Annals of Warsaw University of Life Sciences – SGGW
Land Reclamation No 45 (1), 2013: 71–83
(Ann. Warsaw Univ. of Life Sci. – SGGW, Land Reclam. 45 (1), 2013)

Using support vector regression to predict direct runoff, base flow and total flow in a mountainous watershed with limited data in Uttaranchal, India

JAN ADAMOWSKI
Department of Bioresource Engineering, Faculty of Agricultural and Environmental Sciences, McGill University

Abstract: *Using support vector regression to predict direct runoff, base flow and total flow in a mountainous watershed with limited data in Uttaranchal, India.* In the ecologically sensitive Himalayan region, land transformations and utilization of natural resources have modified water flow patterns. To ascertain future sustainable water supply it is necessary to predict water flow from the watersheds as affected by rainfall and morphological parameters. Although such predictions may be made using available process--based models, in mountainous and hilly areas it is extremely difficult to determine the numerous parameters needed to run such models, thus limiting their applicability. Artificial intelligence (AI) based models are a possible alternative in such circumstances. In this study an AI technique, support vector machines (SVM), was used for modeling the rainfall-runoff relationship from three hilly watersheds in the state of Uttaranchal, India. Different SVM models were developed to predict direct runoff, base flow, and total flow based on the daily rainfall, runoff, and morphological parameters collected from each watershed. The results confirm the potential of SVM models in the prediction of runoff, base flow, and total flow in hilly areas.

Key words: support vector machine, artificial intelligence, runoff, base flow, total flow, Himalayas, watersheds

INTRODUCTION

In the Himalayan region, agricultural activities are the main source of the population's livelihood. While these activities and other land uses have supported the population they have also placed significant pressure on the land, water, and other natural resources creating ecological imbalances. This poses a serious threat to the sustainable development of the region's water resources (Samra et al. 1999). Given this situation, it is necessary to understand the water flow behaviour in this region.

Rainfall-runoff relationships are a complex hydrological phenomenon influenced by temporal and spatial variability in the watershed characteristics, uncertainty in rainfall patterns, and changes in soil cover and morphological parameters (Tokar and Johnson 1999). To simulate the behaviour of this complex system several conceptual process--based models have been developed to mathematically simulate rainfall-runoff processes on a watershed scale. These models are constructed to approximate the general internal sub-process and physical mechanisms that govern the hydrologic cycle, and are based on important hydrological processes. However, it is difficult to translate these processes into mathematical form and in practical situations such models may not be accurate at prediction as they may require

several input parameters that cannot be accurately determined due to spatial and temporal variability. In addition, in hilly regions it is often extremely difficult to collect information due to problematic topography. In such situations, it may be preferable to consider a direct mapping between the readily measurable inputs and outputs by implementing a simpler data driven model with little consideration of the complex processes.

Artificial intelligence (AI) methods such as artificial neural network (ANN) models may be used in runoff prediction without prior knowledge of the actual complex processes involved. A number of researchers have used ANN models for studying rainfall-runoff relationships and found promising results compared to conceptual models (Smith and Eli 1996; LeRoy et al. 1996; Tokar and Johnson 1999; Gautam et al. 2000; Dibike et al. 2001; Jain and Prasad Indurthy 2003; Castellano-Méndez et al. 2004; Nilsson et al. 2005; Adamowski 2008; Adamowski and Sun 2010). Support vector machines (SVM) are another AI method that has recently been employed for solving hydrological problems. Mukherjee et al. (1997) applied SVM regression on chaotic time series and compared the results with those obtained with different prediction methods such as polynomial, rational, local polynomial, multi-quadrics radial basis functions and neural networks. They reported that SVM provided significantly better results as compared to other methods. Dibike et al. (2001) developed both ANN and SVM to predict stream flow discharge at the watershed level using daily rainfall and evaporation as inputs, reporting a 15% increase in accuracy of runoff estimation with the SVM model over the ANN model. Bray and Han (2004) explored the applicability of the SVM model for flood forecasting, concentrating on determining a suitable model structure and appropriate parameter values for rain-fall-runoff modeling. They addressed the complexity of the SVM optimization with manual based operation of the method and concluded that for appropriate and effective application of SVM in rainfall-runoff modeling more research is needed. Asefa et al. (2005) used SVM models to predict at seasonal and hourly time scales, reporting promising SVM model performance. Behzad et al. (2009) compared SVM against ANN and ANN-GA models and reported the prediction accuracy of SVM to be as good as or better than those models. Wang et al. (2009) developed SVM and three other artificial intelligence models for the same data set, comparing their predictive ability and reporting strong predictive accuracy from the SVM model.

The overall goal of this paper is to investigate the applicability of SVMs in the prediction of runoff, base flow, and total flow for hilly watersheds. Three small watersheds in the Uttaranchal state of India, located in the mid-Himalayan region, were selected and their rainfall and morphological data used to build and test the models. The data for the three watersheds were randomized for training and testing, thus building models that could generalize predictions for similar geographic/climatic watersheds.

MATERIAL AND METHODS

Site description and data

The study data for this research were three watersheds located in the hilly terrain of Uttaranchal, India. The Central Soil and Water Conservation Research and Training Institute in Dehradun, Uttaranchal recorded the necessary watershed data. The three watersheds have been denoted as WS1, WS2 and WS3 with areas of 255, 52 and 163 ha, respectively. All three watersheds have varying morphological characteristics with similar steep slopes of 62–66%. WS1 is predominantly mixed forest and scrub, WS3 is mainly a forested watershed, while WS2 consists mainly of agricultural land. More detailed information on these watersheds is given in Table 1. Total flow of all three watersheds was recorded at the outlet with runoff and base flow computed from total flow. Data were recorded on a daily basis for the three-year period between July 1, 2001 and June 30, 2004.

Model description

SVMs are based on Vapnik's statistical learning theory (Vapnik 1995). Let us consider a simple problem where the data set has a linear relationship with M observations. Each observation consists of a pair: a vector $x_i \in R^n$; $i = 1, ..., M$, and the corresponding response variable y_i. The final objective of SVM regression is to develop a linear function that can make the best approximation of the dependent response variable. The function can be formulated as follows:

$$y = f(x) = < w \cdot x > + b \quad (1)$$

where:
w, b – regression parameters,
$< w \cdot x >$ – the dot product of w and x.

TABLE 1. General characteristics of the three Sainji watersheds

Category	Watershed characteristics	WS1	WS2	WS3
General features	area [ha]	255	52	163
	length [m]	2,950	1,360	2,100
	relief [m]	1,020	635	870
Shape indicators	circulatory ratio [-]	0.553	0.704	0.705
	compactness coefficient [-]	1.34	1.19	1.18
	elongation ratio [-]	0.610	0.598	0.686
Drainage pattern	drainage density [km/km^2]	2.76	3.83	2.2
	time of concentration [min]	14	6.76	9.86
	length of streams [m]	7,050	2,010	3,595
	main channel length [m]	2,950	1,360	2,100
Landuse pattern	agriculture [%]	16.55	22.94	14.87
	forest [%]	36.53	0.64	54.01
	scrubs [%]	46.92	76.42	29.12
Hydrologic soil cover complex	weighted curve number	64.99	69.57	62.57

The optimal regression function can be obtained, according to Gunn (1998) and Cristianini and Shawe-Taylor (2000), by minimizing a function, Ψ, as follows:

minimize

$$\Psi(w,\lambda) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{M}(\lambda_i^- + \lambda_i^+) \quad (2)$$

such that $(w \cdot x_i + b - y_i) + \lambda_i^+ \geq \varepsilon$

where:
C – regularization constant,
λ_i^-, λ_i^+ – slack variables that represent the upper and lower constraint on the regression function.

To optimize this function SVM regression uses a loss function that shows the maximum allowed deviation of the predicted values from the measured one. Some of the commonly used loss functions are Quadratic, Laplace, Huber, and ε-insensitive (Gunn 1998). Among these, the ε-insensitive loss function was proposed by Vapnik (1995) as a robust loss function to reduce sensitivity to the outliers by focusing on optimizing a bound around the regression function. For this study, the ε-insensitive loss function was selected. A SVM regression model based on this function calculates the difference between the predicted and the actual values, and if the differences are less than ε, the regression function is considered to be acceptable (Schmola and Scholkopf 1998).

Using Lagrangian multipliers, the solution to the optimization problem of equation (2) can be written as follows:

minimize

$$\varepsilon\sum_{i=1}^{M}(\alpha_i + \alpha_i^*) + \\ + \frac{1}{2}\sum_{i,j=1}^{M}\left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right)(x_i \cdot xj) + \\ + \sum_{i=1}^{M} y_i(\alpha_i - \alpha_i^*) \quad (3)$$

subject to:

$$\sum_{i=1}^{M}(\alpha_i - \alpha_i^*) = 0$$

and $0 \leq \alpha_i, \alpha_i^* \leq C$

where α_i, α_i^* – the Lagrange multipliers.

To handle non-linear regression cases the data is linearized by mapping it into a higher dimensional space using Lagrange transformations incorporating kernel functions, so that linear regression functions can be applied. The commonly used kernels are the radial basis function (RBF) kernels, sigmoid kernels, and polynomial kernels (Gunn 1998; Chang and Lin 2001). The RBF kernel, most commonly used in SVM approaches, is defined as follows:

$$K(x,y) = e^{-\gamma(x-y)^2} \quad (4)$$

where γ – kernel parameter.

Using the kernel function, the above-mentioned optimization function equation can be rewritten as:

minimize

$$\frac{1}{2}\sum_{i,j=1}^{M}\left(\alpha_i-\alpha_i^*\right)\left(\alpha_j-\alpha_j^*\right)K(x_i,x_j)+$$
$$+\sum_{i=1}^{M}y_i(\alpha_i-\alpha_i^*)+\varepsilon\sum_{i=1}^{M}(\alpha_i+\alpha_i^*) \quad (5)$$

subject to

$$\sum_{i=1}^{M}(\alpha_i-\alpha_i^*)=0$$

and $0\le\alpha_i,\alpha_i^*\le C$

The solution of this problem will yield α_i and α_i^* for all i = 1,2, ..., M. It should be mentioned that all the training points within the ε-sensitive zone will yield α_i and α_i^* equal to zero. The type of kernel function to be used is selected by the user. The user also needs to adjust the kernel-specific parameters, the values of parameters γ, C, and ε. The selection of the optimal values of these parameters determines the success of the SVM approach for a given problem. For more detailed information, readers are referred to Vapnik (1995), Burges (1998), and Cristianini and Shawe-Taylor (2000).

The SVM regression model is trained using a portion of the data set (e.g., 80%) containing the dependent and independent variables. The remaining 20% of the data (unseen data) is used for testing the predictive accuracy or performance of the developed model. The model is run with different sets of values of γ, ε, and C, using the training data set, and the optimal values are selected by optimizing the cross-validation error using a five-fold cross-validation technique. Next, the SVM regression model is built based on these optimum values. The generalization ability and predictive accuracy of the model is determined by using the test data set.

Method description

In this study SVM models were trained using a set of data containing both independent and dependent variables. The data set contained seventeen independent variables (inputs), namely: day of the year, rainfall, antecedent precipitation index (API_5), watershed area, length of the watershed, relief, circulatory ratio (the ratio of the watershed area to the area of a circle of the same perimeter as that of the watershed), compactness coefficient (a ratio between the watershed perimeter and the circumference of a circle with the same area), elongation ratio (the ratio of the diameter of a circle with the same area as the watershed area to the maximum length of the watershed), drainage density (which is a ratio between the total length of the drainage channel to the drainage area of a watershed), time of concentration, length of streams, main channel length, percentage of agricultural area, percentage of forested area, percentage of scrubs area, and runoff calculated with the curve number method. The dependent variables (outputs) were runoff, base flow and total flow. Each output was modeled separately as the SVM structure allows modeling of one output at a time.

The data collected over the three watersheds were collated and randomized, generating data sets that did not represent a single watershed but rather the characteristics of all three. This allowed

development of a generalizing model, that if proven accurate would allow its use in predicting flow for ungauged watersheds of similar geographical and climatic characteristics where past rainfall/runoff are not available. This generalizing ability was based upon the work of Sharda et. al (2006) in determining the most important watershed features in the watersheds that affect the rainfall-runoff relationships (curve number, rainfall, antecedent moisture condition and day of the year).

The available data for modeling were limited to just three watersheds over a period of three years, which is not sufficient for modeling hydrologic behaviours. However, the available data were sufficient for evaluating the potential of the SVM method in modeling rainfall--runoff in hilly watersheds. For comprehensive model development, data from a greater number of watersheds over a longer period would need to be collected. Because of the small available data set a five-fold cross-validation procedure was applied to check the generalization ability of the model. In this procedure, the data was randomized and divided into five equal parts. The models were trained using four parts of the data (80%), and tested with the remaining "unseen" fifth part (20%) to evaluate model performance. The procedure was repeated for all five possible combinations.

The performance of each model was evaluated by regression analysis of the simulated results over observed data. The intercepts and the slopes of the best-fit regression lines were determined and compared with their ideal values of 0 and 1, respectively. Other statistical parameters such as, root mean square error (RMSE), mean bias error (MBE) and modeling efficiency (EF), were also used for the comparison of the estimated and measured data. RMSE represents the mean distance between measured and estimated data (Kobayashi and Salam 2000), and is generally very sensitive to extreme values. Mean bias error (MBE) displays various features of the overall deviation between estimated and measured data, and can show the general bias of the model predictions. The optimum value of MBE is zero and it can be either positive or negative. Modeling efficiency can be viewed as an indicator of the overall correspondence between the measured and estimated values, and explains the goodness of fit of the predicted values with observed values. In the case of biased data, if the data is strongly correlated, EF gives more acceptable analysis than R^2. A negative value of EF shows the model prediction to be very poor.

To further explore the ability of SVM models in predicting runoff, base flow and total flow models were developed using separate data categorized for years. Year-based models used two-year data sets for model development and the third "unseen" year data was used for model testing. This resulted in a total of three pairs of training and testing data sets.

RESULTS AND DISCUSSION

Prediction of runoff, base flow and total flow using the total data set

Runoff prediction

The statistical results from the five-fold cross-validation for surface runoff are presented in Table 2a. For both training and testing data sets, correlation coef-

TABLE 2. Summary of the results obtained from the SVM method applied on watershed data with the five-fold cross-validation procedure, for training and testing (randomized data)

Runoff												
Fold	Training						Testing					
	r	Slope	Inter-cept	RMSE	MBE	EF	r	Slope	Inter-cept	RMSE	MBE	EF
1	0.907	0.688	0.066	0.644	0.0294	0.800	0.956	0.865	0.019	0.682	–0.011	0.912
2	0.937	0.789	0.034	0.644	0.003	0.870	0.789	0.737	0.060	0.579	0.033	0.601
3	0.935	0.780	0.036	0.650	0.003	0.864	0.863	0.695	0.036	0.551	0.006	0.741
4	0.940	0.837	0.022	0.492	0.002	0.881	0.876	0.574	0.031	1.235	–0.055	0.719
5	0.930	0.770	0.027	0.679	–0.008	0.855	0.960	1.135	0.037	0.332	0.048	0.869
Base flow												
Fold	Training						Testing					
	r	Slope	Inter-cept	RMSE	MBE	EF	r	Slope	Inter-cept	RMSE	MBE	EF
1	0.827	0.662	0.442	0.435	–0.020	0.683	0.770	0.635	0.468	0.495	–0.043	0.586
2	0.823	0.654	0.458	0.444	–0.020	0.676	0.743	0.589	0.539	0.498	–0.017	0.599
3	0.822	0.649	0.469	0.443	–0.016	0.674	0.790	0.630	0.522	0.463	0.024	0.623
4	0.812	0.626	0.482	0.445	–0.029	0.656	0.814	0.617	0.495	0.481	–0.047	0.656
5	0.820	0.652	0.462	0.442	–0.018	0.672	0.783	0.554	0.567	0.485	–0.037	0.605
Total flow												
Fold	Training						Testing					
	r	Slope	Inter-cept	RMSE	MBE	EF	r	Slope	Inter-cept	RMSE	MBE	EF
1	0.925	0.756	0.414	0.692	0.052	0.843	0.937	0.855	0.231	0.888	–0.005	0.878
2	0.928	0.792	0.254	0.784	–0.064	0.854	0.835	0.779	0.321	0.740	–0.002	0.688
3	0.964	0.923	0.109	0.544	–0.009	0.930	0.822	0.907	0.179	0.856	0.045	0.595
4	0.930	0.839	0.199	0.632	–0.041	0.863	0.923	0.607	0.499	1.239	–0.135	0.779
5	0.915	0.776	0.301	0.842	–0.042	0.833	0.879	1.191	–0.234	0.867	–0.039	0.546

r – correlation coefficient, RMSE – root mean square error [mm], MBE – mean bias error [mm], EF – modelling efficiency.

ficients between the observed and predicted runoff were consistently high in the five-fold test (training: 0.91 to 0.94, testing: 0.79 to 0.96), highlighting the ability of the SVM method to learn the input-output relationship and predicting runoff. The value of the intercepts were close to the ideal value of 0, however, the slopes were slightly lower than the ideal value of 1, which showed that the model under-estimated values for the high runoff events. This may be explained by the small number of high runoff events in the training data. There were only 123 runoff events out of 3,288 data, with only five events having runoff higher than 20 mm, although this may be attributed both to the monsoonal nature of the wa-

tershed and difficulty in collecting data. The RMSE values for the training and testing sets were 0.49 to 0.68 mm and 0.33 to 1.24 mm, respectively. The corresponding MBE values were –0.01 to 0.03 mm and –0.06 to 0.05 mm. These reasonably low values indicate a close agreement between the observed and simulated runoff for both training and testing data sets (Table 2a). The EF values were high (0.80 to 0.88 for training and 0.62 to 0.92 for testing data), which confirm good model performance in runoff prediction.

Base flow prediction

The statistical results from the five-fold cross-validation for base flows are presented in Table 2b. The correlation coefficients for both training and testing data sets were consistently high (training: 0.81 to 0.83, testing: 0.74 to 0.81), which implies that the prediction of base flow is quite satisfactory. The low values of RMSE (0.44 to 0.45 mm and 0.46 to 0.50 mm for the training and testing sets, respectively) and MBE (–0.03 to –0.02 mm and –0.05 to 0.02 mm) indicate that the observed and simulated values of base flow were very well matched. The slope and the intercept values (0.63 to 0.66 for slope and 0.44 to 0.48 for intercept) of the best-fit lines indicate that the model slightly underestimated the base flow for higher base flow events and overestimated for lower base flow events. The moderate values of EF (0.66 to 0.68 and 0.59 to 0.66 for training and testing sets, respectively) confirm that the SVM model performance was acceptable in estimating base flow.

Total flow prediction

The statistical results from the five-fold cross-validation for total flows are presented in Table 2c and illustrated in Figures 1 and 2. The high value of the correlation coefficients for both training (0.92 to 0.96) and testing data sets (0.82 to 0.94) indicated that the SVM models consistently predicted total flows very well. The low values of RMSE (training: 0.54 to 0.84 mm, testing: 0.74 to 1.24 mm) and MBE (training: –0.06 to

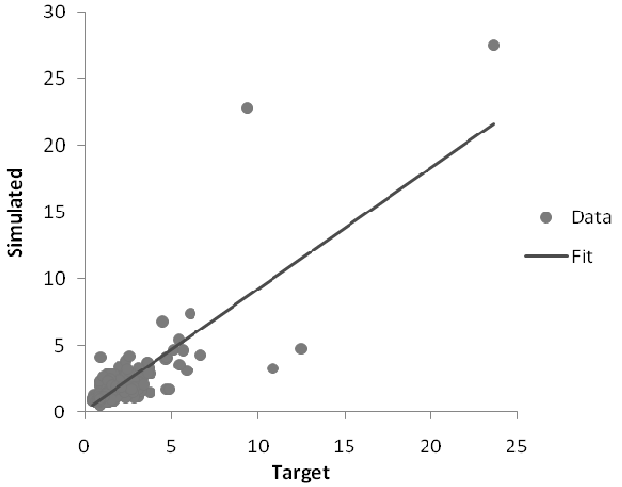

FIGURE 1. Total flow: simulated versus observed values [mm]

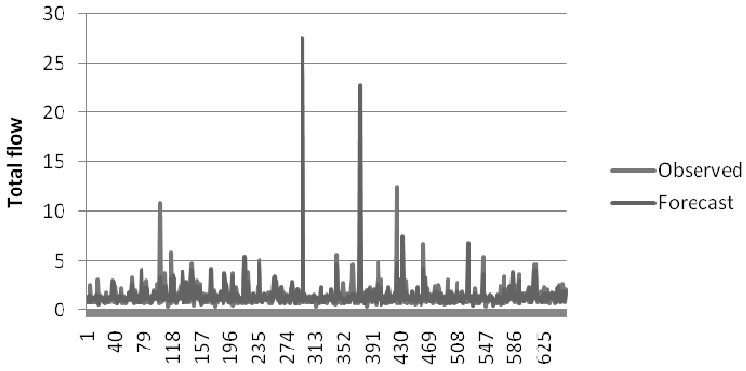

FIGURE 2. Observed and simulated total flow values plotted for each event [mm]

0.05, testing: –0.14 to 0.04 mm) along with the high values of EF (training: 0.83 to 0.93 mm, testing: 0.55 to 0.88) also indicate that the SVM model performance was acceptable in predicting total flows. Figure 1 plots the measured values against the predicted values, with the slope of the line of best fit illustrating the R value. Figure 2 illustrates the measured and predicted values for each flow event.

Prediction of runoff, base flow and total flow using year-based data

Runoff prediction

The statistical results from the model training and validation for runoff prediction using year-based data are reported in Table 3a. The correlation coefficients between the observed and predicted runoff for training data were high (above 0.93), with the slope and intercept values close to their ideal values for the best fit line. The RMSE, MBE and EF values were closer to their ideal values. Thus, the SVM models were able to develop a very good relationship in training between rainfall-runoff based on the rainfall and watershed morphological characteristics (Table 3a).

For the test data, the SVM models resulted in high correlation coefficients (above 0.72), but the slopes of the line of best fit were not very good for all the three possible combinations. The RMSE, MBE and EF values indicated a mismatch between the observed and predicted values of runoff. Such results may be explained by the limited data available in testing, accentuated when the model extrapolates for predictions.

Taking into account the complexity of the procedure of relating rainfall data to runoff, and the limited (three years) data, the results were in the expected range. It is evident that SVM models perform satisfactorily in predicting runoff when the training data set contains extreme runoff values (as was the case with year 1 and 3 for training and year 2 for testing).

Base flow prediction

The statistical results from the model training and validation for base flow prediction using year-based data are reported in Table 3b. For all three combinations consistently high correlation coefficients (above 0.80) were obtained

TABLE 3. Summary of the results obtained from the SVM method applied on watershed data with year-based training and testing data

Runoff													
Training							Testing						
Year	r	Slope	Inter-cept	RMSE	MBE	EF	Year	r	Slope	Inter-cept	RMSE	MBE	EF
Y1&Y2	0.978	0.941	0.303	0.421	0.297	0.915	Y3	0.727	0.164	0.324	1.703	0.167	0.271
Y1&Y3	0.930	0.755	–0.009	0.707	–0.043	0.851	Y2	0.903	0.664	0.009	0.554	–0.037	0.787
Y2&Y3	0.951	0.885	–0.046	0.516	–0.064	0.902	Y1	0.928	0.345	–0.012	1.105	–0.072	0.550
Base flow													
Training							Testing						
Year	r	Slope	Inter-cept	RMSE	MBE	EF	Year	r	Slope	Inter-cept	RMSE	MBE	EF
Y1&Y2	0.871	0.733	0.346	0.367	–0.007	0.757	Y3	0.607	0.369	0.801	0.661	–0.134	0.341
Y1&Y3	0.798	0.616	0.513	0.461	–0.017	0.636	Y2	0.794	0.665	0.429	0.479	–0.028	0.628
Y2&Y3	0.838	0.678	0.434	0.439	–0.021	0.700	Y1	0.750	0.619	0.585	0.474	0.098	0.537
Total flow													
Training							Testing						
Year	r	Slope	Inter-cept	RMSE	MBE	EF	Year	r	Slope	Inter-cept	RMSE	MBE	EF
Y1&Y2	0.968	0.889	0.273	0.462	0.114	0.930	Y3	0.659	0.197	1.184	1.880	–0.158	0.299
Y1&Y3	0.935	0.796	0.257	0.757	–0.053	0.866	Y2	0.885	0.782	0.284	0.753	–0.042	0.782
Y2&Y3	0.939	0.863	0.168	0.676	–0.050	0.881	Y1	0.855	0.416	0.865	0.676	–0.050	0.881

r – correlation coefficient, RMSE – root mean square error (mm), MBE – mean bias error (mm), EF – modelling efficiency.

between the observed and predicted base flow for training data. This shows that the SVM models were able to establish a reasonable link between rainfall-runoff events. However, for testing the correlation coefficient values were lower (0.61 to 0.79). Similarly, the slope and intercept values of the best-fit line, RMSE, MBE and EF for the training data sets were consistently closer to their ideal values, whereas this was not the case with the testing data sets.

Total flow prediction

The statistical results from the model training and validation for total flow prediction using year-based data are reported in Table 3c. The training data sets reported high correlation coefficients (0.94 to 0.97), along with slope and intercepts close to their ideal values. The good RMSE (less than 0.76 mm), MBE (less than 0.11 mm) and EF values (higher than 0.87) also confirmed that the SVM models were able to establish rainfall-total flow relationships in conjunction with the watershed rainfall and morphological characteristics during training. The results for the testing data sets were relatively poor compared to the training data sets; however, the results were satisfactory when the testing data were within the range of the data used for model training.

CONCLUSIONS

In this study it was found that SVM models have the capability to develop relationships between rainfall and watershed characteristics and runoff, base flow and total flow in hilly watersheds.

The results from the five-fold cross-validations for the training data sets were encouraging, and the results indicate that the models can learn the hidden relationships between the inputs and all the three outputs: runoff, base flow and total flow. The five-fold cross-validations also produced encouraging results when testing the data against "unseen" data.

For the development of SVM models on a yearly basis the training correlations produced were encouraging. However, when tested against "unseen" data, the SVM models did not produce very accurate results. This result may be explained by the smaller data set available for training of these models, and given the encouraging results from training and the five-fold validation tests there is reason to expect that more accurate SVM models for yearly basis may be developed given sufficient training data.

It is concluded that SVM models have the potential to be developed for and used in the accurate prediction of runoff, base flow and total flow in hilly areas. However, it is recommended that further work be undertaken. The utilization of larger data sets to further investigate yearly models in different watersheds needs to be explored. As well, uncertainty related to the forecasting needs to be explored, as does comparing SVM models with other state of the art forecasting methods such as wavelet neural network models.

Acknowledgements

Financial support for this study was partially provided by an NSERC Discovery Grant (RGPIN 382650-10), as well as a grant provided by the Cyprus Institute, both held by Jan Adamowski.

REFERENCES

ADAMOWSKI J. 2008. Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. *Journal of Hydrology* 353, 247–266.

ADAMOWSKI J., SUN K. 2010: Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology* 390, 85–91.

ASEFA T., KEMBLOWSKI M., MCKEE M., KHALIL A. 2005: Multi-time scale stream flow predictions: The support vector machines approach. *Journal of Hydrology* 318: 7–16.

BRAY M., HAN D. 2004: Identification of support vector machine for runoff modeling. *Journal of Hydroinformatics* 6 (4), 265–280.

BEHZAD M., ASGHARI K., EAZI M., PALHANG M. 2009: Generalization performance of support vector machines and neural networks in runoff modeling. *Expert Systems with Applications* 36, 7624–7629.

BURGES C. 1998. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery* 2 (2), 121–167.

CASTELLANO-MÉNDEZM.,GONZÁLEZ--MANTEIGAW.,FEBRERO-BANDEM., PRADA-SÀNCHEZ J.M. LOZANO--CALDEŔON R. 2004: Modelling of the monthly and daily behavior of the runoff of the Xallas River using Box-Jenkins and Neural Networks methods. *Journal of Hydrology* 296 (1): 38–58.

CHANG C., LIN C. 2001: *LIBSVM: a library for support vector machines*. http://www.csie.ntu.edu.tw/~cjlin/libsvm/. [Last Modification 4.03.2313].

CRISTIANINI N., SHAWE-TAYLOR J. 2000: *An Introduction to support vector machines and other Kernel-based learning methods*. Cambridge University Press, New York.

DIBIKE Y.B., VELICKOV S., SOLOMATINE D., ABBOTT M.B. 2001: Model induction with support vector machines. *ASCE Journal of Computing in Civil Engineering* 15 (3), 208–216.

GAUTAM M.R., WATANABE K., SEAGUSA H. 2000: Runoff analysis in humid forest catchment with Artificial Neural Network. *Journal of Hydrology* 235, 117––136.

GUNN S. 1998: *Support vector machines for classification and regression*. Technical report, ISIS, Department of Electronics and Computer Science, University of Southampton, Southhampton.

JAIN A., PRASAD INDURTHY S.K.V. 2003: Comparative analysis of event-based rainfall-runoff modeling techniques-deterministic, statistical, and Artificial Neural Networks. *Journal of Hydrologic Engineering* 8 (2), 93–98.

KOBAYASHI K., SALAM M.U. 2000: Comparing Simulated and measured values using mean squared deviation and its components. *Agronomy Journal* 92, 345–352.

LeROY P.N., TOKAR A.S., JOHNSON P.A. 1996: Stream hydrological and ecological response to climate change assessed with an Artificial Neural Network. *Limnology and Oceanography* 41 (5), 857–864.

MUKHERJEE S., OSUNA E., GIROSI F. 1997: *Nonlinear prediction of chaotic time series using support vector machines*. Proceedings of IEEE NNSP VII, 24–26. 09.1997 Amelia Island, 511–520.

NILSSON P., UVO C., BERNDTSSON R. 2005: Monthly runoff simulation: comparing and combining conceptual and neural network models. *Journal of Hydrology* 321 (4): 344–363.

SAMRA J.S., DHYANI B.S., SHARMA A.R. 1999: *Problems and prospects of natural resource management in Indian Himalayas – a base paper*. Hill and Mountain Agro-Ecosystem Directorate, NATP. CSWCRTI, 218 Kaulagarh Road, Dehradun.

SCHMOLA A. SCHOLKOPF A. 1998: *A Tutorial on Support Vector Regression*, NeuroCOLT2 Technical Report NC2-T-R-1998-030.

SHARDA V.N., PATEL R.M., PRASHER S.O., OJASVI P.R., PRAKASH C. 2006: Modeling runoff from middle Himalayan watersheds employing artificial intelligence techniques. *Journal of Agricultural Water Management* 83, 233–242.

SMITH J., ELI R.N. 1996: Neural-network models of rainfall-runoff process. *Journal of Water Resources Planning and Management* 121 (6), 499–608.

TOKAR A.S., JOHNSON P.A. 1999: Rainfall-runoff modeling using Artificial Neural networks. *Journal of Hydrologic Engineering* 4 (3), 232–239.

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

WANG W., CHAU K., CHENG C., QIU L. 2009: A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology* 374, 294–306.

Streszczenie: *Wykorzystanie wektorów wspierających w zależnościach regresyjnych do prognozowania odpływu bezpośredniego i całkowitego w zlewniach górskich przy ograniczonej liczbie danych w zlewni Uttaranchal, Indie.* Na obszarach wrażliwych, jakim są Himalaje, zmiany w wykorzystaniu powierzchni obszarów górskich oraz zasobów przyrodniczych modyfikują warunki kształtowania się odpływu. Dla zrównoważonego gospodarowania zasobami wodnymi w tym regionie koniecznym jest prognozowanie odpływu ze zlewni na podstawie opadu i warunków morfologicznych obszaru. Prognozowanie odpływu przy wykorzystaniu modeli deterministycznych jest dosyć trudne i ograniczone ze względu na trudności w identyfikacji wielu parametrów. W pracy zastosowano modele wykorzystujące techniki sztucznej inteligencji (AI) za pomocą wektorów wspierających (SVM) jako alternatywę do modelowania zależności opad-odpływ dla trzech zlewni górskich w stanie Uttaranchal, Indie. Wyniki zawarte w pracy potwierdzają możliwość wykorzystanie metody SVM do prognozowania charakterystycznych wielkości odpływu w warunkach górskich.

Słowa kluczowe: metoda wektorów wspierających, sztuczna inteligencja, modelowanie, odpływ, Himalaje, zlewnia górska

MS. received 28 June 2013

Author's address:
Department of Bioresource Engineering
Faculty of Agricultural and Environmental Sciences
McGill University
21111 Lakeshore Road, Ste-Anne de Bellevue, Que., Canada H9X 3V9
e-mail: jan.adamowski@mcgill.ca