CrossMark

ORIGINAL ARTICLE

Short-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet transforms and machine learning methods

A. Belayneh[1] · J. Adamowski[2] · B. Khalil[2]

Abstract Droughts can cause significant damage to agricultural and other systems. An important aspect of mitigating the impacts of drought is an effective method of forecasting future drought events. In this study, five methods of forecasting drought for short lead times were explored in the Awash River Basin of Ethiopia. The Standard Precipitation Index (SPI) was the drought index chosen to represent drought in the basin. Machine learning techniques including artificial neural networks (ANNs) and support vector regression (SVR) were compared with coupled models (WA-ANN and WA-SVR) which pre-process input data using wavelet analysis (WA). This study proposed and tested the SVR and WA-SVR methods for short term drought forecasting. This study also used only the approximation series (derived via wavelet analysis) as inputs to the ANN and SVR models, and found that using just the approximation series as inputs for models gave good forecast results. The forecast results of all five data driven models were compared using several performance measures (RMSE, MAE, R^2 and a measure of persistence). The forecast results of this study indicate that the coupled wavelet neural network (WA-ANN) models were the best models for forecasting SPI 3 and SPI 6 values over lead times of 1 and 3 months in the Awash River Basin.

✉ A. Belayneh
antenehbelayneh@cmail.carleton.ca

J. Adamowski
jan.adamowski@mcgill.ca

[1] School of Public Policy and Administration, Carleton University, Ottawa, ON K1S 5B6, Canada

[2] Department of Bioresource Engineering, Faculty of Agricultural and Environmental Sciences, McGill University, Quebec H9X 3V9, Canada

Introduction

Drought is a natural phenomenon that occurs when precipitation is significantly lower than normal (Belayneh et al. 2014; Saadat et al. 2011). These deficits may cause low crop yields for agriculture, reduced flows for ecological systems, loss of biodiversity and other problems for the environment, in addition to adversely impacting the hydroelectric industry, as well as causing deficits in the drinking water supply which can negatively affect local populations. The less predictable characteristics of droughts such as their initiation, termination, frequency and severity can make drought both a hazard and a disaster. Drought is characterized as a hazard because it is a natural accident of unpredictable occurrence, but of recognizable recurrence (Mishra and Singh 2010). Drought is also characterized as a disaster because it corresponds to the failure of the precipitation regime, causing the disruption of the water supply to natural and agricultural ecosystems as well as to other human activities (Mishra and Singh 2010). 22 % of the global damage caused by natural disasters can be attributed to droughts (Keshavarz et al. 2013).

Droughts have also had a great impact in Africa. The Sahel has experienced droughts of unprecedented severity in recorded history (Mishra and Singh 2010). The impacts in sub-Saharan Africa are more severe because rain-fed agriculture comprises 95 % of all agriculture (Keshavarz et al. 2013). In 2009, reduced rainfall levels led to an increase in the frequency in droughts and resulted in an

increase of 53 million food insecure people in the region (Husak et al. 2013). Due to the changes in climate in Africa and around the world (Adamowski et al. 2009, 2010; Nalley et al. 2012, 2013; Pingale et al. 2014), it is likely that droughts will become more severe in the future.

Due to their slow evolution in time, droughts are phenomena whose consequences take a significant amount of time with respect to their inception to be perceived by both ecological and socio-economic systems. Due to this characteristic, effective mitigation of the most adverse drought impacts is possible, more than in the case of the other extreme hydrological events such as floods, earthquakes or hurricanes, provided a drought monitoring system which is able to promptly warn of the onset of a drought and to follow its evolution in space and time is in operation (Rossi et al. 2007). An accurate selection of indices for drought identification, providing a synthetic and objective description of drought conditions and future drought conditions, represents a key point for the implementation of an efficient drought warning system (Cacciamani et al. 2007). This can help local stakeholders to try and adapt to the effects of droughts in an effective and sustainable manner (Halbe et al. 2013; Kolinjivadi et al. 2014a, b; Straith et al. 2014; Inam et al. 2015; Butler and Adamowski 2015).

In this study, the drought index chosen to forecast drought is the Standardized Precipitation Index (SPI), which was developed to quantify a precipitation deficit for different time scales (Guttman 1999). The Awash River Basin was the study basin explored in this study and the SPI index was used to forecast drought mainly because the SPI drought index requires precipitation as its only input. Furthermore, it has been determined that precipitation alone can explain most of the variability of East African droughts and that the SPI is an appropriate index for monitoring droughts in East Africa (Ntale and Gan 2003).

In hydrologic drought forecasting, stochastic methods have been traditionally used to forecast drought indices. Markov Chain models (Paulo et al. 2005; Paulo and Pereira 2008) and autoregressive integrated moving average models (ARIMA) (Mishra and Desai 2005, 2006; Mishra et al. 2007; Han et al. 2010) have been the most widely used stochastic models for hydrologic drought forecasting. The major limitation of these models is that they are linear models and they are not very effective in forecasting nonlinearities, a common characteristic of hydrologic data (Tiwari and Adamowski 2014; Campisi et al. 2012; Adamowski et al. 2012; Haidary et al. 2013).

In response to non-linear data, researchers in the last two decades have increasingly begun to forecast hydrological data using artificial neural networks (ANNs). ANNs have been used to forecast droughts in several studies (Mishra and Desai 2006; Morid et al. 2007; Bacanli et al. 2008; Barros and Bowden 2008; Cutore et al. 2009; Karamouz

et al. 2009; Marj and Meijerink 2011). However, ANNs are limited in their ability to deal with non-stationarities in the data, a weakness also shared by ARIMA and other stochastic models.

Support vector regression (SVRs) are a relatively new form of machine learning that was developed by Vapnik (1995), and which have been recently used in the field of hydrological forecasting. There are several studies where SVRs were used in hydrological forecasting. Khan and Coulibaly (2006) found that a SVR model was more effective at predicting 3–12 month lake water levels than ANN models. Kisi and Cimen (2009) used SVRs to estimate daily evaporation. Finally, SVRs have been successfully used to predict hourly streamflow (Asefa et al. 2006), and were shown to perform better than ANN and ARIMA models for monthly streamflow prediction (Wang et al. 2009; Maity et al. 2010), respectively. SVRs have also been applied in drought forecasting (Belayneh and Adamowski 2012).

Wavelet analysis, an effective tool to deal with non-stationary data, is an emerging tool for hydrologic forecasting and has recently been applied to: examine the rainfall–runoff relationship in a Karstic watershed (Labat et al. 1999), to characterize daily streamflow (Saco and Kumar 2000) and monthly reservoir inflow (Coulibaly et al. 2000), to evaluate rainfall–runoff models (Lane 2007), to forecast river flow (Adamowski 2008; Adamowski and Sun 2010; Ozger et al. 2012; Rathinasamy et al. 2014; Nourani et al. 2014), to forecast groundwater levels (Adamowski and Chan 2011), to forecast future precipitation values (Partal and Kisi 2007) and for the purposes of drought forecasting (Kim and Valdes 2003; Ozger et al. 2012; Mishra and Singh 2012; Belayneh and Adamowski 2012).

The effectiveness of these data-driven models and wavelet analysis, coupled with ANN and SVR models, has been shown in a variety of study locations. Kim and Valdes (2003) used WA-ANN models to forecast drought in the semi-arid climate of the Conchos River Basin of Mexico. Mishra and Desai (2006) used ANN models to forecast drought in the Kansabati River Basin of India. Bacanli et al. (2008) forecast the SPI in Central Anatolia where the precipitation was concentrated in the spring and winter and where the temperature difference between summer and winter was extremely high. Ozger et al. (2012) coupled wavelet analysis with artificial intelligence models to forecast long term drought in Texas, while Mishra and Singh (2012) investigated the relationship between meteorological drought and hydrological drought using wavelet analysis in different regions of the United States. While the principal reason for the use of these models in the aforementioned areas is the susceptibility of these regions to drought, the variability in climatic conditions highlights

how versatile and effective these new forecasting methods are.

Belayneh and Adamowski (2012) forecast SPI 3 and SPI 12 in the Awash River Basin using ANN, SVR and WA-SVR models. This study complements that study by forecasting the SPI over a larger selection of stations in the same area, and by coupling, SVR models with wavelet transforms. The main objective of the present study was to compare traditional drought forecasting methods such as ARIMA models with machine learning techniques such as ANNs and SVR, along with ANNs with data pre-processed using wavelet transforms (WA-ANN), SVR, and the coupling of wavelet transforms and support vector regression (WA-SVR) for short-term drought forecasting. The SPI, namely SPI 3 and SPI 6, were forecast using the above mentioned methods for lead times of 1 and 3 months in the Awash River Basin of Ethiopia. Both SPI 3 and SPI 6 are short-term drought indicators, and forecast lead times of 1 and 3 months represent the shortest possible monthly lead time and a short seasonal lead time, respectively.

Theoretical development

Development of SPI series

The SPI was developed by McKee et al. (1993). A number of advantages arise from the use of the SPI index. First, the index is based on precipitation alone making its evaluation relatively easy (Cacciamani et al. 2007). Secondly, the index makes it possible to describe drought on multiple time scales (Tsakiris and Vangelis 2004; Mishra and Desai 2006; Cacciamani et al. 2007). A third advantage of the SPI is its standardization which makes it particularly well suited to compare drought conditions among different time periods and regions with different climates (Cacciamani et al. 2007). A drought event occurs at the time when the value of the SPI is continuously negative; the event ends when the SPI becomes positive. Table 1 provides a drought classification based on SPI. Details regarding the

Table 1 Drought classification based on SPI (McKee et al. 1993)

SPI values	Class
>2	Extremely wet
1.5–1.99	Very wet
1.0–1.49	Moderately wet
−0.99 to 0.99	Near normal
−1 to −1.49	Moderately dry
−1.5 to −1.99	Very dry
<−2	Extremely dry

computation of the SPI can be found in Belayneh and Adamowski (2012) and Mishra and Desai (2006).

Autoregressive integrated moving average (ARIMA) models

Autoregressive integrated moving average models are amongst the most commonly used stochastic models for drought forecasting (Mishra and Desai 2005, 2006; Mishra et al. 2007; Cancelliere et al. 2007; Han et al. 2010).

The general non-seasonal ARIMA model may be written as (Box and Jenkins 1976):

$$z_t = \frac{\theta(B)a_t}{\phi(B)\nabla^d} \tag{1}$$

$$\phi(B) = (1 - \phi_t B - \phi_2 B^2 - \cdots - \phi_p B^p) \tag{2}$$

and

$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) \tag{3}$$

where z_t is the observed time series and B is a back shift operator. $\phi(B)$ and $\theta(B)$ are polynomials of order p and q, respectively. The orders p and q are the order of non-seasonal auto-regression and the order of non-seasonal moving average, respectively. Random errors, a_t are assumed to be independently and identically distributed with a mean of zero and a constant variance. ∇^d describes the differencing operation to data series to make the data series stationary and d is the number of regular differencing.

The time series model development consists of three stages: identification, estimation and diagnostic check (Box et al. 1994). In the identification stage, data transformation is often needed to make the time series stationary. Stationarity is a necessary condition in building an ARIMA model that is useful for forecasting (Zhang 2003). The estimation stage of model development consists of the estimation of model parameters. The last stage of model building is the diagnostic checking of model adequacy. This stage checks if the model assumptions about the errors are satisfied. Several diagnostic statistics and plots of the residuals can be used to examine the goodness of fit of the tentative model to the observed data. If the model is inadequate, a new tentative model should be identified, which is subsequently followed, again, by the stages of estimation and diagnostic checking.

Artificial neural network models

ANNs are flexible computing frameworks for modeling a broad range of nonlinear problems. They have many features which are attractive for forecasting such as their rapid development, rapid execution time and their ability to handle large amounts of data without very detailed

90

Sustain. Water Resour. Manag. (2016) 2:87–101

knowledge of the underlying physical characteristics (ASCE 2000a, b).

The ANN models used in this study have a feed forward multi-layer perceptron (MLP) architecture which was trained with the Levenberg–Marquardt (LM) back propagation algorithm. MLPs have often been used in hydrologic forecasting due to their simplicity. MLPs consist of an input layer, one or more hidden layers, and an output layer. The hidden layer contains the neuron-like processing elements that connect the input and output layers given by (Belayneh and Adamowski 2012):

$$y'_k(t) = f_0 \left[\sum_{j=1}^{m} w_{kj} \cdot f_n \left(\sum_{i=1}^{N} w_{ji} x_i(t) + (w_{j0}) + w_{k0} \right) \right] \quad (4)$$

where N is the number of samples, m is the number of hidden neurons, $x_i(t) =$ the ith input variable at time step t; $w_{ji} =$ weight that connects the ith neuron in the input layer and the jth neuron in the hidden layer; $w_{j0} =$ bias for the jth hidden neuron; $f_n =$ activation function of the hidden neuron; $w_{kj} =$ weight that connects the jth neuron in the hidden layer and kth neuron in the output layer; $w_{k0} =$ bias for the kth output neuron; $f_0 =$ activation function for the output neuron; and $y'_k(t)$ is the forecasted kth output at time step t (Kim and Valdes 2003).

MLPs were trained with the LM back propagation algorithm. This algorithm is based on the steepest gradient descent method and Gauss–Newton iteration. In the learning process, the interconnection weights are adjusted using the error convergence technique to obtain a desired output for a given input. In general, the error at the output layer in the model propagates backwards to the input layer through the hidden layer in the network to obtain the final desired output. The gradient descent method is utilized to calculate the weight of the network and adjusts the weight of interconnections to minimize the output error.

Support vector regression models

Support vector regression (SVRs) was introduced by Vapnik (1995) in an effort to characterize the properties of learning machines so that they can generalize well to unseen data (Kisi and Cimen 2011). SVRs embody the structural risk minimization principle, unlike conventional neural networks which adhere to the empirical risk minimization principle. As a result, SVRs seek to minimize the generalization error, while ANNs seek to minimize training error.

In regression estimation with SVR the purpose is to estimate a functional dependency $f(\vec{x})$ between a set of sampled points $X = \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_l$ taken from R^n and target

values $Y = \{y_1, y_2, \ldots, y_l\}$ with $y_i \in R$ [the input and target vectors (x_i's and y_i's) refer to the monthly records of the SPI index]. Detailed descriptions of SVR model development can be found in Cimen (2008).

Wavelet transforms

The first step in wavelet analysis is to choose a mother wavelet (ψ). The continuous wavelet transform (CWT) is defined as the sum over all time of the signal multiplied by scaled and shifted versions of the wavelet function ψ (Nason and Von Sachs 1999):

$$W(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t - \tau}{s} \right) dt \quad (5)$$

where s is the scale parameter; τ is the translation and * corresponds to the complex conjugate (Kim and Valdes 2003). The CWT produces a continuum of all scales as the output. Each scale corresponds to the width of the wavelet; hence, a larger scale means that more of a time series is used in the calculation of the coefficient than in smaller scales. The CWT is useful for processing different images and signals; however, it is not often used for forecasting due to its complexity and time requirements to compute. Instead, the successive wavelet is often discrete in forecasting applications to simplify the numerical calculations. The discrete wavelet transform (DWT) requires less computation time and is simpler to implement. DWT scales and positions are usually based on powers of two (dyadic scales and positions). This is achieved by modifying the wavelet representation to (Cannas et al. 2006):

$$\psi_{j,k}(t) = \frac{1}{\sqrt{|s_0^j|}} \psi \left(\frac{t - k\tau_0 s_0^j}{s_0^j} \right) \quad (6)$$

where j and k are integers that control the scale and translation respectively, while $s_0 > 1$ is a fixed dilation step (Cannas et al. 2006) and τ_0 is a translation factor that depends on the aforementioned dilation step. The effect of discretizing the wavelet is that the time–space scale is now sampled at discrete levels. The DWT operates two sets of functions: high-pass and low-pass filters. The original time series is passed through high-pass and low-pass filters, and detailed coefficients and approximation series are obtained.

One of the inherent challenges of using the DWT for forecasting applications is that it is not shift invariant (i.e. if we change values at the beginning of our time series, all of the wavelet coefficients will change). To overcome this problem, a redundant algorithm, known as the à trous algorithm can be used, given by (Mallat 1998):

$$C_{i+1}(k) = \sum_{l=-\infty}^{+\infty} h(l) c_i(k + 2^i l) \qquad (7)$$

where h is the low pass filter and the finest scale is the original time series. To extract the details, $w_i(k)$, that were eliminated in Eq. (7), the smoothed version of the signal is subtracted from the coarser signal that preceded it, given by (Murtagh et al. 2003):

$$w_i(k) = c_{i-1}(k) - c_i(k) \qquad (8)$$

where $c_i(k)$ is the approximation of the signal and $c_{i-1}(k)$ is the coarser signal. Each application of Eqs. (7) and (8) creates a smoother approximation and extracts a higher level of detail. Finally, the non-symmetric Haar wavelet can be used as the low pass filter to prevent any future information from being used during the decomposition (Renaud et al. 2002).

Study areas

In this study, the SPI was forecast for the Awash River Basin in Ethiopia (Fig. 1). The Awash River Basin was separated into three smaller basins for the purpose of this study on the basis of various factors such as location, altitude, climate, topography and agricultural development. The statistics of each station is shown in Table 2. Drought is a common occurrence in the Awash River Basin (Edossa et al. 2010). The heavy dependence of the population on rain-fed agriculture has made the people and the country's economy extremely vulnerable to the impacts of droughts. The mean annual rainfall of the basin varies from about 1600 mm in the highlands to 160 mm in the northern point of the basin. The total amount of rainfall also varies greatly from year to year, resulting in severe droughts in some
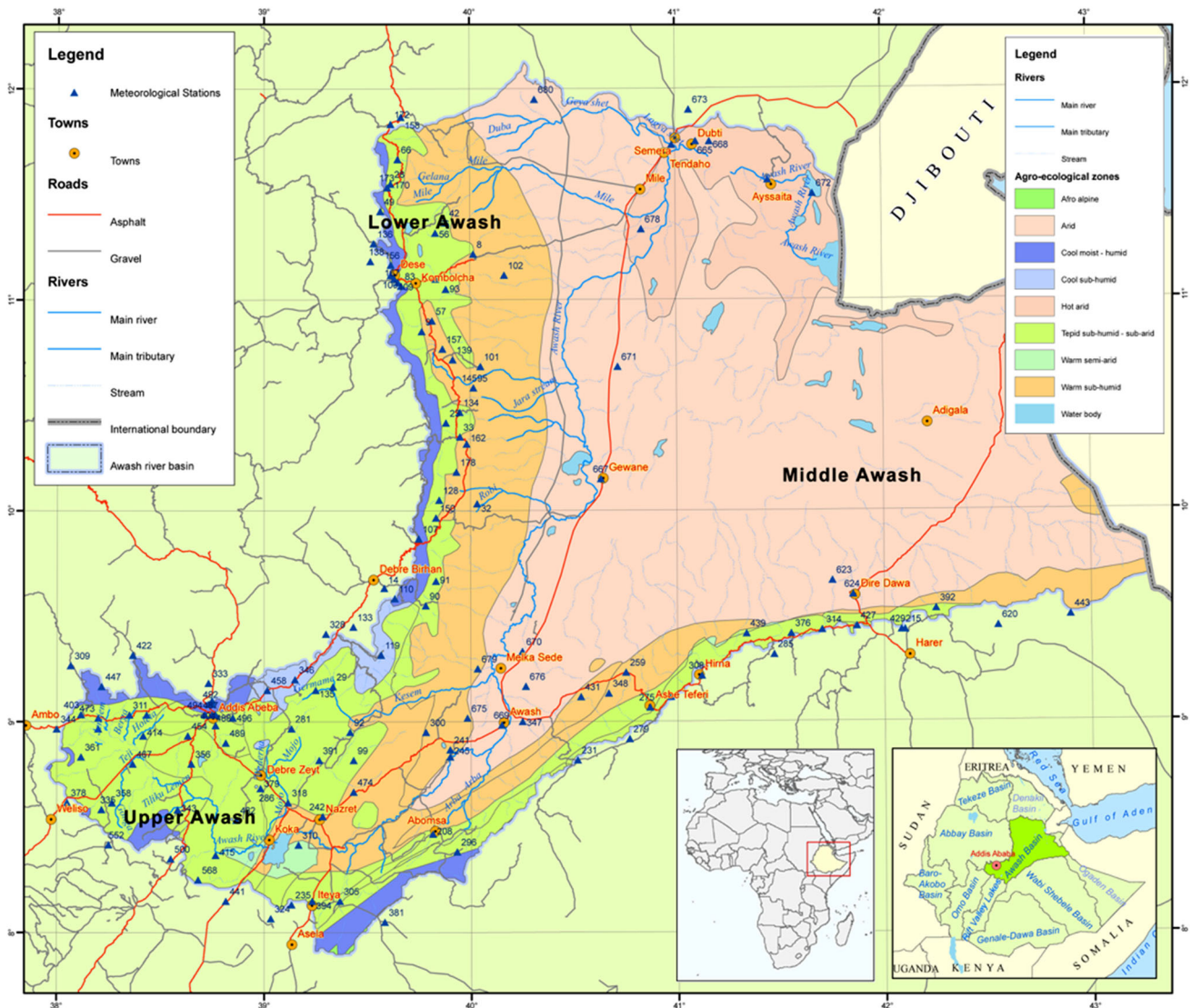

Fig. 1 Awash River Basin (Source: Ministry of Water Resources, Ethiopia. Agricultural Water Management Information System. http://www.mowr.gov.et/AWMISET/images/Awash_agroecologyv3.pdf. Accessed 06-June-2013)

92

Sustain. Water Resour. Manag. (2016) 2:87–101

Table 2 Descriptive statistics of the Awash River Basin

Basin	Station	Mean annual precipitation (mm)	Max annual precipitation (mm)	Standard deviation (mm)
Upper Awash Basin	Bantu Liben	91	647	111
	Ginchi	97	376	90
	Sebeta	111	1566	172
	Ejersalele	67	355	75
	Ziquala	100	583	110
Middle Awash Basin	Modjo	76	542	92
	Wolenchiti	76	836	95
	Gelemsso	77	448	75
	Dire Dawa	51	267	54
Lower Awash Basin	Eliwuha	44	374	57
	Dubti	87	449	89
	Bati	26	268	40

years and flooding in others. The total annual surface runoff in the Awash Basin amounts to some 4900×10^6 m^3 (Edossa et al. 2010). Effective forecasts of the SPI can be used for mitigating the impacts of drought that manifests as a result of rainfall shortages in the area. Rainfall records from 1970 to 2005 were used to generate SPI 3 and SPI 6 time series. The normal ratio method, recommended by Linsley et al. (1988), was used to estimate the missing rainfall records at some stations.

Methodology

ARIMA model development

Based on the Box and Jenkins approach, ARIMA models for the SPI time series were developed based on three steps: model identification, parameter estimation and diagnostic checking. The details on the development of ARIMA models for SPI time series can be found in the works of Mishra and Desai (2005) and Mishra et al. (2007).

In an ARIMA model, the value of a given times series is a linear aggregation of p previous values and a weighted sum of q previous deviations (Mishra and Desai 2006). These ARIMA models are autoregressive to order p and moving average to order q and operate on dth difference of the given times series. Hence, an ARIMA models is distinguished with three parameters (p, d, q) that can each have a positive integer value or a value of zero.

Wavelet transformation

When conducting wavelet analysis, the number of decomposition levels that is appropriate for the data must be chosen. Often the number of decomposition levels is chosen according to the signal length (Tiwari and Chatterjee 2010) given by $L = \text{int}[\log(N)]$ where L is the level of decomposition and N is the number of samples. According to this methodology the optimal number of decompositions for the SPI time series in this study would have been 3. In this study, each SPI time series was decomposed between 1 and 9 levels. The best results were compared at all decomposition levels to determine the appropriate level. The optimal decomposition level varied between models. Once a time series was decomposed into an appropriate level, the subsequent approximation series was either chosen on its own, in combination with relevant detail series or the relevant detail series were added together without the approximation series. With most SPI time series, choosing just the approximation series resulted in the best forecast results. In some cases, the summation of the approximation series with a decomposed detail series yielded the best forecast results. The appropriate approximation was used as an input to the ANN and SVR models. As discussed in "Wavelet transforms", the 'a trous' wavelet algorithm with a low pass Haar filter was used.

ANN models

The ANN models used to forecast the SPI were recursive models. The input layer for the models was comprised of the SPI values computed from each rainfall gauge in each sub-basin. The input data was standardized from 0 to 1.

All ANN models, without wavelet decomposed inputs, were created with the MATLAB (R.2010a) ANN toolbox. The hyperbolic tangent sigmoid transfer function was the activation function for the hidden layer, while the activation function for the output layer was a linear function. All the ANN models in this study were trained using the LM back propagation algorithm. The LM back propagation algorithm

was chosen because of its efficiency and reduced computational time in training models (Adamowski and Chan 2011).

There are between 3 and 5 inputs for each ANN model. The optimal number of input neurons was determined by trial and error, with the number of neurons that exhibited the lowest root mean square error (RMSE) value in the training set being selected. The inputs and outputs were normalized between 0 and 1. Traditionally the number of hidden neurons for ANN models is selected via a trial and error method. However a study by Wanas et al. (1998) empirically determined that the best performance of a neural network occurs when the number of hidden nodes is equal to $\log(N)$, where N is the number of training samples. Another study conducted by Mishra and Desai (2006) determined that the optimal number of hidden neurons is $2n + 1$, where n is the number of input neurons. In this study, the optimal number of hidden neurons was determined to be between $\log(N)$ and $(2n + 1)$. For example, if using the method proposed by Wanas et al. (1998) gave a result of four hidden neurons and using the method proposed by Mishra and Desai (2006) gave seven hidden neurons, the optimal number of hidden neurons is between 4 and 7; thereafter the optimal number was chosen via trial and error. These two methods helped establish an upper and lower bound for the number of hidden neurons.

For all the ANN models, 80 % of the data was used to train the models, while the remaining 20 % of the data was divided into a testing and validation set with each set comprising 10 % of the data.

WA-ANN models

The WA-ANN models were trained in the same way as the ANN models, with the exception that the inputs were made up from either, the approximation series, or a combination of the approximation and detail series after the appropriate wavelet decomposition was selected. The model architecture for WA-ANN models consists of 3–5 neurons in the input layer, 4–7 neurons in the hidden layer and one neuron in the output layer. The selection of the optimal number of neurons in both the input and hidden layers was done in the same way as for the ANN models. The data was partitioned into training, testing and validation sets in the same manner as ANN models.

Support vector regression models

All SVR models were created using the OnlineSVR software created by Parrella (2007), which can be used to build support vector machines for regression. The data was partitioned into two sets: a calibration set and a validation set. 90 % of the data was partitioned into the calibration set while the final 10 % of the data was used as the validation

set. Unlike neural networks the data can only be partitioned into two sets with the calibration set being equivalent to the training and testing sets found in neural networks. All inputs and outputs were normalized between 0 and 1.

All SVR models used the nonlinear radial basis function (RBF) kernel. As a result, each SVR model consisted of three parameters that were selected: gamma (γ), cost (C), and epsilon (ε). The γ parameter is a constant that reduces the model space and controls the complexity of the solution, while C is a positive constant that is a capacity control parameter, and ε is the loss function that describes the regression vector without all the input data (Kisi and Cimen 2011). These three parameters were selected based on a trial and error procedure. The combination of parameters that produced the lowest RMSE values for the calibration data sets were selected.

WA-SVR models

The WA-SVR models were trained in exactly the same way as the SVR models with the OnlineSVR software (2007) with the exception that the inputs were wavelet decomposed.

The data for WA-SVR models was partitioned exactly like the data for SVR. The optimal parameters for the WA-SVR models were chosen using the same procedure used to find the parameters for SVR models.

Performance measures

To evaluate the performances of the aforementioned data driven models the following measures of goodness of fit were used:

The coefficient of determination $(R^2) = \dfrac{\sum_{i=1}^{N} (\hat{y}_i - \bar{y}_i)}{\sum_{i=1}^{N} (y_i - \bar{y}_i)^2}$

$$(9)$$

where $\bar{y}_i = \dfrac{1}{N} \sum_{i=1}^{N} y_i$ $\qquad (10)$

where $\bar{y}_i$ is the mean value taken over N, y_i is the observed value, $\hat{y}_i$ is the forecasted value and N is the number of samples. The coefficient of determination measures the degree of correlation among the observed and predicted values. It is a measure of the strength of the model in developing a relationship among input and output variables. The higher the value of R^2 (with 1 being the highest possible value), the better the performance of the model.

The Root Mean Squared Error (RMSE) $= \sqrt{\dfrac{\text{SSE}}{N}}$ $\quad (11)$

where SSE is the sum of squared errors, and N is the number of data points used. SSE is given by:

$$\text{SSE} = \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 \qquad (12)$$

with the variables already having been defined. The RMSE evaluates the variance of errors independently of the sample size.

$$\text{The Mean Absolute Error (MAE)} = \sum_{i=1}^{N} \frac{|\hat{y}_i - y_i|}{N} \qquad (13)$$

The MAE is used to measure how close forecasted values are to the observed values. It is the average of the absolute errors.

Results and discussion

In the following sections, the forecast results for the best data driven models at each sub-basin are presented. The forecasts presented are from the validation data sets for time series of SPI 3 and SPI 6, which are mostly used to describe short-term drought (agricultural drought).

SPI 3 forecasts

The SPI 3 forecast results for all data driven models are presented in Tables 3 and 4. As the forecast lead time is increased, the forecast accuracy deteriorates for all stations. In the Upper Awash basin, the best data driven model for SPI 3 forecasts of 1 month lead time was a WA-ANN model. The WA-ANN model at the Ziquala station had the best results in terms of RMSE and MAE, with forecast results of 0.4072 and 0.3918, respectively. The Ginchi station had the best WA-ANN model in terms of R^2, with forecast results of 0.8808. When the forecast lead time is increased to 3 months, the best models remain WA-ANN models. The Bantu Liben station had the model with the lowest RMSE and MAE values of 0.5098 and 0.4941, respectively. The Sebeta station had the best results in terms of R^2, with a value of 0.7301.

In the Middle Awash basin, for forecasts of 1 month lead time, WA-ANN and WA-SVR models had the best forecast results. The WA-ANN model at the Modjo station had the best results in terms of R^2 with a value of 0.8564. However, unlike the Upper Awash basin, the best forecast results in terms of RMSE and MAE were from a WA-SVR model. The WA-SVR model at the Modjo station had the lowest RMSE and MAE values of 0.4309 and 0.4018, respectively. For forecasts of 3 months lead time, WA-ANN models had the best results across all performance measures with the Modjo station having the highest value

of R^2 at 0.6808 and the Gelemsso station having the lowest RMSE and MAE values of 0.5448 and 0.5334, respectively.

In the Lower Awash basin, for forecasts of 1 month lead time, the best results were from WA-ANN and WA-SVR models, similar to the Middle Awash basin. The highest value for R^2 was 0.7723 and it was from the WA-ANN model at the Eliwuha station. The lowest values for RMSE and MAE were 0.4048 and 0.3873, and were from the WA-SVR model at the Eliwuha station. For forecasts of 3 months lead time the best results were observed at the Bati station in terms of R^2. The WA-SVR model at this station had the highest R^2 value of 0.5915 and the WA-SVR model at the Eliwuha station had the lowest RMSE and MAE values of 0.5159 and 0.5129 respectively.

SPI 6 forecasts

The SPI 6 forecast data is presented in Tables 5 and 6. The forecast results for SPI 6 are significantly better than SPI 3 forecasts according to all three performance measures. In the Upper Awash Basin a WA-ANN model at the Bantu Liben station had the best forecast result in terms of RMSE with a result of 0.3438. Another WA-ANN model at the Ginchi station provided the lowest MAE value of 0.3212. The best forecast result in terms of R^2, 0.9163, was from a WA-SVR model at the Ziquala station. The forecast results in Tables 5 and 6 show that WA-ANN and WA-SVR models provide the best SPI 6 forecasts. Neither method is significantly better than the other. This is best illustrated by the fact the models provide very similar results according to the performance measures used.

As the forecast lead time is increased the forecast accuracy of all the models declines. This decline is most evident in the ARIMA, ANN and SVR models. The forecast accuracy for 3 month lead time forecasts are still significantly better for WA-ANN and WA-SVR models. In the Upper Awash Basin the best model in terms of R^2 is a WA-ANN model at the Sebeta station and has a forecast result of 0.7723. The best model in terms of RMSE and MAE is a WA-SVR model at the Ginchi station with forecast results of 0.4224 and 0.3864 respectively. In the Middle Awash Basin the best model in terms of R^2 is also a WA-ANN model and the Modjo station with a result of 0.7414 and the best model in terms of RMSE and MAE is a WA-SVR model at the Dire Dawa station with forecast results of 0.4049 and 0.3897 respectively. In the Lower Basin the best result in terms of R^2 and MAE is from a WA-SVR model at the Eliwuha and Dubti stations respectively and the best model in terms of RMSE is from a WA-ANN model at the Dubti station.

Table 3 The best ARIMA, ANN and SVR models for 1 and 3 month forecasts of SPI 3

Basin	Station	ANN models	R^2	RMSE	MAE	ARIMA models	R^2	RMSE	MAE	SVR (γ, C, ε)	R^2	RMSE	MAE
1 month lead time													
Upper	Bantu Liben	3-4-1	0.777	0.729	0.713	(5, 1, 0)	0.743	0.846	0.825	0.8, 98, 0.008	0.804	0.729	0.719
	Ginchi	4-4-1	0.725	0.744	0.729	(3, 0, 0)	0.698	0.832	0.824	0.5, 95, 0.007	0.753	0.703	0.792
	Sebeta	4-4-1	0.848	0.74	0.721	(3, 1, 1)	0.768	0.822	0.817	0.4, 90, 0.008	0.746	0.735	0.723
	Ejersalele	5-4-1	0.718	0.751	0.739	(1, 0, 0)	0.713	0.853	0.846	0.5, 95, 0.005	0.782	0.709	0.703
	Ziquala	3-4-1	0.735	0.715	0.704	(1, 0, 0)	0.718	0.846	0.836	0.6, 93, 0.004	0.792	0.739	0.714
Middle	Modjo	3-4-1	0.783	0.72	0.733	(5, 0, 1)	0.721	0.835	0.816	0.5, 94, 0.004	0.733	0.732	0.692
	Wolenchiti	3-4-1	0.739	0.704	0.702	(3, 0, 2)	0.71	0.851	0.844	0.7, 98, 0.004	0.744	0.709	0.67
	Gelemsso	4-4-1	0.808	0.74	0.689	(3, 0, 0)	0.716	0.852	0.835	0.4, 88, 0.007	0.832	0.739	0.711
	Dire Dawa	3-4-1	0.777	0.707	0.697	(1, 1, 1)	0.711	0.852	0.833	0.6, 93, 0.004	0.8	0.723	0.681
Lower	Eliwuha	4-4-1	0.709	0.732	0.723	(1, 1, 1)	0.701	0.856	0.853	0.6, 94, 0.008	0.731	0.714	0.703
	Dubti	3-5-1	0.706	0.737	0.72	(2, 0, 1)	0.694	0.853	0.838	0.4, 96, 0.006	0.749	0.706	0.695
	Bati	3-4-1	0.73	0.744	0.724	(3, 1, 0)	0.726	0.851	0.849	0.6, 88, 0.007	0.754	0.715	0.702
3 month lead time													
Upper	Bantu Liben	4-4-1	0.5	0.978	0.954	(4, 0, 0)	0.46	1.345	1.314	0.7, 99, 0.006	0.5	0.984	0.871
	Ginchi	5-4-1	0.516	0.949	0.942	(3, 0, 1)	0.454	1.263	1.254	0.8, 84, 0.004	0.475	0.936	0.818
	Sebeta	5-4-1	0.587	0.925	0.908	(3, 1, 2)	0.446	1.313	1.287	0.7, 87, 0.005	0.448	0.995	0.89
	Ejersalele	4-4-1	0.512	0.961	0.957	(1, 0, 0)	0.452	1.293	1.272	0.6, 93, 0.008	0.462	0.981	0.966
	Ziquala	4-4-1	0.462	0.996	0.974	(1, 1, 0)	0.446	1.293	1.265	0.85, 90, 0.007	0.466	0.939	0.927
Middle	Modjo	4-4-1	0.406	0.963	0.947	(2, 1, 0)	0.406	1.316	1.315	0.65, 97, 0.007	0.5	0.949	0.933
	Wolenchiti	4-4-1	0.492	0.961	0.952	(1, 1, 1)	0.453	1.323	1.314	0.55, 85, 0.004	0.45	0.944	0.93
	Gelemsso	5-4-1	0.424	0.994	0.971	(3, 0, 0)	0.414	1.325	1.305	0.6, 90, 0.007	0.494	0.932	0.919
	Dire Dawa	4-4-1	0.435	0.986	0.959	(1, 0, 0)	0.416	1.327	1.325	0.6, 86, 0.01	0.524	0.949	0.933
Lower	Eliwuha	5-4-1	0.48	0.962	0.945	(1, 0, 0)	0.456	1.352	1.349	0.7, 92, 0.01	0.611	0.946	0.936
	Dubti	4-5-1	0.44	0.958	0.94	(2, 0, 2)	0.427	1.314	1.305	0.6, 93, 0.009	0.565	0.945	0.926
	Bati	4-4-1	0.461	0.941	0.903	(3, 0, 1)	0.454	1.354	1.348	0.8, 79, 0.001	0.559	0.942	0.927

Column 3 is the ANN architecture detailing the number of nodes in the input, hidden and output layers respectively

In column 11 the parameters of the SVR models are given

Discussion

As shown in the forecast results for both SPI 3 and SPI 6, the use of wavelet analysis increased forecast accuracy for both 1 and 3 month forecast lead times. This pattern is similarly shown in SPI forecasts within the Awash River Basin for lead times of 1 and 6 months in Belayneh and Adamowski (2012). Once the original SPI time series was decomposed using wavelet analysis it was found that the approximation series of the signal was disproportionally more important for future forecasts compared to the wavelet detail series of the signal. Irrespective of the number of decomposition levels, an absence of the approximation series would result in poor forecast results. Adding the approximation series to the wavelet details did not noticeably improve the forecast results compared to

using the approximation series on its own in most models. Traditionally, the number of wavelet decompositions is either determined via trial and error or using the formula $L = \log [N]$, with N being the number of samples. Using this formula the optimal number of decompositions would be $L = 3$. In this study, the above method was repeated for wavelet decomposition levels 1 through 9 until the appropriate level was determined using the aforementioned performance measures.

In general, WA-ANN and WA-SVR models were the best forecast models in each of the sub-basins. Wavelet-neural networks were also shown to be the best forecast method in forecasting the SPI at 1 and 6 months lead time in Belayneh and Adamowski (2012). Unlike the study by Belayneh and Adamowski (2012), this study also coupled wavelet transforms with SVR models. Coupled WA-SVR

96

Sustain. Water Resour. Manag. (2016) 2:87–101

Table 4 The best WA-ANN and WA-SVR models for 1 and 3 month forecasts of SPI 3

Basin	Station	WA-ANN	R^2	RMSE	MAE	WA-SVR	R^2	RMSE	MAE
1 month lead time									
Upper	Bantu Liben	4-7-1	0.7952	0.43	0.3922	0.8, 98, 0.008	0.7228	0.4461	0.4038
	Ginchi	4-7-1	0.8808	0.4663	0.4598	0.5, 95, 0.007	0.8462	0.4293	0.4188
	Sebeta	4-5-1	0.8722	0.4742	0.4118	0.4, 90, 0.008	0.8029	0.4183	0.3961
	Ejersalele	7-4-1	0.7797	0.47	0.4012	0.5, 95, 0.005	0.8193	0.4829	0.4711
	Ziquala	6-4-1	0.7997	0.4072	0.3918	0.6, 93, 0.004	0.7628	0.483	0.4627
Middle	Modjo	5-5-1	0.8564	0.4975	0.458	0.5, 94, 0.004	0.8398	0.4309	0.4018
	Wolenchiti	4-4-1	0.7632	0.4781	0.4202	0.7, 98, 0.004	0.8429	0.4839	0.4683
	Gelemsso	6-4-1	0.813	0.4409	0.4248	0.4, 88, 0.007	0.8287	0.4389	0.4104
	Dire Dawa	4-5-1	0.7816	0.4893	0.4735	0.6, 93, 0.004	0.8245	0.4738	0.4596
Lower	Eliwuha	5-4-1	0.7723	0.4217	0.3996	0.6, 94, 0.008	0.7911	0.4048	0.3873
	Dubti	7-4-1	0.7495	0.4127	0.4067	0.4, 96, 0.006	0.7238	0.4079	0.3934
	Bati	5-4-1	0.7507	0.4985	0.4875	0.6, 88, 0.007	0.7311	0.4281	0.3852
3 month lead time									
Upper	Bantu Liben	5-7-1	0.5566	0.5098	0.4941	0.7, 99, 0.006	0.5817	0.5012	0.4871
	Ginchi	5-7-1	0.6204	0.5708	0.5143	0.8, 84, 0.004	0.5171	0.5782	0.5638
	Sebeta	5-5-1	0.7301	0.582	0.5345	0.7, 87, 0.005	0.5281	0.5827	0.5739
	Ejersalele	7-4-1	0.7178	0.5843	0.5406	0.6, 93, 0.008	0.5812	0.5922	0.5782
	Ziquala	7-4-1	0.5598	0.5568	0.5343	0.85, 90, 0.007	0.5821	0.5881	0.5781
Middle	Modjo	6-5-1	0.6808	0.5996	0.5604	0.65, 97, 0.007	0.6093	0.5921	0.5829
	Wolenchiti	5-4-1	0.6059	0.5809	0.5738	0.55, 85, 0.004	0.6018	0.5921	0.5829
	Gelemsso	7-4-1	0.5285	0.5498	0.5334	0.6, 90, 0.007	0.5728	0.5721	0.5617
	Dire Dawa	5-5-1	0.6256	0.5966	0.5828	0.6, 86, 0.01	0.6092	0.5921	0.5817
Lower	Eliwuha	6-4-1	0.5792	0.5908	0.5783	0.7, 92, 0.01	0.5056	0.5159	0.5129
	Dubti	7-4-1	0.5621	0.5884	0.5725	0.6, 93, 0.009	0.5051	0.5253	0.519
	Bati	6-4-1	0.563	0.5993	0.5403	0.8, 79, 0.001	0.5915	0.5494	0.5287

Column 3 is the ANN architecture detailing the number of nodes in the input, hidden and output layers respectively

In column 7 the parameters of the SVR models are given

models had improved results compared to SVR models and outperformed WA-ANN using some of the performance measures models at some stations.

While both the WA-ANN and WA-SVR models were effective in forecasting SPI 3, most WA-ANN models had more accurate forecasts. In addition, as shown by Figs. 2 and 3, the forecast from the WA-ANN model seems to be more effective in forecasting the extreme SPI values, whether indicative of severe drought or heavy precipitation. While the WA-SVR model closely mirrors the observed SPI trends, it seems to underestimate the extreme events, especially the extreme drought event at 170 months.

The reason why WA-ANN models seem to be slightly more effective than WA-SVR models for forecasts of SPI 3, and seem to be more effective in forecasting extreme events, is likely due to the inherent effectiveness of ANNs compared to SVR models, such as their simplicity in terms of development and their reduced computation time, as the wavelet analysis used for both machine learning techniques

is the same. This observation is further supported by the fact that most ANN forecasts have better results than SVR models as shown in Table 3. Theoretically, SVR models should perform better than ANN models because they adhere to the structural risk minimization principle instead of the empirical risk minimization principle. They should, in theory, not be as susceptible to local minima or maxima. However, the performance of SVR models is highly dependent of the selection of the appropriate kernel and its three parameters. Given that there are no prior studies on the selection of these parameters for forecasts of the SPI, the selection was done via a trial and error procedure. This process is made even more difficult by the size of the data set (monthly data from 1970 to 2005), which contributes to the long computation time of SVR models. The uncertainty regarding the three SVR parameters increases the number of trials required to obtain the optimal model. Due to the long computational time of SVR models the same amount of trials cannot be done as for ANN models. For ANN models, even in complex systems, the relationship between

Table 5 The best ARIMA, ANN and SVR models for 1 and 3 month forecasts of SPI 6

Basin	Station	ANN	R^2	RMSE	MAE	ARIMA $(p, 0, q)$	R^2	RMSE	MAE	SVR (γ, C, ε)	R^2	RMSE	MAE
1 month lead time													
Upper	Bantu Liben	3-4-1	0.785	0.6903	0.6874	(3, 0, 1)	0.7521	0.8136	0.8024	0.2, 98, 0.02	0.8233	0.6851	0.6706
	Ginchi	4-3-1	0.7321	0.7006	0.6885	(4, 1, 1)	0.7056	0.8064	0.7896	0.1, 94, 0.05	0.776	0.6923	0.6728
	Sebeta	5-4-1	0.8407	0.7058	0.6836	(3, 0, 0)	0.7832	0.8013	0.7853	0.6, 100, 0.001	0.7852	0.6841	0.6742
	Ejersalele	5-4-1	0.8262	0.7134	0.6991	(3, 1, 1)	0.7702	0.7952	0.7714	0.5, 98, 0.007	0.8035	0.6828	0.6837
	Ziquala	3-4-1	0.7469	0.6905	0.6758	(1, 0, 1)	0.7166	0.8164	0.7907	0.8, 98, 0.006	0.8107	0.685	0.6602
Middle	Modjo	3-4-1	0.8014	0.6888	0.6704	(1, 1, 1)	0.7514	0.8218	0.8041	0.8, 92, 0.008	0.7751	0.6864	0.6745
	Wolenchiti	4-3-1	0.7612	0.6769	0.6594	(4, 1, 1)	0.7293	0.8233	0.8057	0.7, 98, 0.004	0.796	0.6765	0.6597
	Gelemsso	4-4-1	0.8122	0.7032	0.6812	(2, 0, 1)	0.7741	0.8335	0.8155	0.3, 90, 0.03	0.8421	0.6966	0.6847
	Dire Dawa	5-4-1	0.8054	0.6841	0.6605	(1, 0, 1)	0.7648	0.8327	0.8164	0.9, 98, 0.01	0.8313	0.6855	0.672
Lower	Eliwuha	5-4-1	0.7322	0.6974	0.6703	(3, 1, 1)	0.7116	0.8162	0.8024	0.6, 88, 0.06	0.7853	0.6956	0.6871
	Dubti	3-5-1	0.7791	0.7046	0.6752	(1, 1, 1)	0.7432	0.8321	0.8026	0.9, 90, 0.005	0.7744	0.7359	0.7226
	Bati	3-4-1	0.7867	0.7142	0.6818	(2, 0, 1)	0.7335	0.8207	0.8063	0.5, 91, 0.07	0.7862	0.7065	0.6807
3 month lead time													
Upper	Bantu Liben	5-4-1	0.5637	0.8826	0.8693	(2, 1, 0)	0.4963	0.8831	0.8749	0.5, 90, 0.02	0.5746	0.8873	0.8751
	Ginchi	4-4-1	0.5347	0.8954	0.8806	(1, 0, 1)	0.5007	0.9366	0.9168	0.6, 91, 0.006	0.553	0.8967	0.8816
	Sebeta	6-4-1	0.5871	0.8924	0.8774	(4, 1, 0)	0.4982	0.9832	0.9632	0.9, 96, 0.004	0.5604	0.9023	0.8864
	Ejersalele	3-5-1	0.577	0.9126	0.9008	(1, 0, 1)	0.5214	0.9763	0.9547	0.5, 91, 0.006	0.6178	0.8865	0.8746
	Ziquala	4-4-1	0.5864	0.9351	0.9162	(2, 1, 2)	0.5284	0.9614	0.9462	0.6, 93, 0.002	0.6182	0.9081	0.8895
Middle	Modjo	3-4-1	0.5205	0.9058	0.8865	(1, 0, 1)	0.4844	0.9741	0.9571	0.05, 100, 0.003	0.6137	0.8943	0.8873
	Wolenchiti	5-4-1	0.5427	0.9125	0.9033	(2, 1, 1)	0.5039	0.9701	0.952	0.08, 95, 0.002	0.5805	0.8868	0.8761
	Gelemsso	4-4-1	0.5534	0.8974	0.8774	(5, 0, 2)	0.513	0.9777	0.9637	0.04, 94, 0.003	0.5814	0.9268	0.9028
	Dire Dawa	3-4-1	0.5493	0.9301	0.9106	(2, 0, 1)	0.4829	0.9862	0.9657	0.05, 85, 0.009	0.5469	0.8962	0.8867
Lower	Eliwuha	5-4-1	0.5341	0.8873	0.8639	(1, 0, 1)	0.4944	0.9869	0.9662	0.09, 90, 0.007	0.6165	0.8716	0.8564
	Dubti	4-4-1	0.5756	0.8792	0.8661	(1, 1, 1)	0.5238	0.9455	0.9212	0.05, 100, 0.005	0.5835	0.9014	0.8961
	Bati	4-3-1	0.5706	0.8864	0.8642	(1, 0, 0)	0.5184	0.9502	0.9384	0.04, 98, 0.005	0.5622	0.8761	0.8576

Column 3 is the ANN architecture detailing the number of nodes in the input, hidden and output layers respectively

In column 11 the parameters of the SVR models are given

Table 6 The best WA-ANN and WA-SVR models for 1 and 3 month forecasts of SPI 6

Basin	Station	WA-ANN	R^2	RMSE	MAE	WA-SVR	R^2	RMSE	MAE
1 month lead time									
Upper	Bantu Liben	5-7-1	0.9064	0.3438	0.3271	0.2, 86, 0.002	0.8872	0.3983	0.3782
	Ginchi	4-4-1	0.8686	0.3483	0.3212	0.08, 90, 0.005	0.8777	0.3478	0.3348
	Sebeta	4-5-1	0.9006	0.4057	0.3826	0.0, 90, 0.001	0.8938	0.3569	0.3464
	Ejersalele	7-4-1	0.8614	0.4035	0.3864	0.05, 98, 0.007	0.8981	0.3871	0.3617
	Ziquala	6-4-1	0.8867	0.3652	0.3551	0.0, 88, 0.008	0.9163	0.3765	0.3647
Middle	Modjo	5-5-1	0.9074	0.3872	0.3731	0.08, 90, 0.08	0.8879	0.389	0.3673
	Wolenchiti	6-4-1	0.9155	0.346	0.3129	0.07, 92, 0.004	0.8983	0.347	0.337
	Gelemsso	6-4-1	0.9252	0.3484	0.312	0.05, 93, 0.003	0.8765	0.3489	0.3324
	Dire Dawa	4-5-1	0.8735	0.3534	0.3432	0.02, 96, 0.007	0.8983	0.3643	0.3547
Lower	Eliwuha	5-4-1	0.894	0.3659	0.3495	0.06, 98, 0.06	0.8606	0.338	0.3228
	Dubti	7-4-1	0.8582	0.3538	0.3353	0.09, 100, 0.008	0.9193	0.3421	0.3311
	Bati	5-4-1	0.8849	0.3836	0.3621	0.06, 91, 0.007	0.8793	0.3785	0.3679
3 month lead time									
Upper	Bantu Liben	5-7-1	0.742	0.4476	0.4351	0.45, 86, 0.02	0.7291	0.4236	0.4201
	Ginchi	5-7-1	0.7036	0.4323	0.3948	0.6, 90, 0.006	0.6938	0.4224	0.3864
	Sebeta	5-5-1	0.7723	0.4378	0.4173	0.9, 90, 0.008	0.6748	0.4453	0.4201
	Ejersalele	7-4-1	0.7307	0.4257	0.4105	0.8, 96, 0.02	0.6981	0.4234	0.4159
	Ziquala	7-4-1	0.7403	0.4921	0.479	0.55, 93, 0.004	0.6838	0.4565	0.4372
Middle	Modjo	6-5-1	0.7414	0.4568	0.4507	0.5, 100, 0.003	0.721	0.433	0.4249
	Wolenchiti	5-4-1	0.7244	0.4556	0.4365	0.8, 95, 0.002	0.7234	0.4304	0.4192
	Gelemsso	7-4-1	0.6829	0.448	0.4371	0.35, 94, 0.03	0.7123	0.443	0.4232
	Dire Dawa	6-5-1	0.6655	0.4851	0.4793	0.75, 95, 0.09	0.7241	0.4049	0.3897
Lower	Eliwuha	6-4-1	0.7126	0.4372	0.4149	0.9, 90, 0.007	0.7483	0.4378	0.3283
	Dubti	7-4-1	0.6997	0.4204	0.4035	0.3, 85, 0.009	0.6599	0.4293	0.3158
	Bati	6-4-1	0.6712	0.46	0.4487	0.5, 100, 0.005	0.6418	0.4357	0.4175

Column 3 is the ANN architecture detailing the number of nodes in the input, hidden and output layers respectively

In column 7 the parameters of the SVR models are given

**Fig. 2** SPI 3 forecast results for the best WA-ANN model at the Bati station for 1 month lead time
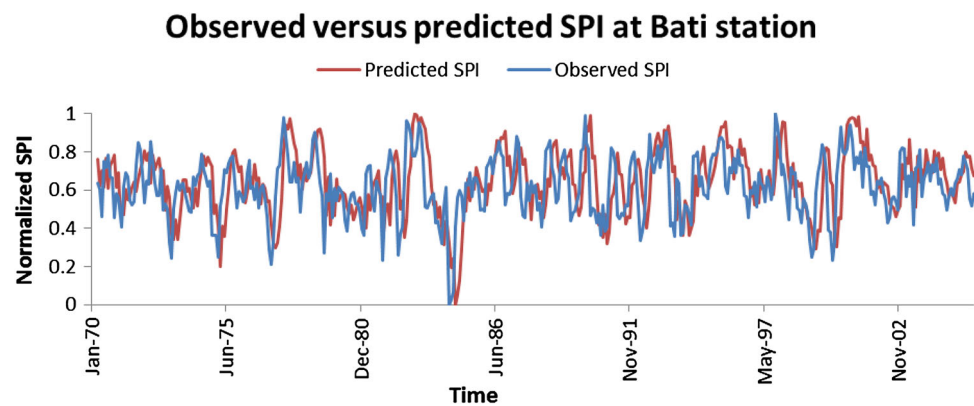
input and output variables does not need to be fully understood. Effective models can be determined by varying the number of neurons within the hidden layer. Producing several models with varying architectures is not computationally intensive and allows for a larger selection pool for the optimal model. In addition, the ability of wavelet analysis to effectively forecast local discontinuities likely reduces the susceptibility in ANN models when they are coupled.

This study also shows that the à trous algorithm is an effective tool for forecasting SPI time series. The à trous algorithm de-noises a given time series and improves the

Fig. 3 SPI 3 forecast results for the best WA-SVR model at the Bati Station for 1 month lead time
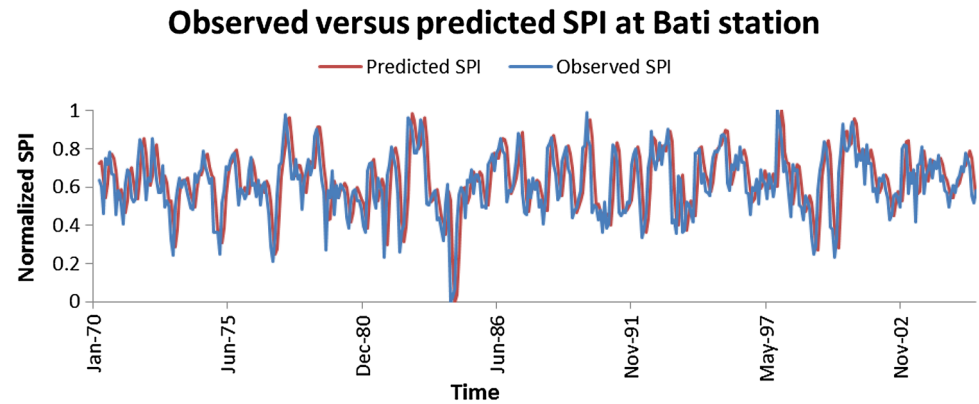
performances of both ANN and SVR models. The à trous algorithm is shift invariant, making it more applicable for forecasting studies, which includes drought forecasting. The fact that wavelet based models had the best results is likely due to the fact that wavelet decomposition was able to capture non-stationary features of the data.

Conclusion

This study explored forecasting short-term drought conditions using five different data driven models in the Awash River basin, including newly proposed methods based on SVR and WA-SVR. With respect to wavelet analysis, this study found, for the first time, that the use of only the approximation series was effective in de-noising a given SPI time series. SPI 1 and SPI 3 were forecast over lead times of 1 and 3 months using ARIMA, ANN, SVR, WA-SVR and WA-ANN models. Forecast results for SPI 1 were low in terms of the coefficient of determination, likely a result of the low levels of autocorrelation of the data sets compared to SPI 3. Overall, the WA-ANN method, with a new method for determining the optimal number of neurons within the hidden layer, had the best forecast results with WA-SVR models also having very good results. Wavelet coupled models consistently showed lower values of RMSE and MAE compared to the other data driven models, possibly because wavelet decomposition de-noises a given time series subsequently allowing either ANN or SVR models to forecast the main signal rather than the main signal with noise.

Two of the three Awash River Sub-basins have semi-arid climates. The effectiveness of the WA-ANN and WA-SVR models indicate that these models may be effective forecast tools in semi-arid regions. Studies should also focus on different regions and try to compare the effectiveness of data driven methods in forecasting different drought indices. The forecasts did not show a particular trend with respect to a particular sub-basin. The climatology of a given sub-basin did not significantly affect the forecast results for any particular station. This study has not found a clear link between a particular sub-basin and performance indicating the need for further studies in different climates to determine whether there is a significant link between forecast accuracy and climate. The coupling of these data driven models with uncertainty analysis techniques such as bootstrapping should be investigated. In addition, coupling SVR models with genetic algorithms to make parameter estimation more efficient could be explored.

References

Adamowski J (2008) Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. J Hydrol 353:247–266

Adamowski J, Chan HF (2011) A wavelet neural network conjunction model for groundwater level forecasting. J Hydrol 407:28–40

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 390:85–91

Adamowski K, Prokoph A, Adamowski J (2009) Development of a new method of wavelet aided trend detection and estimation. Hydrol Process 23:2686–2696

Adamowski J, Adamowski K, Bougadis J (2010) Influence of trend on short duration design storms. Water Resour Manage 24:401–413

Adamowski J, Chan H, Prasher S, Sharda VN (2012) Comparison of multivariate adaptive regression splines with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds with limited data. J Hydroinform 3:731–744

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000a) Artificial neural networks in hydrology. I. Preliminary concepts. J Hydrol Eng 5(2):124–137

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000b) Artificial neural networks in hydrology. II. Hydrologic applications. Hydrol Eng 5(2):115–123

Asefa T, Kemblowski M, McKee M, Khalil A (2006) Multi-time scale stream flow 505 predicitons: the support vector machines approach. J Hydrol 318(1–4):7–16

Bacanli UG, Firat M, Dikbas F (2008) Adaptive neuro-fuzzy inference system for drought forecasting. Stoch Env Res Risk Assess 23(8):1143–1154

Barros A, Bowden G (2008) Toward long-lead operational forecasts of drought: an experimental study in the Murray-Darling River Basin. J Hydrol 357(3–4):349–367

Belayneh A, Adamowski J (2012) Standard Precipitation Index drought forecasting using neural networks, wavelet neural networks, and support vector regression. Appl Comput Intell Soft Comput 2012:794061. doi:10.1155/2012/794061

Belayneh A, Adamowski J, Khalil B, Ozga-Zielinski B (2014) Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet-support vector regression models. J Hydrol 508:418–429

Box GEP, Jenkins GM (1976) Time series analysis. Forecasting and control. Holden-Day, San Francisco

Box GEP, Jenkins GM, Reinsel GC (1994) Time series analysis. Forecasting and control. Prentice Hall, Englewood Cliffs

Butler C, Adamowski J (2015) Empowering marginalized communities in water resources management: addressing inequitable practices in participatory model building. J Environ Manage 153:153–162

Cacciamani C, Morgillo A, Marchesi S, Pavan V (2007) Monitoring and forecasting drought on a regional scale: Emilia-Romagna region. Water Sci Technol Libr 62(1):29–48

Campisi S, Adamowski J, Oron G (2012) Forecasting urban water demand via wavelet- denoising and neural network models. Case study: city of Syracuse, Italy. Water Resour Manag 26:3539–3558

Cancelliere A, Di Mauro G, Bonaccorso B, Rossi G (2007) Stochastic forecasting of drought indices. In: Rossi G, Vega T, Bonaccorso B (eds) Methods and tools for drought analysis and management. Springer, Netherlands

Cannas B, Fanni A, Sias G, Tronci S, Zedda MK (2006) River flow forecasting using neural networks and wavelet analysis. In: Proceedings of the European Geosciences Union

Cimen M (2008) Estimation of daily suspended sediments using support vector machines. Hydrol Sci J 53(3):656–666

Coulibaly P, Anctil F, Bobee B (2000) Daily reservoir inflowvforecasting using artificial neural networks with stopped trainingvapproach. J Hydrol 230:244–257

Cutore P, Di Mauro G, Cancelliere A (2009) Forecasting palmer index using neural networks and climatic indexes. J Hydrol Eng 14:588–595

Edossa DC, Babel MS, Gupta AD (2010) Drought Analysis on the Awash River Basin, Ethiopia. Water Resour Manag 24:1441–1460

Guttman NB (1999) Accepting the standardized precipitation index: a calculation algorithm. J Am Water Resour Assoc 35(2):311–322

Haidary A, Amiri BJ, Adamowski J, Fohrer N, Nakane K (2013) Assessing the impacts of four land use types on the water quality of wetlands in Japan. Water Resour Manage 27:2217–2229

Halbe J, Pahl-Wostl C, Sendzimir J, Adamowski J (2013) Towards adaptive and integrated management paradigms to meet the challenges of water governance. Water Sci Technol Water Supply 67:2651–2660

Han P, Wang P, Zhang S, Zhu D (2010) Drought forecasting with vegetation temperature condition index. Wuhan Daxue Xuebao (Xinxi Kexue Ban)/Geomat Inf Sci Wuhan Univ 35(10):1202–1206 + 1259

Husak GJ, Funk CC, Michaelsen J, Magadzire T, Goldsberry KP (2013) Developing seasonal rainfall scenarios for food security early warning. Theor Appl Climatol. doi:10.1007/s00704-013-0838-8

Inam A, Adamowski J, Halbe J, Prasher S (2015) Using causal loop diagrams for the initialization of stakeholder engagement in soil salinity management in agricultural watersheds in developing countries: a case study in the Rechna Doab watershed, Pakistan. J Environ Manage 152:251–267

Karamouz M, Rasouli K, Nazil S (2009) Development of a hybrid index for drought prediction: case study. J Hydrol Eng 14:617–627

Keshavarz M, Karami E, Vanclay F (2013) The social experience of drought in rural Iran. J Land Use Policy 30:120–129

Khan MS, Coulibaly P (2006) Application of support vector machine in lake water level prediction. J Hydrol Eng 11(3):199–205

Kim T, Valdes JB (2003) Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. J Hydrol Eng 8:319–328

Kisi O, Cimen M (2009) Evapotranspiration modelling using support vector machines. Hydrol Sci J 54(5):918–928

Kisi O, Cimen M (2011) A wavelet-support vector machine conjunction model for monthly streamflow forecasting. J Hydrol 399:132–140

Kolinjivadi V, Adamowski J, Kosoy N (2014a) Juggling multiple dimensions in a complex socioecosystem: the issue of targeting in payments for ecosystem services. Geoforum 58:1–13

Kolinjivadi V, Adamowski J, Kosoy N (2014b) Recasting payments for ecosystem services (PES) in water resource management: a novel institutional approach. Ecosyst Serv 10:144–154

Labat D, Ababou R, Mangin A (1999) Wavelet analysis in Karstic hydrology 2nd Part: rainfall–runoff cross-wavelet analysis. Comptes Rendus de l'Academie des Sci Ser IIA Earth Planet Sci 329:881–887

Lane SN (2007) Assessment of rainfall–runoff models based upon wavelet analysis. Hydrol Process 21:586–607

Linsley RK, Kohler MA, Paulhus JLH (1988) Hydrology for engineers. International edition. McGraw-Hill, Singapore

Maity R, Bhagwat PP, Bhatnagar A (2010) Potential of support vector regression for prediction of monthly streamflow using endogenous property. Hydrol Process 24:917–923

Mallat SG (1998) A wavelet tour of signal processing. Academic, San Diego, p 577

Marj AF, Meijerink AM (2011) Agricultural drought forecasting using satellite images, climate indices and artificial neural network. Int J Remote Sens 32(24):9707–9719

McKee TB, Doesken NJ, Kleist J (1993) The relationship of drought frequency and duration to time scales. In: Paper presented at 8th conference on applied climatology. American Meteorological Society, Anaheim

Mishra AK, Desai VR (2005) Drought forecasting using stochastic models. Stoch Env Res Risk Assess 19(5):326–339

Mishra AK, Desai VR (2006) Drought forecasting using feed-forward recursive neural network. Ecol Model 198(1–2):127–138

Mishra AK, Singh VP (2010) A review of drought concepts. J Hydrol 391(1–2):202–216

Mishra AK, Singh VP (2012) Simulating hydrological drought properties at different spatial units in the United States based on wavelet-bayesian approach. Earth Interact 16(17):1–23

Mishra AK, Desai VR, Singh VP (2007) Drought forecasting using a hybrid stochastic and neural network model. J Hydrol Eng 12:626–638

Morid S, Smakhtin V, Bagherzadeh K (2007) Drought forecasting using artificial neural networks and time series of drought indices. Int J Climatol 27(15):2103–2111

Murtagh F, Starcj JL, Renaud O (2003) On neuro-wavelet modeling. Decis Support Syst 37:475–484

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

Nalley D, Adamowski J, Khalil B, Ozga-Zielinski B (2013) Trend detection in surface air temperature in Ontario and Quebec, Canada during 1967–2006 using the discrete wavelet transform. J Atmos Res 132(133):375–398

Nason GP, Von Sachs R (1999) Wavelets in time-series analysis. Philos Trans R Soc A Math Phys Eng Sci 357(1760):2511–2526

Nourani V, Baghanam A, Adamowski J, Kisi O (2014) Applications of hybrid wavelet-artificial intelligence models in hydrology: a review. J Hydrol 514:358–377

Ntale HK, Gan TY (2003) Drought indices and their application to East Africa. Int J Climatol 23(11):1335–1357

Ozger M, Mishra AK, Singh VP (2012) Long lead time drought forecasting using a wavelet and fuzzy logic combination model: a case study in Texas. J Hydrometeorol 13(1):284–297

Parrella F (2007) Online support vector regression. Master Thesis, University of Genoa

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

Paulo A, Pereira L (2008) Stochastic predictions of drought class transitions. Water Resour Manag 22(9):1277–1296

Paulo A, Ferreira E, Coelho C, Pereira L (2005) Drought class transition analysis through Markov and Loglinear models, an approach to early warning. Agric Water Manag 77(1–3):59–81

Pingale S, Khare D, Jat M, Adamowski J (2014) Spatial and temporal trends of mean and extreme rainfall and temperature for the 33 urban centres of the arid and semi-arid state of Rajasthan, India. J Atmos Res 138:73–90

Rathinasamy M, Khosa R, Adamowski J, Ch S, Partheepan G, Anand J, Narsimlu B (2014) Wavelet-based multiscale performance analysis: an approach to assess and improve hydrological models. Water Resour Res 50(12):9721–9737. doi:10.1002/2013WR014650

Renaud O, Starck J, Murtagh F (2002) Wavelet-based forecasting of short and long memory time series. Department of Economics, University of Geneva

Rossi G, Castiglione L, Bonaccorso B (2007) Guidelines for planning and implementing drought mitigation measures. In: Rossi G, Vega T, Bonaccorso B (eds) Methods and tools for drought analysis and management. Springer, Netherlands, pp 325–247

Saadat H, Adamowski J, Bonnell R, Sharifi F, Namdar M, Ale-Ebrahim S (2011) Land use and land cover classification over a large area in Iran based on single date analysis of satellite imagery. J Photogramm Remote Sens 66:608–619

Saco P, Kumar P (2000) Coherent modes in multiscale variability of streamflow over the United States. Water Resour Res 36(4):1049–1067

Straith D, Adamowski J, Reilly K (2014) Exploring the attributes, strategies and contextual knowledge of champions of change in the Canadian water sector. Can Water Resour J 39(3):255–269

Tiwari M, Adamowski J (2014) Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. Water Resour Res 49(10):6486–6507

Tiwari MK, Chatterjee C (2010) Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach. J Hydrol 394:458–470

Tsakiris G, Vangelis H (2004) Towards a drought watch system based on spatial SPI. Water Resour Manage 18(1):1–12

Vapnik V (1995) The nature of statistical learning theory. Springer, New York

Wanas N, Auda G, Kamel MS, Karray F (1998) On the optimal number of hidden nodes in a neural network. In: Proceedings of the IEEE Canadian conference on electrical and computer engineering, pp 918–921

Wang WC, Chau KW, Cheng CT, Qiu L (2009) A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. J Hydrol 374(3–4):294–306

Zhang GP (2003) Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50:159–175