Multi-step streamflow forecasting using data-driven non-linear methods in contrasting climate regimes

Daniel J. Karran, Efrat Morin and Jan Adamowski

ABSTRACT

Considering the popularity of using data-driven non-linear methods for forecasting streamflow, there has been no exploration of how well such models perform in climate regimes with differing hydrological characteristics, nor has the performance of these models, coupled with wavelet transforms, been compared for lead times of less than 1 month. This study compares the use of four different models, namely artificial neural networks (ANNs), support vector regression (SVR), wavelet-ANN, and wavelet-SVR in a Mediterranean, Oceanic, and Hemiboreal watershed. Model performance was tested for 1, 2 and 3 day forecasting lead times, measured by fractional standard error, the coefficient of determination, Nash–Sutcliffe model efficiency, multiplicative bias, probability of detection and false alarm rate. SVR based models performed best overall, but no one model outperformed the others in more than one watershed, suggesting that some models may be more suitable for certain types of data. Overall model performance varied greatly between climate regimes, suggesting that higher persistence and slower hydrological processes (i.e. snowmelt, glacial runoff, and subsurface flow) support reliable forecasting using daily and multi-day lead times.

Key words | artificial neural networks, climate regime, forecasting, streamflow, support vector regression, times series analysis

Daniel J. Karran
Jan Adamowski (corresponding author)
Department of Bioresource Engineering,
McGill University,
21 111 Lakeshore Road,
Ste. Anne de Bellevue,
QC,
Canada H9X 3V9
E-mail: *jan.adamowski@mcgill.ca*

Efrat Morin
Geography Department,
Hebrew University of Jerusalem, 91905,
Israel

INTRODUCTION

Forecasting daily streamflow with a reasonable level of accuracy plays a key role in the management of water resource systems. Reliable forecasts can be used as a tool by water authorities to more effectively allocate the resource among competing users (e.g. domestic, agriculture, environment, hydroelectric power), as well as to plan for the future expansion and/or reduction of water resources infrastructure. Forecasting is important in all climatic regions of the world; however, the performance of forecasting methods varies considerably depending on the characteristics of the watershed. For example, methods that are found effective for forecasting streamflow in relatively water abundant regions may, in fact, be unsuitable for use in dryer watersheds, where water scarcity is a reality due to the intermittent nature of streams. These climate characteristics, and others, may dramatically affect the performance of various forecasting methods in different watersheds and this area of research still requires much more exploration.

Due to the sensitivity of water resources in many areas around the world, it is becoming increasingly important to ensure that water is managed in a sustainable manner. To do this requires an understanding of stream flow dynamics, which are governed by various physical mechanisms acting on a wide range of temporal and spatial scales (Sivakumar 2003). Modeling these relationships can be done using either a physical, conceptual or data-driven approach. Although physical and conceptual models are good at providing physical interpretation and insight into watershed processes, they have been criticized for a number of reasons that include: being difficult to implement for real-time forecasting applications; requiring many different types of data that are often difficult to obtain; being difficult to construct; and, resulting in models that are overly complex, leading to

doi: 10.2166/hydro.2013.042

problems of over parameterization and equifinality (Beven 2006). This is in contrast to data-driven models, which have found appeal due to their minimum information requirements, rapid development times, simplicity, and accuracy in streamflow forecasting (Adamowski 2008a). That being said, data-driven models do have limitations, which become increasingly apparent as the data become more complex.

When using a data-driven approach to forecast streamflow, one must consider that the data are usually non-linear and non-stationary. Traditional data-driven streamflow forecasts have used statistical models such as multiple linear regression (MLR) and autoregressive integrated moving average (ARIMA) models. Although MLR and ARIMA have been shown to perform fairly well for long-term forecasts (e.g. Krstanovic & Singh 1991; Modini 2000), both are limited by their inherent assumptions that the data are linear. Consequently, non-linear models that use machine learning techniques such as artificial neural networks (ANNs) and, more recently, support vector machines (SVMs) have found popularity for use in hydrologic forecasts.

ANNs used for hydrological applications were first reported by Daniel (1991), followed by Kang *et al.* (1993), who found the method useful for forecasting hourly and daily streamflows. Since then, there have been many studies to confirm the usefulness of ANNs in streamflow forecasting. The most popular type of ANN that appears in the literature is the multi-layer perceptron (MLP) optimized with a back propagation algorithm. Using the MLP, ANNs have been shown to produce improved short-term forecasts compared to ARIMA and MLR models (e.g. Abrahart & See 1998; Birikundavyi *et al.* 2002). ANNs have been used to forecast streamflow at a variety of lead times with monthly (Jain *et al.* 1999), weekly (Zealand *et al.* 1999), daily (Jeong & Kim 2005; Chen & Chang 2009; Tiwari *et al.* 2013) and hourly (Lekkas *et al.* 2001; Besaw *et al.* 2010) forecasts. Additional studies where ANNs have been used to forecast streamflow include Cigizoglu (2005), Kişi & Cigizoglu (2007), Kişi (2009), Adamowski & Sun (2010), and Abudu *et al.* (2011).

SVMs are a relatively new form of machine learning that was developed by Vapnik (1995) for use in the telecommunications industry. Early on, a type of SVM called support vector regression (SVR) was researched for its application in financial forecasting (Cao & Tay 2001) and electrical load forecasting (Mohandes 2002). SVR has since gained popularity for use in environmental applications. One of the first such uses of SVR was for air pollutant forecasting, where Lu *et al.* (2002) showed their effectiveness and Wang *et al.* (2003) showed that SVR outperformed the ANN models. There are also a number of studies where SVR was used in hydrological forecasting. Khan & Coulibaly (2006) found that an SVR performed better than MLP ANNs in 3–12 month predictions of lake water levels. Yu *et al.* (2006) were successful in using SVR for predicting flood stages with 1–6 hour lead times and Han *et al.* (2007) found that SVR performed better than other models for flood forecasting. Rajasekaran *et al.* (2008) used SVR successfully for storm surge predictions and Kişi & Çimen (2009) used SVR to estimate daily evapotranspiration. Finally, SVR has been successfully used to predict hourly streamflow by Asefa *et al.* (2006), and was shown to perform better than ANN and ARIMA models for monthly streamflow prediction by Wang *et al.* (2009) and Maity *et al.* (2010), respectively.

Even though machine learning methods have shown improved results over traditional methods, they have limitations processing non-stationary data. This has led to the recent creation of hybrid models, where data are pre-processed for non-stationary characteristics and then run through machine learning models to cope with the non-linearity. One of the most promising pre-processing techniques is the use of the wavelet transform, which can decompose a time series into a comprehensible time-frequency representation at different scales. Wavelet transforms are useful for identifying variability and trends in time series data. They have been used in hydrology in a variety of ways. Smith *et al.* (1998) used the discrete wavelet transform (DWT) to analyze streamflow variability, and Coulibaly & Burn (2004) did the same using the continuous wavelet transform (CWT). Khaliq *et al.* (2006) used wavelets for a frequency analysis on hydro-meteorological extremes. Adamowski (2007, 2008a,b) used the CWT and cross-wavelet analysis as a stand alone technique for streamflow and flood forecasting. Labat (2008) used wavelets to analyze the discharge of the world's largest rivers and Adamowski *et al.* (2009) developed a wavelet-aided technique for trend

detection in monthly streamflows. More recently, wavelet transforms have been used for the synthetic generation of streamflow by Wang *et al.* (2011) and to predict uncertainty in daily streamflows by Dhanya & Kumar (2011).

Wavelet transformation combined with machine learning methods has been shown to provide highly accurate and reliable short-term forecasts. The most researched hybrid model is the wavelet transform coupled with a wavelet-ANN (WANN). One of the earliest hydrological applications of the WANN model was by Kim & Valdes (2003) to forecast drought in the Conchos River Basin, Mexico. Since then, there have been a number of applications of the WANN in streamflow forecasting. Typical lead times for these studies usually range from daily (Anctil & Tape 2004) to monthly (Cannas *et al.* 2006; Partal 2009; Wei *et al.* 2012) flows and, to date, there have been several studies where WANN models have been used to forecast flows in intermittent streams (Kişi 2009; Adamowski & Sun 2010). In all of the aforementioned studies, the WANN models have outperformed the stand alone ANNs.

Another hybrid method that has recently been proposed is the wavelet transform coupled with wavelet-SVR (WSVR). To the best knowledge of the authors, there has been very little research into the application of this technique for streamflow forecasting. Kişi & Çimen (2011) and Guo *et al.* (2011) both applied WSVR models with different methodologies to forecast monthly streamflow, and both found that the WSVR models outperformed the stand alone SVR. Nevertheless, to date, there has been no research that: (1) explores the use of WSVR methods for streamflow forecasting of lead times less than 1 month; and (2) compares the WSVR method to other hybrid methods, most notably WANNs.

Finally, the majority of the aforementioned studies for forecasting streamflow use either lagged precipitation, lagged streamflow, or both as inputs, and rarely is temperature ever used except in a few cases. Furthermore, all of the studies test the methods in one or, at most, two watersheds that are always in the same climate regime, providing little insight into the performance of these methods when climatic conditions change. The goal of the current research was to compare the forecasting performance of hybrid methods (i.e. WANN and WSVR) and their stand alone counterparts (i.e. ANN and SVR) using multiple lead times with daily data in Mediterranean, Oceanic, and Hemiboreal climates.

STUDY AREAS

To allow for a more robust test of the different modeling methods, three climates – Mediterranean, Oceanic, and Hemiboreal – were selected because each has varying degrees of annual precipitation, streamflow, and temperature. All climates are classified according to the Koppen-Gieger climate classification system described by Peel *et al.* (2007) and the watersheds presented in Figure 1 were chosen because each existed in one of the selected climate regimes and there was at least 30 years of continuous records available. The watersheds selected include the Alexander Stream in Israel, the Koksilah River in British Columbia, Canada and the Upper Bow River in Alberta, Canada. Lead times of 1, 2 and 3 days were selected for this study because daily data were the only data available in all three watersheds. These lead times were also deemed appropriate when considering the size of the watersheds and some of the faster processes (i.e. surface runoff) that characterise streamflow in the Mediterranean watershed. Four sets of data were collected for each watershed, including: total daily streamflow volume (m^3), total daily precipitation (mm), minimum daily temperature (°C), and maximum daily temperature (°C). These variables were selected as inputs for the models as they typically provide a good representation of climate regime and have considerable influence on streamflow dynamics (Hurkmans *et al.* 2010). For each of the variables, 30 years of data were collected for the period 1970–1999, except for the Hemiboreal case study, where the period was 1965–1994. The rationale for choosing these time intervals pertained to the availability of data in each watershed. Each of the data sets was split into two: a training/validation set (first 80% of the data) and a testing set (last 20% of the data).

Figure 1 shows a map of each watershed and Table 1 shows the different characteristics of each watershed. Hydrographs for all of the watersheds during the selected time period are shown in Figure 2 and Tables 2–4 show

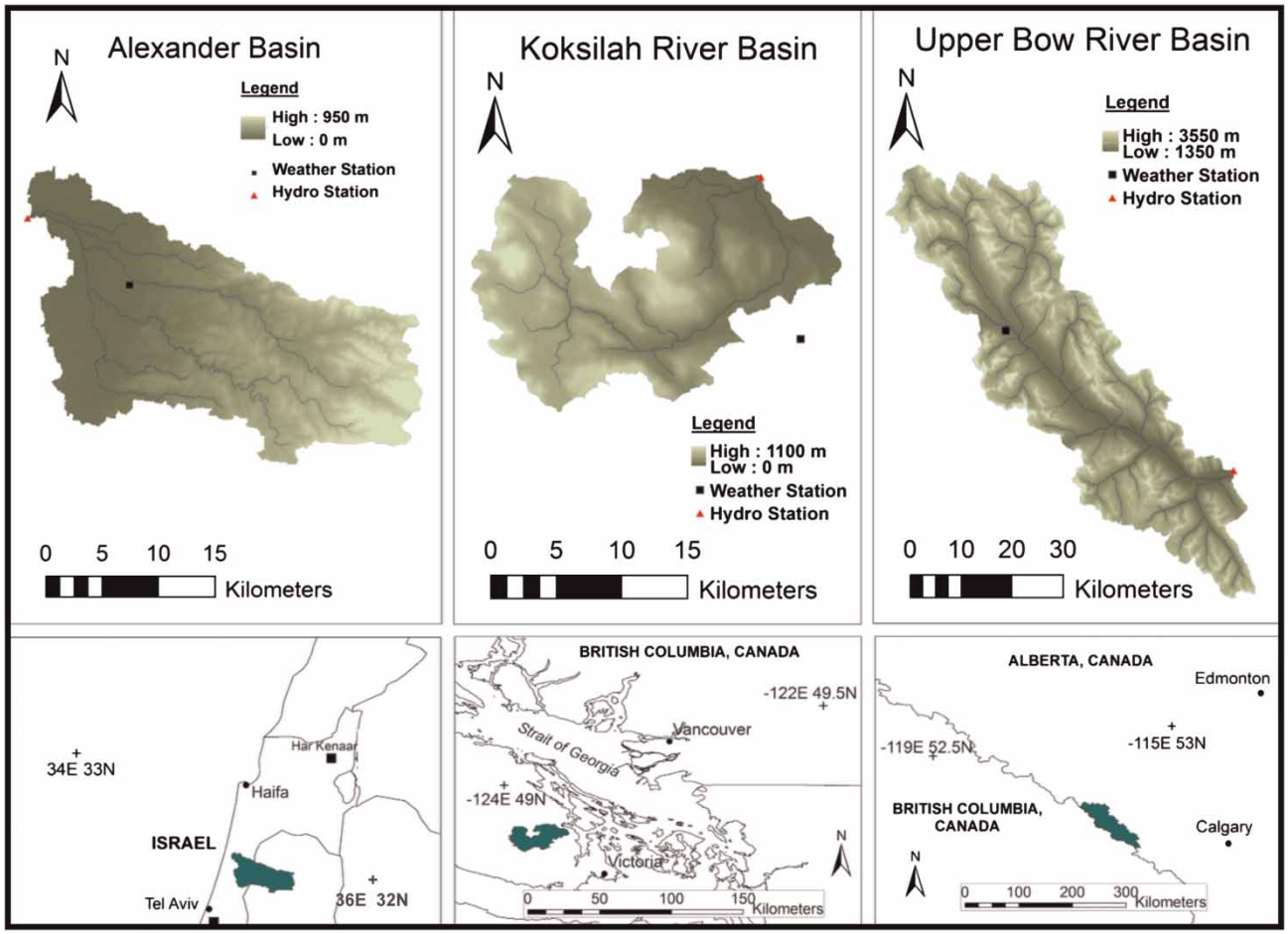

Figure 1 | Maps of the watersheds selected for this study and their corresponding locations and climate regimes.

the descriptive statistics for each of the variables in each of the watersheds.

The Mediterranean watershed was chosen for this study because it is intermittent and is subject to fast peaks and declines in the annual hydrograph. This is portrayed in Table 2 by the average daily streamflow minimums (0 m^3) and maximums (~9 m^3), and the relatively high skewness (19.90), respectively. Note how lagged precipitation shows the highest correlation with streamflow, even more so than the streamflow itself, suggesting that it is most dependent on precipitation, likely in the form of surface runoff. Lagged temperature shows the least correlation with streamflow out of all of the watersheds.

The Oceanic watershed was chosen because it is comparable in many ways to the Mediterranean (e.g. streamflow contributions); however, it is situated in a more temperate climate with more consistent precipitation during the year (Table 3). Although the Oceanic watershed is smaller in area than the Mediterranean watershed, it carries a far greater volume of streamflow with minimums averaging 950 m^3 and maximums averaging ~23 million m^3. The peaks and declines in this watershed are much more gradual than in the Mediterranean watershed, shown by the reduced skewness (4.75), and there is much more correlation between all of the lagged variables and streamflow overall.

Finally, the Hemiboreal watershed was selected for this study because it differs from the other two watersheds in that a significant amount of streamflow originates from snowpack, and there is continuous ice cover during the winter months. It is the largest watershed and it carries the most streamflow (Table 4) throughout the year, with

Table 1 | Characteristics of each watershed studied

Climate regime[a]	Watershed name	Annual flow characteristics	Origin of streamflow	Data collection location
Mediterranean (Csa)	Alexander Stream, Central Israel	The stream is intermittent; it typically only flows during the rainy months of November through April.	Contributions to the stream are entirely from runoff; however, in the last 3 years of the data set, small volumes of wastewater were discharged into the stream.	Hydrometric Elyashiv station Precipitation Yad Hanna station Temperature Bet Dagan
Oceanic (Cfb)	Koksilah River, British Columbia, Canada	Continuous flow throughout the year. Peak flows occur from November to January as a result of heavy rainfall.	The contributions to the river are mostly from rain and snowfall which accumulates during the winter. Contributions enter the river either directly from surface runoff or are released more slowly via groundwater aquifers.	Hydrometric Cowichan station Precipitation Shawnigan Lake station Temperature Shawnigan Lake station
Hemiboreal (Dfb)	The Upper Bow River, Alberta, Canada	Continuous flow throughout the year. Peak flows occur in June and low flows in January. During the winter months there is continuous ice covering the river.	The river contributions originate from snowpack (80%) and rainwater (20%) (Bow River Project Research Consortium 2010).	Hydrometric Banff station Precipitation Lake Louise station Temperature Lake Louise station

[a]Climate regimes classified according to Koppen-Geiger climate classification system described by Peel *et al.* (2007).

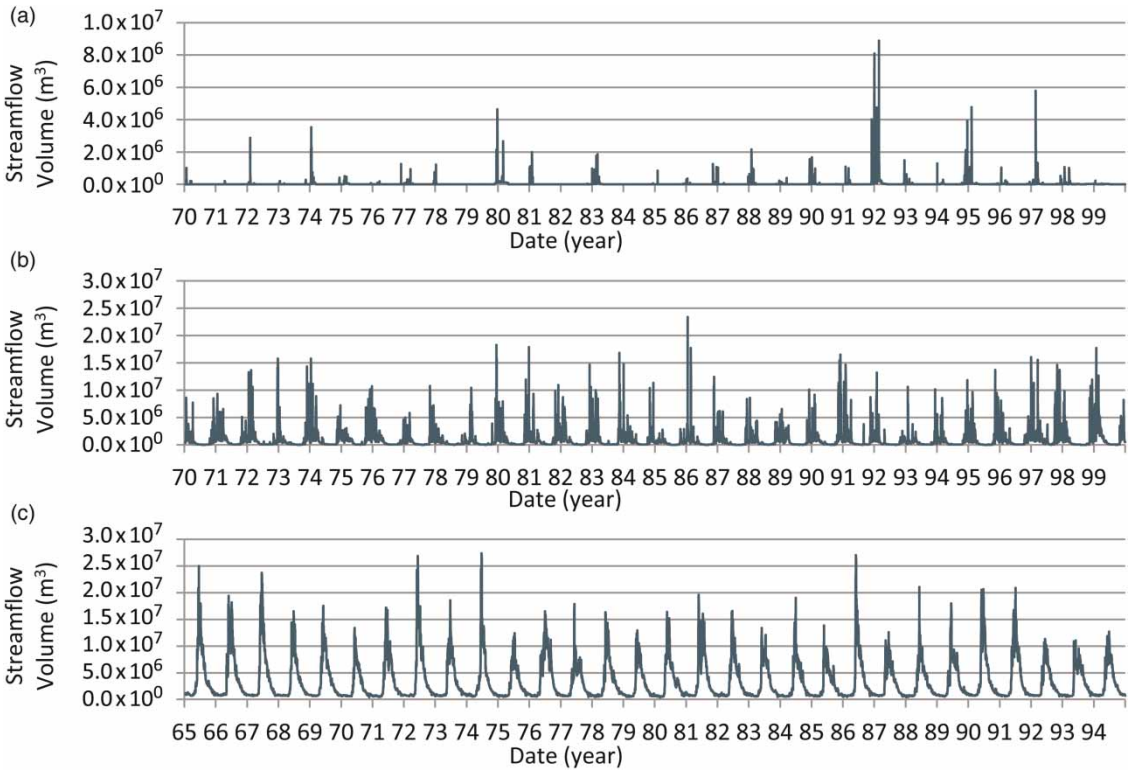

Figure 2 | Total daily streamflow (m^3) in the (a) Mediterranean, (b) Oceanic, and (c) Hemiboreal watersheds for the 30 year study period. Note the change in y-axis between watersheds.

Table 2 | Descriptive statistics and lagged correlation coefficients for the variables in Mediterranean watershed

Catchment area = 492 km²		90th percentile of total daily streamflow in testing period (10^6 m³) = 3.29×0^{-2}							
Variable	**Data sets**	x_{mean}	x_{max}	x_{min}	s_x	c_{sx}	$r_{(t-1)}$	$r_{(t-2)}$	$r_{(t-3)}$
Daily flow volume (10^6 m³)	Training	2.67×10^{-2}	8.90	0.00	0.22	21.11	0.53	0.20	0.15
	Testing	3.61×10^{-2}	5.80	0.00	0.16	16.20	0.54	0.12	0.06
	Entire	2.86×10^{-2}	8.90	0.00	0.23	19.90	0.53	0.18	0.13
Daily precipitation (mm)	Training	1.73	126	0.00	7.04	4.23	0.58	0.30	0.17
	Testing	1.72	121	0.00	7.58	3.79	0.70	0.37	0.14
	Entire	1.73	126	0.00	7.15	4.13	0.61	0.31	0.17
Minimum daily temperature (°C)	Training	13.19	27.00	− 2.20	5.53	− 0.14	− 0.11	− 0.11	− 0.12
	Testing	14.76	26.80	− 0.80	5.73	− 0.11	− 0.11	− 0.11	− 0.12
	Entire	13.50	27.00	− 2.20	5.61	− 0.12	− 0.11	− 0.11	− 0.12
Maximum daily temperature (°C)	Training	25.12	45.60	7.10	5.67	0.19	− 0.20	− 0.17	− 0.16
	Testing	25.52	42.00	10.80	5.65	0.22	− 0.21	− 0.18	− 0.15
	Entire	25.20	45.60	7.10	5.67	0.20	− 0.20	− 0.17	− 0.16

x_{mean} = Daily mean; x_{max} = Daily maximum; $x_{min=}$Daily minimum; s_x = Standard deviation; c_{sx} = Skewness.
$r_{(t-1)}$ = Correlation coefficient between variable time series delayed 1 day and total daily streamflow at time (t).
$r_{(t-2)}$ = Correlation coefficient between variable time series delayed 2 days and total daily streamflow at time (t).
$r_{(t-3)}$ = Correlation coefficient between variable time series delayed 3 days and total daily streamflow at time (t).

Table 3 | Descriptive statistics and lagged correlation coefficients for the variables of the Oceanic watershed

Catchment area = 209 km²		90th percentile of total daily streamflow in testing period (10^6 m³) = 2.89							
Variable	**Data sets**	x_{mean}	x_{max}	x_{min}	s_x	c_{sx}	$r_{(t-1)}$	$r_{(t-2)}$	$r_{(t-3)}$
Daily flow volume (10^6 m³)	Training	0.78	23.41	9.50×10^{-3}	1.49	5.25	0.71	0.52	0.44
	Testing	1.08	17.71	1.24×10^{-2}	1.83	3.49	0.78	0.59	0.49
	Entire	0.84	23.41	9.50×10^{-3}	1.57	4.75	0.73	0.54	0.46
Daily precipitation (mm)	Training	3.30	95.6	0.00	7.61	4.23	0.69	0.52	0.40
	Testing	4.05	91.6	0.00	8.79	3.79	0.69	0.59	0.43
	Entire	3.45	95.6	0.00	7.87	4.13	0.69	0.54	0.41
Minimum daily temperature (°C)	Training	5.17	20.00	− 15.00	5.27	− 0.22	− 0.24	− 0.27	− 0.30
	Testing	5.49	18.50	− 13.50	5.12	− 0.16	− 0.27	− 0.32	− 0.36
	Entire	5.24	20.00	− 15.00	5.24	− 0.21	− 0.24	− 0.28	− 0.31
Maximum daily temperature (°C)	Training	13.75	36.00	− 11.00	7.26	0.19	− 0.32	− 0.35	− 0.37
	Testing	14.54	33.50	− 5.00	7.27	0.22	− 0.38	− 0.41	− 0.44
	Entire	13.90	36.00	− 11.00	7.27	0.20	− 0.33	− 0.36	− 0.38

x_{mean} = Daily mean; x_{max} = Daily maximum; x_{min} = Daily minimum; s_x = Standard deviation; c_{sx} = Skewness.
$r_{(t-1)}$ = Correlation coefficient between variable time series delayed 1 day and total daily streamflow at time (t).
$r_{(t-2)}$ = Correlation coefficient between variable time series delayed 2 days and total daily streamflow at time (t).
$r_{(t-3)}$ = Correlation coefficient between variable time series delayed 3 days and total daily streamflow at time (t).

average minimums of 360,000 m³ and maximums of ~27 million m³. The peak flows and declines are much slower than the other watersheds as it is the least skewed (1.97), and all the lagged variables are strongly correlated with streamflow, except for precipitation, where there is very little correlation at all. This demonstrates how important temperature is in this watershed because precipitation that falls during the winter months, when temperatures are below freezing, is not making a meaningful contribution to streamflow.

Table 4 | Descriptive statistics and lagged correlation coefficients for the variables of the Hemiboreal watershed

Catchment area = 2,209 km²		90th percentile of total daily streamflow in testing period (10^6 m³) = 8.38							
Variable	**Data sets**	x_{mean}	x_{max}	x_{min}	s_x	c_{sx}	$r_{(t-1)}$	$r_{(t-2)}$	$r_{(t-3)}$
Daily flow volume (10^6 m³)	Training	3.38	27.39	0.36	3.90	2.01	0.99	0.97	0.94
	Testing	3.32	20.91	0.49	3.60	1.73	0.99	0.97	0.95
	Entire	3.36	27.39	0.36	3.84	1.97	0.99	0.97	0.94
Daily precipitation (mm)	Training	1.58	53.50	0.00	3.66	4.06	0.09	0.08	0.06
	Testing	1.57	46.00	0.00	3.48	4.38	0.05	0.05	0.03
	Entire	1.58	53.50	0.00	3.63	4.12	0.08	0.07	0.05
Minimum daily temperature (°C)	Training	− 7.95	13.00	− 47.00	11.02	− 0.82	0.59	0.59	0.58
	Testing	− 7.69	12.00	− 45.00	10.99	− 0.86	0.63	0.63	0.62
	Entire	− 7.89	13.00	− 47.00	11.01	− 0.83	0.60	0.60	0.59
Maximum daily temperature (°C)	Training	6.91	31.70	− 32.20	11.18	0.19	0.64	0.65	0.65
	Testing	7.40	31.00	− 31.00	11.17	0.22	0.64	0.65	0.65
	Entire	7.01	31.70	− 32.20	11.18	0.20	0.64	0.65	0.65

x_{mean} = Daily mean; x_{max} = Daily maximum; x_{min}=Daily minimum; s_x = Standard deviation; c_{sx} = Skewness.
$r_{(t-1)}$ = Correlation coefficient between variable time series delayed 1 day and total daily streamflow at time (t).
$r_{(t-2)}$ = Correlation coefficient between variable time series delayed 2 days and total daily streamflow at time (t).
$r_{(t-3)}$ = Correlation coefficient between variable time series delayed 3 days and total daily streamflow at time (t).

MODEL DEVELOPMENT

Four different types of model were developed in this study: ANNs, WANNs, SVRs, and WSVRs. For each of the four variables discussed in the previous section (i.e. total streamflow volume, total precipitation, minimum daily temperature, maximum daily temperature), two sets of inputs were created: the variables themselves delayed by 1 ($t-1$), 2 ($t-2$), and 3 ($t-3$) days; and, the same delayed variables decomposed by wavelet transformation into their respective high- and low-frequency components. The rationale for delaying the variables and wavelet sub-time series up to 3 days originated from studies that show improved model performance when using such a procedure with precipitation (Anctil & Tape 2004) and streamflow (Kişi 2008; Kişi & Çimen 2011) variables. The delayed variables become the inputs for the ANNs and SVRs, whereas, the delayed wavelet sub-time series are the inputs for the WANNs and WSVRs.

The creation of wavelet sub-time series inputs for the WANN and WSVR models

DWTs were used in this study to decompose the original time series into a time-frequency representation at different scales (wavelet sub-times series). This area of study has been widely researched over the past two decades; however, readers who are new to this area of research are directed to Shensha (1992) for more details on the theoretical background of wavelet transformation.

The number of decomposition levels was selected according to signal length (Wang & Ding 2003; Nourani *et al.* 2008; Tiwari & Chatterjee 2010), given by:

$$L = \text{int}[\log(N)] \quad (1)$$

where L is the level of decomposition and N is the length of the signal. In this study, the training and validation set comprises $N = 8752$ samples; therefore, the level decomposition is $L = 3$. Considering that $\log(N)$ is much closer to 4 than it is to 3, the models were also tested using $L = 4$, but under these conditions, model performance decreased in all cases, suggesting that $L = 3$ is a more appropriate level of decomposition.

The signals were decomposed using the redundant à trous algorithm according to Murtagh *et al.* (2003) in conjunction with the non-symmetric Haar wavelet as the low-pass filter. Four sets of wavelet sub-time series were created, including: a low-frequency component (Approximation) that uncovers the signal's trend, and three sets of high-frequency

components (Details), which reveal the periodicity at each dyadic scale of 2, 4 and 8 days.

The motivation for using the à trous algorithm with the Haar wavelet is to overcome two inherent problems with using DWTs for forecasting applications, namely 'shift variance' and the inclusion of future data as inputs to the models. Classical DWTs use 'decimation', or the retaining of one sample out of every two, to decompose a signal. This means that only half of the coefficients of the details are left at the current level and half of the coefficients of the approximation are recursively processed using high-pass and low-pass filters for coarser resolution levels. This has many advantages for some applications like compression but it presents a challenge for forecasting applications in that it makes the signal 'shift variant' (i.e. if we change the values at the beginning of our time series, all of the wavelet coefficients will change). To overcome this challenge, we use the redundant or non-decimated à trous algorithm. Instead of decimating the signal, the à trous algorithm dilates the mother wavelet by inserting zeros, thereby creating an approximation and detail component that are the same length as the original signal. Thus, changing any of the values in the time series will have no effect on any of the other values or subsequent details, and the original signal can always be recreated by simply summing the details with the smoothest approximation of the signal. Furthermore, the Haar wavelet was selected as the low-pass filter because it only uses data obtained previously in time to calculate each of the wavelet coefficients, ensuring that future information is not used in the model's forecasts. For more information on these problems, readers are directed to Renaud *et al.* (2002) and Maheswaren & Khosa (2012).

Finally, because the wavelet transformation necessitates a convolution of the Haar filter, careful attention must be given to the boundary conditions at the beginning of the newly created inputs. To compensate, the first eight samples of each newly created input were discarded so as to remove any coefficients that were not created entirely with original data.

The selection of significant inputs

The next step in the model development process (for all models) is to determine which inputs are significant. This is typically done with cross-correlation analysis (Sudheer *et al.* 2002; Partal & Kişi 2007; Tiwari & Chatterjee 2010). Specifically, an analysis is done on the correlation between each individual input and the single output of daily streamflow. Once the correlation between the inputs and outputs is found, a threshold must be set for what is significant and insignificant. In this study, it was found that setting the threshold too high eliminates too much valuable information from the signal, especially in the less correlated catchments. Therefore, the threshold for this study was set at $|r| = 0.10$; all inputs with a correlation above this threshold were considered significant and all below were discarded. The correlation coefficients between the undecomposed variables and the output are presented in Tables 2–4. The correlation coefficients between the wavelet sub-time series and the output are presented in Table 5, where A3, D1, D2, and D3 are the level 3 approximation, levels 1–3 details, respectively.

Randomization of the training/validation data set

Both the ANN and SVR models train in epochs and stop training when model performance does not continue to improve. Each epoch is a single presentation of all the input vectors matched to their corresponding targets (observed values) for both the training and validation data sets. In this study, it was found that 'how' the data were partitioned into the training and validation sets had a significant impact on model performance. To clarify, the inputs are arranged in a matrix where the rows are the individual variables and columns are the time series of each variable. Deciding how the data will be partitioned is a matter of assigning which column vectors will be used for training and which will be used for validation. When the column vectors were assigned randomly, performance was improved.

Finding the optimum random index was done with the SVR models due to the robustness of their output compared to ANNs. To clarify, if SVR models are given the same random index and same parameters, they will always train with the same number of epochs and produce the exact same results. This is unlike the ANN models, which produce results that vary as a consequence of their internal mechanics (discussed more below).

Table 5 | Correlation coefficients between lagged wavelet sub-time series for each variable in each watershed and observed total daily streamflow. Correlation coefficients in bold indicate those that exceed the significance threshold of 0.10

	Total daily streamflow			Total daily precipitation			Min daily temp			Max daily temp		
	$r_{(t-1)}$	$r_{(t-2)}$	$r_{(t-3)}$	$r_{(t-1)}$	$r_{(t-2)}$	$r_{(t-3)}$	$r_{(t-1)}$	$r_{(t-2)}$	$r_{(t-3)}$	$r_{(t-1)}$	$r_{(t-2)}$	$r_{(t-3)}$
MD												
A3	**0.30**	**0.21**	**0.19**	**0.37**	**0.24**	**0.18**	**−0.13**	**−0.13**	**−0.13**	**−0.18**	**−0.17**	**−0.17**
D1	0.04	−0.06	**−0.10**	−0.07	−0.04	−0.04	**0.10**	**0.11**	**0.10**	**0.13**	**0.13**	**0.14**
D2	0.02	−0.08	−0.07	0.07	0.00	−0.04	**0.11**	**0.11**	**0.11**	**0.12**	**0.13**	**0.13**
D3	−0.01	−0.06	−0.06	0.05	−0.02	−0.04	**0.12**	**0.12**	**0.13**	**0.13**	**0.13**	**0.13**
OC												
A3	**0.56**	**0.49**	**0.44**	**0.68**	**0.57**	**0.49**	**−0.36**	**−0.38**	**−0.39**	**−0.42**	**−0.43**	**−0.44**
D1	**−0.32**	**−0.31**	**−0.29**	**−0.27**	**−0.20**	**−0.20**	**0.28**	**0.30**	**0.32**	**0.35**	**0.37**	**0.38**
D2	**−0.23**	**−0.24**	**−0.22**	**−0.17**	**−0.18**	**−0.17**	**0.33**	**0.34**	**0.35**	**0.39**	**0.39**	**0.39**
D3	**−0.14**	**−0.16**	**−0.17**	−0.08	−0.09	**−0.11**	**0.39**	**0.38**	**0.37**	**0.42**	**0.41**	**0.40**
HB												
A3	**0.94**	**0.93**	**0.91**	**0.11**	**0.10**	0.09	**0.64**	**0.63**	**0.63**	**0.67**	**0.67**	**0.66**
D1	**−0.94**	**−0.92**	**−0.90**	−0.04	−0.02	−0.02	**−0.54**	**−0.53**	**−0.53**	**−0.62**	**−0.61**	**−0.60**
D2	**−0.88**	**−0.86**	**−0.85**	−0.02	−0.02	−0.02	**−0.52**	**−0.51**	**−0.51**	**−0.61**	**−0.59**	**−0.58**
D3	**−0.79**	**−0.79**	**−0.77**	−0.02	−0.02	−0.03	**−0.51**	**−0.51**	**−0.51**	**−0.57**	**−0.56**	**−0.55**

MD = Mediterranean; OC = Oceanic; HB = Hemiboreal.
A3 = Level 3 approximation; D1 = Level 1 details; D2 = Level 2 details; D3 = Level 3 details.
$r_{(t-1)}$ = Correlation coefficient between sub-time series delayed 1 day and total daily streamflow at time (t).
$r_{(t-2)}$ = Correlation coefficient between sub-time series delayed 2 days and total daily streamflow at time (t).
$r_{(t-3)}$ = Correlation coefficient between sub-time series delayed 3 days and total daily streamflow at time (t).

For each watershed, the SVR model was run 100 times, each time with a new randomized order of inputs. The observed values that the model uses as its target were also divided according to the same random index. The random index that produces the lowest testing root mean squared error (RMSE) was selected as the optimum index for each watershed and is the same one used to train the ANN, WANN, SVR, and WSVR models. RMSE is a measure of model precision, expressed as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}}\sum_{i=1}^{n}(y_i - \grave{y}_i)^2} \quad (2)$$

where y_i and $\grave{y}_\mathrm{i}$ are the observed and forecasted streamflows, respectively, and n is the number of samples. The model becomes more precise as the RMSE decreases, with a perfect forecasting capability expressed as RMSE = 0.

ANN models

The ANN used in this study was a feed forward MLP architecture trained with the Levenberg–Marquardt (LM) back propagation algorithm. MLPs are often used in hydrology due to their simplicity; consisting of an input layer, one or more hidden layers, and an output layer. The hidden layer contains the neuron-like processing elements that connect the input and output layers, which is described in more detail by Kim & Valdes (2003).

All regular ANN models (i.e. those without wavelet decomposed inputs) were constructed with the MATLAB® (v.7.10.0) ANN toolbox. The activation function for the hidden neuron is the tan-sigmoid function and the activation function for the output layer is the pure-linear function. The LM back propagation algorithm was selected to train all ANN models because of its efficiency and reduced computational time (Kişi 2009; Adamowski & Chan 2011), and its

ability to skip local minima better than other algorithms (Hagan & Menhaj 1994). The only important parameter to select when using the LM algorithm is the number of hidden neurons and this is usually decided by trial and error (Yuan *et al.* 2003).

There are up to a maximum of 12 inputs (4 variables × 3 lag periods) for each of the ANN models. The inputs that meet the significance criteria were normalized between −1 and 1 and the output data were normalized between 0 and 1 according to Equation (3) below:

$$y_{\text{norm}} = (y_{\max} - y_{\min}) * \frac{(x - x_{\min})}{(x_{\max} - x_{\min})} + y_{\min} \quad (3)$$

Selecting the number of neurons was done with an optimization program, written in MATLAB®, which trains the model for 20 iterations for each number of hidden neurons ranging between 1 and 20. The number of neurons with the lowest mean RMSE (Equation (2)) for generalization (testing) was chosen as the optimum number for each model. Table 6 shows the number of hidden neurons that were selected for each of the ANN and WANN models. There was no need to go beyond 20 hidden neurons in any of the ANN models due to the large increase in generalization RMSE after this point.

As previously mentioned, 80% of the data was used for training/validation and the last 20% was used for testing. The training/validation set was further divided into 80% training and 20% validation, so overall, 64% of the data was used for training, 16% was used for validation, and 20% was used for testing. The partitioning of the training and validation data sets was decided according to the optimum random index selected for each watershed, discussed above.

Table 6 | Number of inputs and optimum number of hidden neurons for each of the ANN and WANN models

Watershed	Model	# of inputs	Number of hidden neurons
Mediterranean	ANN	12	6
	WANN	31	1
Oceanic	ANN	12	7
	WANN	46	2
Hemiboreal	ANN	9	19
	WANN	38	17

Every time the model was trained, there was a mechanism to stop training early when the model had converged. This mechanism is a default operation of the ANN toolbox. Model convergence was determined when the training and validation error reached a minimum and performance did not continue to increase. Depending on where the model begins on the error surface, the initial setting of the weights and biases, and whether the model gets stuck in a local minima or not, convergence times will differ and produce different outcomes. These are elements inherent to the ANN used for this study that prevent the performance of the training and simulation from being exactly the same each time. To overcome this and better quantify model performance, each ANN model was trained and simulated 500 times and the performance results were analyzed to construct a 99% confidence interval. Out of all of the results, the widest confidence interval was found to be less than 0.01% of the mean and, therefore, negligible, so only the mean values of the 500 simulations are reported.

WANN models

The procedure for the WANN models is exactly the same as for the ANN models except for the number and type of inputs. Instead of inputs that consist of undecomposed variables, the inputs were the significant sets of wavelet sub-time series. In previous studies, the wavelet sub-time series for a given signal were often summed once the insignificant coefficients were discarded (e.g. Partal 2009; Kişi & Çimen 2011); however, for this study each set of wavelet coefficients were used as inputs into the model. Consequently, there are up to a maximum of 48 inputs (4 variables × 4 wavelet sub-time series × 3 lag periods) for each of the models. The reason for this procedure is twofold: (1) to allow the model to better extract the non-stationary components from the data; and (2) to prevent a recreation of the original signal when no details are discarded. Finally, the training and validation data sets were partitioned according to the optimum random index selected for each watershed.

SVR models

In SVR, the goal is to find a functional dependency $f(\vec{x})$ between the input variables $X = (\vec{x}_1, \vec{x}_2, \ldots \ldots \vec{x}_n)$ and the target values

$Y = (y_1, y_2, \ldots \ldots y_n)$, where $\vec{x}_i \in X \subseteq R^m$, $y_i \in Y \subseteq R$, n is the total number of training samples, and m is the number of dimensions in the feature space, that has a prescribed maximum deviation (ε) and, at the same time, remains as flat as possible. Readers are directed to Vapnik (1995), Smola (1996), and Smola & Scholkopf (2004) for a comprehensive theoretical background on these techniques.

All SVR models were created using the LIBSVM (v.3.1) software created by Chang & Lin (2011). Similar to the ANN models, there are a maximum of 12 inputs for each SVR model. All inputs are normalized between −1 and 1 and the output is normalized between 1 and 0. The data are divided into 64% training, 16% validation, and 20% testing; however, because of the nature of SVR relative to ANNs, the validation data are used to define the optimum parameters instead of determining model convergence. The partitioning of the training and validation data sets was decided according to the optimum random index selected for each watershed.

The non-linear RBF kernel was selected for this study, which requires that three parameters are selected by the user, namely: gamma (γ), cost (C), and epsilon (ε). The procedure for finding these parameters consisted of a number of steps. First, each of the parameters was adjusted through a trial and error process to get an approximate idea of the optimum range. Second, an optimization program, created in MATLAB®, was run to test the different combination of values within the ranges identified in the first step. The combination of parameters that produces the lowest training and validation RMSE (Equation (2)) is chosen as the best combination. Finally, the selected combination is adjusted with even more precision through a trial and error process for a more localized optimization of the model parameters. Table 7 shows the optimum parameters selected for each of the SVR and WSVR models.

WSVR models

The WSVR models were created in the same manner as the WANN models. Significant sets of wavelet sub-time series were used as inputs for the SVR models and the training and validation data sets were partitioned according to the optimum random index selected for each watershed, as discussed above. Lastly, the procedure for finding the optimum parameters was the same as before and the selection of parameters is shown in Table 7.

Model performance comparison

From a technical perspective, one is concerned with a model's ability to forecast individual values with reduced error, but from a hydrological perspective, one is more concerned with a model's ability to forecast extreme events, such as peak flows and droughts. Considering this, a number of statistical performance measures were selected to meet two objectives: (1) measure the model's capability to forecast individual values; and (2) measure the model performance in a meaningful way for hydrological interpretation. The model's ability to forecast individual values was measured by fractional standard error (FSE) and the coefficient of determination (R^2). The criteria selected for hydrological interpretation were the Nash–Sutcliffe model efficiency (E), bias (B), probability of detection (POD), and false alarm rate (FA). Finally, each model was tested for lead times of 1, 2 and 3 days, so that model accuracy with relation to the lead time window could be analyzed.

Table 7 | Number of inputs and optimum parameters for each of the SVR and WSVR models

Watershed	Model	# of Inputs	γ	ε	C
Mediterranean	SVR	12	0.013	0.0003	97
	WSVR	31	0.069	0.0014	92
Oceanic	SVR	12	0.032	0.0030	102
	WSVR	46	0.017	0.0010	102
Hemiboreal	SVR	9	0.082	0.0010	102
	WSVR	38	0.050	0.0010	102

Performance criteria for prediction of individual values

FSE is the RMSE (Equation (2)) divided by the corresponding mean of the targets (observed values). It is a scalable measure of model precision, expressed as:

$$\mathrm{FSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\bar{y}} \tag{4}$$

where y_i, $\hat{y}_i$, and $\bar{y}$ are the observed, forecasted, and mean of the observed streamflows, respectively. The model becomes more precise as the FSE reaches zero.

The coefficient of determination (R^2) is expressed as:

$$R^2 = \left(\frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \hat{y}_{\mathrm{mean}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \hat{y}_{\mathrm{mean}})^2}} \right)^2 \tag{5}$$

where $\hat{y}_{\mathrm{mean}}$ is the mean of the forecasted streamflows. The R^2 shows how much variability in the data set is accounted for by the model and provides a measure of how likely future outcomes will be forecasted. Values for R^2 range from 0 to 1, with 1 showing perfect forecasting ability.

Performance criteria for hydrological interpretation

The Nash–Sutcliffe model efficiency (E) is expressed as:

$$E = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{6}$$

E is used widely in hydrology because it measures the ability of the model to forecast values that are better than the mean. Values of E range from $-\infty$ to 1, with 1 showing perfect model performance.

Multiplicative bias (B), is expressed as:

$$B = \frac{\sum_{i=1}^{n}\hat{y}_i}{\sum_{i=1}^{n} y_i} \tag{7}$$

B provides a good measure of whether the model is overestimating ($B > 1$) or underestimating ($B < 1$) compared to observed values. $B = 1$ indicates perfect model performance.

POD and FA depend on a threshold determined by the user. Typically, the threshold is set to test the ability of the model to forecast certain events (e.g. streamflow peaks) in relation to observed events. Due to the differences in data among the catchments, it was very hard to set a threshold that is representative of all events in all catchments; therefore, the threshold for this study was set at the 90th percentile of total daily streamflow (m^3) for the testing period of each catchment. The 90th percentile was selected because it eliminates those events that are more periodic than others. The values for each threshold are reported in Tables 2–4.

The POD is expressed as:

$$\mathrm{POD} = \frac{\#\text{ of times for all } i\ (\hat{y}_i \geq b | y_i \geq b)}{\#\text{ of times for all } i (y_i \geq b)} \tag{8}$$

where $b =$ threshold, $\hat{y}_i =$ forecasted output, and y_i the observed output. POD values, for this study, are ratios between 0 and 1 expressing the percentage of times the forecasted model was correct in forecasting events greater than the 90th percentile of total daily streamflow (m^3). The FA rate is given by:

$$\mathrm{FA} = \frac{\#\text{ of times for all } i\ (\hat{y}_i \geq b | y_i < b)}{\#\text{ of times for all } i (y_i < b)} \tag{9}$$

FA indicates how many times the model forecasts events greater than the 90th percentile of total daily streamflow when there is no observation of such event.

RESULTS

Mediterranean climate regime

The model performance statistics for the Mediterranean watershed are reported in Table 8. Although the SVR model had the highest testing R^2 (0.695), overall the WANN model outperformed all of the other models for all three lead times. Figure 3 shows a comparison of observed versus forecasted values for the WANN model and Figure 4

Table 8 | Model performance in the Mediterranean watershed

Model	Lead Time	Training FSE	Testing FSE	Testing R^2	Testing *E*	Testing *B*	Testing POD[a]	Testing FA[a]
ANN	1 day	2.833	4.293	0.620	0.601	0.878	0.680	0.023
	2 day	4.561	6.806	0.076	0.020	1.013	0.270	0.020
	3 day	4.637	6.905	0.021	0.012	1.003	0.834	0.283
WANN	1 day	3.100	3.854	0.691	0.681	0.855	0.653	0.028
	2 day	4.481	6.598	0.082	0.079	1.003	0.425	0.070
	3 day	4.557	6.873	0.026	0.021	1.003	0.693	0.140
SVR	1 day	5.567	4.292	0.695	0.604	0.682	0.735	0.008
	2 day	7.341	6.679	0.075	0.056	1.006	0.498	0.018
	3 day	8.425	6.936	0.015	0.004	1.006	0.449	0.018
WSVR	1 day	4.700	4.440	0.629	0.574	0.617	0.685	0.053
	2 day	6.813	6.743	0.046	0.038	1.005	0.609	0.098
	3 day	8.814	6.956	0.011	0.002	1.005	0.595	0.088

[a]Threshold for POD and FA is 90th percentile of total daily streamflow in testing period (10^6 m^3) = 5.65×10^{-2}.
All performance measures, except for POD and FA, for ANN and WANN models are mean values of 500 iterations.

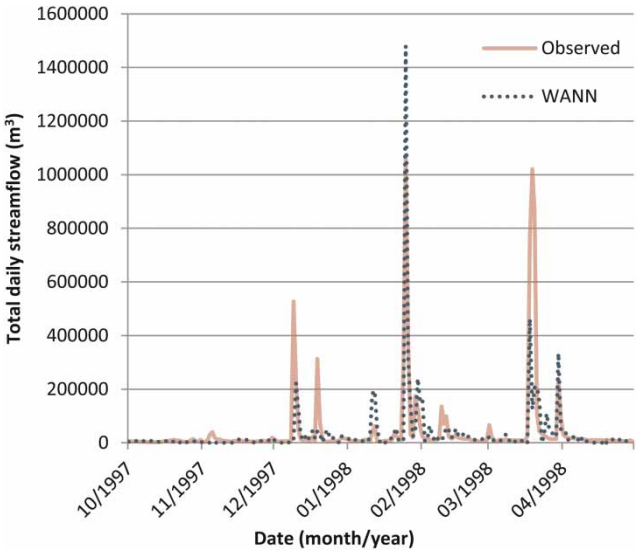

Figure 3 | Observed versus the WANN model in the Mediterranean watershed.

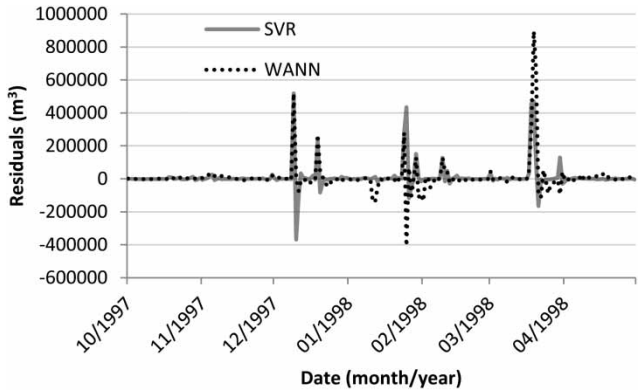

Figure 4 | Residuals of best and second best performing models in the Mediterranean watershed.

shows a residual comparison of the WANN model versus the second best performing model, the SVR. Coupling the ANN with wavelet transformed data improved model performance by 7%; however, it deteriorated the model performance of the SVR models by 7%. Some possible explanations are that the optimum parameters for the WSVR were not found in this case, or that the regression-based models were not as capable as the ANN models to create a function with wavelet transformed data, which represented the Mediterranean watershed's intermittency.

Overall, the models were able to forecast the stream's intermittency and capture the regular and no flow periods, but did not capture the majority of peak flows. Furthermore, an analysis of model performance with relation to the lead time suggests that performance deteriorates almost completely as the lead time increases, suggesting that lead times greater than 1 day are not suitable for this watershed.

Oceanic climate regime

The model performance statistics for the Oceanic watershed are reported in Table 9. The SVR model outperformed all of the other models for all three lead times. Figure 5 shows a comparison of observed versus forecasted values for the

Table 9 | Model performance in the Oceanic watershed

Model	Lead Time	Training FSE	Testing FSE	Testing R^2	Testing *E*	Testing *B*	Testing POD[a]	Testing FA[a]
ANN	1 day	0.726	0.650	0.858	0.856	0.976	0.781	0.011
	2 day	1.150	1.115	0.579	0.571	0.998	0.907	0.107
	3 day	1.362	1.325	0.415	0.395	1.017	0.858	0.111
WANN	1 day	0.784	0.692	0.842	0.838	0.938	0.758	0.010
	2 day	1.164	1.133	0.568	0.557	1.015	0.800	0.065
	3 day	1.382	1.331	0.408	0.389	1.019	0.838	0.089
SVR	1 day	0.772	0.586	0.888	0.883	0.922	0.799	0.007
	2 day	1.397	1.179	0.581	0.520	1.027	0.793	0.033
	3 day	1.581	1.402	0.418	0.322	1.037	0.693	0.018
WSVR	1 day	0.818	0.632	0.880	0.863	0.903	0.795	0.010
	2 day	1.379	1.210	0.561	0.495	1.028	0.791	0.312
	3 day	1.547	1.415	0.381	0.310	1.036	0.707	0.020

[a]Threshold for POD and FA is 90th percentile of total daily streamflow in testing period (10^6 m^3) = 5.65×10^{-2}.
All performance measures, except for POD and FA, for ANN and WANN models are mean values of 500 iterations.

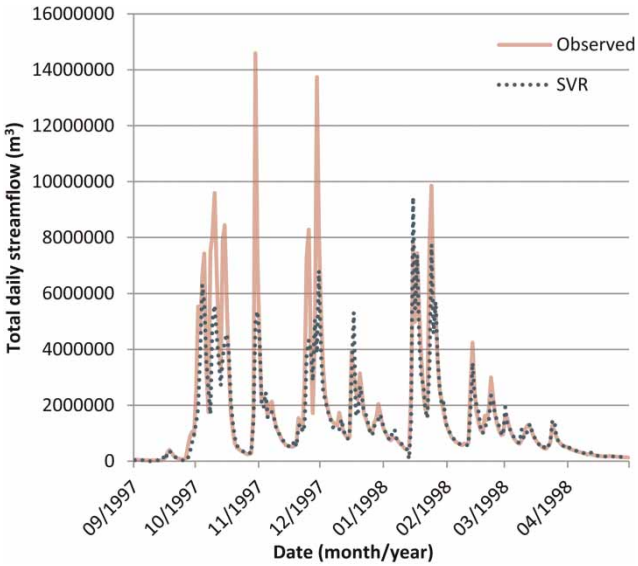

Figure 5 | Observed versus the SVR model in the Oceanic watershed.

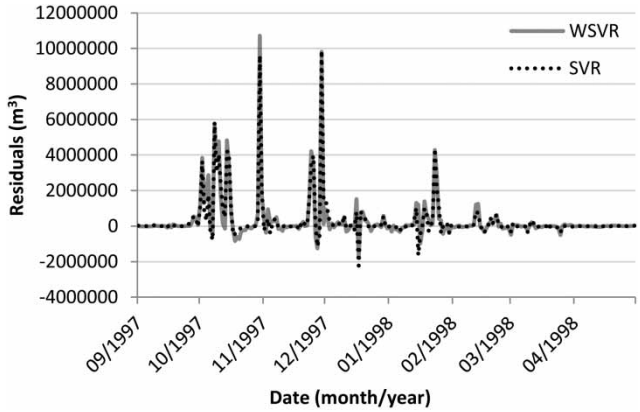

Figure 6 | Residuals of best and second best performing models in the Oceanic watershed.

SVR model and Figure 6 shows a residual comparison of the SVR model versus the second best performing model, the WSVR. For the 1 day lead time, coupling the ANN and SVR models with wavelet transformed data decreased model performance by 1.6 and 0.8%, respectively.

Overall, all of the models forecasted well for 1 day lead times and much better than similar forecasts in the Mediterranean watershed. The models also captured more of the peak flow periods than models in the Mediterranean watersheds, but did not capture all of them. Furthermore, an analysis of model performance with relation to the lead time shows that performance deteriorates with increasing lead times, but more gradually than in the watersheds with less correlated inputs. One and 2 day lead time forecasts are still useable; however, 1 day lead times are more reliable.

Hemiboreal climate regime

The model performance statistics for the Hemiboreal watershed are reported in Table 10. The WSVR model outperformed all of the other models in this watershed. The WSVR produced the lowest training and testing FSE (0.101 and 0.110, respectively), had the highest testing R^2 (0.992), and was the best at forecasting outcomes different

Table 10 | Model performance in the Hemiboreal watershed

Model	Lead Time	Training FSE	Testing FSE	Testing R^2	Testing E	Testing B	Testing POD[a]	Testing FA[a]
ANN	1 day	0.115	0.123	0.990	0.990	1.000	0.927	0.011
	2 day	0.198	0.201	0.966	0.966	0.997	0.958	0.009
	3 day	0.264	0.271	0.938	0.938	0.998	0.962	0.123
WANN	1 day	0.099	0.114	0.991	0.991	1.000	0.950	0.008
	2 day	0.198	0.210	0.964	0.963	0.991	0.960	0.010
	3 day	0.271	0.277	0.935	0.935	1.001	0.952	0.010
SVR	1 day	0.125	0.118	0.991	0.991	0.988	0.918	0.008
	2 day	0.214	0.198	0.968	0.967	1.009	0.945	0.006
	3 day	0.299	0.273	0.940	0.937	1.013	0.942	0.006
WSVR	1 day	0.101	0.110	0.992	0.992	0.991	0.932	0.006
	2 day	0.186	0.187	0.971	0.970	1.006	0.944	0.006
	3 day	0.267	0.266	0.941	0.940	1.009	0.938	0.008

[a]Threshold for POD and FA is 90th percentile of total daily streamflow in testing period (10^6 m^3) = 5.65×10^{-2}.
All performance measures, except for POD and FA, for ANN and WANN models are mean values of 500 iterations.

than the mean ($E = 0.992$). Overall, the forecast from the WSVR model slightly underestimated observed values ($B = 0.991$), and was able to capture events greater than the threshold 93% of the time with FAs less than one percent of the time. Figure 7 shows a comparison of observed versus forecasted values for the WSVR model and Figure 8 shows a residual comparison of the WSVR versus the second best performing model, the WANN. Coupling the ANN and SVR models with wavelet transformed data improved model performance slightly (<1%) in all cases.

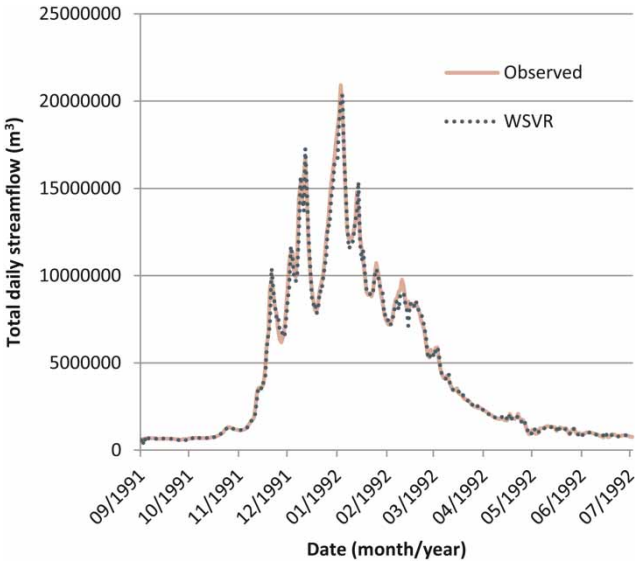

Figure 7 | Observed versus the WSVR model in the Hemiboreal watershed.

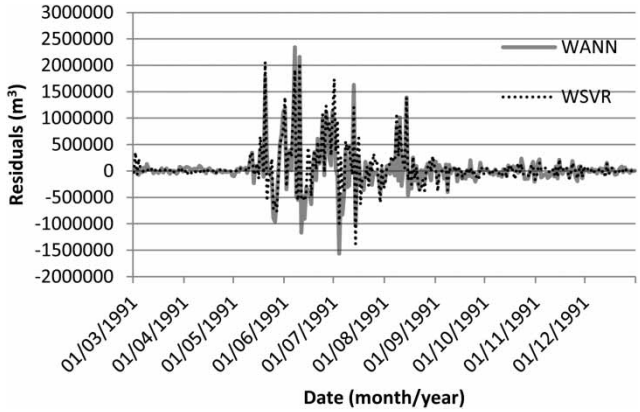

Figure 8 | Residuals of best and second best performing models in the Hemiboreal watershed.

The precipitation variable for the Hemiboreal watershed did not show correlation high enough to meet the significance threshold; therefore, it was discarded for the ANN and SVR. After decomposition, the precipitation approximation sub-time series (A3 in Table 5) delayed by 1 and 2 days met the significance criteria and was included as inputs for the WANN and WSVR. Adding the extra correlated inputs is probably the reason for the slight increase in performance of the hybrid models.

Overall, all the models were successful in predicting the large majority (>99%) of information, including the peak flows. Furthermore, although model performance deteriorated as the lead time increased, the decrease is smaller

than all of the other watersheds and performance is still relatively good at a 3 day lead time.

Overall model performance

Using the methodology described in this paper, all of the models performed within a close range of one another (0–10%) and there was no one model that outperformed the others in the majority of cases. The same is true considering those models that performed best and second best; however, SVR-based models outperformed ANN-based models in four out of six cases. The improved performance of the SVR models likely reflects their ability to better find the global minima, whereas the ANN models may have been occasionally getting stuck in local minima.

Using data pre-processed with wavelet transformation increased model performance in two out of three watersheds. Considering those models that performed best and second best, wavelet transformation was shown to increase model performance in three out of six cases. Wavelet transformation improved ANN model performance in two watersheds (i.e. Mediterranean, Hemiboreal) and only improved SVR performance in one (Hemiboreal). Some reasons for this may be that, in some cases, adding many more inputs may make it more difficult to find the model's optimum parameters. Because the ANNs only have one parameter (i.e. number of hidden neurons), they are more easily calibrated; whereas, the SVR models have three parameters (i.e. cost, gamma, and epsilon), making it much more difficult to define exactly what each of them are. All in all, this suggests that, although wavelet transformation can be effective at increasing model performance, this is not always true and may depend on the characteristics of the data and the chosen model.

In all of the models, increasing the lead time from 1 day to 2 days and 3 days deteriorated model performance. The amount of deterioration was different in each watershed and was relative to the lag correlation coefficients presented in Tables 2–4, suggesting that these models cannot be expected to make useful forecasts with inputs that are not well correlated to the variable being forecasted. The correlation in this case is likely attributed to the difference in persistence in each watershed, which is discussed more in the next section.

Model performance between climate regimes

Forecast performance, ranked from highest to lowest, is as follows: (1) Hemiboreal, (2) Oceanic, and (3) Mediterranean. A likely explanation for this is the difference in persistence of the three watersheds as indicated by the lag correlation coefficients for streamflow undecomposed (Tables 2–4) and the approximation wavelet sub-time series (Table 5). Streamflow in the Oceanic and Hemiboreal climate regimes is, by far, more persistent, in which case, forecasting next day streamflow is in general an 'easier' task for the models. Also, persistence likely explains the quick deterioration of model performance relative to the lead time window in the Mediterranean catchment. A previous day's streamflow alone does not contain enough useful information to make reliable predictions about the following day's streamflow; instead, precipitation is a far more important variable for daily forecasts. This reflects the dependency of flow on surface runoff, the short time of concentration and how quickly peak flow events happen in this watershed; therefore, a stronger correlation might be found using hourly data, if it should be available. This is in contrast to the high persistence of the Oceanic and Hemiboreal watersheds, where flow contributions come from slower processes, most notably snowmelt, glacial runoff, and subsurface flow. Support for this finding includes: (1) the higher correlation of temperature and streamflow data for the Oceanic and Hemiboreal as compared to the Mediterranean watershed (Tables 2–4); and (2) the fact that precipitation had very little correlation with the streamflow in the Hemiboreal watershed (Table 4). Furthermore, the correlation for temperature in the Oceanic and Hemiboreal watersheds increases with the lead time, indicating the potential improvement that temperature data may have for larger lead times in these climate regimes.

SUMMARY AND CONCLUSIONS

The application of new data-driven methods for forecasting daily streamflow in three different climate regimes was analyzed in this study. Four models were tested in each watershed: ANN, SVR, WANN, and WSVR. The inputs to

the ANN, and SVR models, consisted of a combination of total daily streamflow, total daily precipitation, daily minimum temperature and daily maximum temperature delayed for 1, 2 and 3 days. The inputs for the WANN and WSVR models consisted of the same data decomposed by wavelet transformation into their corresponding wavelet sub-time series. Inputs were deemed significant if, after correlation analysis, they exceeded the threshold of 0.10. The WANN, SVR, and WSVR models all performed best in the Mediterranean, Oceanic, and Hemiboreal watersheds, respectively; whereas, the ANN was outperformed by all of the other models in every case. SVR models outperformed ANN-based models in the majority of cases (4/6), but the use of wavelet sub-time series as inputs only improved performance in half (3/6) of the cases.

Three watersheds in three different climate regimes were used for this study. Performance in the Oceanic and Hemiboreal watersheds far exceeded that in the Mediterranean watershed, most likely because of the much higher persistence, slower processes (i.e. snowmelt, glacial runoff, etc.) that contribute to streamflow, and the highly correlated temperature data in those watersheds. That being said, even though model performance in the Mediterranean watershed was relatively less reliable, the models were still able to capture the majority of information and performance was improved by as much as 7% by using wavelet sub-time series.

Further research into the performance of these techniques in different climate regimes should be investigated with a much larger sample size to determine if ANN or SVR-based models are more appropriate for certain types of data. Considering the promise of using SVR-based methods for forecasting streamflow, it may be useful to combine this method with bootstrapping in an effort to better address uncertainty. Although this study showed model performance for lead times of 1, 2 and 3 days, it is likely beneficial to shorten the lead time for watersheds wherein flow originates almost entirely from surface runoff, and lengthen the lead time for watersheds wherein flow originates from a variety of sources. Ideally, one would derive the appropriate lead time according to the watershed's characteristics, but that was not possible in this study due to the unavailability of data. It should also be noted that while long records (e.g. 30 years) of daily data are widely available, such records of data with shorter time intervals are much rarer, implying that the selected station is considerably more distant from the watershed.

ACKNOWLEDGEMENTS

All Canadian hydrological and meteorological data were acquired online from Environment Canada by Deasy Nalley, Faculty of Bioresource Engineering, McGill University, Canada.

All Israeli hydrological and meteorological data were provided by the Israeli Meteorological Service and the Israeli Hydrological Service. Funding for this study was provided by the Tim Casgrain Water Resources Management Fellowship at McGill University.

REFERENCES

Abrahart, R. J. & See, L. 1998 Neural Network vs. ARMA modeling: constructing benchmark case studies of river flow prediction. In: *Proceedings of the 3rd International Conference on Geocomputation*, University of Bristol, England.

Abudu, S., King, J. P. & Bawazir, A. S. 2011 Forecasting monthly streamflow of spring-summer runoff season in Rio Grande headwaters basin using stochastic hybrid modeling approach. *Journal of Hydrologic Engineering* **16**, 384–390.

Adamowski, J. 2007 *Development of a Short-term River Flood Forecasting Method Based on Wavelet Analysis*. Polish Academy of Sciences Publication, Warsaw, 172 pp.

Adamowski, J. 2008a 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. 2008b River flow forecasting using wavelet and cross-wavelet transform models. *Journal of Hydrological Processes* **22**, 4877–4891.

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

Adamowski, K., Prokoph, A. & Adamowski, J. 2009 Development of a new method of wavelet aided trend detection and estimation. *Journal of Hydrological Processes* **23**, 2686–2696.

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.

Anctil, F. & Tape, D. 2004 An exploration of artificial neural network rainfall–runoff forecasting combined with wavelet decomposition. *Journal of Environmental Engineering and Science* **3**, 121–128.

Asefa, T., Kemblowski, M., McKee, M. & Khalil, A. 2006 Multi-time scale stream flow 505 predictions: The support vector machines approach. *Journal of Hydrology* **318**, 7–16.

Besaw, L. E., Rizzo, D. M., Bierman, P. R. & Hackett, W. R. 2010 Advances in ungauged streamflow prediction using artificial neural networks. *Journal of Hydrology* **386**, 27–37.

Beven, K. 2006 A manifesto for the equifinality thesis. *Journal of Hydrology* **320**, 18–36.

Birikundavyi, S., Labib, R., Trung, H. & Rousselle, J. 2002 Performance of neural networks in daily streamflow forecasting. *Journal of Hydrologic Engineering* **7**, 392–398.

Bow River Project Research Consortium 2010 *Bow River Project: Final Report*. Alberta Water Research Institute, Edmonton.

Cannas, B., Fanni, A., Sias, G., Tronci, S. & Zedda, M. K. 2006 River flow forecasting using neural networks and wavelet analysis. In: *Proceedings of the European Geosciences Union*, 2–7 April, Vienna, Austria.

Cao, L. & Tay, F. E. H. 2001 Financial forecasting using support vector machines. *Neural Computing and Applications* **10**, 184–192.

Chang, C.-C. & Lin, C.-J. 2011 LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* **2**, 27:1–27:27. Software available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Chen, Y. H. & Chang, F. J. 2009 Evolutionary artificial neural networks for hydrological systems forecasting. *Journal of Hydrology* **367**, 125–137.

Cigizoglu, H. K. 2005 Application of generalized regression neural networks to intermittent flow forecasting and estimation. *Journal of Hydrologic Engineering* **4**, 336–342.

Coulibaly, P. & Burn, D. H. 2004 Wavelet analysis of variability in annual Canadian streamflows. *Water Resources Research* **40**, W031051–W031051. American Geophysical Union.

Daniel, T. M. 1991 Neural networks – Applications in hydrology and water resources engineering. International Hydrology and Water Resources Symposium, 2–4 October 1991, Perth, pp. 791–802.

Dhanya, C. T. & Kumar, D. N. 2011 Predictive uncertainty of chaotic daily streamflow using ensemble wavelet networks approach. *Water Resources Research* **47**, W06507.

Guo, J., Zhou, J., Qin, H., Zou, Q. & Li, Q. 2011 Monthly streamflow forecasting based on improved support vector machine model. *Expert Systems with Applications* **38**, 13073–13081.

Hagan, M. T. & Menhaj, M. B. 1994 Training feedforward networks with the Marquardt Algorithm. *IEEE Transactions on Neural Networks* **5**, 989–993.

Han, D., Kwong, T. & Li, S. 2007 Uncertainties in real-time flood forecasting with neural networks. *Hydrological Processes* **21**, 223–228.

Hurkmans, R., Terink, W., Uijlenhoet, R. & Torfs, P. 2010 Changes in streamflow dynamics in the Rhine Basin under three high-resolution regional climate scenarios. *Journal of Climate* **23**, 679–699.

Jain, S. K., Das, A. & Srivastava, D. K. 1999 Application of ANN for reservoir inflow prediction and operation. *Journal of Water Resources Planning and Management* **125**, 263–271.

Jeong, D. & Kim, Y. 2005 Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction. *Hydrological Processes* **19**, 3819–3835.

Kang, K. W., Kim, J. H., Park, C. Y. & Ham, K. J. 1993 Evaluation of hydrological forecasting system based on neural network model. In: *Proceedings of the 25th Congress of the International Association for Hydraulic Research*, Delft, The Netherlands, pp. 257–264.

Khaliq, M. N., Ouarda, T. B. M. J., Ondo, J.-C., Gachon, P. & Bobee, B. 2006 Frequency analysis of a sequence of dependant and/or non-stationary hydro-meteorological observations: A review. *Journal of Hydrology* **329**, 534–552.

Khan, M. S. & Coulibaly, P. 2006 Application of support vector machine in lake water level prediction. *Journal of Hydrologic Engineering* **11**, 199–205.

Kim, T. W. & Valdes, J. B. 2003 Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. *Journal of Hydrologic Engineering, ASCE* **6**, 319–328.

Kişi, Ö. 2008 Stream flow forecasting using neuro-wavelet technique. *Hydrological Processes* **22**, 4142–4152.

Kişi, Ö. 2009 Neural networks and wavelet conjunction model for intermittent streamflow forecasting. *Journal of Hydrologic Engineering* **14**, 773–783.

Kişi, Ö. & Cigizoglu, H. K. 2007 Comparison of different ANN techniques in river flow prediction. *Civil Engineering and Environmental Systems* **3**, 211–231.

Kişi, Ö. & Çimen, M. 2009 Evapotranspiration modelling using support vector machines. *Hydrological Science Journal* **54**, 918–928.

Kişi, Ö. & Çimen, M. 2011 A wavelet-support vector machine conjunction model for monthly streamflow forecasting. *Journal of Hydrology* **399**, 132–140.

Krstanovic, P. F. & Singh, V. P. 1991 A univariate model for long-term streamflow forecasting. *Stochastic Hydrology and Hydraulics* **5**, 189–205.

Labat, D. 2008 Wavelet analysis of the annual discharge records of the world's largest rivers. *Advances in Water Resources* **31**, 109–117.

Lekkas, D. F., Imrie, C. E. & Lees, M. J. 2001 Improved non-linear transfer function and neural network methods of flow routing for real-time forecasting. *Journal of Hydroinformatics* **3**, 153–164.

Lu, W., Wang, W., Leung, A. Y. T., Lo, S., Yuen, R. K. K., Xu, Z. & Fan, H. 2002 Air pollutant parameter forecasting using support vector machines. Neural networks, 2002. In: *Proceedings of the 2002 International Joint Conference*, Honolulu, Hawaii, pp. 603–635.

Maheswaren, R. & Khosa, R. 2012 Wavelet-Volterra coupled model for monthly stream flow forecasting. *Journal of Hydrology* **450–451**, 320–335.

Maity, R., Bhagwat, P. P. & Bhatnagar, A. 2010 Potential of support vector regression for prediction of monthly streamflow using endogenous property. *Hydrological Processes* **24**, 917–923.

Modini, G. C. 2000 Long-lead precipitation outlook augmentation of multi-variate linear regression streamflow forecasts. In: *Proceedings of the 68th annual Western Snow Conference*, Port Angeles, Washington, pp. 57–68, US Army Corps of Engineers.

Mohandes, M. 2002 Support vector machines for short-term electrical load forecasting. *International Journal of Energy Research* **26**, 335–345.

Murtagh, F., Starcj, J. L. & Renaud, O. 2003 On neuro-wavelet modeling. *Decision Support Systems* **37**, 475–484.

Nourani, V., Alami, M. T. & Aminfar, M. H. 2008 A combined neural–wavelet model for prediction of Ligvanchai watershed precipitation. *Engineering Applications of Artificial Intelligence* **16**, 1–12.

Partal, T. 2009 River flow forecasting using different artificial neural network algorithms and wavelet transform. *Canadian Journal of Civil Engineering* **36**, 26–38.

Partal, T. & Kişi, Ö. 2007 Wavelet and neuro-fuzzy conjunction model for precipitation forecasting. *Journal of Hydrology* **342**, 199–212.

Peel, M. C., Finlayson, B. L. & McMahon, T. A. 2007 Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences* **11**, 1633–1644.

Rajasekaran, S., Gayathri, S. & Lee, T. L. 2008 Support vector regression methodology for storm surge predictions. *Ocean Engineering* **36**, 1578–1587.

Renaud, O., Starck, J.-L. & Murtagh, F. 2002 *Wavelet-based Forecasting of Short and Long Memory Time Series*. Department d'Econometrie, University of Geneva, 2002.04. Available online at: http://www.unige.ch/ses/metri/cahiers/2002_04.pdf.

Sivakumar, B. 2003 Forecasting monthly 601 streamflow dynamics in the western United States: a nonlinear dynamical approach. *Environmental Modelling & Software* **18**, 721–728.

Shensha, M. J. 1992 The discrete wavelet transform: wedding the à trous and Mallat algorithms. *Transactions on Signal Processing* **40**, 2464–2482.

Smith, L. C., Turcotte, D. L. & Isacks, B. 1998 Stream flow characterization and feature detection using a discrete wavelet transform. *Hydrological Processes* **12**, 233–249.

Smola, A. J. 1996 Regression Estimation with Support Vector Learning Machines. MSc Thesis, Technische Universitat Munchen, Germany.

Smola, A. J. & Scholkopf, B. 2004 A tutorial on support vector regression. *Statistics and Computing* **14**, 199–222.

Sudheer, K. P., Gosain, A. K. & Ramasastri, K. S. 2002 A data-driven algorithm for constructing artificial neural network rainfall–runoff models. *Hydrological Processes* **16**, 1325–1330.

Tiwari, M. K. & Chatterjee, C. 2010 Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach. *Journal of Hydrology* **394**, 458–470.

Tiwari, M. K., Song, K. Y., Chatterjee, C. & Gupta, M. M. 2013 Improving reliability of river flow forecasting using neural networks, wavelets and self organising maps. *Journal of Hydroinformatics* **15**, 486–502.

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

Wang, W. & Ding, S. 2003 Wavelet network model and its application to the predication of hydrology. *Nature and Science* **1**, 67–71.

Wang, W., Xu, Z. & Lu, J. W. 2003 Three improved neural network models for air quality forecasting. *Engineering Computations* **20**, 192–210.

Wang, W. C., Chau, K. W., Cheng, C. T. & Qiu, L. 2009 A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology* **374**, 294–306.

Wang, W., Hu, S. & Li, Y. 2011 Wavelet transform method for synthetic generation of daily streamflow. *Water Resources Management* **25**, 41–57.

Wei, S., Zuo, D. & Song, J. 2012 Improving prediction accuracy of river discharge time series using a Wavelet-NAR artificial neural network. *Journal of Hydroinformatics* **14**, 974–991.

Yu, P. S., Chen, S. T. & Chang, I. F. 2006 Support vector regression for real-time flood stage forecasting. *Journal of Hydrology* **328**, 704–716.

Yuan, H. C., Xiong, F. L. & Huai, X. Y. 2003 A method for estimating the number of hidden neurons in feed-forward neural networks based on information entropy. *Computers and Electronics in Agriculture* **40**, 57–64.

Zealand, C. M., Burn, D. H. & Simonovic, S. P. 1999 Short term streamflow forecasting using artificial neural networks. *Journal of Hydrology* **214**, 32–48.

First received 10 April 2013; accepted in revised form 11 September 2013. Available online 6 November 2013