_M)]^T$ and $K(x_n, x_i)$ is a kernel function.

For MODEL I, $x = [R_T, F]$ and $y = [R_E]$.

For MODEL II, $x = [R_T, F, S_r]$ and $y = [R_E]$.

The likelihood of the complete dataset is given below:

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

Automatic Relevance Detection (ARD) prior is set over the weights for preventing overfitting.

$$p(w|\alpha) = \prod_{i=0}^{N} N\left(w_i|0, \alpha_i^{-1}\right) \tag{3}$$

Where α is a hyperparameter vector that controls how far from zero each weight is allowed to deviate [22]. The following expression is obtained by combining the likelihood and prior within Bayes' rule

$$p\left(w, \alpha, \sigma^2/y\right) = \\ = \frac{p\left(y/w, \alpha, \sigma^2\right).p(w, \alpha, \sigma)}{\int p(y/w, \alpha, \sigma^2)\, p(w, \alpha, \sigma^2)\, dw d\alpha d\sigma^2} \tag{4}$$

$p\left(w / y, \alpha, \sigma^2\right)$ follows Gaussian distribution. So, the expression of $p\left(w / y, \alpha, \sigma^2\right)$ is given below.

$$p\left(w/y, \alpha, \sigma^2\right) \sim N(\mu, \Sigma) \tag{5}$$

Where μ is mean and $\sum$ is covariance. The expression of μ and $\sum$ is given below.

$$\mu = \sigma^{-2} \sum \Phi^T y \tag{6}$$

$$\sum = \left(\sigma^{-2}\Phi^T\Phi + A\right)^{-1} \tag{7}$$

with diagonal $A = diag(\alpha_0, \ldots, \alpha_N)$.

For uniform hyperpriors over α and σ^2, one needs only maximize the term $p\left(t / \alpha, \sigma^2\right)$:

$$p\left(y/\alpha, \sigma^2\right) = \int p\left(y/w, \sigma^2\right) p(w/\alpha)\, dw = \\ = \left(\frac{(2\pi)^{\frac{-N}{2}}}{\sqrt{\left|\sigma^2 + \Phi A^{-1}\Phi^T\right|}}\right) \times exp\left\{-\frac{1}{2}y^T\left(\sigma^2 + \Phi A^{-1}\Phi^T\right)^{-1} y\right\} \tag{8}$$

The outcome of this optimization is that many elements of α go to infinity such that w will have only a few nonzero weights that will be considered as relevant vectors. Training and testing datasets have been required for developing the RVM. This article uses 165 datasets as training datasets. The remaining 71 datasets have been adopted as testing dataset. The datasets are normalized between 0 and 1. Radial basis function has been adopted as a kernel function. The expression of radial basis function is given below

$$K(x, x_i) = \exp\left\{-\frac{(x_i - x)(x_i - x)^T}{2\sigma^2}\right\} \tag{9}$$

where σ is width of radial basis function. Fig. 1 shows the flow chart of the RVM. The program of RVM has been constructed in MATLAB environment.

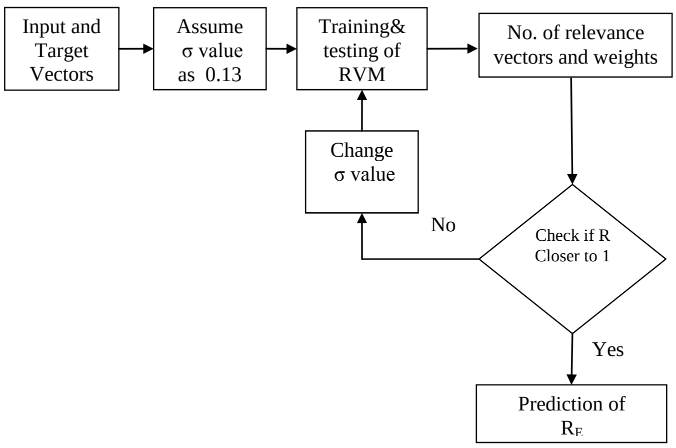

Fig. 1. Flow chart of the RVM.

3 Details of MPMR

MPMR is constructed based on minimax probability machine classification by using kernel function. In MPMR, the relation between input(x) and output(y) is given below:

$$y = \sum_{i=1}^{N} \beta_i K(x_i, x) + b \tag{10}$$

where $K(x_i, x)$ is kernel function and β, b are output of the MPMR algorithm.

Tab. 1. Statistical parameters of the dataset.

Variable	mean	Standard deviation	Skewness	Kurtosis
F (%)	28.36	23.29	1.11	4.24
S_r (%)	62.19	25.71	0.09	1.66
R_T (°C·m/W)	5.47	7.30	4.32	26.34
R_E (Ω ·m)	31.65	57.23	4.98	34.83

For MODEL I, $x = [R_T, F]$ and $y = [R_E]$. For MODEL II, $x = [R_T, F, S_r]$ and $y = [R_E]$.

To develop MPMR, one data set is obtained by shifting all of the datasets $+\varepsilon$ along the output. The other dataset is obtained by shifting all of the datasets $-\varepsilon$ along the output. The regression surface is the classification boundary between these two classes. MPMR separates the training dataset into the following two classes.

$$u_i = (y_i + \varepsilon, x_{i1}, x_{i2}, \ldots, x_{in}). \quad (11)$$

$$v_i = (y_i + \varepsilon, x_{i1}, x_{i2}, \ldots, x_{in}). \quad (12)$$

The classification boundary between u_i and v_i is the regression surface. The details of MPMR are given by Strohmann and Grudic [18]. MPMR uses radial basis function as kernel function. MPMR adopts the same training dataset, testing dataset and normalization technique as used by the RVM model. Fig. 2 shows flow chart of the MPMR for prediction of R_E.

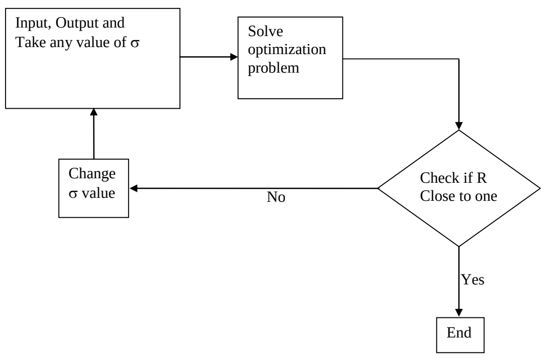

Fig. 2. Flow chart for prediction of R_E.

The program of MPMR is constructed by using MATLAB.

4 Results and Discussion

For developing RVM, the design value of σ has been determined by trial and error approach. For MODEL I, the design value of σ is 0.6. The performance of training dataset has been depicted in Fig. 3.

Fig. 4 illustrates the performance of testing dataset. The performance of developed models has been assessed in terms of Coefficient of Correlation(R) value.

For a good model, the value of R should be close to one. It is observed from Figs.1 and 2 that the value of R is close to one. The developed RVM gives the following equation for prediction

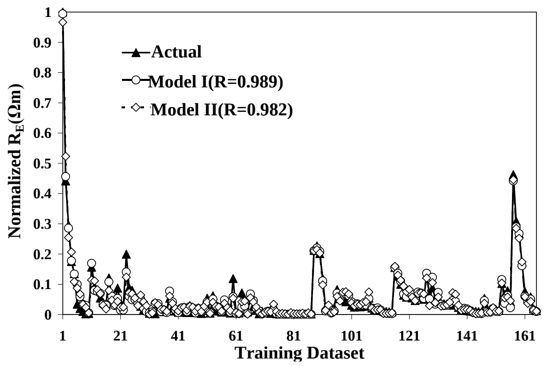

Fig. 3. Performance of training dataset for the RVM.

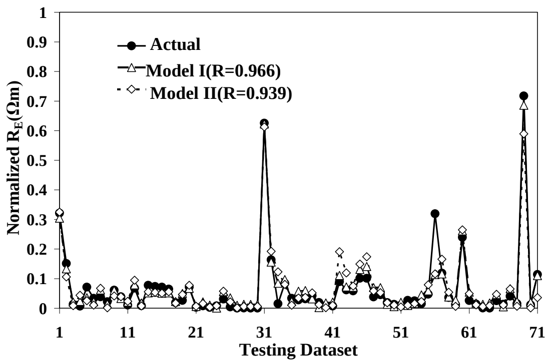

Fig. 4. Performance of testing dataset.

of R_E.

$$R_E = \sum_{i=1}^{165} w_i \exp\left[\frac{-(x_i - x)(x_i - x)^T}{0.72}\right] \quad (13)$$

Fig. 5 shows the value of w.

For MODEL II, the design value of σ is 0.4. The performances of training and testing datasets have been shown in Fig. 3 and 4 respectively. The value of R is close to one for training as well as testing datasets. MODEL II gives the following equation for prediction of R_E.

$$R_E = \sum_{i=1}^{165} w_i \exp\left[\frac{-(x_i - x)(x_i - x)^T}{0.32}\right] \quad (14)$$

The values of w have been shown in Fig. 5. For developing MPMR, the design value ε and σ have been determined by trial and error approach. For MODEL I, the design values of ε and σ are 0.003 and 0.7 respectively. The performance of training dataset has been depicted in Fig. 6.

It is also clear from Fig. 4 and 5 that the value of R is close to one for training as well as testing dataset. For MODEL II, the design values of ε and σ are 0.005 and 0.2 respectively. The

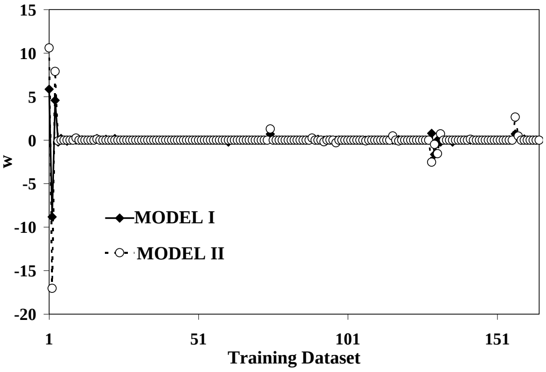

Fig. 5. Values of w for the RVM.

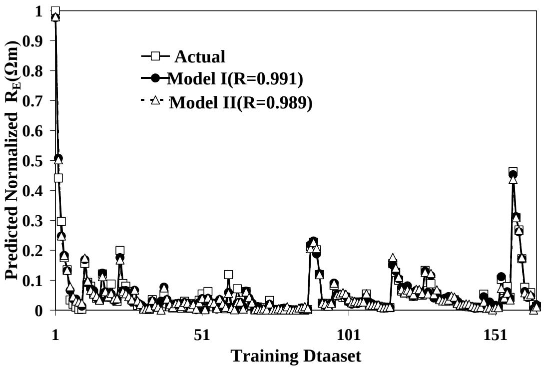

Fig. 6. Performance of training dataset for MPMR.

performance of training and testing dataset has been shown in Fig. 6 and 7 respectively. The value of R is close to one for training as well as testing dataset. Therefore, the developed MPMR predicts R_E reasonable well.

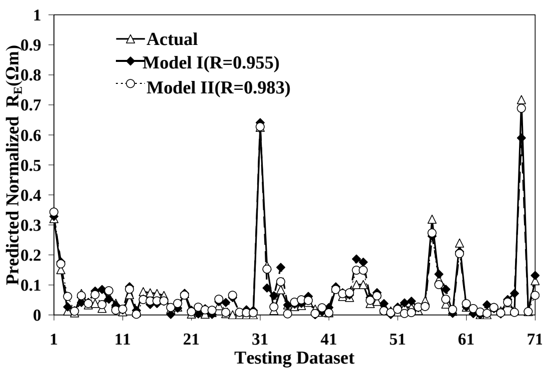

Fig. 7. Performance of testing dataset for MPMR.

Fig. 8 shows the bar chart of R values of ANN, RVM and MPMR models. The comparison has been done for testing dataset.

It is clear from Fig. 6 that the performances of ANN, RVM and MPMR are almost same. This article uses Root Mean Square Error (RMSE), Mean Absolute Error (MAE), coefficient of efficiency (E), root mean square error to observation's standard deviation ratio (RSR), variance account for(VAF), performance index(ρ) and normalized mean bias error (NMBE) to asses the performance of the RVM and MPMR models.

The expressions of RMSE, MAE, E, RSR, VAF, ρ, and NMBE are given below [23–27].

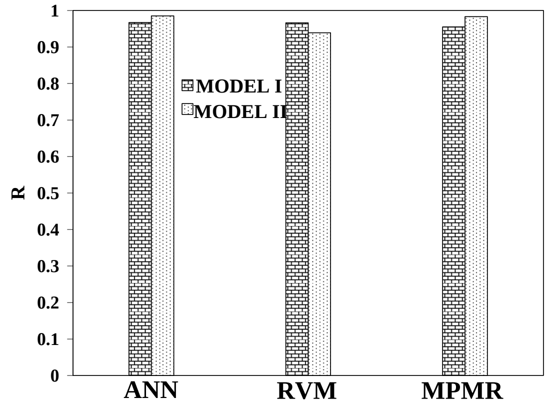

Fig. 8. Comparison between the ANN,RVM and MPMR models.

Table 2 shows the values of different parameters. For a good model, the value of RMSE and MAE should be close to zero.

$$MAE = \frac{\sum_{i=1}^{N} |A_i - P_i|}{N} \tag{15}$$

$$E = 1 - \frac{\sum_{i=1}^{N} (A_i - P_i)^2}{\sum_{i=1}^{N} \left(A_i - \overline{A}\right)^2} \tag{16}$$

$$RSR = \frac{RMSE}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(A_i - \overline{A}\right)^2}} \tag{17}$$

$$\rho = \frac{RMSE}{\overline{A}(R+1)} \tag{18}$$

$$VAF = (1 - (var(A_i - P_i) / var(A_i)))\,100 \tag{19}$$

Where A is actual value, P is predicted value, N is the Number of dataset, var is variance, $\overline{A}$ is the mean and p is the number of predictor variable. purpose. The The developed RVM only shows under-prediction for MODEL I. For a good model, the value of RSR and ρ should be low. The developed models show low value of RSR and ρ. For a good accuracy of model, the value of VAF should be close to 100. The value of VAF is close to 100 for all the developed models. For an accurate model, the value of E is close to one. The developed models show the value of E is close to one. Hence, the developed models prove their capability for prediction of electrical resistivity of soil. The developed RVM and MPMR use less tuning parameters compare to the ANN model. The developed MPMR and RVM models are probabilistic model. However, ANN is not a probabilistic model. Kernel function has been adopted for developing the RVM and MPMR models. For developing ANN, kernel function is not required.

5 Conclusions

This article examines the capability of RVM and MPMR for prediction of R_E of soil. Different input variables have been

Tab. 2. Values of different error parameters of the developed RVM and MPMR models.

Models	RMSE		MAE		E		RSR		NMBE(%)		ϱ		VAF	
	Training Dataset	Testing Dataset	Training Dataset	Testing Dataset	Training Dataset	Testing Dataset	Training Dataset	Testing Dataset	Training Dataset	Testing Dataset	Training Dataset	Testing Dataset	Training Dataset	Testing Dataset
RVM(MODEL I)	7.948	16.310	5.330	9.672	0.978	0.934	0.147	0.256	-0.352	5.171	0.133	0.233	97.834	93.498
RVM(MODEL II)	14.019	21.370	6.753	12.359	0.966	0.886	0.259	0.336	0.823	5.171	0.235	0.310	96.612	88.723
MPMR(MODEL I)	7.727	17.309	5.019	12.077	0.979	0.925	0.143	0.272	0.481	10.635	0.129	0.249	97.953	92.937
MPMR(MODEL II)	7.117	10.961	4.475	8.438	0.982	0.970	0.131	0.172	0.603	4.484	0.119	0.155	98.263	97.088

tried to get best performance. The developed RVM and MPMR predict R_E of soil reasonable well. The developed equation can be used for practical purpose. The performance of RVM and MPMR is comparable with the ANN model. The developed RVM produces sparse solution. There is no sparseness in the MPMR. This article gives practical tools based on RVM and MPMR for prediction of R_E of soil.

References

1 **Del Mar WA, Burrell RW, Bauer CA**, *Soil types identification and physical properties-II, soil thermal characteristics in relation to underground power cables*, AIEE Committee Report, (1960), 795–803.

2 **King SY, Halfter NA**, *Under Ground Power Cables*, Longman; London, 1982.

3 **Syed Osman S B, Tuan Harith ZZ**, *Correlation of Electrical Resistivity with Some Soil Properties in Predicting Factor of Safety in Slopes Using Simple Multi Meter*, UTP Institution repository. University Teknologi Petronas. ID Code 5649, 2010.

4 **Abu Hassanein ZS**, *Use of electrical resistivity measurement as a quality control tool for compacted clay liners. M.S. Thesis*, University of Wisconsin; Madison, 1994.

5 **Narain Singh D, Kuriyan SJ, Chakravarthy Manthena K**, *A generalised relationship between soil electrical and thermal resistivities*, Experimental Thermal and Fluid Science, **25**(3-4), (2001), 175–181, DOI 10.1016/S0894-1777(01)00082-6.

6 **Salomone LA, Marlow JI**, *Soil Rock Classification According to Thermal Conductivity*, EPRI CU-6482, Electric Power Research Institute; Palo Alto, CA, 1989.

7 **Erzin Y, Rao BH, Patel A, Gumaste SD, Singh DN**, *Artificial neural network models for predicting electrical resistivity of soils from their thermal resistivity*, International Journal of Thermal Sciences, **49**(1), (2010), 118–130, DOI 10.1016/j.ijthermalsci.2009.06.008.

8 **Park D, Rilett LR**, *Forecasting freeway link travel times with a multi-layer feed forward neural network*, Computer Aided Civil and infra Structure Eng, **14**, (1999), 358–367.

9 **Kecman V**, *Learning and Soft Computing: Support Vector Machines.Neural Networks, and Fuzzy Logic Models*, MIT Press; Cambridge, 2001.

10 **Tipping M**, *The Relevance Vector Machine*, Neural Information Processing Systems, 2000.

11 **Tipping M**, *Sparse Bayesian learning and the relevance vector machine*, Journal of Machine Learning Research, **1**, (2001), 211–244.

12 **Li RA**, *Fuzzy progressive transductive relevance vector machine classifier for network attack detection*, Journal of Information and Computational Science, **8**(15), (2011), 3445–3451.

13 **Wang ZA**, *Hybrid model of rough sets and relevance vector machine for intrusion detection of internet of things*, Journal of Computational Information Systems, **23**, (2012), 9881–9886.

14 **Batt HA, Stevens DK**, *Relevance Vector Machine Models of Suspended Fine Sediment Transport in a Shallow Lake—I: Data Collection*, Environmental Engineering Science, **30**(11), (2013), 681–688, DOI 10.1089/ees.2012.0487.

15 **Cui S, Rajan N, Maas SJ, Youn E**, *An automated soil line identification method using relevance vector machine*, Remote Sensing Letters, **5**(2), (2014), 175–184, DOI 10.1080/2150704X.2014.890759.

16 **Wang B, Liu S-L, Zhang H-L, Jiang C**, *Advances about relevance vector machine and its applications in machine fault diagnosis*, Journal of Vibration and Shock, **34**(3), (2015), 145–167.

17 **Lanckriet G, El Ghaoui L, Bhattacharyya C, Jordan MA**, *Robust minimax approach to classification*, Journal of Machine Learning Research, **3**, (2002), 555–582.

18 **Strohmann TR, Grudic GZA**, *Formulation for minimax probability machine regression*, In: **Dietterich TG, Becker S, Ghahramani Z** (eds.), Advances in Neural Information Processing Systems (NIPS) 14, MIT Press; Cambridge, 2002.

19 **Sun J, Bai Y, Luo J, Dang J**, *Modelling of a chaotic time series using a modified minimax probability machine regression*, Chinese Journal of Physics, **47**(4), (2009), 491–501.

20 **Yang L, Wang L, Sun Y, Zhang R**, *Simultaneous feature selection and classification via Minimax Probability Machine*, International Journal of Computational Intelligence Systems, **3**(6), (2010), 754–760, DOI 10.1080/18756891.2010.9727738.

21 **Zhou Z, Wang Z, Sun X**, *Face recognition based on optimal kernel minimax probability machine*, Journal of Theoretical and Applied Information Technology, **48**(2), (2013), 1645–1651.

22 **Scholkopf B, Smola AJ**, *Learning with kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press; Cambridge, 2002.

23 **Kisi O, Shiri J, Tombul M**, *Modeling rainfall-runoff process using soft computing techniques*, Computers & Geosciences, **51**(51), (2013), 108–117, DOI 10.1016/j.cageo.2012.07.001.

24 **Srinivasulu S, Jain A**, *A comparative analysis of training methods for artificial neural network rainfall–runoff models*, Applied Soft Computing, **6**(3), (2006), 295–306, DOI 10.1016/j.asoc.2005.02.002.

25 **Chen H, Xu C-Y, Guo S**, *Comparison and evaluation of multiple GCMs, statistical downscaling and hydrological models in the study of climate change impacts on runoff*, Journal of Hydrology, **434-435**, (2012), 36–45, DOI 10.1016/j.jhydrol.2012.02.040.

26 **Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL**, *Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations*, Transactions of the ASABE, **50**(3), (2007), 885–900, DOI 10.13031/2013.23153.

27 **Nash JE, Sutcliffe JV**, *River flow forecasting through conceptual models part I — A discussion of principles*, Journal of Hydrology, **10**(3), (1970), 282–290, DOI 10.1016/0022-1694(70)90255-6.