Multiscale nonlinear model for monthly streamflow forecasting: a wavelet-based approach

Maheswaran Rathinasamy and Rakesh Khosa

ABSTRACT

The dynamics of the streamflow in rivers involve nonlinear and multiscale phenomena. An attempt is made to develop nonlinear models combining wavelet decomposition with Volterra models. This paper describes a methodology to develop one-month-ahead forecasts of streamflow using multiscale nonlinear models. The method uses the concept of multiresolution decomposition using wavelets in order to represent the underlying integrated streamflow dynamics and this information, across scales, is then linked together using the first- and second-order Volterra kernels. The model is applied to 30 river data series from the western USA. The mean monthly data series of 30 rivers are grouped under the categories low, medium and high. The study indicated the presence of multiscale phenomena and discernable nonlinear characteristics in the streamflow data. Detailed analyses and results are presented only for three stations, selected to represent the low-flow, medium-flow and high-flow categories, respectively. The proposed model performance is good for all the flow regimes when compared with both the ARMA-type models as well as nonlinear models based on chaos theory.

Key words | forecasting, mean monthly streamflow, multiscale nonlinear models, Volterra models, wavelets

Maheswaran Rathinasamy (corresponding author)
Rakesh Khosa
Department of Civil Engineering,
Indian Institute of Technology Delhi,
New Delhi 110016,
India
E-mail: *maheswaran27@yahoo.co.in*
or *rakesh.khosa@gmail.com*

NOMENCLATURE

ANN Artificial Neural Networks
ANFIS Artificial Neural Fuzzy Inference System
ARMA Auto-Regressive Moving Average

INTRODUCTION

Understanding streamflow dynamics plays an important role in the proper management of our water resources. Simulation and forecasting of the mean monthly streamflows plays a vital role in the operation of reservoirs and hydropower generation units. Streamflow dynamics are governed by various physical mechanisms acting on a wide range of temporal and spatial scales (Sivakumar 2003). Further, the presence of nonlinearity at different scales in streamflow observations has been a subject of scrutiny in numerous studies including the study by Wang *et al.* (2006). This latter study has affirmed the presence of nonlinearity in streamflow observed at monthly scales. Advances in research on the nonlinear features of streamflow processes notwithstanding, an objective understanding of the precise nature of these nonlinearities is not easily discernible.

The presence of long and short memory components at the observed scale, as reported by Corduas & Piccolo (2006), and other changes that take place at the local and global levels, also influence the evolving characteristics of the observed time series and, together, all these factors make streamflow modelling a nontrivial and complex problem.

During the past few decades, a great deal of research has been devoted to the formulation and development of modelling approaches to understand streamflow dynamics and significant progress has indeed been made. These methods can be broadly classified into: (i) physical-based models, (ii) conceptual models, and (iii) data-driven black box models. Among these, the data-driven models are often preferred because of their simplicity and predictive accuracy in

doi: 10.2166/hydro.2011.130

streamflow forecasting. These data-driven models are often conveniently classified into the following types: (i) linear stochastic models, (ii) nonlinear models and (iii) artificial intelligence models.

The linear stochastic analysis methods (Box *et al.* 1970; Carlson *et al.* 1970; Obeysekara & Salas 1986; Haltiner & Salas 1988; Hipel & McLeod 1994; María *et al.* 2004) do not predict accurately the real-world runoff sequences because of their nonlinearity in nature as suggested by available evidence and has accordingly led to a heightened interest in nonlinear studies and application of nonlinear models to modelling streamflow time series and are reviewed in detail in Wu *et al.* (2009). Other prominent studies include Tong & Lim (1980), Noakes *et al.* (1985) and Muftuoglu (1991). The study by Tong & Lim (1980) saw the development of the Threshold Autoregressive model (TAR) in which threshold value characteristics divide the data into different linear sets and for each set a piecewise linear AR model is considered. The TAR-based approach has been used in several studies such as Kapetanios (2003) and Chen *et al.* (2008). Noakes *et al.* (1985) developed PARMA (for Periodic Auto-Regressive Moving Average) models for seasonal modelling of monthly flows and, in contrast, Muftuoglu (1991) developed nonlinear models using the modified Volterra functions for runoff forecasting at a monthly scale.

Chaos theory has also found numerous applications to the problem of streamflow modelling and included Wilcox *et al.* (1991), Jayawardena & Lai (1994), Porporato & Ridolfi (1997), Elshorbagy *et al.* (2000) and Sivakumar *et al.* (2001). Interestingly, these studies demonstrate that the assumption of the presence of chaos-like features is indeed credible and, as a further validation, most of these approaches seem to perform reasonably well. Further, with the advent of techniques like ANN (French *et al.* 1992; Chang & Chang 2001; Dawson & Wilby 2001; Tesong *et al.* 2001; Jain *et al.* 2004; Huang & Catherine 2008; Wu & Chau 2010), ANFIS (Firat & Güngör 2007; Chen *et al.* 2008; Firat & Turan 2010), genetic algorithms (Chen *et al.* 2008) and support vector machines (Sivapragasam *et al.* 2001; She & Basketfield 2005; Asefa *et al.* 2006), data-driven models have begun to dominate streamflow forecasting effort. However, while these methods seem to perform well for a given dataset, they lack the ability to discern and identify dominant features that are present at different scales but whose individual, scale-specific, characteristic details get camouflaged in the integrated observation. It is now recognised that the apparently irregular behaviour, as is often seen in the time series of streamflows, could be the outcome of simple deterministic processes that are influenced by a few nonlinear interdependent components happening at different scales.

Recent developments in wavelet theory have prompted a significant shift in emphasis while designing a suitable approach to modelling time series of streamflows and, indeed, other geo-physical phenomena. With the capability to enable multiscale resolution and frequency localization in time, wavelets offer the advantage of facilitating a decomposition of the given time series of flows into its various, but scale-specific, dynamic components as surrogates of the corresponding physical processes at those scales.

Numerous studies based on wavelet applications to modelling geo-physical time series have been reported in the recent past. For example, Venugopal *et al.* (1999) used wavelet packets to study the energy distribution of rainfall over time, frequency and scale in an effort to gain more insight into the rainfall-generating mechanism. Labat *et al.* (2000) applied wavelet analysis to investigate the multiscale phenomena in the karstic regions. Gaucherel (2002), Lafrenière & Sharp (2003) and Anctil & Coulibaly (2004) have used wavelet analysis to interpret temporal patterns of different basin responses that include either rapid processes or slow recharges. Apart from simulation studies, there has also been a heightened interest in wavelets and its application for developing forecasting models of various geo-physical series. These include studies by Partal & Kisi (2007), Kisi (2009a, b) and Adamowski & Sun (2010) wherein models that combine wavelet analysis with ANNs have been developed and investigated on the assumption that such a combination approach would improve the forecasts of the modelled hydrologic time series. However, in the context of these models, Kisi (2009a, b, 2010) comments that these models do not overcome the disadvantages that are normally attributed to the ANN-based model. As an alternative approach, the latter author has proposed a linear wavelet regression model for monthly streamflow forecasting and was shown to perform better than the ANN models (Kisi 2009a, b, 2010).

In order to address the issue of nonlinear behaviour of some geo-physical state variables, including streamflow,

Volterra models offer a generic representation of these nonlinear systems and have indeed been widely applied in the area of rainfall–runoff modelling. The results of these studies have been promising and prominent amongst these studies are Diskin & Boneh (1973), Muftuoglu (1991) and Chou (2007).

In the present study, wavelets, with their capability of multiresolution decomposition, have been combined with a Volterra model in order to develop a forecasting model for univariate time series of streamflow. Detailed study is done on monthly streamflows observed at 30 stations spread throughout 10 states in the western USA and compared with results on the specific datasets analysed in Sivakumar (2003). In particular, a special focus was devoted to data for those 30 stations (out of a total of 79) investigated in the study by Sivakumar (2003) for which the nonlinear approximation method, as proposed by the latter author, did not perform satisfactorily.

The present paper is organised as follows: the next section gives a brief description of wavelet analysis followed by a description of the multiscale nonlinear model in the third section. Then the following section presents details of the study area and data considered for analysis in this research. Details of the analyses carried out and the results obtained are presented in the fifth section while the final section presents some of the important conclusions drawn from the present study and potential areas for further research.

WAVELET ANALYSIS

Wavelet analysis has become an important milestone in spectral analysis due to its multiresolution and localization capability both in time and frequency domains and has been extensively applied in the area of time series analysis and prediction. Wavelet decompositions at various scales (frequencies) often reveal the underlying low and high frequency components of the observed series and are, importantly, localized in time. Several algorithms have been designed that enable such decompositions and the selection of any particular approach depends on the application in hand.

For example, the Continuous Wavelet Transform (CWT), used mainly in the processing of medical images and seismic signals, calculates the wavelet transform as an integral product of the given signal and the wavelet function. The coefficients of the wavelet transform of a square-integrable continuous-time signal, $f(t)$, are defined by the linear integral operator:

$$C(a,\tau) = \int_{-\infty}^{\infty} f(t)\psi_{a,\tau}(t)\mathrm{d}t \quad \text{where} \quad \psi_{a,\tau}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-\tau}{a}\right) \tag{1}$$

The function $\psi(t)$, which can be real or complex, plays the role of a convolution-kernel and is called a wavelet. The parameter a can be interpreted as a dilation ($a > 1$) or a contraction ($a < 1$) factor of the wavelet function $\psi(t)$ corresponding to different scales of observation. The parameter τ can be interpreted as a temporal translation or shift of the function $\psi(t)$ and allows discrimination of the signal $f(t)$ locally around time τ. The wavelet function $\psi(t)$ is designed with the following properties (Burrus *et al.* 1998):

1. The function integrates to zero:

$$\int_{-\infty}^{\infty} \psi(t)\mathrm{d}t = 0 \tag{2a}$$

2. The function is square integrable or, equivalently, has finite energy:

$$\int_{-\infty}^{\infty} |\psi(t)|^2\mathrm{d}t < \infty \tag{2b}$$

A disadvantage of these non-orthogonal wavelets is that the CWT of a given signal is characterised by redundancy of information among the wavelet coefficients. This redundancy, on account of the correlation between coefficients, is intrinsic to the wavelet kernel and not a characteristic of the analysed signal. As an alternative, for practical applications (as in the study of noise reduction models for communication systems and image and signal compression), the Discrete Wavelet Transform (DWT) is usually preferred. In this approach, wavelet coefficients are calculated at every dyadic step, i.e. the operation of WT is carried out at dyadic

dilations and integer translations. The wavelet function in its dyadic form can be represented as:

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k) \tag{3}$$

In Equation (3), $\psi(t)$ is the mother wavelet, and j and k are the translation and dilation indices.

In DWT, decimation is carried out so that only half of the coefficients of the detailed component are left at the current level and half of the coefficients of the smooth version are recursively processed using high-pass and low-pass filters for coarser resolution levels. Due to the decimation, the number of the wavelet coefficients is halved with each move to a coarser level. The consequence is that, at the coarser level, there is lesser information available to train the forecasting model and, consequently, leading to overall predicting inaccuracy.

This problem, caused by decimation, may be overcome by introducing the stationary or, alternatively, *à trous* wavelet transform (Shensa 1992). The basic idea of the *à trous* wavelet transform is to fill the resulting gaps using redundant information obtained from the original series. Corresponding to the original series to be $x(t)$, smoother versions of $x(t)$ are defined at different scales as given by Equations (4) and (5):

$$c_0(t) = x(t) \tag{4}$$

$$c_i(t) = \sum_{l=-\infty}^{\infty} h(l)c_{j-1}(t + 2^{i-1}l) \tag{5}$$

In the preceding Equation (5), i takes values from 1 to p and h is a low pass filter with compact support such as the B_3 spline defined as (1/16, 1/4, 3/8, 1/4, 1/16) and Haar defined as (1/2, 1/2).

The detail component of $x(t)$ at level i is defined as:

$$d_i(t) = c_{i-1}(t) - c_i(t) \tag{6}$$

The set $\{d_1, d_2, \ldots, d_p, c_p\}$ represents the additive wavelet decompositions of the data up to the resolution level p. The term c_p is the residual component or the approximation. Accordingly, for reconstruction, the inverse transform is given by:

$$x(t) = c_p(t) + \sum_{i=1}^{p} d_i(t) \tag{7}$$

Here, unlike the classical DWT, the decimation is left out, which renders the components at different scales to be of the same length.

Treatment of boundary details

Estimation of wavelet coefficients requires the selection of appropriate boundary conditions and this exercise does indeed require special care. From Equation (5), it is seen that estimation of the wavelet coefficient at time t uses observations over the neighbourhood between $(t-p)$ and $(t+p)$, depending on the filter length. In general wavelet applications, various kinds of boundary conditions such as: (i) periodic boundary, (ii) reflective boundary extension, and (iii) constant extension are usually used for extending the series up to $(t-p)$ and $(t+p)$ (Strang & Nguyen 1996). However, in the case of forecasting models, these extensions cannot work (Renaud *et al.* 2005).

For clarity, consider a simple example where the B3-spline wavelet is used for computing $c(t)$ from the series $x(t)$. At each time step, according to Equation (5), calculation of $c_1(n)$ requires $x(n+1)$ and $x(n+2)$. Similarly, $c_2(n)$ requires $x(n+1)$, $x(n+2), \ldots, x(n+6)$. In general, computation of $c_i(n)$ would require $x(t)$ defined at $t = n+1,\ n+2, \ldots, n+(2^{i+1}-2)$ and clearly are not available prior to their actual realization in time as these refer to future observations.

Alternatives that do not involve the use of future values while calculating the wavelet coefficients include the use of causal filters as these use only the past values to calculate wavelet coefficients at time t. Renaud *et al.* (2005) proposed a redundant Haar wavelet transform and Luan (2005) proposed a shifted B_3-spline wavelet so that the filtering is done with the past values of $x(t)$ and not with future ones.

Selection of the wavelet

Selection of the 'most appropriate' wavelet for the application at hand is indeed very intriguing and requires a

prior understanding of the important attributes of the candidate wavelet such as region of support and vanishing moments required for that application. The energy content of the Haar wavelet is concentrated over the narrowest support band and, therefore, has good localization properties. This attribute makes it the most suitable for change detection studies (Ahuja *et al.* 2005; Khosa *et al.* 2005). On the other hand, the B3-spline wavelets have a wider support and result in the dispersion of its energy content over a wider region and, consequently, reduced energy density. Appropriately, therefore, this particular wavelet is ideally suited to studies that require long-term averaging. The present study employs the shifted B_3-spline wavelet for multiresolution decomposition based on the perception that, in order to understand the internal dynamics of the underlying generating system, the past history over a reasonable duration is required while seeking to estimate amplitudes of the various derived component processes at any time position. A brief description of the formulation is presented below.

Let $X = C_0 = X_1, X_{2,\ldots}, X_p$ represent the original time series of observations on the state variable under scrutiny. The first level wavelet decomposition of this series, denoted as C_1, can be obtained from the original series C_0 by convolving the latter with $h = \{1/16,\ 1/4,\ 3/8,\ 1/4,\ 1/16\}$.

Mathematically

$$\left.\begin{aligned} C_1(n) &= \frac{1}{16}C_0(n) + \frac{1}{4}C_0(n-1) + \frac{3}{8}C_0(n-2) \\ &\quad + \frac{1}{4}C_0(n-3) + \frac{1}{16}C_0(n-4) \\ w_1(n) &= C_0(n) - C_1(n) \end{aligned}\right\} \tag{8}$$

More generally

$$\left.\begin{aligned} C_{i+1}(n) &= \frac{1}{16}C_i(n) + \frac{1}{4}C_i(n-(2^i\times 1)) + \frac{3}{8}C_i(n-(2^i\times 2)) \\ &\quad + \frac{1}{4}C_i(n-(2^i\times 3)) + \frac{1}{16}C_i(n-(2^i\times 4)) \\ w_{i+1}(n) &= C_i(n) - C_{i+1}(n) \end{aligned}\right\} \tag{9}$$

From (9) it can be seen that at any time point, n, the decompositions do not use prospective information (time-wise) beyond n for calculating wavelet coefficients, thus illustrating the contention that the '*à trous* transform' provides a convincing and computationally very straightforward solution while, at the same time, avoiding the troublesome boundary effects.

MULTISCALE NONLINEAR MODEL DEVELOPMENT

Consider a multiscale system whose internal (underlying) processes respond to the stimulus of external causal factors to produce the observable output variable. The scheme of this system is shown in Figure 1(a). Here, the processes $u_1, u_2, \ldots, u_5$ depict the unobservable processes at different scales. $W(t)$ is the unobservable noise-free system output, $v(t)$ is the noise that contaminates the system output and $y(t)$ is the observable, noise-contaminated, system output. The foregoing depiction can be assumed to be a reasonable representation of the streamflow dynamics where unknown component processes combine together at different scales to produce the observed streamflow. The wavelet-based approach to obtain such a multicomponent representation is developed as described below.

Methodology

Let $\{X\}$ be the time series of observations on streamflows which carries the information about the system. Let the corresponding descriptive wavelet coefficients at each scale be denoted by $\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_J$ and the scaling coefficients be denoted by $\hat{u}_{J+1}$ where J is the level of decomposition. The time series of actual observations may be assumed to be a result of integration of these individual decompositions which, in turn, are assumed to be outputs of individual Volterra generators and integrated together in a nonlinear manner. Accordingly, therefore, the wavelet and scaling coefficients may be combined nonlinearly to produce $y(t)$, a particular value of $\{X\}$, using the Volterra model within a multiple-input–single-output framework. The multiscale nonlinear model formulation is

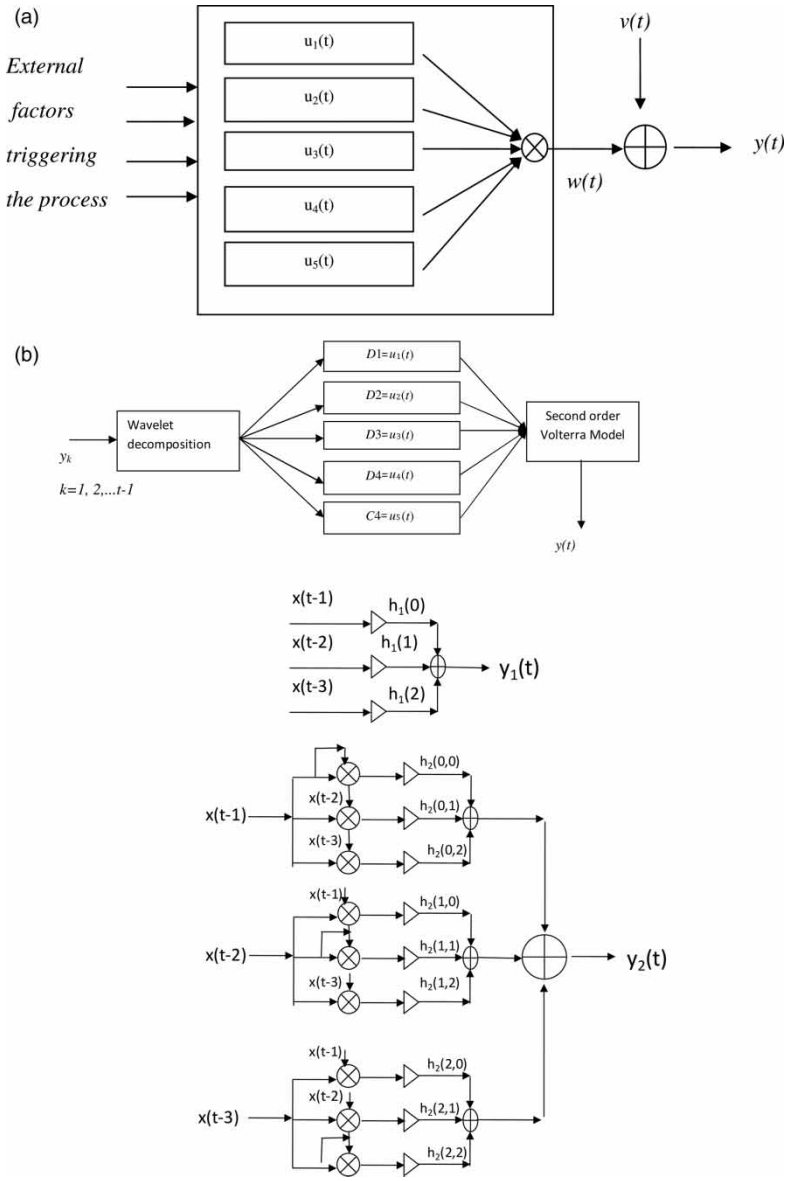

Figure 1 | (a) Multiscale system. (b) Model scheme for the multiscale nonlinear model.

given by:

$$y(t) = \sum_{n=1}^{J+1}\sum_{\tau=1}^{m} h_1^{(n)}(\tau)u_n(t-\tau) + \sum_{n=1}^{J+1}\sum_{\tau_1=1}^{m}\sum_{\tau_2=1}^{m} h_{2s}^{(n)}(\tau_1,\tau_2)u_n(t-\tau_1)u_n(t-\tau_2) + \sum_{n_1=1}^{J+1}\sum_{n_2=1}^{n_1-1}\sum_{\tau_1=1}^{m}\sum_{\tau_2=1}^{m} h_{2x}^{(n_1,n_2)}(\tau_1,\tau_2)u_{n_1}(t-\tau_1)u_{n_2}(t-\tau_2) \quad (10)$$

The first-order kernel, $h_1^{(n)}$, describes the linear relation between the nth input u_n and y. The second-order self-kernel function $h_{2s}^{(n)}$ describes the second-order nonlinear relation between the nth input u_n and y, respectively. The second-order cross-kernels $h_{2x}^{(n_1,n_2)}$ describe the second-order nonlinear interactions between each unique pair of inputs (u_{n1} and u_{n2}) as they affect y and m denotes the memory length of the process.

Combining the last two terms we get:

$$y(t) = \sum_{n=1}^{J+1}\sum_{\tau=1}^{m} h_1^{(n)}(\tau)u_n(t-\tau) + \sum_{n_1=1}^{J+1}\sum_{n_2=1}^{J+1}\sum_{\tau_1=1}^{m}\sum_{\tau_2=1}^{m} h_2^{(n_1,n_2)}(\tau_1,\tau_2)u_{n_1}(t-\tau_1)u_{n_2}(t-\tau_2) \quad (11)$$

Estimation of kernels h_1 and h_2

The four summation terms in Equation (11) may be modified by denoting each of the variables $u_1(t-1)$, $u_1(t-\tau)$, $u_2(t-1)$, $u_2(t-\tau)$, …, as a new set of variables denoted as $x_1(t)$, $x_2(t)$, $x_3(t)$,…, $x_M(t)$. Equation (11) may now be written as:

$$\left.\begin{aligned} y(t) &= \sum_{l=1}^{P} h_1(l)x_l(t) + \sum_{l_1=1}^{P}\sum_{l_2=1}^{P} h_2(l_1,l_2)x_{l_1}(t)x_{l_2}(t) \\ y(t) &= y_1(t) + y_2(t) \end{aligned}\right\} \quad (12)$$

$$\begin{aligned} x_l(t) &= \{u_k(t) \quad 1 \le k \le J+1; \quad 1 \le l \le J+1 \\ x_l(t) &= \{u_k(t-\tau) \quad 1 \le k \le J+1; \quad J+1 < l < M; \\ &\qquad \tau = 1, 2, 3 \ldots q \end{aligned}$$

$q = q^{\text{th}}$ lagged value.

P = total number of input values.

The entire model scheme is shown in Figure 1(b). The top block diagram shows the overall model scheme and the middle and bottom parts show the linear and second-order kernels of the Volterra model. The conventional method for identifying the Volterra model is inefficient and inaccurate (Nikolaou & Mantha 2000) on account of the large number of parameters that must be estimated. In this regard, Nikolaou & Mantha (2000) have demonstrated how to use wavelets for reparameterization of second-order Volterra models in terms of a substantially smaller number of coefficients. The resulting structure retains several of the advantages of the Volterra structure while, at the same time, being parsimonious. In the present study too, the method using the wavelet-based least squares and described by Nikolaou & Mantha (2000) has been used to estimate the unknown kernel functions in (12). Additionally, the optimal value of: (i) the level of decomposition, and (ii) the memory 'm' at each scale also needs to be established and is a trial-and-error-based iterative process.

Performance measures

In addition to factors such as model structure, complexity and computational requirements, forecast accuracy may be the most important consideration that provides an objective basis for comparison between various competing models as a measure of their respective performances. Various measures to evaluate model performance are available in the published literature and amongst these the Nash–Sutcliffe criterion (Nash & Sutcliffe 1970) has been widely used. A disadvantage of the Nash–Sutcliffe efficiency criterion that is often quoted in the literature arises from the use of squared differences that causes the criterion to accord more weightage to observations that have larger numerical values as compared to lower magnitude values. The choice of 'no model' forecast is another issue of concern with regards to the Nash–Sutcliffe criterion and has been discussed extensively in the published literature (see, for example, Garrick *et al.* (1978)). The present study has also used performance criteria based on Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and the Correlation Coefficient (CC). Karunanithi *et al.* (1994) suggested that RMSE is a good measure for indicating goodness of fit at high flows, while the MAE provides a more balanced perspective of the goodness of fit at moderate flows. In general, RMSE > MAE, and the degree to which RMSE exceeds MAE is an indicator of the extent to which large outliers (discrepancy between the observed and the forecasted values) exist in the evaluation set.

Equations (13)–(16) below have been used to estimate these performance measures:

$$\text{NSC} = R^2 = E = 1 - \frac{\sum_{t=1}^{N}(Q_o^t - Q_m^t)^2}{\sum_{t=1}^{N}(Q_o^t - \bar{Q}_o)^2} \quad (13)$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(Q_o^t - Q_m^t)^2} \quad (14)$$

$$\text{MAE} = \sum_{t=1}^{N}|Q_o^t - Q_m^t| \quad (15)$$

$$CC = \sum_{t=1}^{N}\frac{(Q_o^t - \bar{Q}_o)(Q_m^t - \bar{Q}_m)}{\sqrt{\sum_{t=1}^{N}(Q_o^t - \bar{Q}_o)^2 \sum_{t=1}^{N}(Q_m^t - \bar{Q}_m)^2}} \quad (16)$$

where Q_m is model output, Q_o^t is observed discharge at time t and $\bar{Q}_o$ denotes the mean discharge of the observed series for the corresponding month.

DATA USED FOR THE STUDY

Average monthly streamflow data collected over a period of 60 years (October 1931 to September 1993) at 30 stations in the western USA have been studied. The data were obtained from www.usgs.com and Table 1 presents the details of these stations (i.e. station numbers, states in which they are located, catchment area and elevation) and also some important statistics of the observed streamflows, namely mean, standard deviation, maximum and minimum values. In the present study, for convenience and brevity, the original station numbers are replaced by serial numbers starting from 01 to 30. The data for these stations show a large range of variation and may be attributed mainly to significant differences in the prevailing climate. Most of the

Table 1 | Statistics of streamflow data observed at 30 stations in the western USA

Serial number	Station number	State	Area (km^2)	Elevation (m)	Mean (m^3/s)	SD (m^3/s)	Maximum value (m^3/s)	Minimum value (m^3/s)
01	6191500	MT	6,794	2,573	86.4	97.3	563.8	11.6
02	6207500	MT	2,989	2,265	26.0	35.6	187.8	127.6
03	6214500	MT	30,549	939	199.0	210.3	1,164.4	38.6
04	6620000	CO/WY	3,706	2,713	11.4	15.1	103.3	0.7
05	14321000	OR	9,539	756	212.1	213.0	1,447.5	22.0
06	6725500	CO	93	3,170	1.5	2.2	11.2	0.1
07	7203000	NM	780	2,850	0.5	0.9	10.5	0.0
08	9299500	UT	293	3,161	3.2	3.9	33.4	0.5
09	10128500	UT	420	2,771	5.8	8.2	49.9	0.8
10	11058500	CA	23	1,067	0.2	0.3	2.9	0.0
11	11237500	CA	60	2,316	1.2	2.3	18.3	0.0
12	11264500	CA	469	2,743	10.1	14.7	93.9	0.1
13	11266500	CA	831	2,682	17.7	25.8	177.8	0.2
14	11281000	CA	225	1,585	2.8	3.7	21.5	0.0
15	11282000	CA	192	1,707	2.3	3.4	24.8	0.0
16	11522500	CA	1,945	610	51.1	50.7	318.8	2.8
17	11532500	CA	1,577	457	106.8	118.4	668.8	5.2
18	12010000	WA	142	277	12.2	11.1	71.6	0.6
19	12134500	WA	1,386	1,128	112.3	71.4	463.4	9.8
20	12306500	ID	1,476	1,484	19.8	29.0	145.3	1.1
21	12321500	ID	251	1,417	5.6	8.1	47.0	0.4
22	12401500	WA	5,750	1,390	43.9	65.3	295.6	1.8
23	12404500	WA	9,842	435	82.6	116.4	511.7	3.7
24	12414500	ID	2,668	1,283	66.8	76.0	393.6	6.7
25	12445000	WA	18,803	262	84.2	99.8	784.9	6.5
26	12451000	WA	831	1,564	40.3	40.4	219.1	3.2
27	13082500	ID	1,639	1,838	1.3	1.6	17.7	0.0
28	13313000	ID	552	2,185	9.8	14.7	99.9	1.2
29	13317000	ID	35,094	2,048	317.7	363.5	2,338.9	70.4
30	13336500	ID	4,947	1,719	106.2	135.6	690.1	9.2

drainage basins are medium to smallsized (less than approximately 1,000 km^2) and are located in elevation zones higher than approximately 500 m (Sivakumar 2003).

For the present analysis, the 30 stations are grouped under three categories based on the magnitude of mean streamflow as follows: (1) low-flow stations (mean streamflow values less than 2.832 m^3/s); (2) high-flow stations (mean values more than 28.32 m^3/s); and (3) medium-flow stations (mean values greater than 2.832 m^3/s but lower than 28.32 m^3/s). For more details about the station details the reader is referred to Sivakumar (2003).

MODEL APPLICATION

The multiscale nonlinear model described in the foregoing discussion was use to forecast monthly streamflows at each of the 30 stations for which data were available. Model design was based on the 'split sample' approach in which the available data for 700 months was split into two segments. The first segment, comprising of data for the first 500 months, was used to calibrate the model and the second segment, comprising of data for the remaining 200 months, was set aside for model validation.

Model performance is evaluated based on one-time-step-ahead forecast (i.e. a lead time of one month) for each of the 30 sites. Detailed analyses and results are presented only for stations 6, 20 and 29 (three stations in all), whereas the final results are presented for all 30 stations. It is important to note that the catchment areas corresponding to these observation stations are 93, 1,476 and 35,094 km^2, respectively. Each of the aforementioned stations (nos. 6, 20 and 29) are taken as being a representative member of their respective group of low- medium- and high-flow categories. Figures 2(a)–(c) show the hydrograph of monthly streamflows respectively for stations 6, 20 and 29 while Figures 3(a)–3(c) show their corresponding wavelet decomposition.

For each station, deemed to be a general characteristic of their respective flow regime, the models are also calibrated for their specific levels of decomposition and

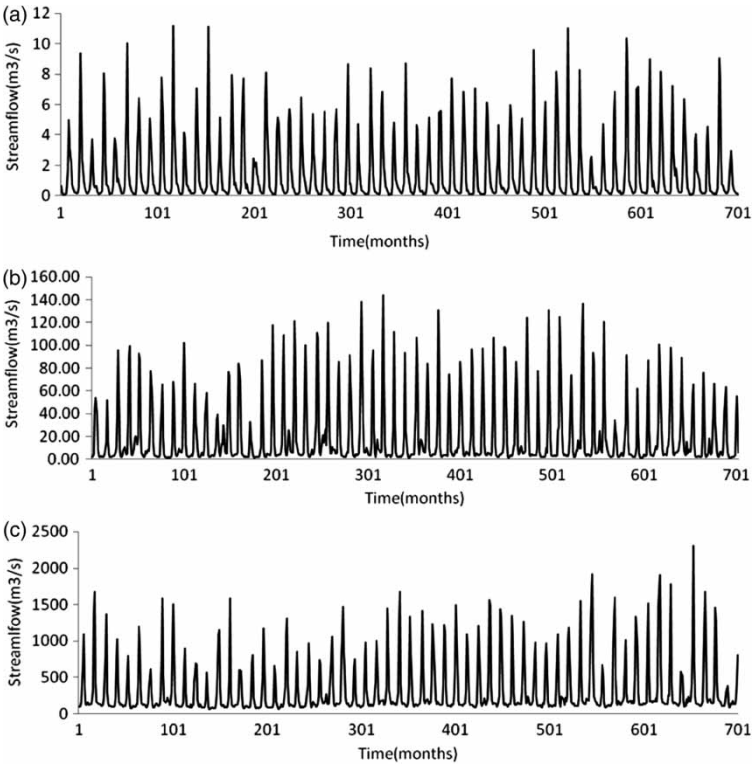

Figure 2 | Variation of monthly streamflow series in the western USA representing: (a) low-flow (station 06), (b) medium-flow (station 20) and (c) high-flow (station 29).

memory. The optimal structure was obtained using a trial-and-error approach and Table 2 presents some combinations of parameters and corresponding results. It can be seen that in most of the cases the level of decomposition was found to be 4 and the memory at each level depends on the time series. The values in the braces show the number of lagged values that were taken at each level (for the details and the approximation component).

Table 3 shows the optimal model structure for the three stations under consideration. In general, for the station corresponding to the low-flow regime, the optimal number of decomposition levels (J) was found to be 4 and the memory at each level had an optimal value of {2,2,1,1,2}. For the station corresponding to the medium-flow regime, the number of decomposition levels, J, was found to be equal to 4 and memory, M, equal to 1 for the high-frequency details (D1, D2) and was equal to 2 for D3 and equal to 3 for D4 and C4. In the case of the station corresponding to the high-flow regime, the optimal value for J was equal to 4; whereas M was

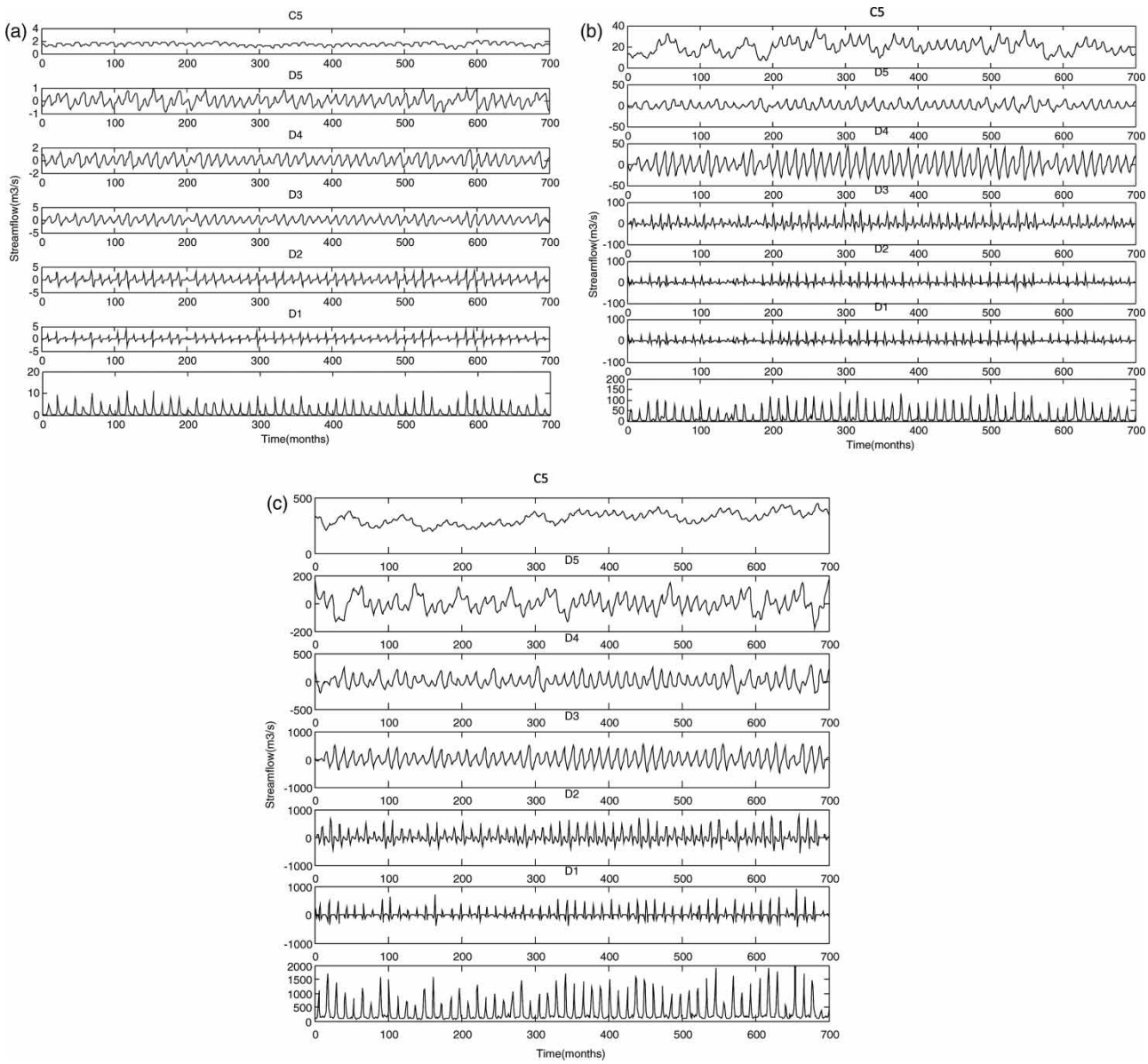

Figure 3 | Wavelet decomposition of the monthly streamflow series: (a) low-flow (station 06), (b) medium-flow (station 20) and (c) high-flow (station29).

Table 2 | Different wavelet models for the various flow regimes

Flow regime	Level of decomposition	Memory (m) at each level {DW_i's , C}	RMSE (m^3/s)	MAE (m^3/s)
Low-flow regime				
6	3	{2,2,2,1}	0.87	0.46
	4	{2,2,1,1,2}	0.70	0.31
	3	{1,1,1,2}	1.17	0.69
7	3	{2,2,2,1}	0.72	0.32
	3	{1,2,1,2}	0.79	0.43
	4	{1,2,2,1,1}	0.60	0.26
	5	{1,1,1,2,2,2}	0.82	0.44
Medium-flow regime				
09	3	{2,2,2,1}	4.20	2.29
	4	{1,2,2,3,3}	3.12	1.96
	5	{1,1,2,2,2,2}	4.62	2.68
	4	{1,1,1, 2,1}	5.94	4.06
20	3	{1,2,2,2}	12.71	7.42
	4	{1,1,2,3,3}	9.79	6.51
	5	{1,1,2,2,2,1}	13.00	8.57
	4	{1,1,2,2,2}	20.19	13.61
High-flow regime				
30	3	{2,2,2,1}	61.96	35.65
	4	{2,2,1,1,2}	57.54	31.65
	4	{2,2,2,3,3}	47.90	26.30
	5	{1,1,2,2,3,1}	100.23	55.23
29	2	{3,2,1}	250.23	130.25
	3	{1,2,2,3}	165.83	79.31
	4	{2,2,2,3,3}	141.74	67.23
	4	{1,3 2,2,1}	163.17	77.97

Table 3 | Wavelet model structure for the various stations

Station	Levels of decomposition (J)	Memory (M) at each level
6	4	{2,2,1,1,2}
20	4	{1,1,2,3,3}
29	4	{2,2,2,3,3}

equal to 2 for D1 and D2, D3 and M was equal to 3 for D4, C4.

Figures 4(a)–(c) present a comparison, using time series plots, between the forecasts and observed streamflow values for the above three stations, respectively, and, as is evident, the forecasts are in reasonable agreement with the observations. From these figures it is observed that, while the peak values are predicted to a satisfactory level, however, there is a lag between the observed and forecast peaks at some points. Additionally, it is also seen that some of the forecast values are either zero or negative.

In order to rectify this, as the next stage of model building, log transformation has been used as a pre-processor for the data. The corresponding results are presented in Figures 5(a)–(c) and show a significant improvement in performance in terms of a much improved match in both the peak values as well as time of occurrence and, importantly, without resulting in negative forecasts. Further, a closer look at the observed time series and its model-derived forecasts reveals that the proposed multiscale nonlinear model captures, to a reasonable detail, not only the major trends but also the minor fluctuations in the streamflow. The summary statistics for the wavelet model with log transformation is given in Table 4 for the three stations namely 6, 20 and 29.

Similar results were achieved for the remaining 27 stations (figures not shown) and are presented in Table 5 in terms of the CC and the other model performance measures along with the results for stations 6, 20, and 29 (all 30 stations) and are grouped under low-flow, medium-flow and high-flow regimes. As can be seen, for streamflow from all the stations, irrespective of the flow regime, the CC values are higher than 0.80.

Results for 'high-flow' regime

The 1-month-ahead multiscale wavelet-based nonlinear forecast model for station 29 (as a representative member of the 'high-flow' regime category, is presented below as Equation (17):

$$\begin{aligned} Y_{t+1} = {} & 0.57D1_{t-1} + 0.082D1_{t-2} - 0.07D2_{t-1} - 0.08D2_{t-4} \\ & - 3.23D3_{t-4} + 1.63D4_{t-1} + 1.02C4_{t-1}0.08D2_{t-1} \\ & \times D4_{t-1} + 0.0235D4_{t-1}D4_{t-4} - 0.076D2_{t-4}D4_{t-2} \\ & + 0.05C4_{t-1}D1_{t-1} - .036D3_{t-1}D4_{t-1} \end{aligned} \quad (17)$$

In general, for the high-flow regimes, it can be seen that for all stations, barring two, the model performance in terms of CC were near about 0.90. For the remaining two stations in this category, the value was nearly not as good but still yielded a CC value in the vicinity of a high of 0.85. Notably, the results for four stations, namely 03 (USGS:6214500),

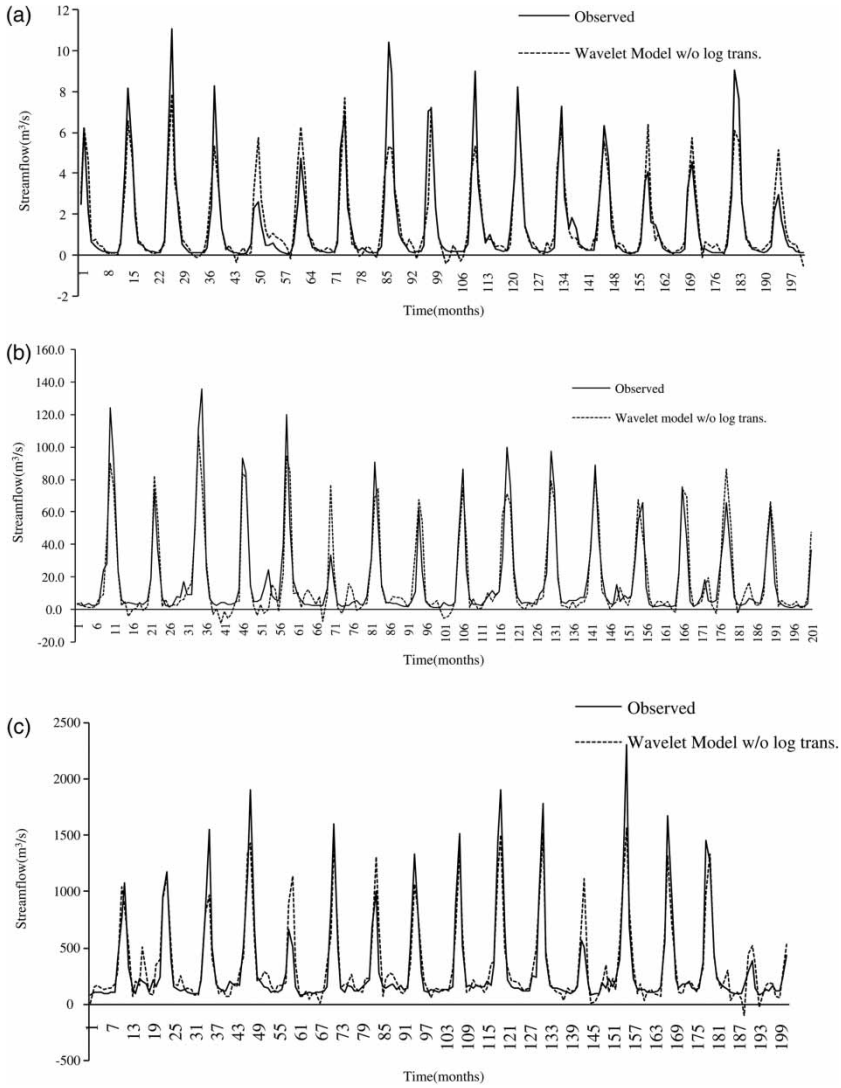

Figure 4 | Validation results of the wavelet models (without log) for various flow regimes: (a) low-flow (station 06), (b) medium-flow (station 20) and (c) high-flow (station 29).

29 (USGS:13317000), 22 (USGS:12401500) and 01 (USGS:6191500), show a marked improvement with the application of wavelet analysis when compared with results obtained with ARMA-type models and presented in Table 5. It can be seen that, for the four stations, ARMA models produced CC values of 0.71, 0.72, 0.74 and 0.77, respectively, while the nonlinear approximation method of Sivakumar (2003) produced CC values of 0.773, 0.795, 0.82 and 0.824, respectively, for these stations. In contrast, using the proposed wavelet-based, multiscale nonlinear model, the CC values have increased beyond 0.92 for all these cases.

Results for 'medium-flow' regime

Equation (18) presents the derived 1-month-ahead forecast model for station 20 as a candidate member of the 'medium-flow' regime category:

$$\begin{aligned} Y_{t+1} = {} & 1.044D1_{t-1} - 1.82D1_{t-2} + 3.72D2_{t-1} \\ & + 3.130D2_{t-4} - 1.6012D3_{t-4} + 1.9061D4_{t-1} \\ & - 0.7061C4_{t-1} - 0.60D1_{t-1}D3_{t-1} \\ & + 0.48D2_{t-1}^2 + 0.64D3_{t-4}D4_{t-1} - 0.65D3_{t-1}D4_{t-4} \end{aligned} \quad (18)$$

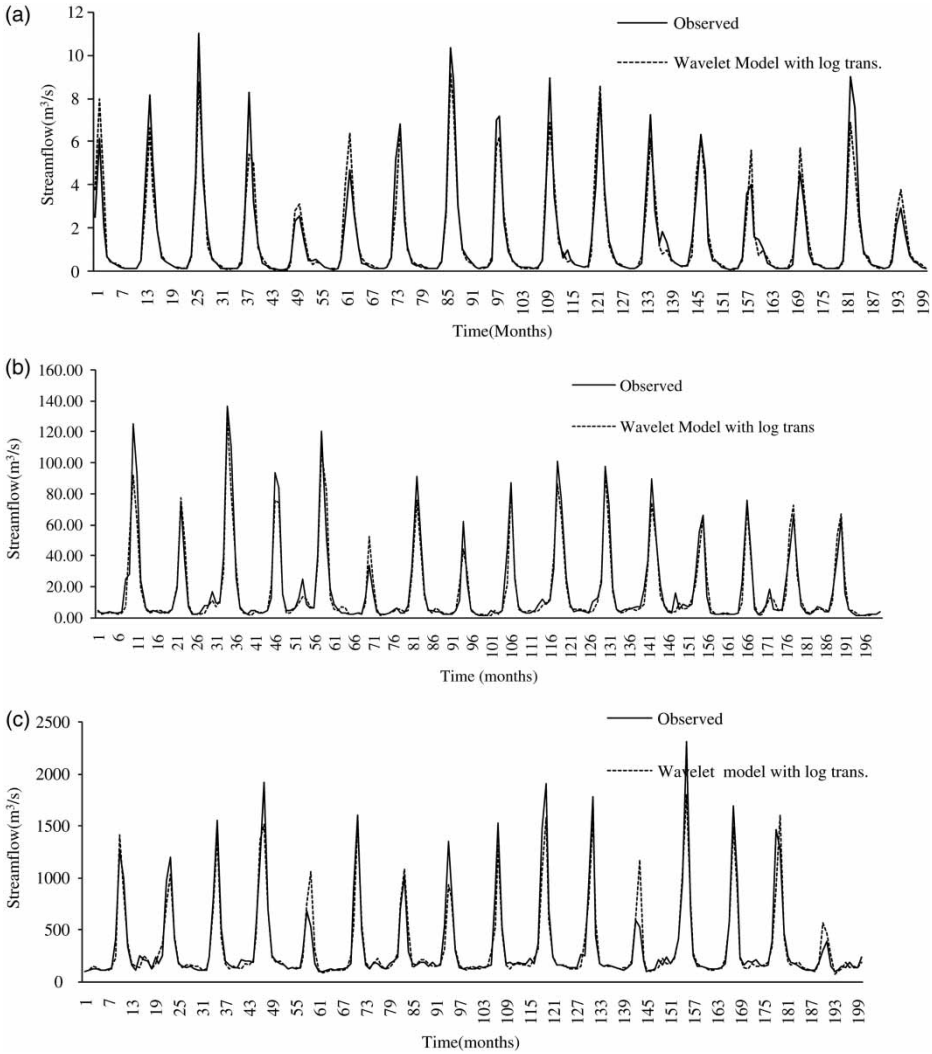

Figure 5 | Validation results of the wavelet models (with log) for various flow regimes: (a) low-flow (station 06), (b) medium-flow (station 20) and (c) high-flow (station 29).

Table 4 | Summary statistics of the wavelet-based model for stations 6, 20 and 29

Serial no.	Mean (m^3/s) Observed	Model	SD (m^3/s) Observed	Model	Maximum value (m^3/s) Observed	Model	Minimum value (m^3/s) Observed	Model	Coefficient of variation Observed	Model
6	1.48	1.52	2.2	2.10	11.03	9.6	0.073	0.074	1.47	1.4
20	18.0	17.03	27.0	25.32	136.16	131.68	1.08	1.32	1.50	1.48
29	343.81	332.65	411.88	365.92	2,312.8	1,800.23	92.85	71.57	1.197	1.10

In the case of the medium-flow regimes, the results show that six stations have CC greater than 0.90 and two stations greater than 0.80 for the wavelet model with log transformation. Similar results were obtained using the wavelet-based-without-log transformation. The RMSE values were also comparatively the same. From these

Table 5 | Result statistics of the wavelet model for the various rivers in the US

		Wavelet–Volterra model								
		With log transformation			Without log transformation			Nonlinear approx. method	Deseasonalized ARMA model	
Flow regimes	Serial number	R^2	RMSE (m^3/s)	MAE (m^3/s)	CC	RMSE (m^3/s)	CC	CC	Model order (CC)	RMSE (m^3/s)
Low-flow regimes	6	0.860	0.757	0.385	0.956	0.817	0.930	0.832	(4,1) 0.72	1.02
	7	0.73	0.72	0.322	0.86	0.603	0.322	0.862	(2,0) 0.75	1.13
	10	0.79	0.1818	0.092	0.88	2.460	0.661	0.877	(3,0) 0.71	0.260
	11	0.702	1.280	0.575	0.83	1.370	0.808	0.841	(1,0) 0.75	1.60
	27	0.730	0.670	0.412	0.87	0.690	0.852	0.764	(2,1) 0.74	0.84
	4	0.77	0.56	0.413	0.88	0.166	0.73	0.951	(4,0) 0.71	1.06
	14	0.84	0.684	0.408	0.942	0.714	0.92	0.93	(1,1) 0.84	1.12
	15	0.865	0.590	0.358	0.91	0.620	0.89	0.87	(2,0) 0.81	0.75
Med.-flow regimes	2	0.814	13.230	7.745	0.930	14.230	0.920	0.802	(3,1) 0.75	15.23
	08	0.79	2.020	1.02	0.860	2.220	0.850	0.850	(2,2) 0.81	2.81
	09	0.780	3.381	2.04	0.90	3.781	0.890	0.701	(1,0) 0.67	4.51
	20	0.84	10.23	6.87	0.93	10.89	0.923	0.810	(2,0) 0.76	12.56
	21	0.82	2.98	1.91	0.92	3.313	0.916	0.746	(6,0) 0.71	3.86
	28	0.835	5.84	3.60	0.925	6.050	0.920	0.757	(5,0) 0.70	7.12
	12	0.85	4.91	2.58	0.89	5.025	0.880	0.853	(3,1) 0.80	6.12
	13	0.87	6.12	3.49	0.942	6.45	0.93	0.922	(2,1) 0.82	7.98
	18	0.854	5.07	2.95	0.932	5.32	0.92	0.873	(3,0) 0.81	7.01
High-flow regimes	1	0.85	35.010	16.32	0.940	35.010	0.921	0.824	(2,0) 0.77	40.12
	3	0.815	86.93	50.4	0.920	87.330	0.916	0.773	(3,0) 0.71	97.42
	05	0.846	40.640	32.50	0.93	40.320	0.932	0.857	(4,0) 0.72	48.23
	16	0.854	50.14	28.72	0.942	51.32	0.937	0.873	(5,0) 0.80	55.14
	17	0.76	72.181	44.69	0.83	77.181	0.776	0.830	(3,1) 0.69	80.19
	19	0.69	44.361	31.141	0.85	46.361	0.838	0.820	(4,1) 0.74	50.47
	22	0.86	22.127	14.06	0.94	24.127	0.931	0.801	(4,0) 0.78	28.23
	23	0.86	42.190	22.750	0.94	43.190	0.930	0.834	(3,0) 0.72	48.44
	24	0.863	40.13	22.350	0.943	41.300	0.940	0.837	(3,0) 0.70	45.19
	25	0.81	48.360	34.120	0.885	49.360	0.875	0.853	(2,0) 0.71	55.47
	26	0.81	42.540	31.5	0.92	44.540	0.910	0.880	(4,0) 0.74	53.23
	29	0.83	151.740	72.230	0.930	152.740	0.910	0.795	(2,0) 0.71	164.23
	30	0.823	51.900	29.30	0.925	53.900	0.901	0.845	(1,0) 0.78	67.23

observations, it seems that for the moderate flows the log transformation is not having much influence.

Further, the results obtained by the nonlinear approximation method of Sivakumar (2003) are also seen to be inferior to those obtained by the proposed wavelet-based multiscale nonlinear model. For example, for stations 28 (USGS:13313000), 21 (USGS:12321500) and 20 (USGS:12306500), the nonlinear approximation method based on chaos theory (Sivakumar 2003) yielded CC values of 0.757, 0.746 and 0.802 and are markedly lower than the corresponding values of CC (0.92, 0.916 and 0.923) obtained using the proposed wavelet-based multiscale nonlinear model. Interestingly, the deseasonalized ARMA-type model (Hipel & McLeod 1994) yielded CC values that were a lowly 0.71, 0.70 and 0.67, respectively.

Results for 'low-flow' regime

The forecast model for 1-month-ahead lead time as derived for the monthly streamflow time series for station 6 is given below as Equation (19):

$$\begin{aligned} Y_{t+1} = {} & 0.4251D1_{t-1} - 0.181D1_{t-2} + 0.014D2_{t-1} + 1.00D3_{t-1} \\ & + 0.9605D3_{t-4} + 0.98D4_{t-1}0.437C4_{t-1} \\ & - 0.75D1_{t-1}D3_{t-1} + 0.0018D3_{t-1}^2 \\ & + 0.14D3_{t-4}D3_{t-1} - 0.825D3_{t-1}D4_{t-4} \end{aligned} \quad (19)$$

It is seen from the results presented in Table 5 that out of eight stations in the low-flow category, three stations have CC above 0.90 while the remaining five stations have CC between 0.80 and 0.90. In comparison, however, the CC values are significantly less when implemented without log transformation as seen in the case of stations 7 and 10 and may be attributed to the very low flows, near to zero, observed at these sites and, consequently, introducing errors in calibration as well as resulting in some forecasts to assume negative values.

DISCUSSION

It can be seen from Figures 3(a)–(c) that in all three cases (stations 6, 20 and 29), the overall time series carry a strong influence from the third- and fourth-level components. This is further confirmed by the wavelet spectra of the streamflow series as shown in Figures 7(a)–(c) as well as the correlations between these individual components and the parent time series for which these components have been obtained. Table 6 presents the estimated correlations and it is seen that, for station 29, the correlation coefficients between D3 and D4 levels and the overall flows are 0.417 and 0.835, respectively. These correlations are even higher for the flows corresponding to stations 6 and 20 as seen from Table 6.

It is also seen from Figures 7(a) and (b) that, in the case of stations 6 and 20, there is a strong subannual component (green band along the 4–8 months scale) which is influencing the flow, apart from the annual dynamics (black band along the 12 month scale). However, for station 29 (see Figure 7(c)), the subannual component was comparatively less significant than in the other two cases.

Table 6 | Correlation between the components and observed flow at each station

Sub-component	Station 6	Station 20	Station 29
D1	0.2870	0.2179	0.1959
D2	0.2934	0.2440	0.1793
D3	0.8407	0.7300	0.417
D4	0.7419	0.6850	0.835
C4	0.434	0.421	0.44

Further, from these results, it is concluded that the proposed wavelet-based, multiscale nonlinear model shows a significant improvement in the model performance over the chaos-theory-based nonlinear approximation method of Sivakumar (2003) as well as the conventional ARMA-type models for all 30 stations across the three flow regimes. Additionally, the proposed model, in contrast to the chaos-theory-based model of Sivakumar (2003), is amenable to on-line updating in real time.

Also, the superior performance of both the proposed models as well as the nonlinear approximation model over the linear ARMA model does suggest, albeit indirectly, the presence of nonlinear features in all 30 modelled time series of monthly streamflows and merits a further discussion.

Note on nonlinearity in the analyzed time series of monthly streamflows

Table 7 gives a comparison between the estimated values of the fraction of total variances explained by: (i) linear wavelet-based forecast model, and (ii) wavelet-based model coupled with the nonlinear Volterra kernels. It is seen from these results that the former implementation is able to achieve values of R^2 (Equation (13)) equal to 0.66, 0.70 and 0.73, respectively, for stations 6, 20 and 29. However, with the latter model having nonlinear Volterra kernels, these fractional values are significantly higher at 0.86, 0.84, and 0.83, respectively, for these stations. These results clearly establish the significant contribution of the nonlinear Volterra kernels in the forecast models given by Equations (17)–(19) wherein the cross-product terms such as $D1 \times D3$ in Equation (19), amongst other similar terms in Equations (17) and (18), highlight the impacts of degree of nonlinearity and interdependence among the scales.

Figures 6(a)–(c) show the Volterra kernels for the three representative station nos. 6, 20 and 29 selected from the low-, medium- and high-flow categories, respectively. From Figures 6(a)–(c), it is seen that numerical values of the second-order kernel is largest for station 6 (belonging to the 'low-flow' regime category) with a catchment area equal to 93 km^2 and is smallest for flows observed at station 29 and having the largest catchment area equal to 35,094 km^2. The second-order kernel values corresponding

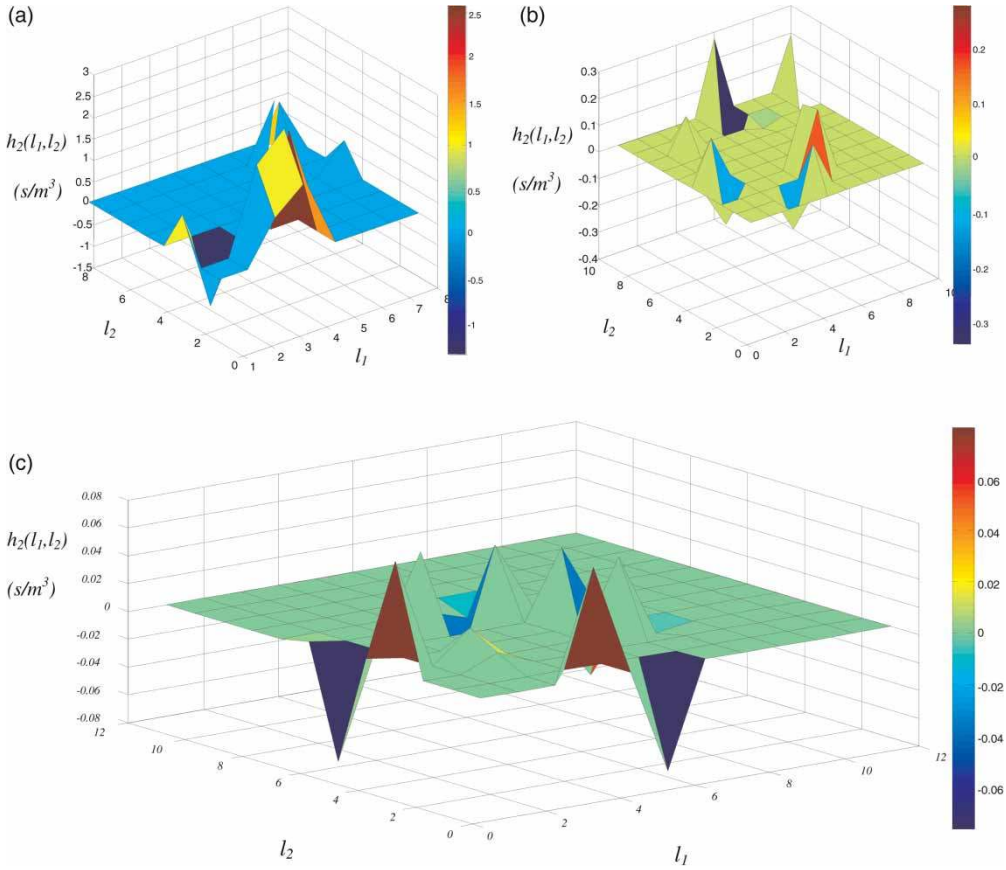

Figure 6 | Second-order volterra kernels for first-step forecasting model for (a) low-flow (station 06), (b) medium-flow (station 20) and (c) high-flow (station 29).

to station 20 are of intermediate magnitude and interestingly the catchment area drained by the river at this site also has an intermediate value between the limits defined by stations 6 (low value) and station 29 (high value). These results show that the degree of nonlinearity is maximum in the case of flows observed at station 6 and, amongst the three stations, would, therefore, be least amenable to prediction by a linear model. For station 29, based on the relatively small magnitude of the second-order kernel as has been obtained, the linear model is expected to perform the best amongst the three stations under scrutiny. This situation is also verified by the values obtained for the fraction of total variance, equal to 0.73, that could be explained by the linear component of the forecast model. In comparison, the fraction of total variance that the linear model explains for station 6 is the least at 0.66. This result is indeed consistent with the general understanding that, in large catchments, the storage-induced inertia moderates the nonlinear features and makes these catchments more amenable to modelling using linear models as compared with small catchments where nonlinear features are more pronounced and clearly evident.

Computational load and model complexity

In addition to its performance efficiency, model complexity and its corresponding computational load are concerns that also determine, to an extent, the model's acceptability. It is also reasonable to aver that computational effort, measured in terms of usage of CPU time, is a credible surrogate for model complexity.

Accordingly, therefore, CPU time usage was recorded for various model runs and a comparison of these times suggests that the proposed method is indeed simple as it involves a straightforward estimation of the Volterra kernels. While the average computational time required for the deseasonalized ARMA model was found to be 0.45 s

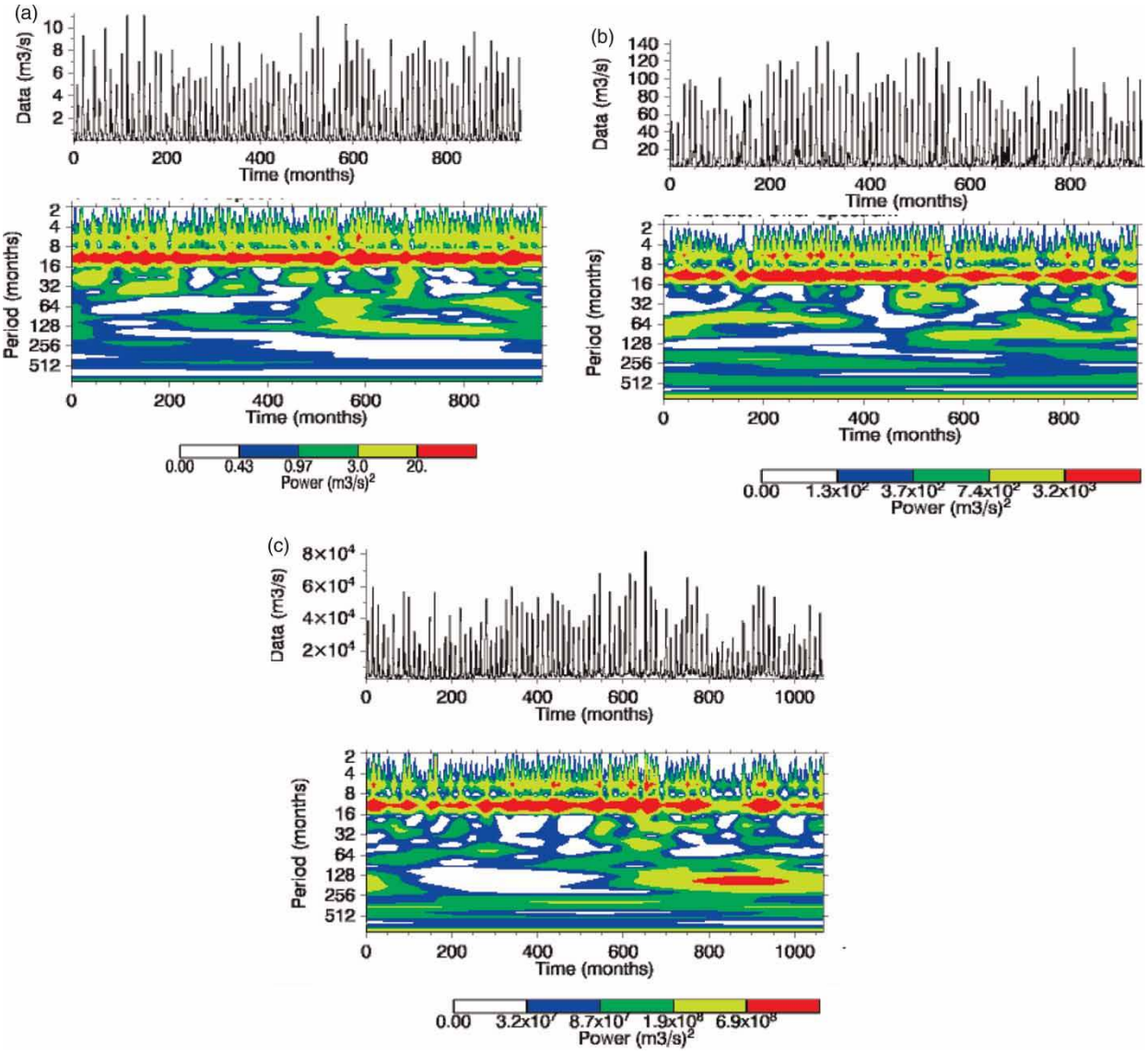

Figure 7 | Wavelet analysis of the observed streamflow series at (a) station 6, (b) station 20 and (c) station 29.

Table 7 | Fraction of total variance explained by linear and second-order kernels

Station	Variance explained by wavelet linear models	Variance explained by wavelet Volterra models
6	0.66	0.86
20	0.70	0.84
29	0.73	0.83

using a 4 GB RAM and 2.66 GHz processor-based machine, the wavelet-based nonlinear model needed an average processing time of only 0.156 s.

CONCLUSIONS

The above study has established that the modelled time series of monthly streamflows are a result of multiscale and nonlinear phenomena. Wavelets are a useful tool to represent this multiscale nature through its multiresolution and time localization capability while the Volterra representation enables the capture of the nonlinear phenomenon detected in these flow sequences and the interdependence that exist across scales.

A multiscale nonlinear model has been designed to developed 1-month-ahead forecasts of monthly streamflows for 30 stations across the western USA. The stations are maintained by USGS and drain catchments that have distinct characteristics.

The approach was based on the concept of multiscale representation of a univariate series to represent the underlying dynamics and a Volterra nonlinear model was used to integrate together the individually significant components to make predictions about future realizations of the time series of flows. The proposed multiscale nonlinear model was able to capture the major trends, wherever these were present, minor fluctuations and the peak values and yielded good results for most stations and significantly increased the forecast accuracy when compared with the performances of ARMA-type models as well as chaos-theory-based nonlinear approximation models proposed by Sivakumar (2003).

Further, in this contribution, the study has applied the *à trous* wavelet transforms, instead of the classical DWT which has the disadvantage of boundary distortion (when used for forecasting). Additionally, an examination of the Volterra kernels provides an insight into the internal dynamics of the underlying process.

As a future scope, the performance of the proposed model could be compared with that of (i) wavelet–neural network model and (ii) cross-wavelet-based model for one-step- and multistep-ahead forecasting. The proposed wavelet-based nonlinear model can also be applied for streamflow forecasting on a daily scale where the influence of nonlinearity is indeed expected to be significant (Wang *et al*. 2006). Further, it may indeed be desirable to develop precipitation–runoff forecast models that explicitly accommodate the role of causal variables such as rainfall and temperature amongst others, as may be deemed necessary.

As a further development, the improved estimation of Volterra kernels by heuristics-based evolutionary optimization techniques like genetic algorithms and particle swarm optimization may also be explored.

REFERENCES

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. *J. Hydrol.* **390** (1–2), 85–91.

Ahuja, N., Lertrattanapanich, S. & Bose, N. K. 2005 Properties determining choice of mother wavelet. *Vision Image Signal Process. IEE Proc.* **152** (5), 659–664.

Anctil, F. & Coulibaly, P. 2004 Wavelet analysis of the interannual variability in southern Québec streamflow. *J. Climate* **17** (1), 163–173.

Asefa, T., Kemblowski, M., Mckee, M. & Khalil, A. 2006 Multi-time scale stream flow predictions: the support vector machines approach. *J. Hydrol.* **318** (1–4), 7–16.

Box, G., Jenkins, G. & Reinsel, G. 1970 *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.

Burrus, C., Gopinath, R. & Guo, H. 1998 *Introduction to Wavelets and Wavelet Transforms*. Prentice-Hall, Upper Saddle River, NJ.

Carlson, R., MacCormick, A. & Watts, D. G. 1970 Application of linear random models to four annual streamflow series. *Wat. Res. Res.* **6** (4), 1070–1078.

Chang, L. C. & Chang, F. J. 2001 Intelligent control for modelling of real-time reservoir operation. *Hydrol. Process.* **15**, 1621–1634.

Chen, C., Liu, C. & Su, H. C. 2008 A nonlinear time series analysis using two-stage genetic algorithms for streamflow forecasting. *Hydrol. Process.* **22** (18), 3697–3711.

Chou, C. M. 2007 Efficient nonlinear modelling of rainfall-runoff process using wavelet compression. *J. Hydrol.* **332** (3–4), 442–455.

Corduas, M. & Piccolo, D. 2006 Short and long memory unobserved components in hydrological time series. *Physics and Chemistry of the Earth* **31** (18), 1099–1106.

Dawson, C. W. & Wilby, R. L. 2001 Hydrological modeling using artificial neural networks. *Prog. Phys. Geogr.* **25** (1), 80–108.

Diskin, M. & Boneh, A. 1973 Determination of optimal kernels for second order stationary surface runoff systems. *Wat. Res. Res.* **9** (2), 311–325.

Elshorbagy, A., Simonovic, S. & Panu, U. S. 2000 Noise reduction in chaotic hydrologic time series: facts and doubts. *J. Hydrol.* **256** (3–4), 147–165.

Firat, M. & Güngör, M. 2007 River flow estimation using adaptive neuro fuzzy inference system. *Math. Comput. Simul.* **75** (3–4), 87–96.

Firat, M. & Gungor, M. 2011 *River Flow Estimation using Adaptive Neuro-Fuzzy Inference System, Mathematics and Computers in Simulation* in press.

Firat, M. & Turan, M. E. 2010 Monthly river flow forecasting by an adaptive neuro-fuzzy inference system. *Wat. Environ. J.* **24** (2), 116–125.

French, M. N., Krajewski, W. F. & Cuykendall, R. R. 1992 Rainfall forecasting in space and time using a neural network. *J. Hydrol.* **137**, 1–31.

Garrick, M., Cunnane, C. & Nash, J. E. 1978 A criterion of efficiency for rainfall-runoff models. *J.Hydrol.* **36**, 375–381.

Gaucherel, C. 2002 Use of wavelet transform for temporal characterisation of remote watersheds. *J. Hydrol.* **269** (3–4), 101–121.

Haltiner, J. & Salas, J. 1988 Short-term forecasting of snowmelt runoff using ARMAX models1. *J. AWWA* **24** (5), 1083–1089.

Hipel, K. & McLeod, A. 1994 *Time Series Modelling of Water Resources and Environmental Systems.* Elsevier, Amsterdam.

Huang, W. R. & Catherine, M. 2008 Multiple-station neural network for modeling tidal currents across Shinnecock Inlet, USA. *Hydrol. Process.* **22**, 1136–1149.

Jain, A., Sudheer, K. P. & Sreenivasalu, S. 2004 Identification of physical processes inherent in artificial neural network rainfall runoff models. *Hydrol. Process.* **18** (3), 571–581.

Jayawardena, A. & Lai, F. 1994 Analysis and prediction of chaos in rainfall and stream flow time series. *J. Hydrol.* **153** (1–4), 23–52.

Kapetanios, G. 2003 Threshold models for trended time series. *Empirical Econ.* **28** (4), 687–707.

Karunanithi, N., Grenney, W. J., Whitley, D. & Bovee, K. 1994 Neural networks for river flow prediction. *J. Comp. Civil Engng.* **8** (2), 201–220.

Khosa, R., Joshi, S. D. & Kiran, G. 2005 Identification of non-stationarities in hydrological time series using wavelets. *Hydrol. J.* **28**, 1–2.

Kisi, O. 2009a Wavelet regression model as an alternative to neural networks for monthly streamflow forecasting. *Hydrol. Process.* **23** (25), 3583–3597.

Kisi, O. 2009b Neural networks and wavelet conjunction model for intermittent streamflow forecasting. *J. Hydrol. Engng.* **14**, 773–782.

Kisi, O. 2010 Wavelet regression model for short-term streamflow forecasting. *J. Hydrol.* **389** (3–4), 344–353.

Labat, D., Ababou, R. & Mangin, A. 2000 Rainfall-runoff relations for karstic springs. Part II: Continuous wavelet and discrete orthogonal multiresolution analyses. *J. Hydrol.* **238** (3–4), 149–178.

Lafrenière, M. & Sharp, M. 2003 Wavelet analysis of inter annual variability in the runoff regimes of glacial and nival stream catchments, Bow Lake, Alberta. *Hydrol. Process.* **17** (6), 1093–1118.

Luan, Y. 2005 Multiresolution traffic prediction: combine RLS algorithm with wavelet transform. In: *Proc. International Conference on Information Networking*, Jeju, Korea, 31 January–2 February. LNCS 3391, Springer-Verlag, Berlin, pp. 321–331.

Maria, C. M., Wenceslao, G. M., Manuel, F. B., Jose, M. P. S. & Roman, L. C. 2004 Modelling of the monthly and daily behaviour of the discharge of the Xallas river using Box–Jenkins and neural networks methods. *J. Hydrol.* **296**, 38–58.

Muftuoglu, R. F. 1991 Monthly runoff generation by non-linear models. *J. Hydrol.* **125**, 277–291.

Nash, J. & Sutcliffe, J. 1970 River flow forecasting through conceptual models part I – A discussion of principles. *J. Hydrol.* **10** (3), 282–290.

Nikolaou, M. & Mantha, D. 2000 Efficient nonlinear modelling using wavelet compression. Nonlinear model predictive control. In: *Non linear Model and Predictive Control* (F. Allgower & A. Zheng, eds.). Birkauser-Verlag, Basel, pp. 313–314.

Noakes, D., McLeod, A. & Hipel, K. W. 1985 Forecasting monthly river flow time series. *Int. J. Forecast.* **1** (2), 179–190.

Obeysekera, J. & Salas, J. 1986 Modeling of aggregated hydrologic time series. *J. Hydrol.* **86** (3–4), 197–219.

Partal, T. & Kisi, O. 2007 Wavelet and neuro-fuzzy conjunction model for precipitation forecasting. *J. Hydrol.* **342** (1–2), 199–212.

Porporato, A. & Ridolfi, L. 1997 Nonlinear analysis of river flow time sequences. *Wat. Res. Res.* **33** (6), 1353–1367.

Renaud, O., Starck, J. & Murtagh, F. 2005 Wavelet-based combined signal filtering and prediction. *IEEE Trans. Syst. Man Cybern., Part B: Cybern.* **35** (6), 1241–1251.

She, N. & Basketfield, D. 2005 Long range forecast of streamflow using support vector machine. In *Proc. ASCE World Water and Environmental Resources Conference*. American Society of Civil Engineers, Anchorage, AK.

Shensa, M. J. 1992 The discrete wavelet transform: wedding the a trous and Mallat algorithms. *IEEE Trans. Signal Process.* **40** (10), 2464–2482.

Sivakumar, B. 2003 Forecasting monthly streamflow dynamics in the western United States: a nonlinear dynamical approach. *Environ. Modell. Software* **18** (8–9), 721–728.

Sivakumar, B., Berndtsson, R., Olsson, J. & Jinno, K. 2001 Evidence of chaos in the rainfall-runoff process. *Hydrol. Sci. J.* **46** (1), 131–146.

Sivapragasam, C., Liong, S. Y. & Pasha, M. F. K. 2001 Rainfall and runoff forecasting with SSA-SVM approach. *J. Hydroinf.* **3** (3), 141–152.

Strang, G. & Nguyen, T. 1996 *Wavelets and Filter Banks*. Cambridge University Press, Cambridge.

Tesong, H., Lam, K. C. & Ng, S. T. 2001 River flow time series prediction with a range dependent neural network. *Hydrol. Sci. J.* **46** (5), 729–745.

Tong, H. & Lim, K. 1980 Threshold autoregression, limit cycles and cyclical data. *J. R. Statist. Soc. Ser. B.* **42** (3), 245–292.

Venugopal, V., Georgiou, E. F. & Sapozhnikov, V. 1999 Evidence of dynamic scaling in space-time rainfall. *J. Geogr. Res.* **104** (D24), 31599–31610.

Wang, W., Vrijling, J., Pieter, H. A. J., Gelder, V. & Ma, J. 2006 Testing for nonlinearity of streamflow processes at different timescales. *J. Hydrol.* **322** (1–4), 247–268.

Wilcox, B., Seyfried, M. & Matison, T. H. 1991 Searching for chaotic dynamics in snowmelt runoff. *Wat. Res. Res.* **27** (6), 1005–1010.

Wu, C. L. & Chau, K. W. 2010 Data-driven models for monthly streamflow time series prediction. *Engng. Appl. Artif. Intell.* **23** (8), 1350–1367.

Wu, C. L., Chau, K. W. & Li, Y. S. 2009 Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques. *Wat. Res. Res.* **45** (8), W08432.

First received 16 October 2010; accepted in revised form 16 June 2011. Available online 22 October 2011