Prediction of daily crop reference evapotranspiration (ET_0) values through a least-squares support vector machine model

Xianghong Guo, Xihuan Sun and Juanjuan Ma

ABSTRACT

Real-time prediction of daily reference crop evapotranspiration (ET_0) is the basis for estimating crop evapotranspiration and for computing crop irrigation requirements. In recent years, least-squares support vector machines (LSSVMs) have been applied for forecasting in many fields of engineering. In this paper, LSSVMs are applied to forecast ET_0 using public weather forecasting data (minimum and maximum temperature, average relative humidity, wind scale and weather conditions). LSSVM-estimated ET_0 is compared with Penman–Monteith (PM)-estimated ET_0 using measured meteorological data. Based on a comparison between LSSVM and PM over a 2 month period, the results show that the root mean square error and mean absolute error are less than 0.5 and 0.4 mm d^{-1}, respectively, and that the model efficiency is greater than 90%. This indicates that ET_0 can be successfully estimated using public weather forecasts through the LSSVM approach.

Key words | least-square support vector, reference crop evapotranspiration, weather forecast prediction model

Xianghong Guo
Xihuan Sun (corresponding author)
Juanjuan Ma
College of Water Resource Science and Engineering,
Taiyuan University of Technology,
Taiyuan 030024,
China
E-mail: *sunxihuan@tyut.edu.cn*

INTRODUCTION

Real-time forecast of irrigation is essential for compiling a dynamic plan of irrigation water use. It can save water, increase crop yields and increase economic benefits in an irrigation area. Real-time forecast of water requirements for crops is the most essential and most difficult part of real-time forecast of irrigation. Real-time prediction of daily reference crop evapotranspiration (ET_0) is the basis for estimating crop evapotranspiration (ET_c) and for computing crop irrigation requirements. The Penman–Monteith (PM) method is recommended by the Food and Agricultural Organization (FAO) of the United Nations as the standard ET_0 method (Allen *et al.* 1998). A number of studies indicate that the PM method is able to provide precise ET_0 estimates in many regions and climates (Allen 1986; Allen *et al.* 1989, 2005, 2006; Jensen *et al.* 1990; De Souza & Yoder 1994; Chiew *et al.* 1995). However, the FAO PM method requires numerous daily meteorological data such as maximum and minimum air temperature (T_{max} and T_{min}), relative humidity (RH), solar radiation (R_s) and wind speed (u). Such meteorological data are often incomplete or inaccessible.

Presently, weather forecasting data are accessible through public media which usually provides information such as rainfall amount and probability, T_{max} and T_{min}, RH, weather conditions (e.g. clear sky, clear to cloudy) and wind scale. Therefore, the application of public weather forecast information for the estimation of ET_0 may potentially be useful for real-time forecast of irrigation requirements.

The support vector machine (SVM) was developed by Vapnik (1995) to solve problems related to small samples, nonlinearity, high dimension and local minimum point efficiently. They are also able to solve actual problems with the advantage of rapid calculation of global minimum and

doi: 10.2166/nh.2011.072

convergence rate. Least squares support vector machines (LSSVMs) were introduced by Suykens & Vandewalle (1999, 2000) as a reformulation of the standard SVM. The training process of standard SVM was simplified by replacing inequality constraints with equality constraints. In recent years, SVMs have been successfully applied to handwriting recognition, particle identification, digital image identification, text categorization, bioinformatics, function approximation and regression and database marketing, among others. Studies on SVM application in the area of hydrology include soil moisture prediction (Gill 2006), rainfall-runoff modelling (Lin & Cheng 2006), sediment-carrying capacity forecasting (Xiong & Li 2005), soil erosion prediction (Li *et al.* 2007) and flood stage forecasting (Liong & Chandrasekaran 2002). LSSVMs have not YET been used for the forecast of ET_0, the objective of this research.

The specific objectives of this study are (1) to explore the use of the LSSVM model for ET_0 prediction based on weather forecast data and (2) to evaluate the performance of the model using measured meteorological data.

THEORY

LSSVMs are reformulations of the standard SVMs. It results in a set of linear equations that eliminate the problem of quadratic programming of SVM.

Consider the given training set

$$\left\{(x_i, y_i) | x_i \in R^l, y_i \in R\right\}_{i=1,2,\ldots l}$$

where x_i and y_i are the input and the output of the ith example and l denotes the number of samples. The support vector method approach aims to construct a regression function with the form:

$$y = \omega^T \varphi(x) + b \tag{1}$$

where $\varphi(\cdot)$ is the nonlinear map of the input space to a usually high-dimensional feature space (can be of infinite dimensions), $\omega \in R^l$ is the coefficient vector and $b \in R$ is a bias term. The unknown parameters ω and b can be obtained by solving the optimization problem (Suykens & Vandewalle 1999, 2000):

$$\begin{aligned} &\min_{\omega,\xi_i} J(\omega, \xi) = \frac{1}{2}\omega^T\omega + \frac{1}{2}\gamma\sum_{i=1}^{l}\xi_i^2 \\ &\text{s.t.} \quad y_i\left[\omega^T\varphi(x) + b\right] = 1 - \xi_i \\ &\qquad \xi_i \geq 0 \quad i = 1, 2, \ldots, l \end{aligned} \tag{2}$$

where γ is a positive real constant and ξ_i are slack variables. The Lagrangian corresponding to Equation (2) can be defined:

$$\begin{aligned} L(\omega, \xi, \alpha) = &\frac{1}{2}\omega^T\omega + \frac{1}{2}\gamma\sum_{i=1}^{l}\xi_i^2 \\ &- \sum_{i=1}^{l}\alpha_i\left(\omega^T\varphi(x_i) + b + \xi_i - y_i\right) \end{aligned} \tag{3}$$

where $\alpha_i (i = 1, 2, \ldots, N)$ are the Lagrange multipliers. The Kuhn–Tucker conditions (Fletcher 1987) can be expressed

$$\begin{cases} \dfrac{\partial L}{\partial \omega} = 0 \to \omega = \sum_{i=1}^{l}\alpha_i\varphi(x_i) \\ \dfrac{\partial L}{\partial b} = 0 \to \sum_{i=1}^{l}\alpha_i = 0 \\ \dfrac{\partial L}{\partial \xi_i} = 0 \to \alpha_i = \gamma\xi_i \quad i = 1, \ldots, l \\ \dfrac{\partial L}{\partial \alpha_i} = 0 \to \omega^T\varphi(x_i) + b + \xi_i - y_i = 0 \quad i = 1, \ldots, l. \end{cases} \tag{4}$$

Values of ω and ξ are obtained by solving Equation (4):

$$\begin{bmatrix} 0 & \vec{1}^T \\ \vec{1} & \mathbf{z}^T\mathbf{z} + \gamma^{-1}\mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \tag{5}$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{z} = [\varphi(x_1), \ldots, \varphi(x_l)]$; $\mathbf{y} = [y_1, \ldots, y_l]^T$; $\vec{1} = [1, \ldots, 1]^T$ and $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_l]^T$.

Solution of $\boldsymbol{\alpha}$ and b can be obtained by solving Equation (5) and substituting it into Equation (1) which becomes:

$$y(x) = \sum_{i=1}^{l}\alpha_i\varphi(x_i)^T\varphi(x) + b. \tag{6}$$

According to the Mercer rule, the kernel function $K(x_i, x_j)$ can be expressed:

$$K(x_i, x_j) = \varphi(x_i)^T\varphi(x_j). \tag{7}$$

Equation (6) is combined with Equation (7), resulting in the equation:

$$y(x) = \sum_{i=1}^{l} \alpha_i K(x_i, x) + b. \tag{8}$$

SIMULATION AND RESULTS

Collection of datasets

Daily climatic data of minimum and maximum temperatures, average RH, wind speed and sunshine duration in Changwu County, Shanxi Province, China (35°12′N, 107° 42′E, altitude 1220 m) were collected during the periods of 1 January 2005–30 June 2007 and 1August 2009–30 September 2009. For the latter period, records of daily weather forecast data T_{max} and T_{min}, average RH, wind scale and weather conditions were also available from http://www.weather.com.cn/html/weather/101110209.shtml. Daily ET_0 values were estimated using the PM56 method because the lysimeter-measured ET_0 values were not available during the period. The PM56-estimated ET_0 values were considered as standard because the values ranked first for both humid and arid regions (Jensen *et al.* 1990).

Determination of input variables

The main task of the LSSVM model for ET_0 is to identify the input vector (dependent variables) and the output. ET_0 mainly depends on temperature, humidity, wind speed and sunshine duration. This study shows that a greater number of dependent variables (temperature, humidity, wind speed and sunshine duration) can produce a more precise ET_0 prediction. Weather parameters considered in this study are based on public weather forecast data, such as maximum and minimum air temperature, RH, relative sunshine duration and wind speed. The output of LSSVM is the daily value of ET_0.

Preparation of training datasets

Public weather forecast information usually cannot provide all the required data, particularly wind speed and sunshine. However, forecast information can provide data such as wind scale and weather conditions. In studies by Cai *et al.* (2005) and Xu *et al.* (2006), wind scale and weather conditions were transformed to wind speeds and sunshine duration; these data are listed in Tables 1 and 2.

The dataset was split into three parts, namely training set, test set 1 and test set 2. The training set included 730 days (1 January 2005–31 December 2006), test set 1 included 181 days (1 January 2007–30 June 2007) and test set 2 included 61 days (1 August 2009–30 September 2009). The purpose of this paper is to establish an LSSVM-estimated ET_0 model based on public weather forecast data. However, public weather forecast data were not available for the training set and test set 1. Wind scale and weather conditions were converted to adopted values combining minimum and maximum temperature and RH as the model input. The training set was used to establish LSSVM for ET_0. Test set 1 and test set 2 were used for checking the performance of LSSVM for estimating ET_0.

LSSVM training

The kernel function of Equation (8) can have many forms such as linear, polynomial or radial basis function (RBF) (Suykens & Vandewalle 1999, 2000). In this paper RBF kernels were selected, of the form:

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

Table 1 | Relationship between wind scales and wind speeds at 2 m heights

Wind Scale	0	1	2	3	4	5	6
Wind speeds	0.0–0.1	0.2–1.1	1.2–2.5	2.6–4.0	4.1–5.9	6.0–8.0	8.1–10.3
Adopted values	0	0.7	1.5	3	5.2	6.7	9

Table 2 | Relationship between weather condition and relative sunshine duration

Weather condition	Clear sky	Clear to cloudy	Cloudy	Overcast sky	Rainy
n/N	1–0.8	0.8–0.6	0.6–0.4	0.4–0.2	0.2–0
Adopted values	0.9	0.7	0.5	0.3	0.1

Note: n is actual sunshine duration; N is theory sunshine duration.

Two parameters, including γ in Equation (2) and σ^2 in the RBF kernel function, should be determined before the application of LSSVM. Optimization of the modelling parameters is of considerable importance for LSSVM models. The combination of the two-step grid search approach and fast leave-one-out cross-validation (Suykens & Vandewalle 1999, 2000) is used for global optimization of the parameters. Using the training set, the optimal values of γ

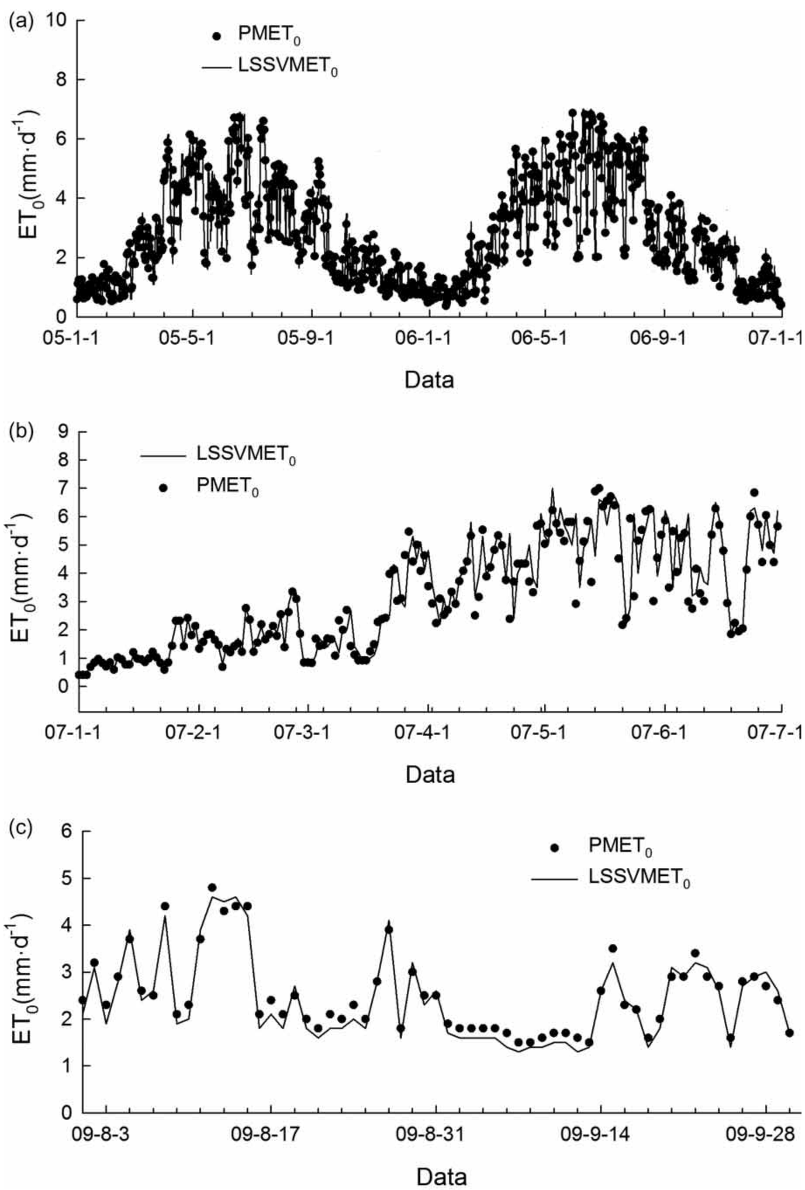

Figure 1 | Comparisons between the PM-estimated ET_0 and LSSVM-estimated ET_0: (a) training set; (b) test set 1; (c) test set 2.

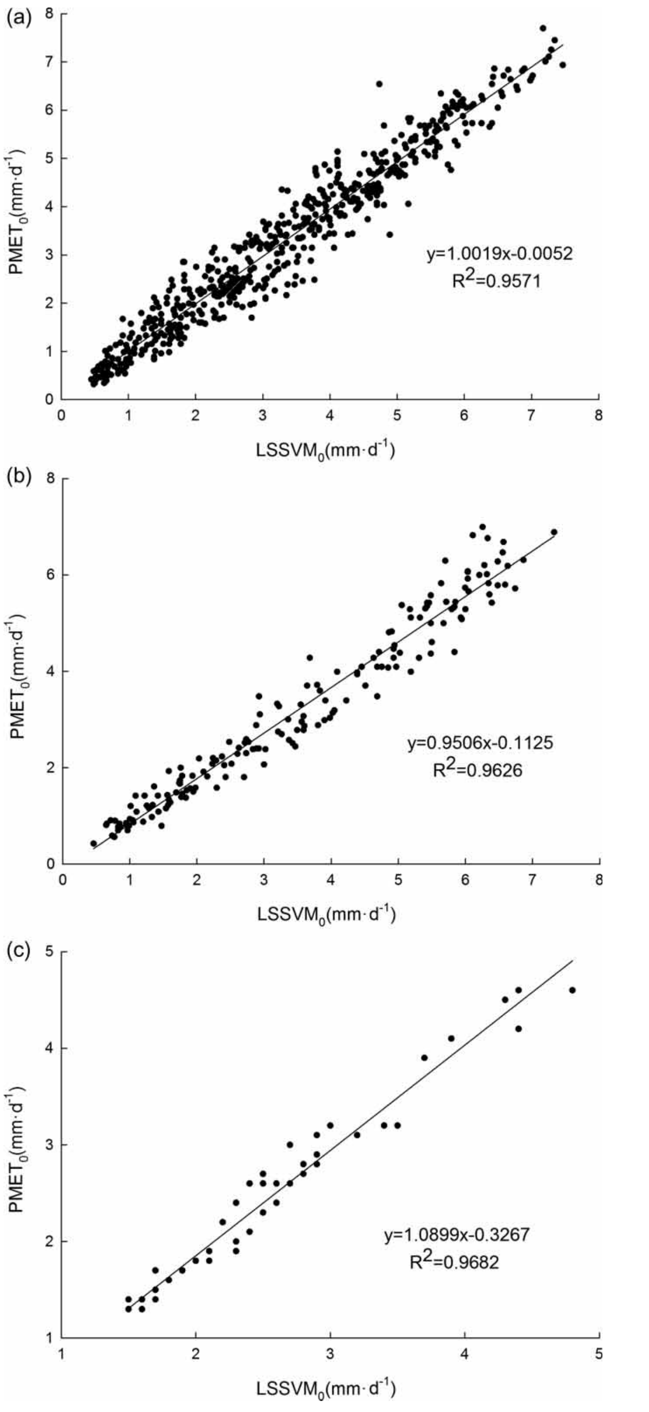

Figure 2 | Scatter plot of the PM-estimated ET_0 and LSSVM-estimated ET_0: (a) training set; (b) test set 1; (c) test set 2.

and σ^2 were calculated to be 232.09 and 24.71, respectively. These values were applied in the LSSVM model used in this study.

All calculations implementing LSSVM were performed using the Matlab toolbox. The toolbox of LSSVM was downloaded from http://www.esat.kuleuven.ac.be/sista/lssvmlab/, which was provided by Suykens and his colleagues. The entire computational program was run in the software system Matlab 7.0 (Mathwork Inc.).

RESULTS AND DISCUSSION

Figure 1 compares the PM56-estimated ET_0 with the LSSVM-estimated ET_0 for the training set, test set 1 and test set 2. Figure 1 shows that the LSSVM-estimated ET_0 values correspond well with the PM56-estimated ET_0 values and follow the same trend in the training set, test set 1 and test set 2.

Figure 2 shows a scattered plot between the PM56-estimated ET_0 and the LSSVM-estimated ET_0 for training set, test set 1 and test set 2. Linear regression analysis was used to test the agreement between LSSVM-estimated ET_0 with PM-estimated ET_0. The regression equation had the form:

$$ET_0^P = b_0 + b_1 ET_0^L \tag{10}$$

where ET_0^P is the PM-estimated ET_0, ET_0^L is the LSSVM-estimated ET_0, b_0 is the intercept and b_1 is the slope. Values of the slope (b_1) in the regression equation, the intercept (b_0) and the coefficients of correlation between the PM56-estimated ET_0 and LSSVM-estimated ET_0 are presented in Table 3. The correlation coefficients are significantly high, the slope is close to unity and the intercept is close to zero for training set, test set 1 and test set 2. The LSSVM-estimated ET_0 using weather forecast data is therefore close to the PM56-estimated ET_0 using actual weather data.

Table 3 | Regression and correlation Coefficients

	b0	b1	Correlation coefficient
Training set	−0.005	1.002	0.978
Test set 1	−0.113	0.951	0.981
Test set 2	−0.327	1.090	0.984

Results were further analyzed using statistical indices. The statistical indices used in the analysis were root mean square error (RMSE), mean absolute error (MAE) and model efficiency (EF). The RMSE, MAE and EF are estimated as:

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{l} \frac{(ET_{0i}^{L} - ET_{0i}^{P})^2}{l}} \tag{11}$$

$$\mathrm{MAE} = \frac{1}{l}\sum_{i=1}^{l} \left|ET_{0i}^{L} - ET_{0i}^{P}\right| \tag{12}$$

$$\mathrm{EF} = 1 - \frac{\sum_{i=1}^{l} (ET_{0i}^{L} - ET_{0i}^{P})^2}{\sum_{i=1}^{l} (ET_{0i}^{P} - \overline{ET}_{0}^{P})^2} \tag{13}$$

where l is the total number of data, ET_{0i}^{L} is the ith LSSVM-estimated ET_0, ET_{0i}^{P} is the ith PM-estimated ET_0 and $\overline{ET}_{0}^{P}$ is the mean of PM estimated ET_0. Values of RMSE, MAE, and EF for the LSSVM model are calculated and presented in Table 4.

To compare the LSSVM- and PM56-estimated ET_0 model accuracy for weather forecast data input, values of RMSE, MAE and EF for the LSSVM model are also presented in Table 4. The RMSE and MAE values are lower for training set, test set 1 and test set 2, and the maximum values of RMSE and MAE are less than 0.53 and 0.46 mm d^{-1} for the LSSVM model and PM56* (PM56 with weather forecast data input), respectively. The EF is greater than 89%. However, the RMSE and MAE of the LSSVM model are less than that of PM56* and the EF of the LSSVM model is greater than that of PM56* for test set 2. These results indicate that LSSVM-estimated ET_0 has a higher precision than PM56* using weather forecast data.

Table 4 | Performance indices for LSSVM model and PM56*

	Models	RMSE(mm)	MAE(mm)	EF(%)
Training set	LSSVM	0.36	0.26	95.71
Test set 1	LSSVM	0.47	0.37	93.76
Test set 2	LSSVM	0.39	0.39	93.87
Test set 2	PM56*	0.52	0.46	89.93

Note: EMSE is root mean square error; MAE is mean absolute error (MAE); EF is the model efficiency; PM56*

CONCLUSIONS

The LSSVM was applied to estimate ET_0 using public weather forecast data. Model procedures were compared by ET_0 estimated from measured meteorological data. Results from this study indicate that ET_0 can be successfully estimated using public weather forecast through LSSVM approach. However, the study only used data from one region for a limited period, and further studies using more data may be required to strengthen these conclusions.

ACKNOWLEDGEMENTS

This research was supported by the National Natural Science Foundation of China (NSFC) Project No. 50979065, Shanxi Province Science and Technology key problem project of China No. 2007031069 as well as the college students' innovation and entrepreneurship projects of Taiyuan City Science and Technology Bureau (100115109).

REFERENCES

Allen, R.G. 1986 A Penman for all seasons. *J. Irrig. Drain. Eng.* **112** (4), 348–368.

Allen, R. G., Jensen, M. E., Wright, J. L. & Burman, R. D. 1989 Operational estimates of reference evapotranspiration. *J. Agron.* **81**, 650–662.

Allen, R. G., Pereira, L. S., Raes, D. & Smith, M. 1998 *Crop evapotranspiration: guidelines for computing crop requirements*. FAO Irrigation and Drainage Paper No. 56. FAO, Rome, Italy.

Allen, R. G., Clemmens, A. J., Burt, C. M., Solomon, K. & O' Halloran, T. 2005 Prediction accuracy for project wide evapotranspiration using crop coefficients and reference evapotranspiration. *J. Irrig. Drain. Eng.* **131** (1), 24–36.

Allen, R. G., Pruitt, W. O., Wright, J. L., Howell, T. A., Ventura, F., Snyder, R., Itenfisu, D., Steduto, P., Berengena, J., Beselga, J., Smith, M., Pereira, L. S., Raes, D., Perrier, A., Alves, I., Walter, I. & Elliott, R., 2006 A recommendation on standardized surface resistance for hourly calculation of reference ET0 by the FAO56 Penman–Monteith method. *Agric. Water Manage.* **81**, 1–22.

Cai, J., Liu, Y., Lei, T. & Xu, D. 2005 Daily reference evapotranspiration estimation from weather forecast messages. *Transactions of the CSAE* **21** (11), 11–15.

Chiew, F. H. S., Kamaladassa, N. N., Malano, H. M. & McMahon, T. A. 1995 Penman-Monteith, FAO-24 reference crop evapotranspiration and class-A pan data in Australia. *Agric. Water Manage*. **28**, 9–21.

De Souza, F. & Yoder, R. E. 1994 ET estimation in the north east of Brazil: Hargreaves or Penman-Monteith equation. Tech. Paper *ASAE Int. Winter Meeting*, American Society of Agricultural Engineers, St. Joseph, Mich.

Fletcher, R. 1987 *Practical Methods of Optimization*. John Wiley & Sons, Chichester.

Jensen, M. E., Burman, R. D. & Allen, R. G. 1990 *Evapotranspiration and Irrigation Water Requirements*. ASCE Manual and Report on Engineering Practice No. 70, ASCE, New York.

Gill, M. K., Asefa, T., Kemblowski, M. W. & McKee, M. 2006 Soil moisture prediction using support vector machines. *Journal of the American Water Resources Association* **42** (4), 1033–1046.

Li, B., Zheng, F., Long, D. & Li, J. 2007 Soil erosion and sediment prediction at watershed under single rainfall event based on support vector regression. *Bulletin of Soil and Water Conservation* **27** (6), 120–124.

Lin, J. & Cheng, C. 2006 Application of support vector machine method to long-term runoff forecast. *Journal of Hydraulic Engineering* **37** (6), 681–686.

Liong, S. & Chandrasekaran, S. 2002 Flood stage forecasting with support vector machines. *Journal of the American Water Resources Association* **38** (1), 173–186.

Suykens, J. A. K. & Vandewalle, J. 1999 Least squares support vector machine classifiers. *Neural Processing Letters* **9** (3), 293–300.

Suykens, J. A. K. & Vandewalle, J. 2000 Recurrent least squares support vector machines. *IEEE Transactions on Circuits and Systems-I* **47** (7), 1109–1114.

Vapnik, V. 1995 *The Nature of Statistical Learning Theory*, Springer, New York.

Xiong, J. & Li, Z. 2005 Sediment-carrying capacity forecasting based on support vector machine. *Journal of Hydraulic Engineering* **36** (10), 1–6.

Xu, J., Peng, S., Zheng, R. & Li, D. 2006 Neural network model for reference crop evapotranspiration prediction based on weather forecast. *Journal of Hydraulic Engineering* **37** (3), 376–379.

First received 30 June 2009; accepted in revised form 7 July 2010