Hydrological Sciences Journal

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/thsj20

Hydrologic simulation approach for El Niño Southern Oscillation (ENSO)-affected watershed with limited rain gauge stations

Suresh Sharma\textsuperscript{a}, Puneet Srivastava\textsuperscript{b}, Xing Fang\textsuperscript{c} & Latif Kalin\textsuperscript{d}

\textsuperscript{a} Civil and Environmental Engineering Program, Youngstown State University, Youngstown, OH, 44555, USA
\textsuperscript{b} Biosystems Engineering Department, Auburn University, Auburn, AL, 36849, USA
\textsuperscript{c} Civil Engineering Department, Auburn University, AL, 36849, USA
\textsuperscript{d} School of Forestry and Wildlife Sciences, Forestry and Wildlife Building, Auburn, AL, USA

Accepted author version posted online: 14 Aug 2014.

To cite this article: Suresh Sharma, Puneet Srivastava, Xing Fang & Latif Kalin (2014): Hydrologic simulation approach for El Niño Southern Oscillation (ENSO)-affected watershed with limited rain gauge stations, Hydrological Sciences Journal, DOI: 10.1080/02626667.2014.952640

To link to this article: http://dx.doi.org/10.1080/02626667.2014.952640

Disclaimer: This is a version of an unedited manuscript that has been accepted for publication. As a service to authors and researchers we are providing this version of the accepted manuscript (AM). Copyediting, typesetting, and review of the resulting proof will be undertaken on this manuscript before final publication of the Version of Record (VoR). During production and pre-press, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal relate to this version also.

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions
Hydrologic Simulation Approach for El Niño Southern Oscillation (ENSO)-Affected Watershed with Limited Rain Gauge Stations

Suresh Sharma\textsuperscript{1*}, Puneet Srivastava\textsuperscript{2}, Xing Fang\textsuperscript{3}, Latif Kalin\textsuperscript{4}

\textsuperscript{1}Assistant Professor, Civil and Environmental Engineering Program, Youngstown State University, Youngstown, OH, 44555, (330)-941-1741, E-mail: ssharma06@ysu.edu.

\textsuperscript{2} Professor, Biosystems Engineering Department, Auburn University, Tom E. Corley Building, Auburn, AL, 36849, (334)-844-7426, E-mail: srivapu@auburn.edu.

\textsuperscript{3} Professor, Civil Engineering Department, Auburn University, 229 Harbert Engineering Center Auburn, AL, 36849, (334)-844-8778, E-mail: xing.fang@auburn.edu.

\textsuperscript{4} Associate Professor, School of Forestry and Wildlife Sciences, Forestry and Wildlife Building, Auburn, AL, (334)-844-4671, 36849, E-mail: kalinla@auburn.edu.

*Corresponding Author.

Abstract

Since the performance of hydrological models relies on numerous factors, the selection of an appropriate modeling approach for hydrological study has always been a crucial issue. The major objective of this research is to demonstrate that data driven models such as the Adaptive Neuro-Fuzzy Inference system (ANFIS) are more suitable in a region where spatially distributed precipitation datasets are not available. Since precipitation has a teleconnection with the El Niño Southern Oscillation (ENSO) in different parts of the world, the sea surface temperatures (SSTs) and sea level pressures (SLPs) of the equatorial Pacific can be expected to act as surrogates for the precipitation if there are insufficient rain gauge stations in the watershed. Moreover, in contrary to conceptual and physically-based models, data driven models can incorporate SST and SLP in their input vectors, and hence additional forcing of SST with precipitation has been experimented with in past studies. Therefore, our second objective is to test whether the
additional forcing of SST and SLP will improve the hydrologic simulation. For this, various ANFIS models for the winter season were developed considering 10 rain gauge stations situated at various locations from the watershed. Rainfall from each rain gauge station was considered in the ANFIS model one at a time with and without SST/SLP. The results show that the performance of the ANFIS model improved with the additional fusion of SST and SLP, especially when a rain gauge station from a remote location was considered. However, this improvement was observed when the analysis was primarily focused on the winter season which is a period with a strong ENSO signal.

**Key Words:** ENSO, ANN, Fuzzy logic, ANFIS, spatial variability, hydrologic modeling

**Introduction**

Hydrologic simulation is one of the most important topics in hydrology due to its wide application, e.g. in flood and drought forecasting and in the design of hydraulic structures. Furthermore, hydrologic simulation is a complex process, relying on various linear and non-linear parameters (Singh 1988). Accordingly, various approaches of modeling such as process-based models, conceptual models, or data-driven/black-box models (Jayawardena et al. 2006) have been proposed for streamflow simulations for different historical time periods. Different studies in the past have reported that the performance of the models varies according to several factors including the availability of the data, user’s knowledge, the accuracy needed and the availability of time and resources. Therefore, it is difficult to generalize the most suitable modeling approach (Jayawardena et al. 2006).

Irrespective of the modelling approach used, hydrologic simulation is affected significantly by the quality of the precipitation data (Sharma et al. 2012a). A fully physically-
based model representing complex land use and soil characteristics with detailed input is also not capable of simulating the streamflow well if the spatial variability of precipitation is not adequately incorporated. On the other hand, spatially distributed rainfall data needed for physically-based or distributed and semi-distributed hydrologic models are rarely available. Generally, it is difficult to capture the spatial variability of the rainfall in watershed modeling unless an adequate number of rain gauge stations is available (Sharma et al. 2012a). In addition, the conventional method of using tipping bucket rain gauges are actually point measurements and do not capture the actual rainfall occurring on the surface unless a dense enough network of rain gauges is available.

Since variations in the sea surface temperature (SST) and sea level pressure (SLP) of various regions of the equatorial Pacific have significant influences on inter-annual variation of precipitation and temperature in different parts of the world (Chiew et al. 1998, Piechota and Dracup 1996, Rajagopalan and Lall 1998, Kulkarni 2000, Hansen et al. 2001, Pascual et al. 2000, Keener et al. 2007, Roy 2006, Barsugli et al. 1999, McCabe and Dettinger 1999), SST and SLP can be additional forcing variables in data driven models to improve hydrologic simulation. This variation in SST and SLP along the equatorial pacific is called the El Niño Southern Oscillation (ENSO).

Different ENSO phases (La Niña, El Niño and Neutral) are identified using various ENSO indicators (SSTs and SLPs) in the equatorial Pacific Ocean. Considering the potential teleconnection between streamflow and ENSO indicators, it may be possible to explicitly utilize these indicators in rainfall-runoff modeling. In other words, SST and SLP can be expected to act as surrogates for the precipitation, particularly for a watershed where spatially distributed precipitation data are not available. However, physically-based and conceptual models cannot
incorporate SST and SLP data as an input, and hence data driven models such as artificial neural networks (ANNs) and fuzzy logic-based models could be appropriate choices in such cases.

Although ANN and fuzzy logic have been proven to be effective and useful when compared with conventional modeling due to their ability to handle large amounts of data, especially when the underlying physical processes are not understood or parameterized (Nayak et al. 2004), scientists have shown an interest in combining these two models. For example, a neuro-fuzzy system known as an Adaptive Neuro-Fuzzy Inference System (ANFIS) (Jang 1993) has been developed and extensively applied in hydrological (Pramanik and Panda 2009, Mukerji et al. 2009) and water quality modeling (Yan et al. 2010).

This study explored the suitability of ANFIS models for simulating streamflow using potential ENSO indicators (SST and SLP) of the equatorial Pacific as input variables. Few studies in the past have used ENSO indicators in data driven models (Khalil et al. 2005, Makkeasorn et al. 2008) for hydrologic study. Thus, the objective of this paper is to: (1) demonstrate that the ANFIS model is a suitable choice for ENSO affected watersheds with limited rain gauge stations, and ii) incorporate SST and SLP into an ANFIS model, comparing the performance with and without these additional inputs.

**Theoretical Background**

ENSO, which is measured using various ENSO indicators, is considered as one of the most reliable phenomenon for relating inter-annual climate variability in streamflow at the local as well as global scales (Ropelewski and Halpert 1986). For ENSO prediction, the National Oceanic and Atmospheric Administration (NOAA) constantly monitors the Niño 3.4 region which is located between 5°N - 5°S and 120°-170°W in the equatorial pacific. An ENSO phase is classified based on the Niño 3.4 index which is calculated based on a 3 month running mean of
ERSST.v3b SST anomalies in this region (Trenberth and Stepaniak 2001). ENSO indicators are incorporated in the ANFIS model which is described in the following section.

**ANFIS**

ANFIS is a multi-layer feed-forward network that utilizes neural network and fuzzy logic to map an input vector to an output vector by using back propagation or a combined algorithm. A generic example of an ANFIS model using two inputs and one output is demonstrated in Figure 1. This model is comprised of five layers with two rules, and two membership functions (MFs) for each input (Figure 1).

The ANFIS model comprising two fuzzy if-then rules can be written as follows:

Rule 1: If X is P1 and Y is Q1, then 
\[ f_1 = a_1 X + b_1 Y + r_1 \]  

Rule 2: If X is P2 and Y is Q2, then 
\[ f_2 = a_2 X + b_2 Y + r_2 \]  

where P1 and P2 are the MFs for input X, Q1 and Q2 are the MFs for input Y, a1, b1, r1 and a2, b2, r2 are the parameters to be determined for the output function. The brief operation of the ANFIS model is explained in the following section.

Layer 1  
In this layer, each node with membership functions is called an adaptive node.

\[ O_{i}^{1} = \mu_{P_{i}}(X), \text{ for } i = 1,2 \]  

\[ O_{i}^{1} = \mu_{Q_{i-2}}(Y), \text{ for } i = 3,4 \]  

Here, X and Y are inputs, \( \mu \) represents the membership function and \( P_{i} \) and \( Q_{i-2} \) are fuzzy sets.

Layer 2  
In this layer, a single output representing the result of the predecessor (firing strength) of that rule is obtained. Hence, the outputs \( O_{i}^{2} \) of this layer can be expressed as follows.
\[ O_i^2 = w_i = \mu_{p_i}(X)\mu_{q_{i-2}}(Y), \quad \text{for } (i) = 1, 2 \]  

Layer 3  
In this layer, every node is labeled as N, and the ratio of firing strength corresponding to each \( i^{th} \) rule to the summation of firing strength that corresponds to all rules is computed.

\[ O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2} \]  
where \( i = 1, 2 \)

Layer 4  
In this layer, the input contribution of each \( i^{th} \) rule for the output of the model is computed using the following equation.

\[ O_i^4 = \bar{w}_i f_i = \bar{w}_i \sum_{i=1}^{2}(a_iX + b_iY + r_i) \]  
where \( \bar{w}_i \) is the output from layer 3 and \([a_i, b_i, r_i]\) are the model parameters.

Layer 5  
The final output is computed as the combination of all receiving signals

\[ O_i^5 = \sum_{i}^{2} \bar{w}_i f_i = \frac{\sum_{i}^{2} w_i f_i}{\sum_{i}^{2} w_i} \]  

The ANFIS model applies a hybrid-learning algorithm (i.e. the combination of the gradient descent method and the least-squares method) to update the model parameters. In order to reduce the number of rules and for effective partitioning, several clustering methods have been suggested to organize the data and construct the rules. Some of the most widely used approaches are grid partitioning (Jang and Sun 1995), the subtractive clustering algorithm (SCA) (Chiu 1996), and fuzzy C-means (FCM) (Dunn 1973).

**Study Area and Data**

The study area is located in the Chickasaw Creek watershed (Figure 2) of Mobile County of Southern Alabama (USA) in the Mobile River Basin. The watershed is 714 km\(^2\) in size and starts
at Citronelle, AL in the north, and eventually drains into Mobile Bay. The watershed is dominated by coastal plain geology with an elevation ranging from 0 m to 13.11 m above mean sea level. The average annual precipitation in the watershed (1651 mm) is relatively higher than in other parts of Alabama. Air temperature and precipitation data sets were available from the weather station at the Mobile Regional Airport (Coop ID-015478), downloaded from the National Climatic Data Center (NCDC). This is the closest rain gauge station to the watershed which is still 32 km away from the watershed boundary. Streamflow data, recorded since 1952, were available at the USGS gauge (02471001) which drains 357 km² of the watershed.

The climate of this region, especially the study area, is highly affected by ENSO, which has been reported in various past studies (Hansen et al. 2001, Sharma et al. 2012b). Winter is characterized by high precipitation and streamflow, and low temperature during the El Niño phase, which is just opposite to the conditions during La Niña as shown in Figure 3. The ENSO signal is comparatively better in winter and spring compared to summer. Similarly, the ENSO correlation with precipitation and temperature is opposite in nature during a period from August to October compared to winter and spring (Figure 4).

Average SST, SLP differences (between Tahiti and Darwin) and the trade wind index (Lukas et al. 1984) in the Niño 3.4 region (5° N to 5° S, 120° W to 170° W) (Trenberth and Stepaniak 2001) were available since 1950, and downloaded from NOAA’s NCDC and National Center for Environment Prediction (NCEP) website. The data used in the study with their sources and formats are reported in Table 1.
Methodology

**Estimation of Parameters**

Based on the simple architecture of the ANFIS model (designed for the given input parameters), the ANFIS structure was further extended to several other inputs. For the present study, \( SF = f \) (SST, SLP, T, PCP), where \( SF \), PCP and T are streamflow (m\(^3\)/s), precipitation (mm) and air temperature (\(^\circ\)C), respectively. For simplicity, an example of an ANFIS having two inputs with two rules is presented below for illustration. Assuming \( P_1:P_2 \) and \( Q_1:Q_2 \) were the respective membership functions for two input variables (SST and PCP) and \( a_1:b_1, a_2:b_2 \) were the model parameters associated with the output function, then the resultant output was computed as linear combinations of subsequent parameters. The equation to predict \( SF_i \) (output), i.e. the streamflow at time \( t_1 \), was simplified as follows.

\[
SF_i = (w_1 S_{ST})a_1 + (w_1 PCP)b_1 + w_1 r_1 + (w_2 S_{ST})a_2 + (w_2 PCP)b_2 + w_2 r_2
\]

(9)

All equations to predict the streamflow time series (\( SF_1, SF_2 \ldots SF_n \)) were arranged in matrix form as follows:

\[
A = \begin{bmatrix}
    w_1 S_{ST_1} & w_1 PC_{P_1} & w_1 S_{ST_2} & w_1 PC_{P_2} & \cdots & w_1 S_{ST_n} & w_1 PC_{P_n} \\
    w_2 S_{ST_1} & w_2 PC_{P_1} & w_2 S_{ST_2} & w_2 PC_{P_2} & \cdots & w_2 S_{ST_n} & w_2 PC_{P_n} \\
    \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
    w_n S_{ST_1} & w_n PC_{P_1} & w_n S_{ST_2} & w_n PC_{P_2} & \cdots & w_n S_{ST_n} & w_n PC_{P_n}
\end{bmatrix}
\]

\[
Z = \begin{bmatrix}
a_1 \\
b_1 \\
ar_1 \\
a_2 \\
b_2 \\
r_2
\end{bmatrix}
\quad \text{and} \quad
B = \begin{bmatrix}
SF_1 \\
SF_2 \\
SF_3 \\
\vdots \\
SF_n
\end{bmatrix}
\]
\[ AZ = B \]  

where \( Z \), as an unknown model parameter matrix, was derived from training data sets solving the following equation:

\[ Z = (A^T A)^{-1} A^T B \]  

where \( A^{-1} \) is the inverse of matrix “A” and \( A^T \) is the transpose of the matrix “A”. In order to solve this matrix, we used the hybrid learning algorithm which combines both the least-squared method and the back propagation algorithm. The hybrid approach trains the data rapidly and also converges much faster.

**Input Data Selection**

The selection of input variables is a crucial part of ANFIS modeling. Optimum inputs should be selected in order to best capture the input and output relation. Since streamflow of this month may also be associated with the temperature and precipitation of an earlier month, the cross correlation technique was used to decide on the appropriate inputs. Moreover, streamflow is an integral form of the quick response component of the watershed, i.e. surface runoff, and the slow response component, i.e. baseflow. Baseflow might be associated with the precipitation of many time steps ahead whereas surface runoff is associated with the current time step. Therefore, it is always preferable to partition the streamflow into base-flow (BF) and surface runoff (SR) components. Accordingly, we used the base flow filter program (Lim et al. 2005) ([https://engineering.purdue.edu/~what/](https://engineering.purdue.edu/~what/)), developed at Purdue University for online hydrograph separation. We used the Recursive Digital Filter method to derive separate SR and BF data sets from streamflow data to develop separate SR and BF models. The input variables for the SR model were “estimated surface runoff”, temperature, SST and SLP at different lead times (Table 2). The estimated surface runoff was computed from precipitation using the SCS curve number
equation (Bosznay 1989). The curve number (CN) was estimated based on the SSURGO soil (SSURGO 2010) database and the 2001 land cover data set (NLCD 2010). Likewise, the model input variables for the BF model were SST, SLP, air temperature and precipitation at different lead times (Table 2).

In the next step, multiple linear regression (MLR) models were fit to different sets of inputs in order to identify the significant model inputs. It was important to evaluate whether SSTs and SLPs were the significant model inputs in addition to temperature and precipitation. In fact, precipitation and temperature are also affected by SST and SLP. One way to see whether the additional forcing of SST and SLP are beneficial is to evaluate the adjusted $R^2$ from the MLR model. If the additional forcing of SST and SLP is not beneficial, those inputs may not be statistically significant in the MLR model, and there will be no additional improvement in the adjusted $R^2$. Even though MLR and ANFIS are two different modeling approaches, the parameters which are not statistically significant in MLR may not be appropriate inputs in the ANFIS model as well. For this reason, a MLR model was developed, and the model depicted that the parameters were significant for $P$ value $< 0.1$. A sensitivity analysis was also conducted by excluding those parameters one at a time from the input vectors and evaluating the performance. The results suggested that SST and SLP were two of the most sensitive model inputs. The optimal lead time and the number of inputs were decided after the sensitivity analysis. Various combinations of inputs were experimented with at different lead times to determine the best inputs. Any additional inputs beyond the optimum sets did not improve the model performance.

The analysis suggested that both SST and SLP data were important, especially in the seasons that ENSO showed a clear signature with streamflow (winter, spring and early fall).
Each model was comprised of separate SR and BF models. The ANFIS model, without SST and SLP, was developed to evaluate the importance of SST and SLP inputs.

Model Development and Implementation

The most sensitive sets of inputs giving the best result in the MLR model were employed in the ANFIS model. The final form of the equation for the SR ANFIS model is given as:

$$SR(t) = f(SST(t-1), SLP(t), T(t-1), T(t), SR_E(t-1), SR_E(t))$$  \hspace{1cm} (12)

where SR is monthly average surface runoff in m$^3$/s, t is the monthly time step, SR$_E$ is the estimated surface runoff using the SCS curve number method based on precipitation, land use/land cover and soil. The model inputs for both the SR and BF models are also presented in Table 2. The ANFIS model was extended to a daily scale with some modification in its input datasets. For the SR model, while simulating at a daily scale, we used the same inputs that we used for the monthly SR model. However, baseflow at the daily scale may depend on the precipitation characteristics of the preceding few weeks due to the time lag of the groundwater contribution. Therefore, the combination of precipitation sets at different lead times were experimented with in the ANFIS model until the model performance improved.

The MATLAB fuzzy logic toolbox was selected for ANFIS simulation using 50 years of observed data; 80% of the total data were used for model training, 10% for model validation and 10% for model testing. Both SCA and FCM clustering methods were experimented with for the selection of the initial model parameters. For this particular study and the given datasets, SCA was found to be more suitable to cluster the datasets because it provided consistently better performance in training, validation and testing. The parameters to be fixed for SCA were clustering radius, acceptance ratio and rejection ratio. We varied the clustering radius from 0.3 to 1 using 0.05 as the step size to determine the optimal parameters which were ascertained by
minimizing the root mean squared error. Other parameters such as the number of epochs, range
of influence of the cluster center, the acceptance ratio, and the rejection ratio of the model were
optimally ascertained through a repeated trial and error. In order to avoid overtraining of the
model, the simulation was stopped when the validation error started increasing.

The performance of each ANFIS model was measured via three non-dimensional
measures including $R^2$ (coefficient of determination), the Nash-Sutcliffe efficiency, $NSE$ (Nash
and Sutcliffe 1970), and the mass balance error ($MBE$) and the dimensional measure, root mean
square error ($RMSE$). These details of these performance measures are described in many other
articles (Moriasi et al. 2007, Abrahart and See 2007, Dawson et al. 2007).

Since our second objective was to test whether the SST and SLP improve streamflow
simulation or not, several winter season ANFIS models were developed. We employed three
techniques to evaluate the additional benefit of SST and SLP for streamflow simulation. First, we
compared the performance of the ANFIS model including SST and SLP with one excluding SST
and SLP from its input vector for the entire season both in SR and BR models. Secondly, 20
ANFIS models were developed for the winter season with and without SST/SLP by replacing the
rain gauge considered for the watershed (mobile, Coop ID-015478) by other rainfall stations
located at different locations, one after the other. Third, 20 additional ANFIS models were
developed with and without SST/SLP considering only La Niña and El Niño periods. Since
neutral period characteristics are associated with the immediate preceding season, the neutral
period does not show any trend. That is, the neutral period followed by El Niño is different to a
neutral period followed by La Niña. Therefore, in the third case, we developed the model using
only the La Niña and El Niño phases to further diagnose how SST/SLP play a role for
simulation, particularly when it enters the La Niña and El Niño phases. For the second and third
case, we developed models with 10 rain gauge stations as inputs including the rain gauge considered for the watershed and then randomly selected 9 additional rainfall stations at various locations in three different States of Alabama, Mississippi and Louisiana with closest proximity to the watershed, which furnish long-term climate data since the 1950s (Figure 2). These analyses were conducted in order to evaluate the model performance with the rainfall located at a significant distance from the watershed. The underlying assumption was that the SST and SLP could be helpful for satisfactory simulation even if the spatially distributed precipitation was not available. Therefore, in each case, the rain gauge station considered for the watershed was replaced by those additional rain gauge stations one at a time, and the model performance was evaluated with and without SST/SLP. The reason for developing a winter season model is to test the sensitivity of SST and SLP during the winter period because ENSO has a clear signature with precipitation and temperature in winter. The ENSO signal is comparatively stronger in the winter and spring seasons in this region (Figure 3). Also, the ENSO signal is clear in August to October but opposite in characteristics to that of winter and spring (Figure 4).

Since we had to develop 40 different ANFIS models for the second and third analysis, we limited our analysis to the winter seasons, and also we used temperature, precipitation, SST and SLP at the current time level as inputs. This is because the effect of ENSO indicators may not be distinct due to too many confounded variables if the inputs at different lag times were considered. Moreover, our objective in this analysis was to verify that SST and SLP are beneficial for streamflow simulation. Therefore, we developed a single SF model due to the computational efforts required, particularly for the second and third analyses. The second and third analyses, which comprise a type of sensitivity analysis of the ANFIS model, were also developed using 80% training data, 10% validation and 10% testing data.
Result and Discussion

Performance of the ANFIS Model

The overall performance of the simulated streamflow model, i.e., the combined SR and BF models, against observed data, is presented in Table 3. The model simulated streamflow with satisfactory performance in each stage of training, validation and testing (Table 3). From the visual inspection, it can be concluded that the model simulation matched well with the observed monthly streamflow (Figure 5). Furthermore, the ANFIS model simulation was at a daily time step. It is interesting to note that SLP, SST and SST (t-1) were the most sensitive inputs for the base flow model at a daily temporal resolution as well. The overall ANFIS model performance including training, validation and testing was also satisfactory at a daily temporal resolution (NSE = 0.68, MBE = 0.6%, RMSE = 7 m³/s).

Model performance was assessed through the statistical parameters, i.e. $NSE$, $R^2$ and $MBE$. ANFIS was able to capture some of the highest flow associated with ENSO events of January 1954 (El Niño), April 1980 (Neutral), May 1991 (El Niño), July 1997 (El Niño) and September 1998 (La Niña) (Figure 5). Because SST and SLP were directly implemented in the ANFIS model, it was interesting for us to explore how this model would simulate streamflow at corresponding major historic ENSO events, which are listed in Table 4 from 1950 onwards. The ANFIS model performance was satisfactory at different ENSO events (Table 5) due to the training of the ANFIS model with long term data sets. The long term time series of ANFIS model simulation of historic ENSO events is shown in Figure 5.

Since SST, SLP and the trade wind index are the parameters related to the basic ENSO phenomenon in the equatorial Pacific, the trade wind index was also a sensitive input for the model developed from 1979 to 2011 (not shown). This indicates that hydrologic analysis in
ENSO-affected coastal areas can be better addressed using SST, SLP and the trade wind index. Because we wanted to run the model for 50 years to evaluate the performance of the model in different ENSO events, we could not include the trade wind index as model inputs as these data were available only after 1979.

The performance of the ANFIS model, for the given datasets in this watershed, has been compared with a physically based model, the Loading Simulation Program C++ (LSPC), and reported in Sharma (2012). The comprehensive model comparison of ANFIS with the LSPC model suggests that ANFIS can perform equally or better than this physically-based watershed model, especially when sufficient rain gauge stations are not available (Sharma 2012).

**ANFIS model and ENSO Indicators**

Having analyzed the model performance at different scales for different events, it can be concluded that the ANFIS model can simulate the watershed response well, partly due to the direct incorporation of the dominant ENSO phenomenon, and partly due to the ANFIS model skill in using ENSO inputs. Moreover, it was important to confirm that the improvement of the model simulation was due to the application of SST/SLP. This was shown in three ways. First, we assessed the performance of the ANFIS model with and without SST and SLP in the input vector, and found that the ANFIS model tended to perform better when SST and SLP were included (Table 6). Since the ENSO characteristic is completely different from season to season (Figure 4), the model developed for the entire season using SST did not improve the results significantly. This is mainly because SST/SLP has a positive correlation with streamflow during winter and spring but the opposite correlation with streamflow from August to October, i.e. separate models should be trained in winter, spring and from August to October. For this reason,
the second and third analyses were conducted primarily for the winter season in order to see the actual effect of SST/SLP on the streamflow simulation.

For the second case, we developed models with and without SST/SLP for the winter season, considering the precipitation from each rain gauge station one at a time. For example, we replaced the watershed rain gauge station by the precipitation from Fairhope (Figure 2). Unsurprisingly the simulation performance was poor (mobile, Coop ID-015478) as it was located approximately 37 km away from the rain gauge station in the study area (mobile, Coop ID-015478). However, when SST/SLP was considered as an additional input in addition to the Fairhope precipitation, the model performance improved (Table 7). In fact, this model performance is almost similar to the model performance when the watershed rain gauge station (mobile, Coop ID-015478) was used without SST/SLP. This indicates that precipitation from the farthest location can also be utilized, resulting in similar model performance provided that the SST and SLP are incorporated. Similarly, we used precipitation data from a rain gauge station located in mid Alabama (Auburn), which is 341.5 km from the watershed rain gauge station (Table 7). Since the actual rainfall characteristic of the watershed was not represented by the Auburn rainfall, the model simulation was poor. However, when SST/SLP was added to the same precipitation from Auburn, the overall model performance improved (Table 7). Similarly, we used rainfall data from Louisville and Meridian, which were the second and third furthest locations from the watershed (Figure 2), respectively. In each case, the model performance improved with the additional forcing of SST and SLP. Despite being closer to the watershed rain gauge station, two rain gauge stations, Laurel and Leakesville, which were at 135 and 60 km away from the watershed rain gauge station (mobile, Coop ID-015478), did not show better performance without SST and SLP, especially compared to one of the farthest rain gauge
stations, Auburn. However, performance improved in both cases after incorporating SST and SLP. The third remote location was Meridian where the performance without SST and SLP was not satisfactory. Even though this station is nearer than Auburn, the performance is not better. One reason could be the precipitation quality of the station. At times, two rain gauge stations which are virtually located in the same place but monitored and operated by different agencies can show poor correlations (Sharma et al. 2012a) indicating that a rain gauge station may not represent the true precipitation in the watershed if it is not maintained properly. Moreover, poor recording and missing data are quite often found in rain gauge stations. In this context, the additional forcing of SST and SLP became crucial to improve the simulation while using precipitation data from the Meridian station. The performance improved for the majority of cases and the effect of SST and SLP was more important, especially when rain gauge stations were considered from the remotest locations.

Even though the effect of SST/SLP improved the model performance in the majority of cases, we experimented further considering only El Niño and La Niña phases to further diagnose the effect of SST when precipitation from different locations was considered. For the third analysis, we repeated the same method but with the data taken from the El Niño and La Niña period because the neutral period is associated with tremendous uncertainty. The performance while using rainfall data from Pascagoula, Louisville, Leaksille and Laurel were poor, especially when ENSO indicators were removed from the input vectors (Table 8). However, their performance improved after incorporating ENSO indicators. This was true while using precipitation from the majority of the stations indicating the model tendency to perform better while incorporating SST and SLP (Table 8).
In addition, the winter season simulation was tested using a MLR model. It was found that the model improved in the winter season due to the direct application of SST and SLP. We want to clearly infer that application of SST and SLP is beneficial only for a period when ENSO shows a stronger signature with streamflow. The readers may also refer to the comprehensive analysis regarding the importance of SST in data-driven models in ENSO-affected regions (Sharma 2012). Since ENSO has a better signature with precipitation in winter and spring seasons but not in summer in the study watershed (Sharma et al. 2012b), we only tested the model performance with and without SST/SLP for these seasons.

Since there was only one rain gauge station (mobile, Coop ID-015478) in the Chickasaw Creek watershed, it is clear that the model developed for the entire season considering only one watershed rain gauge station suffered from the lack of spatially distributed precipitation data. This was clear through the close inspection of precipitation and corresponding streamflow over a long period of record. Some of the streamflow events did not correspond to the observed precipitation. There could be many reasons for the unusual streamflow for a given precipitation record. Some of the major reasons are: (1) the precipitation recorded at the station was erroneous on that day, (2) the precipitation measured at the station did not match with the actual precipitation that occurred in the watershed, (3) there could be error while measuring the observed data (e.g. precipitation and streamflow). SST and SLP are physically associated and bring about rainfall variation. Hence additional inclusion of these variables may not be required. However, the precipitation input in this model did not adequately represent the actual precipitation characteristic of the watershed. Hence, the underlying assumption was that by incorporating SST and SLP, it would be helpful to eliminate the error associated with point
measurements of rainfall, any recording errors in the precipitation, and missing precipitation data which were filled with data from the other precipitation stations.

**Summary and Conclusion**

In this paper, we implemented the SST and SLP difference between Tahiti and Darwin directly in the ANFIS model as model inputs and investigated model performance over 50 years of simulation. The performance of the ANFIS model was comprehensively evaluated with observed streamflow using different statistical criteria at different periods and seasons. Moreover, observed and simulated streamflow were compared in different La Niña and El Niño events throughout the 50 year simulation period. The model simulated the monthly streamflow satisfactorily, with reasonable accuracy. ANFIS models were developed with and without SST and SLP in order to evaluate the influence of additional forcing of SST and SLP in data driven models. Since the ENSO characteristics are opposite in nature from season to season, the model developed for the entire season did not improve significantly. However, the performance improved during the ENSO period in the winter season. The relevancy of applying SST and SLP was tested using precipitation data at various locations, developing 20 ANFIS models and considering 10 rain gauge stations. The performance of the ANFIS model improved when SST and SLP were applied irrespective of the rain gauge station locations. Moreover, the capability of the ANFIS model to capture the low and high flows during winter La Niña and El Niño events suggest the application of SST and SLP is useful for long term continuous streamflow simulation. This research opens a new window for future studies to develop neuro-fuzzy computational techniques by using long term recorded climate data in the equatorial Pacific for rainfall-runoff modeling to capture the dominant climate pattern, ENSO. As discussed in different literatures, there are various modeling approaches and the supremacy of one modeling
approach over another cannot be generalized. However, this research highlights the application of data-driven modeling approaches using SST and SLP data in ENSO-affected coastal regions, which is especially beneficial when the number of rain gauge stations is not sufficient. More importantly, it provides a viable alternative for streamflow simulation and justifies the relevancy of traditional data-driven modeling approaches.

Since the rainfall-runoff modeling that is applicable in much of the US is not applicable to the coastal plain (Sheridan et al. 2002), this approach is much more relevant for coastal watersheds that are affected by ENSO climate patterns.
Acknowledgement

The authors would like to acknowledge the corrections and suggestions made by the anonymous reviewers and the associate editor to improve this manuscript. The comments of the reviewers and associate editor were very helpful in addressing numerous issues which were ignored in our initial submission.
References


Table 1. Data used in the study with their sources and format

<table>
<thead>
<tr>
<th>Data</th>
<th>Source</th>
<th>Additional information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weather gauge station</td>
<td>NOAA, National Climatic Data Center,</td>
<td>Coop ID-015478</td>
</tr>
<tr>
<td></td>
<td><a href="http://ncdc.noaa.gov">http://ncdc.noaa.gov</a></td>
<td></td>
</tr>
<tr>
<td>Streamflow</td>
<td><a href="http://www.al.water.usgs.gov">www.al.water.usgs.gov</a></td>
<td>(USGS gage 02471001)</td>
</tr>
<tr>
<td>Hydrologic (stream network) data</td>
<td><a href="http://www.aces.edu/waterquality/gisdata">www.aces.edu/waterquality/gisdata</a></td>
<td></td>
</tr>
<tr>
<td>GIS layers/TIGER files</td>
<td><a href="http://www.AlabamaView.org">www.AlabamaView.org</a></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Input variables for the ANFIS model (monthly data)

<table>
<thead>
<tr>
<th>Model</th>
<th>Model inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR Model</td>
<td>SST (t-1)  SLP(t)  T(t-1)  T(t)  SR_{E}(t-1)  SR_{E}(t)</td>
</tr>
<tr>
<td>BF Model</td>
<td>SST(t)  SLP(t)  T(t-1)  P(t-1)  P (t)</td>
</tr>
</tbody>
</table>

Note: SR_{E}, T(t), P(t), and SST(t), SLP(t) denote “estimated surface runoff”, temperature, precipitation, sea surface temperature, and sea level pressure difference between Tahiti and Darwin at the month “t”, respectively.

Table 3. Performance of the ANFIS model in different modeling stages (monthly simulation from 1952-2002)

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Streamflow Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE (m^3/s)</td>
<td>2.63</td>
</tr>
<tr>
<td>NSE</td>
<td>0.8</td>
</tr>
<tr>
<td>MBE</td>
<td>0%</td>
</tr>
</tbody>
</table>
Table 4. El Niño and La Niña years (Dec-April) from 1950 to 2003.

<table>
<thead>
<tr>
<th>La Niña Year</th>
<th>El Niño Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1950</td>
<td>1958</td>
</tr>
<tr>
<td>1951</td>
<td>1966</td>
</tr>
<tr>
<td>1955</td>
<td>1969</td>
</tr>
<tr>
<td>1956</td>
<td>1973</td>
</tr>
<tr>
<td>1968</td>
<td>1983</td>
</tr>
<tr>
<td>1971</td>
<td>1987</td>
</tr>
<tr>
<td>1974</td>
<td>1988</td>
</tr>
<tr>
<td>1975</td>
<td>1992</td>
</tr>
<tr>
<td>1976</td>
<td>1995</td>
</tr>
<tr>
<td>1985</td>
<td>1998</td>
</tr>
<tr>
<td>1989</td>
<td>2003</td>
</tr>
<tr>
<td>1996</td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. ANFIS model performance for 25 ENSO events (monthly simulation)

For La Niña and El Niño events (since 1952)

<table>
<thead>
<tr>
<th>Statistics</th>
<th>ANFIS-Simulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSE</td>
<td>0.83</td>
</tr>
<tr>
<td>MBE</td>
<td>3.10</td>
</tr>
<tr>
<td>R²</td>
<td>0.83</td>
</tr>
</tbody>
</table>

14 La Niña events (Table 7)

<table>
<thead>
<tr>
<th>Statistics</th>
<th>ANFIS-Simulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSE</td>
<td>0.86</td>
</tr>
<tr>
<td>MBE</td>
<td>6%</td>
</tr>
<tr>
<td>R²</td>
<td>0.87</td>
</tr>
</tbody>
</table>

11 El Niño events (Table 7)

<table>
<thead>
<tr>
<th>Statistics</th>
<th>ANFIS-Simulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSE</td>
<td>0.76</td>
</tr>
<tr>
<td>MBE</td>
<td>-3.12%</td>
</tr>
<tr>
<td>R²</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Table 6. Performance of the ANFIS model during 50 years simulation at monthly time step

<table>
<thead>
<tr>
<th>Statistics</th>
<th>ANFIS (With SST and SLP)</th>
<th>ANFIS (Without SST and SLP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSE</td>
<td>0.79</td>
<td>0.74</td>
</tr>
<tr>
<td>MBE</td>
<td>0.60%</td>
<td>1.27%</td>
</tr>
<tr>
<td>R²</td>
<td>0.79</td>
<td>0.74</td>
</tr>
</tbody>
</table>
Table 7. Sensitivity analysis of the ANFIS model performance measured through $R^2$ using precipitation from various rain gauge stations for winter season with and without SST/SLP

<table>
<thead>
<tr>
<th>Stations</th>
<th>With SST/SLP</th>
<th>Without SST/SLP</th>
<th>Distance from Mobile Station (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Validation/Testing</td>
<td>Training</td>
</tr>
<tr>
<td>Bogalusa</td>
<td>0.82</td>
<td>0.1</td>
<td>0.54</td>
</tr>
<tr>
<td>Fairhope, AL</td>
<td>0.77</td>
<td>0.32</td>
<td>0.63</td>
</tr>
<tr>
<td>Laurel</td>
<td>0.73</td>
<td>0.06</td>
<td>0.44</td>
</tr>
<tr>
<td>Lakesville</td>
<td>0.75</td>
<td>0.35</td>
<td>0.60</td>
</tr>
<tr>
<td>Louisville</td>
<td>0.86</td>
<td>0.22</td>
<td>0.68</td>
</tr>
<tr>
<td>Meridian</td>
<td>0.84</td>
<td>0.26</td>
<td>0.52</td>
</tr>
<tr>
<td>Mobile, AL</td>
<td>0.91</td>
<td>0.52</td>
<td>0.70</td>
</tr>
<tr>
<td>Pascagoula, MS</td>
<td>0.72</td>
<td>0.58</td>
<td>0.56</td>
</tr>
<tr>
<td>Robertsdale</td>
<td>0.74</td>
<td>0.34</td>
<td>0.50</td>
</tr>
<tr>
<td>Auburn</td>
<td>0.68</td>
<td>0.28</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Table 8. Sensitivity analysis of the ANFIS model performance measured through $R^2$ using precipitation from various rain gauge stations during the El Niño and La Niña period of winter season with and without SST/SLP

<table>
<thead>
<tr>
<th>Stations</th>
<th>With SST/SLP</th>
<th>Without SST/SLP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training</td>
<td>Validation/Testing</td>
</tr>
<tr>
<td>Bogalusa</td>
<td>0.44</td>
<td>0.54</td>
</tr>
<tr>
<td>Fairhope, AL</td>
<td>0.66</td>
<td>0.6</td>
</tr>
<tr>
<td>Laurel</td>
<td>0.45</td>
<td>0.35</td>
</tr>
<tr>
<td>Leakesville</td>
<td>0.49</td>
<td>0.38</td>
</tr>
<tr>
<td>Louisville</td>
<td>0.42</td>
<td>0.14</td>
</tr>
<tr>
<td>Meridian</td>
<td>0.46</td>
<td>0.36</td>
</tr>
<tr>
<td>Mobile, AL</td>
<td>0.65</td>
<td>0.58</td>
</tr>
<tr>
<td>Pascagoula, MS</td>
<td>0.43</td>
<td>0.51</td>
</tr>
<tr>
<td>Robertsdale</td>
<td>0.41</td>
<td>0.51</td>
</tr>
<tr>
<td>Auburn</td>
<td>0.52</td>
<td>0.34</td>
</tr>
</tbody>
</table>
Figure 1. ANFIS architecture for two inputs (X and Y) with two rules and two membership functions (P1, P2 and Q1, Q2) for each rule.
Figure 2. Chickasaw Creek watershed in Mobile County of South Alabama, USA, showing the USGS gauging station and land use distribution within the watershed. The land use classification is as per NLCD land use categories.
Figure 3. The variation of monthly average precipitation and temperature in the El Niño and La Niña phase for (a) the winter (Jan-March) season and (b) for the spring (April-June) season.

Figure 4. Pattern of streamflow in two different ENSO phases.
Figure 5. Monthly streamflow simulated for 52 years using the ANFIS model compared with observed monthly streamflow.