Adaptive Neuro-Fuzzy Inference System for Drought Forecasting in the Cai River Basin in Vietnam

Nguyen, Luong Bang State Key Lab Hydrology Water Resources & Hydraulic Engineering, Hohai University | Thuyloi University

Li, Qiong Fang State Key Lab Hydrology Water Resources & Hydraulic Engineering, Hohai University

Ngoc, Trieu Anh Thuyloi University

Hiramatsu, Kazuaki Laboratory of Water Environment Engineering, Division of Bioproduction Environmental Sciences, Department of Agro-environmental Sciences, Faculty of Agriculture, Kyushu University

https://doi.org/10.5109/1543403

出版情報:九州大学大学院農学研究院紀要. 60 (2), pp.405-415, 2015-09-18. Faculty of Agriculture, Kyushu University バージョン: 権利関係:

Adaptive Neuro–Fuzzy Inference System for Drought Forecasting in the Cai River Basin in Vietnam

Luong Bang NGUYEN^{1,2}*, Qiong Fang LI¹, Trieu Anh NGOC² and Kazuaki HIRAMATSU

Laboratory of Water Environment Engineering, Division of Bioproduction Environmental Sciences, Department of Agro–environmental Sciences, Faculty of Agriculture, Kyushu University, Fukuoka 812–8581, Japan (Received April 6, 2015 and accepted May 19, 2015)

In order to achieve effective agricultural production, the impact of drought must be mitigated. An important requirement for mitigating the impact of drought is an effective method of forecasting future drought events. This paper presents the correlations between sea surface temperature anomalies (SSTA) and both the Standardized Precipitation Index (SPI) and the Standardized Precipitation Evapotranspiration Index (SPEI) at four areas monitoring El Nino–Southern Oscillation (ENSO) activities at the Cai River basin in Vietnam. The correlation analyses for selecting potential variables serves as a forecasting mechanism, and SSTAs events in NinoW and Nino4 zones are used to construct Adaptive Neuro–Fuzzy Inference System (ANFIS) forecasting models. Different ANFIS forecasting models for SPI and SPEI (1–, 3–, 6–, and 12–month) are trained and tested. The results of our research show that the best performing models are M5, M11, and M13. For drought forecasting in the short–term (1– or 3–month models), the SPI should be used, because it has a better performance than the SPEI. Drought forecasting with seasonal or long–term indexes (6– or 12–month models) should use the SPEI, because the SPEI performs better than SPI in these cases. We find that the ANFIS forecasting model (M11) for SPEI–12 is the best forecasting model. Furthermore, the ANFIS method with input variables constituting SSTA events can be successfully applied in order to establish accurate and reliable drought forecasting models.

Key words: Drought forecast, Adaptive Neuro–Fuzzy Inference System, SSTA, Cai River basin.

INTRODUCTION

Drought occurs around the globe, affecting human lives more than any other major natural hazard. It is widely considered to be the most complex and least understood of all the natural hazards (Dai et al., 2004; He et al., 2011). One of the major challenges presented by agricultural systems is how to mitigate the impact of droughts. Droughts affect agricultural systems economically as well as environmentally. With respect to economic impacts, droughts damage agricultural production, and can cause economic damage to industries that are connected to agricultural production, in addition to leading to unemployment because of reduced production. From an environmental perspective, droughts can deprive crops and soils of essential precipitation, as well as increasing the salt content in soils and irrigation systems (Mishra and Singh, 2010).

In order to mitigate the impact of drought, an effective and timely monitoring system is required. The effective monitoring of droughts can aid in the development of an early warning system. An objective evaluation of the drought condition in a particular area is the first step towards planning water resources, in order to prevent and mitigate the effects of future occurrences of drought. The evaluation and forecasting of drought is made possible by the use of drought indices. Several drought indices are commonly employed, such as the Palmer Index, the Crop Moisture Index, the Standardized Precipitation Index (SPI), and the Standardized Precipitation Evapotranspiration Index (SPEI). The Palmer Index and SPI are traditionally the most popular indices for forecasting drought, due to their standardization. For the purposes of comparing drought conditions from different areas, which often have different hydrological balances, the most important characteristic of a drought index is its standardization (Bonaccorso et al., 2003). Standardization of a drought index ensures independence from geographical position, as the index in question is calculated with respect to the average precipitation in the same location (Cacciamani et al., 2007). One of the differences between the Palmer Index and SPI is that the characteristics of the Palmer Index vary from site to site, while those of SPI do not. Another difference is that the Palmer Index has a complex structure with a very long memory, while SPI is an easily interpreted and simple moving average process (Tsakiris and Vangelis, 2004). This characteristic makes SPI useful as the primary drought index, because it is simple, spatially invariant in its interpretation, and probabilistic, allowing for its use in risk and decision making analyses. In addition, SPI is more representative of short-term precipitation than the Palmer Index, and is therefore a better indicator of soil moisture variation and soil wetness (Mishra and Singh, 2010). Given the focus of this paper on short-term drought, this characteristic makes the use of SPI advantageous. SPI also provides a better spatial standardization than the

¹ State Key Lab Hydrology Water Resources & Hydraulic Engineering, Hohai University, Nanjing, China

² Thuyloi University, Hanoi, Vietnam

^{*} Corresponding author (E-mail: bangdhtl@yahoo.com)

Palmer Index does, with respect to extreme drought events (Lloyd–Hughes and Saunders, 2002). Furthermore, SPI has been found to be better than the Palmer Index in detecting the onset of a drought event (Hayes et al., 2010). The SPEI index is calculated based on precipitation and evapotranspiration. Like SPI, SPEI is useful as the primary drought index because it is simple, spatially invariant in its interpretation, and probabilistic, and so can be used in risk and decision making analyses (Vicente-Serrano et al., 2010). Given the aforementioned advantages of SPI over the Palmer Index, and the shared characteristics of SPI and SPEI, the SPI and SPEI indices are chosen for application to drought forecasting in the present research.

El Nino-Southern Oscillation (ENSO), the most well-known coupled ocean-atmosphere phenomenon, is one of the most important regarding internal climate variations (Philander, 1990). El Nino refers to the phenomenon of abnormal increase in sea surface temperature (SST), from the equator to the east of the Pacific Ocean, which creates a hot water stream spreading from the equator along the coast of Peru and Ecuador to the south. This hot water stream often reaches peak intensity by the end of December (Wyrtki, 1975). Countering El Nino is the phenomenon in which the SST of the equatorial zone of the Pacific Ocean becomes exceptionally cold, which is known as La Nina, and causes climatic anomalies in many regions. Southern Oscillation (SO), regularly existing in the atmosphere around the South Pacific, is the cause of air exchange between the eastern and western hemispheres. This phenomenon was observed late in the previous century, prior to being described in detail by Walker and Bliss (1932) in terms of scale, specific characteristics, and confirmation of its relationship with the variation of temperature and rainfall in Oceania, South Asia, and some other regions of the world. This was based on SSTA for the observation of ENSO activities, at four specific zones of the equatorial Pacific Ocean; namely, the Nino12 zone (0°N-10°S; 90°W-80°W), Nino3 zone (5°N-5°S; 150°W-90°W), Nino 4 zone (5°N-5°S; 160°E-150°W), and NinoW zone (15°N-0°S; 130°E–150°E). In these four zones, the temperature change tendency in the NinoW zone often opposes that of the remaining three zones. According to absolute values, standard deviations will reduce gradually from the Nino12 zone to the NinoW zone (Xiao and Mechoso, 2009; Balmaseda *et al.*, 2009).

The area that the research was conducted lay on Indochina, whose climate is directly affected by the South Asian summer monsoon (SASM) (Chen and Yoon, 2000; Zhang *et al.*, 2002; Yanai *et al.*, 1992). Because SASM is known to interact with ENSO (Walker, 1924; Yasunari, 1990; Chung and Nigam, 1999), changes in SSTA events of the Pacific Ocean will effect SASM, and alter the precipitation of Indochina (Goswami *et al.*, 1999; Chen and Yoon, 2000; Wang *et al.*, 2003). Therefore, in this research, the author will study the correlation between SSTA at four observation zones for ENSO activity, and the SPI and SPE indices. Correlation analyses for the selection of potential variables serves as a method of drought forecast.

New techniques, such as artificial neural networks (ANN), Fuzzy Logic (FL), and Adaptive Neuro-Fuzzy Inference Systems (ANFIS), have recently been accepted as efficient alternative tools for the modeling of complex hydrologic systems, and widely employed for forecasting. Some specific applications of ANN to hydrology include the modeling of the rainfall-runoff process (Jeong and Kim, 2005; Senthil Kumar et al., 2005; Rajurkar et al., 2004), hydrologic time series modeling (Jain and Kumar, 2007), sediment concentration estimation (Nagy et al., 2002), the estimation of heterogeneous aquifer parameters (Mantoglou, 2003), and runoff and sediment yield modeling (Agarwal et al., 2006). Morid et al. (2007) examined the utility of the ANN approach for medium and long-term forecasting of both the likelihood of drought events and their severity. Mishra and Desai (2006) applied the feed-forward recursive neural network and ARIMA models for drought forecasting using the SPI series as a drought index. The results have demonstrated that the neural network method can be successfully applied to drought forecasting. Moreover, ASCE Task Committee reports (ASCE, 2000) carried out a comprehensive review of applications of ANN in the context of hydrological forecasting. In addition, several studies have been carried out using FL in hydrology and water resources planning (Mahabir et al., 2003; Liong et al., 2000; Nayak et al., 2005; Altunkaynak et al., 2005). In recent years, ANFIS, which constitutes an integration of the ANN and FL methods, has been applied to the modeling of nonlinear engineering and water resources problems (Chang and Chang, 2006; Nayak et al., 2004; en and Altunkaynak, 2006; Firat and Güngör, 2007; 2008). Furthermore, Chou and Chen (2007) have used neurofuzzy computing techniques for the development of a drought early warning index. With this aim, an approach has been proposed for the development of a drought early warning index (DEWI) in southern Taiwan, to detect droughts in advance in order to set up proper plans for reducing the impact of water shortages.

Traditional methods, such as regression analysis and autoregressive moving average models, are commonly used in the estimation of hydrological processes. However, the FL and ANN methods offer real advantages over conventional modeling, especially when the underlying physical relationships are not fully understood. FL is employed in order to describe human thinking and reasoning, within a mathematical framework. The main problem with FL is that no systematic procedure exists to define the MF parameters and to design fuzzy rules. To construct fuzzy rules it is necessary to define premises and consequences as fuzzy sets. In this paper, an Adaptive Neuro-Fuzzy Inference System (ANFIS), which is an integration of the ANN and FL methods, is proposed as an alternative to the traditional methods for drought forecasting, using SPI/SPEI for multiple time scales. The main contribution of the ANFIS method is that it eliminates the basic problems in fuzzy modeling (the definition of the membership function parameters and the design of fuzzy if-then rules), using the learning capability of

ANN for automatic fuzzy rule generation and parameter optimization. SSTA events are used to construct the ANFIS forecasting models, and the best fit forecasting model structure is determined by comparing the forecasted and observed values.

MATERIALS AND METHODS

Study area

The basin that we study lies between 12°02′49″-12°36′13″N and 108°40′03″-109°11′38″E in the Khanhhoa Province, Vietnam, with a total area of 1,889 km² (Fig. 1). This basin is located in a tropical monsoon zone that exhibits rather unique deformational features, and an oceanic climate. The average rainfall in the period 1982-2012 was 1,616 mm, and there are two distinct seasons: rainy and dry. The rainy season usually lasts from May to December, and rainfall is largely concentrated in September, October, and November, comprising 55% of the average annual rainfall. During many dry season months, there is no rain. With a high temperature foundation, the average annual temperature during many years is 26.7°C, and the difference in temperature between the months is relatively small. The potential evaporation in this region is high, averaging approximately 1,200-1,600 mm/year.



Fig. 1. Cai River basin, Vietnam.

Datasets

Rainfall and temperature datasets

Rainfall and temperature data were recorded at the Khanhvinh Meteorological Station in the Khanhhoa Province, Vietnam. These datasets are under the management of the Vietnam Institute of Meteorology and Hydrography. The data quality is assured, and reliable enough for calculations. The period covered by the data that was mined for use in the research was January 1983 to December 2012.

SSTA datasets

The Four SSTA datasets in the Nino12, Nino3, Nino4, and NinoW zones were standardized by dividing standard deviations. The data was collected from 1982 to 2012, a duration of 31 years. SSTA data was collected from the Comprehensive Ocean Atmosphere Data Set (COADS) (Reynolds *et al.*, 2002), via the website (APDRC, 2013). For the purposes of this research, the field of SSTA used in the Nino12, Nino3, Nino4, and NinoW zones was within the corresponding coordinates of $(0^{\circ}N-10^{\circ}S; 90^{\circ}E-80^{\circ}W)$, $(5^{\circ}N-5^{\circ}S; 150^{\circ}W-90^{\circ}W)$, $(5^{\circ}N-5^{\circ}S; 160^{\circ}E-150^{\circ}W)$, and $(15^{\circ}N-0^{\circ}S; 130^{\circ}E-150^{\circ}E)$. A 2° latitude × 2° longitude resolution was applied, by using averaged computation from the original 1×1 grid point datasets.

Correlation between the SSTA and SPI/SPEI drought indices

The SPI and SPEI drought indices

SPI (Standardized Precipitation Index) was first introduced by McKee et al. (1993). Based on the high conformance of the gamma distribution with rain data indices over time in many places, McKee et al. (1993) developed SPI in the form of a stochastic variable with a normal distribution. The SPI index is based on precipitation alone, making its evaluation relatively easy in comparison with other drought indices, namely the Palmer Index and the crop moisture index (Cacciamani et al., 2007). In addition to the advantages mentioned earlier, a major benefit of the SPI index is that it makes it possible to describe drought on multiple time scales (Cacciamani et al., 2007; Mishra and Desai, 2006; Tsakiris and Vangelis, 2004). SPEI (Standardized Precipitation Evapotranspiration Index) was introduced by Vicente-Serrano et al. (2010), in order to study and follow up on the drought process under the effect of the global warming phenomenon. It is based on the index (D), the difference between rainfall (P) and potential evaporation (PET). SPEI is calculated in the same way as SPI.

SPI/SPEI constitutes an easily interpreted, simple moving average process. This characteristic makes SPI/ SPEI useful as the primary drought index, because it is simple, spatially invariant in its interpretation, and probabilistic, allowing it to be used in risk and decision making analysis. For these reasons, it was selected for this study.

In this study, the author used the SPI/SPEI program available on the Website (Beguería and Vicente Serrano, 2009), in order to calculate SPI and SPEI at different time scales (i.e., 1–, 3–, 6–, and 12–month) (WMO, 2012). According to the World Meteorological Organization (WMO), (drought or wet) climatic conditions have been classified as per SPI and SPEI as presented in Table 1 (WMO, 2012).

Correlation between SSTA and SPI/SPEI

Correlation equations are widely used in sciences today as a measure of linear dependence between two

SPI/SPEI	Classification	SPI/SPEI	Classification
2.00 or more	Extremely wet	-0.50 to -0.99	Mild drought
1.50 to 1.99	Very wet	-1.00 to -1.49	Moderate drought
1.00 to 1.49	Moderately wet	-1.5 to -1.99	Severe drought
0.50 to 0.99	Mildly wet	-2,0 or less	Extreme drought
-0.49 to 0.49	Normal		

Table 1. Standard Precipitation Index (SPI) and Standardized Precipitation Evapotranspiration Index (SPEI) classifications

variances. This procedure was developed by Karl Pearson, from a related idea that was introduced by Francis Galton in the 1880s (Rodgers and Nicewander, 1988; Stigler, 1989). We use the correlation equation between SSTA and SPI/SPEI for determining correlation coefficients. If the correlation coefficient is positive (reflecting a covariant relationship, then the absolute value of correlation coefficient is greater, and the level of linear dependence between SSTA and SPI/SPEI is greater. The reverse holds if the correlation coefficient is negative (reflecting a counter covariant relationship).

We will employ the correlation between an SSTA data chain $x_1, x_2, ..., x_n$ and an SPI/SPEI data chain $y_1, y_2, ..., y_n$, with n observatory value pairs $\{x_i, y_i\}$ (i=1,..., n), to establish a correlation matrix between SSTA and SPI/SPEI. The data chains used for our calculation are described as follows:

(1) Our SSTA data chain includes factors from SSTA in the Nino12, Nino3, Nino4, and NinoW zones with time scales of 1–, 3–, 6–, and 12–months. These are designated for the respective zones as N12_j, N3_j, N4_j, and NW_j. The index signs j = 1, 2, ..., 12 indicate SSTA data at the jth time (j = 1 at SSTA indicates the same period as SPI/SPEI, j = 2 indicates SSTA at 1 month ahead of SPI/SPEI, j =12 indicates SSTA at 11 months prior to SPI/SPEI).

(2) The comparative element series is given by the SPI/SPEI value series at the case study site, whose objects are denoted by SPI_k , $SPEI_k$, where index k = 1, 2, 3, 4. Here, k = 1 when SPI/SPEI is for 1-month time scales, k = 2 when SPI/SPEI is for 3-month time scales, k = 3 when SPI/SPEI is for 6-month time scales, and k = 4



Fig. 2. Correlation between SSTA and SPEI/SPI.

when SPI/SPEI is for 12-month time scales.

The correlation matrix results between SSTA and SPI/SPEI presented in Fig. 2 show that the correlation between SSTA and SPI is greater than the correlation between SSTA and SPEI. However, there is one notable shared characteristic. That is, the correlation between SSTA and SPI/SPEI decreases gradually from NinoW to Nino4, then to Nino3, and is smallest in Nino12 zones. The correlation between SSTA and SPI/SPEI in NinoW is positive, while it is negative in most other cases. The results in Fig. 2 also show that the correlation between SSTA and SPI/SPEI over 12–month time scales is most often the greatest, while it is smallest over 1–month time scales.

The results for the correlation coefficient value between the SSTA and SPI/SPEI indices show that the correlation at Nino3 and Nino2 zones is weak, while the correlation at NinoW and Nino4 zones is relatively high. Furthermore, the highest correlation coefficient always occurs when SPI/SPEI is larger than SSTA from a range of 1 to 3 months. It can be seen that the level of linear dependence between SPI/SPEI and SSTA in the NinoW and Nino 4 zones is relatively high. Therefore, in this research we will use SSTA data from these zones with a high coefficient of correlation with SPI/SPEI as input variables to build drought forecast models for the area under consideration.

ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM (ANFIS) MODELS

The ANFIS model, proposed by Jang (Jang (1993); Jang et al. (1997)), has been applied in order to study many problems. The ANFIS model is based on a fuzzy interface system, which is trained by a learning algorithm derived from neural network theory. A detailed treatment of ANFIS is provided by Ying (2000). The ANFIS of Sugeno-type systems with three inputs and two rules, which was used in this study, is shown in Fig. 3. The square and circle nodes are used to reflect different properties of adaptive learning. Square nodes (adaptive nodes) have parameters, while circle nodes (fixed nodes) do not. Each node has an own node function. The node functions vary from node to node. The connection between the two nodes indicates the direction of signal. No weight is associated with the connection.

In Fig. 3, the ANFIS has three input variables: $SSTA_1$ (indicated by x), $SSTA_2$ (y), and $SSTA_3$ (z); and one out-



Fig. 3. The scheme of the Adaptive Neuro-Fuzzy Inference System.

put variable: SPI/SPEI. Each input variable has two membership functions, given by (A_1, A_2) , (B_1, B_2) , and (C_1, C_2) , respectively.

The Sugeno-type fuzzy if – then rule for linear output is set up as:

Rule1: if x equal to A₁, y equal to B₁, and z equal to C₁,
then
$$f_1 = p_1 x + q_1 y + r_2 z + s_1$$
 (1)

Rule2: if x equal to A₂, y equal to B₂, and z equal to C₂, then $f_2 = p_2 x + q_2 y + r_2 z + s_2$ (2)

In the above, A_i , B_i , and C_i are fuzzy sets (a fuzzy set is uniquely specified by its membership function); f_i is the output within the fuzzy region; and p_i , q_i , r_i , and s_i are the consequence parameters determined during the training process, where i runs from one to two.

A membership function (MF) is a curve that defines how strongly a point $\mathbf{x}' \in X$ belongs to a set, by assigning a membership degree between 0 and 1, $\mu_z'(\mathbf{x}') : X \to \{0,1\}$. There are several possible MFs, such as a bell MF, Gaussian MF, two–side Gaussian MF, triangular MF, trapezoidal MF, and pi–shape MF. In the paper, the Gaussian membership function is introduced as an example. The Gaussian MF takes the form:

$$\mu_{z'}(x') = f(x',\sigma) = \exp(-(x' c/\sigma)^{2}), \qquad (3)$$

where x' is the value of the input to node i, and c and σ are the center and the width of the Gaussian curve of the fuzzy set Z', respectively. c and σ are also called the premise parameters.

The first layer of Fig. 3 is contains adaptive nodes represented by i, whose outputs are computed by the function member:

$$O_{1,i} = \mu_{Ai}(x), i = 1, 2$$
 (4)

$$O_{1,i} = \mu_{Bi-2}(y), \, i = 3, 4 \tag{5}$$

$$O_{1,i} = \mu_{Ci-4}(z), \, i = 5, \, 6 \tag{6}$$

where $A_i, B_{i\cdot 2}, C_{i-4}$ and are linguistic variables (in this study, the linguistic variables are 'low', 'medium', and 'high') associated with that node, and μ is the membership function. As such, $O_{1,i}$ represents the membership grade of a fuzzy set $Z'(=A_1, A_2, B_1, B_2, C_1, \text{ or } C_2)$, and

specifies the degree to which each input x, y, z satisfies the fuzzy set A.

The second layer is the rule layer. Every node in this layer is a fixed node, indicated as Π , which are called rule nodes. An output from the rule nodes represents a product of the input signals. That is, the fixed node receives inputs from the respective adaptive nodes, and computes the firing strength of the given rule:

$$O_{2,i} = w_1 = \mu_{Ai}(x)\mu_{Bi}(y)\mu_{Ci}(z), i = 1, 2$$
(7)

In the third layer, every node is a fixed node, and is indicated as N. The number of nodes in this layer is the same number nodes as in the rule layer. The ith node in this layer receives inputs from all nodes in the rule layer, and calculates the normalized firing strength of the rules:

$$O_{3,i} = \overline{w}_1 = \frac{w_i}{\Sigma w_i}, i = 1, 2$$
(8)

In the fourth layer, every node is an adaptive node, and is indicated as Z. The number of nodes in this layer is the same as the number of nodes in the third layer. The output from each node is the weighted consequent value of a given rule:

$$O_{4,i} = \overline{w}_1 f_1 = \frac{w_i}{\Sigma w_i} (p_1 x + q_1 + r_1 z + s_i), i = 1, 2 \quad (9)$$

where \overline{w}_1 is a normalized firing strength from the third layer, and $\{p_i, q_i, r_i, s_i\}$ represents the sequence of linear parameters of node i.

The fifth layer contains only one output node, and is called the summation layer. This single node is a fixed node, and indicated as Σ . This node computes the overall ANFIS output, as the sum of the outputs of all the adaptive nodes in the fourth layer, SPI/SPEI:

$$O_{5,i} = \sum_{i=1}^{2} O_{4,i} = \sum_{i=1}^{2} \overline{w}_{i} f_{i} = \overline{w}_{i} f_{1} + \overline{w}_{i} f_{2} = \frac{\sum_{i=1}^{2} w_{i} f_{i}}{\sum_{i=1}^{2} w_{i}}$$
(10)
SPI/SPEI= $f(x, y, z) = \frac{w_{1} f_{1} + w_{2} f_{2}}{w_{1} + w_{2}} =$

$$= \frac{w_{1}(SSTA_{1}, SSTA_{2}, SSTA_{3}) f_{1} (SSTA_{1}, SSTA_{2}, SSTA_{3}) + w_{1}(SSTA_{1}, SSTA_{2}, SSTA_{3}) f_{2} (SSTA_{1}, SSTA_{2}, SSTA_{3}) f_{2} (SSTA_{1}, SSTA_{2}, SSTA_{3})$$

ANFIS uses a hybrid learning algorithm to calibrate the network. The combination of a back-propagation algorithm with a least-squares approximation or back propagation algorithm is used in the hybrid learning algorithm to optimize the parameters in layers 1 and 4, respectively. The mathematical details of these algorithms are given in Jang *et al.* (1997), Nayak *et al.* (2004), and Bacanli *et al.* (2009).

DROUGHT FORECASTING BY ANFIS

Input variables

In this study, SSTA events (1-, 3-, 6-, and 12-month) are used to generate a drought estimation model with the ANFIS method. For this, SPI/SPEI (1-, 3-, 6-, and 12-month) outputs are considered. In the construction of the estimation models, again, different models are generated for each of the SPI/SPEI (1-, 3-, 6-, and 12-month) outputs. The datasets are divided into two subsets: training and testing datasets. The training dataset includes data records measured between the years 1983 and 2008. In order to achieve a more reliable evaluation and comparison, the models are tested by the evaluation if a dataset that was not used during the training process. The testing dataset consists of data records observed between 2009 and 2012.

Model structures

One of the most important steps in developing a satisfactory forecasting model is the selection of the input variables. This is because these variables determine the structure of the forecasting model, and affect the weighted coefficients and the results of the model. Presently, different estimation models are constructed for each phase. The models for 1, 3, 6, and 12 months are named as SPI/SPEI–1, SPI/SPEI–3, SPI/SPEI–6, and SPI/ SPEI–12, respectively. Here, SPI/SPEI–1, SPI/SPEI–3, and SPI/SPEI–6 are considered as the indices for short term or seasonal variation, and SPI/SPEI–12 is considered as the long-term drought index. Eighteen models with different input numbers and structures are constructed for each phase, using these variables. In this study, forecasting models based on various combinations of SSTA events in NinoW and Nino4 are constructed (Table 2). The SSTA events in NinoW and Nino4 are named as $SSTA_w$ and $SSTA_4$.

In each model, every input variable must be clustered into several class values in layer 1, to build up fuzzy rules. In addition, each fuzzy rule will be constructed through several parameters of the membership function in layer 2. As the number of parameters increases with the fuzzy rule increment, the model structure becomes more complicated. In this study, the subtractive fuzzy clustering function is used to establish the fuzzy rule, based on the relationship between the input-output variables. In order to determine the nonlinear input and linear output parameters, a hybrid algorithm was used. The learning procedure and the construction of the rules are provided by this algorithm. The performance of the ANFIS models for the training and testing datasets is evaluated according to statistical criteria, such as the Correlation Coefficient (CORR), Efficiency (E), and Root Mean Square Error (RMSE). CORR is a commonly used statistic, providing information on the strength of the linear relationship between the observed and computed values. E is one of the most widely employed statistics for evaluating model performance. Values of CORR and E close to 1.0 indicate a good model performance. The RMSE statistic indicates a model's ability to predict a value separated from the mean.

RESULTS AND DISCUSSION

When comparing the performances of the ANFIS models for SPI/SPEI, it is seen that the performance of

Model	Input structure	Output
M1	$SSTA_w(t-1), SSTA_w(t-2)$	SPI(t), SPEI(t)
M2	$SSTA_w(t-1), SSTA_w(t-2), SSTA_w(t-3)$	SPI(t), SPEI(t)
M3	$SSTA_w(t-1), SSTA_w(t-2), SSTA_w(t-3), SSTA_w(t-4)$	SPI(t), SPEI(t)
M4	$SSTA_w(t-1), SSTA_w(t-2), SSTA_w(t-3), SSTA_w(t-4), SSTA_w(t-5)$	SPI(t), SPEI(t)
M5	$SSTA_w(t-1), SSTA_w(t-2), SSTA_w(t-3), SSTA_w(t-4), SSTA_w(t-5), SSTA_w(t-6)$	SPI(t), SPEI(t)
M6	$SSTA_w(t-2), SSTA_w(t-3), SSTA_w(t-4), SSTA_w(t-5)$	SPI(t), SPEI(t)
M7	$SSTA_w(t-2), SSTA_w(t-2), SSTA_w(t-4), SSTA_w(t-5), SSTA_w(t-6)$	SPI(t), SPEI(t)
M8	$\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!2),\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!3),\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!4),\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!5),\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!6),\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!7)$	SPI(t), SPEI(t)
M9	$SSTA_w(t-1), SSTA_w(t-2), SSTA_w(t-3), SSTA_4(t-1)$	SPI(t), SPEI(t)
M10	$\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!1),\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!2),\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!3),\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!4),\mathrm{SSTA}_{\!\scriptscriptstyle 4}(t\!-\!1)$	SPI(t), SPEI(t)
M11	$\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!1), \mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!2), \mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!3), \mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!4) \ \mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!5), \mathrm{SSTA}_{\!\scriptscriptstyle 4}(t\!-\!1)$	SPI(t), SPEI(t)
M12	$SSTA_w(t-1), SSTA_w(t-2), SSTA_w(t-3), SSTA_i(t-1)), SSTA_i(t-2)$	SPI(t), SPEI(t)
M13	$\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!1),\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!2),\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!3),\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!4),\mathrm{SSTA}_{\!\scriptscriptstyle 4}(t\!-\!1),\mathrm{SSTA}_{\!\scriptscriptstyle 4}(t\!-\!2)$	SPI(t), SPEI(t)
M14	$SSTA_w(t-2), SSTA_w(t-3), SSTA_4(t-1), SSTA_4(t-2)$	SPI(t), SPEI(t)
M15	$\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!2),\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!3),\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!4),\mathrm{SSTA}_{\!\scriptscriptstyle 4}(t\!-\!1),\mathrm{SSTA}_{\!\scriptscriptstyle 4}(t\!-\!2)$	SPI(t), SPEI(t)
M16	$SSTA_w(t-2), SSTA_w(t-3), SSTA_w(t-4), SSTA_i(t-2)$	SPI(t), SPEI(t)
M17	$\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!2),\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!3),\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!4),\mathrm{SSTA}_{\!\scriptscriptstyle W}(t\!-\!5),\mathrm{SSTA}_{\!\scriptscriptstyle 4}(t\!-\!2)$	SPI(t), SPEI(t)
M18	$SSTA_w(t-2), SSTA_w(t-3), SSTA_4(t-2), SSTA_4(t-3)$	SPI(t), SPEI(t)

Table 2. Structures of the forecasting models

the ANFIS model for SPI/SPEI–12 is better than that of the models for SPI/SPEI–1 to SPI/SPEI–6. However, the ANFIS models for SPI–1 and SPI–3 show a higher performance than for SPEI–1 and SPEI–3. Furthermore, the ANFIS models for SPI–6 and SPI–12 show an even lower performance. The average performance of all models is shown in Table 3. When the results for the ANFIS models are compared, models with small input variables usually show a lower performance. The model with the highest performance is most often a model with a high number of input variables, such as M5, M11, or M13. The ANFIS model for SPI/SPEI-1 with the highest performance is M5. For SPI/SPEI-3 and SPI/SPEI-12, it is M11, and for SPI/

Drought Index -		Training set			Testing set	
	CORR	Ε	RMSE	CORR	Ε	RMSE
SPI-1	0.73	0.55	0.58	0.45	0.16	0.71
SPI-3	0.79	0.63	0.58	0.65	0.36	0.52
SPI-6	0.84	0.70	0.53	0.75	0.47	0.39
SPI-12	0.88	0.77	0.47	0.68	0.51	0.30
SPEI-1	0.71	0.54	0.63	0.47	0.19	0.82
SPEI-3	0.78	0.63	0.58	0.69	0.38	0.63
SPEI-6	0.84	0.72	0.51	0.75	0.48	0.46
SPEI-12	0.88	0.77	0.46	0.60	0.53	0.32

Table 3. Average performances of all the models

Table 4. Performances of the most suitable models for SPI/SPEI

Drought Index (best model)		Training set		_	Testing set	
	CORR	Е	RMSE	CORR	Е	RMSE
SPI-1 (M5)	0.9402	0.8827	0.3098	0.45	0.16	0.71
SPEI-1 (M5)	0.9089	0.8251	0.4027	0.65	0.36	0.52
SPI-3 (M11)	0.9451	0.8923	0.3231	0.75	0.47	0.39
SPEI-3 (M11)	0.9448	0.8919	0.3213	0.68	0.51	0.30
SPI-6 (M13)	0.9527	0.9071	0.3052	0.47	0.19	0.82
SPEI-6 (M13)	0.9572	0.9158	0.2855	0.69	0.38	0.63
SPI-12 (M11)	0.9486	0.8997	0.3173	0.75	0.48	0.46
SPEI-12 (M11)	0.9598	0.9210	0.2775	0.60	0.53	0.32



Fig. 4. Comparison of the performances of the ANFIS Models for SPI/SPEI-6.



Fig. 5. Performances of the ANFIS models for SPI/SPEI-1, SPI/SPEI-3, SPI/SPEI-6, and SPI/SPEI-12.

SPEI-6, it is M13. The performances for the most suitable models are shown in Table 4.

As it would require too much space to show the model results for each phase, having 18 models overall, only the results for SPI/SPEI–6 are presented. The training performances of the ANFIS models for SPI/SPEI–6 are given in Fig. 4. This figure shows that the performance of the ANFIS Models for SPEI–6 is higher than for SPI–6, but there is also a shared feature, in that the highest performance is in the M13 case.

Fig. 5 shows the performances of the best fit models

for the data from 1-month to 12-month (SPI/SPEI-1 to SPI/SPEI-12). In this figure, the variations of the CORR and RMSE criteria for SPI/SPEI-1 to SPI/SPEI-12 during the training period are presented.

The values of CORR and E for the ANFIS models for SPI/SPEI–1 are lower than those for the other models. The reason that the ANFIS models developed using the SPI/SPEI outputs of 12 months show a better performance is that SPI/SPEI values calculated for such a long term include dry and wet periods for a longer duration. Short-term periods, such as one or three months, may



Fig. 6. The results of the ANFIS models for SPI; (a) SPI-1; (b) SPI-3; (c) SPI-6; (d) SPI-12.

include a wet or a dry period for only a short time. For example, in a three months period, drought occurs more frequently and for a shorter time, and when the period increases the duration of drought increases, but its frequency decreases. This means that for shorter periods, the SPI/SPEI values may contain a one-month dry and a one-month wet period, and this causes instability. Passages between positive and negative values occur more frequently, and this also results in the instability. For this reason, the ANFIS estimation models constructed with SPI/SPEI values calculated for shorter periods cannot catch dry and wet periods, and therefore yield unsuccessful results. In addition, the SPI/SPEI outputs for 12 months have a more stable run. Therefore, the ANFIS models developed using SPI/SPEI outputs for 12 months can catch dry and wet periods and yield better results. Fig. 6 shows the results for the best fit models for SPI-1 to SPI-12, and Fig. 7 shows the results of the best fit models for SPEI-1 to SPEI-12.

From Fig. 6 and Fig. 7 is it seen that for forecasting droughts with a short-term index (1–, 3–month), the SPI index delivers a higher performance than SPEI, and for





forecasting drought with a seasonal or long-term index (6-, 12-month), the SPEI index performs better than SPI. The M11 model SPEI-12 shows the best performance. As a result, we conclude that that ANFIS can be successfully applied and provide a high accuracy and reliability for drought forecasting.

CONCLUSIONS

SPI /SPEI is one of the most widely employed methods in relation to drought. An accurate and reliable estimation of SPI/SPEI is highly important. Traditional methods, such as regression analysis and autoregressive moving average models, are commonly used in the estimation of hydrological processes. In this paper, Adaptive Neuro– Fuzzy Inference System (ANFIS) was proposed as an alternative drought–forecasting tool to the traditional methods.

In order to illustrate the applicability of the ANFIS method in forecasting drought, SSTA data at NinoW and Nino4 zones was selected for input variables, in order to forecast drought in the region chosen for our research. Various ANFIS forecasting models for SPI/SPEI (1–, 3–, 6–, and 12–month) were trained and tested, and the results showed that the models with the highest performance are M5, M11, and M13. The ANFIS models for SPI/SPEI–1 that deliver the highest performance are M5, for SPI/SPEI–3 and SPI/SPEI–12 they are M11, and for SPI/SPEI–6 they are M13.

The ANFIS models for SPI/SPEI-12 perform better than those for SPI/SPEI-1 to SPI/SPEI-6. The reason this is that SPI/SPEI values calculated for long periods contain longer dry and wet periods. This means that for shorter periods, the SPI/SPEI values may contain only a 1-month dry or 1-month wet period, and this causes instability. In addition, passages between positive and negative values occur more frequently, and this also results in instability. For this reason, the ANFIS estimation models constructed with SPI/SPEI values calculated for shorter periods cannot catch dry and wet periods, and yield unsuccessful results. In addition, the SPI/SPEI outputs for 12 months have more stable run. Therefore, the ANFIS models developed using SPI/SPEI outputs for 12 months can catch dry and wet periods, and yield better results.

For forecasting drought with a short-term index (1-, 3-month), the SPI index performs better than SPEI, and for forecasting drought with a seasonal or long-term index (6-, 12-month), the SPEI index performs better than SPI. The M11 model SPEI-12 gives the best forecasting performance. As a result, it we can conclude that the ANFIS method with SSTA events as input variables can be successfully applied to establish accurate and reliable drought forecasting models.

REFERENCES

Agarwal, A., S. K. Mishra, S. Ram and J. K. Singh 2006 Simulation of Runoff and Sediment Yield using Artificial Neural Networks. *Biosystems Engineering*, 94(4): 597–613

- Altunkaynak, A., M. Özger and M. Çakmakci 2005 Water Consumption Prediction of Istanbul City by Using Fuzzy Logic Approach. Water Resources Management, 19(5): 641–654
- APDRC 2013 APDRC LAS7 for public (SSTA). Retrieved from http://apdrc.soest.hawaii.edu/las/v6/constrain?var=295
- ASCE 2000 Artificial Neural Networks in Hydrology. II: Hydrologic Applications. Journal of Hydrologic Engineering, 5(2): 124–137
- Bacanli, U., M. Firat and F. Dikbas 2009 Adaptive Neuro–Fuzzy Inference System for drought forecasting. Stochastic Environmental Research and Risk Assessment, 23(8): 1143–1154
- Balmaseda, M. A., O. J. Alves, A. Arribas, T. Awaji, D. W. Behringer, N. Ferry, Y. Fujii, T. Lee, M. Rienecker, T. Rosati and a. D. Stammer 2009 Ocean initialization for seasonal forecasts. *Oceanography*, **22**(3): 154–159
- Beguería, S. and S. M. Vicente Serrano 2009 SPEI and SPI Calculator. Retrieved from http://digital.csic.es/handle/10261/ 10002
- Bonaccorso, B., I. Bordi, A. Cancelliere, G. Rossi and A. Sutera 2003 Spatial Variability of Drought: An Analysis of the SPI in Sicily. Water Resources Management, 17(4): 273–296
- Cacciamani, C., A. Morgillo, S. Marchesi and V. Pavan 2007 Monitoring and Forecasting Drought on a Regional Scale: Emilia–Romagna Region. *In* "Methods and Tools for Drought Analysis and Management", Vol. 62, ed. by Rossi, G., T. Vega and B. Bonaccorso, Springer Netherlands, pp. 29–48
- Chang, F.–J. and Y.–T. Chang 2006 Adaptive neuro–fuzzy inference system for prediction of water level in reservoir. *Advances in Water Resources*, **29**(1): 1–10
- Chen, T.–C. and J.–h. Yoon 2000 Interannual Variation in Indochina Summer Monsoon Rainfall: Possible Mechanism. *Journal of Climate*, **13**(11): 1979–1986
- Chou, F. N. F. and B. P. T. Chen 2007 Development of Drought Early Warning Index: Using Neuro–Fuzzy Computing Technique. In "8th international symposium on advanced intelligence systems 2007", Paper No: A1469, Korea
- Chung, C. and S. Nigam 1999 Asian Summer Monsoon—ENSO Feedback on the Cane–Zebiak Model ENSO. Journal of Climate, 12(9): 2787–2807
- Dai, A., K. E. Trenberth and T. Qian 2004 A Global Dataset of Palmer Drought Severity Index for 1870–2002: Relationship with Soil Moisture and Effects of Surface Warming. *Journal* of Hydrometeorology, 5(6): 1117–1130
- Firat, M. and M. Güngör 2007 River flow estimation using adaptive neuro fuzzy inference system. *Mathematics and Computers in Simulation*, **75**(3–4): 87–96
- Firat, M. and M. Güngör 2008 Hydrological time-series modelling using an adaptive neuro-fuzzy inference system. *Hydrological Processes*, **22**(13): 2122–2132
- Goswami, B. N., V. Krishnamurthy and H. Annamalai 1999 A broad scale circulation index for the interannual variability of the Indian summer monsoon. *Quarterly Journal of the Royal Meteorological Society*, **125**(554): 611–633
- Hayes, M., M. Svoboda, N. Wall and M. Widhalm 2010 The Lincoln Declaration on Drought Indices: Universal Meteorological Drought Index Recommended. Bulletin of the American Meteorological Society, 92(4): 485–488
- He, B., A. Lü, J. Wu, L. Zhao and M. Liu 2011 Drought hazard assessment and spatial characteristics analysis in China. *Jurnal of Geographical Sciences*, **21**(2): 235–249
- Jain, A. and A. M. Kumar 2007 Hybrid neural network models for hydrologic time series forecasting. Applied Soft Computing, 7(2): 585–592
- Jang, J. S. R. 1993 ANFIS: adaptive-network-based fuzzy inference system. Systems, Man and Cybernetics, IEEE Transactions on, 23(3): 665–685
- Jang, J. S. R., C. T. Sun and E. Mizutani 1997 Neuro-Fuzzy and Soft Computing, ed. Prentice Hall Englewood Cliffs, New Jersey, USA
- Jeong, D.-I. and Y.-O. Kim 2005 Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction.

Hydrological Processes, 19(19): 3819–3835

- Liong, S.-Y., W.-H. Lim, T. Kojiri and T. Hori 2000 Advance flood forecasting for flood stricken Bangladesh with a fuzzy reasoning method. *Hydrological Processes*, 14(3): 431–448
- Lloyd–Hughes, B. and M. A. Saunders 2002 A drought climatology for Europe. International Journal of Climatology, 22(13): 1571–1592
- Mahabir, C., F. E. Hicks and A. R. Fayek 2003 Application of fuzzy logic to forecast seasonal runoff. *Hydrological Processes*, 17(18): 3749–3762
- Mantoglou, A. 2003 Estimation of heterogeneous aquifer parameters from piezometric data using ridge functions and neural networks. Stochastic Environmental Research and Risk Assessment, 17(5): 339–352
- McKee, T. B., N. J. Doesken and J. Kleist 1993 The relationship of drought frequency and duration to time scales. In "8th Conf. on Applied Climatology", Anaheim, California, pp. 179–184
- Mishra, A. K. and V. R. Desai 2006 Drought forecasting using feedforward recursive neural network. *Ecological Modelling*, 198(1–2): 127–138
- Mishra, A. K. and V. P. Singh 2010 A review of drought concepts. Journal of Hydrology, **391**(1–2): 202–216
- Morid, S., V. Smakhtin and K. Bagherzadeh 2007 Drought forecasting using artificial neural networks and time series of drought indices. *International Journal of Climatology*, 27(15): 2103–2111
- Nagy, H. M., K. Watanabe and M. Hirano 2002 Prediction of Sediment Load Concentration in Rivers using Artificial Neural Network Model. *Journal of Hydraulic Engineering*, **128**(6): 588–595
- Nayak, P. C., K. P. Sudheer and K. S. Ramasastri 2005 Fuzzy computing based rainfall–runoff model for real time flood forecasting. *Hydrological Processes*, **19**(4): 955–968
- Nayak, P. C., K. P. Sudheer, D. M. Rangan and K. S. Ramasastri 2004 A neuro–fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology*, **291**(1–2): 52–66
- Philander, S. G. H. 1990 El Nino, La Nina and the Southern Oscillation, ed. Academic Press San Diego
- Rajurkar, M. P., U. C. Kothyari and U. C. Chaube 2004 Modeling of the daily rainfall–runoff relationship with artificial neural network. *Journal of Hydrology*, 285(1–4): 96–113
- Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes and W. Wang 2002 An Improved In Situ and Satellite SST Analysis for Climate. *Journal of Climate*, **15**(13): 1609–1625
- Rodgers, J. L. and W. A. Nicewander 1988 Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1): 59–66

- Şen, Z. and A. Altunkaynak 2006 A comparative fuzzy logic approach to runoff coefficient and runoff estimation. *Hydrological Processes*, **20**(9): 1993–2009
- Senthil Kumar, A. R., K. P. Sudheer, S. K. Jain and P. K. Agarwal 2005 Rainfall–runoff modelling using artificial neural networks: comparison of network types. *Hydrological Processes*, **19**(6): 1277–1291
- Stigler, S. M. 1989 Francis Galton's Account of the Invention of Correlation. *Statistical Science*, 4(2): 73–79
- Tsakiris, G. and H. Vangelis 2004 Towards a Drought Watch System based on Spatial SPI. Water Resources Management, 18(1): 1–12
- Vicente–Serrano, S. M., S. Begueria and J. I. Lopez–Moreno 2010 A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *Journal* of Climate, 23(7): 1696–1718
- Walker, G. T. 1924 Correlation in seasonal variations of weather, IV, A further study of world weather, Mem. Indian Meteorol. Memoirs of the India Meteorological Department, 24(9): 275–332
- Walker, G. T. and E. W. Bliss 1932 World Weather V. Memoirs of the Royal Meteorological Society, 4(36): 53–84
- Wang, B., R. Wu and T. Li 2003 Atmosphere–Warm Ocean Interaction and Its Impacts on Asian–Australian Monsoon Variation*. *Journal of Climate*, 16(8): 1195–1211
- WMO 2012 Standardized Precipitation Index User Guide In "WMO–No. 1090", Vol. ed. by World Meteorological Organization, Geneva 2, Switzerland
- Wyrtki, K. 1975 El Niño—The Dynamic Response of the Equatorial Pacific Oceanto Atmospheric Forcing. Journal of Physical Oceanography, 5(4): 572–584
- Xiao, H. and C. R. Mechoso 2009 Seasonal Cycle–El Niño Relationship: Validation of Hypotheses. *Journal of the Atmospheric Sciences*, **66**(6): 1633–1653
- Yanai, M., C. Li and Z. Song 1992 Seasonal heating of the Tibetan Plateau and its effects on the evolution of the Asian summer monsoon. *Journal of the Meteorological Society of Japan*, 70(1B): 319–351
- Yasunari, T. 1990 Impact of Indian monsoon on the coupled atmosphere/ocean system in the tropical pacific. *Meteorology and Atmospheric Physics*, **44**(1–4): 29–41
- Ying, H. 2000 Fuzzy Control and Modeling: Analytical Foundations and Applications, ed. Wiley–IEEE Press, pp. i – xxxii
- Zhang, Y., T. Li, B. Wang and G. Wu 2002 Onset of Asian summer monsoon over Indo-china and its interannual variability. *Journal of Climate*, 15(22): 3206–3221