



# Modeling Short Term Rainfall Forecast Using Neural Networks, and Gaussian Process Classification Based on the SPI Drought Index

Saeed Azimi<sup>1</sup> · Mehdi Azhdary Moghaddam<sup>1</sup>

Received: 26 September 2019 / Accepted: 28 January 2020 /

Published online: 16 February 2020

© Springer Nature B.V. 2020

## Abstract

Determining the quantitative occurrence of droughts, discovering the spatial correlation of droughts, and predicting the occurrence of undesirable classes of drinking water quality and aquifer farming is of high importance. In this research, the Standardized Precipitation Index (SPI) was calculated and analyzed with a monthly survey of 26,027 wells in 609 study areas over a period of 20 years. After analyzing the missing data, the annual rainfall was forecasted in 362 synoptic stations of the country based on an artificial intelligence model. In addition, statistical relationships were extracted in order to achieve a comprehensive and historical map of the state of shortages and surpluses of water resources, as well as verification of artificial intelligence relationships in predicting base data cultivars. The results indicated that the “mild drought” indicator was steeper than the “near-normal drought” indicator. Eventually, the southern and eastern regions and certain parts of the northeast of the country in the period from 2005 to 2015 were placed in the 7th and 8th classes, which indicates severe drought. The analysis of the period 1994–2014 showed that the plains of the Sistan and Baluchestan Province in the south-east region of the country have been significantly more affected by the droughts. With the exception of the central parts of Khorasan, the general eastern, southeast, and southern regions of the country can be considered as an absolute drought class for the long term.

**Keywords** Meteorological drought · Short term rainfall forecast · Artificial intelligence · Artificial neural networks · Gaussian process classification

---

✉ Mehdi Azhdary Moghaddam  
mazhdary@eng.usb.ac.ir

Saeed Azimi  
saed.azimi@pgs.usb.ac.ir

<sup>1</sup> Department of Civil Engineering, Faculty of Engineering, University of Sistan and Baluchestan, P.O. Box 9816745563-161, Zahedan, Iran

## 1 Introduction

Drought is characterized by water scarcity in one or more components of the hydrological cycle such as rainfall and river flow flux. This phenomenon occurs when the available water of a system is insufficient to meet the needs of at least one of the biological, economic, and social sectors over a significant period of time (Tsakiris et al. 2007). Although there is generally no global definition of drought, droughts can still be used as a climate indicator. Because the indicators provide a quantitative method for determining the beginning and end of the drought and their amount indicates the level of drought severity (Tabari et al. 2013).

For a quantitative analysis of drought, the existence of a specific indicator for the exact determination of wet and dry periods is necessary (da Silva 2004). The greatest focus of scientific references in drought assessment is on climate drought, which can be explained by the wider and more comprehensive information as well as the load factor of rainfall in drought factors (Ferral et al. 2017). Today, drought analysis based on rainfall data is used as the most important factor for the investigation of different types of droughts (Brekhovskikh et al. 2009).

The requirement for the calculation of the Standardized Precipitation Index (SPI) is the availability of analyzing comprehensive rainfall data sets for the region under study. This process involves the restoration of missing data and the utilization of statistical forecasting methods. One of the common tasks in water resource planning is to predict, simulate or construct a model of some hydrological variables such as rainfall, river flows, and flood flows, which are potential or accidental phenomena. For each of the hydrological variables, various factors are measured and recorded. By analyzing these factors, which have occurred and measured in the past, one can come to the conclusion that if it is generalized to the future, it will simplify decision making or simulation of the basin's behavior. Given the technological advances, although this modeling and forecasting in the domain of time and space is not possible, it has many complexities, since this behavior itself depends on many uncertain factors, including pressure, temperature, velocity, and wind direction (Khuram et al. 2017). Due to limitations such as lack of rainfall information on the spatial and temporal scales and the aforementioned complexities, it is practically impossible to use physical models. Today, along with the existing models, newer methods are developed for prediction. Neural networks are new tools that can be used for analyzing and simulating nonlinear and indeterminate systems, where the relationships between components and system parameters are not well-described (Kohzadi et al. 1995).

Past researchers have developed various indicators to monitor the state of drought and to study the effects of this phenomenon. From various drought indicators, the Palmer drought severity index (PDSI), precipitation anomaly index (RAI), deciles, DI, standardized rainfall index (SPI), normal percentage (PN), and RDI have been used in previous studies (Tsakiris et al. 2007).

Understanding that rain has various effects on water resources such as groundwater, surface water, and snow, has led to the formulation and presentation of the SPI index. Currently, the SPI indicator is widely used in research and practice to monitor droughts. This indicator is known as the most suitable indicator for drought analysis, in particular, spatial analysis, due to the simplicity of computing, the use of available rainfall data, the ability to calculate for any desired time scale and the great potential for spatial comparison of results. The most important advantage of the SPI index is its ability to be calculated for different time scales. This indicator can affect the short-term periods of water reserves (including soil moisture) and the effects of long periods of water resources (such as groundwater reservoirs, surface water reservoirs, and

river flow). The shortage of rainfall on a short-term basis causes fluctuations in soil moisture, and in longer periods, causes changes in groundwater resources and reservoir levels (Mishra and Singh 2010).

This study contributes by presenting a hybrid neural network and Gaussian process classification for short term rainfall forecast using based on the SPI drought index. The proposed framework stands for a prediction model based on extensive historical data associated with the status of shortages and surpluses of climate water resources.

To calculate SPI, first, a gamma distribution is fitted to the data of the monthly rainfall or total rainfall in each desired time interval. Afterward, its cumulative probability function is calculated, and then, the values of the SPI index is computed by transferring the accumulated probability to the standardized cumulative distribution (Vlček and Huth 2009).

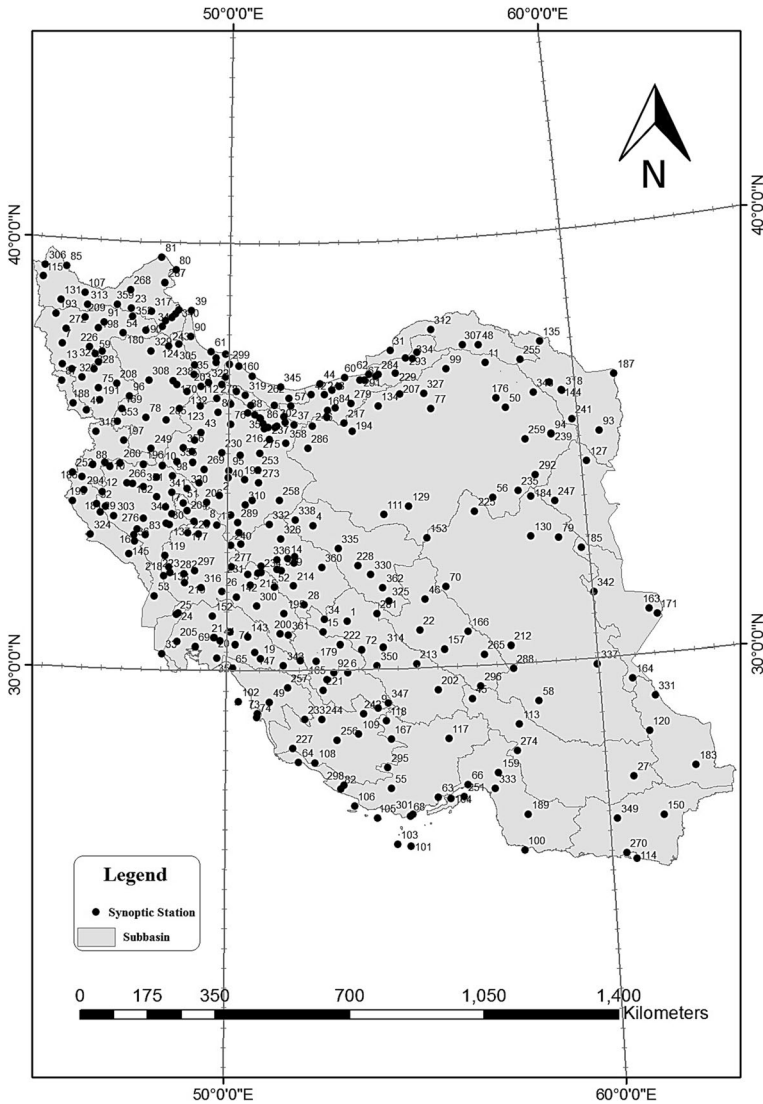
In the calculation of the SPI index, the length of the statistical period and the type of distribution fitted to the frequency of rainfall data are of great importance. Non-selection of long-term courses and the inappropriateness of gamma distribution for rainfall data can lead to an estimation of incorrect values from the SPI index (Mishra and Singh 2010).

The probability distribution function is an effective and useful tool for a comprehensive description of any meteorological or hydrologic variable (Vlček and Huth 2009). In hydrological studies, we try to fit the probabilistic functions to empirically measured and recorded data, and the best function that matches the data is chosen as the probability distribution function (Keyantash and Dracup 2002). Various distribution functions have been used to fit rainfall data, some of which are gamma distributions (Mishra and Singh 2010), Pearson distribution (Guttman 1999), gamma Poisson distribution (Lana et al. 2001), Weibull distribution (Castellví et al. 2004), Lognormal distribution (Shoji and Kitaura 2006), exponential distribution and exponential distribution functions (Wu et al. 2007), and combined exponential distribution (Schoof 2008).

## 1.1 Literature Review

Bhuiyan (2004a) developed the standard water level index (SWI) to monitor rainfall fluctuations in rain drought surveys (Bhuiyan 2004b). Lloyd-Hughes and Saunders (2002) stated that in arid and semi-arid regions, where rainfall distribution is seasonal and rainfall is not normal in some seasons, in some chapters, there will be a lot of zero in the time series. In such areas, the SPI values calculated in short-term timescales due to the high skewness of the rain data may not be distributed normally and the gamma distribution cannot fit well to the rainfall data. Therefore, this can cause errors in determining the data distribution function in arid regions with small data samples. (Vlček and Huth 2009) examined the compliance of daily rainfall in each year and in different chapters of gamma distribution in 90 stations in Europe based on the K-S and corrected K-S tests. They stated that the amount of winter precipitation in more than 40% of the stations studied did not match the two-parameter gamma distribution.

Tigkas et al. (2015) analyzed the drought based on some new drought indices, i.e., Precipitation Deciles (PD) and Streamflow Drought Index (SDI). Shirisha et al. (2019) proposed a real-time rainfall prediction model under data uncertainty for Indian divisions. The developed modeling framework includes an adaptive forecasting module based on grey theory, a runoff component, and a fuzzy updating system. Fu et al. (2019) studied the real-time forecasting problem for the river stage with regression tree analysis. The model calibration was handled using data stream values between 2005 and 2009 in a River Basin in Taiwan.



**Fig. 1** Location of Synoptic stations and the sub-basins of the country

The main aim of this study is to present a modeling framework for short term rainfall forecast using Gaussian process classification (GPC) and backpropagation (BP) artificial neural network (ANN) based on the SPI drought index. The proposed methodology can be used to create a decision-making framework for reducing uncertainty in water resource management calculations, in particular for optimizing the management of groundwater drinking water sources. The calculations here are based on the data obtained from previous calculations of rainfall data and water levels in the aquifer of the country in the period between 2004 to 2015 for the maximum of synoptic stations and observation wells in Iran. Under-ground water levels in 609 study plains in Iran were used to predict drought over the test period extending from 2017 to 2021. The artificial intelligence methods, including the BP-

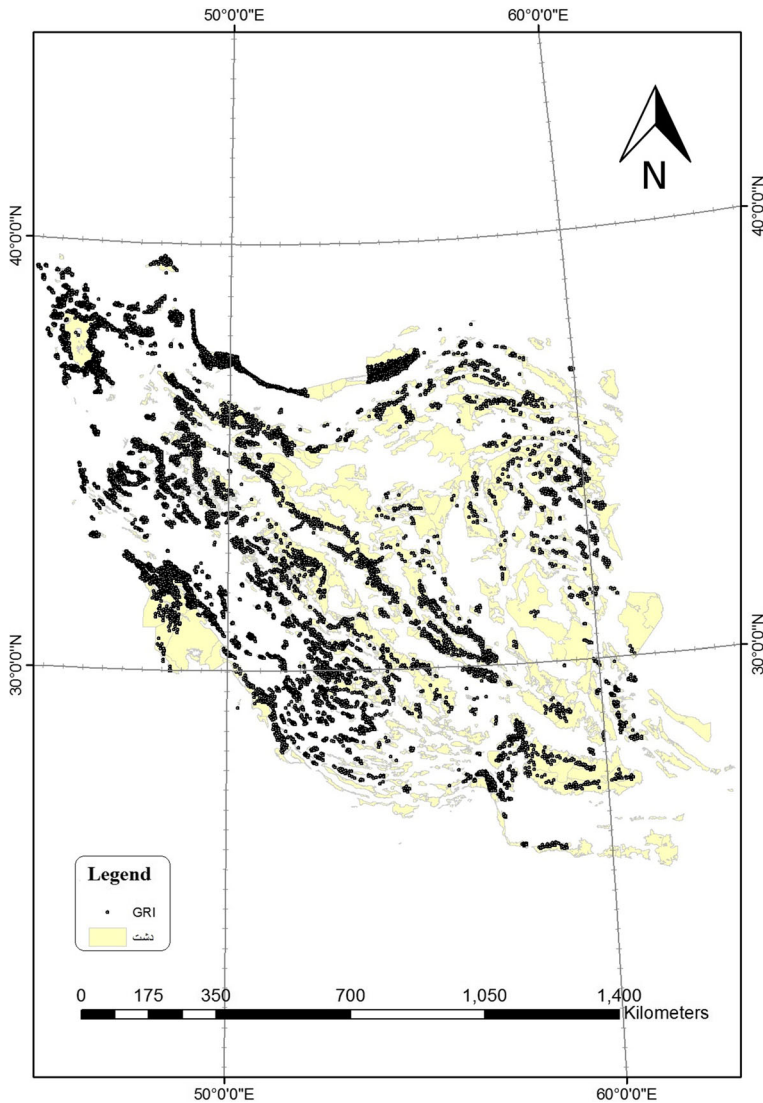


Fig. 2 Location of the Observational wells in the country’s plains

ANN, are implemented in the Python coding environment to achieve an annual precipitation rate. A statistical summary of the rasterized cells of zoning maps is used in order to validate the prediction results. At the end of this study, it is expected that higher risk aquifers, as well as certain areas of Iran that are exposed to severe drought stresses, will be detected with the lowest overall error. Considering the relationship between water quality reductions in Iranian aquifers due to the occurrence of groundwater drought periods, the results are validated by analysis of the effect of climate drought using the SPI index on the occurrence of observed droughts with the GRI index.

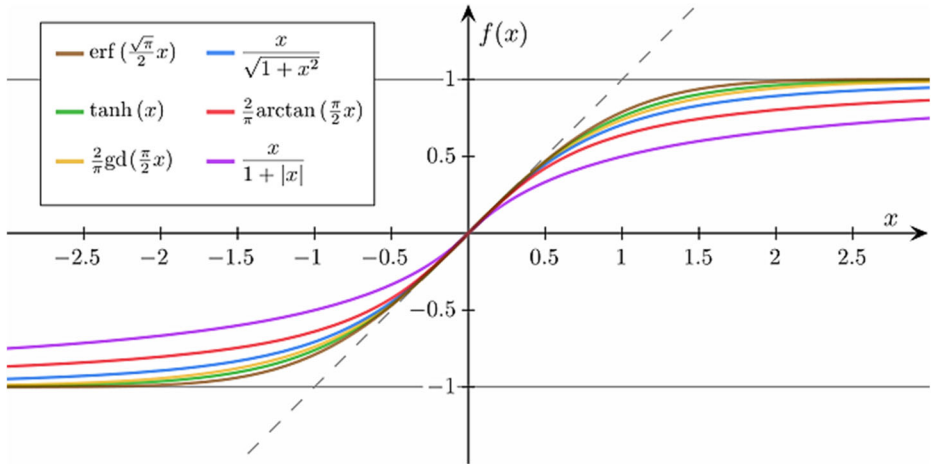


Fig. 3 Sigmoid functions

## 2 Materials and Methods

The study area spans the entire geographical range of Iran (25 to 40 degrees north latitude and 44 to 64 degrees east longitude) and zoning is based on peeling 609 plains in the country. The situation includes climatic variation and climate change in warm and dry areas in the middle, to wet and overcast areas in northern regions. In addition, the unique geological structure and type of each aquifer are enclosed to half free and fully open. The natural condition simultaneously makes all aquifers impossible. In areas with higher population density, like the western regions of Iran, underground aquifers are directly affected by human factors. Quantitative changes followed by the quality of aquifer feeding waters, while in the eastern and central regions, generally, surface currents so much on underground water.

In addition, in the aquifer section, the unique geological form and structure of each aquifer, from enclosed to semi-free and free-form, has made it impossible for natural and simultaneous studies of all aquifers. In areas with higher population density, such as the western regions of Iran, underground aquifers are directly affected by human factors, with a small change in the water supply of aquifers, while in the eastern and central regions, generally, underground is not effective. Figures 1, 2 show the position of Iran’s plains and the position of synoptic stations and observation wells, respectively. More specifically, Fig. 1 shows the geographical distribution of synoptic stations and the sub-basins of the study area.

Table 1 Determination of drought and wetness index of GRI Groundwater (Mendicino et al. 2008)

Class number	Description	Range GRI
1	very intense wetness	[2,∞]
2	high wetness	[1.5, 1.99]
3	Gentle mild wetness	[1, 1.49]
4	Near normal - timid	[0, 0.99]
5	Close to normal - drought	[-0.99, 0]
6	Mild drought	[-1.49, -1]
7	Severe drought	[-1.99, -1.5]
8	Very severe drought	[-∞, -2]

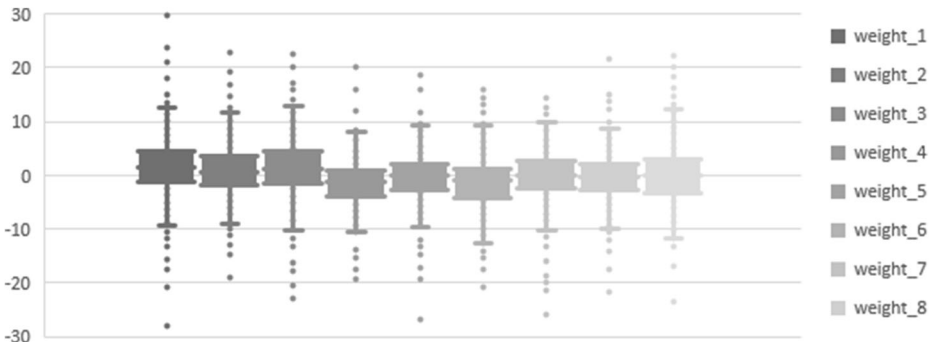


Fig. 4 Box and whisker chart of artificial neural network weights for 362 synoptic stations

In order to predict the annual precipitation digit using post-propagation artificial neural network, it was necessary to prepare the input layer figures. In order to provide the maximum statistical correlation in the input layer column, each time series in 362 synoptic stations of the country was subdivided into  $n$  rows and  $m$  matrix columns. The interval is 0 to 1. In this way, the implementation of the training repetition for the input examples resulted in the optimal extracted relationship. In each implementation of the Python code, the post-propagation ANN was used to generate random weights, which results in the execution of the code for the averaged data. There were 362 stations with precipitation. The output is generated after 1000 training iterations. Thus, new and corrected weights are assigned per year in the GIS environment which can be accessed using scattered points in the coordinates defined by the intermediary tool and continuous raster maps. Finally, we obtained the SPI index from the prediction data.

### 2.1 Backpropagation (BP)

Backpropagation (BP) is a technique that is processed in ANN to calculate the contribution of each neuron error after a batch of data. This is a special case of the older and more general method named automatic differentiation (Basirati et al. 2019). In the learning context, the BP algorithm is typically adjusted using the gradient calculation of the loss function. This method also sometimes has the backpropagation

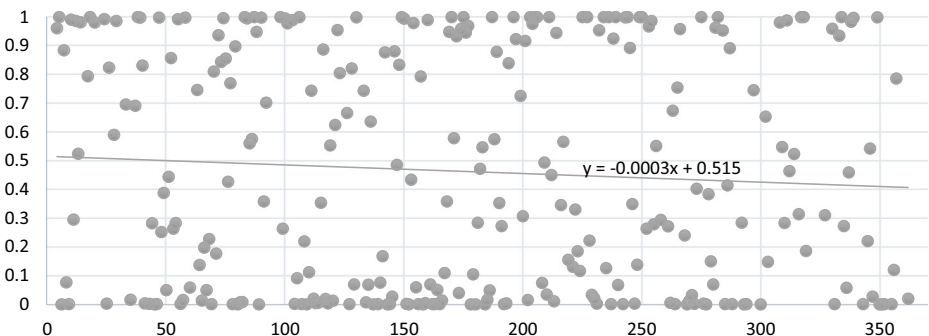


Fig. 5 Distributed Spotted Points of Normal Cultivars Precipitation forecast in 362 synoptic stations

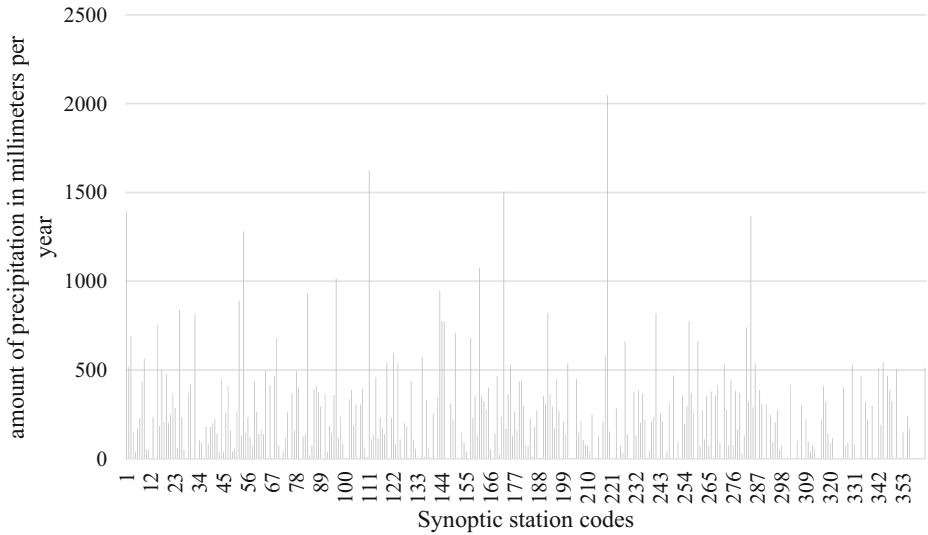


Fig. 6 The forecast of precipitation in 362 synoptic stations of the country (mm per year)

of errors because the error in the output is computed and distributed through network layers (Zhang et al. 1998).

The purpose of any supervised learning algorithm is to find a function that maps the set of inputs to its correct output. The main purpose of the backpropagation is to calculate partial derivatives or gradients,  $\partial E/\partial w$  of a function based on the loss of  $E$  according to any weights,  $w$ , in the network (Schmidhuber 2015). The artificial neural network BP has the following formulation (Nielsen 2015):

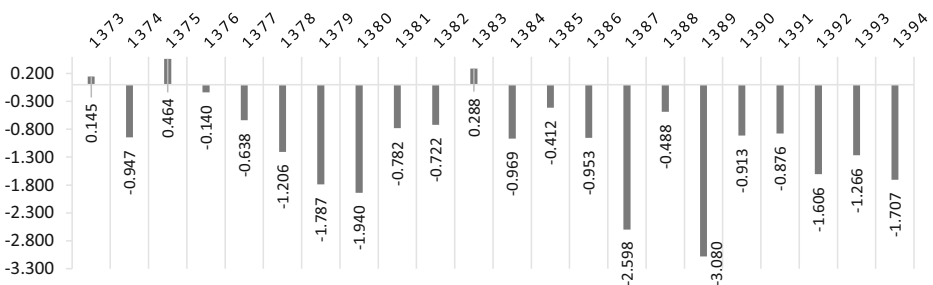


Fig. 7 Average SPI in 362 synoptic stations in the country, 22-year interval



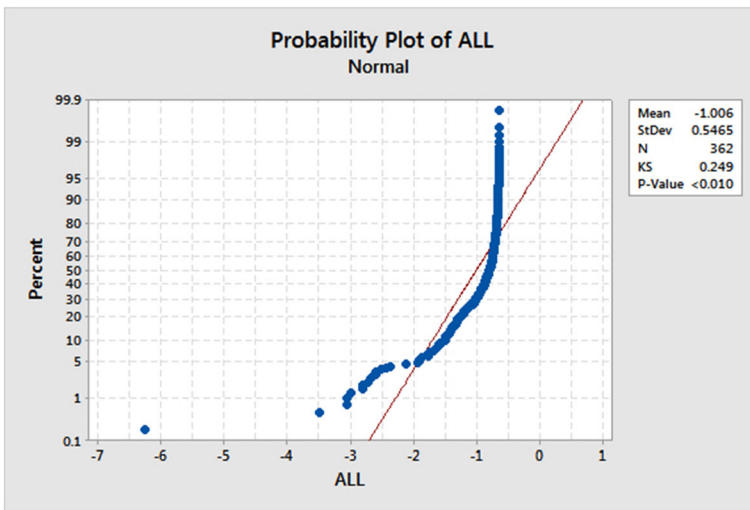


Fig. 8 Distribution of abnormal mean data in interval 1994–2015

$$\begin{aligned}
 a_0^{(2)} &= g\left(\Theta_{00}^{(1)}x_0 + \Theta_{01}^{(1)}x_1 + \Theta_{02}^{(1)}x_2\right) = g\left(\Theta_0^T x\right) = g\left(z_0^{(2)}\right) \\
 a_1^{(2)} &= g\left(\Theta_{10}^{(1)}x_0 + \Theta_{11}^{(1)}x_1 + \Theta_{12}^{(1)}x_2\right) = g\left(\Theta_1^T x\right) = g\left(z_1^{(2)}\right) \\
 a_2^{(2)} &= g\left(\Theta_{20}^{(1)}x_0 + \Theta_{21}^{(1)}x_1 + \Theta_{22}^{(1)}x_2\right) = g\left(\Theta_2^T x\right) = g\left(z_2^{(2)}\right) \\
 h_\Theta(x) &= a_1^{(3)} = g\left(\Theta_{10}^{(2)}a_0^{(2)} + \Theta_{11}^{(2)}a_1^{(2)} + \Theta_{12}^{(2)}a_2^{(2)}\right)
 \end{aligned}
 \tag{1}$$

In eq. (1),  $g$  is a sigmoid function that refers to a particular state of the logistic function and is defined as Eq. (2).

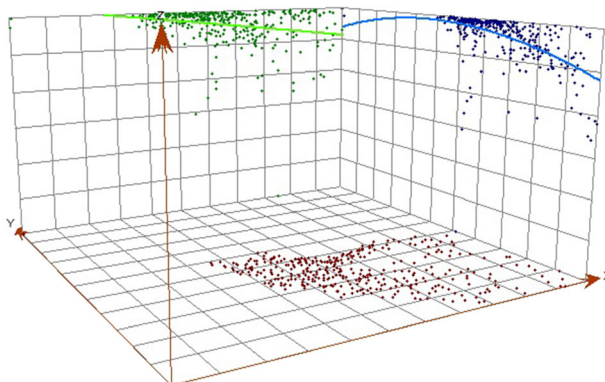


Fig. 9 Trend in the average in interval 1994–2015

**Table 2** Comparison of errors between different methods for the production of continuous interpolation

Error index	Type of interpolation method						
	Kriging						EBK
	Ordinary	Simple			Universal		
	Stable	J-Bessel	Gaussian	Spherical			
Mean	0.005	0.002-	0.003-	0.002-	0.002-	0.005	0.002
RMS	0.499	0.496	0.495	0.496	0.496	0.586	0.514
MS	0.012	0.005-	0.007-	0.005-	0.005-	0.005-	0.003
RMSS	1.120	1.021	1.017	1.021	1.016	1.210	1.059
ASE	0.444	0.485	0.486	0.485	0.488	0.512	0.481
Error index	Type of interpolation method						
	IDW			GPI			RBF
	Power one	Power two	Power three	First grade	second grade	third grade	
Mean	0.005-	0.013-	0.012-	0.001	0.001-	0.000	0.001-
RMS	0.518	0.541	0.559	0.507	0.498	0.496	0.519

$$g(z) = \frac{1}{1 + e^{-z}} \tag{2}$$

The sigmoid is a bounded and positive derivative function. One of the reasons for using the sigmoid function, which is one of the earliest selections of neural networks, is that its derivative has a very good quality. In many of the weight update algorithms, the need to know the derivate. In all of these cases, the derivative function can be expressed on the basis of  $f$  and  $1-f$ . In fact, this is the only class of functions that is desirable:  $f'(t) = f(t)(1-f(t))$ . However, usually, weights are more important than the particular function. The sigmoid functions are very similar, and the output differences are small. In Fig. 3, different types of sigmoid functions are illustrated. In the post-normalized learning method, the normalization of input vectors is not a requirement; however, normalization can improve performance (LeCun et al. 2015).

If the matrix representation is used, the equations of the previous section are converted into eqs. (3) to (8).

$$x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix} \quad z^{(2)} = \begin{bmatrix} z_0^{(2)} \\ z_1^{(2)} \\ z_2^{(2)} \end{bmatrix} \tag{3}$$

$$z^{(2)} = \Theta^{(1)}x = \Theta^{(1)}a^{(1)} \tag{4}$$

$$a^{(2)} = g\left(z^{(2)}\right) \tag{5}$$

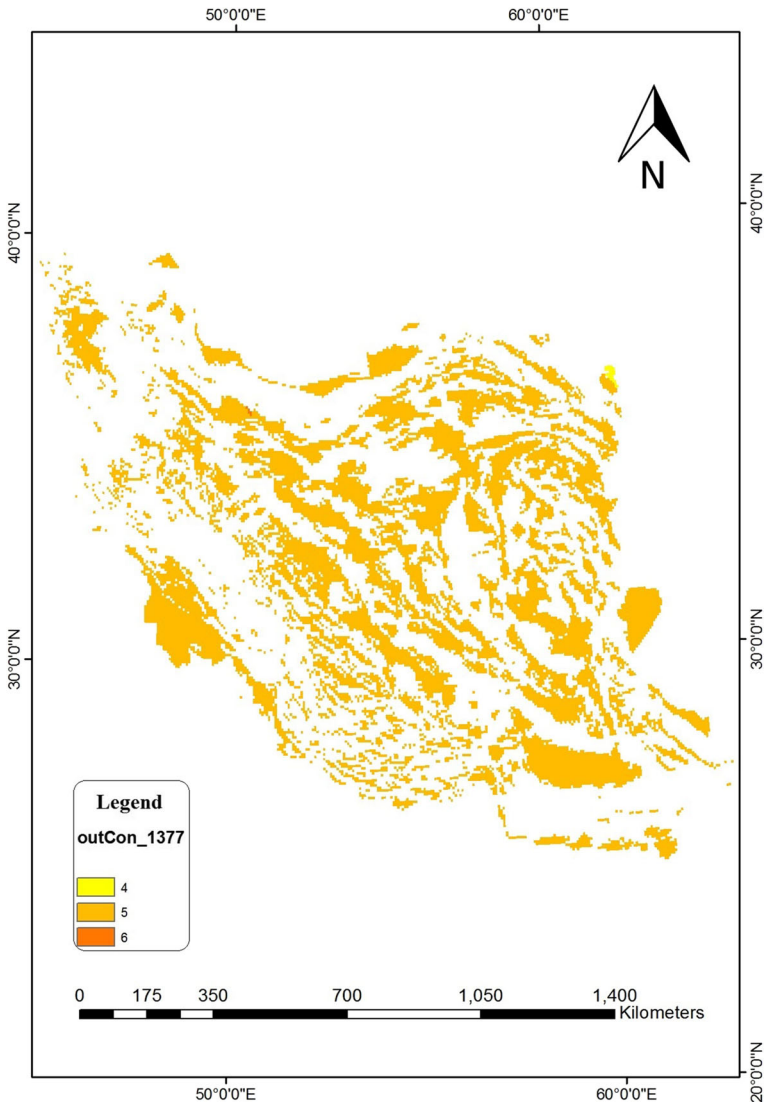


Fig. 10 SPI calculation in the year1998

$$a_0^{(2)} = 1.0 \tag{6}$$

$$z^{(3)} = \Theta^{(2)} a^{(2)} \tag{7}$$

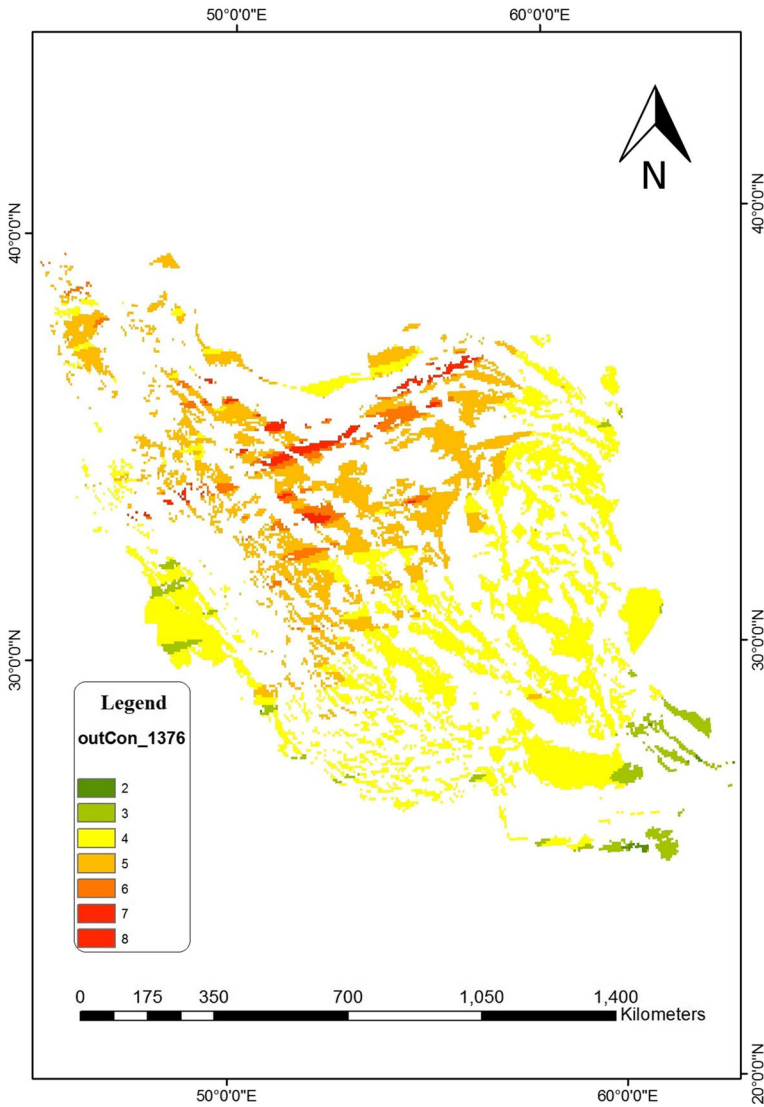


Fig. 11 SPI calculation in the year1997

$$h_{\Theta}(x) = a^{(3)} = g(z^{(3)}) \tag{8}$$

The learning algorithm can be divided into two stages: 1) release and 2) update of weights (Li et al. 2009). The update process includes the removal of a gradient of weight. This percentage affects the speed and quality of learning, which is also called the learning rate. The gradient characteristic of a weight that indicates that the error is increasing, is why weight should be updated in the opposite direction. Stages 1 and 2 are repeated until the network performance is satisfactory. For the output unit ( $L = 3$ ), if an error from node  $j$  in layer  $l$  is denoted by  $\delta_j^{(L)}$ , the

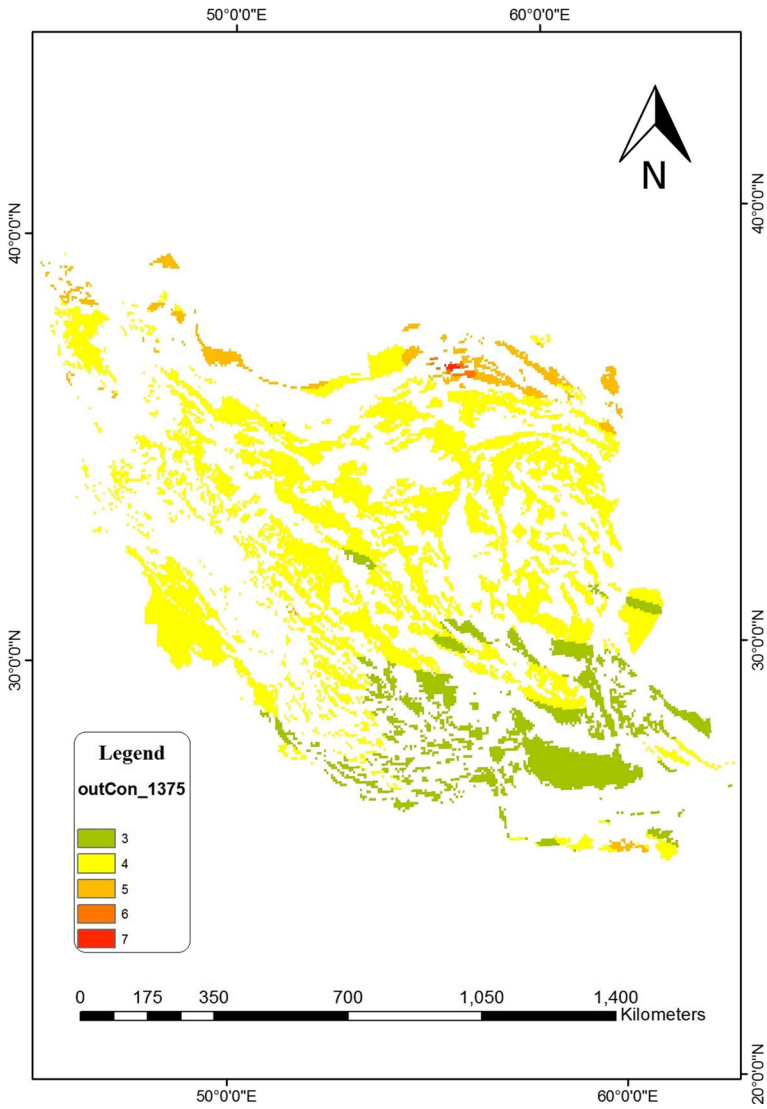


Fig. 12 SPI calculation in the year1996

actual value activation is equal to:

$$\delta_j^{(3)} = a_j^{(3)} - y_j = h_{\Theta}(x) - y_j \tag{9}$$

If the vector format is used, then:

$$\delta^{(3)} = a^{(3)} - y \tag{10}$$

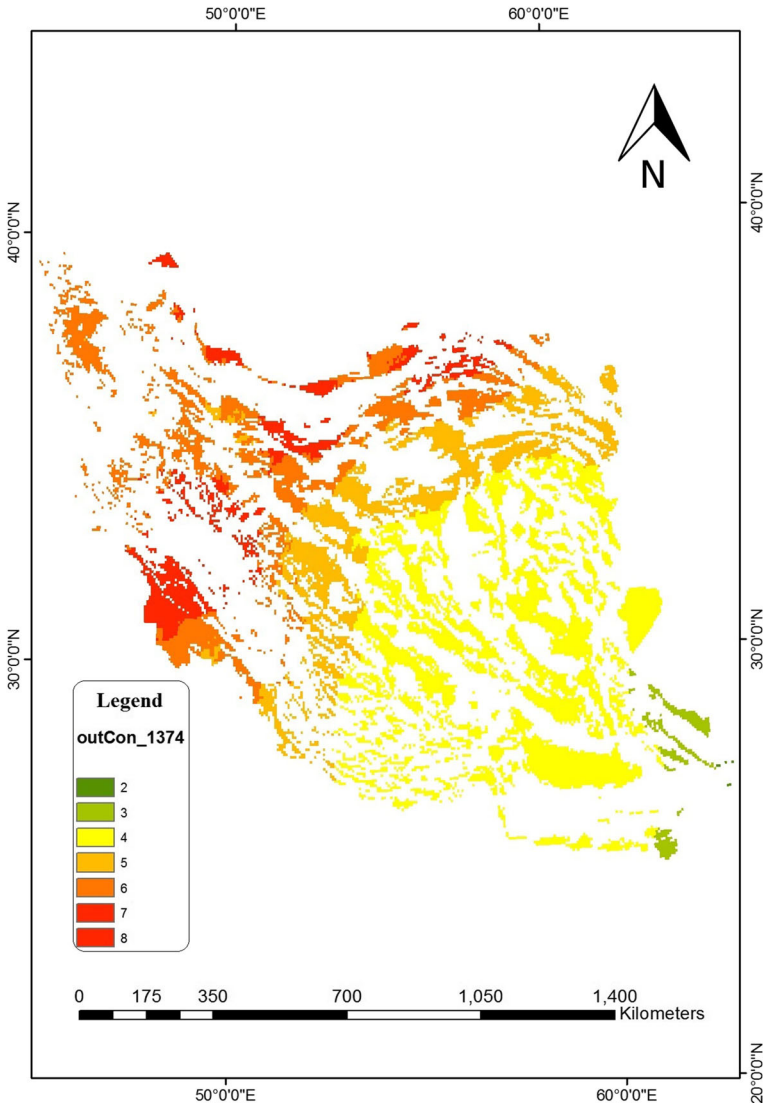


Fig. 13 SPI calculation in the year1995

$$\delta^{(2)} = \left(\Theta^{(2)}\right)^T \delta^{(3)} .g' \left(z^{(2)}\right) \tag{11}$$

It should be noted that the condition  $\delta^{(1)}$  does not exist because the input layer is considered as observational values and is used as a training set. Therefore, there are no errors with input (Schmidhuber 2015). Correspondingly, the derivative of the cost function can be expressed as Eq. (6).

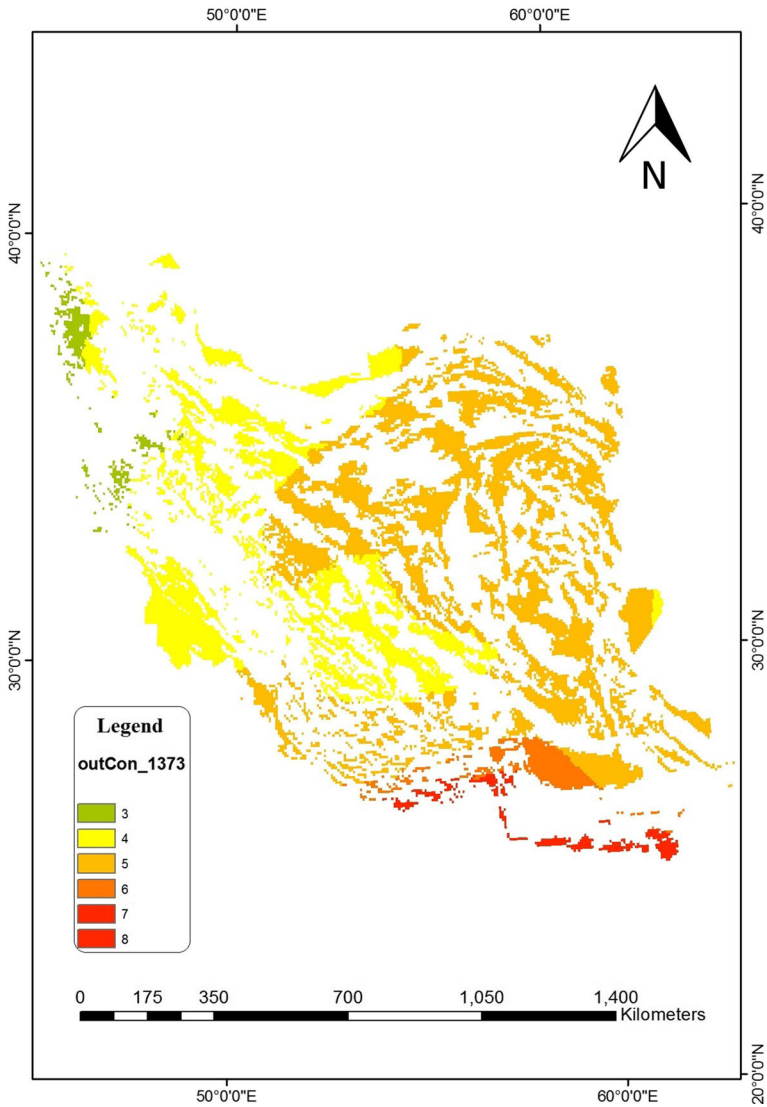


Fig. 14 SPI calculation in the year 1994

$$\frac{\delta}{\delta \Theta_{ij}^l} J(\Theta) = a_j^{(l)} \delta_i^{(l+1)} \tag{12}$$

This amount is used to update the weight and also the training rate can be multiplied by the weight adjustment.

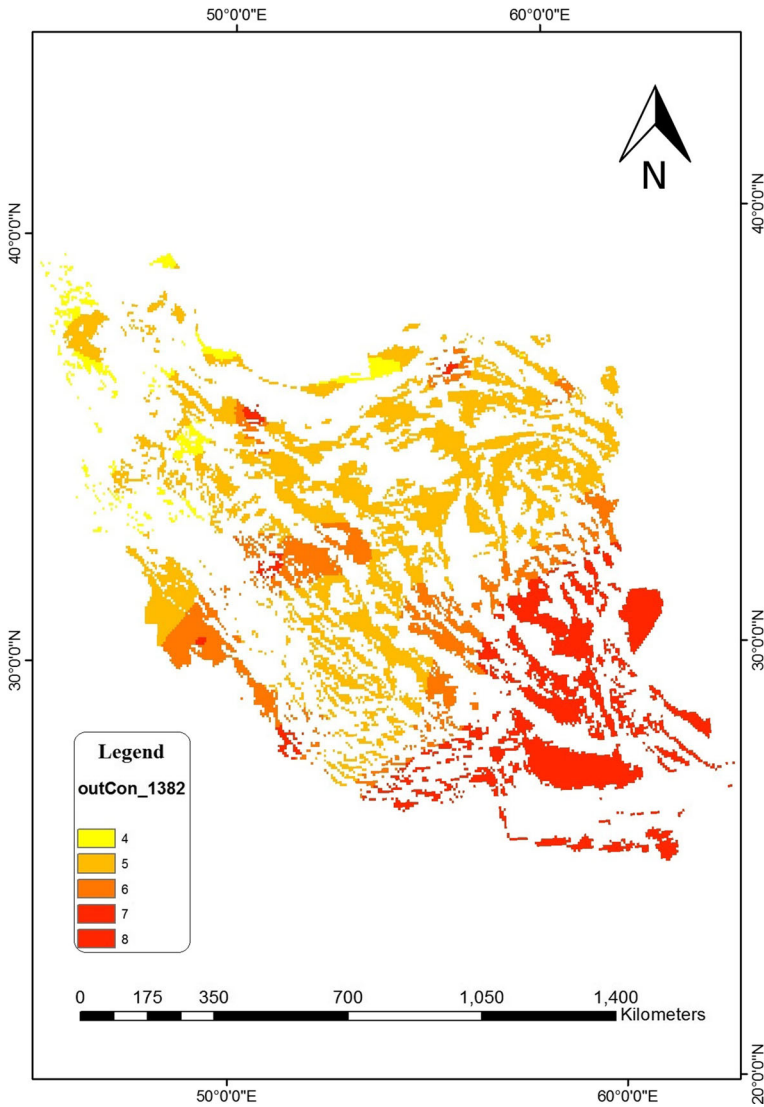


Fig. 15 SPI calculation in the year 2003

### 2.2 Mathematical Theory of the SPI Index

The value of the index based on the probability density function of the gamma probability distribution for  $x > 0$  and is calculated by:

$$g(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \tag{13}$$

where  $\alpha > 0$  is the shape parameter,  $\beta > 0$  is the gamma- parameter and  $x > 0$  is the cumulative precipitation value. In this regard,  $I(x) < 0$  is also a gamma function that can be defined by:



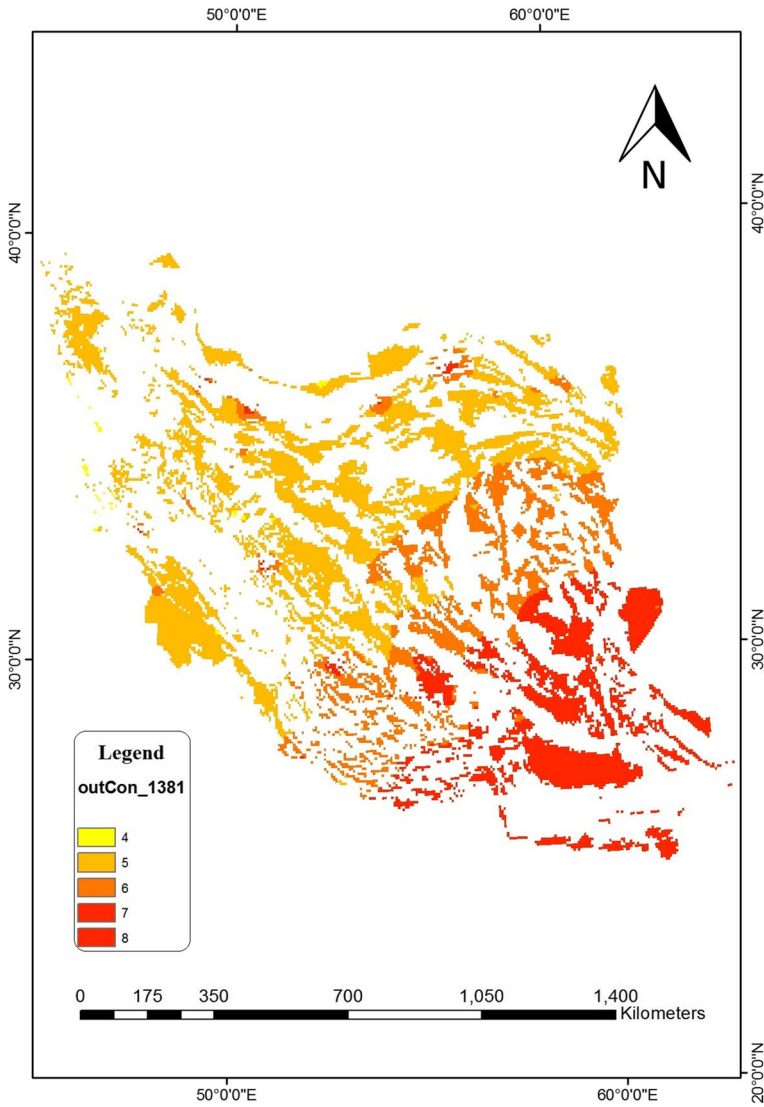


Fig. 16 SPI calculation in the year 2002

$$\Gamma(\alpha) = \lim_{n \rightarrow \infty} \prod_{v=0}^{n-1} \frac{n!n^{y-1}}{y+v} \equiv \int_0^{\infty} y^{\alpha-1} e^{-y} dy \tag{14}$$

In order to fit a gamma distribution to a dataset, it is necessary to estimate  $\alpha$  and  $\beta$ . (Edwards 1997) proposed the use of the exponential maximal method for estimating these two parameters:

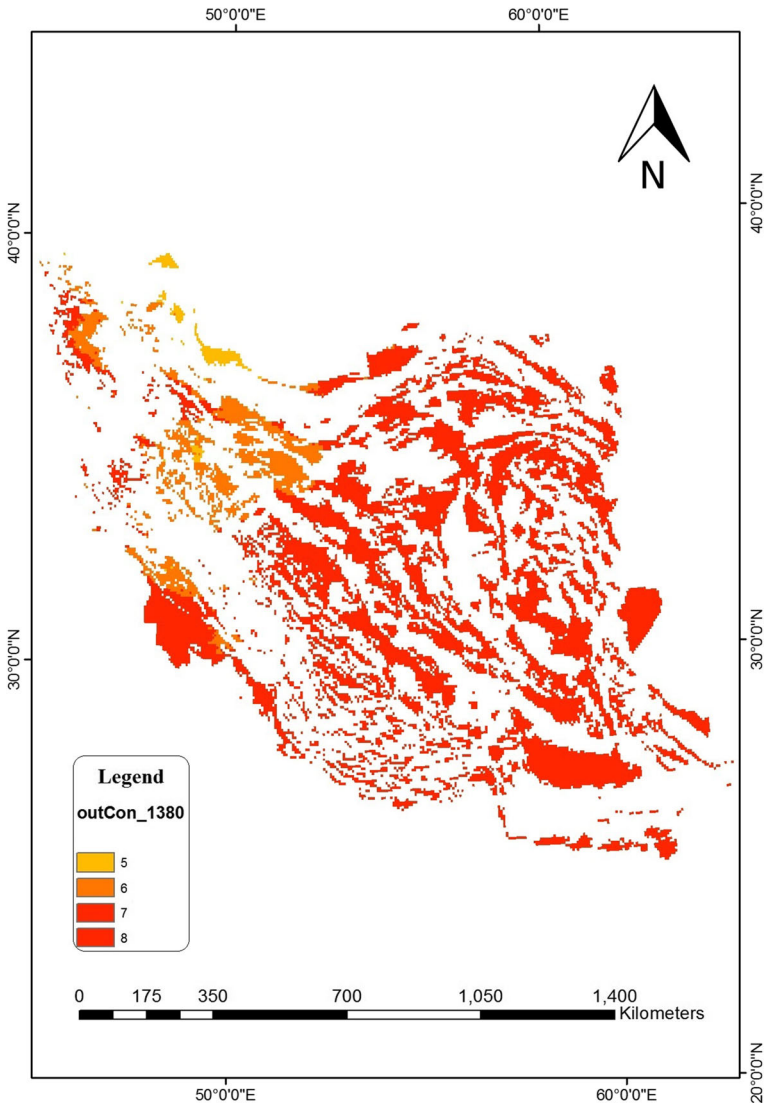


Fig. 17 SPI calculation in the year 2001

$$\tilde{\alpha} = \frac{1}{4A} \left( 1 + \sqrt{1 + \frac{4A}{3}} \right) \tag{15}$$

$$\tilde{\beta} = \frac{\bar{X}}{\tilde{\alpha}} \tag{16}$$

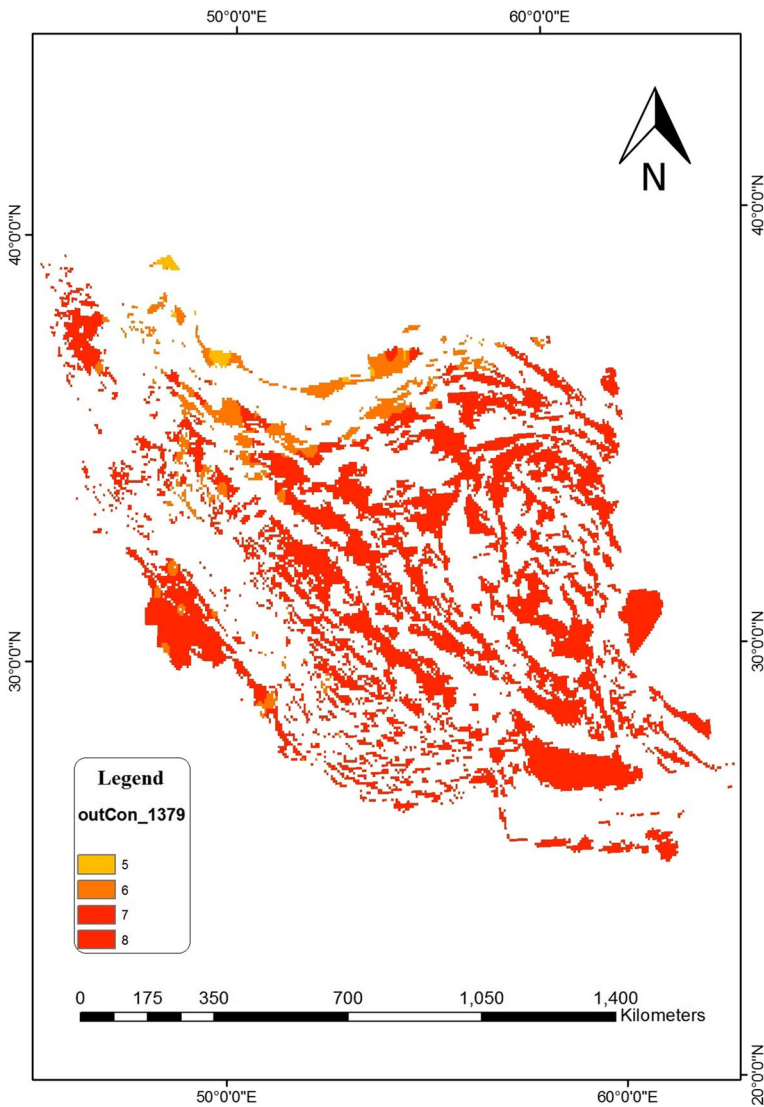


Fig. 18 SPI calculation in the year 2000

In Eq. (15), the value of A for  $n$  observations is estimated using:

$$A = \lim \left( \bar{x} \right) - \frac{\sum \ln(x)}{n} \tag{17}$$

In this equation,  $n$  is the number of observations in which the desired data is available, and  $\bar{x}$  is also the average of the data for the desired time interval (e.g. monthly, quarterly, yearly, etc.). Using the estimated parameters in eqs. (15) and (16), we can calculate the cumulative probability of the data at the desired scale for each reconstruction parameter. Assuming that  $t = \frac{x}{\beta}$ , then the cumulative probability becomes an incomplete gamma function which is

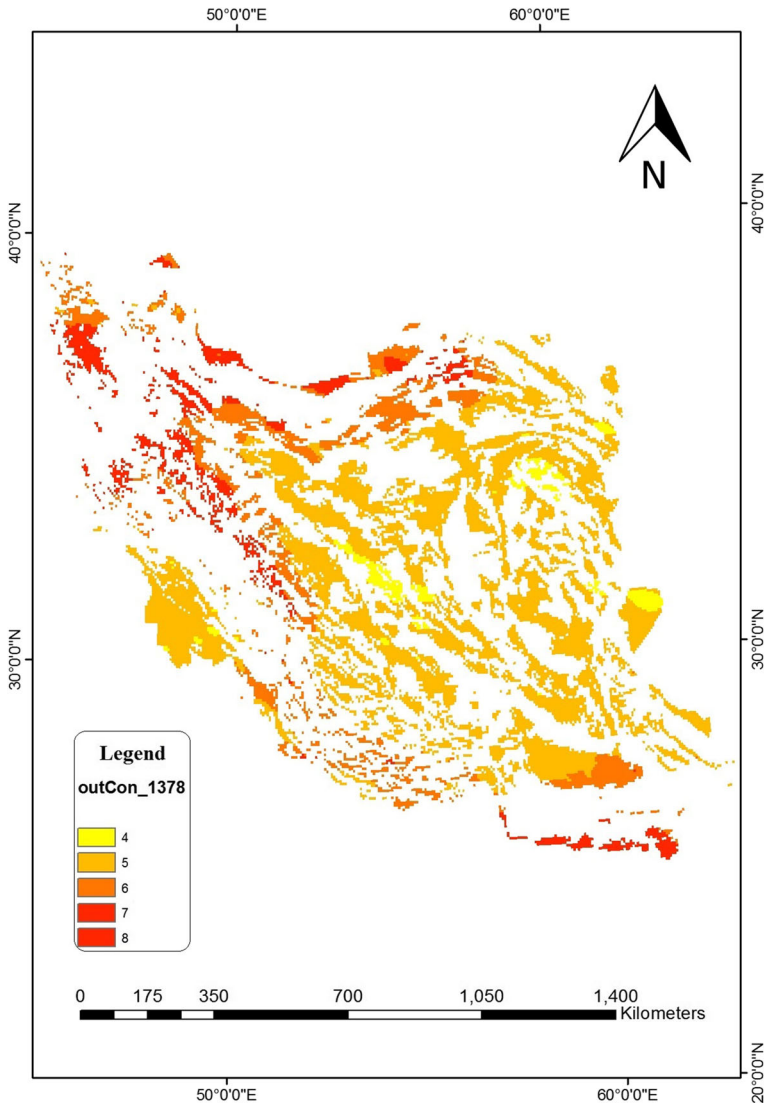


Fig. 19 SPI calculation in the year 1999

defined by:

$$G(x) = \int_0^x g(x) dx = \frac{1}{\Gamma(\tilde{\alpha})} \int_0^x t^{\tilde{\alpha}-1} e^{-t} dt \tag{18}$$

Since the gamma function for  $x = 0$  cannot be defined and precipitation data always includes a large number of observations with a value of 0, the cumulative probability of the data is converted into eq. (19).

$$H(x) = q + (1-q)G(x) \tag{19}$$

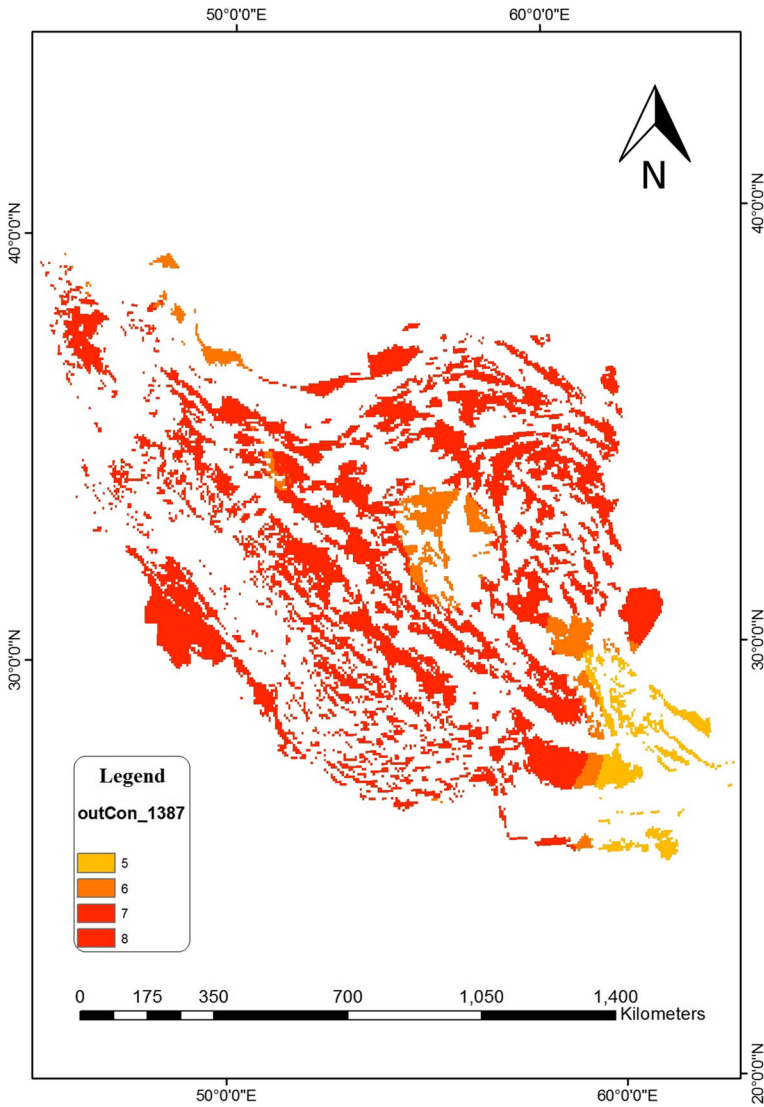


Fig. 20 SPI calculation in theyear2008

where the value of  $q$  is equal to the probability of the value of the data. Finally, using eqs. (20) and (21),  $H(x)$ , is transmitted to the standard normal distribution with mean zero and standard deviation of 1, which results in the SPI value.

$$Z = \text{SPI} = -\left( t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3} \right) \quad 0 < H(x) < 0.5 \quad (20)$$

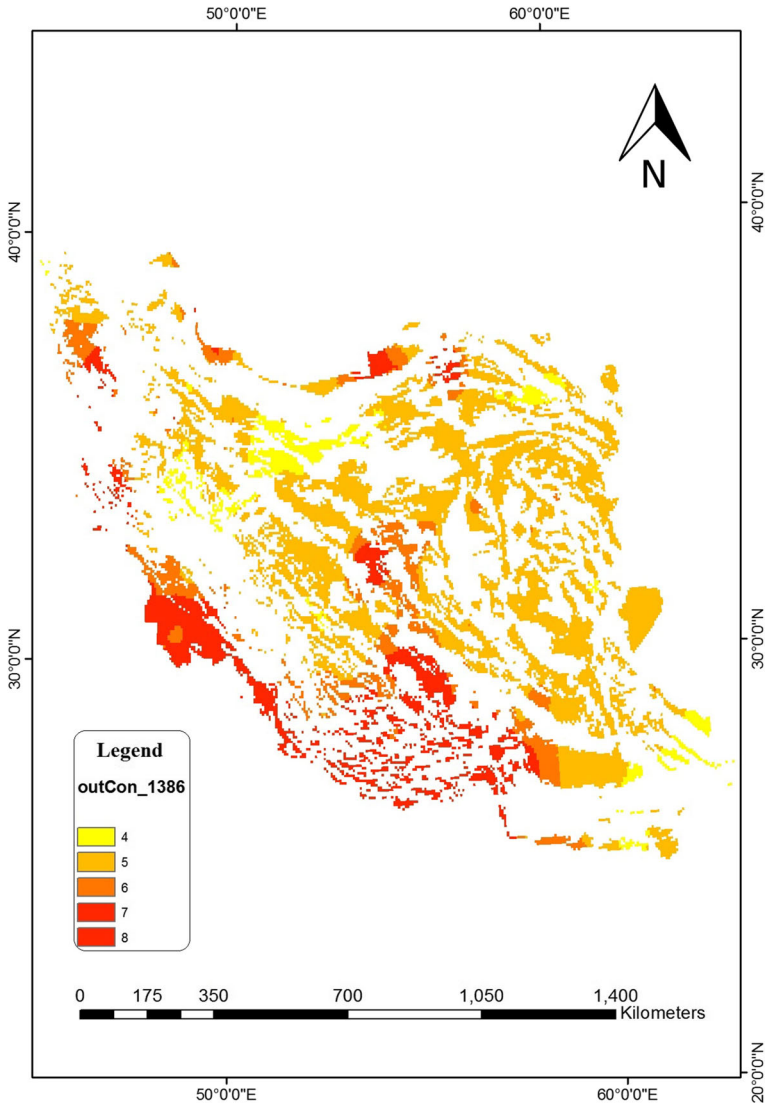


Fig. 21 SPI calculation in theyear2007

$$Z = \text{SPI} = + \left( t - \frac{c_0 + c_1t + c_2t^2}{1 + d_1t + d_2t^2 + d_3t^3} \right) 0.5 < H(x) < 1 \tag{21}$$

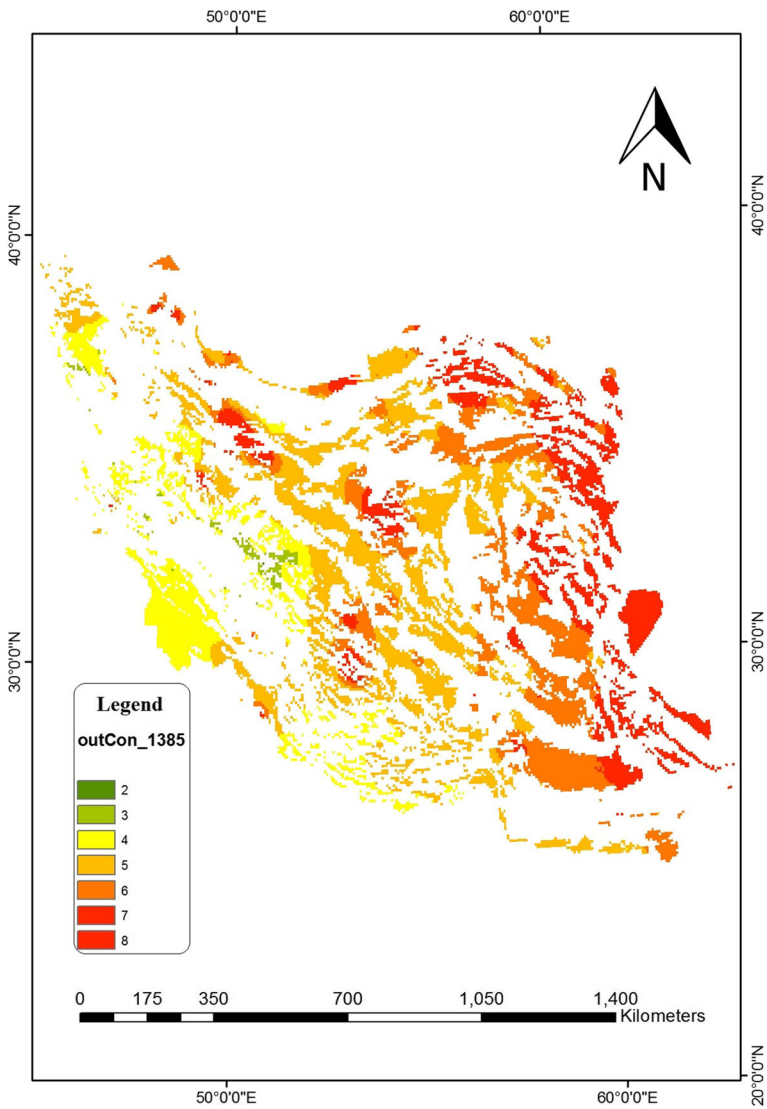


Fig. 22 SPI calculation in theyear2006

The t component is also obtained from eqs. (22) and (23), respectively.

$$t = \sqrt{\ln\left(\frac{1}{(H(x))^2}\right)} \tag{22}$$

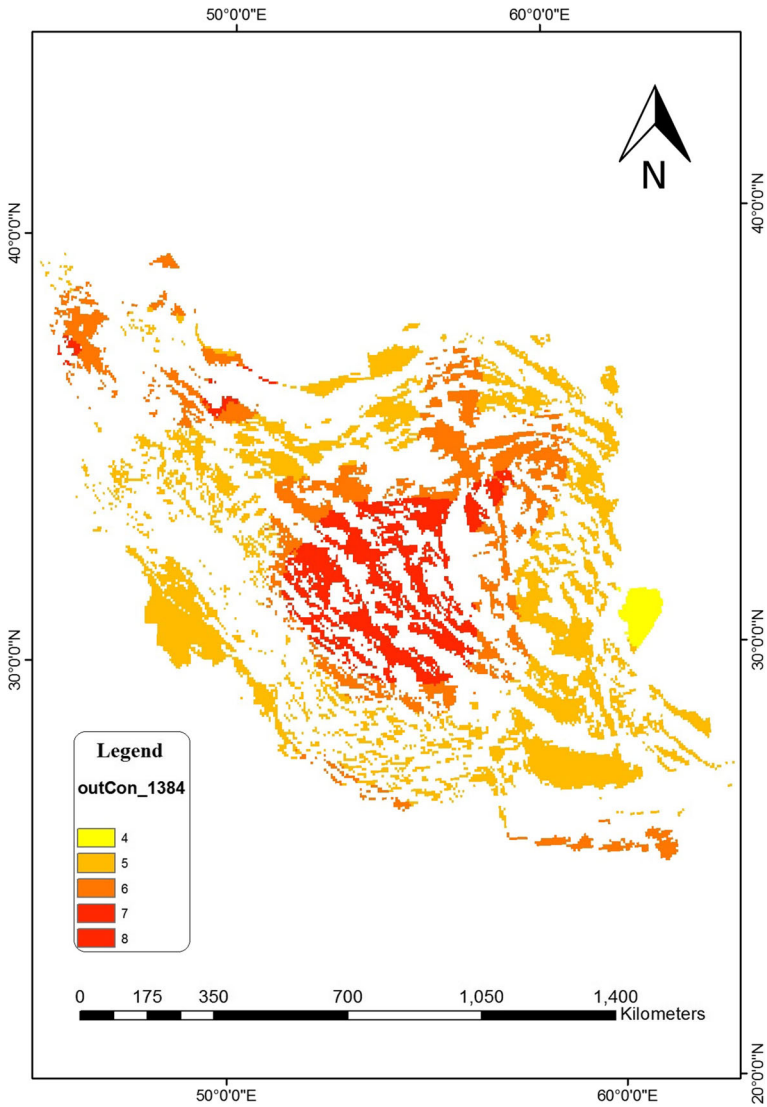


Fig. 23 SPI calculation in theyear2005

$$t = \sqrt{\ln\left(\frac{1}{(1-H(x))^2}\right)} \tag{23}$$

The components  $c_0, c_1, c_2, d_1, d_2, d_3$  are also constant coefficients. Following this approach, the SPI values are equal to the standard score in the standard normal distribution and can be classified as in Table 1.



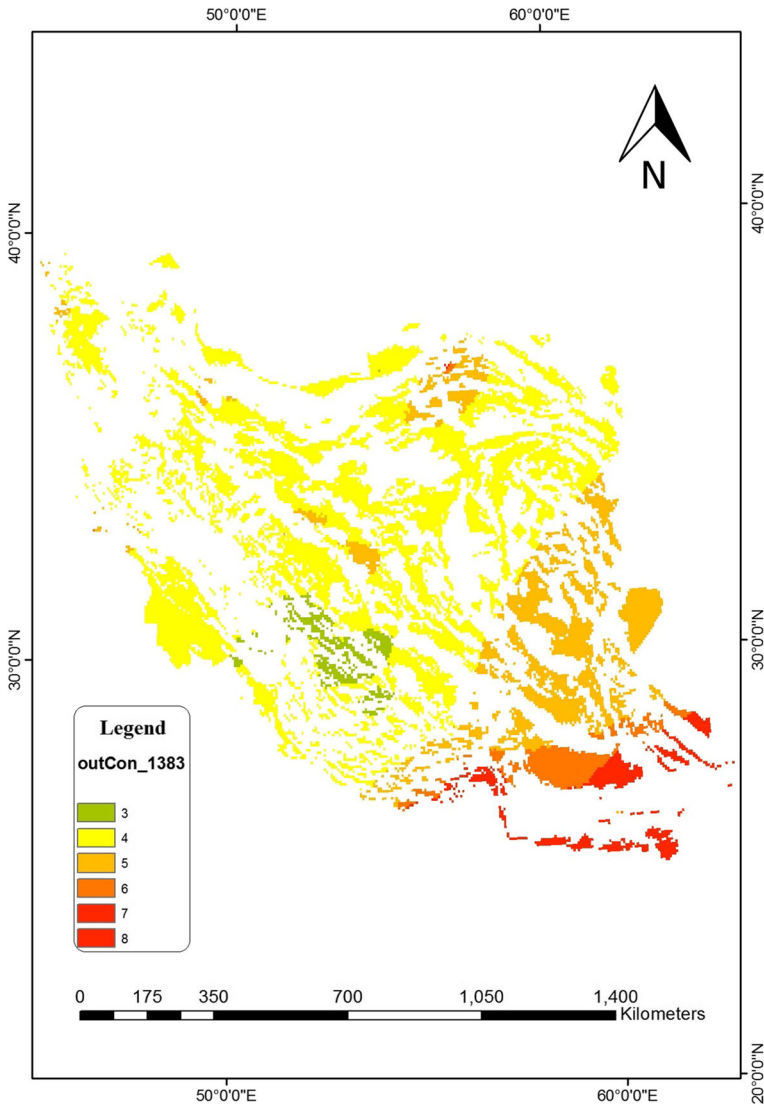


Fig. 24 SPI calculation in theyear2004

### 3 Results and Discussion

In order to predict the annual precipitation rate in the year 2015, there is a need to provide the format of the input layer digits. This step was performed using the Python code of the artificial post-propagation network, according to the aforementioned descriptions in the methods section. In order to provide the maximum level of statistical relationships in the input layer column, each time series in 362 synoptic stations of the country was divided into 12 rows and 9 columns. In each data section, the cumulative rainfall data were used as an output, and for each element of output,

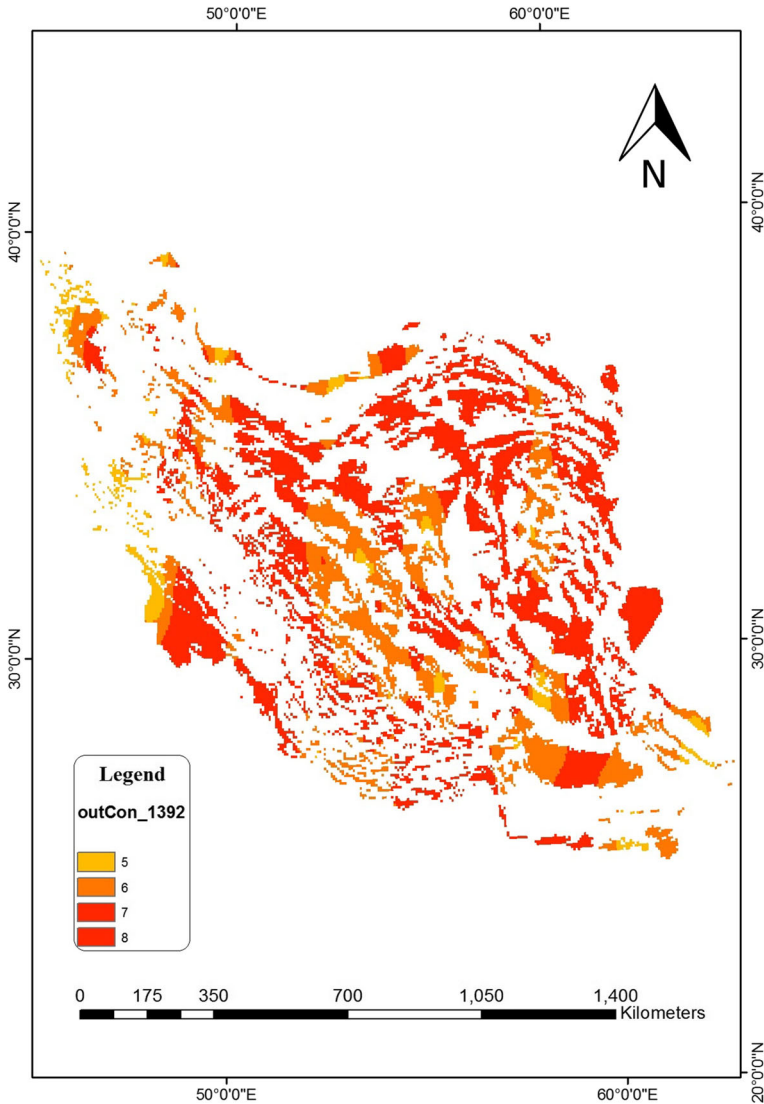


Fig. 25 SPI calculation in theyear2013

nine successive rainfall variables were used. Also, one of the requirements for using post-propagation neural network code is to supply input variables in the normal structure and in the range of 0 to 1. By implementing an algorithm for the training data set, an optimal relationship was obtained. Finally, for all 362 synoptic stations in the country, normalized tables are obtained for educational examples. In each run of the post-propagation artificial neural network model, a randomized weighing order was used. The output of this command was normalized using 1000 training repetitions and the allocation of new and corrective weights with a value of 0.862 and equal to 3,780,300 mm per year for the average of all stations. Similarly, the random weight is

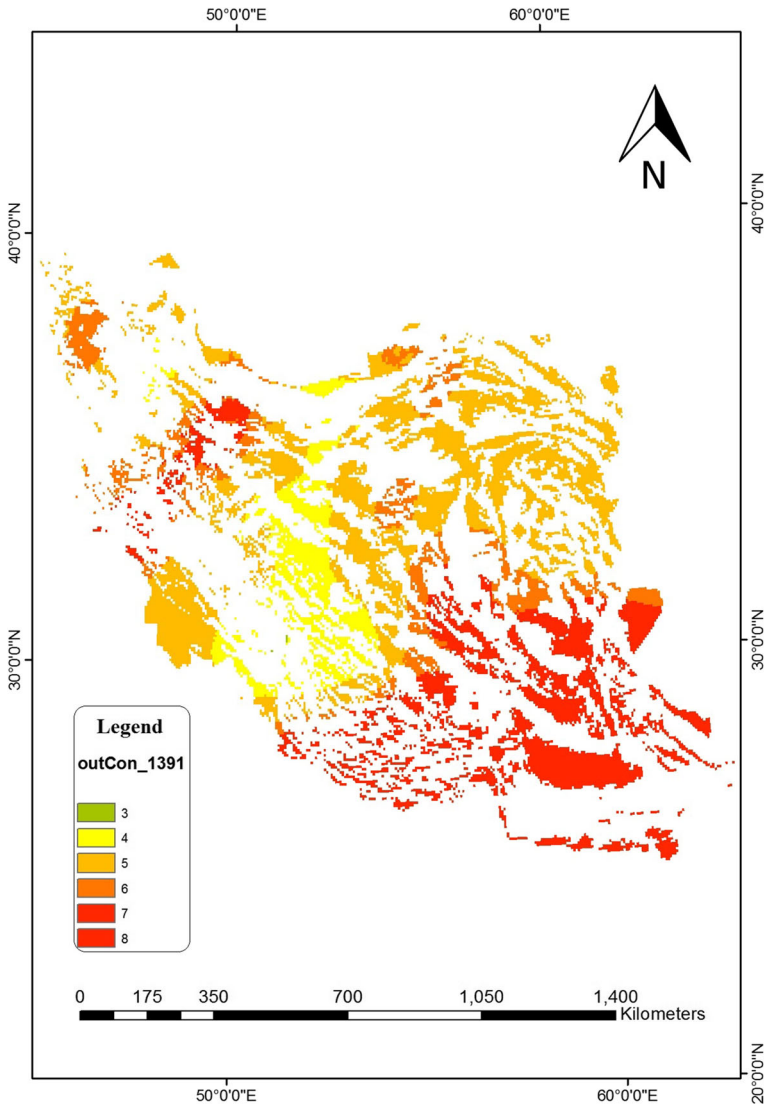


Fig. 26 SPI calculation in theyear2012

generated for all 362 synoptic stations as illustrated in Fig. 4. Cultivars can show the trend of changes to the artificial neural network model. In Fig. 4, the maximum range of variations is for weight No. 1 and the lowest is for weight No. 6. The Box and whisker chart shows that the average weight variations are around 0 and the range of variations in all cases is approximately the same and at the same time has been symmetric for the positive and negative range. This can put the output guess around an average.

As a result of the implementation of training and extracting the solution for the final matrix, as specified in Fig. 5, each of the points in the standard normal space ranging from 0 to 1,

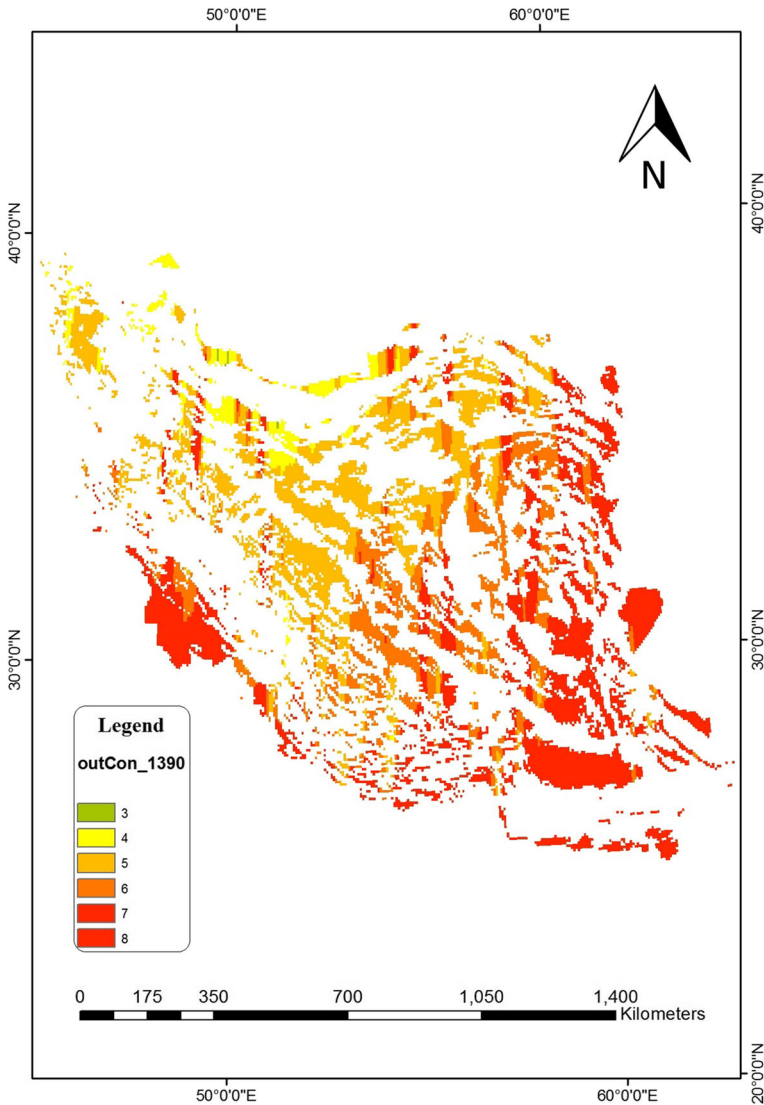


Fig. 27 SPI calculation in theyear2011

specifies a predicted digit that corresponds to the values for all synoptic stations in 2013, as shown in Fig. 6. The scattered points, which are arranged in an ascending order based in the area of the plain, indicate that the accumulation and multiplication of more than half of the predicted values on the right side of the graph (wide plains) is less than the mean long term, and in Meanwhile, with a negative slope of close to zero, the variables predicted a relatively straight regression line. This suggests a repeat of the relative increase in drought in 2015 for the eastern and southeastern parts of the country, which generally have a larger share of the vast plains.

The predicted precipitation levels, before the precipitation change in millimeters per year, is roughly equal to the long-term average and, with a relative decline for vast plains. However,

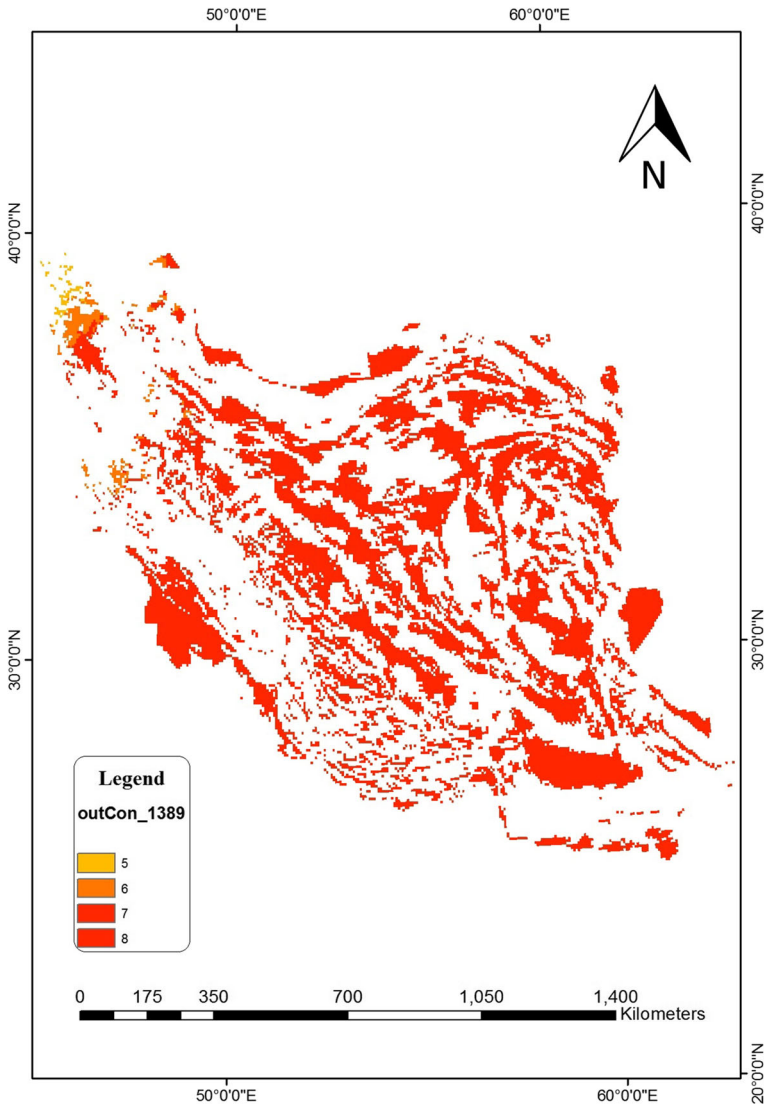


Fig. 28 SPI calculation in theyear2010

the process of change is not necessarily linear and varies from one station to another. This change after the extraction of the SPI drought index is derived from the standard normal transformations based on the sigmoid function. According to the results of the model, the range of drought to areas of the country that was located in near-normal conditions in 2014 has increased at the time of forecasting. After the development of the SPI index model in the spreadsheet environment, the average annual aggregated rainfall drought index for all stations and in all computational years was calculated and shown in Fig. 7.

With the purpose of describing the spatial distribution of the SPI index around the study area for all stations, the production of interpolation maps required the analysis and processing

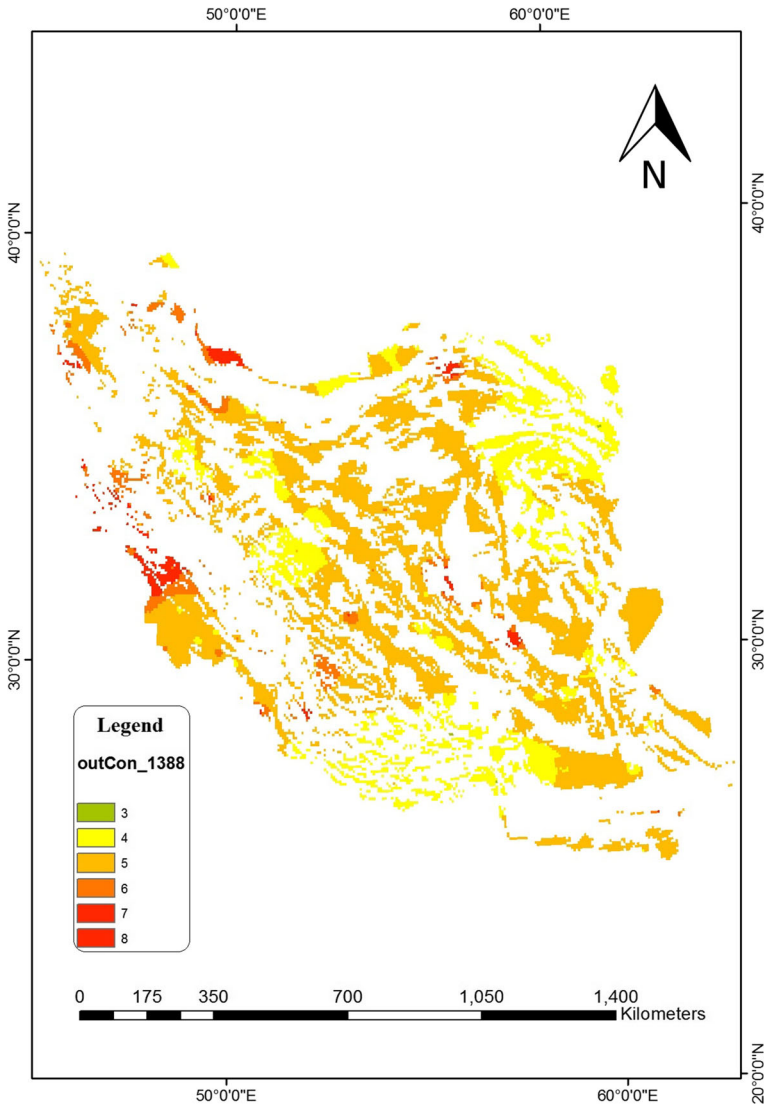


Fig. 29 SPI calculation in theyear2009

of raw data. In this approach, firstly, through a survey tool for non-random processes in data, this value was obtained in a set of three-dimensional diagrams. Figure 7 shows the existence of a non-randomized second-order trend in the direction of north to south and east to west in the initial mean data of the SPI index in the long term.

Given the initial assumptions in the Kriging method, the normalization of data is a fundamental condition. In this study, Minitab software was used to analyze the statistical distribution of data. According to the results of the analysis, the outcomes confirmed the distribution of abnormalities in all years. Figure 8 illustrates the distribution of abnormal mean data in interval 1994–2015. Also, Fig. 9 shows the

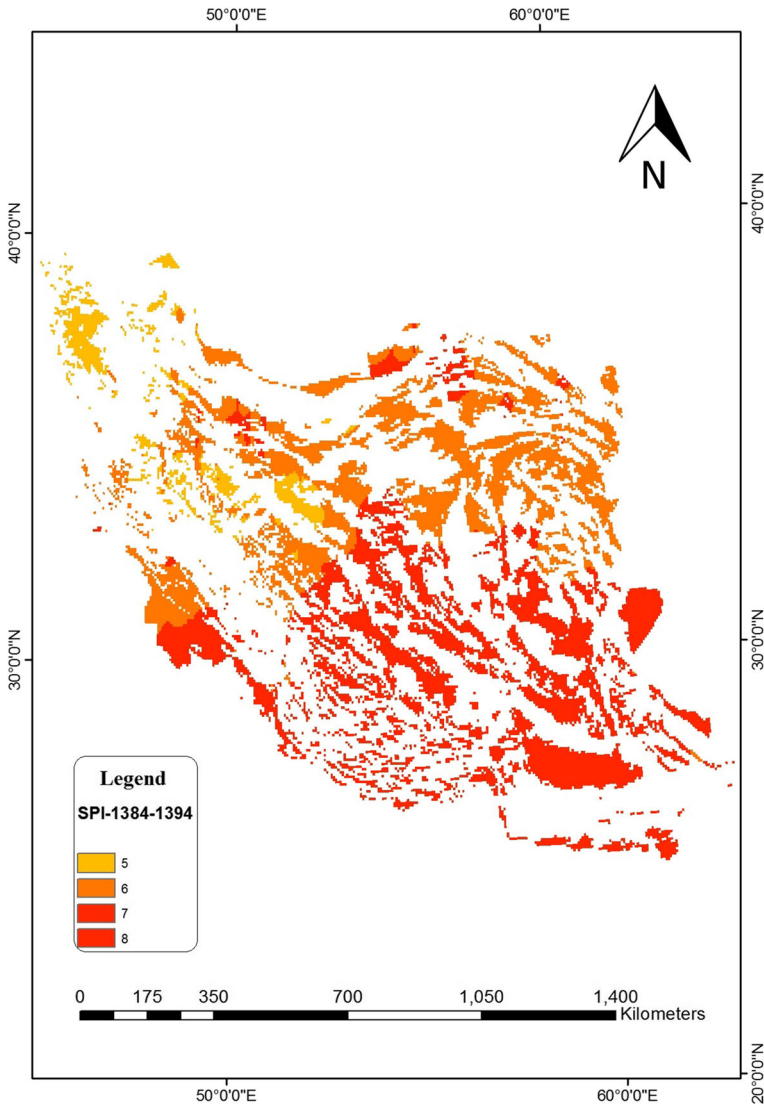
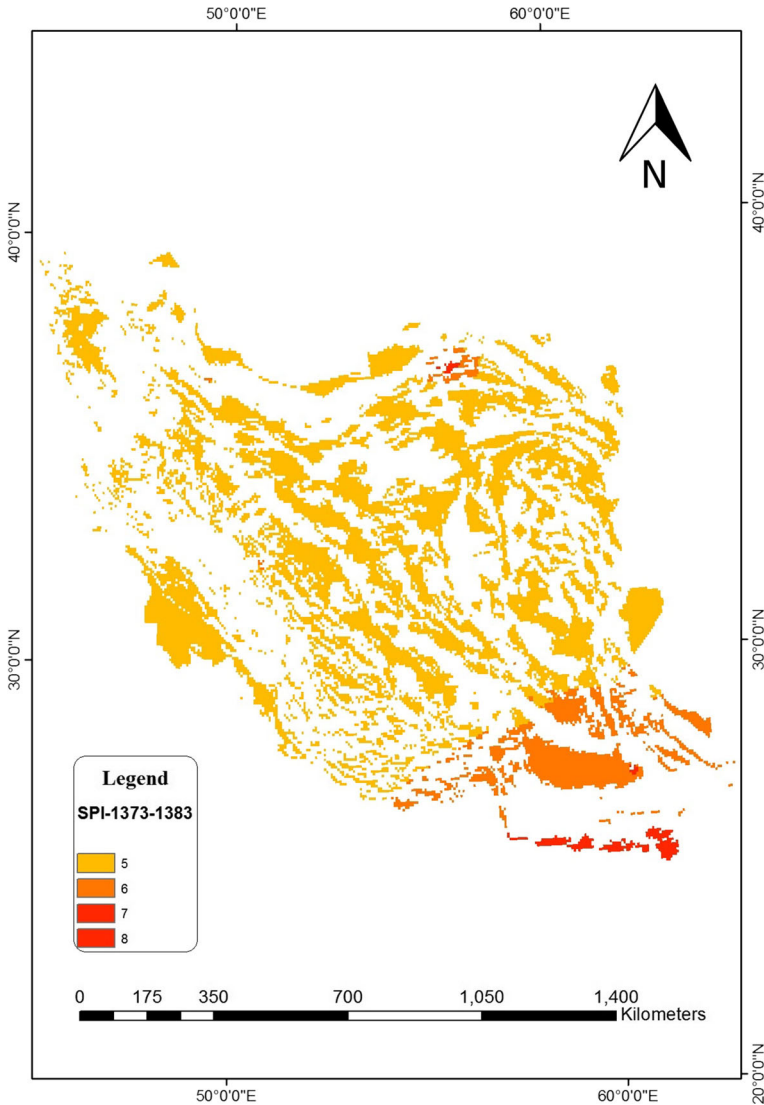


Fig. 30 SPI calculation in years2005–2015

unpredictable statistical distribution of the average SPI index data in a long-term interval.

The main factors of creating interlayer layers have been recognized using the GeoStatistical tool. Accordingly, the continuous levels of static Markov chain of these indexes were developed for ten year intervals. The choice of optimal method was based on a comparison between the produced samples during the range from 1994 to 2015. The results of this experiment are presented in Table 2. For this purpose, two distinct methods of Inverse Distance Weighting (IDW) with three potentials 1, 2 and 3 were used. In addition, using the GPI method with degrees 1, 2 and 3, the RBF method and the Ordinary and Universal Kriging as well as



**Fig. 31** SPI calculation in years 1994–2004

Simple Kriging with four Stable, J-Bessel, Gaussian, and Spherical functions were continuously mapped.

In the Empirical Bayesian Kriging method, using 100 repetitions of the Semivariogram, a continuous layer was generated which, in comparison with all samples, ultimately reduced the estimation errors. More specifically, the mean error was reduced to 0.002, the root mean square error (RMSE) to 0.514, standardized error to 0.003, Standardized Square Root Error to 1.059, and finally Standard Mean Error to 0.481. The error was due to the simple Kriging method selection with the Stable function as the final approach. In Figs. 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, the result of the conditional code



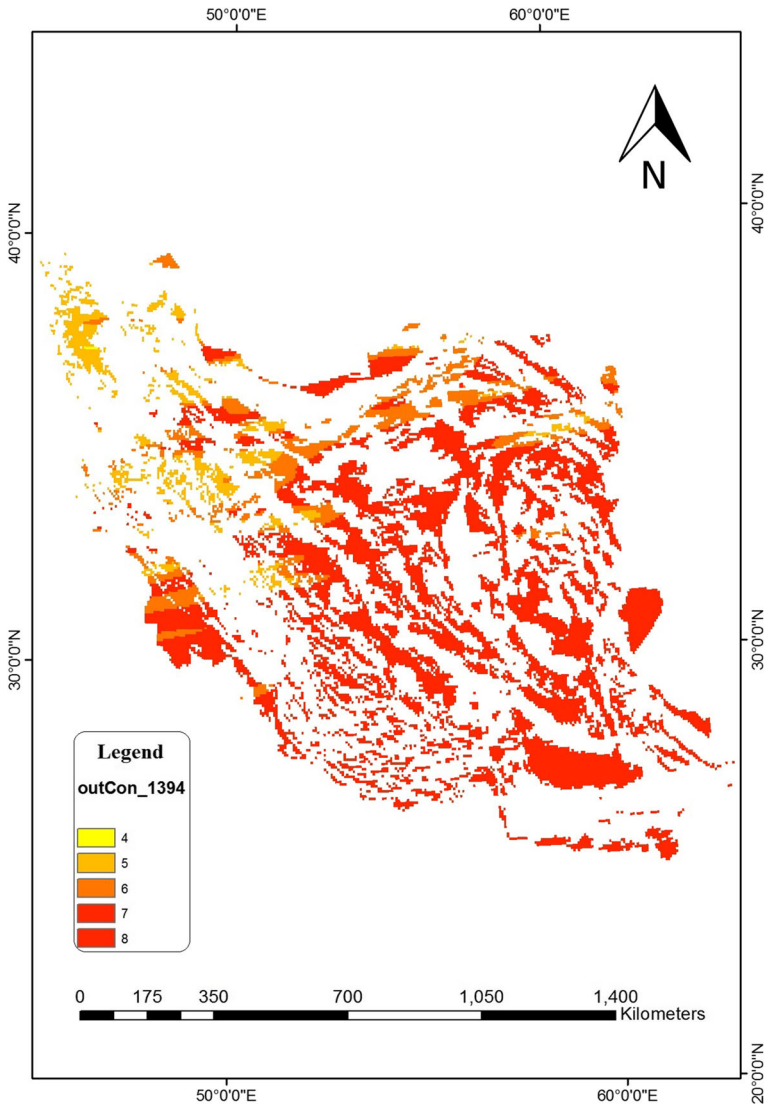
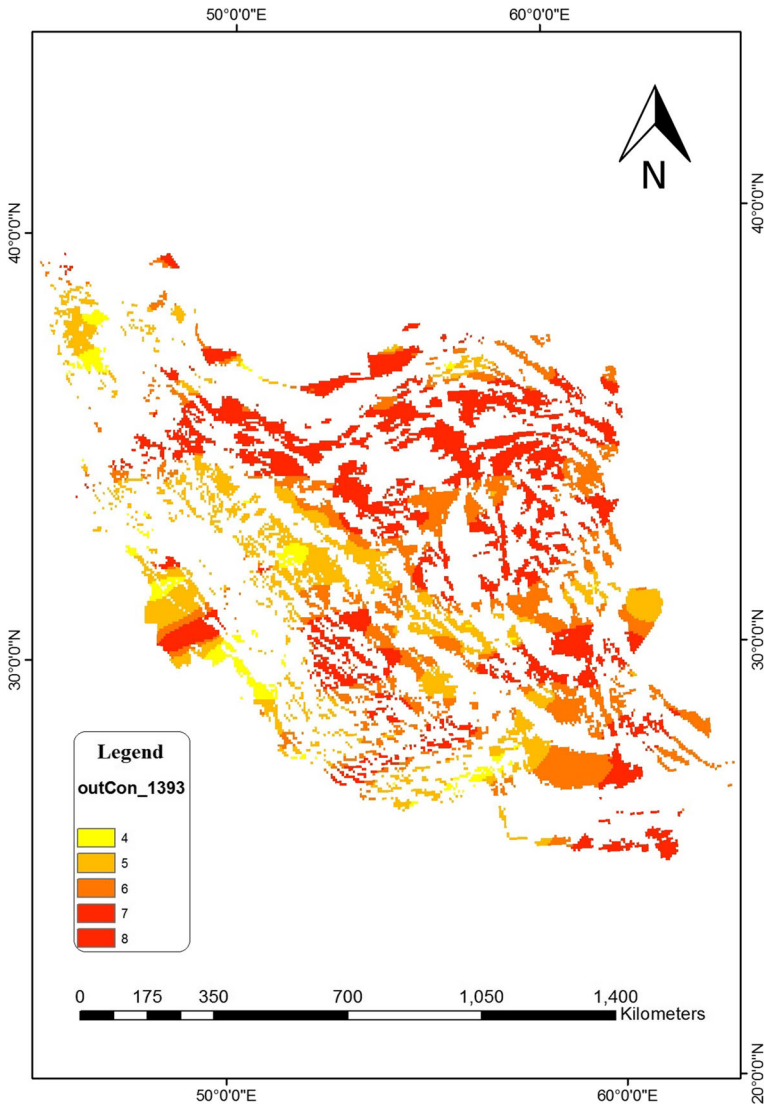


Fig. 32 SPI calculation in theyear2015

for calculation of the SPI Index classification model is provided. In these Figures, a class of color scheme was also used. Accordingly, visual comparisons between the variations of the Raster cells are provided.

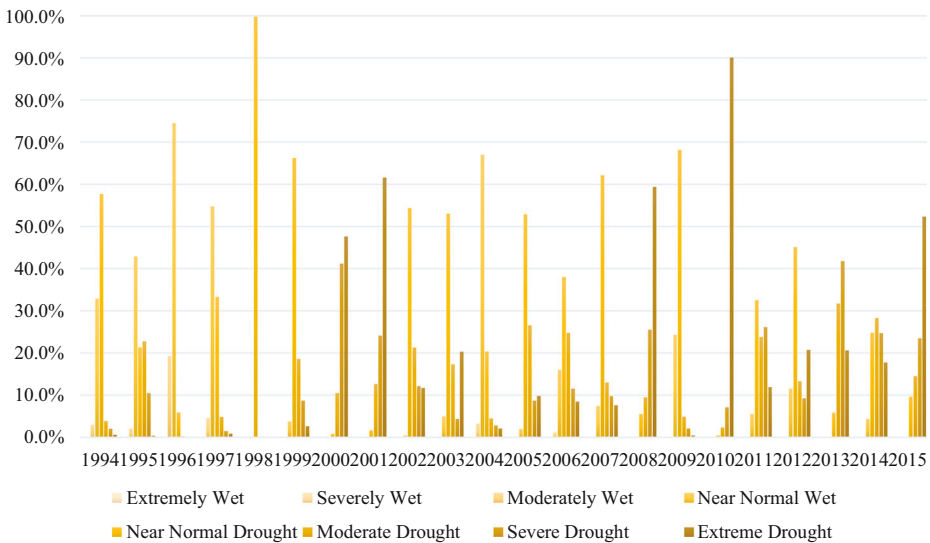
Based on continuous maps, it can be concluded that meteorological droughts have been repeated from the beginning of the study period for a large area in the south-east of the country, and in particular in Sistan and Baluchestan province. Exceptionally, between 1995, 1996, and 1997, a period of near-normal, as well as gentle mildness occurred, which has reversed since 1998. The major occurrence of droughts in the study area shows a more intense tendency towards southeastern regions. However, in areas with higher population density and higher



**Fig. 33** SPI calculation in theyear2014

harvesting of water resources, such as northwest, north, and west of the country, although disturbingly, in 1995, 1999, 2000, 2001, 2008 and 2010 is characterized by severe drought and severe drought, but the trend of droughts has not been the same as in southeastern and eastern parts of the country.

Though, drought investigation in the central, eastern and southeastern parts of the country can be more important due to longer and more severe occurrences and also due to special regional conditions. For example, the magnitude of droughts can lead to the loss of surface water resources, which itself feeds on groundwater reservoirs and reservoirs. Also, the severity of resource dependence in the eastern and southeastern parts of the country is much higher due



**Fig. 34** Comparison graph of the percentage allocation of each class of the SPI index

to the small amount of precipitation and rainfall season. Investigating the changes in the SPI Meteorological Drought Index indicates that the trend towards higher grade classes has increased in most of the country as the severity of the drought in the final years of the study period. In recent maps, with the increase of the numerical value of the class number, the status of the changes indicated a shortage of rainfall observations and as a result of the increased drought. Fig. 34 shows the 22-year data series of the dataset represents the long-term variation of the rainfall drought index (SPI).

In order to provide an overview of the drought situation in the country’s plains, the Raster calculation was performed in the GIS environment. Accordingly, three continuous maps were developed for the ten years period from 1994 to 2004 and the second ten-year period from 2005 to 2015 Also, using the same tool, for the period from 1994 to 2015, the average of the time series maps were calculated. Results clearly indicate the occurrence of drought in the second interval (Table 3).

**Table 3** Statistical summary of SPI drought index

Year - interval	Number of cells in qualitative class							
	Very high wetness	high wetness	mild wetness	Close to normal wetness	mild drought	severe drought	Very severe drought	
2004–1994	0	0	0	0	13,743	1864	235	44
2015–2005	0	0	0	0	1315	6545	6719	1307
2015–1994	0	0	0	0	5848	8664	1356	18
Year - interval	Percentage in qualitative class							
	Very high wetness	high wetness	mild wetness	Close to normal wetness	mild drought	severe drought	Very severe drought	
2004–1994	0.0	0.0	0.0	0.0	86.5	11.7	1.5	0.3
2015–2005	0.0	0.0	0.0	0.0	8.3	41.2	42.3	8.2
2015–1994	0.0	0.0	0.0	0.0	36.8	54.5	8.5	0.1

## 4 Conclusion

The scarcity of groundwater happens when aquifers as an important source of supplying water, are affected by long-term drought. Given the importance of this issue, in this paper, intelligent forecasting models were proposed to predict the negative annual changes in the groundwater quality in the large area of Iran.

In the present study, an artificial intelligence model was developed for extracting statistical relationships and establishing a general and historical understanding of the status of shortages and surpluses of climate water resources. Python programming environment was used to validate this model and to predict the base data digits for 2015. In addition, the results of this study were used to explore the spatial correlation and also to predict the unfavorable classes of drinking water quality and aquifer farming on the basis of climate drought.

From the beginning of the study period, up to the year 1997, the occurrence of a very close to normal range with a completely different distribution in the country has been observed; however, since 1998, the major share of the index in the higher classes indicates the occurrence of drought. This condition has been reversed in the year 2002 to 2007, and in particular, in the year 2004, it has been closely related to normal and close to a normal drought. However, from 2008 to 2015, the stability trend was not observed and the index varied from 24.3% in 2009 to 0% in 2015. Noteworthy here, the trend of rainfall shortage in the last years is very severe and very stressful. In 2013, more than 52% of the index share was in class No. 8.

Based on the results of SPI in the second period, the “mild drought” indicator has a more rapid trend than “near normal” drought. Due to these changes, the southern and eastern parts of the country, as well as certain parts of the northeast of the country during the period from 2005 to 2015, have been occurred in two classes as “severe drought” and “very severe drought”. Also, an examination of the first observation period indicates that the plains of Sistan and Baluchestan Province in the south-east of the country have been facing a much longer period of drought. The average SPI index for the period 1994 to 2015 in Rudbar Jiroft Plain, Hamoon Jazmourian, Dahang Chigi, Chah Hashim, Bazman Sedghal, Iranshahr Bampur, Meshkatan, and Southern Plains of Sistan and Baluchestan provinces faces with severe and very severe droughts. With the exception of the central parts of Khorasan, it is possible to calculate the total area of the eastern, southeast, and southern regions of the country along with the areas near Qazvin plain in the northwest of the country for the long-range in the absolute drought class.

In order to develop the present research, future research topics are suggested as follows: (1) modeling uncertainty analysis using numerical computational methods, (2) the extension of the method of predicting groundwater quality loss by analyzing the groundwater level changes in the long-term periods.

## Compliance with Ethical Standards

**Conflict of Interest** No potential conflict of interest was reported by the authors.

## References

- Basirati M, Jokar MRA, Hassannayebi E (2019) Bi-objective optimization approaches to many-to-many hub location routing with distance balancing and hard time window Neural Computing and Applications:1–22

- Bhuiyan C (2004a) Various drought indices for monitoring drought condition in Aravalli terrain of India. In: 2004. pp 12–23
- Bhuiyan C (2004b) Various drought indices for monitoring drought condition in Aravalli terrain of India. In: XXth ISPRS Congress, Istanbul, Turkey, pp 12–23
- Brekhovskikh V, Volkova Z, Lomova D (2009) Assessment of water quality in the rivers of northern European Russia by using oxygen demand indices. *Russ Meteorol Hydrol* 34:321–330
- Castellvi F, Mormeneo I, Perez P (2004) Generation of daily amounts of precipitation from standard climatic data: a case study for Argentina. *J Hydrol* 289:286–302
- da Silva VDPR (2004) On climate variability in northeast of Brazil. *J Arid Environ* 58:575–596
- Edwards DC (1997) Characteristics of 20th century drought in the United States at multiple time scales. AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH,
- Ferral A, Solis V, Frery A, Orueta A, Bemasconi I, Bresciano J, Scavuzzo CM (2017) Spatio-temporal changes in water quality in an eutrophic lake with artificial aeration. *J Water Land Dev* 35:27–40
- Fu J-C, Huang H-Y, Jang J-H, Huang P-H (2019) River stage forecasting using multiple additive regression trees. *Water Resour Manag* 33:4491–4507
- Guttman NB (1999) Accepting the standardized precipitation index: a calculation algorithm. *JAWRA J Am Water Resour Assoc* 35:311–322
- Keyantash J, Dracup JA (2002) The quantification of drought: an evaluation of drought indices. *Bull Am Meteorol Soc* 83:1167–1180
- Khuram I, Barinova S, Ahmad N, Ullah A, Din SU, Jan S, Hamayun M (2017) Ecological assessment of water quality in the Kabul River, Pakistan, using statistical methods. *Oceanol Hydrobiol Stud* 46:140–153
- Kohzadi N, Boyd MS, Kaastra I, Kermanshahi BS, Scuse D (1995) Neural networks for forecasting: an introduction. *Can J Agric Econ/Revue canadienne d'agroeconomie* 43:463–474
- Lana X, Serra C, Burgueño A (2001) Patterns of monthly rainfall shortage and excess in terms of the standardized precipitation index for Catalonia (NE Spain). *Int J Climatol* 21:1669–1691
- LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444
- Li Y, Fu Y, Li H, Zhang S-W (2009) The improved training algorithm of back propagation neural network with self-adaptive learning rate. In: *Computational Intelligence and Natural Computing*. CINC'09. International Conference on, 2009. IEEE, pp 73–76
- Lloyd-Hughes B, Saunders MA (2002) A drought climatology for Europe. *Int J Climatol* 22:1571–1592
- Mendicino G, Senatore A, Versace P (2008) A groundwater resource index (GRI) for drought monitoring and forecasting in a mediterranean climate. *J Hydrol* 357:282–302
- Mishra AK, Singh VP (2010) A review of drought concepts. *J Hydrol* 391:202–216
- Nielsen MA (2015) Neural networks and deep learning vol 2018. Determination press San Francisco, CA, USA
- Schmidhuber J (2015) Deep learning in neural networks: an overview. *Neural Netw* 61:85–117
- Schoof J (2008) Application of the multivariate spectral weather generator to the contiguous United States. *Agric For Meteorol* 148:517–521
- Shirisha P, Reddy KV, Pratap D (2019) Real-time flow forecasting in a watershed using rainfall forecasting model and updating model. *Water Resour Manag* 33:4799–4820
- Shoji T, Kitaura H (2006) Statistical and geostatistical analysis of rainfall in Central Japan. *Comput Geosci* 32: 1007–1024
- Tabari H, Nikbakht J, Talaei PH (2013) Hydrological drought assessment in northwestern Iran based on streamflow drought index (SDI). *Water Resour Manag* 27:137–151
- Tigkas D, Vangelis H, Tsakiris G (2015) DrimC: a software for drought analysis based on drought indices. *Earth Sci Inf* 8:697–709
- Tsakiris G, Pangalou D, Vangelis H (2007) Regional drought assessment based on the reconnaissance drought index (RDI). *Water Resour Manag* 21:821–833
- Vlček O, Huth R (2009) Is daily precipitation gamma-distributed?: adverse effects of an incorrect use of the Kolmogorov–Smirnov test. *Atmos Res* 93:759–766
- Wu H, Svoboda MD, Hayes MJ, Wilhite DA, Wen F (2007) Appropriate application of the standardized precipitation index in arid locations and dry seasons. *Int J Climatol* 27:65–79
- Zhang G, Patuwo BE, Hu MY (1998) Forecasting with artificial neural networks: the state of the art. *Int J Forecast* 14:35–62