Journal of Research in Ecology

An International Scientific Research Journal

ISSN No: Print: 2319 –1546; Online: 2319– 1554

Case Study

A survey of the occurrence and persistence probability of rainy days using Markov chain model (Case study of Shiraz city, Iran)

Authors:
Nima Tavanpour[1], Ali Asghar Ghaemi[2], Tooraj Honar[2] and Amin Shirvani[2]

Institution:
1. PhD Candidate of Water Engineering, Department of Water Engineering, Faculty of Agriculture, Shiraz University, Shiraz, Iran

2. Associate Professor, Department of Water Engineering, Faculty of Agriculture, Shiraz University, Shiraz, Iran

Corresponding author:
Ali Asghar Ghaemi

ABSTRACT:

One of the basic needs in the planning of water resources is prediction of water amount for agricultural, industrial and urban consumption. Thus, it is required to predict the water capability of each region at different time intervals for efficient planning via reliable and suitable methods. Probable analyses are useful methods to recognize and predict some phenomenon including precipitation. One of the methods commonly used is Markov chain. Markov chain is a special state of models in which the current state of a system depends upon its previous states. The present study studied the frequency and persistence of rainy days in Shiraz city, Iran by the existing statistics of the daily precipitation of 62 years (1956-2016) meteorology stations in Shiraz city using Markov chain model. In this study, due to a few numbers of daily precipitations in June to September days, these months are not considered. The daily precipitation data are ordered based on the frequency matrix of the state change of occurrence of dry and wet days and transition matrix is calculated based on the maximum likelihood method. In the present study, by exact statistical methods, the suitable order of Markov chain is determined and applied. The stationary probability matrices and return period of persistence of rainfall days (2-5 days) were calculated for the mentioned months. The results showed that precipitation occurrence probability in each day was 0.164 and the probability of precipitation non-occurrence was 0.836. Also, it was shown that the highest occurrence probability of precipitation days was in January and February and it was observed that precipitation in Shiraz city had heterogeneous time distribution.

Keywords:
Markov chain, Rainy day, Dry day, Persistence, Return period.

Article Citation:
Nima Tavanpour, Ali Asghar Ghaemi, Tooraj Honar and Amin Shirvani
A survey of the occurrence and persistence probability of rainy days using Markov chain model (Case study of Shiraz city, Iran)
Journal of Research in Ecology (2018) 6(1): 1462-1474

Dates:
Received: 17 Aug 2017 **Accepted:** 21 Oct 2017 **Published:** 20 Mar 2018

Web Address:
http://ecologyresearch.info/documents/EC0480.pdf

INTRODUCTION

The increase or reduction of rainfall increasing flood risk and drought compared to the normal condition has different socio-economic outcomes. Thus, the awareness of probability distribution of rainfall has provided a good ground for the planning of water resources (Yusefi *et al*., 2007). Precipitation as one of the basic effective variables on water resources in Iran has unbalanced time and place distribution. The temporal and spatial variability of precipitation is of great importance in the evaluation of water resources of basin and the relative study of local and regional water resources (Mirmoosavi and Zohrevandi, 2011). One of the basic requirements in the planning of water resources is the prediction of water amount for agricultural, industrial and urban consumption. Thus, it is required to predict the water strength of each region for different time phases for efficient planning via suitable and reliable methods (Razie *et al*., 2003). As Iran is located in the arid and semi-arid climate, the long-term prediction of precipitation is of great importance for planning and management of water resources. The decision makers of water resources need the reliable prediction for managerial decisions (Sedaghatkerdar and Fatahi, 2008).

A stochastic process is a collection of random variables indexed by time. ‘Xt’ variable is performed over time ‘t’ in ‘T’ range. The collection of ‘T’ is indexed collection. Stochastic process is considered as discrete or continuous time (dependent upon ‘T’). The value given to ‘X(t)’ is called process state. The collection of values in which all values of ‘X(t)’ are placed is called space of states. The observed rainfall in a point is a hydrological process as registered continuously. The analysis is performed by transforming the continuous process to a discrete process with ‘Δt’ interval. A real value function as defined on the space of a sample is called random variable. The description of the daily rainfall value as a discrete variable is satisfactory if the registered daily values are considered but it is only an approximate of natural rainfall process (Khanal and Hamrick, 1971). Prediction is possible if there is information about their past. For example, prediction of the precipitation amount is possible if we are aware of the features of past precipitation (Yusefi *et al*., 2007).

The prediction of climatic phenomena is possible by two methods *viz.,* dynamic and statistic. The dynamic models are based on physical rules. The exact recognition of these rules as associated with three phases such as solid, liquid and steam and energy exchange between these three phases and application of these rules are encountered with special problems at real time. Another set of prediction models are statistical models and not considering the physics of the investigated phenomenon explicitly and emphasize only on determining the relationship between inputs and outputs. These models are better than the previous models in terms of easy use. Although it is generally accepted that the results of dynamic models are preferred to the statistical models, this proposition is not true always and its recognition depends upon the recognition of the relevant physical rules, model explanation and its separation power. Thus, using the second group models is unavoidable in some cases (Akan and Houghtalen, 2003).

To calculate the chance of rainfall even occurrence, it is required to use a suitable probability model. For example, time series methods namely Markov chain method are the most suitable and applied methods of statistical prediction in climatic science and the probability methods have been considered in different researches in recent years. Gabrieal and Neumann (1962) is a pioneer using Markov chain model to evaluate the features of daily rainfall occurrence. They showed that the daily rainfall occurrence of Tel Aviv and had the features of two-state first order Markov-chain. It means that there is a correlation between the precipitation occurrences of today with the underlying conditions of yesterday. The present study attempts to use Markov method to estimate the probabilities of precipitation

Figure 1. Location of Shiraz city in Iran

occurrence and finally their prediction for future.

MATERIALS AND METHODS

Area of the study

Shiraz is located at the central of Fars province (in south of Iran), in the mountainous region of Zagros with mild climate. Shiraz is located at 29° north and 52° east and its altitude from the sea level is ranging from 1480 to 1670m at different areas of the city (Figure 1). The maximum absolute temperature is 43.2 and its minimum absolute temperature is -14.4. The mean annual temperature is 18 and the mean annual rainfall of Shiraz city is 337.8mm.

Markov process

Markov (Russian mathematician) presented this hypothesis and the output of the required test is only dependent upon the output of the previous test. In other words:

$$P(A_{k+1}|A_k,A_{k-1},\ldots,A_1) = P(A_{k+1}|A_k) \qquad (1)$$

This hypothesis leads to the formulation of classic concept of a stochastic process as Markov process (for continuous time) or Markov chain (for discrete time) (Sariahmed, 1969). Markov chain is a special state of the model in which current state of a system is dependent upon its previous states. Two factors should be defined in determining system state by this model including: 1- System state at definite time, 2- The probabilities of special state change to other possible states as called transition probabilities. The models of Markov chain have two advantages: First, the predictions exist immediately after the observations, as they use the local weather data as predictors. Second, after processing climatology data, the minimum calculations are required. The order of Markov chain is the number of steps of the past time upon which the current condition of the chain depends. It is attempted to use Markov chains with lower orders as at first, there are a few parameters to be estimated and better estimations are achieved. Second, the next use of fitted model is simpler to calculate other quantities (e.g. the probabilities of long dry periods) (Dash, 2012).

N^{th} order Markov chain model for a discrete stochastic process [Xt, t=0, 1, 2,…] is written mathematically (Khanal and Hamrick, 1971)

$$\begin{aligned}&\Pr\left[X_t = X_t / X_{t-1} = X_{t-1},\ldots,X_1 = X_1\right] \qquad (2)\\ &= \Pr\left[X_t = X_t / X_{t-1} = X_{t-1},\ldots,X_{t-N} = X_{t-N}\right]\end{aligned}$$

For all

$$X_1, X_2, \ldots, X_t \text{ و } t = N+1, N+2, \ldots..$$

First order Markov chain model for N=1 is written as follows:

$$\begin{aligned}&\Pr\left[X_t = X_t / X_{t-1}\ldots\ldots, X_1, = X_1\right]\\ &= \Pr\left[X_t = X_t / X_{t-1} = X_{t-1}\right]\end{aligned} \qquad (3)$$

If Xt-1-=i, Xt=j, then the system has the condition change from condition "i" to "j" in tth stage. The probabilities of change of different conditions as occurred are called transition probability and is written as follows in the first order Markov chain (Khanal and Hamrick, 1971):

$$P_{ij} = P\left[X_t = j / X_{t-1} = i\right] = \frac{\mathrm{P}[X_{t-1} = i, X_t = j]}{\mathrm{P}[X_{t-1} = i]} \qquad (4)$$

In other words, this model states that the prediction of tomorrow's value is performed exclusively by the data of today and the previous day's data are not effective on it.

In this study, the state space (S) at a definite day has one of the two conditions $S=\{W,D\}$ in which 'D' indicates dry day and 'W' indicates wet day. In this research, based on the environmental conditions of Shiraz, 0.1 mm precipitation per day is considered as the threshold of determining wet day compared to that of the dry day.

The matrix of transition probability of two-state Markov chain is defined as follows (Garg and Singh, 2010):

$$P = \begin{array}{c} \\ D \\ W \end{array} \begin{array}{c} \quad D \qquad W \\ \begin{bmatrix} P_{00} & P_{10} \\ P_{01} & P_{11} \end{bmatrix} \end{array} \qquad (5)$$

where, index 0, 1 are indicated to dry and wet days, respectively and probability value shows four states. In other words, the probability of the existence of a dry day after one dry day probability of the existence of a wet day after a dry day , probability of the existence of a wet day after a dry day , probability of the existence of a wet day after a wet day . The following Equations are established (Cazacioc and Cipu, 2004):

$$P_{00} = P\{X_{t+1} = 0|X_t = 0\} \qquad (6)$$
$$P_{01} = P\{X_{t+1} = 1|X_t = 0\} \qquad (7)$$
$$P_{10} = P\{X_{t+1} = 0|X_t = 1\} \qquad (8)$$
$$P_{11} = P\{X_{t+1} = 1|X_t = 1\} \qquad (9)$$

To achieve the matrix of condition change probabilities, at first the matrix of frequency count should be calculated (Selvaraj and Selvis, 2010). In this research, the two-state frequency matrix is made as follows:

$$\begin{array}{c} \\ D \\ W \end{array} \begin{array}{c} \quad D \qquad W \\ \begin{bmatrix} n_{00} & n_{01} \\ n_{10} & n_{11} \end{bmatrix} \end{array} \qquad (10)$$

where, indicates the number of days of a dry day after a dry day, n_{01} is the number of days of a rainy day after a dry day, n_{10} is the number of days of a dry day after rainy day and n_{11} is the number of days in which a rainy day is occurred after a rainy day. Transition probabilities are estimated based on relative frequencies in a long statistical period by maximum likelihood method.

The matrix of transition probability consists of the following features

- For all elements of this matrix, we have
  $\forall i,j \quad 0 \le P_{ij} \le 1$
- The sum of transition probabilities from one state to all possible states during the next time should be equal to 1, here we have:
  $$P_{00} + P_{01} = 1$$
  $$P_{10} + P_{11} = 1 \qquad (11)$$
- If p is the matrix of transition probability of a Markov chain, we have:
  $$P^{(n)} = P^{(0)}P^{n} \qquad (12)$$

where, $P^{(n)}$ indicates the n^{th} transition probability, it (4)
means that after n stage of process initial period, how is the probability to be placed in each of zero or one conditions. $P^{(0)}$ indicates the probabilities of initial stage and $P^{(n)}$ is the n^{th} multiplication of matrix of transition probability p by itself (Cox and Miller, 1977).

As shown, by the maximum likelihood method, we can estimate the values of transition probabilities and the results of the above matrix are as follows (Bakhtiari *et al*., 2014)

$$\hat{P}_{00} = \frac{n_{dd}}{n_{00} + n_{01}} \qquad (13)$$

$$\hat{P}_{01} = \frac{n_{01}}{n_{00} + n_{01}} \qquad (14)$$

$$\hat{P}_{10} = \frac{n_{10}}{n_{10} + n_{01}} \qquad (15)$$

$$\hat{P}_{11} = \frac{n_{11}}{n_{10} + n_{11}} \qquad (16)$$

Now, we evaluate this issue whether Markov chain model is suitable for series of data of precipitation or not. To do this, Chi-square test is used. Null hypothesis of this test is based on this idea that the data series are independent (Hoaglin *et al*., 2011).

$$\begin{cases} H_0: \text{The data series are independent} \\ H_1: \text{The data series follow the first order markov chain} \end{cases}$$ (or)

$$\begin{cases} H_0: P_{ij} = P_{i\cdot} & \forall\, i,j \\ H_1: P_{ij} \neq P_{i\cdot} & \exists\, i,j \end{cases}$$

The statistics of this test under the independence assumption of series of data is:

$$X_0^2 = \sum_{i=0}^{1}\sum_{j=0}^{1} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \qquad (17)$$

where, n_{ij} is the frequency of observations in the element ij^{th} of frequency matrix and e_{ij} is the expected transition frequencies in transition from state 'i' to state 'j' under the independence assumption for the element ij^{th} of frequency matrix and is written as $e_{ij} = \frac{n_{i\cdot}\, n_{\cdot j}}{n_{\cdot\cdot}}$ (i, j=0.1) where $n_{i\cdot} = n_{i0} + n_{i1}$ indicates the number of days in which the day before the data series is in situation i^{th} (i=0,1); $n_{\cdot j} = n_{0j} + n_{1j}$ is the number of day in which the series of data is regarding the current day in j^{th} situation (j=0,1) and n is the number of total days in this study. The calculated X^2_0 of Equation 17 follows the chi-square distribution with degree of freedom 1 and is compared with the number of chi-square distribution table at the significance level 5% (3.84) and if X^2_0 is bigger than this value, the above H_0 is rejected (Zarei, 2004).

As Markov chain order plays an important role in the calculation of return period of different persistence of precipitation, it is required to evaluate the order of Markov chain for different months in the existing data before the calculation of return period. Thus, at first, we perform the test of comparison of first and second order as follows (Yakowitz, 1976):

$$\begin{cases} H_0: \text{The data series follow first order Markov chain} \\ H_1: \text{The data series follow second order markov chain} \end{cases}$$

$$\begin{cases} H_0: P_{ijk} = P_{\cdot jk} & \forall i,j,k \\ H_1: P_{ijk} \neq P_{\cdot jk} & \exists i,j,k \end{cases}$$ (or)

By using likelihood ratio method, the statistic of the above test is:

$$X^2 = 2\sum_{i=1}^{1}\sum_{j=0}^{1}\sum_{k=0}^{1} n_{ijk}\left(\ln\left(\frac{n_{ijk}}{n_{ij\cdot}}\right) - \ln\left(\frac{n_{\cdot jk}}{n_{\cdot j\cdot}}\right)\right) \qquad (18)$$

This statistics is hypothetically framed by chi-square distribution with degree of freedom 2 and based on percentile point of chi-square distribution at the significance level 5% is equal to 5.99, we can reject or accept H_0.

$$P_{ijk} = P\{X_{t+1} = k | X_t = j, X_{t-1} = i\}\ \forall\, i,j,k = 0,1 \qquad (19)$$

Transition probabilities in the second order Markov chain are as follows (Harrison and Waylen, 2000):

Similarly, frequencies of n_{ijk} , n_{ijkm} are counted corresponding with the transition probabilities for a long period. Again, by using the estimation of the maximum likelihood, each of probabilities is estimated by the following formulas (Harrison and Waylen, 2000):

$$\hat{P}_{ijk} = \frac{n_{ijk}}{n_{ij.}} = \frac{n_{ijk}}{n_{ij0} + n_{ij1}} \qquad (20)$$

$$\hat{P}_{ijkm} = \frac{n_{ijkm}}{n_{ijk\cdot}} = \frac{n_{ijkm}}{n_{ijk0} + n_{ijk1}} \qquad (21)$$

As Markov chain is a rank data, to test whether the data show Markov condition or a trend is considered, it is better to use rank method. One of the common methods is using Spearman rank test, in this method, at first the difference between the rank of each value and its order in the series d_i; i=1,…, n is calculated and then Spearman statistic (r_s) is calculated by Equation 22 (WMO, 2000).

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \qquad (22)$$

The investigation of the condition of stationary probability matrix

One of the important issues in Markov chains is to achieve a stationary chain. Stationary of chain means that rainfall occurrence during the study period has no considerable trend. It means that precipitation occurrence probability is the same all around the period. Theoretically, stationary means as after a long period, the system is under the statistical equilibrium condition and it means that transition probabilities are independent from the initial conditions of chain. Thus, if the statistical equilibrium is achieved and for two conditions of dry and wet, the stationary probabilities are denoted by π_0, π_1, respectively. We have a homogenous system of equations in the matrix form of limiting Equation 12 (Cox and Miller, 1977):

$$\pi = \pi P \qquad (23)$$

where, $\pi = (\pi_0, \pi_1)$ is the vector of stationary probabilities and 'p' is the transition probability matrix and it is shown with the general form $P = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}$ as$0 \leq \alpha, \beta \leq 1$. Now, by placing matrix 'P' in equation 23, we achieve $\pi_0\alpha - \pi_1\beta = 0$ and as $\pi_0+\pi_1=1$, thus, stationary probabilities are as follows:

$$\pi_0 = \frac{\beta}{\alpha+\beta} \qquad (24)$$

$$\pi_1 = \frac{\alpha}{\alpha+\beta} \qquad (25)$$

Also, we can achieve stationary probabilities via consecutive multiplication of the initial transition probability matrix by itself as all rows of matrix are equal with each other- and if we multiply it again by the initial matrix, it is not changed (Cox and Miller, 1977).

Precipitation persistence and its estimation

Meteorology observations are not independent from the previous conditions. These conditions are intended to be continued to the next periods at a time. This tendency is recognized as persistence. One of the applications of Markov chain method is the estimation of persistence of m day's periods. Persistence of m days of raining is the number of m consecutive days of rainfall as occurred but before and after m days, no precipitation is occurred (Grace and Eagleson, 1966).

Persistence values

If n_1, N are the number of wet days and total days in the studied period, respectively, then the empirical estimation P, means the rainfall occurrence probability in this period as $P=n_1/N$. If it is assumed $P_1=P_{11}$, it is the occurrence probability of a rainy day after a rainy day and d_1 indicates the number of runs form a dry day to a rainy day. Indeed, d_1 is equal to the number of at least one day runs from one dry day to a rainy day and $d_1P_1$indicate the number of at least two-day runs and thus, we have $d_1-d_1P_1=d_1(1-P_1)$ is the number of exactly one day runs (e.g. the number of total times in which only one-day rain is occurred). Similar to this, we can say $d_1(1-P_1)P_1^{m-1}$, achieves the number of exactly 'm' days runs. Easily, we can predict the day through the following equation (Berger and Goossens, 1983):

$$P_1 = 1 - \frac{d_1}{n_1} \qquad (26)$$

It is worth to mention that $T_1 = \frac{n_1}{d_1} = \frac{1}{1-P_1}$ is the average length of exactly one-day runs (e.g. estimation of one-day persistence) and $T_m = \frac{1}{(1-P_1)P_1^{m-1}}$ is the length of average runs of exactly m day's (e.g. estimation of m day's persistence).

The above method is recognized as the simplest method to estimate persistence under the conditions in which the observations have the first order Markov chain but under the conditions in which we have second order or above Markov chain observations, we should use other equations. At first, we achieve probability function of the duration of precipitation period and return period is calculated based on it. The function of probability of precipitation duration (l) in the second order Markov chain as denoted by $f_L(l)$, we achieve the followings (Akyuz *et al*., 2012).

$$f_L(\ell) = \begin{cases} P(W|WD) & \ell = 1 \\ P(W|DW)P^{\ell-2}(W|WW)P(D|WW) & \ell \geq 2 \end{cases} \qquad (27)$$

Thus, the return period of the second order Markov chain is as:

$$T(\ell) = \frac{2 + \frac{P(D|WD)}{P(W|DD)} + \frac{P(W|DW)}{P(D|WW)}}{P(W|DW)P^{\ell-2}(W|WW)P(D|WW)} \qquad (28)$$

Thus, researchers have proposed other indices to estimate the persistence value.

If the 'persistence ratio' is defined as the ratio of average duration of exactly one-day runs to the average duration of the number of rainy days and it is denoted by R, then (Grace and Eagleson, 1966):

$$R = \frac{\frac{1}{1-P_1}}{\frac{1}{1-P}} = \frac{1-P}{1-P_1} \qquad (29)$$

Besson (1924) applied persistence ratio to introduce its index as follows:

$$R_B = R - 1 = \frac{1-P}{1-P_1} - 1 \qquad (30)$$

In this formula, if P_1=P, it means that there is no persistence. Thus, R_B is zero and if the precipitation occurrence is definite after a rainy day P_1=1, then R_B=+∞. It is worth to mention that if P1 is less than P, is negative and it showed the tendency to intermittence in the series of observations between the occurrence and non-occurrence of precipitation. The more power the value of R_B, the tendency to persistence in observations is higher and if it is negative, the tendency of observation series to consecutive change of state between the occurrence and non-occurrence of precipitation is high. Other introduced statistics to evaluate persistence are as follows (Brooks and Carruthers, 1953):

$$r_B = 1 - \frac{1}{(R_B+1)^2} = 1 - \frac{1}{R^2} \qquad (31)$$

Normally, r_B is ranging zero to one acting similar to serial correlation coefficient (auto-correlation function). On the other h and, the persistence ratio based on r_B is as follows (Grace and Eagleson, 1966):

$$R = 1 + R_B = \frac{1}{\sqrt{1\text{-}r_B}} = 1\text{-}\left(\frac{1\text{-}P_1}{1\text{-}P}\right)^2 \qquad (32)$$

All calculations and analysis were done in this research by using R statistical software (version 3.4.1).

RESULTS

Table 1 shows the statistical characteristics of daily rainfall of Shiraz city during 62 years (from the beginning of January 1956 to the end of December 2016). These data were obtained from Shiraz airport weather station. Due to a few number of precipitation in June, July, August and September, these months are ignored in this study. In Table 1, the number of rainy days, the mean daily precipitation of each month, maximum daily precipitation occurred in each month and standard deviation and the daily precipitation changes coefficient of each month are presented.

The mean of the number of annual rainfall days in Shiraz city is 305 days and total rainy days during the study is 2441 days equal to 16% of total days. Based on the information of this Table, the highest number of wet days is dedicated to January (496 days) and the lowest

Table 1. The statistical characteristics of daily precipitation of Shiraz (1956-2016)

S. No	Month	Number of wet days	The mean daily precipitation (mm)	Maximum daily precipitation (mm)	The coefficient of change of daily precipitation
1	January	496	2.8±8.3	85	3
2	February	430	1.7±5.2	48	3
3	March	427	1.6±5	54	3.1
4	April	315	5.7±4.2	71.3	0.7
5	May	101	3.4±1.4	24.2	0.4
6	October	60	4.8±1.6	32	0.3
7	November	224	7.8±5	99	0.6
8	December	388	10.3±7.7	99	0.8

Table 2. The two state transition probabilities and frequencies of Shiraz (1956-2016)

previous day \ Current day	Dry	Wet
Dry	11135 (0.898)	1266 (0.102)
Wet	1267 (0.520)	1170 (0.480)

number of wet days is dedicated to October (60 days)

The probable condition of wet and dry days

By assuming two states (wet and dry) of frequency table, the change of precipitation condition between two consecutive days in Shiraz city is shown in Table 2. In Table 2, the values inside the parenthesis are the two-state transition probabilities.

As shown in Table 2, the number of condition change from dry day to the next dry day is 11135 days, the number of change of condition from wet day to dry day is 1267 days, the number of condition change from dry day to wet day is 1266 days and the number of condition change from wet day to wet day is equal to 1170 days.

Now, we should investigate whether the probability values in Table 2 follow the two-state Markov chain or not and it is required to test the trend existence. At first, we go to one of the valid tests to evaluate Markov nature of transition probability values, chi-square test. The hypotheses of this test are as follows:

$$\begin{cases} H_0: & \text{The matrix of transition probability shows the independence of days.} \\ H_1: & \text{The matrix of transition probability indicates a Markov chain} \end{cases}$$

The statistics of this test is as in $X^2 = \sum\sum\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$ which O_{ij} is the frequency value of each cell in the frequency matrix in Table 2 and E_{ij} is the expected value under independence assumption in the cell as shown in Table 3.

The test statistics based on the observed and expected values is 2118.2 with the degree of freedom 1 and probability value (p-value) is less than 0.00001and it indicates the support of a very strong Markov relationship in the precipitation data of Shiraz city. The correlation coefficient between precipitation values in the consecutive days based on Spearman method is performed to evaluate the strong relationship in terms of correlation between the consecutive days and the correlation coefficient is 0.3. This shows the non-stationary of data and the correlation of sequence of daily precipitation is significant (The probability value is almost zero).

Table 3. The expected values of two-state frequency matrix of Shiraz (1956-2016)

previous day \ Current day	Dry	Wet
Dry	10365	2036
Wet	2037	400

Thus, it is supported that the matrix of transition probability between wet and dry days followed one by one of Markov chain. It means that $P = \begin{array}{c c} & \begin{array}{cc} D & W \end{array} \\ \begin{array}{c} D \\ W \end{array} & \begin{bmatrix} 0.898 & 0.102 \\ 0.520 & 0.480 \end{bmatrix} \end{array}$ is the matrix of transition probability of two-state Markov chain as the transition probability from one dry day to another one and is equal to P_{00}=P(D/D)=0.898, the transition probability form one wet day to dry day is equal to P_{10}=P(D/W)=0.520 and we have P_{01}=P(W/D)=0.102 and P_{11}=P(W/W)=0.480.

In a general view to the study period, we can say stationary probabilities of the occurrence or non-occurrence of precipitation are 0.164, 0.836, respectively and this shows that in the long-term, we should expect that 0.16 of the days of year, precipitation is occurred in Shiraz city.

Table 4. The values of transition frequencies and second order transition probabilities of Shiraz (1956-2016)

Previous days		Current day	
Two days before	**One day before**	**Dry**	**Wet**
Dry	Dry	10179 (0.904)	1078 (0.096)
Dry	Wet	597 (0.471)	669 (0.528)
Wet	Dry	1079 (0.852)	188 (0.148)
Wet	Wet	669 (0.572)	501 (0.428)

Table 5. The test statistics of likelihood ratio

S. No	Month	Spearman rank test		Likelihood ratio test	
		Statistic amount	P-value	First order	Second order
	January	0.18	<0.001	202.7	12.8
1	February	0.24	<0.001	145.6	11.9
2	March	0.24	<0.001	208.6	3.8
3	April	0.25	<0.001	167.7	1.5
4	May	0.42	<0.001	92.8	2.5
5	October	0.47	<0.001	47.5	8.4
6	November	0.38	<0.001	179	6.4
7	December	0.24	<0.001	195.6	4.7

Table 6. The frequency matrices, transition probabilities, stationary probabilities and persistence ratio indices for Shiraz (1956-2016)

S. No	Month	Frequency matrix	Probability matrix	Stationary probability matrix	R_B	r_B
1	January	$\begin{bmatrix}1153 & 241\\ 242 & 254\end{bmatrix}$	$\begin{bmatrix}0.827 & 0.173\\ 0.488 & 0.512\end{bmatrix}$	$\begin{bmatrix}0.738 & 0.262\\ 0.738 & 0.262\end{bmatrix}$	0.5	0.6
2	February	$\begin{bmatrix}1076 & 225\\ 225 & 205\end{bmatrix}$	$\begin{bmatrix}0.826 & 0.174\\ 0.523 & 0.477\end{bmatrix}$	$\begin{bmatrix}0.750 & 0.250\\ 0.750 & 0.250\end{bmatrix}$	0.4	0.5
3	March	$\begin{bmatrix}1250 & 214\\ 214 & 212\end{bmatrix}$	$\begin{bmatrix}0.854 & 0.146\\ 0.502 & 0.498\end{bmatrix}$	$\begin{bmatrix}0.775 & 0.225\\ 0.775 & 0.225\end{bmatrix}$	0.5	0.6
4	April	$\begin{bmatrix}1312 & 173\\ 172 & 142\end{bmatrix}$	$\begin{bmatrix}0.884 & 0.117\\ 0.548 & 0.452\end{bmatrix}$	$\begin{bmatrix}0.825 & 0.175\\ 0.825 & 0.175\end{bmatrix}$	0.5	0.6
5	May	$\begin{bmatrix}1723 & 66\\ 66 & 35\end{bmatrix}$	$\begin{bmatrix}0.963 & 0.037\\ 0.654 & 0.347\end{bmatrix}$	$\begin{bmatrix}0.947 & 0.053\\ 0.947 & 0.053\end{bmatrix}$	0.5	0.5
6	October	$\begin{bmatrix}1786 & 44\\ 44 & 16\end{bmatrix}$	$\begin{bmatrix}0.976 & 0.024\\ 0.733 & 0.267\end{bmatrix}$	$\begin{bmatrix}0.968 & 0.032\\ 0.968 & 0.032\end{bmatrix}$	0.3	0.4
7	November	$\begin{bmatrix}1481 & 124\\ 124 & 100\end{bmatrix}$	$\begin{bmatrix}0.923 & 0.077\\ 0.553 & 0.446\end{bmatrix}$	$\begin{bmatrix}0.878 & 0.123\\ 0.878 & 0.123\end{bmatrix}$	0.6	0.6
8	December	$\begin{bmatrix}1300 & 202\\ 202 & 186\end{bmatrix}$	$\begin{bmatrix}0.866 & 0.135\\ 0.521 & 0.479\end{bmatrix}$	$\begin{bmatrix}0.795 & 0.205\\ 0.795 & 0.205\end{bmatrix}$	0.5	0.6

Table 7. The frequency matrices, transition probabilities, stationary probabilities and persistence ratio indices for Shiraz (2017)

S. No	Month	Frequency matrix	Probability matrix	Stationary probability matrix	R_B	r_B
1	January	$\begin{bmatrix}22 & 2\\ 2 & 4\end{bmatrix}$	$\begin{bmatrix}0.917 & 0.083\\ 0.333 & 0.667\end{bmatrix}$	$\begin{bmatrix}0.800 & 0.200\\ 0.800 & 0.200\end{bmatrix}$	1.4	0.8
2	February	$\begin{bmatrix}17 & 3\\ 3 & 5\end{bmatrix}$	$\begin{bmatrix}0.895 & 0.105\\ 0.375 & 0.625\end{bmatrix}$	$\begin{bmatrix}0.781 & 0.219\\ 0.781 & 0.219\end{bmatrix}$	0.9	0.7
3	March	$\begin{bmatrix}14 & 6\\ 6 & 4\end{bmatrix}$	$\begin{bmatrix}0.700 & 0.300\\ 0.600 & 0.400\end{bmatrix}$	$\begin{bmatrix}0.667 & 0.333\\ 0.667 & 0.333\end{bmatrix}$	0.1	0.2
4	April	$\begin{bmatrix}26 & 2\\ 2 & 0\end{bmatrix}$	$\begin{bmatrix}0.963 & 0.037\\ 1.000 & 0.000\end{bmatrix}$	$\begin{bmatrix}0.964 & 0.036\\ 0.964 & 0.036\end{bmatrix}$	-0.07	-0.2
5	May	$\begin{bmatrix}23 & 3\\ 3 & 2\end{bmatrix}$	$\begin{bmatrix}0.920 & 0.080\\ 0.600 & 0.400\end{bmatrix}$	$\begin{bmatrix}0.882 & 0.118\\ 0.882 & 0.118\end{bmatrix}$	0.4	0.5

Table 8. P-value amounts of two proportions tests

S. No	Month	Fisher test	Normal approximation test
1	January	0.4	0.3
2	February	1	0.8
3	March	0.1	0.08
4	April	0.8	0.5
5	May	0.08	0.06

Also, the frequencies of second order of chain transition between dry and wet days are shown in Table 4, and the transition probabilities of second order Markov chain based on Formulas 18, 19 were calculated and the results are shown inside the Parenthesis in the cells of Table 4.

For example, the probability that the current day is rainy, on the condition that the previous day is wet and two days before is dry equal to $P_{011}=P(X_{t+1}=1/X_t=1, X_{t-1}=0) = P(W/WD) = 0.528$.

This statistics is hypothetically with chi-square distribution with degree of freedom 2 and based on percentile point of chi-square distribution at the significance level 5% is equal to 9.21 and we can reject or support H_0. The results of test for each month are shown separately in Table 5. According to Spearman rank test, all months follow Markov chain and according to likelihood ratio test months January, February, October and November follow second order Markov chain and other studied months follow the first order Markov chain.

The values of stationary probabilities and the calculation of return periods in different months

As the time distribution and precipitation value in different months of year are different in Shiraz city and for exact investigation of frequency probabilities and persistence of wet days of Shiraz city, the frequencies and transition probabilities as monthly are calculated in Table 6. As shown in this Table, the lowest occurrence probability of consecutive dry days in January and February is about 0.83. The highest probability of occurrence of such condition is dedicated to October about 0.98 and the highest occurrence probability of consecutive wet days is dedicated to January about 0.51 and lowest value is dedicated to October about 0.27. Also, in Table 6, the values of stationary probabilities are based on formulas 24, 25 by which the highest precipitation occurrence probability is dedicated to January (0.262).

In order to validate the results in Table 6, long-term predictions were compared with real precipitation data of January, February, March, April and May 2017 (Table 7) by using two strong proportions tests (normal approximation and Fisher tests) in Minitab software and the results are presented in Table 8.The results showed that the predictions were strongly confirmed. Based on the results of Table 6, we can achieve the return periods as 2, 3, 4, 5 days and its result based on the studied months is shown in Table 9.

DISCUSSION

The present study was aimed to evaluate the frequency and persistence of wet days in Shiraz city. To achieve this purpose, at first based on chi-square and Spearman tests, it was shown that two-state Markov chain was a suitable method to study the precipitation frequency in Shiraz city. Then, by likelihood ration test, this result was achieved that the data follow the second

Table 9. The estimation of return periods of persistence of precipitation 2-5 days in different months of Shiraz (1956-2016)

	Month	January	February	March	April	May	October	November	December
Return period of precipitation (day)	2 days	24	23	6	7	20	210	46	7
	3 days	54	58	26	40	370	840	125	30
	4 days	123	149	113	225	6923	3359	337	146
	5 days	282	380	501	1282	129543	13433	910	709

order Markov chain. To evaluate the precipitation persistence in Shiraz city, the monthly precipitation occurrence probability and then the return period of persistence 2-5 days were studied for the studied months.

In all studies done in Iran, in order to forecast precipitation by using the Markov chain, only the first order of the Markov chain was used which may not be in good agreement with data and resulted to incorrect results. But in this study, by using accurate statistical methods, the appropriate order of the Markov chain was diagnosed to be used.

In a research done by Mohammadi *et al.* (2015) about occurrence and persistence probability of rainy days of Shiraz city, they didn't check the accurate order of Markov chain and only used first-order Markov chain model. For example, they estimated the retain periods of 2-day precipitation persistence of Shiraz five days for January and 35 days for October. Also, they didn't validate their results by comparing with real data.

Dunn (2004) in concurrent modeling of event sand precipitations by using Markov chain models, Poisson distribution and Gamma distribution showed that these models presented suitable precipitation models. Dastidar *et al*. (2010) applied Markov chain model to simulate seasonal rainfall of four meteorology stations in the Bengal of India. These researchers applied Bayesian theory to determine the order of Markov chain model. The results showed that the third order of Markov chain model had the best description of precipitation pattern for all stations except one station. Ng and Panu (2010) by comparison of the traditional-random models for daily rainfall occurrence among some investigated models (geometry distribution, Markov chain, probability matrix), found that Markov chain model has performed well and the daily precipitation event was described well. Bigdeli and Eslami (2010) applied the daily rainfall data for an11-years period to analyze the wet and dry days of Spring and Summer using Markov chain model. The results showed that in summer, the number of dry days was more than that of wet days. Also, the occurrence probability of two consecutive dry days was more than the number of wet days in summer. Alijani *et al*. (2005) analyzed and predicted the precipitation of Larestan and applied first-order Markov chain model as a suitable model to analyze precipitation of this region. Jalali *et al*. (2011) studied the occurrence probability of rainy days in Urmia city using Markov chain model. The results of studies showed that the highest probability of occurrence of rainy days was in (Spring) April.

CONCLUSION

Precipitation occurrence probability in each day was 0.164 and the probability of precipitation non-occurrence was 0.836. According to Spearman rank test, all months follow Markov chain and according to likelihood ratio, test months *viz,* January, February, October and November follow second order Markov chain and other studied months follow the first order Markov chain. The results showed that the highest precipitation occurrence probability in Shiraz was dedicated to January and February and the lowest precipitation occurrence probability in the studied months was dedicated to October. In order to validate the results, long-term predictions were compared with real precipitation data of January, February, March, April and May 2017 by using two proportions tests (normal approximation and Fisher tests) and the results showed that the predictions were strongly confirmed.

REFERENCES

Akan AO and Houghtalen RJ. 2003. Urban hydrology, hydraulics, and storm water quality, John Wiley and sons. Inc, USA. 392 p.

Akyuz DE, Bayazit M and Onoz B. 2012. Markov chain models for hydrological drought characteristics. *Journal of Hydrometeorology,* 13: 298-309

Alijani B, Jafarpour Z and Ghaderi H. 2005. The analysis and prediction of precipitation of Larestan region using Markov chain model. *Journal of Sarzamin* 7: 1-10.

Bakhtiari B, Shahraki N and Ahmadi MM. 2014. Estimation probability of daily precipitation by using Markov chain models in different climates of Iran, *Journal of Iran Water Resources Research*, 10(2): 44-55.

Berger A and Goossens C. 1983. Persistence of wet and dry spells at Uccle (Belgium). *International Journal of Climatology*, 3(1): 21-34.

Besson L. 1924. Sur la probabilite de la pluie, *Comptes Rendus,* 178: 1743-1745.

Bigdeli A and Eslami A. 2010. Analysis of wet and dry periods by Markov chain model in southern of Caspian Sea, International Conference on Environmental Engineering and Applications, 10-12 Sept, Singapore, Singapore, 3: 96-99

Brooks CEP and Carruthers N. 1953. Handbook of statistical methods in meteorology, Her Majesty's Stationary Office, London. 412 p.

Cazacioc L and Cipu EC. 2004. Evaluation of the transition probabilities for daily precipitation time series using a Markov chain model. Proceedings of the 3-rd International Colloquium Mathematics in Engineering and Numerical Physics, October: 7-9.

Cox DR and Miller HD. 1977. The theory of stochastic processes. Chapman and Hall/CRC. 408 p.

Dash PR. 2012. A Markov chain modeling of daily precipitation occurrences of Odisha. *International Journal of Advanced Computer and Mathematical Science,* 3(4): 482-486

Dastidar AG, Gosh D and Dasgupta S. 2010. Higher order Markov chain models for monsoon rainfall over west Bengal, India. *Indian Journal of Radio and Space Physics,* 39(1): 39-44

Dunn PK. 2004. Occurrence and quantity of precipitation can be modeled simultaneously. *International Journal of Climatology,* 24(10):1231-1239.

Gabriel KR and Neumann J. 1962. A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quarterly Journal of the Royal Meteorological Society,* 88(375): 90-95.

Garg VK and Singh JB. 2010. Three – State Markov chain approach on the behavior of rainfall. *Journal of New York Science,* 3(12): 76-81.

Grace RA and Eagleson PS. 1966. The synthesis of short-time-increment rainfall sequences, Report - Hydrodynamics Laboratory, Massachusetts Institute of Technology, Cambridge, USA, 191 p.

Harrison M and Waylen P. 2000. A note concerning the proper choice for Markov model order for daily precipitation in the humid tropics: a case study in Costa Rica. *International Journal of Climatology,* 20(14): 1861-1872.

Hoaglin DC, Mosteller F and Tukey JW. 2011. Exploring data tables, trends, and shapes. John Wiley and Sons, 527 p.

Jalali M, Kargar H and Soltani S. 2011. Investigation of occurrence probability of rainy days in Urmia city by using Markov chain model. *Journal of Geographical Space,* 11(35): 235-257.

Khanal NN and Hamrick RL. 1971. A stochastic model for daily rainfall data synthesis. Central and Southern Florida Flood Control District, 1-28 p.

Mirmosavi H and Zohrevandi H. 2011. The modeling of weekly precipitation probability to analyze the consecutive dry days. Case study of Meteorology station of Nahavand, Hamedan province. The 2nd national conference of applied researches of water resources, Iran. 28, 29 Ordibehesht. Regional Water Company. Zanjan.

Mohammadi H, Mahootchi MH, Khazaei M and Abbasi E. 2015. Survey of the occurrence and persistence probability of rainy days of Shiraz city by using Markov chain model. *Journal of Geographical Data,* 24 (93): 77-90.

Ng WW and Panu US. 2010. Comparisons of traditional and novel stochastic models for the generation of daily precipitation occurrences. *Journal of Hydrology,* 380: 222-236.

Razie T, Shokuhi AR and Saghafian B. 2003. The prediction of intensity, persistence and frequency of droughts using probabilistic methods and time series; case study of Sistan and Baluchestan province. Desert Magazine, 8: 292-310.

Sariahmed A. 1969. Synthesis of sequences of summer thunderstorm volumes for the Atterbury Watershed in the Tucson area. M.Sc. thesis, university of Arizona.

Sedaghatkerdar A and Fatahi E. 2008. The prediction indices of drought in Iran. *Geography and Development,* 11: 59-76.

Selvaraj RS and Selvis T. 2010. Stochastic modelling of daily rainfall at ADUTHURAI. *International Journal of Advanced Computer and Mathematical Sciences,* 1: 52-57.

WMO. 2000. Detecting trend and other change in Hydrological Data, WMO/ TD- NO.1013. 158 p.

Yakowitz SJ. 1976. Small sample hypothesis tests of Markov order, with application to simulated and hydrologic chains. *Journal of the American Statistical Association,* 71: 132-136.

Yusefi N, Hejam S and Irannejad P. 2007. The estimation of probabilities of drought and wet year by Markov chain and normal distribution (case study of Qazvin). *Geographical Researches,* 60: 121-128.

Zarei A. 2004. The engineering statistics. 1st ed., Daneshparvar publications. Tehran