Simulation of Daily Rainfall Through Markov Chain Modeling

N. Sadiq[1]

[1]Institute of Space & Planetary Astrophysics, University of Karachi, Karachi, 75270, Pakistan
[1]nsadiq@uok.edu.pk

Abstract-Being an agricultural country, the inhabitants of dry land in cultivated areas mainly rely on the daily rainfall for watering their fields. A stochastic model based on first order Markov Chain was developed to simulate daily rainfall data for Multan, D. I. Khan, Nawabshah, Chilas and Barkhan for the period 1981-2010. Transitional probability matrices of first order Markov Chain was utilized to generate the daily rainfall occurrence while gamma distribution was used to generate the daily rainfall amount. In order to achieve the parametric values of mentioned cities, method of moments is used to estimate the shape and scale parameters which lead to synthetic sequence generation as per gamma distribution. In this study, unconditional and conditional probabilities of wet and dry days in sum with means and standard deviations are considered as the essential parameters for the simulated stochastic generation of daily rainfalls. It has been found that the computerized synthetic rainfall series concurred pretty well with the actual observed rainfall series.

Keywords-Markov Chain, Daily Rainfall Occurrence, Daily Rainfall Amount, Gamma Distribution

I. INTRODUCTION

Daily rainfall is one of the most vital meteorological parameter for agriculture based economy like Pakistan, as the majority of the populace is straightforwardly or circuitously fastened with the horticulture and associated products [i]. Agricultural productivity is positively or negatively affected due to the availability, unavailability or excessive overflow of water through Indus water network that is vulnerable to daily rainfalls [ii-iii]; mainly in monsoon season [iv-v]. Hence, sufficient and in time rainfalls are not only beneficial to the Kharif but also favorable and advantageous to the upcoming Rabi crops [vi].

Cutting-edge techniques of statistics ameliorated the applied methods' area available for the data analysis especially which are not normally distributed. This progression is of significant importance, as daily rainfalls are evidently not normally distributed [vii-viii]. In the contemporary epoch, the advancements for the developing models of daily rainfall frequency are significantly improved. Furthermore, besides the data generation, it renders the valuable information in agriculture, aviation and in many other related hydrological applications. Hence, an appropriate diurnal rainfall frequency model, in particular the probabilistic distribution of wet and dry spells, is substantial [ix].

Being a random phenomenon of rainfall occurrence, Markov Chain is generally considered as an effective illustration of the rainfall frequency. Then again, amount of daily rainfall is also crucial in weather characteristics, in particular, for agriculture. Both, the occurrence and amount of rainfall are also interconnected to the temperature, humidity and solar radiation as well as soil water balance that are also eminent in respect of growth and then further development of agricultural products, pests and related diseases, and, the weeds control. Because of this reason, in several agro-climatic studies, analysis and associated modeling of rainfall is chiefly focused, in which availability of the archives containing the record of rainfall data is very important.

In the field of consecutive rainfall frequencies, [x] are considered as the pioneers. They fitted the first order Markov Chain model to Tel Aviv data in successful manner. Lately, [xi] found that the same order is compatible for fitting the daily rainfall occurrence over Italy. The base of Markov Chain model is the assumption that occurrence of daily rainfall depends on the previous day's condition (i.e. whether it was dry or wet). It is considered as one of the most challenging problem in the field of hydrometeorology because extreme events like floods and droughts have been also in a relationship and driven by the daily rainfall. Hence, it is not easy task to model the daily rainfall stochastically because of its highly varied inherent nature [xii].

In general, sufficiently long time series of daily rainfall is utilized as input in hydro-meteorological models. In addition, to evaluate the sensitivity and suitability of such models, ensemble datasets are required to assess the long term changes in the rainfall regime. Then, the observed sequence provided a weather process realization that further used as an input data rainfall time series, in detailed impact studies from the simulated scenarios of climate change. Though, numbers of these sequences are still small in range because of the high computational cost of such scenarios. As a consequence, a synthetic sequence of rainfall data, based on the stochastic structure of the process is needed to evaluate the different ranges of outcomes that may be obtained with other equivalent series. This method was adopted by [xiii] to simulate the rainfall data, temperature and solar radiation on daily basis. First order Markov Chain has been utilized for the daily rainfall to describe the precipitation frequency while to approximate the distribution of the amount, exponential distribution has been employed. Afterward, [xiv-xv] replaces the exponential distribution by gamma distribution which has been now commonly adapted for most of the climate change studies.

Dry and wet behavior of weather is of great significance to all directly related and allied fields like agriculture, livestock, hydrology, industry, etc. The appropriate model of daily rainfall may be utilized in agricultural planning, drought management, flood predictions and soil erosion, climate change studies, and some other associated fields. The objective of the study is to develop a Markov Chain model for rainfall occurrence. Moreover, modeling of daily rainfall amount has been made which is desirable for wet days.

II. Materials and Methods

A. Study Area

For the study, five different meteorological stations selected from each of the province i.e. Multan, D. I. Khan, Nawabshah, Barkhan and Chilas from Punjab, Khyber Pakhtunkhwa, Sindh, Balochistan and Gilgit-Baltistan administration, respectively (Fig. 1).

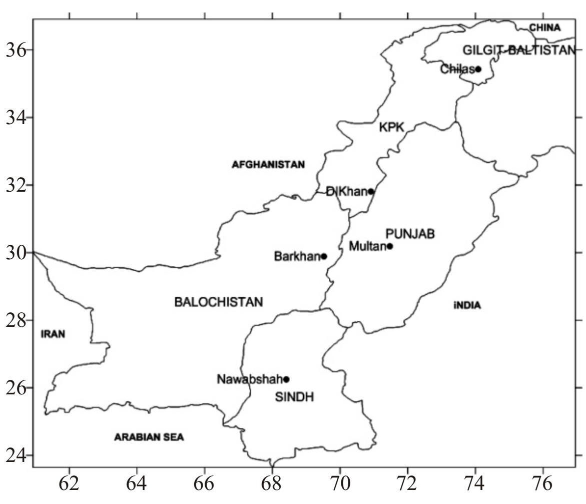


Fig. 1. Map showing the locations of selected cities of Pakistan

Daily rainfall data of 30 years (i.e. 1981-2010) utilized in the study is obtained from the archives of Pakistan Meteorological Department (PMD). The standard SI unit of millimeter (mm) is used in the study. For the studied period, highest and lowest amount of total normal rainfall is found for Barkhan (405.62mm) and Nawabshah (144.3mm), respectively. D. I. Khan bore the highest value of extreme rainfall (150mm) as on 4th August, 2010 while least extreme value is observed over Barkhan (104.2mm) for 23rd August, 1984. Fig. 2 shows the parametric values for all the considered cities of the country.

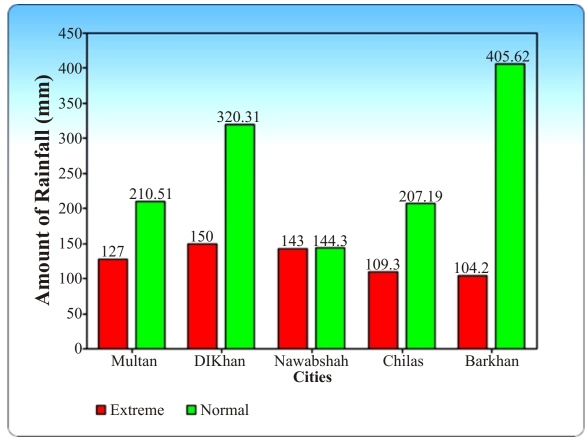


Fig. 2. Normal and extreme rainfalls for the period of 1981-2010

B. The Rainfall Model

a) Rainfall Probability

For the recorded *n* years, if specific event occurred in *m* of these years, subsequently probability of that occurred event in any given year is *m/n*. [xvi] illustrated that a curve may be fitted to the probabilities should be transformed to assure that the fitted curve is one of the probabilities given by

$$F = log(\frac{p}{1-p}) \quad (1)$$

Where $p = 0 \ to \ 1$

If *t* is the day of the year then function F(*t*) may be fitted to the *f* values and thus fitted probabilities may be given by

$$p(t) = exp\,\{\frac{F(t)}{1+exp\,F(t)}\} \quad (2)$$

Fourier series for the function F(*t*) of *n* harmonics is

$$F(t) = \mathrm{a_o} + \mathrm{a_1 sinx} + \mathrm{b_1 cosx} + \mathrm{a_2 sin2x} + \mathrm{b_2 cos2x} + \dots \mathrm{a_n sinnx} + \mathrm{b_n cosnx} \quad (3)$$

Where $x = \frac{\pi t}{366}$

A function given by this expression joins at the beginning and end of the year. This provides the curve with 2*n*+1 coefficient that describe the rainfall pattern here.

b) Rainfall occurrence

Occurrence of Rainfall may be described by two state Markov Chain model i.e. either day is dry or wet. Hence chance of rain or not on a particular day depends on the previous day that whether rain occurred or not. This kind of probabilistic approach based on Markov Chains has been utilized in many studies [xvii-xix] to generate synthetic rainfall sequence.

If random variables *X0, X1, X2, ..., Xn,* are identically distributed by considering only two probabilities, i.e. 0 and 1 such that

$Xn = 0$ *or* 1 if the n[th] day is wet or dry, respectively.

Initially, it may be assumed that,

$$P(Xn+1 = Xn+1 | Xn = xn, Xn-1 = xn-1, \ldots, Xo = xo) = P(Xn+1 = xn+1 | Xn = xn) \quad (4)$$

Where $xo, x1, \ldots, xn+1 \epsilon \{0,1\}$

Thus, it is expected that chance of rainy day depends only on the previous day (i.e. was wet or dry). In this assumption the probability of wetness is independent of further preceding days, therefore, the stochastic process $\{X_n\}$ with $n = 0,1,2, \ldots$ is a Markov Chain [xx].

Comparatively simple model concern with the probabilities of weather conditions depends on the state of preceding day, can be a simple transition matrix as

$$\begin{bmatrix} p_{oo} & p_{01} \\ p_{1o} & p_{11} \end{bmatrix}$$

On condition that $p_{ij} = p(X_1 = j | X_0 = i)$ $i,j = 0, 1$ is the probability, if a given day is type of i then it will be followed by a day of type j, that is p_{01} is the conditional probability of rainy day following a dry one, while p_{11} is the conditional probability of a rainy day following a wet one and so on in case of p_{10} and p_{00}.

Chance of the wet day occurrence may be ascertained by the comparison of a random number generated from a uniform distribution between the numerical values of 0 and 1 to the values of the mentioned transition probabilities p_{01} and p_{11}. If the random number is not larger than p_{01} than current day is wet and the preceding day was dry and the decision process is similar in case if the preceding day was wet. In this manner, once the wet day occurrence is established, the amount of rainfall on that is determined by generating a new random number from a uniform distribution and by solving the inverse cumulative function for the daily rainfalls.

c) Rainfall Amount

Amount of rainfall during rainy days can also be modeled to get the mean rainfall(s) per rain day. It not only set the desired variable for fitting of the amounts but also estimates the shape parameter of the utilized gamma distribution, then

$$F(x) = \frac{(k/\mu)x^{k-1}e^{kx/\mu}}{\Gamma(k)} \quad (5)$$

Where Γ is the gamma function, depends on the two parameters i.e. shape parameter k and the scale parameter μ i.e. mean rainfall per day which is assumed to be constant throughout the year and was estimated from all rainfall amounts. Consequently, with the implication of this condition, four curves were fitted to the mean amount of rain per rain day.

III. RESULTS AND DISCUSSION

The process of Markov Chain represents a system of components making transition from one state to another over time. The technique used in the study involves the utilization of first order transition probability matrix of a Marko Chain in addition to an algorithm to produce the time series of daily rainfall values. The way in which Markov Chain model can be utilized to generate rainfall time series have been illustrated in the subsequent sections.

A. Probability of Rainfall

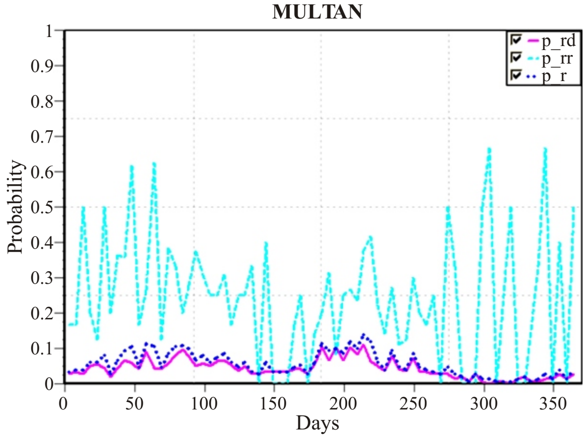
Rainfall probability is a helpful computational tool to study and analyze the distribution of rainfall in time. Several studies, as for instance, [xxi-xxiii] presented that rainfall occurrence probability at specific time is dependent on the preceding day(s) condition, called conditional probability. Contrariwise, independent rainfall occurrence is an unconditional probability (or overall probability) of rainfall.

B. Conditional and Unconditional Probability

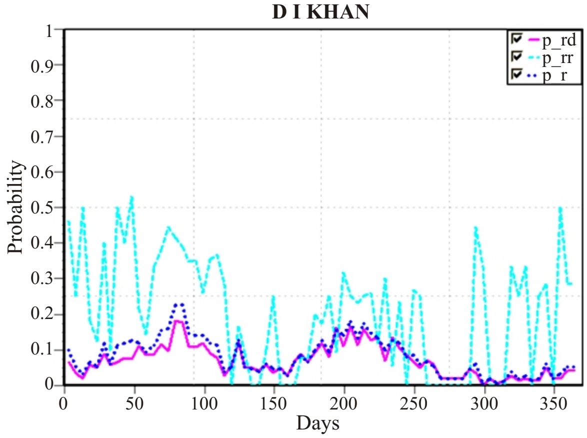
At given place on any given date(s), the proportion of wet days estimated the overall rain probability as (p_r). For the reason that it explore the overall chance of rain in the main (summer monsoon season) and small (winter) rainy season, at that place, such information is meaningful for agricultural planning. The analysis of first order Markov Chain deals with the computational probabilities of rains that depends on the condition whether previous day was rainy (p_rr) or dry (p_rd).

The conditional and unconditional probabilities of daily rainfall for the considered cities are depicted in Fig. 3. Chance of getting rain on a particular day in the small rainy season shows a significant continual increase for all the cities, if followed by a wet day. On the other hand, the probability of getting rain in the small rainy season will decrease significantly, if flowed by a dry day. Hence, the condition explores that the chance of rain occurrence in the small rainy season is significantly depended on whether the previous day was wet or dry. In D. I. Khan and Chilas small rainy season (with amplitude differences) appears with prominent values of probability. While Nawabshah, Barkhan and Multan shows the prominent large season's probabilities. Nawabshah appears with the least values of rr and rd. The D. I. Khan and Chilas appears with dominant small value probabilities as this area is more affected during the western disturbances era rather than eastern summer monsoon times [xxiv]. Also there is a very little demarcation of small and large season for Chilas.

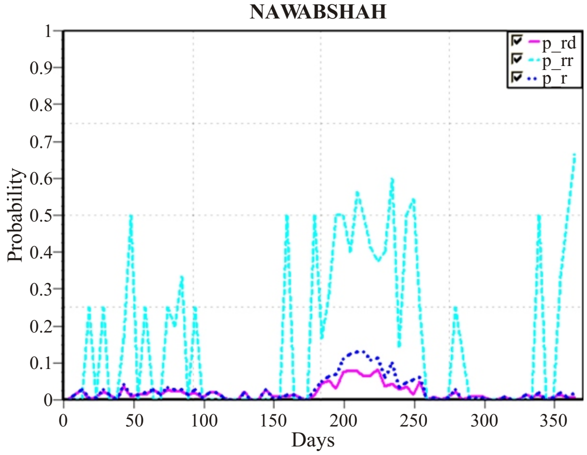
MULTAN



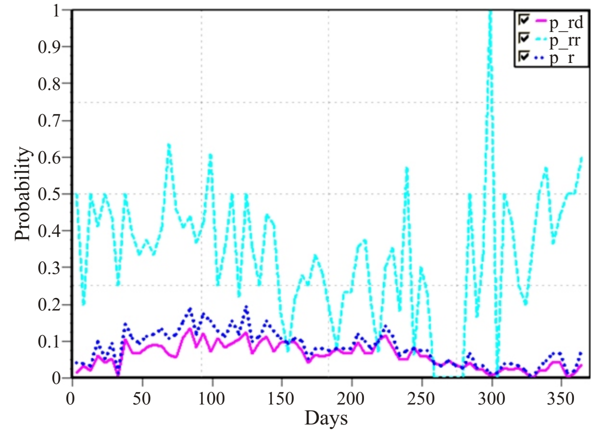
D I KHAN



NAWABSHAH



CHILAS



BARKHAN

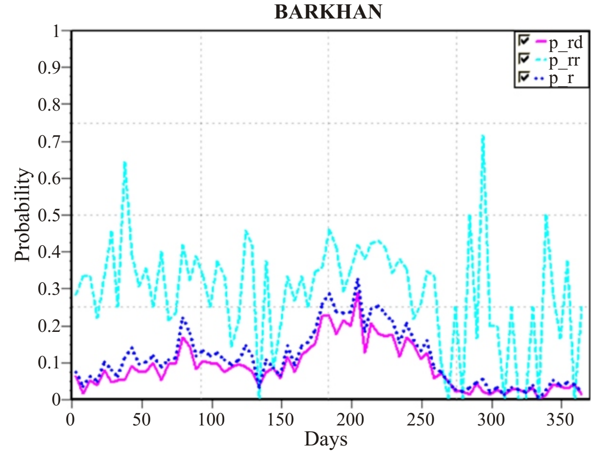


Fig. 3. Overall (p_r) and conditional probabilities (p_rr & p_rd) for the different cities of Pakistan

C. Fitting the probabilities

To observe the smooth and clear state of the rain probability, fitting of the probabilities to the chance of rain with the previous days, whether dry or wet, is performed. Variation in between the treatment implies response variation which is '*explained*' by the factor (i.e. predictor variable) in the model, and its sum of squares summarizes the model variability predictions (the fitted values). Variation within each factor level is '*unexplained*' by the factor in the model and its sum of squares (also called residual sum of squares) construe that how much the response values vary around best prediction of the response for that factor level. Total sum of squares reflects the overall variability of the response via

$$SS_{Total} = SS_{Explained} + SS_{Unexplained} \tag{6}$$

which may be expressed in the form

$$\sum(y_{ij} - \bar{y})^2 = \sum(fit_i - \bar{y})^2 + \sum(y_{ij} - fit_i)^2 \tag{7}$$

Where y_{ij} is the '*raw*' response after assuming randomized data with a categorical factor satisfying a normal model of the form

$$y_{ij} = (explained\ by\ factor) + (unexplained) \qquad (8)$$

or

$$y_{ij} = \mu_i + \varepsilon_{ij} \qquad (9)$$

For $i = 1\ to\ g$ and $j = 1$ to n_i where ε_{ij} ~ normal $(0, \sigma)$

and fitted values for categorical, linear and quadratic models, respectively are given by

$$fit_i = \begin{cases} \bar{y}_i \\ b_o + \quad b_1 x_i \\ b_o + \quad b_1 x_i \quad b_2 x_{i^2} \end{cases} \qquad (10)$$

Obviously each sum of squares is associated with a particular degree of freedom, as under

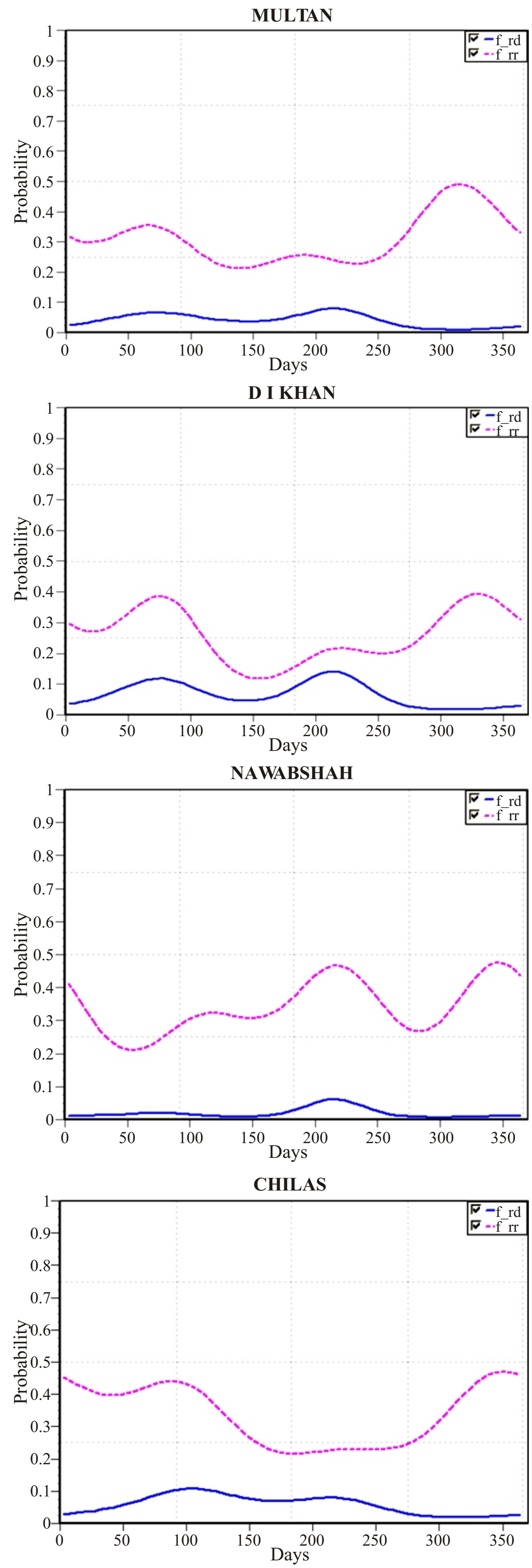
$$\begin{aligned} df_{total} &= n - 1 \\ df_{residual} &= n - k \\ df_{explained} &= k - 1 \end{aligned} \qquad (11)$$

implies that sums of squares are found by dividing each sum of squares by its degrees of freedom. Then *F* ratios are simply calculated by dividing each mean '*explained*' sum of squares by the mean residual sum of squares

$$F = \frac{MSS_{explained}}{MSS_{Unexplained}} \qquad (12)$$

If the added parameter(s) have no effect, the corresponding F ratio is expected to be around 1, though it can be somewhat higher or lower by chance. A p-value assesses whether it is unusually high and is interpreted in a similar way to all other *p*-values -- the closer the *p*-value to zero, the stronger the evidence that the term is needed in the model.

Hence, this results in the fitted probabilities (f_rd & f_rr) for the mentioned cities are drawn as shown in Fig. 4. For D. I. Khan, Nawabshah, Chilas and Barkhan over all probability of rain after a rainy day is less than 50% while for Multan, probability in small rainy season of a rainy day after the previous wet day reaches almost up to 50% to be occurring. For D. I. Khan and Chilas, and, for Multan and Nawabshah, chance of rain after a dry day is less than 10% and 20%, respectively.



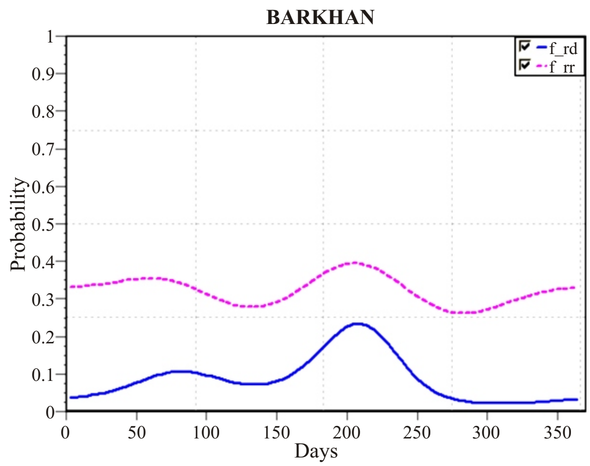
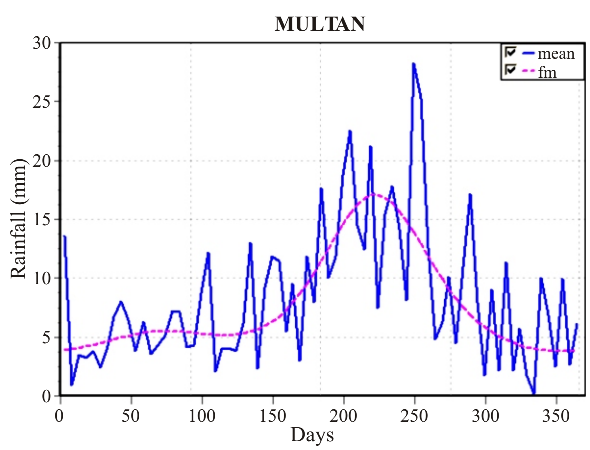
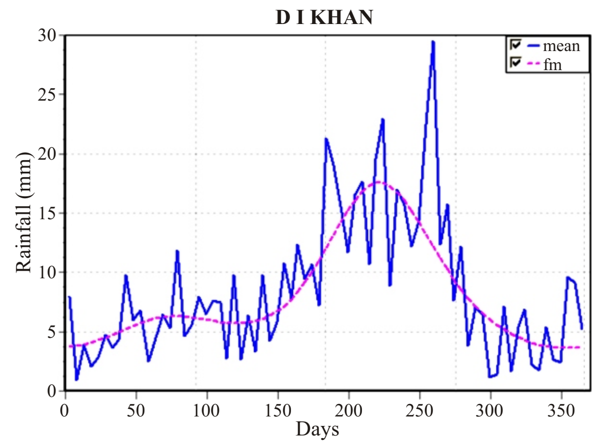


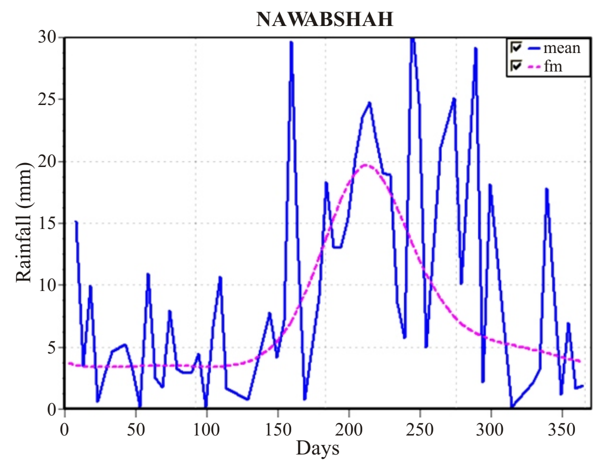
Fig. 4. Fitted means to the conditional probabilities for different cities of Pakistan

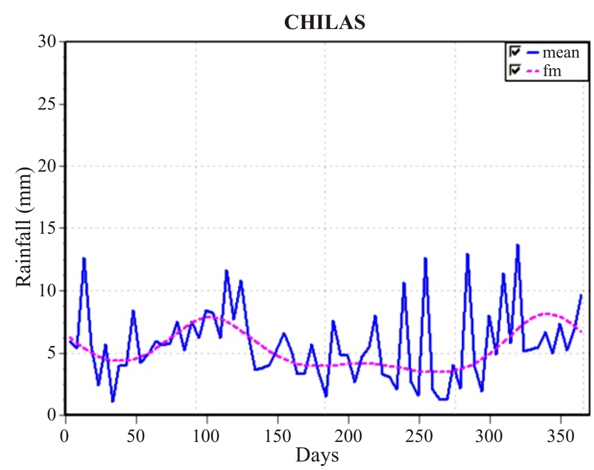
D. Modelling of Rainfall Amount

Gamma distribution is used to described the amount of rainfall. This distribution comprises of two parameters, scale parameter i.e. mean rain per day (μ) and shape parameter of the distribution (k). The mean rain per rain day (μ) is computed and after repeating the fitting process for rainfall amounts, plotted in Fig. 5. The fitted curves shows that mean rain per rain day vary in time and acquire maximum values in the peak months of large season with larger values in Multan and D. I. Khan ($\approx$17mm), Nawabshah ($\approx$20 mm), and Barkhan ($\approx$12 mm) as these stations are more influenced in monsoon season rather than westerlies while rainfall amount of Chilas is greater in small rainy season.









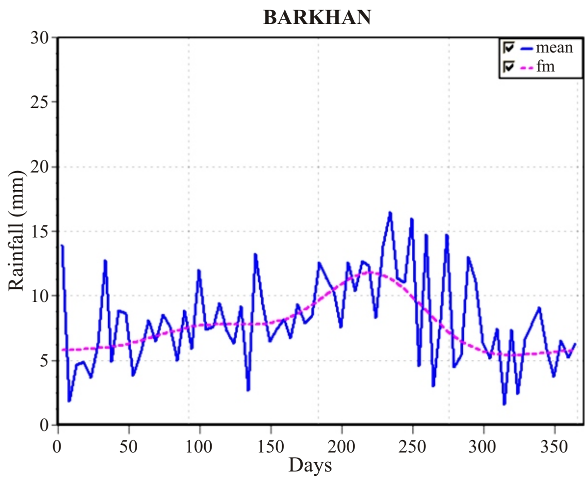


Fig. 5. Overall mean and fitted mean of rainfall amounts for different cities of Pakistan

Through the method of maximum likelihood, k-value for the each city is estimated that is assumed to be constant throughout the year. The calculated shape values have close values for the considered cities though, D. I. Khan (≈0.79) and Multan (≈0.72) appears with maximum and minimum, respectively (Fig. 6). Higher values of the shape parameter show analogous to exponential behavior of rainfall in the considered cities.

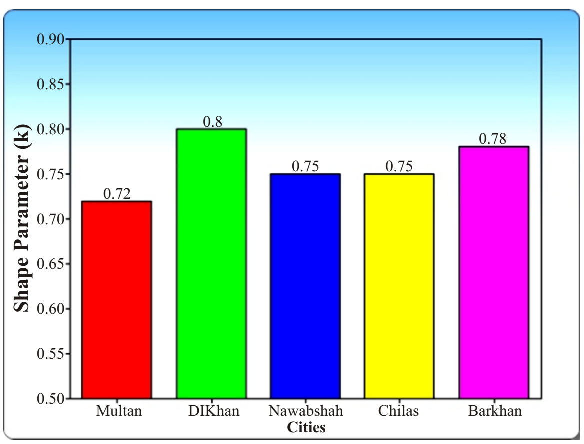
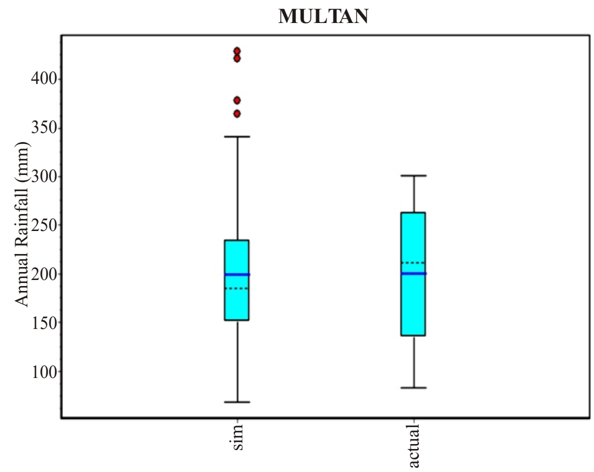
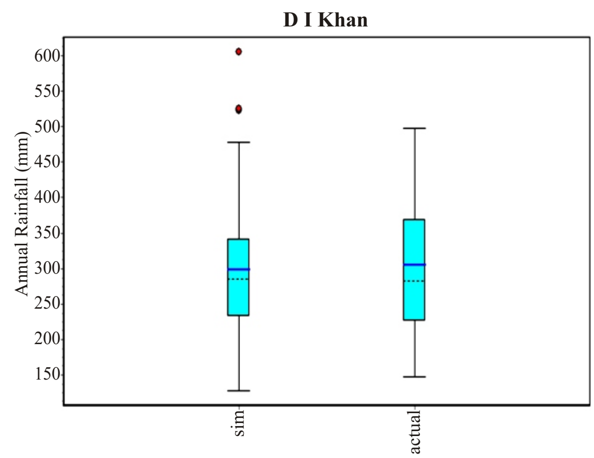


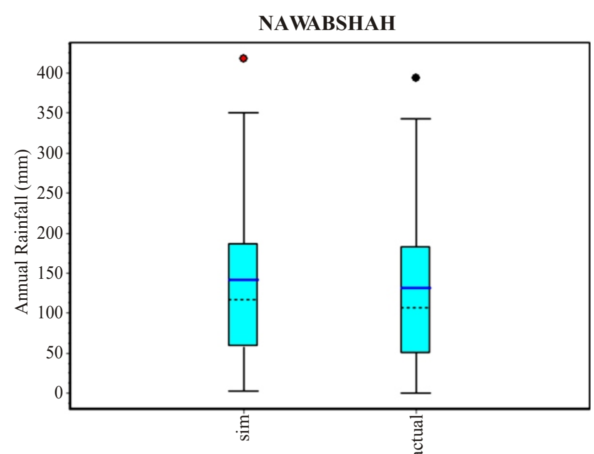
Fig. 6. Annual shape parameter for the respective cities

IV. Validation of the Model

In addition to the recognized procedure illustrated above, the synthetic rainfall time series were examined thoroughly to uncover their ability to uphold the statistical properties and to evaluate the applicability of Markov Chain models for rainfall generation. In this milieu, the significant statistical parameters are the general parameters (i.e. mean, standard deviation, etc.), autocorrelation functions and the probability distribution of the time series.







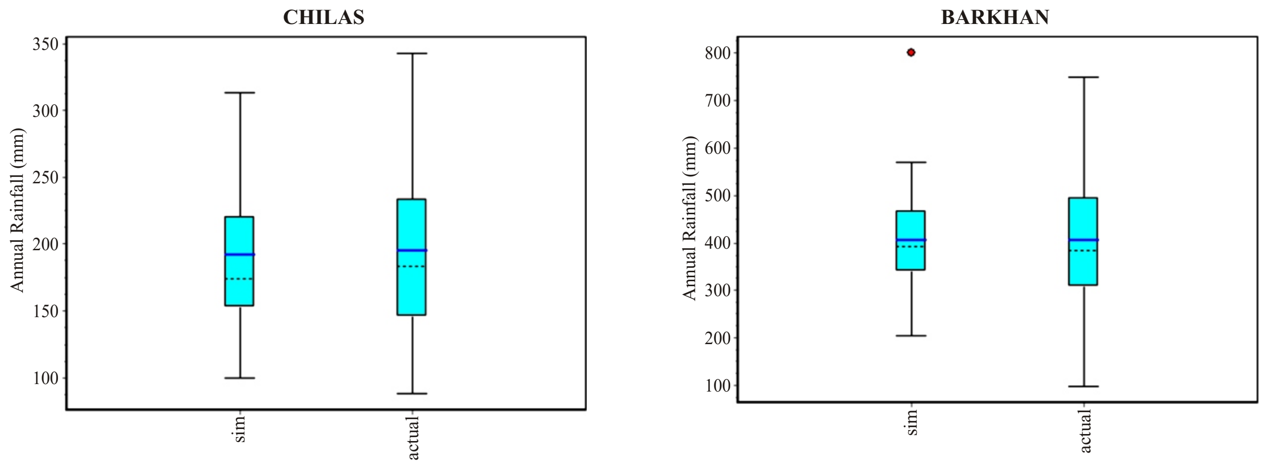


Fig. 7. Actual and simulated rainfall by first order Markov model for different cities of Pakistan

In order to examine the accuracy of first order Markov Chain modeling approach, the general statistical parametric values such as mean, standard deviation and the percentiles of the simulated values are summarized in Table I. Keeping in view the highly varied nature daily rainfalls, the comparison shows that the simulated values by the first order Markov Chain model is sufficient to preserve most of the nature of actual data, which implies the adequacy of the fitted Markov Chain model in respect of daily rainfalls for the selected cities of Pakistan.

TABLE I

SOME CHARACTERIZED STATISTICAL AND ANALYTICAL VALUES FOR SIMULATED AND ACTUAL DATA

City		Mean	Std. Deviation	20^{th} percentile	50^{th} percentile	80^{th} percentile	Min	Max
Multan	Simulated	198.54	75.78	134.32	184.8	264.14	67.831	428.81
	Actual	202.97	71.66	126.14	212.3	287	83	300.7
D I Khan	Simulated	297.7	89.24	222.34	285.73	358.46	127.76	605.09
	Actual	305.27	100.24	216.7	282.1	416.4	147.3	497
Nawabshah	Simulated	141	99.6	49.9	116.4	212	2.6	417.8
	Actual	130.2	110.1	29.6	107.3	242.3	0	393.4
Chilas	Simulated	191.9	58.8	148.3	173.8	253.2	99.6	313.7
	Actual	194.7	69.5	137.8	183.2	250.1	88.1	342.8
Barkhan	Simulated	406.1	101.6	331.3	392.2	476.8	204.7	800.9
	Actual	405.6	160.4	290.3	385.0	510.8	98.1	749.5

V. CONCLUSIONS

The models have utilized 30 years actual rainfalls to simulate the 50 years of synthetic rainfall time series. This study has examined the relative efficiency of the use of rainfall amount and wet days in the determination of generated rainfall at five different cities (Multan, Chitral, Nawabshah, Chilas and Barkhan) of Pakistan. Outcomes of this work show that both methods (the use of rainfall amount and wet days) are equally efficient with respect to the mean rainfall and wet days. The total predicted number of wet days is based on a first-order Markov Chain process for the month and the total amount of monthly rainfall for wet days is determined by gamma distribution. The total normal rainfall over Barkhan and Nawabshah are found to be of the highest (405.62mm) and lowest (144.3 mm) amount, respectively. D. I. Khan bears the highest extreme value (150 mm) while the least extreme appears for Barkhan (104mm) which is very close to Chilas extreme (i.e. 109.3mm). For D. I. Khan and Chilas small rainy seasons are dominated for the probability of rainfall while chance of rainfall in Multan is found up to 50% in the season if the previous day was also rainy. Fitted means exhibit the same scheme in a clearer and smooth way. An overall probability of rainy day after wet day is much greater than the rainy day followed by a dry one. Fitted mean of rainfall amount disclose the value under 10 mm for D. I. Khan and Chilas throughout the year. In respect of Barkhan these values are not below the 6 mm. It has been noted that simulated rainfall time series agreed fairly well with the actual observed rainfall series.

REFERENCES

[i] N. Sadiq and M. S. Qureshi, "Estimating Recurrence Intervals of Extreme Rainfall through a Probabilistic Modeling approach for different urban cities of Pakistan", *Arabian Journal for Science and Engineering,* 39 (1), 191-198, 2014.

[ii] N. Sadiq, A. A. Abbasi and M. S. Qureshi, "Drought Analysis of Sindh using Standardized Precipitation Index (SPI)", *Mausam*, 65 (3), 433-440, 2014.

[iii] F. Van Ogtrop, M. Ahmad & C. Moeller, "Principal components of sea surface temperatures as predictors of seasonal rainfall in rainfed wheat growing areas of Pakistan. Meteorological applications", 21(2), 431-443, 2014.

[iv] N. Faisal and N. Sadiq, "Monsoon onset over selected eastern boundary cities of Pakistan", *The Nucleus*, 49 (3), 239-245, 2012.

[v] N. Faisal and N. Sadiq, "Climatic Zonation of Pakistan through Precipitation Effectiveness Index", *Pak. J. Met*. 6(11): 51- 60, 2009.

[vi] N. Sadiq and M. S. Qureshi, "Climatic Variability and Linear Trend Models for the Five Major Cities of Pakistan", *J. of Geog. & Geol*., 2(1): 83-92, 2010.

[vii] A. Abo-Monasara and M. A. Al-Zahrania, "Estimation of rainfall distribution for the southwestern region of Saudi Arabia", 59 (2), 420-431, 2014.

[viii] R. D. Stern and R. Coe, "The use of rainfall models in agricultural planning", *J. Agril. Met*. 26: 35-50, 1982.

[ix] S. M. Deni and A. A. Jemain, "Fitting the distribution of dry and wet spells with alternative probability models", *Met. & Atmosph. Phy*. 104: 13-27, 2009.

[x] K. R. Gabriel and J. Neumann, "A Markov Chain model for daily rainfall occurrences at Tel Aviv", *Quart. J. Roy. Met. Soc*. 88: 90-95, 1962.

[xi] N. T. Kottegoda, L. Natale and E. Raiteri, "Some considerations of periodicity and persistence in daily rainfalls". *J. Hydrol*. 296: 23-37, 2004.

[xii] M. A. Barkotulla, "Stochastic Generation of the Occurrence and Amount of Daily Rainfall". *Pak. J. stat. oper. Res*., 6(1): 61-73, 2010.

[xiii] C. W. Richardson, "Stochastic simulation of daily precipitation temperature, and solar radiation", *Water Resour. Res*., 17: 182-190, 1981.

[xiv] D. S. Wilks, "Adapting stochastic weather generation algorithms for climate change studies", *Clim. Cha*., 22: 67-84, 1992.

[xv] D. S. Wilks, "Inter annual variability and climate change studies", *Clim. Cha*. 22: 67-84, (1999).

[xvi] R. D. Stern and R. Coe, "The use of rainfall models in agricultural planning", *J. Agril. Met*. 26: 35-50, 1982.

[xvii] G. A. Larsen and R. B. Pense, "Stochastic simulation of daily climatic data for agronomic models", *Agron. J. Hydrol*. 296: 23-37, 1982.

[xviii] J. Roldan and D. A. Woolhiser, "Stochastic daily precipitation models, A comparison of occurrence process", *Water Resour. Res*., 18: 1461-1468, 1982.

[xix] C. W. Richardson, "Stochastic simulation of daily precipitation, temperature and solar radiation", *Wat. Resour*. 17: 182-191, 1981.

[xx] J. Medhi, *Stochastic Process*. John Wiley & Sons, 69-70, 1981.

[xxi] J. L. Chaudhary, "Rainfall Analysis using Markov Chain Modeling for Bastar", *Geog. Rev. of Ind*., 60(1): 74-83, 2006.

[xxii] S. R. Samuel and S. S. Tamil, "Stochastic modeling of daily rainfall at Aduthurai", *Int. J. of Advan. Comp. and Math. Sci*., 1(1): 52-57, 2010.

[xxiii] A. Senthilvelan, A. Ganesh and K. Banukumar, "Markov Chain Model for probability of weekly rainfall in Orathanadu Taluk, Thanjavur District, Tamil Nadu", *Int. J. of Geomat. & Geosci*. 3 (1): 191-203, 2012.

[xxiv] I. Ahmed and N Sadiq, "Spatio-Temporal analysis of Western Disturbances over Pakistan", *The Nucl*, 49 (4): 329-338, 2012.