

A NONLINEAR MODEL FOR DROUGHT FORECASTING BASED ON CONJUNCTION OF WAVELET TRANSFORMS AND NEURAL NETWORKS

Tae-Woong Kim¹ and Juan B. Valdés²

File: Fcst_Asce_Final.doc (05/14/2003)

Accepted for Publication by ASCE, *Journal of Hydrologic Engineering*

¹ Graduate Research Assistant, Department of Civil Engineering and Engineering Mechanics, and Center for Sustainability of Semi-Arid Hydrology and Riparian Areas (SAHRA), The University of Arizona, Tucson, Arizona, 85721-0072 (email: taek@email.arizona.edu)

² Professor and Head, Department of Civil Engineering and Engineering Mechanics, and Center for Sustainability of Semi-Arid Hydrology and Riparian Areas (SAHRA), The University of Arizona, Tucson, Arizona, 85721-0072 (email: jvaldes@u.arizona.edu)

Key Words: drought forecasting, conjunction, wavelet transforms, neural networks, Conchos River Basin, PDSI.

ABSTRACT: Droughts are destructive climatic extreme events, which may cause significant damages both in natural environments and human lives. Drought forecasting plays an important role in the control and management of water resources systems. In this study, a conjunction model is presented to forecast droughts. The proposed conjunction model is based on dyadic wavelet transforms and neural networks. Neural networks have shown great ability in modeling and forecasting nonlinear and nonstationary time series in a water resources engineering, and wavelet transforms provide useful decompositions of an original time series. The wavelet-transformed data aids in improving the model performance by capturing helpful information on various resolution levels. Neural networks were used to forecast decomposed sub-signals in various resolution levels and reconstruct forecasted sub-signals. The performance of the conjunction model was measured using various forecast skill criteria. The model was applied to forecast droughts in the Conchos River Basin in Mexico, which is the most important tributary of the Lower Rio Grande/Bravo. The results indicate that the conjunction model significantly improves the ability of neural networks for forecasting the indexed regional drought.

INTRODUCTION

A drought is generally defined an extreme deficiency of water available in the hydrologic cycle over an extended period of time. Droughts are more destructive than other climatic extreme events such as earthquakes, tornadoes, hurricanes, and floods, which cause significant damages both in natural environments and human lives. Droughts are the most costly natural disaster in the United States (Riebsame et al. 1991). The losses in the United States average \$6-8 billion for severe droughts that extended for 12% or more of the United States on average every year. In recent years, droughts in the United States caused damages as high as \$39 billion for the 1987-

1989 three-year drought (NOAA Paleoclimatology Program 2000). In Florida, \$500 million in damages were the consequence of extensive drought-induced fires, which burned over 475,000 acres during the 1998 drought. Texas lost \$4 billion in direct income with the total impact to the state's economy close to \$11 billion due to the 1996 and 1998 droughts. In Western Canada, continuing uncertainty in drought prediction contributes to crop insurance payouts of over \$175 million per year, and in Mexico, droughts occurred in recent decades have produced extensive damages to crops and cattle in the northwestern states (National Drought Policy Commission 2000; Texas Center for Policy Studies 2001). Woodhouse and Overpeck (1998) reviewed a wide range of the paleoclimatic literature including a variety of data sources like tree-ring data and instrumental records, and suggested that droughts more severe than those of the 1930s and 1950s, which had the most severe impact on the continental United States, were likely to occur in the future.

Hydrologic drought forecasting plays an important role in the mitigation of impacts of drought on water resources systems. Traditionally, statistical models have been used for hydrologic drought forecasting based on time series methods. Simple/multiple regression and autoregressive moving average (ARMA) models are typical models for statistical time series methods for forecasting. However, they are basically linear models assuming that data are stationary, and have a limited ability to capture nonstationarities and nonlinearities in data. Hydrologic variables of interest such as annual and monthly streamflows and precipitation have been extensively modeled by ARMA models, which have been generally accepted by practitioners during the past several decades. However, it is necessary for hydrologists to consider alternative models when nonlinearity and nonstationarity play a significant role in the forecasting.

In recent decades, artificial neural networks (ANNs) have shown great ability in modeling and forecasting nonlinear and nonstationary time series in hydrology and water resources engineering due to their innate nonlinear property and flexibility for modeling (ASCE Task Committee on Application of Artificial Neural Networks in Hydrology 2000). ANNs are nonlinear computational frameworks based on the massive interaction with neurons, whose components have direct analogs to components of an actual human neuron (Tang and Fishwick 1993; Tsoukalas and Uhrig 1996). In general, the advantages of ANNs over other statistical models are that (1) the application of ANNs does not require a prior knowledge of the process because ANNs have black-box properties, (2) ANNs have the inherent property of nonlinearity since neurons activate a nonlinear filter called an activation function, (3) ANNs can have multiple inputs having different characteristics, which can make ANNs able to represent the time-space variability, and (4) ANNs have the adaptability to represent change of problem environments.

However, the training process of ANNs requires significant amounts of data so that the patterns embedded in the system are discovered (Zheng et al. 2000). In addition, Hsu et al. (1995) and Gupta et al. (2000) indicated that the objective function surface of ANNs is typically nonconvex with extensive regions that are insensitive to the variations in the values of the network weights, contains numerous multilocal optima, and is generally of very high dimensions, which make ANNs easily trapped by local optima. They proposed the linear least square simplex (LLSSIM) method for network training by partitioning the weight space into its linear and non-linear components, which provides globally optimal weight estimates with amounts of computation time.

In statistical time series forecasting, decomposition approaches seek to decompose a time series into its major subcomponents. Forecasting using a decomposition method is often more useful in providing forecasts and information regarding the component of a time series than trying to predict a single pattern (Makridakis et al. 1983). Recently wavelet transforms have become a common tool for analyzing local variation in time series (Torrence and Compo 1998), and hybrid models have been proposed for forecasting a time series based on a wavelet transform preprocessing (Aussem and Murtagh 1997; Aussem et al. 1998; Zheng et al. 2000; Zhang and Dong 2001). Wavelet transforms provide useful decompositions of original time series, so that wavelet-transformed data improves the ability of a forecasting model by capturing useful information on various resolution levels. Aussem and Murtagh (1997) and Aussem et al. (1998) improved neural network prediction accuracy using a dynamic recurrent neural network (DRNN) and the “à trous” wavelet transform. Zheng et al. (2000) presented a wavelet transform method for load forecasting based on a decomposition scheme of multi-resolution analysis (MRA). They indicated that the wavelet coefficients are modeled as state variables of Kalman filters. Zhang and Dong (2001) proposed a short-term local forecast model based on neural networks and the multi-resolution wavelet decomposed by autocorrelation shell representation.

In this study, a conjunction model is presented in order to improve forecast accuracy for regional droughts. The conjunction model is a hybrid neural network model combined with dyadic wavelet transforms. Improved forecasts allow water resources decision makers to develop drought preparedness plans far in advance to mitigate social, environmental, and economic costs. The “à trous” algorithm for the dyadic wavelet transform is modified so that future information is not used when decomposing and reconstructing. Neural networks are used to forecast decomposed signals in various resolution levels and reconstruct forecasted decomposed signals

into the original time series. The performance of the model is measured using several forecast skill criteria and compared with simpler models for drought prediction. The model was applied to the Conchos River Basin in Mexico.

DROUGHT ISSUES IN THE CONCHOS RIVER BASIN

The Conchos River is the main tributary of the Lower Bravo/Grande River. The Conchos River Basin, shown in Fig. 1, is totally located in the arid/semiarid area of the northern Mexico. The Conchos River Basin has had significant economic and population growth in the last few decades. The waters of the Conchos River are used primarily for irrigation and generation of the hydroelectric power in the basin. It also supplies municipal, industrial, and agricultural water for Mexican cities such as Chihuahua, Hidalgo del Parral, and Delicias. In addition, the Conchos River is part of an international treaty between the United States and Mexico, which supplies approximately 70-80% of the flow of the Lower Bravo/Grande River above the mainstream binational reservoirs of Amistad and Falcon.

The Conchos River Basin has a limited water supply and suffers from periodic droughts (Kim et al. 2002; Schmandt 2002). Droughts in the last decade in northern Mexico have reduced the Conchos River inflow into the Bravo/Grande River as shown in Fig. 2, which resulted from the lower than normal rainfalls and high surface temperatures in the basin. For example, the 1990s rainfall, shown in Fig. 3, has been significantly below the normal precipitation (long-term mean of annual precipitation during 1955-2000) not only the magnitude but also the timing of precipitation. In 1994, the annual precipitation reached only 48% of normal and less than 11% of the annual precipitation fell in the first five months (Jan-May). During the period of 1992-1997, due to the severe reduction of the Conchos River inflows into the Bravo/Grande River, Mexico

owes the United States approximately $1,240\text{Mm}^3$ (1.024 M acre-ft) with respect to water delivery under the 1944 U.S./Mexico Treaty on the use of the Bravo/Grande River (Texas Center for Policy Studies 2001). The treaty specifies that five-year average flows to be delivered by Mexico are approximately $430\text{Mm}^3/\text{year}$ (350,000 acre-ft/year) (Schmandt 2002). This treaty also allows the flows to be below the minimum in a five-year average under conditions of “extreme drought”. However, the concept of “extreme drought” is not explicitly defined in the treaty. At the close of the third year of the five-year cycle from October 1997 to September 2000, Mexico had delivered a total of 490Mm^3 (0.407 M acre-ft) and total deficit became $1,670\text{Mm}^3$ (1.381 M acre-ft) (International Boundary and Water Commission 2002). The reduction of the Conchos River flows during droughts is a controversial issue, which arose dispute over the water delivery under the treaty.

ARTIFICIAL NEURAL NETWORKS

Multilayered Feed Forward Neural Networks

Artificial neural networks (ANNs) are massively paralleled-distributed data processing systems consisting of a large number of highly interconnected artificial neurons with performance characteristics resembling biological neural networks of the human brain (Haykin 1994; Tsoukala and Uhrig 1996). Due to the advantages of ANNs in modeling they have become extremely popular for the prediction and forecasting of water resources variables.

As shown in Fig. 4, three-layered feed forward neural networks (FFNNs), which have been usually used in forecasting hydrologic time series, provide a general framework for representing nonlinear functional mapping between a set of input and output variables. Three-layered FFNNs

are based on a linear combination of the input variables, which are transformed by a nonlinear activation function. The explicit expression for an output value of FFNNs is given by

$$\hat{y}_k = f_o \left[\sum_{j=1}^M w_{kj} \cdot f_h \left(\sum_{i=1}^N w_{ji} x_i + w_{jo} \right) + w_{ko} \right] \quad (1)$$

where w_{ji} is a weight in the hidden layer connecting the i th neuron in the input layer and the j th neuron in the hidden layer, w_{jo} is the bias for the j th hidden neuron, f_h is the activation function of the hidden neuron, w_{kj} is a weight in the output layer connecting the j th neuron in the hidden layer and the k th neuron in the output layer, w_{ko} is the bias for the k th output neuron, and f_o is the activation function for the output neuron. The weights are different in the hidden and output layer, and their values can be changed during the process of network training.

A Backpropagation Training Algorithm for Three-Layered Neural Networks

Because there are no physical rules between inputs and outputs in designing ANNs, the relationship of the available input variables and output variables is generated by the training process. In this study, the process of training ANNs is accomplished by a backpropagation algorithm, as shown in Fig. 4, which has been applied successfully to solve difficult and diverse problems. This algorithm is based on the error-correction learning rule. Basically, the error-propagation process consists of two passes through the different layers of the network as shown in Fig. 4. In the forward pass, an input vector is applied to the neurons of the network, and its effect propagates through the network layer by layer. A set of output is produced as the actual response of the network. During the backward pass, on the other hand, the weights are all adjusted in accordance with the error-correction rule. The error signal is then propagated

backward through the network. The weights are adjusted so as to make the actual response (\hat{y}_k) of the network closer to the desired response (y_k).

The objective of the backpropagation training process is to adjust the weights of the network to minimize the sum of square errors of the network in Eq. (2), which approximates the model outputs to the target values with a selected error goal.

$$E(n) = \frac{1}{2} \sum_{k=1}^K [y_k(n) - \hat{y}_k(n)]^2 \quad (2)$$

where $y_k(n)$ is the desired target responses and $\hat{y}_k(n)$ is the actual response of the network for the k th neuron at the n th iteration.

To minimize $E(n)$, the gradient-descent method and the derivative chain rule are employed to modify the network weights (Haykin 1994). The pure backpropagation is rarely used to solve practical problems. Momentum allows a network to response not only to the local gradient, but also to recent trends in the error surface. Without momentum a network may get stuck in a local minimum, and with momentum a network can slide through such a minimum. An adaptive learning rate is one of the methods of implement actions using a momentum. At each epoch new weights and biases are calculated using the current learning rate and new output and error are then calculated (Demuth and Beale 1994). It is noted that the momentum must be less than 1.0 for convergence (Dai and Macbeth 1997), and the optimal learning rates are determined by trial and error (Maier and Dandy 2000).

WAVELET TRANSFORMS

Fourier Transforms and Wavelet Transforms

In the last two decades, wavelet transforms have been extensively tested in many fields such as signal processing, image processing, communications, computer science, and mathematics (Rao and Bopardika 1998). Grossmann and Morlet (1984) studied wavelet transforms motivated by the fact that certain seismic signals can be modeled suitably by combining translations and dilations of a simple oscillatory function called a wavelet.

Fourier transforms can also break down a stationary signal into continuous sinusoids of different frequencies. In transforming to the frequency domain, time information is lost. When looking at a Fourier transform of a signal, it is impossible to tell when a particular event took place (Misiti et al. 2000). Gabor (1946) introduced a local Fourier transform to analyze a small section of the signal at a time taking into a sliding window. The Gabor's short-time Fourier transform (STFT) maps a signal into a two-dimensional function of time and frequency, which can provide some information about both when and at what frequencies a signal event occurs. However, this method is only applicable to situation where the coherent time is independent of the frequency, and the limited precision is determined by the size of the window (European Southern Observatory 1998; Misiti et al. 2000).

Wavelets are mathematical functions that give a time-scale representation of the time series and their relationships to analyze time series that contain nonstationarities. Wavelet analysis allows the use of long time intervals for low frequency information and shorter intervals for high frequency information. Wavelet analysis is capable of revealing aspects of data like trends, breakdown points, and discontinuities that other signal analysis techniques might miss. Furthermore, it can often compress or de-noise a signal.

Dyadic Wavelet Transforms

The basic objective of wavelet transforms is to achieve a complete time-scale representation of localized and transient phenomena occurring at different time scale (Labat et al. 2000). The continuous wavelet transform (CWT) is defined as the sum over all time of the signal multiple by scale and shifted versions of wavelet function, ψ ,

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(x) \Psi^* \left(\frac{x-b}{a} \right) dx \quad (3)$$

where a is the scale parameter, b is the position parameter, and $*$ corresponds to the complex conjugate. Several families of wavelets (Ψ) that have proven to be useful for various applications are described in related references (Mallat 1998; Rao and Bopardika 1998; Misiti et al. 2000).

The coefficient plots of the continuous wavelet transform are precisely the time-scale view of the signal. However, calculating wavelet coefficients at every possible scale is time-consuming and generates large amount of information. Then, it makes the use of the continuous wavelet transform for forecasting not practically possible.

Discrete dyadic wavelet transforms are scale samples of the wavelet transform following a geometric sequence of ratio 2 (Mallat 1998). The scale is sampled along a dyadic sequence to simplify the numerical calculations. A discrete dyadic wavelet transform can be computed with a fast filter bank algorithm called the “à trous” algorithm. In this algorithm, the distance between samples increasing by a factor 2 from the scale ($i-1$) to the next one, $c_i(k)$ is given by

$$c_i(k) = \sum_{l=-\infty}^{+\infty} h(l) c_{i-1}(k + 2^{i-1}l) \quad (4)$$

where $c_0(t) = x(t)$, h is the low pass filter, and the distance wavelet transform, $w_i(k)$, is $w_i(k) = c_{i-1}(k) - c_i(k)$. The coefficients, $h(k)$, are derived from the scaling function $\phi(x)$ given by

$$\frac{1}{2}\psi\left(\frac{x}{2}\right) = \phi(x) - \frac{1}{2}\phi\left(\frac{x}{2}\right) \quad (5)$$

where the sampled data, $c_0(k)$, are assumed to be scalar products at pixel k of function $f(x)$ with a scaling function $\phi(x)$, which corresponds to a low pass filter (European Southern Observatory 1998) associated with a wavelet $\psi(x)$.

A CONJUNCTION MODEL FOR FORECASTING

The Palmer Drought Severity Index in the Conchos River Basin

A conjunction model was developed to forecast regional droughts in this study. The Palmer Drought Severity Index (PDSI) is used to represent a regional drought, which measures how much precipitation for a given time period has deviated from normal. In the United States, the PDSI is used by the U.S. Department of Agriculture to determine when to grant emergency drought assistance. The original report by Palmer (1965) and related papers describe the detail procedure to calculate the PDSI (Alley 1984; Kim et al. 2002).

The PDSI has positive characteristics useful for the drought monitoring. For example, the PDSI provides a measurement of the abnormality of weather for a region and an opportunity to place current conditions in historical perspective. In addition, the PDSI is a standardized value available to compare and assess regional drought. However, the PDSI is sensitive to the available water capacity (AWC) of a soil type, which is one of limitations to consider when using the PDSI (Alley 1984). Due to the lack of specific soil type information in the Conchos River Basin, it was assumed that the surface soil layer water capacity is one inch, and the lower soil layer

water capacity is six inches in this study. These values were adopted from the National Climatic Data Center (NCDC, <http://lwf.ncdc.noaa.gov/oa/ncdc.html>). The NCDC uses these values for the AWC in the Texas climatologic region 5, where is the immediate vicinity of the Conchos River Basin.

The regional respective of the PDSI was calculated using the areal mean of precipitation and temperature in the basin. The spatial and recurrence characteristics of the PDSI in the Conchos River Basin have been examined by Kim et al. (2002, 2003).

The Conjunction Model and Forecasting System

The aim of the conjunction model is to predict the t -months ahead PDSI given the current and previous PDSI, $x(k)$, $k = 1, 2, \dots, n$. The schematic representation for the proposed conjunction model is shown in Fig. 5. The “à trous” algorithm for the dyadic wavelet transform, which has been used for time series forecasts (Aussem and Murtagh 1997; Aussem et al. 1998; Zhang and Dong 2001), performs successive convolutions with the discrete low pass filter as shown in Eq. (4). In order to decompose the observed PDSI, the “à trous” wavelet transform with the Mallat’s quadratic spline as a low pass filter was used in this study. Daubechies and Morlet wavelet transforms have been widely introduced for signal and image processing. However, it is reported that they have weaknesses for prediction, since Daubechies wavelets have many different identical events across a given time series, and Morlet wavelets generate more inputs for the model (Aussem et al. 1998). The dyadic wavelet transforms have a more consistent response than Daubechies and Morlet wavelets by taking the mirror operation in the boundary region (Aussem and Murtagh 1997; Aussem et al. 1998). However, the inconsistency of the decomposed sub-signal still remains problematic in a forecasting model. We took an alternative

approach to this issue by taking a convolution value in the “à trous” algorithm fixed to the beginning and the end of the signal, not symmetric like the original dyadic wavelet transform.

Clearly, in forecasting studies, very careful attention must be given to the boundaries of the signal during decomposing and reconstructing in order not to include the future information. In the proposed model, wavelet transforms were performed twice to produce input (step 1) and target (step 2) values of ANNs in a training phase. This is the initial step in the forecasting model, which uses a transforming preprocess in order not to use the future information in the input data. In the forecasting phase, after training the network, ANNs predict sub-signals with a specific lead time. The forecasted decomposed sub-signal (\hat{c}_i) can then be reconstructed by inverting the dyadic wavelet transform (Mallat 1998). The model uses previous values, $x(t)$, $x(t-1)$, $x(t-2)$, ..., and forecasts a value leaped over a lead time, $x(t+m)$. It is difficult to reconstruct the predicted time series leaped over a lead time, because successive convolutions in reconstruction wavelets cannot be carried out between $x(t)$ and $x(t+m)$. In the reconstructing phase, we used ANNs one more time for reconstructing predicted sub-signals from forecasting ANNs. The ANNs used in the reconstructing phase have different architecture and weights from the ANNs used in the training and forecasting phase.

Design of Artificial Neural Networks

The ANNs used in this study are three-layered FFNNs, as shown in Fig. 4, which are typically used in water resources engineering. It has been shown that ANNs with one hidden layer can approximate any function, given that sufficient degrees of freedom are provided (Maier and Dandy 2000). Although the optimal number of hidden neurons is highly dependent on the problem of interest and a matter of experimentation, the architecture with a bottleneck structure,

which has fewer neurons in the hidden layer than in the input layer, has worked well in the past (Fletcher and Goss 1993; Maier and Dandy 2000; Zhang and Dong 2001). In order to determine the optimal network architecture, the number of neurons in the input and hidden layer were determined by experimentation. In the training/forecasting phase (step 3-5), 11 combinations of input neurons (1, 0.5*m*, *m*, 1.5*m*) and hidden neurons (0.5*n*, *n*, 1.5*n*) were examined to consider the nature of the data and the architecture of ANNs. The value *m* indicates the maximum lag within the 95% confidence level of the observed PDSI correlogram, and *n* is the number of input neurons. In the reconstruction phase (step 6 and 7), the input neurons are fixed for decomposition levels, and three architectures of ANNs with different hidden neurons (0.5*n*, *n*, 1.5*n*) were examined.

ANNs were trained for 5000 epochs using the improved backpropagation algorithm with an initial learning rate of 0.01 to reach an error goal of 0.0. Fig. 6 shows training and forecasting results for one-month ahead forecasts with the wavelet transform (level one). The training time increases as the number of input and hidden neurons increase. It is also shown that the RMSE for prediction increases as the RMSE for training decreases. When we choose a network architecture of (11,17) with the lowest RMSE for training, we have no guarantee that the model has sufficient competence for forecasting. When the network architecture of (1,1) with the lowest computing time for training is chosen, the model has no competence for training and forecasting. The RMSE and the normalized computation time (Ntime) for training a network can be considered together by Eq. (6).

$$RSSET = \sqrt{\frac{1}{2}(RMSE_T^2 + Ntime_T^2)} \quad (6)$$

The RSSET was used as criterion to select the optimal ANNs architecture considering both the model competence and efficiency. The RSSET allows to choose the ANN architecture to reduce the risks of overfitting for unknown patterns.

For each forecast lead time, 11 combinations of input and hidden neurons performed the training process, and their RSSETs were calculated. The architecture, which has the lowest RSSET, was chosen. Table 1 shows optimal architectures of the ANNs used in the conjunction model and their RMSEs for training and forecasting. For the training/forecasting phase, the number of hidden neurons of $0.5n$ was chosen in most cases. For the reconstructing phase, the number of hidden neurons was equal to the number of input neurons (decomposition level +1).

FORECASTING RESULTS

The proposed conjunction model (ANN-DD) was evaluated for four forecast lead times (1, 3, 6, and 12 months) and five wavelet decomposition levels (1-5). For each case, data sets during 1957-1990 were used to calibrate the ANNs so as to find efficient architectures of ANNs available for wavelet decomposition levels. The 1991-2000 validation data sets were used to test the performance of the model. During the validation period, the basin has experienced severe droughts in terms of severity and duration as well as extreme wet conditions.

The performance of the proposed conjunction model was measured using various forecast skill criteria and compared with conventional neural network model (ANN), which has an ARX form and no particular data preprocessing except normalizing the PDSI between 0.1 and 0.9.

Forecast Skills

The RMSE was used to measure forecast accuracy, which produces a positive value by squaring the errors. The RMSE increases from zero for perfect forecasts through large positive values as the discrepancies between forecasts and observations become increasingly large (Wilks 1995). Table 1 shows the selected architecture, wavelet decomposition level of the conjunction model (ANN-DD), and their RMSEs during calibration and validation period for four forecast lead times. Fig. 7 shows the time series of the observed and one-, three-, and six-months ahead forecasted values of the Conchos regional PDSI by the conjunction model. The prediction RMSEs for the conjunction model (ANN-DD) are shown in Table 2 compared with reference forecasts and conventional neural network model (ANN). Persistence forecasts are values of the predictand in the previous time period. The one-month ahead forecasts captured the interannual variability and turning points in the time series, which represent the end of dry and wet spells.

Forecast skills related to climatologically average values (climatology) were used to evaluate the performance of forecasting models. The normalized root mean square errors (NRMSE) of the conjunction model referred to the climatology were compared with ANN in Fig. 8. Liu et al. (1998) pointed out that the squared error of the forecast model approaches asymptotically the variance of the time series. In this study, the model holds forecasting skills up to six months of the lead time as shown in Fig. 8. The forecasting skill of all models was lost at 12-months lead time as shown in Table 2. At this situation, the forecast of the system is equivalent to the randomly selecting a value from the past time series (Liu et al. 1998).

The forecast skill is usually presented as a skill score (SS), which is interpreted as a percentage improvement over reference forecasts (Wilks 1995). In this study, the skill score is given by

$$SS_r = \frac{RMSE_m - RMSE_r}{RMSE_p - RMSE_r} = 1 - \frac{RMSE_m}{RMSE_r} \quad (7)$$

where $RMSE_p$ ($=0$), $RMSE_r$, and $RMSE_m$ are the RMSE achieved by perfect forecasts, reference forecasts, and forecast models, respectively. If $SS_r = 0$, it indicates there is no improvement over reference forecasts, and if $SS_r < 0$, the forecast model is inferior to reference forecasts.

Fig. 9 shows the forecasting skill score of models referred to climatology. The conjunction model (ANN-DD) improved 7.3-60% for climatology forecasts. The conjunction model also improved 4.0-7.8% of the forecast performance achieved from the conventional neural network model (ANN). There is clearly significant forecasting skill of the conjunction model above climatology and persistence in the forecasts even at long lead times as shown in Fig. 8 and 9. The high persistence of the PDSI formed the persistence barrier, which other forecasts were not able to overcome. The wavelet decomposition almost eliminated such barrier in the time series by capturing variations in long-term intervals for low frequency information and in short-term intervals for high frequency information.

The PDSI is basically a first order autoregressive process. The long-term memory of the PDSI, highly dependent on antecedent soil and atmospheric moisture conditions, becomes an obstacle in making a valuable forecasting model. However, there are significant increases in the forecast skills for the new proposed model. Fig. 8 and 9 show that the proposed conjunction model (ANN-DD) has forecast skill up to six months and overcome the weakness of other forecasts for a high persistence drought index caused by the long-term memories through wavelet transforms.

Categorical Forecast of the PDSI

PDSI forecasts may be considered using discrete categories. Palmer (1965) categorized a drought condition using the PDSI. For example, in dry spells, a PDSI of -0.5 to 0 is considered near normal, -1.0 to -0.5 incipient drought, -2.0 to -1.0 mild drought, -3.0 to -2.0 moderate drought, -4.0 to -3.0 severe drought, and -4.0 or less extreme drought. The hit rate is the most direct and intuitive measure of the accuracy of categorical forecast (Wilks 1995). Table 3 shows the multi-categorical hit rate using Palmer's 11 classes for the PDSI. In the majority of the cases, the proposed model's performance is more accurate than the other forecasting models.

SUMMARY AND CONCLUSIONS

Droughts are more obstinate and pernicious events than other climatic extreme events, and quantitative drought forecasting remains a difficult but vitally important task for hydrometeorologists and water resources managers. Significant efforts have been made to improve the accuracy of forecasts using statistical and dynamic models. Traditional statistical forecasting methods like linear multiple regressions are unable to capture nonlinearities and nonstationarities in hydrologic variables related to droughts. In this study a forecasting strategy based on the conjunction of wavelet transforms and neural networks is applied to forecast the Palmer drought severity index at various lead times.

The wavelet transform for the forecasting scheme was modified in this study based on the dyadic "à trous" algorithm to reduce the inconsistency of the sub-signal. The ANNs were used to forecast sub-signals from the wavelet decomposition and to reconstruct the forecasted sub-signals. An experimental selection of the neural network architecture was based on a combined

score of RMSE and computing time. Using ANNs allowed to forecast individual wavelet decomposed signals and then reconstruct the forecasts so as to generate a successfully predicted PDSI.

For the particular case of the Conchos River Basin, the overall forecast accuracy of the conjunction model, as measured by RMSE, was better than conventional forecasting models. Moreover the forecasting skill of the conjunction model tested using several skill scores was improved dramatically ranged from 4 to 60% up to six-months lead time when compared with other statistical prediction models.

Linear regression models well represent the Markovian characteristics of the PDSI, and ANNs trained with the improved backpropagation training algorithm have been successfully applied to forecasting models. However, there are restrictions to be applied to the highly persistent drought indexed time series. We conclude that the wavelet transform is a useful tool to forecast time series capturing the dynamics of signals with a multi-resolution of decomposition. The conjunction model can make valuable forecasts for regional droughts through wavelet decompositions, when other forecasting models have limitations due to the embedded unpredictable components in a time series. Accurately predicted droughts allow water resources decision makers to prepare efficient management plans and proactive mitigation programs which can reduce drought-related social, environmental, and economic impact significantly.

ACKNOWLEDGMENTS

This study is based upon works supported by SAHRA (Sustainability of semi-Arid Hydrology and Riparian Areas) at the University of Arizona under the STC Program of the National Science Foundation, Agreement No. EAR-9876800. We are grateful to Dr. Javier

Aparicio (The Mexican Institute of Water Technology), Dr. Hoshin Gupta (The University of Arizona, SAHRA), and Mr. Gi-hyeon Park (The University of Arizona, HyDIS) for assisting with data preparation and providing valuable comments.

REFERENCES

- Alley, W. M. (1984). "The Palmer Drought Severity Index; limitations and assumptions." *Climate Appl. Meteor.* 23, 1100-1109.
- ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000). "Artificial neural networks in hydrology. II: Hydrologic applications." *Journal of Hydrology Engineering*, ASCE, 5(2), 124-137.
- Aussem, A., Campbell, J. and Murtagh, F. (1998). "Wavelet-based feature extraction and decomposition strategies for financial forecasting." *Journal of Computational Intelligence in Finance*, 6(2), 5-12.
- Aussem, A., and Murtagh, F. (1997). "Combining neural network forecasts on wavelet-transformed time series." *Connection Science*, 9(1), 113-121.
- Dai, H. C., and Macbeth, C. (1997). "Effects of learning parameters on learning procedure and performance of a BPNN." *Neural Networks*, 10(8), 1505-1521.
- Demuth, H., and Beale, M. (1994). *Neural Network Toolbox: For Use with MATLAB*, The Math Works, MA, 448.
- European Southern Observatory (1998). *ESO-MIDAS User Guide Volume B Data Reduction*. München, Germany.
- Fletcher, D., and Goss, E. (1993). "Forecasting with neural networks: An application using bankruptcy data." *Information & Management*, 24, 159-167.

- Gabor, D. (1946). "Theory of communication." *J. Inst. Elec. Eng.*, 93, 429-457.
- Grossmann, A., and Morlet, J. (1984). "Decomposition of Hardy functions into square integrable wavelets of constant shape." *SIAM J. Math. Anal.*, 15, 723-736.
- Gupta, H. V., Hsu, K., and Sorooshian, S. (2000). "Effective and efficient modeling for streamflow forecasting." *Artificial Neural Networks in Hydrology*, R. S. Govindaraju and A. R. Rao, ed., Kluwer Academic Publishers, Boston, 7-22.
- Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation*. MacMillan, NY, 696
- Hsu, K., Gupta, H. V., and Sorooshian, S. (1995). "Artificial neural network modeling of the rainfall-runoff process." *Water Resources Research*, 31(10), 2517-2530.
- International Boundary and Water Commission (2002). *Deliveries of Waters Allotted to the United States Under Article 4 of the United States-Mexico Water Treaty of 1944*. US Section Report, El Paso, TX.
- Kim, T., Valdés, J. B., and Aparicio, J. (2002). "Frequency and spatial characteristics of droughts in the Conchos River basin, Mexico." *Water International*, IWRA, 27(3), 420-430.
- Kim, T., Valdés, J. B., and Yoo, C. (2003). "A nonparametric approach for estimating return periods of droughts in arid regions." Accepted for publication by ASCE, *Journal of Hydrologic Engineering*.
- Labat, D., Ababou, R., and Mangin, A. (2000). "Rainfall-runoff relationships for karstic springs. Part II: continuous wavelet and discrete orthogonal multiresolution analyses." *Journal of Hydrology*, 238, 149-178.
- Liu, Z., Valdés, J. B., and Entekhabi, D. (1998). "Merging and error analysis of regional hydrometeorologic anomaly forecasts conditioned on climate precursors." *Water Resources Research*, 34(8), 1959-1969.

- Maier, H. R. and Dandy, G. C. (2000). "Neural networks for the prediction and forecasting of water resources variables; a review of modeling issues and applications." *Environmental Modelling and Software*, 15, 101-124.
- Makridakis, S., Wheelwright, S. C., and McGee, V. E. (1983). *Forecasting: Methods and Applications*, Wiley, NY, 923.
- Mallat, S. G. (1998). *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, 577.
- Misiti, M., Misiti, Y., Oppenheim, G., and Poggi, J. (2000). *Wavelet Toolbox: For Use with MATLAB*, The Math Works, MA, 941.
- National Drought Policy Commission (2000). *Preparing for Drought in the 21st Century*. Washington, D. C.
- NOAA Paleoclimatology Program of the National Climatic Data Center (2000). *North American Drought: A Paleo Perspective*. Boulder, CO.
- Palmer, W. C. (1965). *Meteorological Drought*. Research Paper No. 45, U.S. Weather Bureau, Washington, D.C., 58.
- Rao, R. M., and Bopardikar, A. S. (1998). *Wavelet Transforms: Introduction to Theory and Applications*. Addison-Wesley, MA, 310.
- Riebsame, W. E., Changnon, S. A., and Karl, T. R. (1991). "Drought and natural resources management in the United States: Impacts and implications of the 1987-89 drought." *Westview Press*, 11-92.
- Schmandt, J. (2002). "Bi-national water issues in the Rio Grand/Río Bravo basin." *Water Policy*, 4(2), 137-155.
- Tang, Z., and Fishwick, P. A. (1993) "Feedforward neural nets as models for time series forecasting." *Journal on Computing, ORSA*, 5(4), 374-385.

- Texas Center for Policy Studies (2001). *The Rio Conchos: A Preliminary Overview*. Austin, TX.
- Torrence, C., and Compo, G. P. (1998). "A practical guide to wavelet analysis." *Bulletin of the American Meteorological Society*, 79(1), 61-78.
- Tsoukalas, L. H., and Uhrig, R. E. (1996). *Fuzzy and Neural Approaches in Engineering*. Wiley, NY, 587.
- Wilks, D. S. (1995). *Statistical Methods in the Atmospheric Sciences An Introduction*, Academic Press, San Diego, 467.
- Woodhouse, C. A., and Overpeck, J. T. (1998). "2000 years of drought variability in the Central United States." *Bulletin of the American Meteorological Society*, 79(12), 2693-2714.
- Zhang, B. L., and Dong, Z. Y. (2001). "An adaptive neural-wavelet model for short term load forecasting." *Electric Power Systems Research*, 59, 121-129.
- Zheng, T., Girgis, A. A., and Makram, E. B. (2000). "A hybrid wavelet-Kalman filter method for load forecasting." *Electric Power Systems Research*, 54, 11-17.

Table 1. Optimal architectures of the conjunction model (ANN-DD) and their RMSEs during the calibration and validation period.

Lead Time (month)	Decomposition Level	Architecture (Forecasting Phase)	Architecture (Reconstructing Phase)	RMSE for Calibration (1955-1990)
1	2	(4,2)	(3,3)	1.0767
3	3	(4,2)	(4,4)	1.779
6	3	(4,2)	(4,4)	2.1256
12	2	(11,6)	(3,3)	2.9096

Table 2. RMSEs of reference forecasts and forecasting models (1991-2000)

LT (Month)	Climatology	Persistence	ANN	ANN-DD
1	2.5676	1.0952	1.0962	1.0524
3	2.5676	1.8367	1.9037	1.7544
6	2.5676	2.5202	2.5530	2.3803
12	2.5676	3.3705	3.2808	3.3126

Table 3. Hit rates of multicategorical forecasts of the PDSI (1991-2000)

Model	Forecast lead time (month)			
	1	3	6	12
Climatology	0.0417	0.0417	0.0417	0.0417
Persistence	0.5750	0.1250	0.0500	0.0167
ANN	0.4750	0.1083	0.0333	0.0417
ANN-DD	0.6417	0.3250	0.0917	0.1000

List of Figures

FIG. 1. Conchos River basin, Mexico. (26°N - 30°N , 104°W - 108°W . The basin area is $72,000\text{km}^2$ from head waters to its mouth to the Bravo/Grande River at Ojinaga, Mexico.)

FIG. 2. Annual Conchos River inflows into the Bravo/Grande River at Ojinaga, Mexico, 1955-2000

FIG. 3. Cumulative precipitation in 1990s for the Conchos River basin

FIG.4. Typical three layered feed forward neural networks (FFNNs) with a backpropagation training algorithm

FIG. 5. The schematic representation of the forecasting model with a conjunction of wavelet transforms and neural networks

FIG. 6. The training and forecasting results for one-month ahead forecasts with level one wavelet decomposition

FIG. 7. Time series of observed and forecasted PDSIs in the Conchos River basin during the calibration and validation period. (a) one-month ahead forecasts (b) three-months ahead forecasts (c) six-months ahead forecasts.

FIG. 8. The normalized RMSE (NRMSE) of forecasting models referred to climatologically average values (Climatology)

FIG. 9. The forecast skill score (SS) of forecasting models referred to climatologically average values (Climatology)



FIG. 1. Conchos River basin, Mexico. (26°N-30°N, 104°W-108°W. The basin area is 72,000km² from head waters to its mouth to the Bravo/Grande River at Ojinaga, Mexico.)

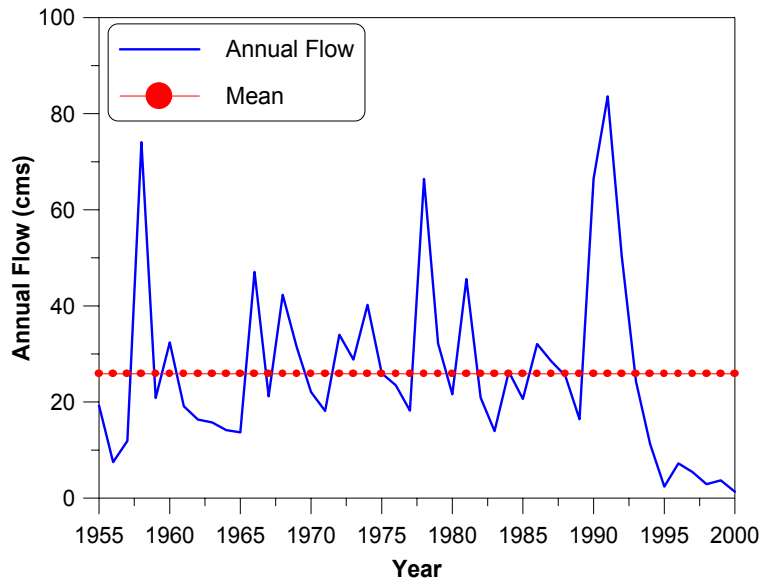


FIG. 2. Annual Conchos River inflows into the Bravo/Grande River at Ojinaga, Mexico, 1955-2000

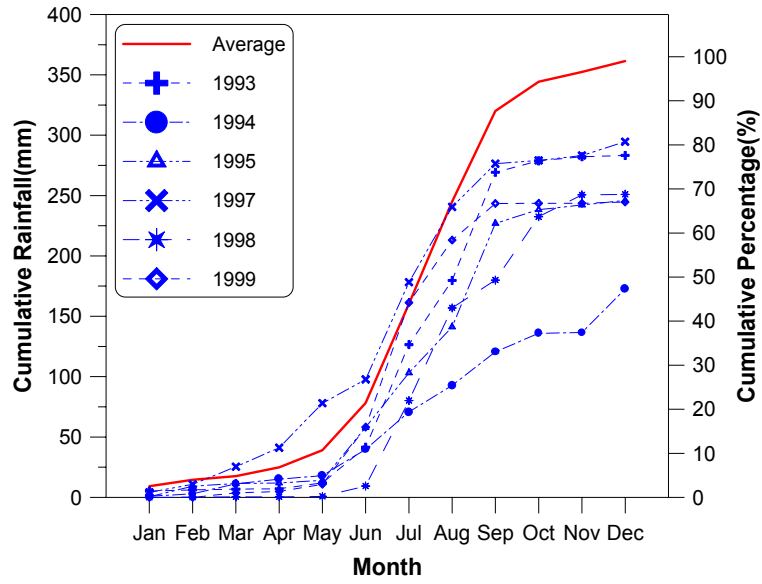


Figure 3. Cumulative precipitation in 1990s for the Conchos River basin

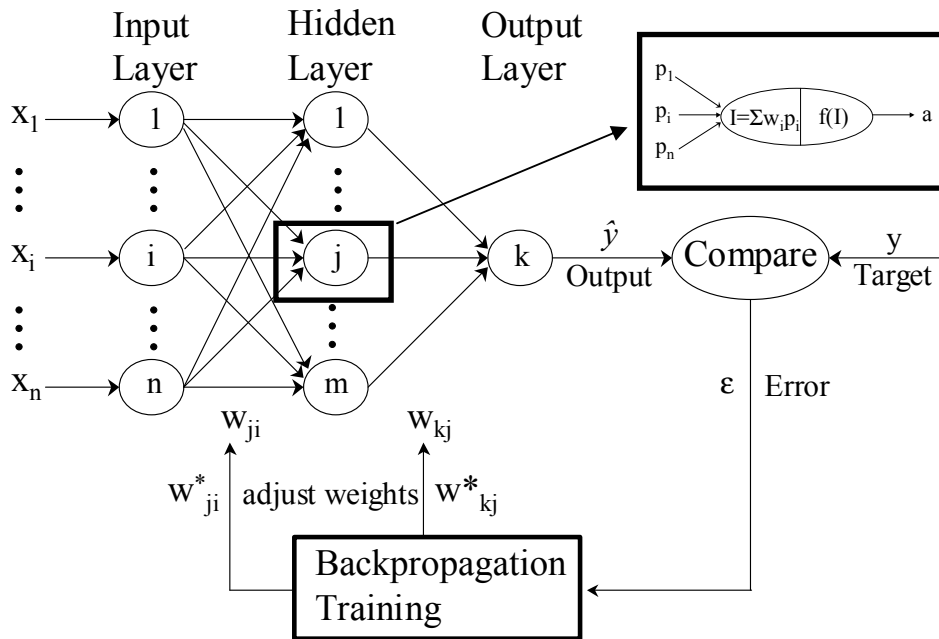


FIG. 4. Typical three layered feed forward neural networks (FFNNs) with a backpropagation training algorithm

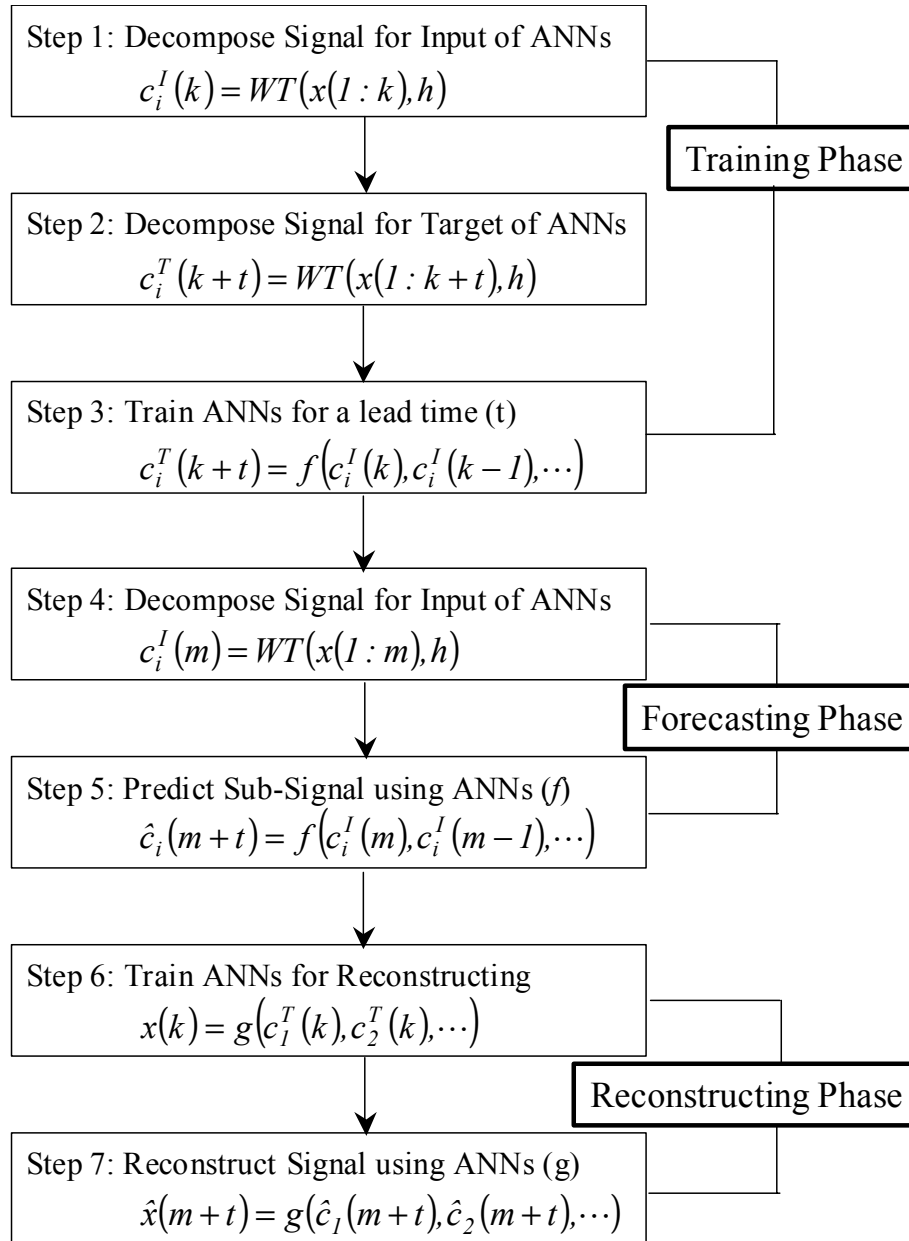


FIG. 5. The schematic representation of the forecasting model with a conjunction of wavelet transforms and neural networks

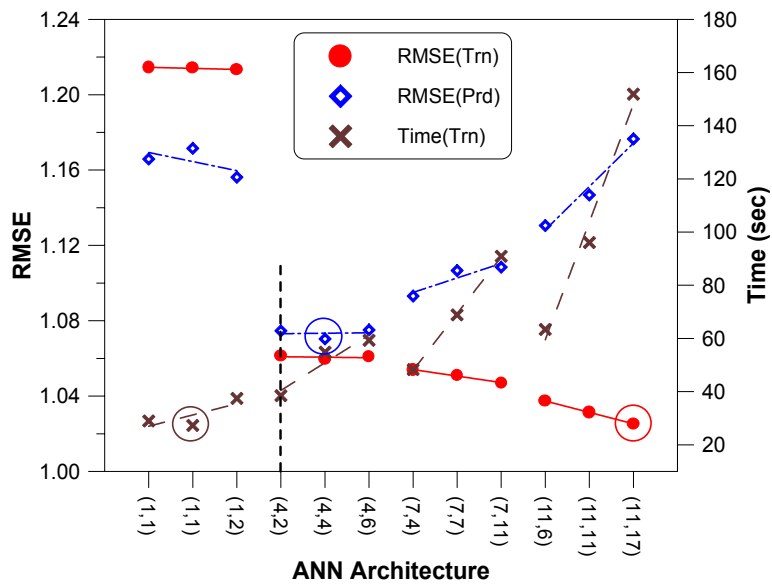
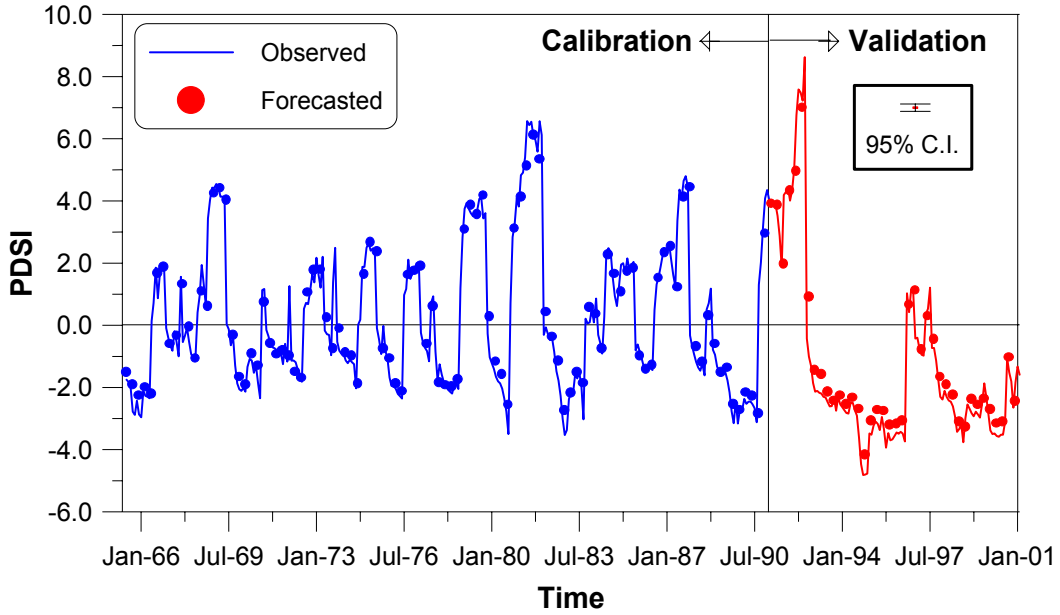
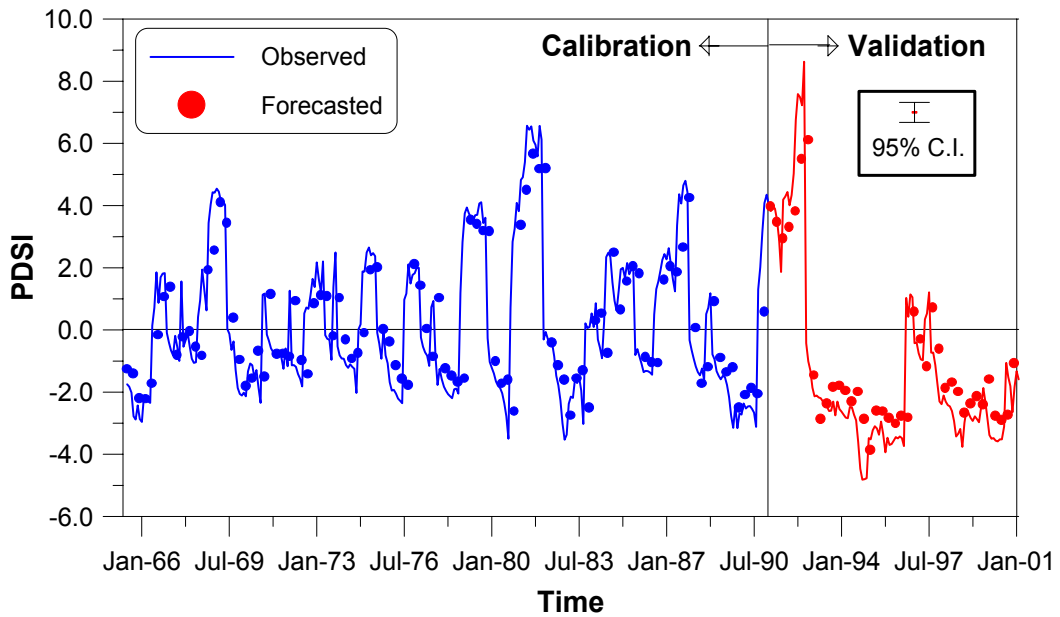


FIG. 6. The training and forecasting results for one-month ahead forecasts with level one wavelet decomposition

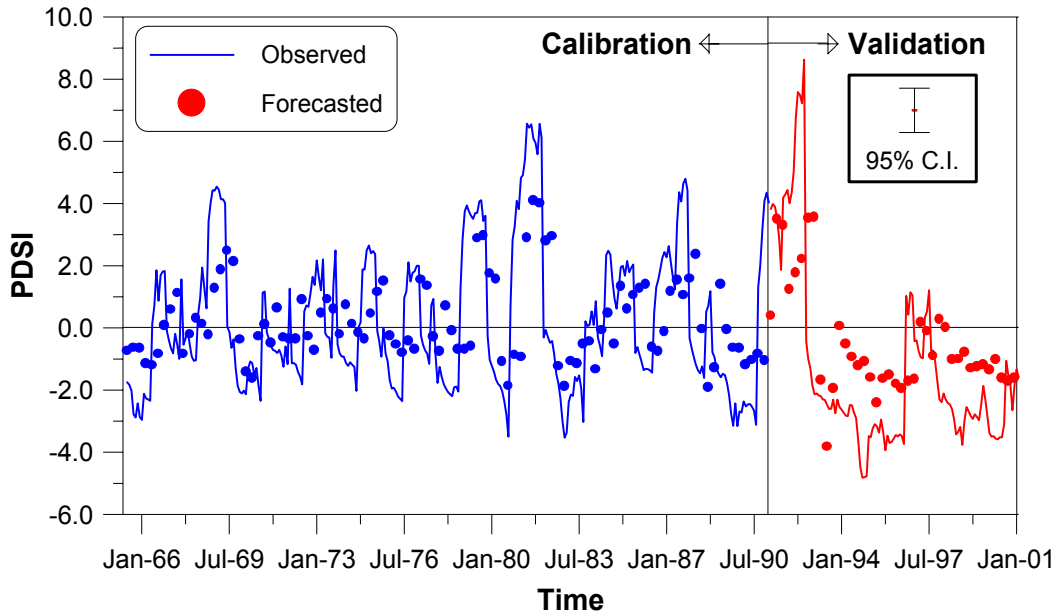


(a)



(b)

FIG. 7. Time series of observed and forecasted PDSIs in the Conchos River basin during the calibration and validation period. (a) one-month ahead forecasts (b) three-months ahead forecasts (c) six-months ahead forecasts. The vertical line in the box represents the respective 95% confidence limit.



(c)

FIG. 7. Time series of observed and forecasted PDSIs in the Conchos River basin during the calibration and validation period. (a) one-month ahead forecasts (b) three-months ahead forecasts (c) six-months ahead forecasts. The vertical line in the box represents the respective 95% confidence limit.

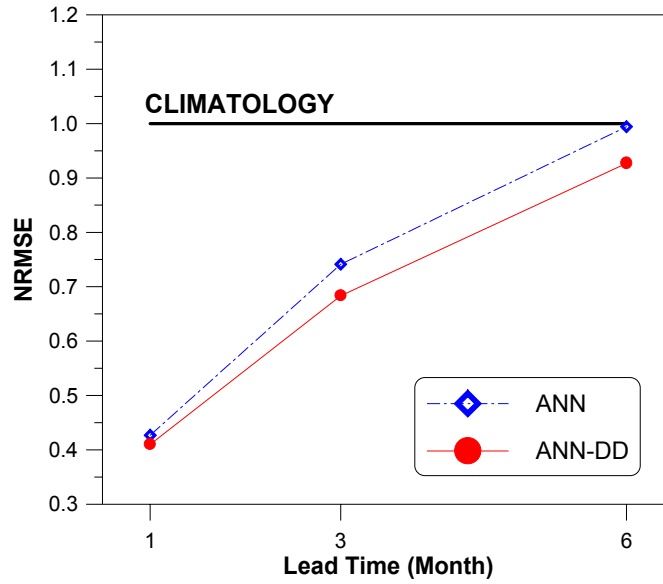


FIG. 8. The normalized RMSE (NRMSE) of forecasting models referred to climatologically average values (Climatology)

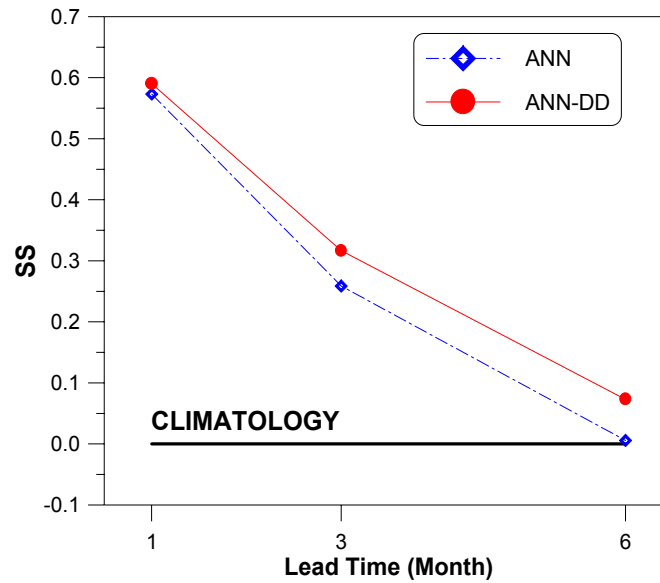


FIG. 9. The forecast skill score (SS) of forecasting models referred to climatologically average values (Climatology). The skill score of climatology is 0.0.