Wavelets-based non-linear model for real-time daily flow forecasting in Krishna River

R. Maheswaran and Rakesh Khosa

ABSTRACT

In this study, a multi-scale non-linear model based on coupling a discrete wavelet transform (DWT) and the second-order Volterra model, i.e. the wavelet Volterra coupled (WVC) model, is applied for daily inflow forecasting at Krishna Agraharam, Krishna River, India. The relative performance of the WVC model was compared with regular artificial neural networks (ANN), wavelet-artificial neural networks (WA-ANN) models and other baseline models such as auto-regressive moving average with exogenous variables (ARMAX) for lead times of 1–5 days. The models were applied for the forecasting of daily streamflow at Krishna Agraharam Station at Krishna River. The WVC performed very well, especially when compared with the WA-ANN model for lead times of 4 and 5 days. The results indicate that the WVC model is a promising alternative to the other traditional models for short-term flow forecasting.

Key words | Krishna River Basin, rainfall–runoff modelling, real-time flood forecasting, Volterra model, wavelet-based models

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

INTRODUCTION

Reservoirs play a vital role in water resources management by supporting demands for irrigation purposes, hydropower, mitigating potentially deleterious environmental impacts of interventions in the natural water cycle, flood mitigation and as a reasonable insurance against droughts. Reliable forecasts of inflows into these reservoirs are an essential prerequisite for an effective operating policy development. Data-driven hydrological methods for forecasting are becoming increasingly popular (Adamowski & Sun 2010) as an alternative to the traditional statistical models that include multiple linear regression (MLR) and autoregressive moving average (ARMA) type models. More recently, artificial neural networks (ANNs) have also been used for flow forecasting applications (Tokar & Johnson 1999; Abrahart & See 2000; Rajurkar *et al*. 2004; Boucher *et al*. 2009; Fernando & Shamseldin 2009; Pramanik & Panda 2009). In general, however, all these approaches are restricted in their scope as these are suitable only for application to a stationary time series and do not perform well when the underlying natural generating systems are seen to be non-stationary (Anctil & Tape 2004; Cannas *et al*. 2006). In such cases, and when the underlying generating system shows a non-linear behaviour, these traditional approaches often do not deliver the required performance standards. As a competing alternative, recent studies demonstrate that wavelet-based approaches may provide a better and a more credible platform for modelling such non-stationary and non-linear systems; the resulting models are therefore expected to demonstrate a much improved performance.

Recent research initiatives in the area of wavelet-based modelling include studies by: (1) Kisi (2008), wherein use of a wavelet-artificial neural network (WA-ANN) model has been explored for flow forecasting in semi-arid watersheds with intermittent flow; (2) Partal (2009), who developed a coupled WA-ANN model for monthly flow forecasting in Turkey; (3) Kisi (2009), for application of a WA-ANN model for daily flow forecasting of intermittent rivers; and (4) Wu *et al*. (2009), who developed a WA-ANN model for 1–3 day forecasting.

doi: 10.2166/hydro.2013.135

In general, these and similar studies have found that WA-ANN-based models outperformed flow forecasting models that are based on, and restricted to, just the ANN-based approach. Similar findings were also reported by Nourani *et al*. (2009), who proposed WA-ANN models for rainfall–runoff modelling and forecasting. Adamowski & Sun (2010) used the coupled WA-ANN models for forecasting the daily streamflows and, based on their study, have shown that coupling these traditional models with wavelets produced better results even up to a lead time of 5 days. More recently, Nourani *et al*. (2009) developed a wavelet-based hybrid artificial intelligence (AI) model for rainfall-runoff modelling.

As a downside however, these models have also been criticized on various aspects, in particular the risk posed by overtraining of the ANN model and the difficulties of parameter estimation using heuristic methods. For example, Kisi (2011) comments that these models do not overcome the disadvantages that are normally attributed to ANN-based modelling approaches and similar sentiments were also expressed by Zhang & Benveniste (1992) who report that, with the implementation of ANNs, the model still suffers from a lack of efficient constructive methods, slow convergence and difficulties encountered in determination of the network structure and its parameters. Importantly, restriction on their use arises from the fact that these models are suitable only for systems whose underlying response characteristics do not change with time. As an alternative method, Maheswaran & Khosa (2012c) proposed a multi-scale non-linear framework using the wavelet Volterra coupled (WVC) approach for monthly streamflow forecasting. The results from the study showed that the WVC model performed better in terms of forecasting accuracy when compared with the baseline models. The model is also more efficient in terms of parsimony and computational time.

In contrast with a monthly time series, modelling a hydrological time series at daily time-steps presents higher-level challenges. For one, the degree of smoothing that results in a time series of hydrologic responses from aggregation (or, alternatively, integration) over a timescale of a month is much higher as compared with the degree of smoothing that accompanies integration over a short time-step of a day. At these latter timescales, many of the individual components of the hydrologic cycle such as interception storage, depression storage and channel flow phase as distinct from the overland flow phase are relatively more significant in influencing the streamflow regime than at timescales of aggregation of a month or higher. Due to the relatively weak influence on the flow regime of these and other individual components at monthly timescales, their combined influence is therefore lumped together in the form of one or more storage elements.

Further, a flow observation is made at a catchment's point of concentration and is clearly an aggregation of dominant elemental contributions from different parts of the catchment at different times. A longer temporal scale of integration also implies relatively larger spatial scale integration when compared with a shorter temporal scale integration of streamflow. A further consequence of the longer timescale lumping, when working with monthly streamflows, is the damping out of the manifestly non-linear nature that characterizes various hydrologic processes of the natural water cycle. It is evident that non-linear features would dominate runoff generation when the process is observed or modelled at the shorter timescale of a day. Similar behaviour was explored by Wang *et al*. (2006), where the authors found that in general the time series at daily timescales have strong non-linearity as compared with the monthly time series of the same. In addition, it is also expected that the system property of inertia would require relatively longer memory models at daily time-steps; at a monthly scale, only a fewer lagged variables are sufficient in order to capture this feature.

Following Todini (1988), modelling at relatively shorter time-steps necessitates use of the highly complex distributed differential-type models as compared with the much simpler, lumped integrated-type models that are deemed to be appropriate at longer timescales such as a month. It is therefore of interest to further validate the proposed WVC model for daily runoff forecasting at higher lead times.

The study presented seeks to address the issues: (1) testing the coupled discrete wavelet Volterra model with a Kalman-filter-based updating procedure for daily flows; (2) forecasting at multiple lead times; and (3) comparison of the results with other alternate models such as wavelet neural networks, neural networks and wavelet linear regression (WLR) models. The study also aims to investigate

the relative advantages of direct versus iterative multi-step forecasting in terms of forecasting accuracy.

STUDY AREA DESCRIPTION

This study seeks to develop a real-time daily flow forecasting framework for the Krishna River that rises in the Western Ghats region of the Indian State of Maharashtra at an elevation of about 1,337 m. It courses in a generally south-easterly direction for about 1,400 km before emptying into the Bay of Bengal. Along its course, the river picks up contributions from major tributaries such as Bhima and Tungabadhra, and enters the State of Andhra Pradesh at 782 km. Along its course, it supports various projects such as the Alamatti, Narayana, Nagarjuna Sagar, Srisailam, Jurala and Prakasam dams. The schematic of the catchment area and the gauging stations are shown in Figure 1 and the geographic locations of some of the various sites of interest are listed in Table 1. In this case study, flow forecasting models are developed for inflows into the Jurala reservoir and include contributions from Bhima Basin and the upper Krishna Basin. The Jurala project is one of the recent multipurpose projects developed across Krishna at a site downstream of its confluence with Bhima. The proposed forecast methodology utilizes feeds from upstream sites, namely Yadgir (on the Bhima River) and Huvinhedigi (on the Krishna River). The model is calibrated using the observed data at Krishna Agraharam.

DATA USED

The data used include mean daily flows for the observation sites at Yadgir, Huvinhedigi and Krishna Agraharam for a period of 7 years from 1981 to 1987, shown in Figure 2. The model calibration was performed using data corresponding to the first 5 years while validation was based on data for the last 2 years. The study was based on this limited dataset as later observations on river flows at these sites were not available. The daily basin-averaged rainfall data were also utilized, and are plotted in Figure 2.

WAVELET ANALYSIS

Wavelet analysis has become an important milestone in spectral analysis due to its multi-resolution and localization capability both in time and frequency domain and has been extensively applied in the area of time series analysis and prediction. For a comprehensive discussion on the theory of wavelets and its applications, refer to Burrus *et al.* (1998).

Wavelet decompositions at various scales (frequencies) often reveal the underlying low- and high-frequency components of the observed series and, importantly, the wavelet-based approach also enables localization of these frequencies in time. For practical applications (as in the study of noise reduction models for communication systems and image and signal compression), discrete wavelet transform (DWT) is usually preferred.

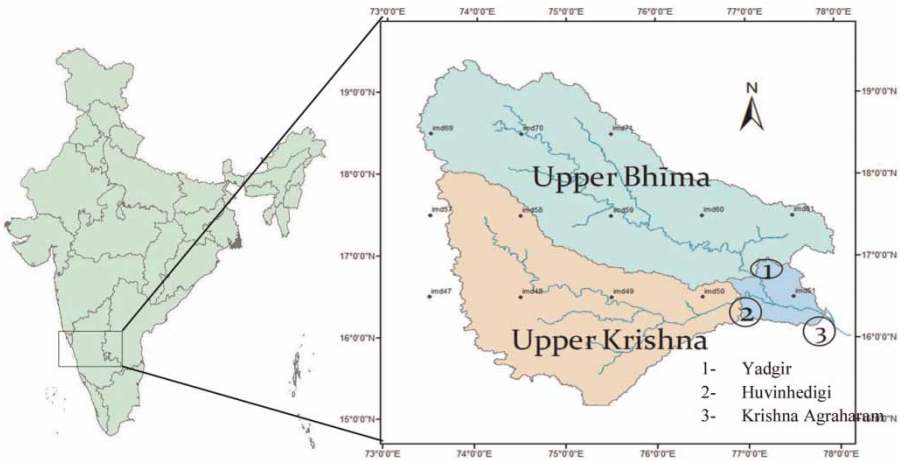

Figure 1 | Map showing the catchment area and the discharges sites.

Table 1 | Geographic locations of the discharge observation sites

			Geographical location					
Station No.	**Name**	**River**	**Latitude**			**Longitude**		
1	Huvinhedigi	Krishna	16	29	07	76	55	07
2	Yadgir	Bhima	16	44	03	77	07	18
3	P.D. Jurala Project	Krishna	16	15	00	77	51	00
4	K. Agraharam	Krishna	16	15	00	77	51	00

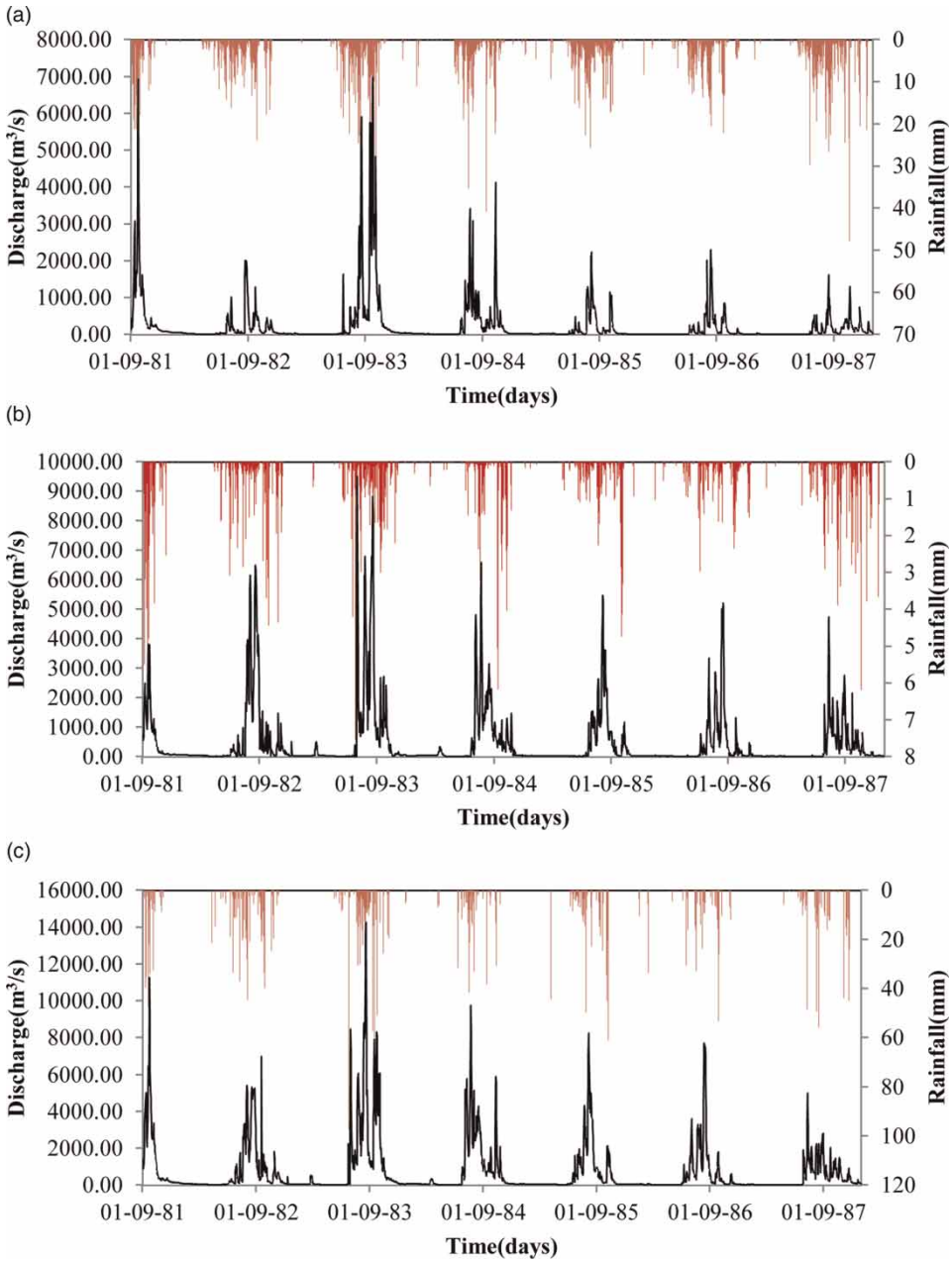

Figure 2 | Runoff hydrograph observed at: (a) Yadgir; (b) Huvinhedigi; and (c) Krishna Agraharam.

The DWT is an orthogonal function which can be applied to a finite group of data. Functionally, it is very much like the discrete Fourier transform, in that: (1) both transforms are convolutions; (2) the transforming functions are orthogonal; (3) a signal passed twice through the transformation is unchanged; and (4) the input signal is assumed to be a set of discrete-time samples. A point of difference, however, is in the nature of the basis function; in the case of a Fourier transform this is a sinusoid, whereas the wavelet basis is a set of functions which are defined by a localized wavelet function.

A typical discrete wavelet function can be represented:

$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k) \tag{1}$$

where $\psi(t)$ is the mother wavelet and j and k are the translation and dilation indices.

The main advantage of DWT is that it retains only a minimal set of wavelet coefficients that are required for full recovery of the signal by employing a process known as decimation. Decimation is carried out in such a way 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.

However, in the case of forecasting studies, the process of decimation results in reduction in information (as the number of wavelet coefficients is halved with each move to a coarser level). As a consequence, there is less information available to train the forecasting model at the coarser level leading, reducing the overall prediction accuracy. The problem of decimation may be overcome by introducing the stationary or *à trous* wavelet transform proposed by Shensa (1992). The basic idea of the stationary wavelet transform is to fill the resulting gaps using redundant information obtained from the original series.

In this approach, the wavelet decomposition is derived by passing the given time series through a low-pass filter; the subsequent derivation of details and the smoothed version then becomes possible. For example, consider the original time series $x(t)$ which may also be denoted c_0 or

$$c_0(t) = x(t) \tag{2}$$

Further smoother versions of $x(t)$ may be derived from:

$$c_i(t) = \sum_{l=-\infty}^{\infty} h(l)c_{j-1}(t + 2^{i-1}l) \tag{3}$$

In Equation (3), l takes the value 1 to j (level of decomposition) and h is a low-pass filter with compact support. The length and characteristics of the low-pass filter will depend on the type of wavelet used. The simplest wavelet is the Haar wavelet with a low-pass filter specification given by (1/2, 1/2). Similarly, the filter values for the B_3 spline wavelet is defined as (1/16, 1/4, 3/8, 1/4, 1/16). Using the smoother versions of $x(t)$ at level i and i–1, the detail component of $x(t)$ at level i is defined:

$$d_i(t) = c_{i-1}(t) - c_i(t) \tag{4}$$

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

$$x(t) = c_p(t) + \sum_{i=1}^{p} d_i(t) \tag{5}$$

Unlike in classical DWT, decimation is avoided here, resulting in components at different scales being the same length.

Treatment of circular effect or boundary of wavelet decomposition

The selection of the suitable wavelet and treatment of the boundary for wavelet decomposition is based on the discussion provided in Maheswaran & Khosa (2012b). The latter issue assumes significance especially when models are designed for forecasting applications.

STATISTICAL TESTING FOR MULTI-SCALE DYNAMICS, NON-LINEARITY AND LONG MEMORY

The applicability of models for a given data series depends on the characteristics of the time series; accordingly, it is

necessary to investigate the properties of the time series in terms of non-linearity, multi-scale non-stationarity and long memory. The following sections provide an overview of the methods used for the analysis followed by a description of the application of these tests for the given data.

Multi-scale analysis using wavelet transforms

Indirect examination for the presence of various stochastic and scale attributes of the rainfall and runoff processes was carried out by deriving wavelet scalograms for each of the time series. An aim of this exercise is the derivation of signature details such as frequency time localization and dominant oscillations of the respective observation sets. The study by Torrence & Compo (1998) formulated significance tests against reasonable background spectra, or an appropriate null hypothesis which encouraged further application of wavelet analysis to hydrologic studies. Examples include studies of Karstic systems (Labat *et al*. 2000), characterization of daily streamflow (Smith *et al*. 1998; Saco & Kumar 2000), study of monthly inflows to a reservoir (Coulibaly *et al.* 2000), study of long range correlations in discharge time series (Audit *et al*. 2002; Kantelhardt *et al*. 2002, 2003) and regionalization of daily streamflow in Australia (Zoppou *et al.* 2002). More recently, Labat *et al.* (2000) investigated the temporal variability of annual discharge series of different rivers and showed that continuous wavelet analysis using Morlet wavelets could identify the underlying periodicities with periods ranging from inter-annual (5–8 years) to decadal (12–15 years), bi-decadal (approximately 28 years) and 4–70-year oscillations. In the present study, significance tests suggested by Torrence & Compo (1998) have been applied to obtain detailed insights into the process under scrutiny.

Investigating for process non-linearity

Exploring and prospecting for different kinds of non-linearity individually, which may be expected in typical hydrologic processes, is generally intractable and perhaps conceptually flawed. Therefore, it is generally recommended that investigations for possible non-linear phenomena should evaluate the overall 'non-linear' content rather than strive for an objective discrimination between each of the hypothesized types. There are a wide variety of methods presently available to test for linearity (or non-linearity) and may be divided into two broad categories: (1) portmanteau tests, which test for departure from linear models without specifying alternative models; and (2) tests designed for some specific alternatives (Wang *et al.* 2006). Patterson & Ashley (2000) tested six portmanteau methods on eight synthetically generated non-linear series of different types and concluded that the BDS (Brock *et al.* 1987) test was better and clearly stood out in terms of overall power against a variety of alternatives. Kim *et al.* (2003) demonstrated the effectiveness of the BDS test when applying the procedure to test for the presence of non-linearity in the residuals of fitted models.

Under the overall objective of estimating the non-linear content in the given time series, Wang *et al.* (2006) have shown that the presence of such non-stationary features as a trend or a level change, or even the presence of periodic features, may affect the estimation; they have therefore recommended that removal of these non-stationary features should precede any study for non-linear content. This generally requires extraction of the linear structure using an estimated linear filter, and then testing the standardized residuals for non-linearity using the BDS test statistic. The standard procedure for applying the BDS test is depicted in Figure 3.

Test for long memory

The presence of long memory features indicates the influence of long past events on current states that the given process is likely to attain. There are several methods available in the literature for estimating the presence of long memory, and such a presence is chiefly evaluated in terms of the Hurst exponent H (e.g. Beran 1994). Further discussions on estimation of the Hurst exponent H have also been presented by Rao & Bhattacharya (1999) and Wang *et al.* (2006). In the present study, the recommendations proposed by the aforementioned investigators have been followed to obtain estimates of the Hurst exponent H for each of the four case studies. Essentially, this analysis required pre-processing of the given time series of observations (log-transformation, deseasonalization and detrending). Table 2 shows the estimated H exponents

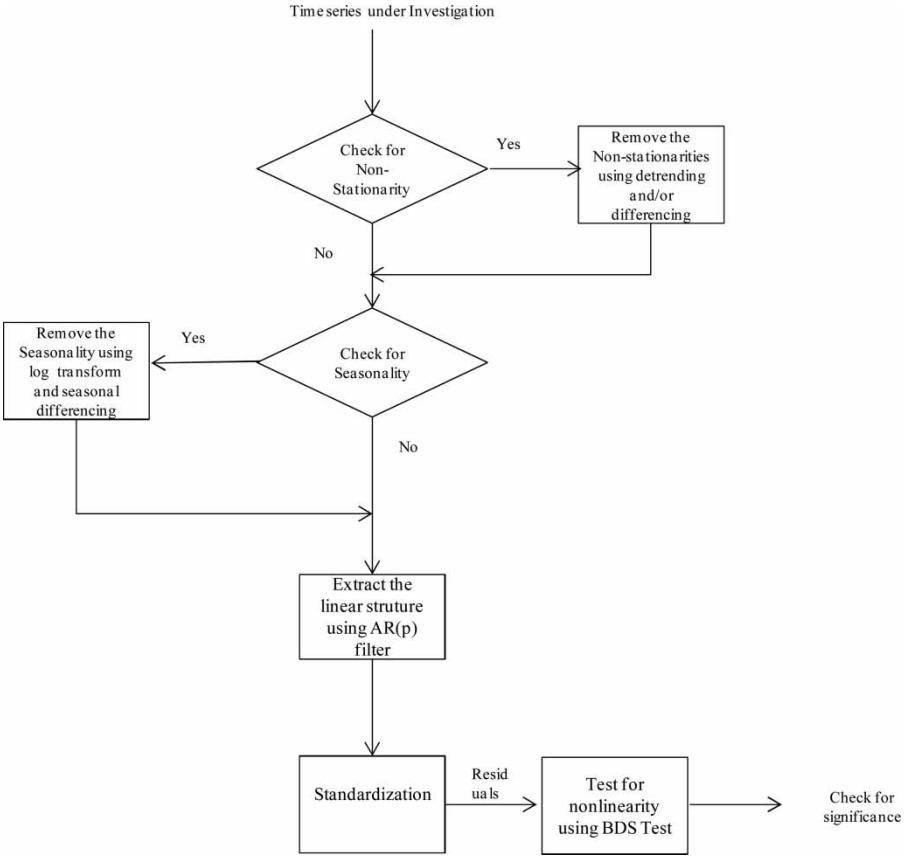

Figure 3 | BDS test procedure.

Table 2 | BDS test results along with the *p*-values; M_e: embedded dimension

		$M_e = 2$		$M_e = 3$		$M_e = 4$	
Case study	**Series**	**Statistic**	***p*-value**	**Statistic**	***p*-value**	**Statistic**	***p*-value**
I	Yadgir	30.41	0	30.09	0	29.31	0
	Huvinhedigi	32.14	0	31.29	0	30.11	0

obtained using two methods: (1) wavelet-based method; and (2) rescale-range-based method. Table 2 also presents a comparison between these estimates obtained for the given observed time series for both cases before and after pre-processing.

APPLICATION TO CASE STUDIES

Figures 4(a) and 4(b) show the wavelet spectrum for the daily basin mean rainfall and runoff observed at Yadgir station, clearly revealing the dominant annual feature (seen as a dark band). There is also evidence of sub-annual cycles with periods of 128–256 days. The analysis of the wavelet spectrum further reveals that high flows occur regularly in a year, but the peak values and duration of high flows is highly variable. Figures 5(a) and 5(b) show the wavelet spectrum for the daily mean rainfall and runoff at Huvinhedigi. It can be seen that the intensity of the annual cycle is decreasing with time, presumably an impact of developmental schemes such the dams at Alamatti and Narayanpur across the Krishna River upstream of Huvinhedigi.

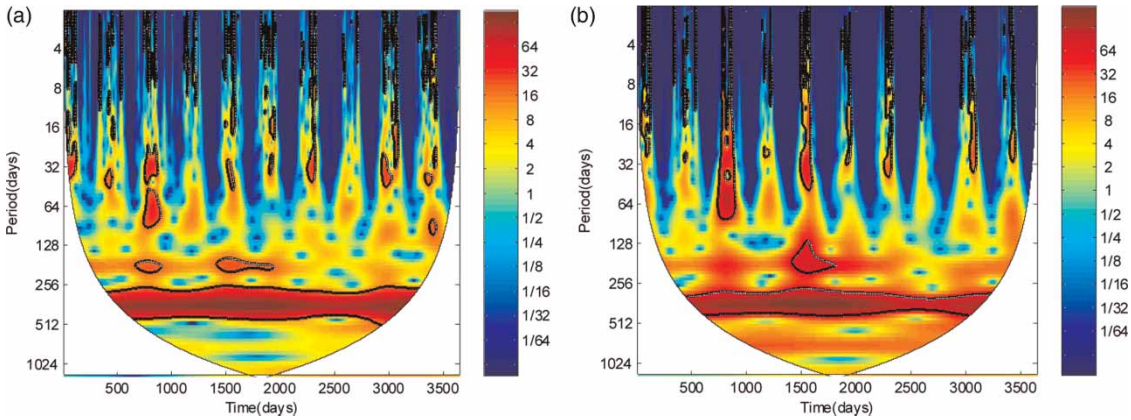

Figure 4 | Normalized local wavelet power spectrum of: (a) rainfall; and (b) runoff observed at Yadgir using Morlet wavelet.

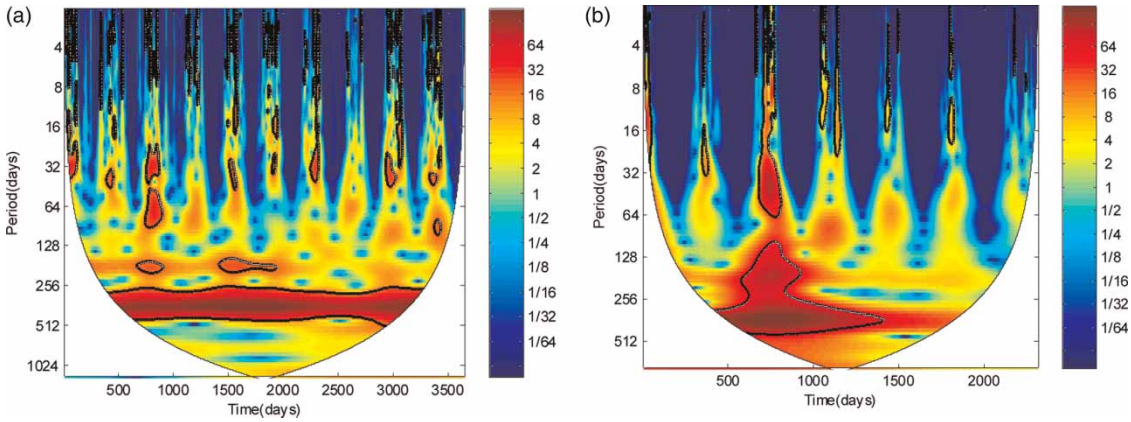

Figure 5 | Normalized local wavelet power spectrum of: (a) rainfall; and (b) runoff observed at Huvinhedigi using Morlet wavelet.

Further, as shown in Table 3, BDS test results for the daily streamflow time series also reveal that there is a stronger non-linear content present in daily streamflow processes, even after discounting for the effects of seasonal variance.

The estimates obtained for the Hurst exponent H (Table 2) are indeed interesting as these reveal the presence of strong long memory features in the daily runoff time series observed at Yadgir and Huvinhedigi. Similar findings have also been reported for daily runoff by other investigators including Lawrance & Kottegoda (1977), who also advocated use of fractional models for these datasets. Conclusive inference about the nature of the underlying generating process is however possible only after a formal examination of the significance of these estimates using prescribed statistical techniques.

Analysis of the results from the BDS test indicates the presence of a strong non-linearity in the daily flows time series, consistent with the general understanding of the processes. Additionally, the multi-scale analysis and the Hurst

Table 3 | Estimation of H for different case studies

		Hurst exponent before pre-processing of time series		Hurst exponent after pre-processing of time series	
Case study	**Series**	**Wavelet-based**	**R/S**	**Wavelet-based**	**R/S**
I	Yadgir	0.91	0.85	0.84	0.771
	Huvinhedigi	0.89	0.87	0.82	0.72

exponent estimation reveal the underlying long-term component at both annual and sub-annual scales.

DESCRIPTION OF MODELS

Wavelet Volterra coupled

Figure 6 presents a general schematic of the structure of the coupled forecast model design. As depicted in the figure, an input signal $Y = (y_1, \ldots y_{n-1})$ is decomposed to obtain wavelet coefficients at different scales using the *à trous* wavelet transform. The resulting wavelet decomposition at various levels are then appropriately integrated using the second-order multiple-input–single-output (MISO) Volterra formulation. The Volterra series representation of a non-linear time-invariant system with memory is based on a simple extension of the Taylor series expansion for non-linear autonomous causal systems with memory. (For more details, see the appendix, available online at http://www.iwaponline.com/jh/15/135.pdf).

To understand the formulation, let $u_1, u_2 \ldots u_J$ denote the wavelet coefficients at each scale and let scaling coefficients be denoted u_{J+1}, where J is the coarsest level of decomposition. The wavelet coefficients and scaling coefficients of the original series are non-linearly convolved using the second-order Volterra representation within a MISO framework. If J denotes the level of decomposition, N the number of inputs, m the memory length at each level and ξ_t the model noise including modelling errors and the unobservable disturbances, the multi-scale non-linear relationship may be written:

$$y(t) = \sum_{n=1}^{J+1}\sum_{\tau=1}^{m} h_1^{(n)}(\tau)u_n(t-\tau) + \sum_{n=1}^{J+1}\sum_{\tau_1=1}^{m}\sum_{\tau_2=1}^{m} h_{2s}^{(n)}(\tau_1,\tau_2)\,u_n(t-\tau_1)\,u_n(t-\tau_2) + \sum_{n_1=1}^{J+1}\sum_{n_2=1}^{n_1-1}\sum_{\tau_1=1}^{m}\sum_{\tau_2=1}^{m} h_{2x}^{(n_1,n_2)}(\tau_1,\tau_2)u_{n_1}(t-\tau_1)u_{n_2}(t-\tau_2)+\xi_t \quad (6)$$

where the first-order kernels $h_1^{(n)}$ describe the linear relationship between the nth input u_n and the output signal y; the second-order self-kernels $h_{2s}^{(n)}$ describe the second-order non-linear relation between the nth input u_n and y; and the second-order cross-kernels $h_{2x}^{(n_1,n_2)}$ describe the second-order non-linear interactions between each unique pair of inputs (u_{n1} and u_{n2}) as they affect y.

Equation (8) can be simplified by combining the last two terms to yield:

$$y(t) = \sum_{n=1}^{J+1}\sum_{\tau=1}^{m} h_1^{(n)}(\tau)u_n(t-\tau) + \sum_{n_1=1}^{J+1}\sum_{n_2=1}^{J+1}\sum_{\tau_1=1}^{m}\sum_{\tau_2=1}^{m} h_2^{(n_1,n_2)}(\tau_1,\tau_2)u_{n_1}(t-\tau_1)u_{n2}(t-\tau_2)+\xi_t \quad (7)$$

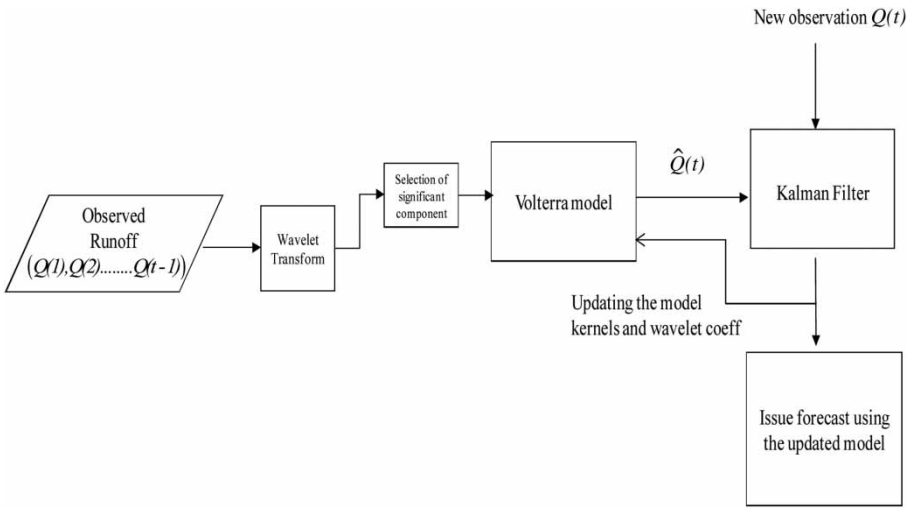

Figure 6 | Wavelet Volterra coupled model.

It now remains to estimate kernels h_1 and h_2. Equation (9) can be further simplified by considering each of the lagged variables $u_1(t–1)$, $u_1(t–\tau)$,… $u_2(t–1)$, $u_2(t–\tau)$,… as separate variables $d_1(t)$, $d_2(t)$, $d_3(t)$,… $d_{N1}(t)$. Equation (9) can be written:

$$y(t) = \sum_{l=1}^{N_l} h_1(l) d_l(t) + \sum_{l_1=l}^{N_l} \sum_{l_2=l}^{N_l} h_2(l_1, l_2) d_{l_1}(t) d_{l_2}(t) \tag{8}$$

More clearly,

$$d_l(t) = \{x_k(t) \quad \text{for} \quad 1 \le k, l \le J+1$$
$$d_l(t) = \{x_k(t-\tau) \quad \text{for} \quad 1 \le k \le J+1 \quad \text{and} \quad J+1 < l \le N_l$$

where τ (=1,…m) is the lagged value; J is the level of decomposition; and N_l is the total number of lagged variables.

In Equation (10), h_1 and h_2 represent the Volterra kernels to be estimated. It is seen that the number of parameters to be estimated increases proportionately with the number of inputs and/or the process memory increases leading to an increased computational burden and a severely compromised estimation. In order to handle these computational difficulties, Chen *et al.* (1989) and later Wei *et al.* (2004) proposed the use of the orthogonal least-squares technique as a preferred estimation approach. This method is better than the ordinary least-squares technique for handling modelling situations where there is a possibility of multi-collinearity among vectors that constitute the coefficient matrix.

Further, as shown in Figure 6, the proposed formulation is recursively updated in real time using the well-known Kalman filter formulation. Here, the Volterra kernel of the second-order Volterra model may be predicted, corrected and updated using the Kalman filter. This recursive updating increases the accuracy of the estimates and also reduces the standard error involved in the estimates as new observations become available. For further details of the model formulation, see Maheswaran & Khosa (2012a, 2012c).

Wavelet neural network

The process of WA-ANN model development begins with the decomposition of the observed process $Q(t)$ into DWT $\{D_1(t), D_2(t),\ldots D_J(t), C_J(t)\}$ by *à trous* algorithm at its various resolution levels J. The modelling philosophy behind this class of models requires use of the aforementioned scale-wise decompositions $D_1(t)$, $D_2(t)$,… $D_J(t)$, $C_J(t)$ at time t as inputs to the ANN component which is then trained appropriately to yield, as the network output, the future states of the observed time series $Q(t+h)$ at t where h is the length of lead time. The number of hidden layers and number of hidden nodes within each hidden layer are determined by trial and error. Further, in this method the weights (parameters) are learned using any of the numerous network learning algorithms that are available in the literature. As explained above, the key feature of these models (also referred to as WA-ANN models) is that the model under discussion is a two-component framework. The first component yields the various wavelet decompositions while the second ANN-based component uses these decompositions as input to yield forecasts on the future states of the given process as observed at regular intervals of time.

The choice of the most 'appropriate' network training algorithm is usually resolved by means of a trial-and-error-based judgement; there are diverse opinions on the use of a specific network algorithm. For example, Wang *et al.* (2006) use the traditional back-propagation algorithm and have endorsed the procedure in terms of its suitability for training WA-ANN models. At the other end of this spectrum is the position taken by Adamowski & Sun (2010) and Adamowski & Karapataki (2010) who rate the Levenberg algorithm as the best for these WA-ANN models. In another study, Tiwari & Chatterjee (2010) applied a bootstrap ANN for training WA-ANN models as their algorithm of choice.

In the present study, WA-ANN models have been implemented with various training algorithms and their performance assessed through a comparative evaluation; only the 'best' results are presented here for brevity. The input variables for the ANN module of the benchmark WA-ANN model are selected from the various wavelet decompositions derived at various scales. The selection of the most significant among these is based on trials that evaluate the relative strengths of individual cross-correlation functions (CCF) between the different wavelet decomposition components and the original process under scrutiny. The WA-ANN model scheme is depicted by Figure 7.

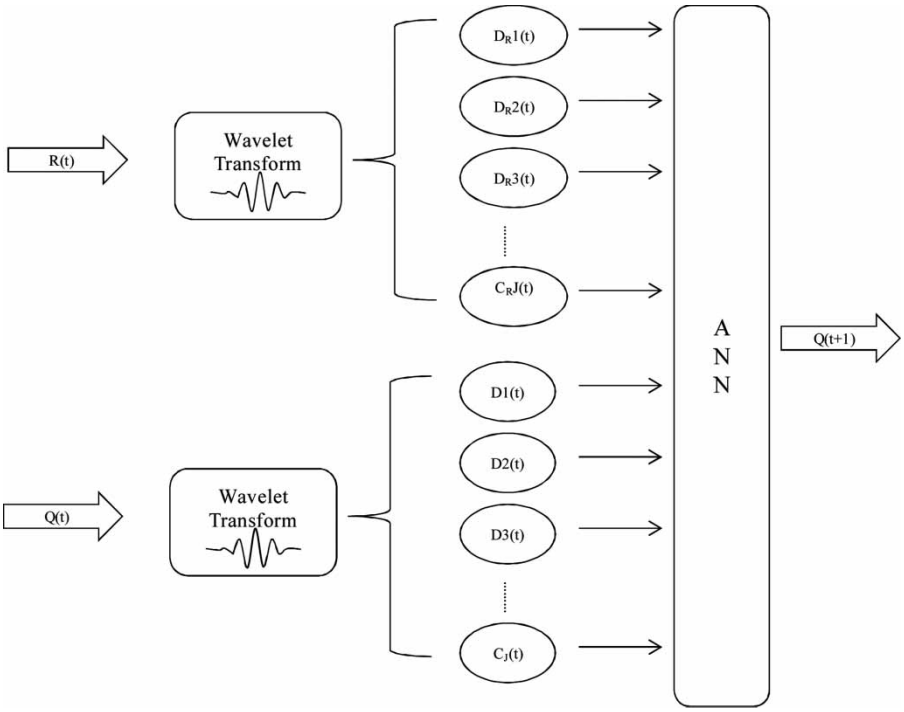

Figure 7 | WA-ANN model for multivariate forecasting.

Artificial neural network

The innate ability of ANNs to 'train' and 'learn' the outputs from a given input render them able to simulate large-scale arbitrarily complex non-linear problems (Rumelhart *et al.* 1994). During the learning process, the network undergoes a loop of calculation such that, for each pass, the network proceeds through a specified sequence of inputs to calculate the outputs and the associated errors. The network adjusts its weight as each input vector is presented in a sequence and the loop of calculations continues until the performance goal is attained. The goal of network training is not only to learn an exact representation of the training data itself, but to build a statistical model of the processes that generate such data. The training process involves adjustment of link weights and the process continues until an acceptable network performance is achieved across all training datasets and a minimum error between the desired output and the neural network output is achieved.

The most widely used neural network structure for hydrological applications is the multilayer perceptron (MLP) which is also capable of non-linear pattern recognition and memory association as discussed by Nayak *et al.* (2004). The MLPs have neurons organized in layers and each neuron is connected only to neurons in adjacent layers. The architecture of a typical feed-forward multilayer neural network is shown in Figure 8. Typically, for forecasting applications the input vector consists of (N–1) past observations of the time series while the output vector is the forecast value of future state. Following pre-processing of the data for standardization, all ANNs in this study were trained using supervised training algorithms

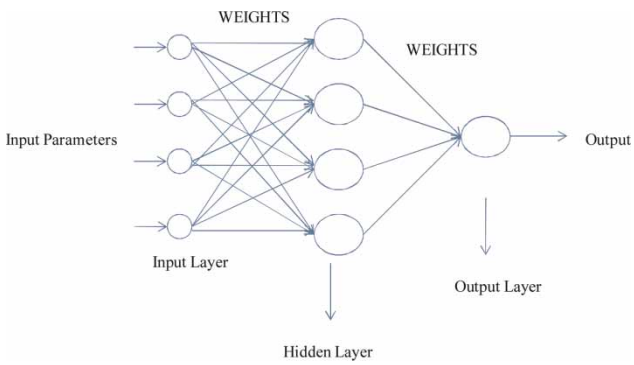

Figure 8 | Artificial neural network model.

which minimized the mean squared error (MSE) objective function. The ANN models were developed using the Neural Networks Toolbox of MatLab version 7.6.0 (R2008a).

Wavelet linear regression

A WLR model is obtained by combining DWT with linear regression (LR). The wavelet regression (WR) model is a LR model that uses sub-time series components derived from DWT of the original observed process. The WR model structure developed in the present study uses a predefined number of the aforementioned sub-time series component details as inputs to the LR component, and its output is deemed to be the computed/simulated output process.

Evidently, the model development process consists of finding a suitable wavelet function and an appropriate number of decomposition levels, as well as their relative effectiveness as a measure of the relative degree to which each component would be influencing the original time series. While there are no clear guidelines available in the literature for selecting the type of wavelets, Nourani *et al.* (2009) provide recommendations on the selection of the optimal number of decomposition levels J. Nourani *et al.* suggest that the optimal level of decomposition J may be fixed according to:

$$J = \log_2(N) \tag{9}$$

The selection of components based on their relative effectiveness has similarly been explored by Kisi (2009) who suggests the use of a correlation coefficient, estimated between each of the various wavelet decomposition components and the given observed process, as the basis for evaluating the relative significance of these individual decomposition components.

MODEL APPLICATION

Wavelet Volterra coupled

In this approach, the rainfall and runoff signals are decomposed into wavelet and scaling components at different scales. The approximation and the detail components represent the slow and fast components that make up the integrated history of the time series of observed flows. These components are combined through the Volterra framework to forecast future flows, followed by a Kalman-filter-based updating procedure that uses the newest (current) observation to update the model for use at the next time-step for further forecasts.

For the current study, the observed flows at Yadgir and Huvinhedgi and representative weighted sub-basin rainfall are the model input variables. For each of these time series, the corresponding scale-specific decompositions are derived using the Haar wavelet transform (selection of the mother wavelet as per the guidelines given by Maheswaran & Khosa 2012b). The significant wavelet components are selected based on the correlation between different components and the observed flows at Krishna Agraharam (see Table 4).

The results presented in Table 4 suggest that the approximation components – namely $AQ_{h,t-1}$, $AQ_{y,t-1}$, $AQ_{h,t-2}$ of the flow time series, $AR_{ka,t-1}$ of rainfall time series and details component of the flow time series $DQ_{h,t-1}$ – made a significant contribution to the overall variability in observed flows at Krishna Agraharam. These components were taken as the inputs for the Volterra framework and orthogonal least squares–error reduction ratio (OLS-ERR) algorithm was used to find the best regressors as per the schematic of Figure 9. Following calibration, the parametric form of the proposed WVC models for lead times of 1, 2 and 3 days are shown in

Table 4 | Correlation analysis of different wavelet components with observed flows at Krishna Agraharam

Wavelet component of different time series	Correlation between D_{t-1} and $Q_{ka,t}$
$DQ_{h,t-1}$	0.26
$AQ_{h,t-1}$	0.87
$AQ_{h,t-2}$	0.82
$DQ_{y,t-1}$	0.09
$AQ_{y,t-1}$	0.82
$AR_{ka,t-1}$	0.56

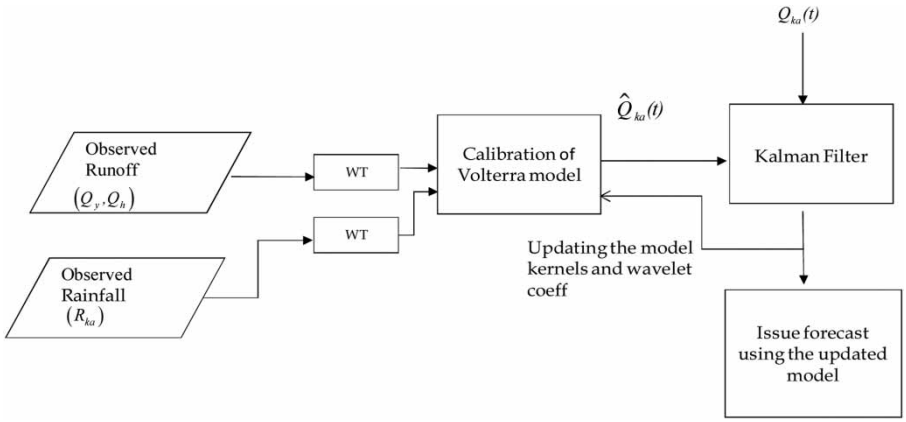

Figure 9 | Wavelet Volterra model for flow forecasting at Krishna Agraharam.

Equations (12–14):

$$Q_{ka,t+1} = 2.45A2Q_{h,t} - 1.18A2Q_{h,t-1} + 0.98A2Q_{y,t} + 1.55D1Q_{h,t} + 0.69D1Q_{h,t-1} + 0.0017A2Q_{h,t}A2Q_{y,t} - 0.00012A2Q_{y,t}A2Q_{h,t-1} - 0.0054A2Q_{y,t}A1R_{ka,t} - 0.0458A1R_{ka,t}D1Q_{y,t} \quad (10)$$

$$Q_{ka,t+2} = 3.4A2Q_{h,t} - 1.8A2Q_{h,t-1} + 0.78A2Q_{y,t} + 1.1D1Q_{h,t} + 0.67D1Q_{h,t-1} + 0.0011A2Q_{h,t}A2Q_{y,t} - 0.0002A2Q_{y,t}A2Q_{h,t-1} - 0.054A2Q_{y,t}A1R_{ka,t} - 0.058A1R_{ka,t}D1Q_{y,t} + 0.04A1R_{ka,t-1}D1Q_{y,t-1} + 0.04A1R_{ka,t-1}D1R_{ka,t-1} \quad (11)$$

$$Q_{ka,t+2} = 1.4A2Q_{h,t} - 0.8A2Q_{h,t-1} + 0.44A2Q_{y,t} + 0.15D1Q_{h,t} + 0.17D1Q_{h,t-1} + 0.0014A2Q_{h,t-1}A2Q_{y,t-1} - 0.002A2Q_{y,t}A2Q_{y,t-1} - 0.041A2Q_{y,t}A1R_{ka,t} - 0.08A1R_{ka,t}D1Q_{y,t} + 0.021A1R_{ka,t-1}D1Q_{y,t-1} + 0.012A1R_{ka,t-1}D1R_{ka,t-1} \quad (12)$$

The forecast results of the WVC model for lead times of 1–3 days are shown in Figures 10(a)–10(c). Examination of these results shows that the WVC model closely picks up the peak discharge values up to a lead time of 3 days. The statistics of the model results are presented in the Results and Discussion sections.

Neural network

For comparison, a feed-forward neural network model with three hidden layers has been adopted in the study and the number of hidden nodes was selected using a trial and error approach. The Levenberg–Marquardt back propagation (LMBP) approach has been adopted as the training algorithm as it has been shown to be faster and finds better optima for a variety of problems than the other methods (Coulibaly & Baldwin 2005). In this study, different input combinations were tested and the best model configuration was selected. The best model in terms of forecast accuracy was found to have a configuration of (4, 3, and 1).

Wavelet neural network

As before, the observed rainfall and runoff time series are decomposed into sets of wavelet coefficients by the *à trous* algorithm at the resolution level J. Subsequently, an ANN model is developed in which the significant wavelet decomposition components at time t are used as inputs to the ANN component to obtain forecasts $Q_{ka}(t+h)$ of the future state of the observed output process at time $(t+h)$, where h is the length of lead time. The number of hidden layers and hidden nodes of each hidden layer are determined by trial and error, while the link weights (parameters) for a selected ANN are determined using the LMBP network algorithm.

In this study, it was found that two levels of decomposition produced optimal results in comparison with the higher levels of decomposition which were used as the input variables for ANN models. For the WA-ANN models, ANN networks consisting of an input layer with

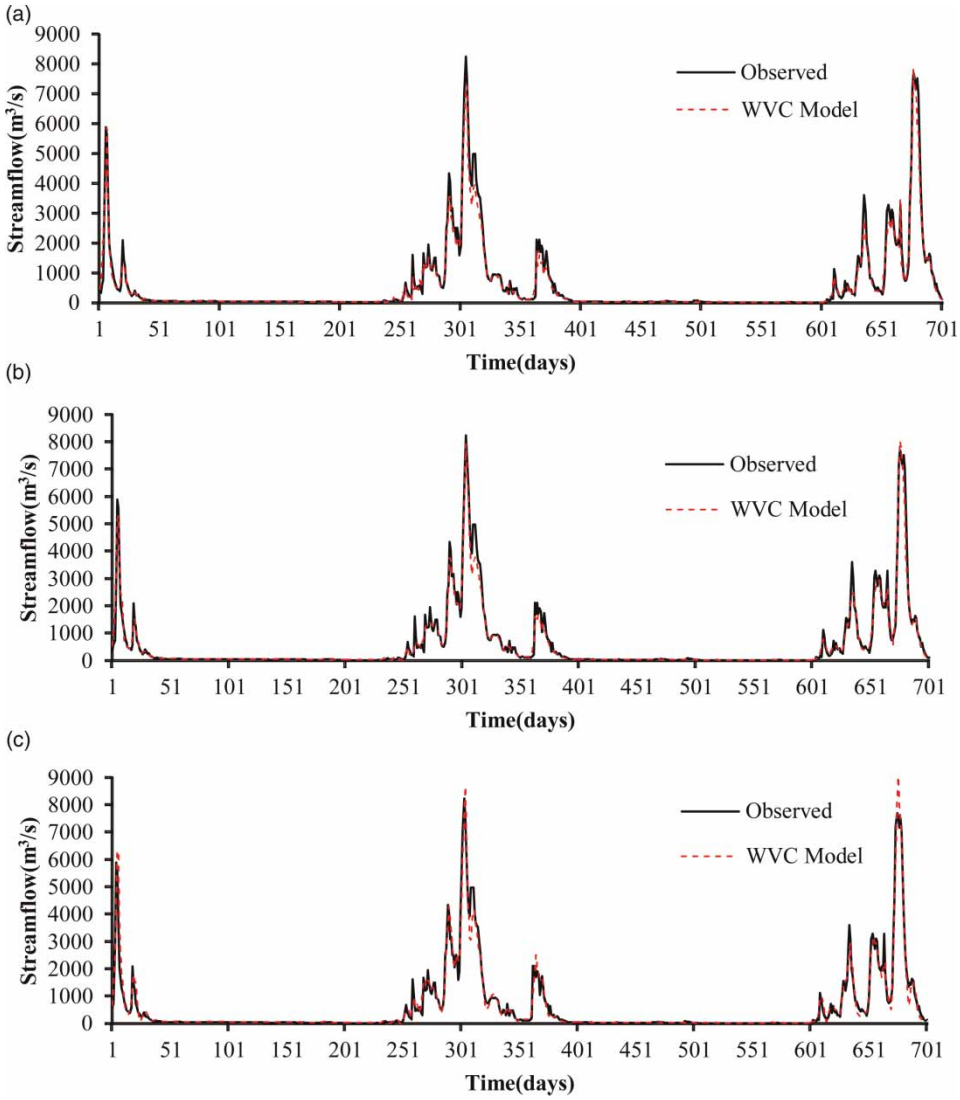

Figure 10 | Comparison of forecasted versus observed flow using the WVC model (Scheme I) for: (a) 1-day-ahead; (b) 2-days-ahead; and (c) 3-days-ahead forecasting for the Krishna River at Krishna Agraharam station (validation period).

1–20 input neurons, one single hidden layer comprising six neurons and one output layer consisting of one neuron (representing the predicted mean daily flow) were developed. A total of six neurons in the hidden layer was seen to be the optimum number for each ANN model, determined via trial and error.

Wavelet LR

The wavelet transforms of observed runoff at Yadgir and Huvinhedigi along with observed rainfall over the intermediate catchment are linearly regressed to forecast future flows at Krishna Agraharam. The forecasting equation using the WLR model formulation under this scheme is given as

$$\hat{Q}_{ka,t} = 2.55D1Q^{h}_{t-1} - 1.42C2Q^{h}_{t-1} - 2.27D1Q^{y}_{t-1} - 1.08C2Q^{h}_{t-1} - 0.02C2Q^{h}_{t-2} + 1.29C2R_{t-1} \quad (13)$$

where $\hat{Q}_{ka,t}$ denotes the one-step-ahead flow forecast at Krishna Agraharam; Di is the ith detail component; and Ci is the ith level approximation component.

ARMAX

Equation (14) describes the auto-regressive moving average with exogenous variables (ARMAX) representation for the inflow forecasting model where the terms q_{ka}, q_h and q_y represent the observed flows at Krishna Agraharam, Yadgir and Huvinahedigi, respectively, r_{ka} denotes the average areal rainfall over the intervening catchment between the upstream and downstream stations and v_t denotes the observation error:

$$Q_{ka,t} = a_{l,t}Q_{ka,t-l} + \ldots + a_{k,t}Q_{y,t-k} + \ldots + a_{j,t}Q_{h,t-j} + \ldots + a_{i,t}R_{ka,t-i} + v_t \quad (14)$$

where $i, j, k, l = 1, \ldots m$.

RESULTS AND DISCUSSION

The forecast results of the models are compiled in Tables 5 and 6. The model results are compared in terms of the Nash–Sutcliffe criteria (NSC, Nash & Sutcliffe 1970) and the root-mean-square error (RMSE) values. Table 5 shows the results for the direct multi-step-ahead models in which a separate model is developed for each lead time, whereas Table 6 shows the results for the recursive multi-step models.

For 1-day-ahead forecasting, the analysis of the results reveals that the WVC model demonstrated superior performance in comparison to the other models, yielding a NSC = 0.9775. The WA-ANN and ANN model performed similarly, resulting in NSC = 0.971 and 0.965, respectively. WLR models perform better than the auto-regressive with exogenous inputs (ARX) models in terms of NSC and

Table 5 | Result statistics for various lead times using direct multi-step models

Lead time (days)		1	2	3	4	5	6
NSC	ANN	0.965	0.91	0.845	0.65	0.60	0.526
	WLR	0.96	0.86	0.83	0.58	0.56	0.50
	ARMAX	0.94	0.83	0.74	0.50	0.46	0.31
	WA-ANN	0.971	0.96	0.94	0.85	0.745	0.66
	WVC	0.9775	0.97	0.954	0.8713	0.75	0.68
RMSE ($m^3 s^{-1}$)	ANN	200.45	400.8	523.24	700.24	740.26	830.23
	WLR	210.7	460.20	513.20	754.98	810.56	872.76
	ARMAX	260.63	520.42	630.20	919.29	966.20	1,023.5
	WA-ANN	193.56	224.45	280.09	460.14	650.12	721.56
	WVC	182.67	217.56	270.45	440.78	630.45	700.98

Table 6 | Result statistics for various lead times using recursive multi-step models

Lead time (days)		1	2	3	4	5	6
NSC	ANN	0.971	0.93	0.87	0.65	0.59	0.53
	WLR	0.96	0.87	0.84	0.58	0.56	0.50
	ARMAX	0.94	0.83	0.74	0.50	0.46	0.31
	WA-ANN	0.971	0.963	0.948	0.86	0.76	0.67
	WVC	0.9823	0.976	0.96	0.890	0.78	0.70
RMSE ($m^3 s^{-1}$)	ANN	200.45	360.81	439.22	700.20	743.60	828.29
	WLR	210.7	440.8	511.56	753.98	800.56	870.76
	ARMAX	260.63	520.42	630.20	916.29	969.20	1,030.25
	WA-ANN	193.56	214.87	270.65	440.91	630.19	700.08
	WVC	182.67	206.76	250.32	400.12	609.97	679.9

RMSE values. The WLR models produced results with a NSC of 0.96 when compared with 0.94 yielded by the ARX models.

In the case of multi-step forecasting, the recursive approach performs better than the direct approach for all the models up to a lead time of 3 days. Beyond that, the performance of both the approaches seems to be similar. Among the different models, the ANN models perform better than the WLR and ARX models. For example, for lead time of 3 days, the RMSE of the forecasts using ANN model was found to be 439.22; WLR and ARX models yielded results with RMSE equal to 511.56 and 580.78 $m^3 s^{-1}$, respectively.

Importantly, however, the results show that the proposed WVC models yield better results than all other models under investigation. The WVC model produced forecasts with NSC value = 0.89 for 4 days lead time, whereas the WA-ANN model forecasts had only a value of 0.86.

Analysis of the direct multi-step model results reveals that the wavelet-based non-linear models perform better than the regular ANN and wavelet linear models. Further, among the wavelet-based non-linear models, the WVC model performs slightly better than the WA-ANN models for all lead times. For example, the WVC model forecast result had a NSC value = 0.871 for lead time of 4 days,

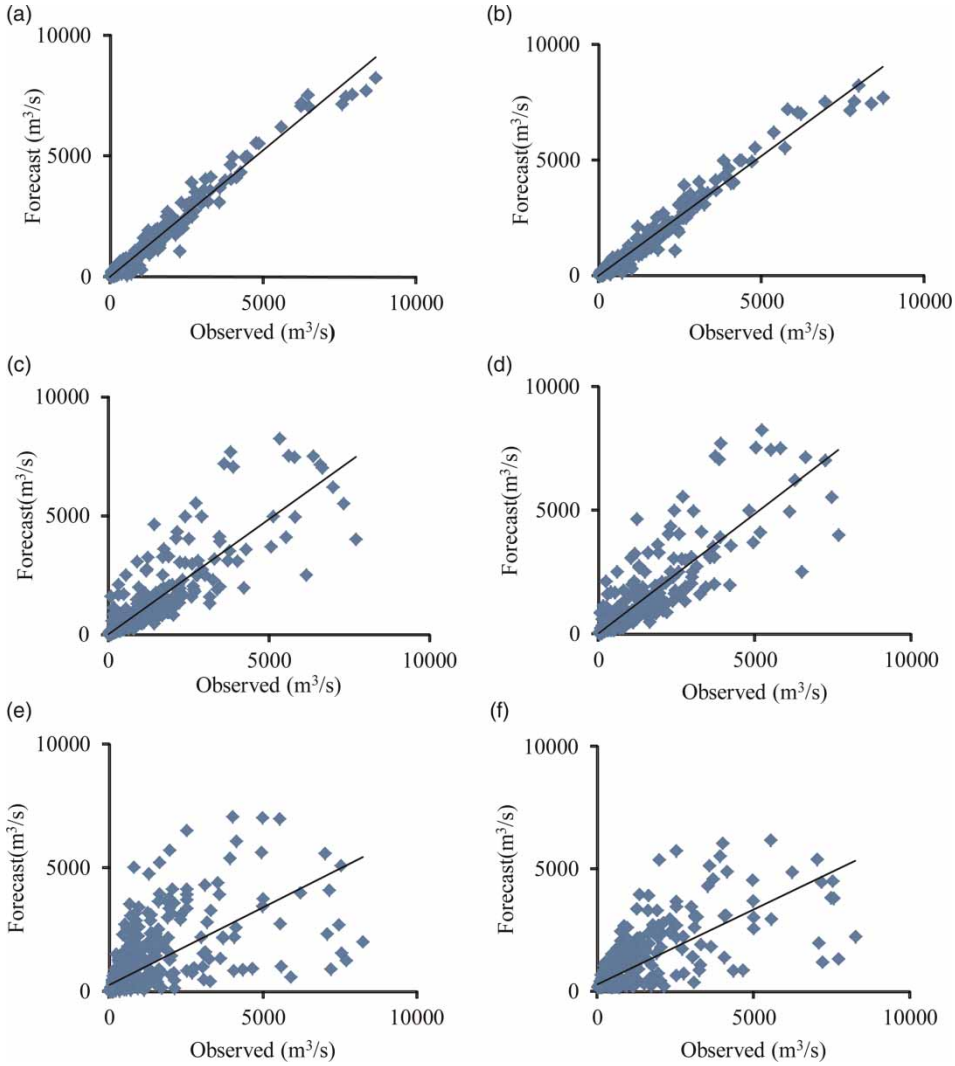

Figure 11 | Scatter plot of the results for (a, c, e) ANN and (b, d, f) WLR models at lead times of 1, 2 and 4 days (from top). The solid line represents the $x = y$ line.

whereas the WA-ANN model yielded results with NSC value = 0.85.

Overall, it can be seen that the coupled wavelet-Volterra model provided more accurate forecasts than the ARMAX- and ANN-based models. It can also be seen that the WVC model performs better than the WA-ANN model in all cases. In a comparison of the direct and recursive multi-step forecasting methods, it can be seen that the recursive multi-step method performs better than the direct method.

Figure 11 is the plot of results of the direct multi-step ANN and WLR models in the form of a scatter plot for lead time of 1, 2 and 4 days. It can be seen that the results are comparable for 1-day-ahead forecast; for higher lead

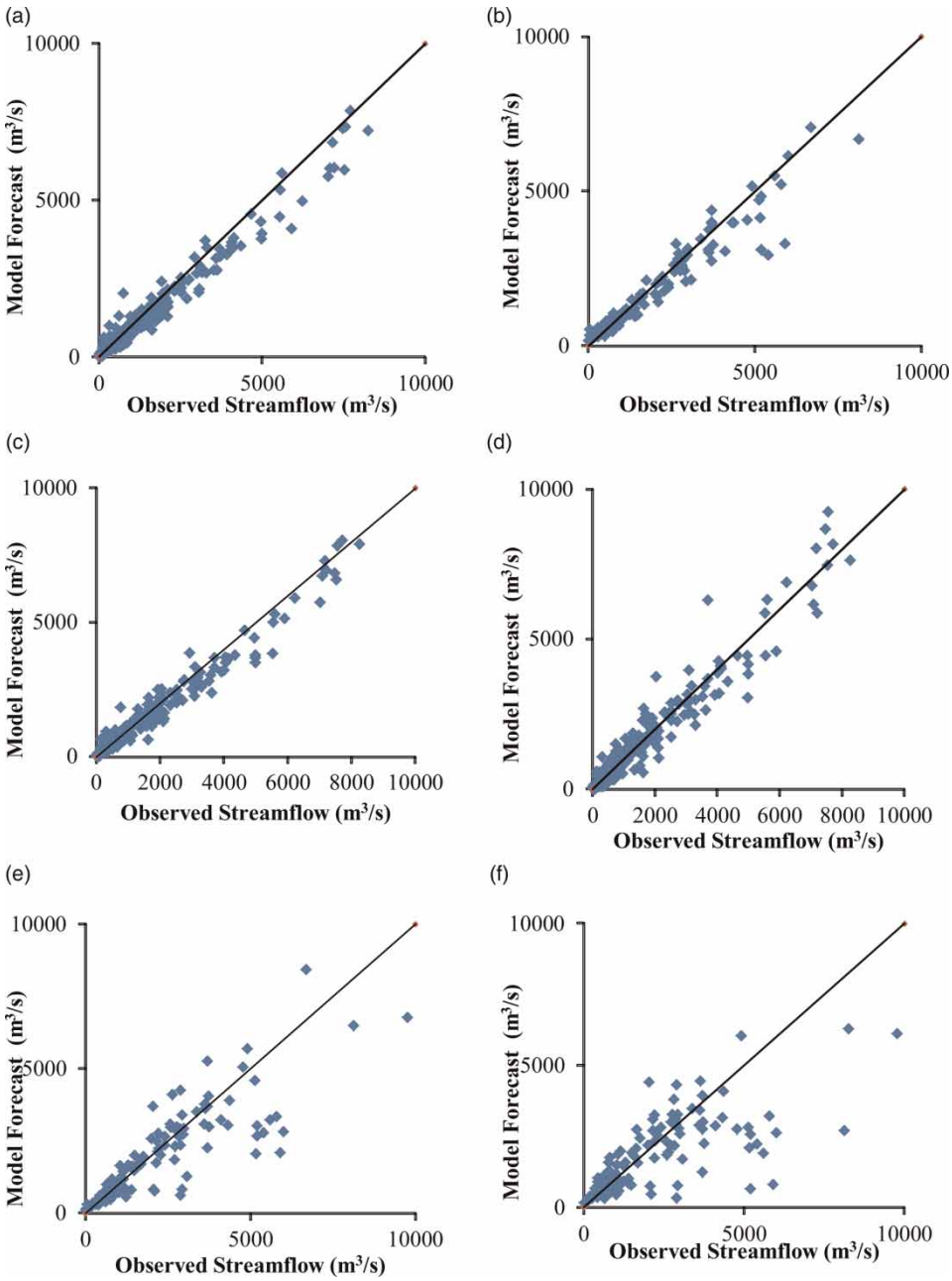

Figure 12 | Scatter plot of the results for (a, c, e) WA-ANN and (b, d, f) WVC models at lead times of 1, 2 and 4 days (from top). The solid line represents the $x = y$ line.

times, however, the WLR model performs poorly when compared with the ANN models due to the higher degree of non-linearity.

A similar comparison between the WVC and WA-ANN models is shown in Figure 12. The investigation of the scatter plot reveals that the WVC model performs better in picking up the peak flow in comparison to the WA-ANN models and, for a lead time of 4 days in particular, it is observed that the WA-ANN model underestimates the flow values in comparison with the WVC models.

A comparison of results shows that the WVC and WA-ANN models demonstrate comparable performances in most test situations explored in this study but, overall, the proposed WVC model is at least as good as the WA-ANN model and, in some tests, actually yielded better results than the latter approach.

With regards to the other models tested in the study, results show that ANN models performed better than the WLR and ARMAX models for all lead times; this may be attributed to the relatively superior capability of ANN models to capture non-linear features of the given process, whose presence are revealed by the BDS test results. However, it is also interesting to note that the WLR model, although a linear model, performed superior to the ARMAX and very close to the ANN model. A reason for the better performance of the wavelet-based model may be the separation of the noise from other, but relatively well behaved, components prior to modelling. Especially in the case of streamflow measurements on a daily scale, noise component of the overall observed process is relatively higher and strong enough to mask the more important underlying dynamics.

The other advantages offered by the proposed WVC model may be summarized as follows:

1. WVC models are simple and versatile;
2. WVC models can be implemented in an adaptive mode whereas WA-ANN models are inherently complex and opaque to scrutiny; and
3. WVC models yield an analytic form of the forecasting model, leading to a better insight into the underlying generating process.

CONCLUSIONS

The potential of WVC models for 1–5-days-ahead flow forecasting was investigated in this study for a site on the River Krishna in India. The WVC models were developed by combining wavelet transforms with the second-order Volterra kernel-based framework.

The WVC models were compared with some of the baseline models for 1–5-days-ahead flow forecasting. The results showed that, for all lead times, the WVC models provided more accurate results than the regular ANN and WLR models. This may be attributed to the ability of the former approach to provide a better scale-specific description of the original time series.

Also, the results obtained using the WA-ANN models were comparable to WVC model results for some test cases. Additionally, it is seen that the WVC models involve less computational effort when compared with the WA-ANN models; the proposed WVC model is therefore a good alternative to the WA-ANN models for flow forecasting.

In a comparison between the direct and recursive multi-step forecasting methods, the latter performs better than the former. In further research, the WVC model could be tested for accuracy in forecasting other hydrological time series such as sediment delivery rates in natural streams or water quality parameters.

REFERENCES

Abrahart, R. J. & See, L. 2000 Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. *Hydrological Processes* **14** (11–12), 2157–2172.

Adamowski, J. & Karapataki, C. 2010 Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: evaluation of different ANN learning algorithms. *Journal of Hydrologic Engineering* **15**, 729.

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** (1–2), 85–91.

Anctil, F. & Tape, D. G. 2004 An exploration of artificial neural network rainfall-runoff forecasting combined with wavelet decomposition. *Journal of Environmental Engineering and Science* **3**, S121–S128.

Audit, B., Bacry, E., Muzy, J.-F. & Arneodo, A. 2002 Wavelet-based estimators of scaling behaviour. *IEEE Transactions on Information Theory* **48** (11), 2938–2954.

Beran, J. 1994 *Statistics for Long-memory Processes*. Chapman & Hall/CRC, New York.

Boucher, M. A., Perreault, L. & Anctil, F. 2009 Tools for the assessment of hydrological ensemble forecasts obtained by neural networks. *Journal of Hydroinformatics* **11** (3–4), 297–307.

Brock, W. A., Dechert, W. & Scheinkman, J. 1987 A test for independence based on the correlation dimension. Working paper, University of Wisconsin at Madison, University of Houston, and University of Chicago.

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

Cannas, B., Fanni, A., See, L. & Sias, G. 2006 Data processing for river flow forecasting using neural networks: wavelet transforms and data partitioning. *Physics and Chemistry of the Earth* **31**, 1164–1171.

Chen, S., Billings, S. A. & Luo, W. 1989 Orthogonal least squares methods and their application to non-linear system identification. *International Journal of Control* **50** (5), 1873–1896.

Coulibaly, P. & Baldwin, C. K. 2005 Nonstationary hydrological time series forecasting using nonlinear dynamic methods. *Journal of Hydrology* **307** (1–4), 164–174.

Coulibaly, P., Anctil, F. & Bobee, B. 2000 Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *Journal of Hydrology* **230** (3–4), 244–257.

Fernando, A. & Shamseldin, A. Y. 2009 Investigation of the internal functioning of the radial basis function neural network river flow forecasting models. *Journal of Hydrologic Engineering* **14** (3), 1–7.

Kantelhardt, J. W., Rybski, D., Zschiegner, S. A., Braun, P., Koscielny-Bunde, E., Livina, V., Havlin, S. & Bunde, A. 2003 Multifractality of river runoff and precipitation: Comparison of fluctuation analysis and wavelet methods. *Physica A* **330**, 240–245.

Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A. & Stanley, H. E. 2002 Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A* **316**, 87–114.

Kim, H. S., Kang, D. S. & Kim, J. H. 2003 The BDS statistic and residual test. *Stochastic Environmental Research and Risk Assessment* **17**, 104–115.

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

Kisi, O. 2009 Neural networks and wavelet conjunction model for intermittent streamflow forecasting. *Journal of Hydrologic Engineering* **14**, 773–782.

Kisi, O. 2011 Wavelet regression model as an alternative to neural networks for river stage forecasting. *Water Resources Management* **25**, 579–600.

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

Lawrance, A. & Kottegoda, N. 1977 Stochastic modelling of riverflow time series. *Journal of the Royal Statistical Society, Series A* (*General*) **140**, 1–47.

Maheswaran, R. & Khosa, R. 2012a Multiscale nonlinear model for monthly stream flow forecasting: a wavelet based approach. *Journal of Hydroinformatics* **14** (2), 424–442.

Maheswaran, R. & Khosa, R. 2012b Comparative study of different wavelets for hydrologic forecasting. *Computers and Geosciences* **46**, 284–295.

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

Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models, Part I - A discussion of principles. *Journal of Hydrology* **10**, 282–290.

Nayak, P., Sudheer, K., Rangan, D. & Ramasastri, K. 2004 A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology* **291** (1–2), 52–66.

Nourani, V., Alami, M. & Aminfar, M. 2009 A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. *Engineering Applications of Artificial Intelligence* **22**, 466–472.

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

Patterson, D. M. & Ashley, R. 2000 *A Nonlinear Time Series Workshop: A Toolkit for Detecting and Identifying Nonlinear Serial Dependence*. Kluwer Academic Publishers, Boston.

Pramanik, N. & Panda, R. K. 2009 Application of neural network and adaptive neuro-fuzzy inference systems for river flow prediction. *Hydrological Science Journal* **54** (2), 247–260.

Rajurkar, M. P., Kothyari, U. C. & Chaube, U. C. 2004 Modeling of daily rainfall-runoff relationship with artificial neural network. *Journal of Hydrology* **285**, 96–113.

Rao, A. R. & Bhattacharya, D. 1999 Hypothesis testing for long-term memory in hydrologic series. *Journal of Hydrology* **216** (3–4), 183–196.

Rumelhart, D. E., Widrow, B. & Lehr, M. A. 1994 The basic ideas in neural networks. *Communications of the ACM* **37** (3), 87–92.

Saco, P. & Kumar, P. 2000 Coherent modes in multiscale variability of streamflow over the United States. *Water Resources Research* **36** (4), 1049–1067.

Shensa, M. J. 1992 The discrete wavelet transform: wedding the à trous and Mallat algorithms. *IEEE Transactions on Signal Processing* **40** (10), 2464–2482.

Smith, L. C., Turcotte, D. L. & Isacks, B. L. 1998 Stream flow characterization and feature detection using a discrete wavelet transform. *Hydrological Processes* **12** (2), 233–249.

Tiwari, M. K. & Chatterjee, C. 2010 Development of an accurate and reliable hourly flood forecasting model using wavelet

bootstrapped ANN (WBANN) hybrid approach. *Journal of Hydrology* **394** (3), 458–470.

Todini, E. 1988 Rainfall-runoff modelling: past, present, and future. *Journal of Hydrology* **100**, 341–352.

Tokar, A. & Johnson, P. 1999 Rainfall-runoff modeling using artificial neural networks. *Journal of Hydrologic Engineering* **4** (3), 232–239.

Torrence, C. & Compo, G. P. 1998 A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society* **79** (1), 61–78.

Wang, W., Vrijling, J., Van Gelder, P. & Ma, J. 2006 Testing for nonlinearity of streamflow processes at different timescales. *Journal of Hydrology* **322** (1–4), 247–268.

Wei, H., Billings, S. & Liu, J. 2004 Term and variable selection for non-linear system identification. *International Journal of Control* **77** (1), 86–110.

Wu, C., Chau, K. & Li, Y. 2009 Methods to improve neural network performance in daily flows prediction. *Journal of Hydrology* **372**, 80–93.

Zhang, Q. & Benveniste, A. 1992 Wavelet networks. *IEEE Transactions on Neural Networks* **3**, 889–898.

Zoppou, C., Nielsen, O. M. & Zhang, L. 2002 Regionalization of daily stream flow in Australia using wavelets and k-means. Available from: http://wwwmaths.anu.edu.au/research.reports/mrr/mrr02.003/abs.html.

First received 14 August 2012; accepted in revised form 14 November 2012. Available online 10 January 2013