ORIGINAL PAPER

Application of several artificial intelligence models and ARIMAX model for forecasting drought using the Standardized Precipitation Index

A. Jalalkamali · M. Moradi · N. Moradi

Received: 29 March 2014/Revised: 17 October 2014/Accepted: 10 November 2014/Published online: 26 November 2014 © Islamic Azad University (IAU) 2014

Abstract Drought is among the most important natural disasters influencing different aspects of human life. In recent decades, intelligent techniques have shown to be highly capable of modeling and forecasting nonlinear and dynamic time series. Hence, the present study aimed to forecast drought using and comparing the multilayer perceptron artificial neural network (MLP ANN), adaptive neuro-fuzzy inference systems (ANFIS), support vector machine (SVM) model, and the autoregressive integrated moving average (ARIMAX) multivariate time series. To this end, the precipitation data obtained from the Yazd synoptic station for a 51-year statistic period were used. Moreover, the humidity levels for short-term (3 and 6 months) and long-term (9, 12, 18, and 24 months) periods were calculated using the Standardized Precipitation Index (SPI). Next, based on the results of calculations, the 1961–2002 period was selected as the control group and the 2003-2012 period was selected as the experimental group. In order to forecast the SPI for the t + 1 period, values of SPI, precipitation, and temperature of previous eras were used. Results indicated that in a 9-months period (as the timescale), the ARIMAX model gives SPI values and forecast drought with more precision than the SVM, ANFIS, and MLP models.

A. Jalalkamali (🖂)

Department of Water Engineering, Islamic Azad University, Kerman Branch, Kerman, Iran e-mail: ajalalkamali@yahoo.com

M. Moradi · N. Moradi Water Resources Engineering, Islamic Azad University of Kerman, Kerman, Iran

N. Moradi Islamic Azad University of Bam, Kerman, Iran

Introduction

All areas across the globe may suffer from drought occasionally, but this phenomenon more commonly occurs in regions that are affected irregularly and randomly by different climatic systems. The chief indication of meteorological drought is the decline of precipitation to less than normal precipitation (long-run average precipitation). Reduction in soil moisture and a decrease in surface and groundwater resources are some of the consequences of reduced precipitation. In this regard, finding an accurate solution to the more precise prediction of drought so as to minimize its adverse effects on the nature and environment is a necessity. The objective of this research was to demonstrate the potential of artificial intelligence techniques for drought forecast. Various indices have been so far introduced for determining the characteristics of meteorological drought. In this research, the Standardized Precipitation Index (SPI) was used as one of the most important and prominent available indices. Studies previously conducted on this concept were as follows:

Bacanli et al. (2008) used the adaptive neuro-fuzzy inference system (ANFIS) to forecast drought in Turkey. Various SPI forecasting models were examined in 1-, 3-, 6-, 9-, and 12-month timescales. Results of examinations revealed that a combination of former amounts of precipitation and SPI enhances the performance of models. Comparisons of the results of the feed-forward neural network (FFNN) and multiple linear regression (MLR) methods revealed the superiority of ANFIS over the other two models. Keskin et al. (2009) also employed the ANFIS



and fuzzy logic models to forecast meteorological drought. They used SPI to analyze drought and also used SPIs of 3-, 6-, 9-, and 12-month periods to better forecast drought. Results of their research also reflected the superiority of ANFIS over the fuzzy logic model. Jalalkamali et al. (2011) used a fuzzy inference system and multilayer perceptron neural networks to predict the level of groundwater resources in Kerman Plain. Results of this research proved the superiority of ANFIS over the multilayer perceptron neural network. Furthermore, Shirmohammadi et al. (2013) used the ANFIS, ANN, Wavelet-ANN, and Wavelet-AN-FIS models to forecast meteorological drought in the next 3 months on the basis of the SPI for Azerbaijan Province (Iran). Results of this research indicated that all of the considered modeling methods were able to forecast SPI, but the hybrid Wavelet-ANFIS model demonstrated a better performance. The multilayer perceptron neural network (MLP) was studied in five Iranian synoptic stations to predict quantitative values of SPI using the optimized Levenberg-Marquardt Algorithm (LMA) and tangent of sigmoid function. The studies managed to calculate values of SPI3, SPI6, SPI9, SPI12, and SPI24. In the course of these studies, the MLP of SPIs of 1 month of prediction was examined as well. Results showed that the MLP model is capable of forecasting 12- and 24-month SPIs with more precision than other SPIs (Rezaeian-zadeh and Tabari 2012). Santos et al. (2009) also worked on forecasting drought based on artificial neural networks (ANNs) of three regions in the San Francisco River basin, Brazil. Results of their research showed that the method is capable of forecasting SPI for the coming month in a 12-month timescale, but it gradually loses its precision over time. Fatehi Marj and Meijerink (2011) used satellite images, climate indices, and artificial neural networks to forecast drought. These researchers found out that artificial neural networks yield highly precise results through accepting NAO and SOI (a year ago) as their inputs. Nikbakht Shahbazi et al. (2011) calculated the SPIs of the catchments of Mamloo and Talegan dams in Tehran, Iran. Based on meteorological variables, including temperature and geopotential height, they found out that SVM often yields highly precise results. Therefore, this method can be used for forecasting nonlinear behavior of meteorological data in short-term statistical periods. In a research, Khan and Coulibaly (2006) studied the support vector machine's potential to forecast the level of the water contained in the Erie Lake in the long run (12 months). They also drew comparisons between SVM, MLP, and an autoregressive model. Results of their studies suggested that SVM demonstrates a performance better than the other two models. Qing et al. (2012) worked on forecasting precipitation on the basis of time series. In the first stage, they applied SVM to environmental factors, and in the second stage, they employed the conditional autoregressive model (CAR). Finally, the reliability of the SVM-CAR method was approved. These researchers reported that the aforementioned method (SVM-CAR) forecasts drought and flood with high precision. Han et al. (2010) used remote sensing data and the autoregressive integrated moving average (ARIMA) model to forecast drought. Results of their attempt indicated that the expanded autoregressive (AR) model can be used to predict drought in Guanzhong Plain. Cancelliere et al. (2006) employed a nonparametrical method as well as SPI to forecast drought in Sicily. They realized that a combination of SPI and ARIMA along with time series of monthly precipitations, as inputs, yields satisfactorily precise results. Durdu (2010) used the linear random model to forecast drought in the city of Buyuk in the Menderes River basin (west of Turkey). Results of this research reflected high precision of the ARIMA model in forecasting drought for two coming months.

The aim of this study was to develop a proper model to forecast meteorological drought using the Standardized Precipitation Index. For this purpose, the time series data on the period between 1961 and 2012 were used in this research. The data were recorded on a monthly basis in the Yazd synoptic station. The Yazd Province is located at the heart of the Iranian plateau. It is situated on northern latitude of $29^{\circ}35''-35^{\circ}7''$ and eastern longitude of $52^{\circ}50''-58^{\circ}16''$ (Fig. 1).

Materials and methods

Standardized Precipitation Index (SPI)

The average precipitation in Yazd Province is between 60 and 80 mm. Recorded data included values of minimum



Fig. 1 Location of the study area in Yazd Province, Iran



Table 1 Statistical parameters of data recorded from the Yazd synoptic station

Parameter	Number of data	Min	Max	Mean	SD	Skewness
Minimum temperature	624	-6.2	35.6	11.9	8.8	0
Maximum temperature	624	3.9	42.3	26.6	9.6	-0.1
Precipitation	624	0	71.2	4.7	8.4	3.2

temperature (Celsius), maximum temperature (Celsius), and precipitation (millimeter). The monthly statistical parameters of the data are given in Table 1.

SPI was calculated using the precipitation data with the DIP (Drought Indices Package) software based on the following equation.

In most cases, the distribution that best models observational precipitation data is the Gamma distribution. The density probability function for the gamma distribution is obtained as follows:

$$g(x) = \frac{1}{\beta^{\alpha} \Gamma(\alpha)} x^{\alpha - 1} e^{\frac{-x}{\beta}} \quad \text{for } x > 0$$
(1)

where a > 0 is the shape parameter, b > 0 is the scale parameter, and x > 0 is the amount of precipitation. $\Gamma(\alpha)$ is the value taken by the standard mathematical function known as the Gamma function defined by the following integral:

$$\Gamma(\alpha) = \lim_{n \to \infty} \prod_{\nu=0}^{n-1} \frac{n! n^{\nu-1}}{\nu + \nu} = \int_{0}^{\infty} y^{\alpha-1} e^{-y} dy$$
(2)

In general, the gamma function is evaluated either numerically or using the values tabulated depending on the value of parameter α . In order to model the data observed with a gamma-distributed density function, it is necessary to estimate appropriately the α and β parameters. Different methods have been so far suggested in the literature for estimation of these parameters. For example, Edwards and McKee (1997) used the Thom (1958) approximation to achieve maximum probability.

$$\hat{\alpha} = \frac{1}{4A} \left(1 + \sqrt{1 + \frac{4A}{3}} 4A \right) \tag{3}$$
$$\hat{\beta} = \frac{\bar{x}}{2} \tag{4}$$

$$\hat{eta} = rac{x}{\hat{lpha}}$$

where for *n* observations:

$$A = \ln(\bar{x}) - \frac{\sum \ln(x)}{n}$$
(5)

Estimation of the parameters can be further improved by using the interactive approach suggested in Wilks (1995).

After estimating coefficients α and β , the density of the probability function g(x) is integrated with respect to x. Consequently, an expression for cumulative probability G(x) is obtained which shows the certain amount of rain has been observed in a given month and a specific timescale.

$$G(x) = \int_{0}^{\infty} g(x) \mathrm{d}x = \frac{1}{\hat{\beta}^{\hat{\alpha}} \Gamma(\hat{\alpha})} \int_{0}^{x} x^{\hat{\alpha}-1} e^{-x/\hat{\beta}_{\mathrm{d}x}}$$
(6)

If $t = x/\hat{B}$, this equation becomes the incomplete gamma function:

$$G(x) = \frac{1}{\Gamma(\hat{\alpha})} \int_{0}^{x} t^{\hat{\alpha}-1} e^{-t} \mathrm{d}t$$
(7)

The gamma function is not defined by x = 0, and since there may be no precipitation, the cumulative probability is as follows:

$$H(x) = q + (1 - q)G(x)$$
(8)

where q is the probability of no precipitation. The cumulative probability is then transformed into a normal standardized distribution with null average and unit variance based on which SPI is obtained (for more details, see Edwards and McKee 1997 or Lloyd-Hughes and Saunders 2002). The above approach, however, is neither practical nor numerically simple to use if there are many grid points or many stations to calculate SPI. In this case, an alternative method was described in Edwards and McKee (1997) using the technique of approximate conversion developed by Abramowitz and Stegun (1965). This technique converts the cumulative probability into a standard variable Z. SPI is thus defined as follows:

$$Z = SPI = -\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right)$$

for $0 < H(x) < 0.5$ (9)

$$Z = SPI = +\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right)$$
for 0.5 < H(x) < 1
(10)

$$t = \sqrt{\ln\left[\frac{1}{(H(x))^2}\right]}$$
 for $0 < H(x) < 0.5$ (11)

$$t = \sqrt{\ln\left[\frac{1}{(1 - H(x))^2}\right]} \quad \text{for } 0.5 < H(x) < 1 \tag{12}$$

where x is precipitation, H(x) is the cumulative probability of precipitation observed, and c^0 , c^1 , c^2 , d^0 , d^1 , d^2 are constants with the following values (Table 2):



 $c^0 = 2.515517, c^1 = 0.802853, c^2 = 0.010328$ $d^0 = 1.432788, d^1 = 0.189269, d^2 = 0.001308$

The adaptive neuro-fuzzy inference system (ANFIS)

The ANFIS is a universal estimator that is capable of approximating any real continuous function in a compact set to any degree of accuracy (Jang et al. 1997). The basic structure of the type of fuzzy inference system could be considered a model mapping input characteristics to input membership functions. This goes on to map input membership functions to rules and the maps rules to a set of output characteristics. Finally, it maps output characteristics to output membership functions and also maps the output membership function to a single-value output or an output-based decision (Jang et al. 1997). Each fuzzy system contains three main components including a fuzzifier, fuzzy database, and defuzzifier. In addition, a fuzzy rule base and inference engine.

Figure 2 represents a typical ANFIS architecture. In layer one, every node is an adaptive node with a node function such as a generalized bell membership function or a Gaussian membership function. In layer two, every node is a fixed node representing the firing strength of each rule and is calculated by the fuzzy and connective of the 'product' of the incoming signals. In layer three, every node is a fixed node showing the normalized firing strength of each rule. The *i*-th node calculates the ratio of the *i*-th rule's firing strength to the sum of two rules firing strengths. In layer four, every node is an adaptive node with a node function indicating the contribution of the *i*-th rule to the overall output. In layer five, the single node is a

Table 2 Classification according to the SPI values (Bacanli et al.2008; McKee et al. 1993)

SPI	Drought category		
2>	Extremely wet		
1.99–1.5	Very wet		
1.49–1.0	Moderately wet		
0.99–(-0.99)	Near normal		
(-1.0)-(1.49)	Moderately dry		
(-1.0)-(-1.99)	Severely dry		
-2<	Extremely dry		



Fig. 2 Typical ANFIS architecture (Jang, 1993)

fixed node indicating the overall output as the sum of all incoming signals (Jang and Sun 1995).

Where x and y are the inputs and z is the final output; A1, A2, B1, and B2 are the linguistic labels (small, large, etc.) associated with this node function, and w_i is the normalized firing strength that is the ratio of the *i*-th rule's firing strength (Wi) to the sum of the first and second rules' firing strengths (W1 and W2). Π is also the node label (Shirmohammadi et al. 2013).

Multilayer perceptron (MLP) model

The MLP is one of the most widely implemented neural network models. These networks take in a set of real inputs, x_i , and use them to compute one or more output values, $f_k(x)$, with one hidden layer, such as that illustrated in Fig. 3. The outputs might be computed as follows:

$$f_k(x) = b_k + \sum_j v_{jk} h_j(x) \tag{13}$$

$$h_j(x) = \tanh\left(a_j + \sum_i u_{ij}x_i\right) \tag{14}$$

where u_{ij} is the weight exerted on the connection by input unit *i* to hidden unit *j*; similarly, v_{ik} is the weight put on the connection by hidden unit j to output unit k. a_i and b_k are biases of the hidden and output units. These weights and biases are the parameters of the networks. Each output value, $f_k(x)$, is just a weighted sum of hidden unit values plus a bias. Each hidden unit computes a similar weighted sum of input values and then passes it in the return of a nonlinear activation function. The activation function selected here is the hyperbolic tangent (tanh) function. The weights and biases in an MLP network are learned based on a set of training cases, $(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})$, giving examples of inputs, $x^{(i)}$, and associated targets, $y^{(i)}$ (both of which have several components). Standard neural network training procedures adjust the weights and biases in the network so as to minimize a measure of "error" in the training cases, which is most commonly the sum of the



Fig. 3 Neural network multilayer perceptron

squared differences between the network outputs and the targets. Finding the weights and biases that minimize the chosen error function is commonly done using some gradient-based optimization methods, derivatives of the error with respect to the weights and biases calculated by back propagation. The detailed theory of the back-propagation algorithm can be found in Haykin (1999), which was beyond the scope of this research. There are typically many local minima, but good solutions often draw on heuristic modifications (Khan and Coulibaly 2006).

The support vector machine (SVM) theory

According to the structural risk minimization (SRM) principle, the generalization ability of learning machines depends more on capacity concepts than the dimensionality of the space or the number of free parameters of the loss function (as espoused by the classical paradigm of generalization). Hence, for a given set of observations $(x^1, y^1), \ldots, (x^n, y^n)$, the SRM principle selects the function $f_\beta b$ in the subset $\{f_\beta: \beta \in A\}$, for which the guaranteed risk bound, as given by Eq. (15), is minimal. In other words, the actual risk is controlled by the two terms in Eq. (15):

$$R(\beta) \le R_{\rm emp}(\beta) + \Omega\left(\frac{n}{h}\right) \tag{15}$$

where the first term is an estimation of the risk and the second term is the confidence interval for this estimation.

The parameter h is called the VC dimension (named after Vapnik and Chervonenkis) of a set of functions. It can be seen as the capacity (or the flexibility of the functional class in fitting the underlying learning problem) of a set of functions implementable by the learning machine. If the function is too complex (for the given amount of training data), chances of over-fitting arise. In the case of ANN, for the selected architecture, capacity is fixed and it is tried to minimize the empirical risk. However, in the case of SVM, the empirical risk term and the capacity term are controlled simultaneously. SVM is an approximate implementation of the SRM principle. The final approximating function for SVM for regression is of the following form:

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) K(x_i, x) + b$$
(16)

where $K(x_i,x) = (\varphi(x).\varphi(x_i))$ is the kernel function, which processes the inner product in feature space, $\varphi(x)$. To act as a kernel, a function needs to satisfy Mercer's condition (discussed in the subsection on the proposed kernel function). Kernel representation offers a powerful alternative for using linear machines for hypothesizing complex real-world problems as opposed to ANN-based learning paradigms, which use multiple layers of threshold nonlinear functions. The approximating function is designed to have the smallest ε deviation (given by Vapnik's ε insensitive loss function) from measured targets, d_i , for all training data. Slack variables, ξ_i and ξ_i^* , are introduced to account for outliers in the training data. The algorithm computes the value of Lagrange multipliers, a_i and a_i^* , by minimizing the following objective function:

Minimize
$$\frac{1}{2} \|a\|^2 + C \sum_{i=1}^{N} \left(\xi_i + \xi_i^*\right)$$

Subject to
$$d_i - (a \cdot x_i + b) \le \varepsilon + \xi_i$$
$$(a \cdot x_i + b) - d_i \le \varepsilon + \xi_i^*$$
$$\xi_i \xi_i^* \ge 0$$
(17)

Expressed in the dual form as follows:

Maximize
$$-\frac{1}{2}\sum_{i,j=1}^{N} (\alpha_{i} - \alpha_{i}^{*}) (\alpha_{j} - \alpha_{j}^{*}) \langle \phi_{i}, \phi_{j} \rangle$$
$$-\varepsilon \sum_{i=1}^{N} (\alpha_{i} + \alpha_{i}^{*}) + \sum_{i=1}^{N} y_{i} (\alpha_{i} - \alpha_{i}^{*})$$
(18)

It is subject to the following constraints:

$$\sum_{i=1}^{n} (\alpha_i - \alpha_i^*) = 0$$

$$0 \leqslant \alpha_i \leqslant C, \quad i = 1, 2, \dots, N$$

$$0 \leqslant \alpha_i^* \leqslant C, \quad i = 1, 2, \dots, N$$
(19)

where *C* is a user-specified constant and it determines the trade-off between the flatness of *f* (*x*) and the amount of allowable deviation. It is worth noting that both in the objective function given by Eq. (18) and in the approximation function given by Eq. (16), the training patterns appear as dot products between the training pairs. The solution to the above problem yields a_i and a_i^* for all i = 1 to *N*. It can be shown that all the training patterns within the ε -insensitive zone yield a_i and a_i^* as zeros. The remaining nonzero coefficients essentially define the final decision function. The training examples corresponding to these non-vanishing coefficients are called support vectors (Sivapragasam et al. 2001).

The values of ε , *C*, and the kernel-specific parameters must be tuned to their optimum by the user to get the final regression estimation. At the moment, identification of optimal values for these parameters is largely a trial and error process. Furthermore, other than the ε -insensitive loss function, quadratic loss function (Fig. 4) may also be used in which case $\varepsilon = 0$. In this study, the quadratic loss function is preferred over the ε -insensitive loss function, as the former is less memory-intensive using computers. Details on SVM can be found, for example, in Vapnik (1995), Drucker et al. (1997), Smola and Scholkopf (1998), Haykin (1999), Vapnik (1999), and Cristianini and Shawe-Taylor (2000).

ARMA and AIRMA processes

In an ARIMA (p, d, q) process, q, d, and p denote the number of autoregressive terms, integration degree, and the





Fig. 4 Illustrative figures for a $\epsilon\text{-insensitive}$ loss function and b quadratic loss function

number of moving average terms, respectively. When d is equal to zero, ARIMA is changed to ARMA (autoregressive moving average). Usually, in order to estimate the ARIMA and ARMA models, the Box–Jenkins approach is employed. This approach is comprised of the following four phases: recognition, estimation, precision measurement, and prediction.

ARIMAX

Recently, the trend of the tertiary industry has been studied by many researchers using time series, but mostly, they use the ARIMA model with one time series as it is an effective way of studying time series with one time series in systems. However, since ARIMA works only with one variable, it could not explain the relationships among system variables. The complex systems always are expressed by more than one variable in practice. According to observations or applications, the multivariate time series could be obtained similar to recording a few different variables of a study system such as the following variables: atmospheric pressure, temperature, humidity in the meteorology, cardiac rate, blood pressure, breathe and oxygen saturation of blood in the physiology, rates of exchange in some currencies, and sub-index in securities markets or in space expanded systems. More than one record could be obtained from the different spaces including over falls, satellite data, electrocardiograms, electroencephalograms, and such. In the case of these multivariate time series, the aforementioned variables are influenced by other variables, while they change their rules as well. It is not possible to express the multivariate time series' changing rules using the ARIMA model with only one time series. The reason is that the mensuration model is imperfect. Therefore, it is necessary to create a model with multivariate ARIMAX model (Fan et al. 2009; Chadsuthi et al. 2012).

Analysis of multivariate time series has been performed so far. Cox and Jenkins modeled steady multivariate time series using ARIMA model and input variables. Technically, it is necessary to stabilize input series and studied



series. The condition limiting the development of the analysis to multivariate time series has to be met as well. Engle and Granger put forward the concept of cointegration. Based on the cointegraty theory, the regression residual time series of input time series and output time series are needed to be stabilized, and thus, it is not necessary to stabilize them. Introduction of the concept of cointegration contributed to the development of the analysis to multivariate time series. The analysis of multiple regression and time series was introduced into the concept to improve the precision of forecasting.

Structure

The model with the following structure was defined as the dynamic regression model, simplified as the ARIMAX model.

$$\begin{cases} y_t = \mu + \sum_{i=1}^k \frac{\Theta_i(B)}{\Phi_i(B)} B^{l_i} x^{it} + \varepsilon_t \\ \varepsilon_t = \frac{\Theta(B)}{\Phi(B)} a_t \end{cases}$$
(20)

where $\Phi_i(B)$ denotes the auto-regression coefficients' multinomial of the *I*-th input time series, $\Theta_i(B)$ denotes the average coefficients' multinomial of the *i*-th input time series, l_i represents the lag degree of the *i*-th input variables, $\{\varepsilon_t\}$ shows the regression residual time series, $\Phi(B)$ denotes residual series' auto-regression coefficients' multinomial, denotes residual series' moving average coefficients' multinomial, and $\{a_t\}$ is white noise time series with zero average.

Concept

The basic concept of the ARIMAX model is as follows: Assuming that both output time series $\{y_t\}$ and input time series $\{x^{it}\}$ (i = 1, 2,..., k) are steady; first, the regression model for output time series and input time series is created:

$$y_t = \mu + \sum_{i=1}^{l} \frac{\Theta_i(B)}{\Phi_i(B)} B^{l_i} x^{it} + \varepsilon_t$$
(21)

And since the linear combinations of steady time series are also steady, the residual time series $\{\varepsilon_t\}$ is steady as well because $\{y_t\}$ and $\{x^{it}\}(i = 1, 2, ..., k)$ are steady.

$$\varepsilon_t = y_t - \left(\mu + \sum_{i=1}^k \frac{\Theta_i(B)}{\Phi_i(B)} B^{l_i} x^{it}\right)$$
(22)

In order to extract the interrelated information from the residual time series $\{\varepsilon_t\}$, the final model (1) will be obtained. The detailed model and mathematical background of ARIMAX can be found in Fan et al. (2009) study.

Table 3 Summary of the advantages and disadvantages of the employed models

Model	Advantages	Disadvantages
ANN	Requiring less formal statistical training, having the capability of implicitly detecting complex nonlinear relationships between dependent and independent variables, having the capability of detecting all possible interactions between predictor variables, and having access to multiple training algorithms	Being of a "black box" nature, greater computational cost, proneness to over-fitting, and the empirical nature of model development
ANFIS	There is a need for adaptability or learning algorithms to produce outputs, whereas a combination of ANN and fuzzy systems, called Neural-Fuzzy system, is capable of eliminating the basic problems in fuzzy system design (generating a set of fuzzy if–then rules), using the learning capability of an ANN for automatic fuzzy if–then rule generation and parameter optimization	ANFIS is sensitive to initial number of fuzzy rules (number of clusters), the computational complexity grows as the number of fuzzy rules increases
SVM	There are four main advantages: First, it has a regularization parameter, which makes the user conscious of avoiding over-fitting. Second, it uses the kernel trick, so it can be built expert knowledge about the problem through engineering the kernel. Third, an SVM is defined by a convex optimization problem (no local minima), for which, there are efficient methods. Fourth, it is an approximation to a bound on the test error rate, and there is a substantial body of theory behind it, suggesting it as a good idea	The biggest limitation of the support vector approach lies in the choice of kernel. The second limitation in concerned with speed and size, both in training and testing stages. Although SVMs have good generalization performance, they can be significantly slow in the testing phase
ARIMAX	The advantage of the model is the wide utilization of times series modeling and forecasting techniques, which makes few parameters adequate for explaining time series	The disadvantages of the method, above all, are difficult interpretation of the model and the need for time series with a minimum of 50 observations, which will sometimes be very problematic to obtain if some values are measured only annually

Comparison of performances of models

Coefficient of determination (R^2) and root-mean-squared error (RMSE) were used to compare the performances of models and select the best one.

$$R = \sqrt{1 - \frac{\sum_{i=1}^{n} \left(\text{SPI}i - \hat{\text{SPI}}i \right)^{2}}{\sum_{i=1}^{n} \left(\hat{\text{SPI}}i - \overline{\text{SPI}}i \right)^{2}}}$$
(23)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(SPIi - S\widehat{P}Ii\right)^{2}}{N}}$$
(24)

where RMSE is the root-mean-square error and R^2 is the coefficient of efficiency. In addition, $S\widehat{PI}_i$ is the measured value, SPI_i is the predicted (estimated) value, and \overline{SPI}_i is the measured values mean. *n* is the total number of events considered. According to previous sections about the structure of the models, a general review is provided to better illustrate the advantages and disadvantages of the models employed in this research as shown in Table 3.

Results and discussion

Combinations presented in Table 4 were calculated using the aforementioned models for 3-, 6-, 9-, 12-, 18-, and 24-month timescales. These combinations were selected

Table 4 Structures of forecasting models

Combination	Input structure	Output
M1	$SPI_{(t)(t-1)(t-2)(t-3)}, \mathbf{P}_{(t)(t-1)}$	SPI _(t+1)
M2	$SPI_{(t)(t-1)(t-2)}, T_{max(t)(t-1)}, T_{min(t)(t-1)}$	$SPI_{(t+1)}$
M3	$SPI_{(t)(t-1)(t-2)}, \mathbf{T}_{max(t)(t-1)(t-2)}, \\ \mathbf{T}_{min(t)(t-1)(t-2)}$	$SPI_{(t+1)}$
M4	$SPI_{(t)(t-1)}, T_{max(t)(t-1)}, T_{min(t)(t-1)}, P_{(t)(t-1)}$	$SPI_{(t+1)}$

from so many other combinations obtained through trial and error. The best combination for each timescale presented in Table 5 was compared to the best results of other scales.

Parameters used in this stage to implement the ANFIS model included the type of Gaussian membership function, number of membership functions, number of iterations obtained through trial and error to prevent over-fitting and under fitting. Moreover, the MLP neural network, sigmoid transfer function, and the number of intermediate layers neurons (which increased by order) were examined for each SPI time series as well. Results showed that MLP artificial neural networks and intermediate layer neurons fewer than input neurons yield the minimum error and maximum correlation coefficient. In this study, the ARI-MAX method was used to forecast drought. The following software was also used to forecast drought and estimate the models: STATISTICA, SPSS, Eviews, DIP. and MATLAB.



 Table 5
 Results of examination of the MLP, SVM, and ANFIS

 models and ARIMAX time series for the best learning case and the
 best combination of various tested time series

Timescale	Model	Input combination	R	RMSE
3	ANFIS	M4	0.852	0.471
3	SVM	M4	0.731	0.621
3	MLP(8-7-1)	M4	0.817	0.508
3	ARIMAX(1,0,1)(1,0,1)	M4	0.895	0.404
6	ANFIS	M3	0.808	0.551
6	SVM	M4	0.713	0.606
6	MLP(8-3-1)	M4	0.774	0.545
6	ARIMAX(1,0,1)(1,0,1)	M4	0.880	0.411
9	ANFIS	M4	0.855	0.423
9	SVM	M4	0.819	0.465
9	MLP(9-5-1)	M3	0.840	0.447
9	ARIMAX(1,0,1)(2,0,1)	M4	0.900	0.313
12	ANFIS	M4	0.790	0.437
12	SVM	M3	0.782	0.458
12	MLP(9-4-1)	M3	0.770	0.458
12	ARIMAX(1,0,1)(1,0,1)	M1	0.847	0.431
18	ANFIS	M4	0.826	0.395
18	SVM	M3	0.732	0.497
18	MLP(7-5-1)	M2	0.791	0.426
18	ARIMAX(1,1,0)(1,0,1)	M3	0.871	0.346
24	ANFIS	M2	0.826	0.392
24	SVM	M3	0.752	0.440
24	MLP(9-6-1)	M3	0.838	0.362
24	ARIMAX(1,0,1)(1,0,0)	M4	0.876	0.354

Table 5 shows values of the R and RMSE parameters for short-term and long-term prediction of SPI along with best combinations of the following models: ANFIS, SVM, MLP, and ARIMA. Short term refers to 3-6 months, while long term refers to 9-24 months. Results presented in this table indicate that the maximum value of *R* obtained by the four models and the aforementioned timescales was associated with SPI 9 and the M4 combination. This value was obtained using the ARIMAX model.

As seen in Table 5, MLP, SVM, ANFIS, and ARI-MAX demonstrated more capability and sensitivity in forecasting drought by SPI 9. In addition, it was observed that in other timescales, the ARIMAX model has higher R compared to the other three models. Moreover, RMSE value of this model is lower than others. As can be seen in Table 5, results from different timescales are very close, and the input combination of M4 in different timescales has the highest frequency. However, it is not recommended to use SVM model in the study area, since it has the weakest results in





Fig. 5 Comparison of four models in the 9-month timescale

different timescales. Furthermore, from the four combinations, combination 4 (M4) demonstrated the best performance of all in most timescale. Hence, a combination of SPI, precipitation, maximum temperature, and minimum temperature has the highest contribution to estimation of results.

According to Table 5 and Fig. 5, the ARIMAX model yields the maximum correlation coefficient and minimum error in all timescales. That is to say, in the 9-month timescale, combination 4 (M4) yields 0.900 and 0.313 for correlation coefficient and error, respectively.

Finally, using the Schwarz Bayesian criterion (SBC) and Akaike information criterion (AIC), the most cost-effective model was selected based on which prediction and data generation were started. The model that gives the minimum AIC and SBC values is selected as the best model. In this research, the minimum values of AIC and SBC were 0.7877 and 0.8570, respectively.

Figure 6 was used to determine the parameters (Q, q, P, p) through the autocorrelation function (ACF) and partial autocorrelation function (PACF). ACF is one of the means of validation. Since the ACF and PACF curves did not exceed the limits, the assumption of independence of residuals holds true.

In the following, the curves of the related time series are included for better analysis of the performance of models using M4 in the 9-month timescale.

As seen in Fig. (7), results of the ARIMAX model were more realistic in a 9-month timescale using M4. Therefore, this model was able to forecast descents (i.e., droughts) better than other models. However, the ANFIS was able to predict ascends (peaks) better than other models. Accordingly, results indicated that time series forecast variations of SPI better than artificial neural networks and neurofuzzy systems. Moreover, short-term timescales are more fluctuating than long-term timescales and also rapidly respond to small changes of monthly precipitation. Since short-term droughts are more frequent, an increase in



Fig. 7 Comparison of performances of the ARIMAX, MLP, SVM, and ANFIS models using actual values in a 9-month timescale in the test period (2003–2012): a SVM, b ANFIS, c MLP, d ARIMAX

timescale leads to a decrease in the frequency of droughts. Hence, it is concluded that in a 9-month timescale, droughts in the study area last longer than droughts in 3and 6-month timescales.

Conclusion

In this study, the precision of the results of the ANFIS, SVM, MLP, and ARIMAX multivariate time series in



forecasting meteorological drought was examined. Results of this study indicated that the ARIMAX(1,0,1)(2,0,1)model demonstrates a performance better than other models in a 9-month timescale in forecasting meteorological drought. Results of this research imply that all the MLP, SVM, ANFIS, and ARIMAX models are useful means of forecasting meteorological drought which work on recorded data. Therefore, it can be noted that these models could be very useful in the continuous monitoring of regional meteorological drought in order to manage limited water resources in different periods. A comparison of the results of models in a 9-month timescale revealed that the highest precision for drought prediction is obtained by the ARI-MAX model which is followed by the ANFIS, MLP, and SVM models. However, with increase in meteorological parameters, the results of the models would more reliable and accurate. According to the results, the MLP, SVM, ANFIS, and ARIMAX models demonstrate more capability and sensitivity to drought forecast in a 9-month period. Based on the results obtained for a 9-month timescale, it was concluded that droughts in the area are less frequent but more lasting.

References

- Abramowitz M, Stegun A (1965) Handbook of mathematical formulas, graphs, and mathematical tables. Dover Publications Inc, New York
- Bacanli UG, Firat M, Dikbas F (2008) Adaptive neuro-fuzzy inference system for drought forecasting. Stoch Environ Res Risk Assess 23(8):1143–1154
- Cancelliere A, Bonaccorso B, Mauro G (2006) A non-parametric approach for drought forecasting through the standardized precipitation index. Metodi statisticie matematici per l Analisi delle serie idrologiche. 1(1):1–8
- Chadsuthi S, Modchang C, Lenbury Y, Iamsirithaworn S, Triampo W (2012) Modeling seasonal leptospirosis transmission and its association with rainfall and temperature in Thailand using timeseries and ARIMAX analyses. Asian Pac J Trop Med 5(7):539–546
- Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines. Cambridge University Press, Cambridge
- Drucker H, Burges C, Kaufman L, Smola A, Vapnik V (1997) Support vector regression machines, advances in neural information processing systems 9. MIT Press, Cambridge, pp 156–161
- Durdu OF (2010) Application of linear stochastic models for drought forecasting in the Buyuk Menderes river basin, western Turkey. Stoch Environ Res Risk Assess 24(8):1145–1162
- Edwards DC, McKee TB (1997) Characteristic of 20th century drought in the United States at multiple timescales, Climatology report. Colorado State University, Fort Collins
- Fan J, Shan R, Cao X (2009) The analysis to Tertiary- industry with ARIMAX model. J Math Res 1(2):156–163

- Fatehi marj A, Meijerink AMJ (2011) Agricultural drought forecasting using satellite images, climate indices and artificial neural network. Int J Remote Sens 32(24):9707–9719
- Han P, Wang PX, Zhang SY, Zhu DH (2010) Drought forecasting based on the remote sensing data using ARIMA model. Math Comput Model 51(11):1398–1403
- Haykin S (1999) Neural networks, a comprehensive foundation, 2nd edn. Prentice-Hall, Englewood Cliffs, pp 135–155
- Jalalkamali A, Sedghi H, Manshouri M (2011) Monthly groundwater level prediction using ANN and neuro-fuzzy models: a case study on Kerman plain, Iran. J Hydro-Inform 13(4):867–876
- Jang JSR (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE 23(3):665–685
- Jang JSR, Sun CT (1995) Neuro-fuzzy modeling and control. Proc IEEE 83:378–406
- Jang JSR, Sun CT, Mizutani E (1997) Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence. Prentice-Hall, Eaglewood cliffs, pp 665–685
- Keskin M, Terzi O, Taylan E, Kucukyaman D (2009) Meteorological drought analysis using data- driven models for the lakes district, turkey. Hydrol Sci 56(6):1114–1124
- Khan MS, Coulibaly P (2006) Application of support vector machine in lake water level prediction. J Hydrol Eng 11(3):199–205
- Lloyd-Hughes B, Saunders MA (2002) A drought climatology for Europe. Int J Climatol 22:1571–1592
- Mckee TB, Doesken NJ, Kleist J (1993) The relationship of drought frequency and duration to time scales. In: Proceedings of the eighth conference on applied climatology. Am Meteor Soc Boston, 179–184
- Nikbakht shahbazi AR, Zahraie B, Sedghi H, Manshouri M, Nasseri M (2011) Seasonal meteorological drought prediction using support vector machine. World Appl Sci J 13(6):1387–1397
- Qing C, Xiaoli Z, Kun Z (2012) Research on precipitation prediction based on time series model. In: International conference on computer distributed control and intelligent environmental monitoring (CDCIEM), Hunan, pp 568–571
- Rezaeian zadeh M, Tabari H (2012) MLP-based drought forecasting in different climatic regions. Theor Appl Climatol 109(1):407–414
- Santos CA, Morais BS, Silva GBL (2009) Drought forecast using an artificial neural network for three hydrological zones in San Francisco River basin, Brazil. In: Proceeding of symposium Hs.2 at the Joint IAHS & IAH convention, Hyderabad, India, 1(1), pp 302–312
- Shirmohammadi B, Moradi HR, Moosavi V (2013) Forecasting of meteorological drought using wavelet-ANFIS hybrid model for different time steps (case study: southeastern part of east Azerbaijan province, Iran). Nat Hazards 69(1):389–402
- Sivapragasam C, Liong SY, Pasha MFK (2001) Rainfall and runoff forecasting with SSA-SVM approach. J Hydroinform 3:141–152
- Smola AJ, Scholkopf B (1998) A tutorial on support vector regression. Technical Report NeuroCOLT Technical Report NC-TR-98-030, Royal Holloway College, University of London, UK. http://www.kernel-machines.org
- Thom HCS (1958) A note on the gamma distribution. Mon Weather Rev 86(4):117–122
- Vapnik V (1995) The nature of statistical learning theory. Springer, New York
- Vapnik V (1999) An overview of statistical learning theory. IEEE Trans Neural Netw 10(5):988–999
- Wilks DS (1995) Statistical methods in the atmospheric sciences. An introduction. Academic Press, London

