Journal of Hydrology 507 (2013) 186–200

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Multiscale streamflow forecasting using a new Bayesian Model Average based ensemble multi-wavelet Volterra nonlinear method

Maheswaran Rathinasamy [a], Jan Adamowski [b,*], Rakesh Khosa [a]

[a] Department of Civil Engineering, Indian Institute of Technology, New Delhi, India

[b] Department of Bioresource Engineering, McGill University, 21111 Lakeshore, Ste Anne de Bellevue, Quebec H9X3V9, Canada

ARTICLE INFO

Article history:
Received 10 May 2013
Received in revised form 12 September 2013
Accepted 19 September 2013
Available online 5 October 2013
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Sheng Yue, Associate Editor

Keywords:
Multiscale streamflow forecasting
Wavelet based nonlinear models
Ensemble forecasting
Bayesian Model Averaging

SUMMARY

Over the last five years, wavelet transform based models have begun to be explored for hydrologic forecasting applications. In general, a particular wavelet transform (and a particular set of levels of decomposition) is selected as the 'optimal' wavelet transform to be used for forecasting purposes. However, different wavelets have different strengths in capturing the different characteristics of particular hydrological processes. Therefore, relying on a single model based on a single wavelet often leads to predictions that capture some phenomena at the expenses of others. Ensemble approaches based on the use of multiple different wavelets, in conjunction with a multi model setup, could potentially improve model performances and also allow for uncertainty estimation. In this study, a new multi-wavelet based ensemble method was developed for the wavelet Volterra coupled model. Different wavelets, levels of decomposition, and model setups are used in this new approach to generate an ensemble of forecasts. These ensembles are combined using Bayesian Model Averaging (BMA) to develop more skilful and reliable forecasts. The new BMA based ensemble multi-wavelet Volterra approach was applied for forecasting stream flow at different scales (daily, weekly and monthly) observed at two stations in the USA. The results of this study reveal that the proposed BMA based ensemble multi-wavelet Volterra nonlinear model outperforms the single best wavelet Volterra model, as well as the mean averaged ensemble wavelet Volterra model.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Wavelet transforms (WT) have begun to be used in several hydrological applications, such as trend analysis (e.g., Kallache et al., 2005; Partal and Küçük, 2006; Adamowski et al., 2009; Nalley et al, 2012; Sang et al., 2012) forecasting (e.g, Wang and Ding, 2003; Nourani et al., 2008; Kisi, 2009; Partal, 2009; Tiwari and Chatterjee, 2010; Adamowski and Sun, 2010; Adamowski and Fung Chan, 2012), time frequency characterisation (e.g., Labat, 2005; Neupauer et al., 2006; and Chen and Xie, 2007), multiscale correlation (e.g., Labat, 2005; and Sang et al., 2010c), and minimizing model parameters (Chou, 2007). In particular, the WT has been shown to be a very useful tool for hydrologic forecasting purposes. Recent publications (e.g., Adamowski and Karapataki, 2010; Kisi, 2011; and Maheswaran and Khosa, 2011c) have shown that the coupling of traditional data driven (or machine learning) models with the wavelet transform significantly increase the accuracy of these models. This has led to the development of coupling nonlinear models (e.g., neural networks, Support Vector Machines), as well as linear regression techniques (e.g., multiple linear regression), with the wavelet coefficients resulting from filter bank decomposition. The majority of studies have found that the wavelet neural network approach provides the most accurate hydrological forecasts. More recently, Maheswaran and Khosa (2011c) developed a new wavelet Volterra coupled model that overcomes the limitations of the wavelet neural network model approach and provides more accurate results than the wavelet neural network approach.

Although significant progress has been made in the use of WTs for hydrologic and other time series forecasting, the choice of appropriate settings for the filter bank is still an open problem. Such settings involve the selection of a wavelet family, such as Daubechies (db), Symlet (sym) or Spline, an order (i.e. wavelet length) within that family, and the number of decomposition levels to be employed. Some authors have opted to test several wavelets, and then select the most appropriate wavelet on the basis of the performance of the resulting model. Maheswaran and Khosa (2011b) tested different wavelets, order of wavelets and levels of decomposition for various types of time series having diverse features. They selected the best wavelet family for a given time series having specific characteristics. For example, it was found that the Haar wavelet can be recommended for time series having transient features, whereas wavelets such as the db2 and Spline b3 can be

* Corresponding author. Tel./fax: +1 514 398 7786.
E-mail address: jan.adamowski@mcgill.ca (J. Adamowski).

0022-1694/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhydrol.2013.09.025

recommended for time steps (e.g. monthly) that have long term features. There are other studies where only a single wavelet was used, such as db1, sym8 or db4 (e.g., Kisi (2009)).

Very little research has been published to date that is aimed at establishing general guidelines regarding the choice of wavelet family, in addition to the order of the wavelet for use in wavelet based forecasting models. The general approach that has been adopted to date is to test for different wavelet families and select the best model based on calibration and validation results. However, the intrinsic problem with this approach is that the selected wavelet may work well in capturing certain features of the input time series, but this may not be the case for other features of the input series. For example, a particular wavelet may capture the high flows of a time series very well, but may not perform well in capturing the low flows. Similarly, another wavelet may perform well in capturing the low flow dynamics, but may not be in a position to capture the high flows. This issue has not been investigated in the literature despite its importance, and as such was investigated in this study.

In addition, it is worth noting that the selection of a suitable number of resolution/decomposition levels has also been an issue that has not been resolved, as reported by Maheswaran and Khosa (2011c). In general, the depth or level of decomposition depends on the features of the time series and the nature of the wavelet used. Therefore, it is recommended that instead of using a single best model, one should use an ensemble of forecasts obtained using a group of models developed using different wavelets and different wavelet decompositions. This has also not been explored to date in the literature, and as such was investigated in this study.

Given the necessity of addressing the above issues, this paper proposes a new approach (i.e., the Bayesian Model Averaging based ensemble multi-wavelet Volterra method) that consists of combining ensembles obtained by using different model settings. The underlying idea is that each of these ensembles may account for a particular set of features in the wavelet domain, and may also be able to capture different features (peaks, low flows, time to peak) at the same time. Therefore, by combining the individual forecasts into an ensemble, a richer exploration of the wavelet domain may be achieved, in addition to increasing model performance.

The concept of combining different model setups into an ensemble has been explored in contexts such as daily stream flow forecasting, prediction of rainfall, and hydrological modelling using physically based models (Shamseldin and O Connor, 1999; Georgakakos et al., 2004; Duan et al., 2007; Dhanya and Kumar, 2011; Fundel and Zappa, 2011; Boucher et al., 2011; Zalachori et al., 2012). The basic idea consists of using a set of individual model forecasts instead of the forecasts derived from a single model selected according to a given criterion. These multi model techniques provide a forecast by linearly combining individual model predictions according to different weighting methods. The general approach is to combine ensembles of forecasts that consist of model mean averaging which results in a point forecast. Techniques such as equal weight, Granger-Ramanathan averaging, and Bates–Granger averaging linearly combine the deterministic model outputs into another single point deterministic forecast (Parrish et al., 2012). Shamseldin and O'Connor (1999) explored the use of ANNs for estimating the weights of the different models. However, Hoeting et al. (1999), Raftery et al. (2005), and Rings et al. (2012)determined that these weights do not necessarily reflect the strength of a particular model's performance, and that these techniques are affected by the presence of outliers. Despite this, the multi model ensemble forecasts from these methods were still better than the single best model forecasts. Recently, Bayesian Model Averaging (BMA) has been found to be useful in diverse areas such as statistics, management, meteorology and hydrology. Bayesian Model Averaging (BMA) is a technique that weighs a model by its performance and likelihood of predicting the observation, resulting in a probabilistic forecast. BMA combines information from a class of models to obtain a probability distribution for a quantity of interest, and accounts for model uncertainty.

Raftery et al. (2005) used the BMA method on a set of meteorological models, while Vrugt and Robinson (2007), Duan et al. (2007) used the BMA technique with hydrologic models. For example, Duan et al. (2007) used BMA to combine different physically based models for hydrologic modelling at a daily scale. The results showed that the BMA based models performed better than the single best model results. More recently, Parrish et al. (2012) used a modified BMA approach that incorporates a sliding window of individual model performances around the forecast. Rings et al. (2012) combined the BMA method with Particle filter and Gaussian mixture modelling with a flexible representation of the conditional probability distribution function. And finally, Zhang and Zhao (2012) improved BMA based Bayesian Neural Networks by combining it with Genetic Algorithms. The above studies demonstrated that the BMA approach is a promising method to combine different models to provide an accurate forecast using an ensemble of forecasts. To date, no research has been published in the literature that explores the use of ensemble coupled wavelet models for short term and long term stream flow forecasting using the Bayesian algorithm. In addition, there are very few papers, if any, that also explore: (i) the selection of the wavelet family; (ii) the order (or length) within the wavelet family; and (iii) the number of wavelet decomposition levels that are appropriate for streamflow forecasting applications. All of these issues are explored in this study for the first time.

The new BMA based ensemble multi-wavelet Volterra method that is proposed in this paper for the first time was tested for stream flow forecasting at monthly, weekly and daily scales. Since the aim of this study was not a comparison of different forecasting methods, but rather to explore, in detail, the above mentioned issues, only the wavelet-Volterra coupled (WVC) method was used as the base model in this study. This particular method was selected since it has already been shown by the first author, in Maheswaran and Khosa (2011c), that the WVC outperforms other new methods that have recently been proposed in the literature for streamflow forecasting (i.e., ANN, wavelet ANN, etc.). The WVC model is simple and versatile and can be implemented in an adaptive mode whereas, in contrast, WANN models are inherently complex and opaque to scrutiny. Further, the proposed WVC based approach yields an analytic form of the forecasting model leading to a better insight into the underlying generating process.

The newly proposed BMA based ensemble multi-wavelet Volterra method is based on the WVC method, but is very different in that it introduces a multi-wavelet BMA based ensemble approach that is coupled with the regular WVC method. In this study, the new BMA based ensemble multi-wavelet Volterra method is compared with the results obtained by individual wavelet-Volterra (WVC) based models that do not have the ensemble algorithm (i.e., the single best WVC model), and the mean averaged ensemble WVC model. The paper is organized as follows. Section 2 provides a brief description of wavelet analysis. The method used for forecasting is described in Section 3. Section 4 presents the BMA methodology. Section 5 describes the study area and the model application. Section 6 provides the validation results of the ensemble models using the BMA technique. Section 7 provides a summary and the conclusions.

2. Wavelet analysis

Wavelet analysis is a powerful tool that is used for function analysis similar to Fourier analysis. While Fourier analysis approximates any periodic function as the sum of sines and cosines by addressing frequency alone, wavelet analysis represents any arbitrary function as the sum of wavelets by addressing both space and scale. The wavelet transform decomposes an arbitrary signal into elementary contributions (or wavelets) which are constructed from one single function called the mother wavelet by adjusting its two parameters, dilation and translation.

For practical applications, the Discrete Wavelet Transform (DWT) is usually preferred. The latter transform is an orthogonal function which can be applied to a finite group of data. Functionally, the DWT has similar attributes to the Discrete Fourier Transform, namely: (i) the transforming function is orthogonal, (ii) a signal passed twice through the transformation is unchanged, (iii) the input signal is assumed to be a set of discrete time samples, and (iv) both transforms are convolutions. However, while the basis function of the Fourier transform is a sinusoid, the wavelet basis, in contrast, is a set of functions that are defined by a localized wavelet function.

A typical discrete wavelet function can be represented as (Labat et al., 2000).

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

where $\psi(t)$ is the mother wavelet and j and k are the translation and dilation indices. In the 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 wavelet coefficients is halved with each move to a coarser level and, as a consequence, there is less information available to train the forecasting model at the coarser level that leads to a reduction in overall forecasting accuracy. This problem, caused by decimation, may be overcome by introducing the stationary or, alternatively, *a trous* wavelet transform proposed by Shensa (1992). The basic idea of the *a trous* 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 and, subsequently, from this the derivation of details and the smoothed version becomes possible. For example, consider the original time series $x(t)$ which may also be denoted by c_o, or

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

Further smoother versions of $x(t)$ may be derived using (Renaud et al., 2005)

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

In the preceding Eq. (3), i takes values from 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 (½, ½). Similarly, the filter values for the B_3 Spline wavelet are 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 as

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

The set $\{d_1, d_2, \ldots d_p, c_p\}$ represents the additive wavelet decompositions of data up to the resolution level of p, and the term c_p in this set denotes the residual component, which is also called the approximation. Accordingly, for reconstruction, the inverse transform is given by (Renaud et al., 2005)

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

Here, unlike the classical DWT, the decimation is avoided, resulting in components at different scales that are of the same length. It should be noted that in this study, the boundary correction was done using the method proposed by Maheswaran and Khosa (2011c).

3. Model used for forecasting – Wavelet Volterra Coupled (WVC) Model

As mentioned earlier, all the models used in this study (i.e., single WVC models, mean average ensemble WVC models, and the newly proposed BMA based ensemble WVC models) are based on the wavelet-Volterra coupled model (WVC) developed by Maheswaran and Khosa (2011a,d). The single WVC model was very recently proposed as a new method for hydrological forecasting by the first author of this paper (Maheswaran and Khosa, 2011a, c) for different hydrological forecasting problems such as streamflow forecasting and groundwater level forecasting, and the results showed that the single WVC method is more accurate than other newly developed machine learning methods such as wavelet neural networks, neural networks, and wavelet linear regression. As such, since the WVC method has already been shown to be more accurate than other new methods for hydrological forecasting, this paper focused on improving the WVC method, in addition to exploring several additional important issues that have not been addressed in the literature regarding the use of wavelet transform based models in hydrological forecasting applications. Since the aim of this study was not to compare different forecasting methods, a common reference model was considered (the single WVC model), and the performance of the newly proposed Bayesian Model Averaging based ensemble multi-wavelet Volterra method was compared with the base model (i.e., single WVC model), as well as the mean average ensemble WVC model.

Fig. 1 presents a general schematic of the structure of the WVC 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 *á trous* wavelet transform, and 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 nonlinear time-invariant system with memory is based on a simple extension of the Taylor series expansion for nonlinear autonomous causal systems with memory. Studies such as Diskin and Boneh (1972), Amorocho (1973), and Labat and Ababou (2001), have led to the general understanding that a second order representation (or, a second order nonlinear Volterra kernel) is usually sufficient for most hydrologic systems. For additional mathematical details, reference may be made to the accompanying appendix.

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

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

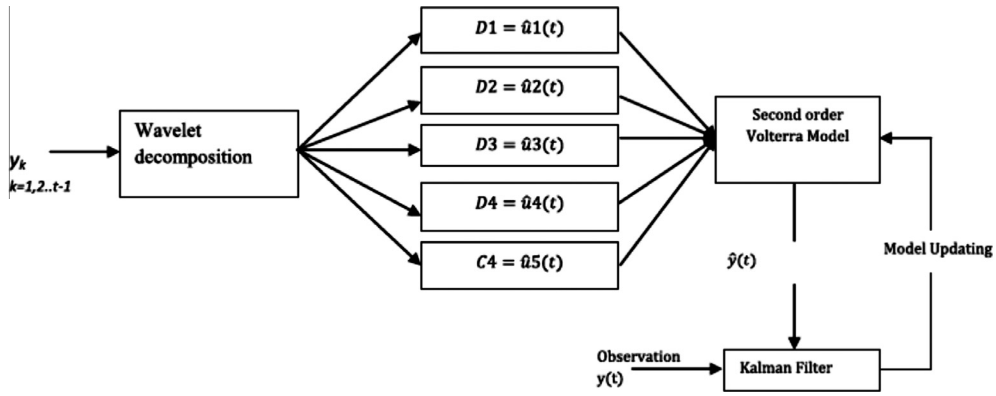

Fig. 1. Wavelet Volterra Coupled (WVC) model.

In Eq. (6), 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 2nd order nonlinear relation between the nth input u_n and y, respectively, and the second order cross-kernels $h_{2x}^{(n_1,n_2)}$ describe the 2nd order nonlinear interactions between each unique pair of inputs (u_{n1} and u_{n2}) as they affect y. Eq. (6) can be simplified by combining the last two terms to yield Eq. (7) (it now remains to estimate kernels h_1 and h_2)

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

The representation of Eq. (7) can be further simplified by considering each of the lagged variables $u_1(t-1), u_1(t-\tau),\ldots, u_2(t-1), u_2(t-\tau),\ldots$ as separate variables $d_1(t), d_2(t), d_3(t),\ldots, d_{Nl}(t)$. Eq. (7) can then be written as (Maheswaran and Khosa 2011c)

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

More clearly

$$d_l(t) = \{x_k(t) 1 \leqslant k \leqslant J+1;\ 1 \leqslant l \leqslant J+1$$

$$d_l(t) = \{x_k(t-\tau), 1 \leqslant k \leqslant J+1;\ J+1 < l \leqslant N_l;\ \tau = 1,2,3\ldots m$$

$\tau = \tau$th lagged value.

$J =$ level of decomposition.

$N_l =$ total number of lagged variables

In the preceding Eq. (8), h_1 and h_2 represent the Volterra kernels to be estimated, and it can be seen that the number of parameters to be estimated increases proportionately as the number of inputs and/or the process memory increases, leading to an increased computational burden and severely compromised estimation. In order to handle these computational difficulties, Chen et al. (1989) and, later, Wei and Billings (2004) proposed the use of the Orthogonal least squares technique as a preferred estimation approach as it is better, in comparison with the Ordinary Least Squares technique, at handling modelling situations where there is a possibility of multi-collinearity amongst vectors that constitute the coefficient matrix.

In addition to the above, as shown in Fig. 1, the proposed formulation is recursively updated in real time using the well known Kalman Filter formulation. For further derivation of the single WVC model formulation (which forms the basis of both the mean average WVC ensemble method, as well as the new BMA based ensemble WVC method, described in subsequent sections), readers are referred to Maheswaran and Khosa (2011c).

4. Ensemble combination

There are different models that researchers can use in data based hydrological forecasting and the choice between them is often not obvious. Thus, the uncertainty over which model to use is an important issue in forecasting. Since model uncertainty critically affects forecasts (and forecast uncertainty), the 'pooling' or combination of forecasts using different models is one approach to address this issue. This pooling can be done in a variety of different ways. The simplest method is the mean averaging of the ensembles (in our study this is the average ensemble WVC model), wherein the weights in combining different forecasts is uniform. Further, advanced techniques can be based on historical performance (Hendry and Clements, 2002). A related methodology for dealing with large numbers of regressors used in macroeconomic forecasting is based on principal components or factors (e.g., Stock and Watson, 2002). Alternatively, the weights used in combining forecasts can be based on posterior model probabilities within a Bayesian framework. This procedure, which is typically referred to as Bayesian Model Averaging (BMA), is the standard approach to model uncertainty within the Bayesian paradigm, where it is natural to reflect uncertainty through probability (Steel, 2011). Thus, it follows from the direct application of Baye's theorem, as is explained in Leamer (1978), Min and Zellner (1993), Raftery et al. (1997), Fernandez et al. (2001) and Ley and Steel (2009), that pooling or combining models using BMA minimizes the expected predictive squared error loss, provided that the set of models under consideration is exhaustive.

This study explored the use of two ensemble methods, namely: the mean average ensemble WVC method, and the newly proposed BMA based ensemble WVC method, and compared these two approaches to the single WVC method.

4.1. Bayesian Model Averaging

Statistical analysis usually assumes that there is a "single best" model, often selected from a class of several possible models. It is typically not the case that the selected model is always the best, and the analysis ignores model uncertainty and assumes that the data obtained from the selected model has no uncertainty. This approach results in overconfident inference and forecasts, especially when substantially different scientific results may be obtained from alternative models (see Hoeting et al., 1999). Bayesian Model

Averaging reduces the potential overconfidence by conditioning not on a single model, but on a class of models. Bayesian Model Averaging (BMA) is a statistical approach for post-processing ensemble forecasts from multiple competing models (Leamer, 1978). It has been widely used in a variety of areas, as mentioned earlier, such as dynamical weather forecasting models (Raftery et al., 2005) and hydrology. For example, in hydrology, Ajami et al. (2007) and Duan et al. (2007) used the BMA approach to provide an ensemble of three different hydrologic models. In their work, they used BMA to combine the model results from three different physically based lumped hydrologic models. They showed that combining the model results via BMA provided better results than the individual models.

The basic principle of the BMA method is to generate an overall forecast probability distribution function (PDF) by taking a weighted average of the individual model forecast PDFs. The weights represent the model performance, or more specifically, the probability that a model will produce the correct forecast. The BMA method assumes that the probability of an observation $y_{obs}(t)$ at time t is given by a weighted sum over a number of probability distributions $g(y_f(k,t))$ from the k different forecasting models

$$P(y_{obs}(t)) = \sum_k w(k) g(y^f(k,t)) \tag{8}$$

The forecast distribution of each individual model is assumed to be a normal distribution with variance $\sigma(k)$

$$g(y^f(k,t)|y_{obs}(t), \sigma(k)) = \frac{1}{\sigma(k)\sqrt{2\pi}} \times \exp\left(-\frac{(y_{obs}(t) - y_f(k,t))^2}{2\sigma(k)^2}\right) \tag{9}$$

The BMA algorithm finds the optimal values for $w(k)$ and $\sigma(k)$, such that the likelihood of the overall PDF (Eq. (9)) is maximal, given a set of historical forecasts and observations. The method does so by optimizing $w(k)$ and $\sigma(k)$ consecutively in an iterative scheme. The first step of the iteration starts with an initial guess for the weights $w(k)$ and $\sigma(k)$ for each of the individual models and estimates the matrix $z(k,t)$ using Eq. (10), which represents the probability that model k gives the best forecast for station s at time t

$$z(k,t) = \frac{g(y_{obs}(t)|y_f(k,t), \sigma(k))}{\sum_k g(y_{obs}(t)|y_f(k,t), \sigma(k))} \tag{10}$$

The second step in the iterative algorithm is to determine the weights $w(k)$ and variances $\sigma(k)$ of each of the model's k, based on the values of $z(k,s,t)$. The weights are estimated using

$$w(k) = \frac{1}{n}\sum_t z(k,t) \tag{11}$$

where n is the number of observations in the training period (t). The variance $\sigma(k)$ is estimated using

$$\sigma^2(k) = \frac{\sum_t z(t)(y_{obs}(t) - y_f(k,t))^2}{\sum_t z(k,t)} \tag{12}$$

The two steps are alternated to convergence; that is, when $w(k)$ and $\sigma(k)$ no longer change after a recalculation of $z(k)$. One can use a convergence criterion or alternatively use a fixed number of iteration cycles that should guarantee convergence. In this study, the Expectation-Maximization (EM) algorithm (Raftery et al., 2005) was used for the purpose of optimization. For further details on the EM algorithm, readers are referred to Raftery et al. (2005). For dynamical model applications, one can use the weights and variances from the previous time step as a starting point for the new iteration.

4.2. Performance evaluation of ensemble forecasts

The performance of forecasting models can be evaluated in terms of goodness of fit after each of the model structures is calibrated. The coefficient of determination (R^2) measures the degree of correlation among the observed and predicted values. R^2 values range from 0 to 1, with 1 indicating a perfect relationship between the data and the line drawn through them, and 0 representing no statistical correlation between the data and a line. The Root Mean Square Error (RMSE) evaluates the variance of errors independently of the sample size. RMSE indicates the discrepancy between the observed and forecasted values. A perfect fit between observed and forecasted values would have a RMSE of 0. The Nash–Sutcliffe model efficiency coefficient (NSC) can also be used to assess the forecasting power of hydrological models. An efficiency of one (NSC = 1) corresponds to a perfect match of forecasted data to the observed data. An efficiency of zero (NSC = 0) indicates that the model predictions are as accurate as the mean of the observed data.

In order to evaluate the quality of ensemble forecasts, the rank histogram (or Talagrand diagram (Talagrand et al., 1999)), can be used. It is a useful tool that allows one to assess the calibration of the predictive distribution. To construct it, the observed time series x_t value is added to the ensemble forecast. That is, if the forecast has n members the new set consists of $n + 1$ values. Then, the rank associated with the observed value is determined. This operation is repeated for all N forecasts and corresponding observations in the archive. The rank histogram is obtained by constructing the resulting N ranks. If the predictive distribution is well calibrated, then the rank histogram should be close to flat (Boucher et al., 2009). An asymmetrical histogram is usually an indication of a bias in the mean of the forecasts. If the rank histogram is symmetric and U shaped, it may indicate that the predictive distribution is under dispersed. If it has an arch form the predictive distribution may be over dispersed.

5. Model application

5.1. Study area

The river flow data used in this study were obtained from the US Geological Survey (USGS). The time series of daily stage data from two stations were used: (i) 12414500 Selway River Nr Lowell ID, USA; and (ii) 13336500 ST Joe River at Calder ID, USA. The above two stations were selected because the flow regimes in both rivers are highly varied, and are therefore useful in testing the capability of the models in a rigorous manner (low flows, high flows). The catchment areas and also the flow statistics are provided in Table 1.

Further information on these two stations can be found on the USGS web server (http://www.usgs.gov). Since the forecasting in this study had three different lead times (daily, weekly and monthly), the observed data (half hourly sampled data) for a period of 45 years was used and the daily, weekly and the monthly mean for these stations were accordingly estimated. 60% of the entire data sets were used for calibration and the remaining data was used for validation purposes.

5.2. Model development – proposed BMA based ensemble multi-wavelet Volterra model

5.2.1. Generation of ensembles

For the purpose of creating an ensemble of forecasts, several WVC models were run by varying the inputs. This was done by

Table 1
Statistics of streamflow data observed at the two stations.

Station no	USGS ID	State	Area (sq km)	Ele (m)	Mean (m^3/s)	SD (m^3/s)	Max. (m^3/s)	Min. (m^3/s)
I	12414500	ID	2668	1283	66.8	76	393.6	6.7
II	13336500	ID	4947	1719	106.2	135.6	690.1	9.2

varying the mother wavelets, length (order) of wavelet within that wavelet family, number of decomposition levels, and the choice of the combination of input wavelet coefficients. The wavelet transform of the flow series were calculated with different filters of the Daubechies (db1,db2, db4,db5, db6,db7), Symlet (sym2, sym4) and Spline (b3-Spline) wavelet families, which are commonly used in wavelet based forecasting (e.g., Nourani et al., 2009; Kisi 2009 and Maheswaran and Khosa 2011c). Overall, 10 filters were employed. In each case, the number of decomposition levels was varied from one to a maximum of J. It is worth noting that the maximum number of decomposition levels depends on the filter length and the time series length. However, the upper limit for the number of decomposition levels was set as 6 for monthly time series, 8 for weekly time series and 10 for daily time series based on previous research (Maheswaran and Khosa, 2011c), where it was observed that increasing the level of decomposition beyond a certain limit deteriorates model performance. The initial set of input wavelet coefficients to be used as the inputs for the models were determined using correlation analysis. The correlation analysis shows the influence of the different decompositions on the original signal. Only wavelet coefficients with correlation coefficients above 0.6 were selected for modelling. The selected inputs were then used to form different combinations of the wavelet coefficients from the different levels of decomposition (note: only within the decomposition from the same wavelet).

The above approach results in a number of input combinations which are then used as inputs for the WVC models used to produce the ensemble of forecasts. In this study, for all the WVC model based approaches (i.e. single, mean average, and BMA WVC models), the stream flow time series is decomposed into wavelet (D_i, $D_{i+1}\ldots D_N$) and scaling (A_N) coefficients 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. Once the WVC models are calibrated for the different input variable combinations, these models are used to produce an ensemble of forecast values.

5.2.2. Single best WVC model

From the set of numerous model runs, the model having the best performance in terms of NSC and Root Mean Square Error (RMSE) was selected and used for forecasting during the validation period.

5.2.3. Ensemble combination by mean averaging

The ensembles of forecasts obtained from the different WVC model setups were combined using mean averaging, and a single point forecast was produced. This was the mean average WVC ensemble model.

5.2.4. BMA WVC approach

From these ensembles of forecasts, only those forecasts that had a certain Nash Sutcliffe Criteria (NSC above 0.75) were selected to form the final set of ensembles. Then the forecast results from these selected models were combined using the BMA technique. The BMA algorithm assigns suitable weights to each of the model setups based on the a priori performance of these models in the calibration or training stage. These weights are then assigned for the corresponding model forecasts during the validation stage to obtain the BMA based forecasts.

Fig. 2 shows the entire schematic of the proposed BMA based ensemble multi-wavelet Volterra method. This approach was adopted for developing models at the different temporal scales (monthly, weekly and daily).

6. Results and discussion

In this section, the results of the different models for forecasting stream flow at monthly, weekly and daily time scales using the single WVC models, the average WVC ensemble models, and the BMA WVC ensemble models are reported.

6.1. Model results at a monthly scale

Varying the wavelets, order of wavelets, decomposition levels, and the input variables resulted in numerous WVC models which were in turn used for producing the ensemble of forecasts. Table 2 provides a detailed report on the sensitivity analysis of the different wavelets, decompositions and input combinations across different scales for the single WVC model.

The analysis of the results from these different WVC model setups provided some interesting findings.

1. It seems that not all of the wavelet filters were able to provide good forecasts. For example, it was observed that higher order wavelets such as db5 and sym4 decreased model performance when compared to models based on lower order wavelets such as Haar, db2 and sym2.
2. It was found that increasing the depth of decomposition (i.e., number of decomposition levels) in some cases (using the 'Daubechies' wavelet and 'Haar' wavelets) increased performance but only to a certain level. For example, in the case of the Haar wavelet the model performance using 4 levels of decomposition was better than that of the model using 3 and 5 levels of decomposition. However, in cases with '*Symlet*' and higher order 'Daubechies' wavelets, the performance of the model decreased significantly when the level of decomposition was increased.
3. It was also observed that the different wavelets captured different features of the time series. Fig. 3 shows the hydrograph of the results from the Haar wavelet based WVC model. It can be seen that the model captured the low flows very well, whereas it could not capture the peak flow. However, the model results from the Symlet and Spline wavelet based WVC model (Figs. 4 and 5, respectively) show that the Symlet wavelet captures the peaks (over estimation), whereas it does not perform well for low flows.

The above analysis shows that the performance of the single WVC model depends on the choice of the wavelet, and also on the level of decomposition. It can also be seen that a 'one best choice' of the model set up is not ideal, and instead a combination of the forecasts from different models would improve the overall

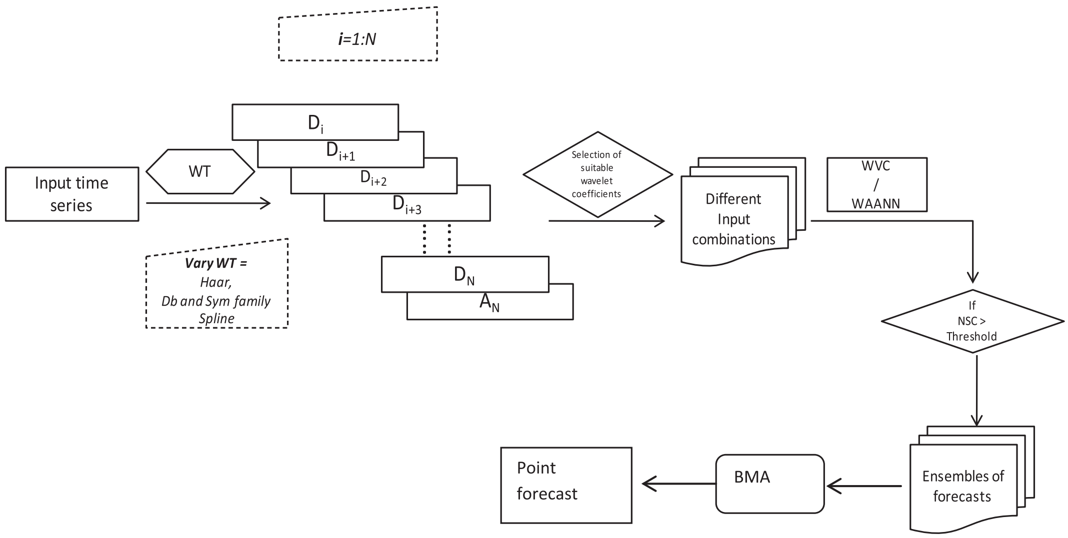

Fig. 2. Proposed BMA WVC algorithm.

Table 2
Summary of model results from various single WVC models for one step ahead forecasting at a monthly scale.

Wavelet used	Level of decomposition	Memory (m) at each level {DW_i's, C}	RMSE (m^3/s)	MAE (m^3/s)	NSC
Station I					
B3-Spline	4	{2,2,2,2,1}	58.82	33.42	0.57
Db4	4	{2,2,1,1,2}	56.6882	32.43	0.73
Db5	3	{1,1,1,2}	56.53	32.85	0.74
Haar	3	{2,2,2,1}	51.30	30.24	0.78
Db2	4	{1,2,2,2,1}	54.93	32.78	0.75
Sym4	5	{1,1,1,2,2,2}	62.27	30.47	0.64
Sym4	6	{1,2,1,1,2,2,2}	84.99	46.5	0.54
Haar	4	{2,2,2,2,1}	57.35	33.43	0.74
Station II					
B3-Spline	3	{2,2,2,1}	37.88	19.88	0.85
Sym4	4	{3,2,2,2,1}	37.01	19.00	0.88
Db2	3	{1,1,1, 2,1}	36.88	18.78	0.89
Haar	3	{1,2,2,2}	37.85	19.7	0.87
Haar	4	{1,1,2,2,2}	36.54	18.35	0.90
Haar	5	{1,1,1,2,2,2}	40.45	21.78	0.54

performance of the model. Therefore, in order to improve the forecasts, instead of taking one best model, the ensembles from different models were combined using: (i) the mean averaging approach and (ii) the newly proposed methodology based on BMA. The BMA technique provides the weights to be assigned to the forecasts from each model setup according to the overall performance of the model during the training stage. For both the stations, a total of 140 model runs were performed, and the best 45 models were selected after applying the threshold of 0.75 (for NSC). The rank histogram for these ensembles is shown in Fig. 6. It can be seen that the histogram for Station I is neither flat nor U shaped, indicating that there is a possible under dispersion of the predictive distributions. However, for Station II, the histogram is relatively flat, indicating that the predictive distribution is properly calibrated.

The weights of the corresponding models obtained via the BMA technique for the two stations are shown in Fig. 7a and b. The

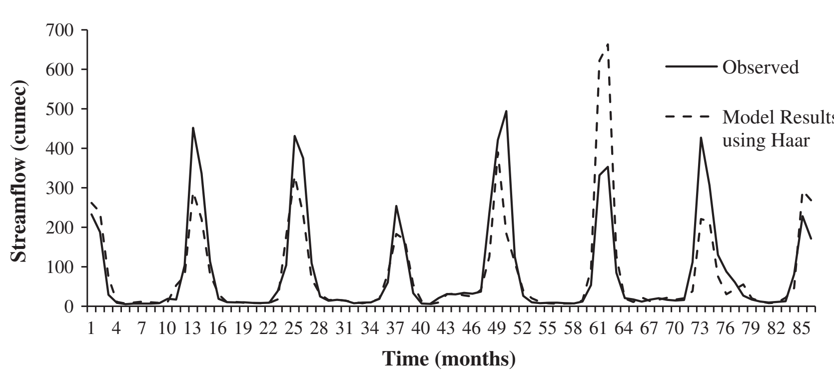

Fig. 3. Model results for single WVC model using Haar wavelet at Station I.

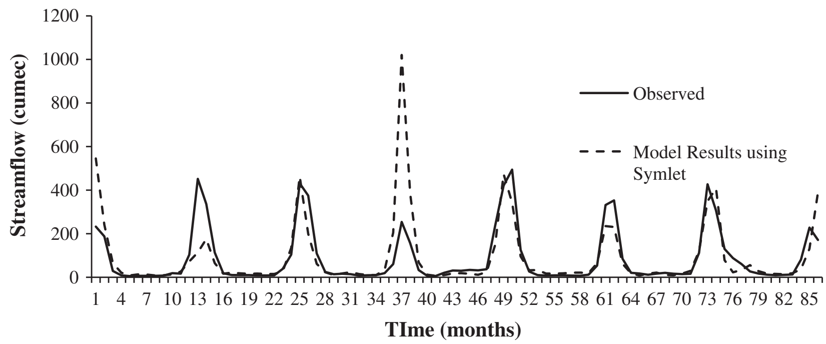

Fig. 4. Model results for single WVC model using Symlet wavelet at Station I.

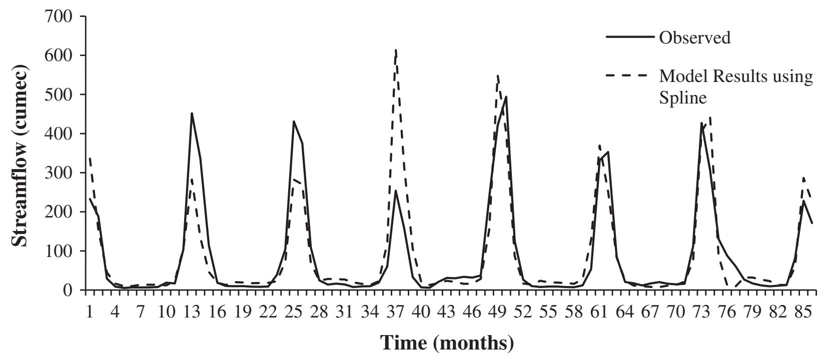

Fig. 5. Model results for single WVC model using Spline wavelet at Station I.

corresponding wavelets and the decomposition levels used in the BMA based ensemble multi-WVC model are also shown in Fig. 7a and b. It can be seen that the Haar wavelet, db2, db4 wavelets and sym4 wavelets contribute significantly to the forecast accuracy. The model results are also plotted in the form of a hydrograph in Figs. 8a and b. It can be seen that the ensemble model based on BMA outperforms the single best WVC model, as well as the average ensemble WVC model. The BMA based ensemble model is able to capture the low and high flows more closely than the other models. The results for Station 2 are also shown in Fig. 8b. Table 3 compares the one month ahead forecast statistics from the best single model, the average ensemble WVC model and the BMA based ensemble WVC model. The results show that using the BMA ensemble method results in higher model accuracy and also higher reliability in the forecast compared to that of the single best model and the mean average ensemble WVC model.

6.2. Model results at a weekly scale

Table 4 shows the results of a few of the single WVC models. Fig. 9a shows the hydrograph of the results from the Haar wavelet (at level 3 decomposition) based single WVC model. It can be seen that the model captured the low flows very well, whereas it could not capture the peak flow. However, the model results from the Symlet and Spline single WVC model (Fig. 9b and c, respectively) show that the Symlet wavelet captures the peaks (over estimation), whereas it does not perform well for low flows. It can be seen that the single WVC models based on Symlet wavelets do not perform well in comparison with the Haar, db2 and db4 based models.

Also, it is to be noted that if the decomposition is performed to a level of 3, the single WVC model is able to pick up the low flows better (see Fig. 9a and d). However, when a 4th or 5th level decomposition is applied, the model is able to capture the underlying seasonality and high flows whereas the low flows are over estimated (see Fig. 9b and c). From these set of models (the best single WVC models), the average WVC ensemble models were formed.

For the BMA WVC models, from the total set of 75 models which were evaluated, those having a NSC $\geqslant 0.75$, (i.e. only 39) were selected for ensemble formation. The rank histogram of these ensembles is shown in Fig. 10. It can be seen that for Station I, the histogram is more or less flat indicating the good calibration of the predictive distribution (Boucher et al., 2009), whereas for Station II, the rank histogram is U-shaped, indicating the under dispersive nature of the predictive distribution. However, Hamil (2001) argues that the U-shaped behaviour may be due to selecting ensemble members with equal prediction probabilities.

The corresponding weights of the BMA WVC models are shown in Fig. 11a (for Station I). Similarly, for Station II the same approach was repeated and the weights of the models were obtained from the BMA algorithm as shown in Fig. 11. The results of the BMA based WVC model, along with the best single WVC model, and the mean average ensemble WVC model, are shown in Table 5. For both the stations, it can be seen that the BMA based WVC model performs better than the best single WVC model, as well as the mean average ensemble WVC model. Figs. 12a and b show the model results of the validation data set for these three models, and it can be observed that the BMA based WVC ensemble model performs better in capturing the peaks as well as the low flows.

6.3. Model results at a daily scale

At the daily scale, a similar approach was used as in the monthly and weekly time scales to develop the single WVC models, the mean average WVC ensemble models, and the BMA WVC

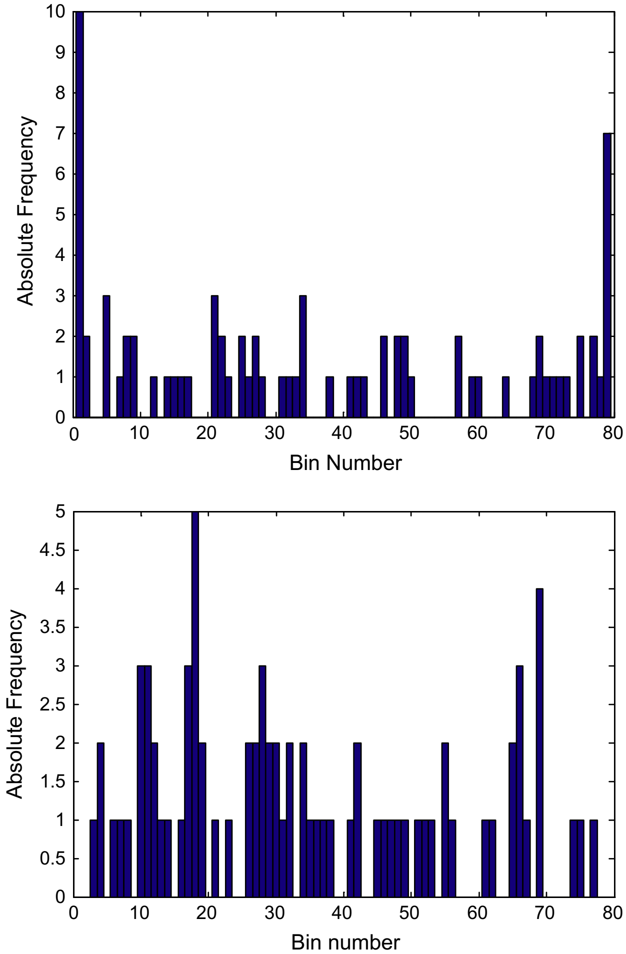

Fig. 6. Rank Histogram of the ensembles obtained from WVC models for (a) Station I and (b) Station II.

ensemble models. Table 6 shows the results of the various single WVC models.

Analysis of these single WVC model results and the corresponding model set ups revealed the following information:

1. It was observed that the lower order wavelet filters such as Haar and db2 provided better forecasts than those obtained using the higher order wavelet filters in the case of daily time scale forecasting. This observation is in congruence with the results of Nourani et al. (2009), who showed that lower order filters performed better than higher order filters and longer wavelets. In our study, it was observed that the db4 and Symlet wavelets performed poorly in comparison with the Haar wavelet. This may be attributed to the fact that the db4, sym4 and sym8 wavelets have broader length (higher order), and are therefore not able to capture the peak flows accurately.
2. With regards to the depth of decomposition (i.e., decomposition level), it was observed that increasing the depth drastically reduced the model performance. It was determined that the maximum depth of decomposition is 2.

A total of 40 different single WVC models were run and 22 models having a threshold of NSC $\geqslant$ 0.90 were selected for combining the forecasts using the BMA WVC model. Figs. 12a and b show the weights that were applied to the model. It can be seen that the Haar wavelet model's contribution is very significant in comparison with the other models. The model results for Station I are in the form of a scatter plot for better clarity (Fig. 13). It can be seen that the BMA algorithm significantly improved WVC model performance when compared to the single WVC model and the mean average WVC ensemble model. Similar analysis was done for Station II and the summary of results obtained is provided for both the stations in Table 7.

It can be seen from the summary of the results that there was a significant improvement in the model performance when the BMA WVC model was used in the case of Station I, compared to using the best single WVC model. However, in the case of Station II, the improvement in model performance was not very significant. This

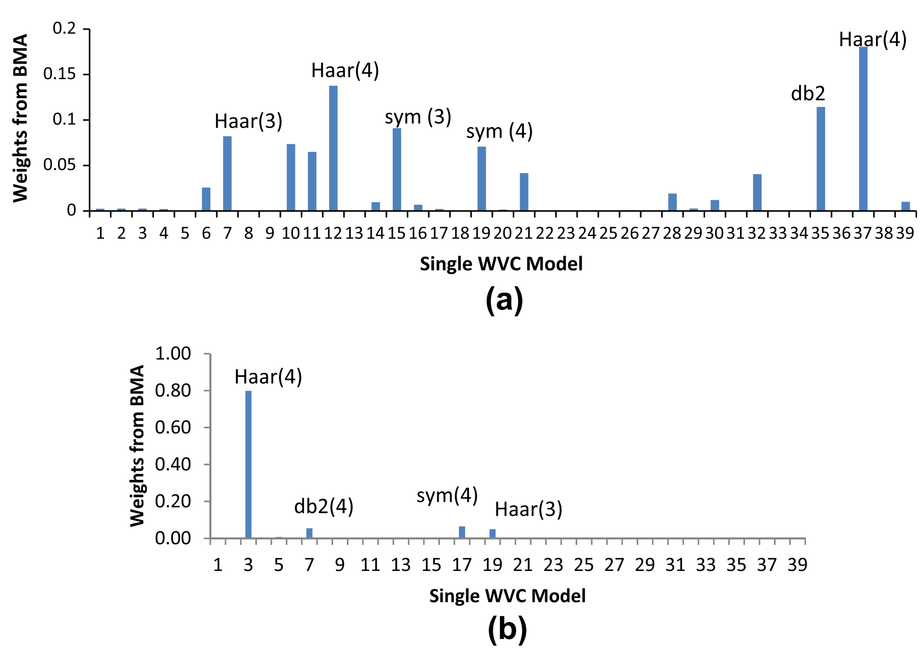

Fig. 7. A single set of BMA weights computed over the entire training period.

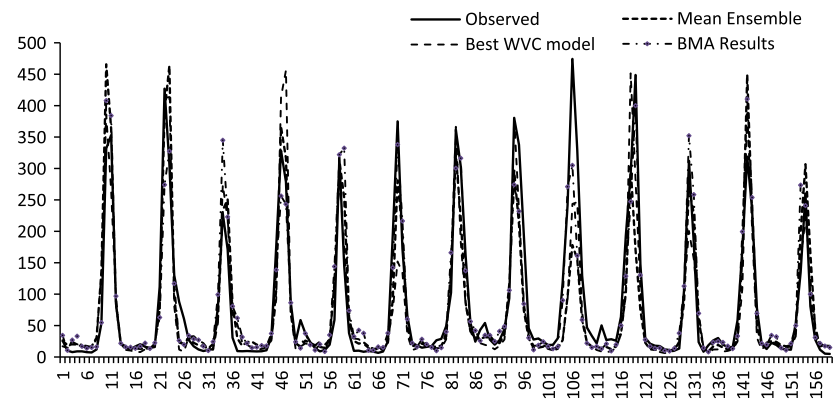

Fig. 8a. Validation results for Station I at a monthly scale.

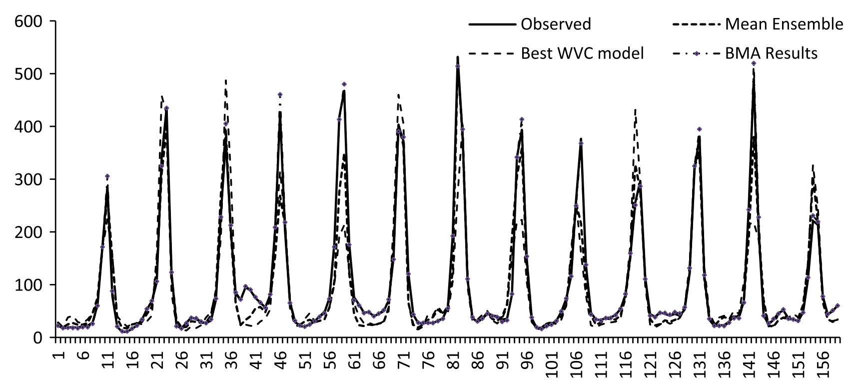

Fig. 8b. Validation results for Station II at a monthly scale.

Table 3
Results for the different models for Stations I and II.

Performance measures	Best single wavelet Volterra (WVC)	Average WVC ensemble	BMA based WVC ensemble
Station I			
RMSE (m^3/s)	51.30	53.69	42.54
MAE (m^3/s)	30.24	28.92	24.56
Correlation coefficient	0.89	0.88	0.92
NSC	0.78	0.77	0.85
Station II			
RMSE (m^3/s)	36.54	34.56	26.56
MAE (m^3/s)	18.35	19.26	13.56
Correlation coefficient	0.94	0.96	0.98
NSC	0.90	0.91	0.95

may be due to the fact that all the models in the ensemble formation may be having the same prediction probabilities. Overall, it can be seen that using the BMA technique increased WVC model performance and reduced the uncertainty in the predictions (see Fig. 14).

7. Summary and conclusions

Different wavelets have unique properties and thus possess different strengths to capture specific aspects of streamflow time series. Therefore, it is highly desirable to have a combined forecasting model that possesses the advantages of different wavelets. In this study, a new wavelet ensemble forecasting approach (i.e., the BMA based ensemble multi-wavelet Volterra approach) that takes advantage of different properties of multiple wavelets was proposed and tested for the first time. This was accomplished through a weighting strategy based on Bayesian Model Averaging. The major findings of this study are:

1. Different wavelets having unique properties possess the ability to capture specific features of real world processes such as streamflows. In this study, it was seen that in the monthly time series, the Haar wavelet and the db wavelet performed well in capturing the low flows, whereas the Symlet wavelet captured the high flows and time to peaks more accurately. It was observed that at a weekly time scale, the Haar, db2, and db4 wavelet based models produced better results than the Symlet wavelet based models. However, in the case of the daily scale, the Haar and db2 wavelet based models outperformed the other wavelet based models for both high and low flows. This issue has not really been explored to date in the literature in any

Table 4
Different single wavelet Volterra models for one step ahead forecasting at a weekly scale.

Wavelet used	Level of decomposition	Memory (m) at each level {DW_i's, C}	RMSE (m^3/s)	MAE (m^3/s)	NSC
Station II					
B3-Spline	3	{2,2,2,1}	55.08	16.89	0.75
B3-Spline	4	{1,2,2,2,1}	58.82	29.91	0.68
B3-Spline	5	{2,2,2,2,2,1}	65.85	32.54	0.58
db2	4	{2,2,1,1,2}	56.4	29.45	0.72
db2	3	{1,1,1,2}	53.15	26.54	0.76
db2	4	{1,2,2,2,1}	56.45	29.04	0.70
Haar	3	{1,1,1,2}	54.56	26.95	0.75
Haar	3	{2,2,2,1}	52.14	26.43	0.79
Sym4	4	{2,2,2,22}	57.45	27.45	0.74
Sym4	5	{1,1,1,2,2,2}	58.45	29.82	0.69
Station I					
B3-Spline	3	{2,2,2,1}	35.08	18.97	0.77
Db5	4	{3,2,2,2,1}	39.02	20.56	0.65
Sym4	3	{1,1,2,1}	34.65	16.7	0.78
Haar	3	{1,2,2,2}	33.78	16.07	0.79
Db4	4	{1,1,2,3,3}	38.2	18.69	0.68
Db2	3	{1,2,2,2}	34.35	16.73	0.79

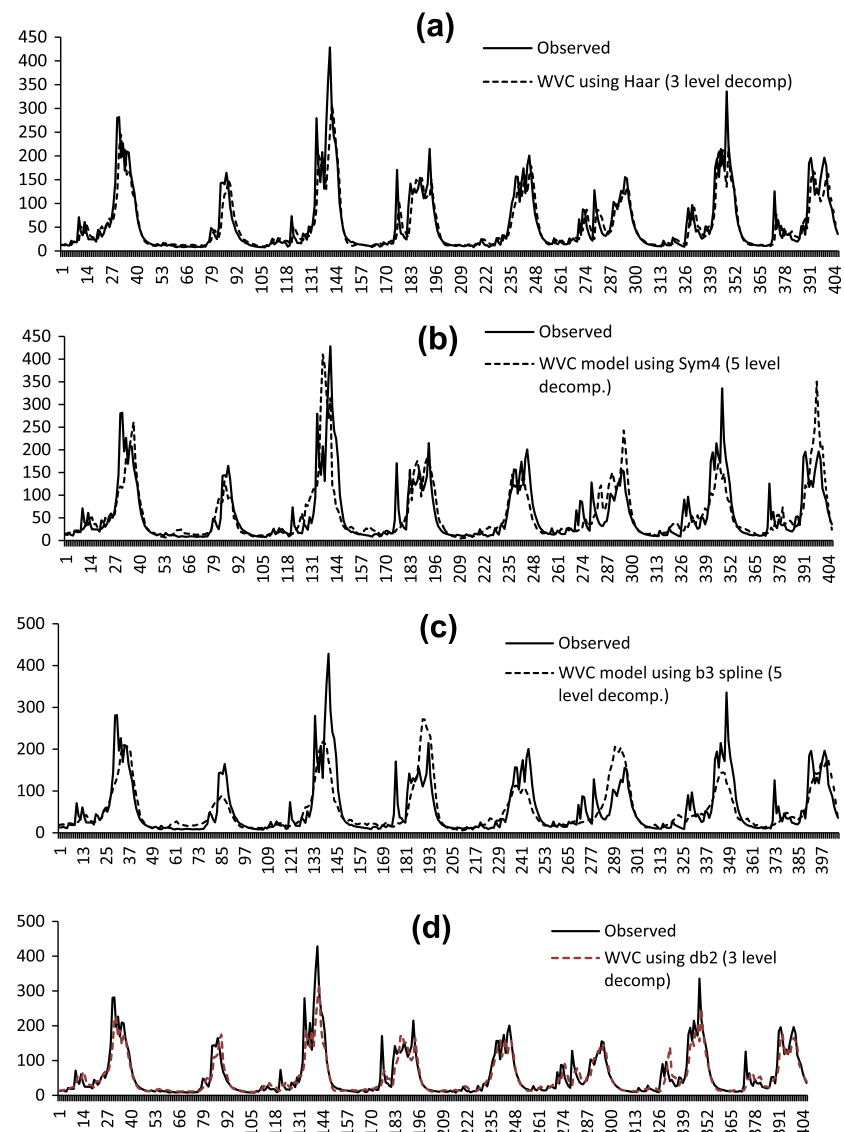

Fig. 9. Model results for WVC model using (a) Haar wavelet, (b) Sym 4, (c) Spline and (d) db2 wavelet at Station I.

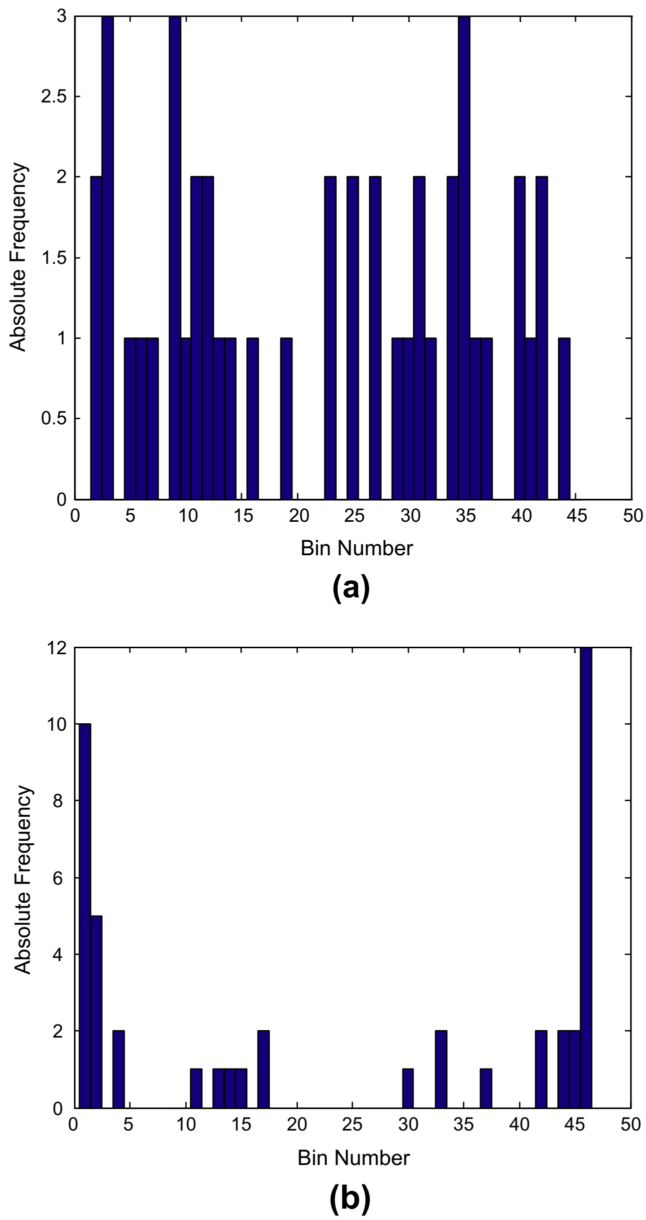

Fig. 10. Rank Histogram of the BMA–WVC ensembles for (a) Station I and (b) Station II.

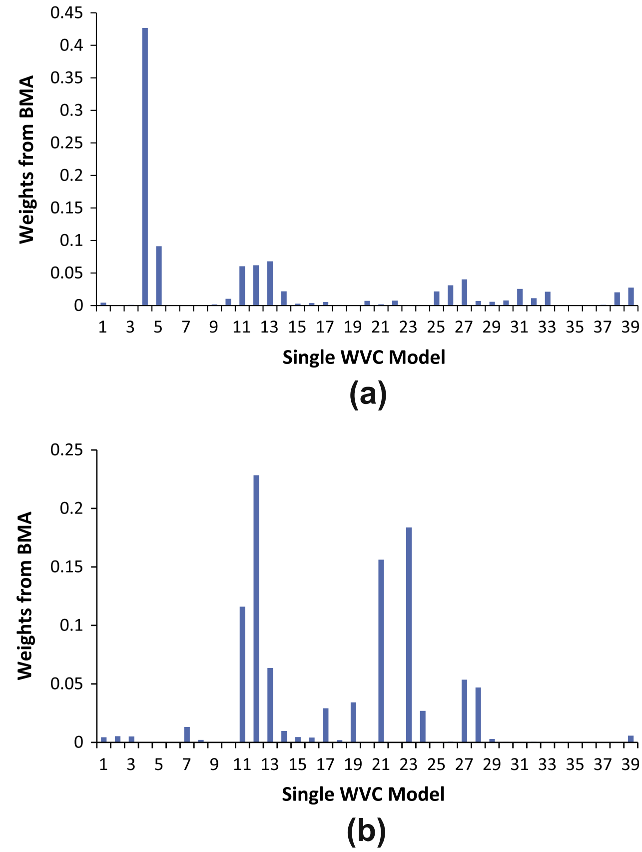

Fig. 11. A single set of BMA weights computed over the entire training period for the BMA WVC models at a weekly scale at (a) Station I and (b) Station II.

detail, so the results of this study should prove to be very useful in helping to select appropriate wavelets for different time scales in streamflow forecasting.

2. It was found that simply averaging the model results by assigning equal weights to the model (i.e. the mean average ensemble WVC approach) does not improve the results in comparison with the best single model. In some cases, averaging decreases the performance of the model. However, weighted averaging based on the BMA technique improved model performance significantly. In the case of monthly forecasting, the BMA based WVC model improved the forecasting performance by 9% and 20% with respect to the best single WVC model for Stations I and II, respectively. Similarly, in the case of the weekly forecasts, the BMA based WVC model improved the forecasting performance by 3.85% and 14% for Stations I and II, respectively. In the case of daily forecasts, the BMA based WVC model produced slightly better results than the best single WVC model. From these results, it appears that the BMA approach is particularly useful for longer lead times.
3. The proposed algorithm combining the BMA technique with the WVC model helped in combining model results from the multiple wavelet model setups and thereby helped in improving the model performance overall. By using the BMA based multiple wavelet ensemble WVC model, different flow regimes were captured properly. Also, it was observed that the weights from the BMA based WVC model provided insights into the type of multi-wavelet combination model that is suitable for a specific time scale. For example, the Haar and db2 wavelets were found to be suitable for daily forecasting, whereas for monthly forecasting, a combination of Haar, db2 and sym2 wavelets provided better forecasts. In the case of weekly forecasting, the model using db2, db4 and Haar wavelets performed the best.

Recommended future research stemming from this study could include:

1. Exploring multi-step ahead forecasting (for example 1, 2, and 3 day ahead; 1, 2, and 3 week ahead; and 1, 2 and 3 month ahead forecasting) using the newly proposed BMA based ensemble multi-wavelet Volterra model.
2. Using different weighting strategies for different flow regimes.
3. Applying the method in other study areas and for different applications (for example groundwater level forecasting, water demand forecasting, etc.) to assess the new BMA WVC approach for different hydrological forecasting applications.

Table 5
Summary results for the different models for Stations I and II at a weekly scale.

Performance measures	Best single wavelet Volterra (WVC)	Average WVC ensemble	BMA based WVC ensemble
Station I			
RMSE (m^3/s)	52.14	54.5	46.58
MAE (m^3/s)	26.43	26.43	25.01
Correlation coefficient	0.901	0.878	0.92
NSC	0.79	0.75	0.81
Station II			
RMSE (m^3/s)	33.18	33.35	26.45
MAE (m^3/s)	16.07	18.28	13.44
Correlation coefficient	0.89	0.87	0.950
NSC	0.81	0.80	0.87

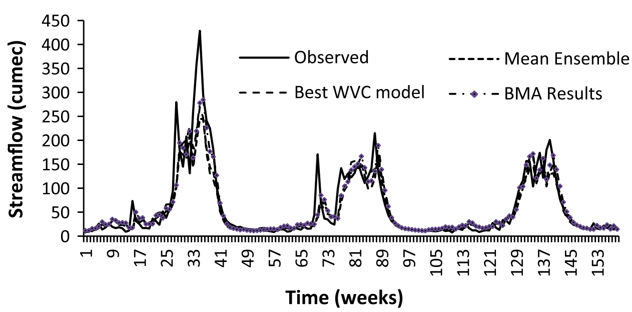

Fig. 12a. Validation results for different models at Station I.

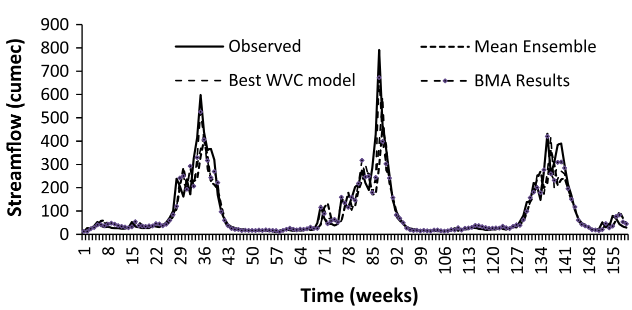

Fig. 12b. Validation results for different models at Station II.

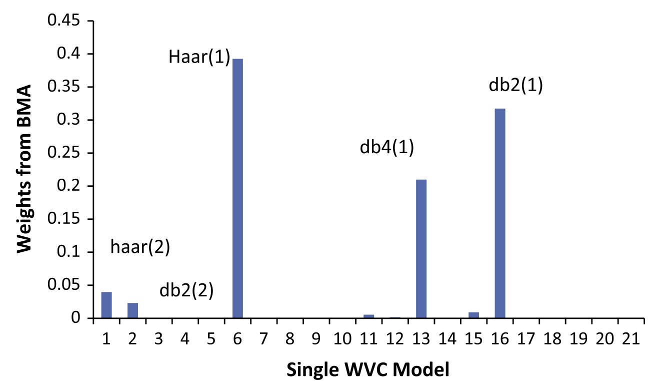

Fig. 13. A single set of BMA weights computed over the entire training period for the BMA WVC models at a daily scale at Station I.

4. Exploring multivariate forecasting (so in addition to flow, including other variables such as precipitation and temperature as input variables).
5. In this study we used existing wavelets for extracting the features from the stream flow time series, however, designing an adaptive wavelet transform matching the characteristics and features of the times series under study could potentially improve forecasting performance significantly.

Table 6
Different single wavelet Volterra models for one step ahead forecasting at a daily scale.

Wavelet used	Level of decomposition	Memory (m) at each level {DW_i's, C}	RMSE (m^3/s)	MAE (m^3/s)	NSC
Station I					
Haar	1	{2,2}	572.11	217.56	0.96
db2	1	{2,2}	790.48	320.12	0.93
db2	1	{3,2}	817.02	330.25	0.917
db2	2	{2,1,2}	891.78	393.28	0.90
db2	3	{1,2,2,1}	1063.7	479.99	0.86
Haar	2	{1,1,1,2}	1300.56	584.74	0.79
Haar	3	{2,2,2,1}	1404.78	700.78	0.72
Sym4	3	{2,2,2,2}	2000.23	1208.3	0.35
Sym5	2	{1,2,2}	2088.63	1131.27	0.45
db4	2	{2,2,2}	1299.89	626.40	0.79
db4	2	{2,2,1}	1331.3	658.47	0.77
db5	1	{2,2}	1071.23	471.87	0.85
db4	1	{3,2}	1006.16	433.32	0.88
db4	3	{2,2,2,2}	1507.87	768.63	0.71
Station II					
Haar	1	{2,2}	669.46	292.91	0.98
db2	1	{2,2}	932.00	418.33	0.96
db2	2	{2,3,1}	1140.5	520.78	0.95
db4	1	{2,1}	1154.03	540.78	0.94
Sym4	1	{2,1}	1825.55	900.68	0.86
Sym5	1	{2,2}	1922.17	1000.7	0.81

Table 7
Summary of results for the different models for Stations I and II at a daily scale.

Performance measures	Best single wavelet Volterra (WVC)	Average WVC ensemble	BMA based WVC ensemble
Station I			
RMSE (m^3/s)	572.11	582.7	512.45
MAE (m^3/s)	217.56	240.5	200.96
Correlation Coefficient	0.98	0.96	0.99
NSC	0.96	0.945	0.98
Station II			
RMSE (m^3/s)	669.46	650.79	600.78
MAE (m^3/s)	292.91	276.97	260.21
Correlation Coefficient	0.99	0.99	0.992
NSC	0.98	0.985	0.99

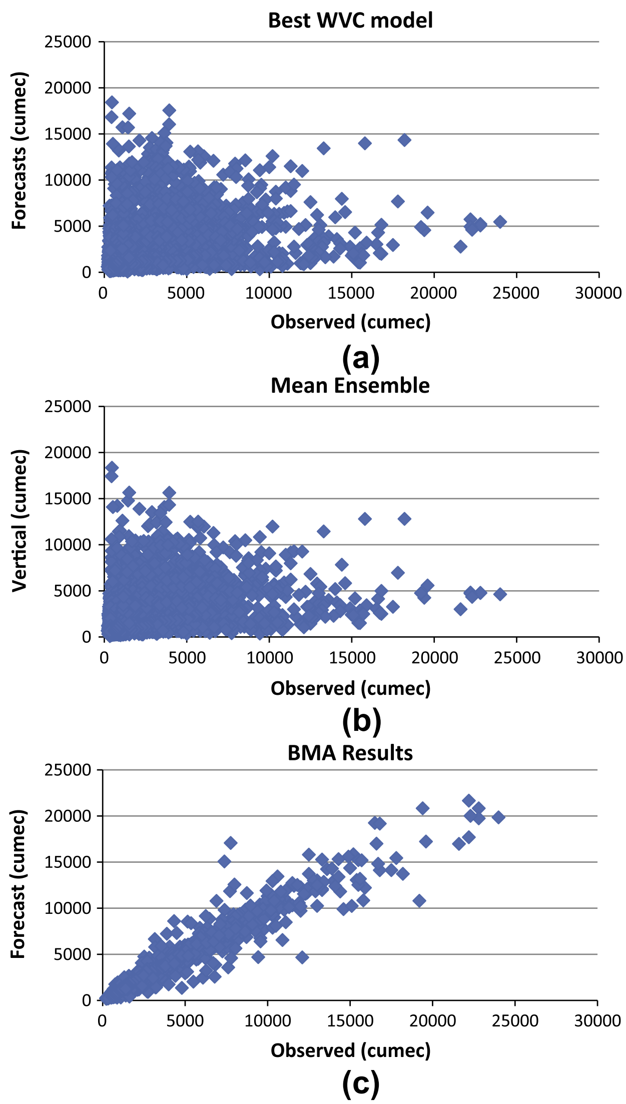

Fig. 14. Scatter plot for one day ahead forecasts at Station I using (a) single best WVC model, (b) mean ensemble WVC model and (c) BMA based WVC model.

Acknowledgements

This research was partially funded by an NSERC Discovery Grant held by Jan Adamowski.

Appendix A. Volterra Model

Volterra series of integral operators was introduced by Volterra (1930) and has since developed into a powerful generic nonlinear modelling tool. The domain of applied hydrology has also seen application of Volterra representation based approach and some notable contributions have come from Amorocho and Orlob (1961) and Amorocho and Brandstetter, (1971).

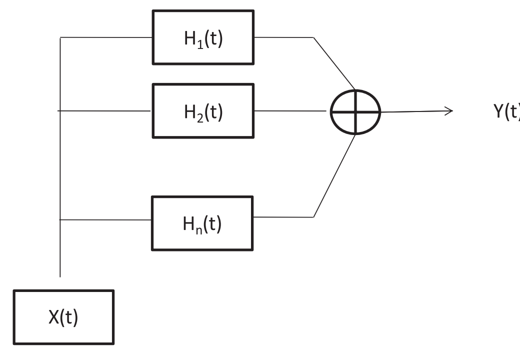

Fig. A.1. n[th] order Volterra Representation.

$$y(t) = \int_0^t h_1(\tau_1)X(\tau-\tau_1)d\tau_1 + \int_0^t\int_0^t h_2(\tau_1,\tau_2)X(\tau-\tau_1) X(\tau-\tau_2)d\tau_1 d\tau_2 + \int_0^t\int_0^t\int_0^t h_2(\tau_1,\tau_2,\tau_3)X(\tau-\tau_1)X(\tau-\tau_2) X(\tau-\tau_3)d\tau_1 d\tau_2 d\tau_3 + \cdots\cdots \quad (A.1)$$

Volterra representation is a simple extension of the Taylor series expansion for nonlinear autonomous causal systems with memory and may be written as

$$y(t) = H_1[x(t)] + H_2[x(t)] + H_3[x(t)] + \cdots\cdots H_n[x(t)] + \cdots \quad (A.2)$$

In the preceding equation, Y(t) represents the system output and $x(t)$ denotes the input to the system. The functions h_n $(\tau_1, \tau_2, \cdots, \tau_n)$ are called the Volterra kernels of the system and the transformation Hn[x(t)] represents a special type of a convolution integral known as the n[th] order *Volterra operator*. Figure A.1 shows a schematic representation of the Volterra model in the form of a block diagram.

References

Adamowski, K., Prokoph, A., Adamowski, J., 2009. Development of a new method of wavelet aided trend detection and estimation. Hydrol. Process. 23 (18), 2686–2696.
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. J. Hydrol. Eng. 15 (10), 729–743.
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.
Adamowski, J., Fung Chan, H., 2012. A wavelet neural network conjunction model for groundwater level forecasting. J. Hydrol. 407 (1–4), 28–40.

Ajami, N.K., Duan, Q., Sorooshian, S., 2007. An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resour. Res. 43, W01403. http://dx.doi.org/10.1029/2005WR004745.

Amorocho, J., 1973. Nonlinear hydrologic analysis. Adv. Hydrosci. 9, 203–251.

Amorocho, J., Brandstetter, A., 1971. Determination of Nonlinear Functional Response Functions in Rainfall-Runoff Processes. Water Resour. Res. 7 (5), 1087–1101.

Amorocho, J., Orlob, G.T., 1961. Non-linear analysis of hydrologic systems, Water Resources Center. Univ. California (Berkeley) 40, 147.

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

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

Chen, G.Y., Xie, W.F., 2007. Pattern recognition with SVM and dual-tree complex wavelets. Image and Vision Computing 25 (6), 960–966.

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

Diskin, M.H., Boneh, A., 1972. Properties of the kernels for time invariant, initially relaxed, second order, surface runoff systems. J. Hydrol 17 (1), 115–141.

Dhanya, Kumar D., 2011. Predictive uncertainty of chaotic daily streamflow using ensemble wavelet networks approach. Water Resour. Res., W06507.

Duan, Q., Ajami, N.K., Gao, X., Sorooshian, S., 2007. Multimodel ensemble hydrologic prediction using Bayesian model averaging. Adv. Water Resour. 30, 1371–1386.

Fernandez, C., Ley, E., Steel, M.F.J., 2001. Benchmark priors for Bayesian model averaging. J. Econometrics 100, 381–427.

Fundel, F., Zappa, M., 2011. Hydrological ensemble forecasting in mesoscale catchments: Sensitivity to initial conditions and value of reforecasts. Water Resour. Res. 47, W09520.

Georgakakos, K.P., Seo, D.J., Gupta, H., Schake, J., Butts, M.B., 2004. Characterizing streamflow simulation uncertainty through multimodel ensembles. J. Hydrol. 298, 222–241.

Hamil, Thomas, M., 2001. Interpretation of Rank Histograms for Verifying Ensemble Forecasts. Mon. Wea. Rev. 129, 550–560.

Hendry, D.F., Clements, M.P., 2002. Pooling of forecasts. Econometrics J. 5, 1–26.

Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial. Stat. Sci. 14 (4), 382–417.

Kallache, M., Rust, H.W., Kropp, J., 2005. Trend assessment: Applications for hydrology and climate research. Nonlinear Processes Geophys. 12, 201–210.

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

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

Labat, D., 2005. Recent advances in wavelet analyses: Part 1. A review of concepts. J. Hydrol. 314 (1–4), 275–288.

Labat, D., Ababou, R., 2001. Introduction of wavelet analyses to rainfall/runoffs relationship for a karstic basin: the case of Licq-Atherey karstic system (France). Ground Water 39 (4), 605–615.

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), 149–178.

Leamer, E.E., 1978. Specication Searches: Ad hoc Inference with Nonexperimental Data. Wiley, New York.

Ley, E., Steel, M.F.J., 2009. On the effect of prior assumptions in Bayesian model averaging with applications to growth regression. J. Appl. Econometrics 24, 651–674.

Maheswaran, R., Khosa, R., 2011a. Multiscale nonlinear model for monthly stream flow forecasting: a wavelet based approach. J. Hydroinform. 14 (2), 424–442.

Maheswaran, R., Khosa, R., 2011b. Comparative study of different wavelets for hydrologic forecasting. Comput. Geosci. 46, 284–295.

Maheswaran, R., Khosa, R., 2011c. Wavelet-Volterra coupled model for monthly stream flow forecasting. J. Hydrol. 450–451, 320–335.

Min, C., Zellner, A., 1993. Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. J. Econometrics 56 (1–2), 89–118.

Nalley, D., Adamowski, J., Khalil, B., 2012. Using discrete wavelet transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954–2008). J. Hydrol. 475, 204–228.

Neupauer, R.M., Powell, K.L., Qi, X., Lee, D.H., Villhauer, D.A., 2006. Characterization of permeability anisotropy using wavelet analysis. Water Resour. Res. 42, W07419.

Nourani, V., Alami, M.T., Aminfar, M.H., 2008. A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. Eng. Appl. Artif. Intel. 22 (3), 466–472.

Nourani, V., Komasi, M., Mano, A., 2009. A multivariate ANN-wavelet approach for rainfall-runoff modeling. Water Resour. Manage. 23 (14), 2877–2894.

Parrish, M.A., Moradkhani, H., DeChant, C.M., 2012. Toward reduction of model uncertainty: integration of Bayesian model averaging and data assimilation. Water Resour. Res. 48, 1–18.

Partal, T., 2009. Modelling evapotranspiration using discrete wavelet transform and neural networks. Hydrol. Process. 23, 3545–3555.

Partal, T., Küçük, M., 2006. Long-term trend analysis using discrete wavelet components of annual precipitations measurements in Marmara region Turkey. Phys. Chem. Earth 31 (18), 1189–1200.

Raftery, A.E., Madigan, D., Hoeting, J., 1997. Bayesian model averaging for linear regression models. J. Am. Stat. Assoc. 92, 179–191.

Raftery, A.E., Gneiting, T., Balabdaoui, F., Polakowski, M., 2005. Using Bayesian model averaging to calibrate forecast ensembles. Mon. Weather Rev. 133, 1155–1174.

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

Rings, J., Vrugt, J.A., Schoups, G., Huisman, J.A., Vereecken, H., 2012. Bayesian model averaging using particle filtering and Gaussian mixture modeling: theory, concepts, and simulation experiments. Water Resour. Res. 48, W05520.

Sang, Y.F., Wang, D., Wu, J.C., Zhu, Q.P., 2010. Wavelet cross-correlation method for hydrologic time series analysis. J. Hydraul. Eng. 41 (11), 1172–1179.

Sang, Y.F., Wang, Z.G., Liu, C.M., 2012. Discrete wavelet-based trend identification in hydrologic time series. Hydrol. Process.

Shamseldin, A.Y., O'Connor, K.M., 1999. A real-time combination method for the outputs of different rainfall-runoff models. Hydrological Sciences Journal 44 (6), 895–912.

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

Steel, M.F., 2011. Bayesian model averaging and forecasting. Bulletin of EU and US Inflation and Macroeconomic Analysis, 200.

Stock, J.H., Watson, M.W., 2002. Macroeconomic forecasting using diffusion indexes. J. Business Econ. Stat. 20, 147–162.

Talagrand, O., Vautard, R., Strauss, B., 1999. Evaluation of probabilistic prediction system. Proc. Workshop on Predictability, Reading, United Kingdom, ECMWF, pp. 1–25.

Tiwari, M.K., Chatterjee, C., 2010. Development of an accurate and reliable hourly flood forecasting model using Wavelet-Bootstrap-ANN hybrid approach. J. Hydrol. 394, 458–470.

Vrugt, J.A., Robinson, B.A., 2007. Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging. Water Resour. Res. 43, W01411.

Wang, W., Ding, J., 2003. Wavelet network model and its application to the prediction of hydrology. Science 1 (1), 67–71.

Wei, H.L., Billings, S.A., 2004. A unified wavelet-based modelling framework for nonlinear system identification: the WANARX model structure. Int. J. Control 77 (4), 351–366.

Zalachori, I., Ramos, M.H., Garçon, R., Mathevet, T., Gailhard, J., 2012. Statistical processing of forecasts for hydrological ensemble prediction: a comparative study of different bias correction strategies. Advances in Science & Research, 8.

Zhang, X., Zhao, K., 2012. Bayesian Neural Networks for Uncertainty Analysis of Hydrologic Modeling: A Comparison of Two Schemes. Water Resour. Manage. 26 (8), 2365–2382.