

Markov Modelling of daily rainfall over different cities of Khyber Pakhtunkhwa (KPK), Pakistan

NAEEM SADIQ

Institute of Space & Planetary Astrophysics, University of Karachi, Karachi, 75270, Pakistan

(Received 16 August 2013, Modified 9 April 2014)

e mail : nsadiq@uok.edu.pk

सार - खैबर पख्तूनख्वा प्रांत का कृषि योग्य भू-भाग दैनिक वर्षा पर अधिक निर्भर हैं जिससे वर्षा होने और वर्षा की मात्रा का मॉडल दोनों प्राचलों के प्रसंभाव्य उत्पत्ति (stochastic generation) से तैयार किए जा सकते हैं। प्रथम श्रेणी का मार्कोव चेन मॉडल वर्षा की तीव्रता का परिणाम (sequence) संक्रांतिक संभावना मैट्रिक्स का उपयोग करके तैयार करने के लिए लागू किया गया जबकि वर्ष 1981-2010 की अवधि की वर्षा की मात्रा ज्ञात करने के लिए गामा वितरण का उपयोग किया गया। इस प्रांत के भिन्न-भिन्न शहरों के प्राचलीय मानों को ज्ञात करने के लिए स्केल पैरामीटर्स तथा आकृति का आकलन करने के लिए क्षणों की पद्धति (Method of moments) का उपयोग किया गया जिसके द्वारा गामा वितरण के अनुसार कृत्रिम परिणाम (sequence) तैयार किया जा सकता है। इस शोध-पत्र में प्रतिबंधित और अप्रतिबंधित संभावनाओं को मध्यमान तथा अन्तर अथवा मानक विचलन के साथ मिलाकर प्रसंभाव्य उत्पत्ति (Stochastic generation) के लिए आवश्यक प्राचलों के रूप में विचार किया गया है। इस अध्ययन का उद्देश्य वर्षा की जाँच-पड़ताल करने के लिए एक समकालिक प्रसंभाव्य (Stochastic) विचार को प्रस्तुत करना है।

ABSTRACT. The cultivated areas in the Province Khyber Pakhtunkhwa (KPK) are highly dependent on the daily rainfall, whose occurrence and amount can be modeled through the stochastic generation of both parameters. First order Markov chain model has been employed to generate the sequence of rainfall frequency by using the transitional probability matrices while gamma distribution is used to generate rainfall amount for the period 1981-2010. Method of moments is used to estimate the shape and scale parameters to find the parametric values at different cities of the province which leads to synthetic sequence generation as per gamma distribution. Conditional and unconditional probabilities altogether with means and variance or standard deviation are being considered as the essential parameters for the stochastic generation in this study. The purpose of this study is to present a contemporary stochastic view of rainfall investigation.

Key words – Daily rainfall occurrence, Daily rainfall amount, Markov Modelling, Gamma distribution, Khyber Pakhtunkhwa.

1. Introduction

Majority of the Pakistan's population is directly or indirectly related to agriculture and related products. In this respect, daily rainfall is one of the most crucial parameter (Faisal and Sadiq, 2009). Major and minor tributaries of the Indus water network gets its major contribution in summer monsoon rainfall that further put on its effects over the different crops of the country (Sadiq and Qureshi, 2010). Hence the crops productivity is positively or negatively affected by daily rainfall and consequently over the Indus water network irrigation system that is vulnerable to daily rainfalls. The province Khyber Pakhtunkhwa (KPK) lies onwards to the plains of Punjab, Sind and beyond the Indus to South Asia. It is located in the north-west of the country. It borders Afghanistan to the north-west, Gilgit-Baltistan (GB) to the north-east, Kashmir to the east and Punjab to the south-east. The climate of KPK varies immensely for a region of its size, most of the many climate types found in Pakistan

as it varies in topography from dry rocky areas in the south to forests and green plains in the north. The province covers the area of 74,521 km² (Sadiq, 2012) and lies in between the ranges of 69° to 74° E longitude and 31° to 38° W latitude (Fig. 1).

Modern statistical methods have substantially ameliorated the area of applied methods available for the data analysis that are not normally distributed. As daily rainfalls are evidently not normally distributed, this evolution is of significant importance (Sadiq and Qureshi, 2014; Stern and Coe, 1982). In the recent era, the development of rainfall frequency models is significantly increased. Moreover, in addition to data generation, it provides the helpful information in agricultural, aviation and hydrological applications. Secernating the appropriate diurnal rainfall frequency model, particularly the probabilistic distribution of dry and wet spells is significant (Deni and Jemain, 2009).

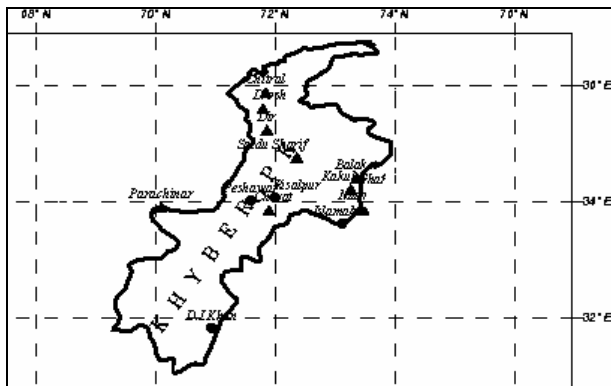


Fig. 1. Geographical location of Khyber Pakhtunkhwa

In fact rainfall occurrence is a random process and as a matter of fact Markov chain is generally accepted as an effective and simple description and illustration of the rainfall frequency. On the other hand, rainfall amount is also an important weather characteristic, particularly, for agriculture. Frequency and amount of rainfall are also related to the humidity, temperature and solar radiation likewise water balance in soil, which are also important in respect of the growth and development of agricultural crops, pests and related diseases, and the weeds control. That is why in many agro-climate studies, analysis and modeling of rainfall is mainly focused in which availability of rainfall data record is substantial.

The groundbreakers in the field of the sequential daily rainfall frequencies are Gabriel and Neumann (1962). They successfully fitted the first order Markov chain model to Tel Aviv data. In the recent times, Kottegoda *et al.* (2004) fitted the same order of daily rainfall occurrence over Italy. The premiss that daily rainfall occurrence is depends on previous day's condition (whether wet or dry) is the base of Markov chain model. As the extreme events like floods and droughts are also interrelated and drive by the rainfall, hence, its analysis and modeling is one of the most challenging problems in the field of hydrometeorology. Also, because of highly varied nature, it is not an easy task to model the daily rainfall stochastically (Barkotulla, 2010).

In Hydro-meteorological models, usually sufficiently long time series of daily precipitation is usually used as an input. Moreover, to assess the suitability and sensitivity of these models, ensemble datasets are needed to evaluate the long term changes in the regime of precipitation. The observed sequence provided a realization of the weather process which is further used as an input data precipitation time series in detailed impact studies from the simulated scenarios of climate change. These numbers of these sequences are still limited due to high computational cost these scenarios. Consequently, based on stochastic

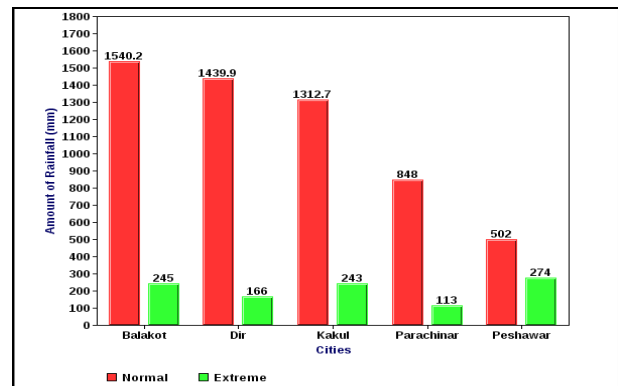


Fig. 2. Normal and extreme rainfalls for the period of 1981-2010

structure of the process, a synthetic sequence of rainfall data is needed to evaluate the different range of results that may be obtained with other equivalent series. Such a technique was adopted by Richardson (1981) to simulate data of rainfall, temperature and solar radiation on daily basis. First order Markov chain has been employed for the rainfall to describe the frequency of precipitation while to approximate the distribution of the amount, exponential distribution has been used. Later, instead of exponential distribution, Wilks (1992 and 1999) used the same model with gamma distribution which has been adapted for climate change studies.

The data regards dry and wet behavior of weather is crucial to all related and allied fields like agriculture, hydrology, industry, livestock, etc. The adequately and appropriately modeled rainfall may be used in agricultural planning, drought management, flood predictions and soil erosion, impacts of climate change studies, sowing and growth of crops, and some other related fields. The objective of the study is not only to analyze the rainfall occurrence (wet or dry) processes, but more specifically, modeling of daily rainfall in addition to its amount which is desirable for wet days. Rainfall analysis suggested that for future rainfall variations, Markov Chain approach provides a modeling alternative.

2. Study area and data

This study has been used the daily rainfall of five major cities of KPK (Balakot, Dir, Kakul, Parachinar and Peshawar) for the period from January 1981 to December 2010. Data utilized in this study were procured from Pakistan Meteorological Department, Karachi. The standard unit of millimeter (mm) is used throughout the study. For the studied period highest and lowest amount of total normal rainfall is found for Balakot (1540.2 mm) and Peshawar (502 mm), respectively. In contrast to total normal rainfall, Peshawar bore the highest value of extreme rainfall (274 mm) while least extreme value is

observed over Parachinar (113 mm). Fig. 2 shows the parametric value of both parameters for all the considered cities of KPK.

3. Methodology and methods

3.1. The rainfall model

3.1.1. Probability of rain

For the recorded n years, if particular event occurred in m of these years then probability of that event occurring in any given year is m/n . A distribution may be fitted to the probabilities should be transformed to ensure that the fitted curve is one of the probabilities given by

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

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 hence fitted probabilities may be given by

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

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

$$F(t) = a_0 + a_1 \sin x + b_1 \cos x + a_2 \sin 2x + b_2 \cos 2x + \dots + a_n \sin nx + b_n \cos nx \quad (3)$$

where, $x = \pi t/366$

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

3.1.2. Rainfall occurrence

Occurrence of Rainfall may be described by two state Markov chain, *i.e.*, either day is dry or wet. Hence chance of rain or not on a given day depends on the previous day that whether rain occurred or not. This type of probabilistic approach based on Markov chains has been used in many studies to generate synthetic rainfall.

If random variables $X_0, X_1, X_2, \dots, X_n$, are identically distributed and taking only two values, *i.e.*, 0 and 1 such that

$X_n = 0$ or 1 if the n th day is dry or wet, respectively.

Firstly, it may be assumed that,

$$P(X_{n+1} = X_n + 1 | X_n = xn, X_n - 1 = xn - 1, \dots, X_0 = x_0) = p(X_{n+1} = xn + 1 | X_n = xn) \quad (4)$$

where, $x_0, x_1, \dots, x_n + 1 \in \{0, 1\}$

Thus, it is assumed that chance of rainy day depends only on the previous day was wet or dry. In this assumption the chance of wetness is independent of further preceding days, hence, the stochastic process $\{X_n\}$ with $n = 0, 1, 2, \dots$ is a Markov chain.

Relatively simple model regards the probabilities of weather conditions depends on the preceding day conditions can be a transition matrix as

$$\begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}$$

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

3.1.3. Amount of rainfall

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

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

where, Γ is the gamma function is of the form

$$\Gamma(k) = \int_0^{\infty} t^{k-1} e^{-t} dt$$

The gamma distribution depends on the two parameters, *i.e.*, shape parameter k and the mean rainfall per day (μ) which is assumed to be constant throughout the year and was estimated from all rainfall amounts.

Consequently, with this condition four curves were fitted to the mean amount of rain per rain day.

As maximum likelihood estimator is considered to be as the most popular estimator for the estimation of α , let

$$F(x) = \mu^{-nk} [\Gamma(k)]^{-n} \left(\prod_{j=1}^n x_j \right)^{k-1} \exp\left(-\frac{1}{\mu} \sum_{i=1}^n x_i\right) \quad (6)$$

With the help of corresponding likelihood function, maximum likelihood estimation of k and μ can be determined by the following equations

$$\ln \mu + \Psi(k) = \sum_{j=1}^n \ln \frac{x_j}{n} \quad (7a)$$

$$k - \sum_{i=1}^n \frac{x_i}{(n\mu)} = 0 \quad (7b)$$

where, $\Psi(k) \cong \frac{d}{dk} \ln \Gamma(k) \cong \ln \Gamma(k)$ is the Euler Psi function. Hence from those equations maximum likelihood estimators (*i.e.*, k^* and μ^*) of k and μ , respectively. To be specific, k^* is the root of the following equation,

$$g(k) \cong \ln k - \Psi(k) = \ln \bar{x} - \frac{1}{n} \sum_{j=1}^n \ln x_j \quad (8)$$

at the same time as

$$\mu^* = \frac{\bar{x}}{k^*} \quad (9)$$

where, \bar{x} is the sample mean, *viz.*,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_k$$

It explores that the function $g(k)$ is rapidly decreasing and takes values in $(0, \infty)$ and hence, the estimator k^* is well defined and unique.

4. Results and discussion

4.1. Probability of rainfall

Rainfall Probability is a useful computational tool to study and analyze the rainfall distribution in time. Some studies (Chaudhary, 2006; Samuel and Tamil 2010; Senthilvelan *et al.*, 2012) presented that the probability of rain occurrence at specific time is dependent on the preceding day(s) condition, called conditional probability. On the other hand independent rainfall occurrence is known as unconditional probabilities (or overall probability) of rainfall.

4.2. Conditional and unconditional probability

At particular place on any given date(s), the proportion of wet days estimated the overall rain probability as (p_r). Because it explore the overall chance of rain in the main (*i.e.*, summer) and small (*i.e.*, winter) rainy season, at that place, such information is significant for agricultural planning. The analysis of first order Markov chain deals with the computational chances of rains that depends on whether previous day was rainy (p_{rr}) or dry (p_{rd}).

The conditional and unconditional probabilities of rain of the mentioned cities are shown 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 Dir, Parachinar and Peshawar small rainy season (with amplitude differences) appears with greater values of probability. The greater probability of rain occurrence followed by previous rainy day is the indication that the province is more affected during the western disturbances era rather than eastern summer monsoon times (Ahmed and Sadiq, 2012).

4.3. Fitting the probabilities

To observe the more clear and smooth 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 means is response variation which is ‘explained’ by the factor in the model, and its sum of squares summarizes the variability of the model 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

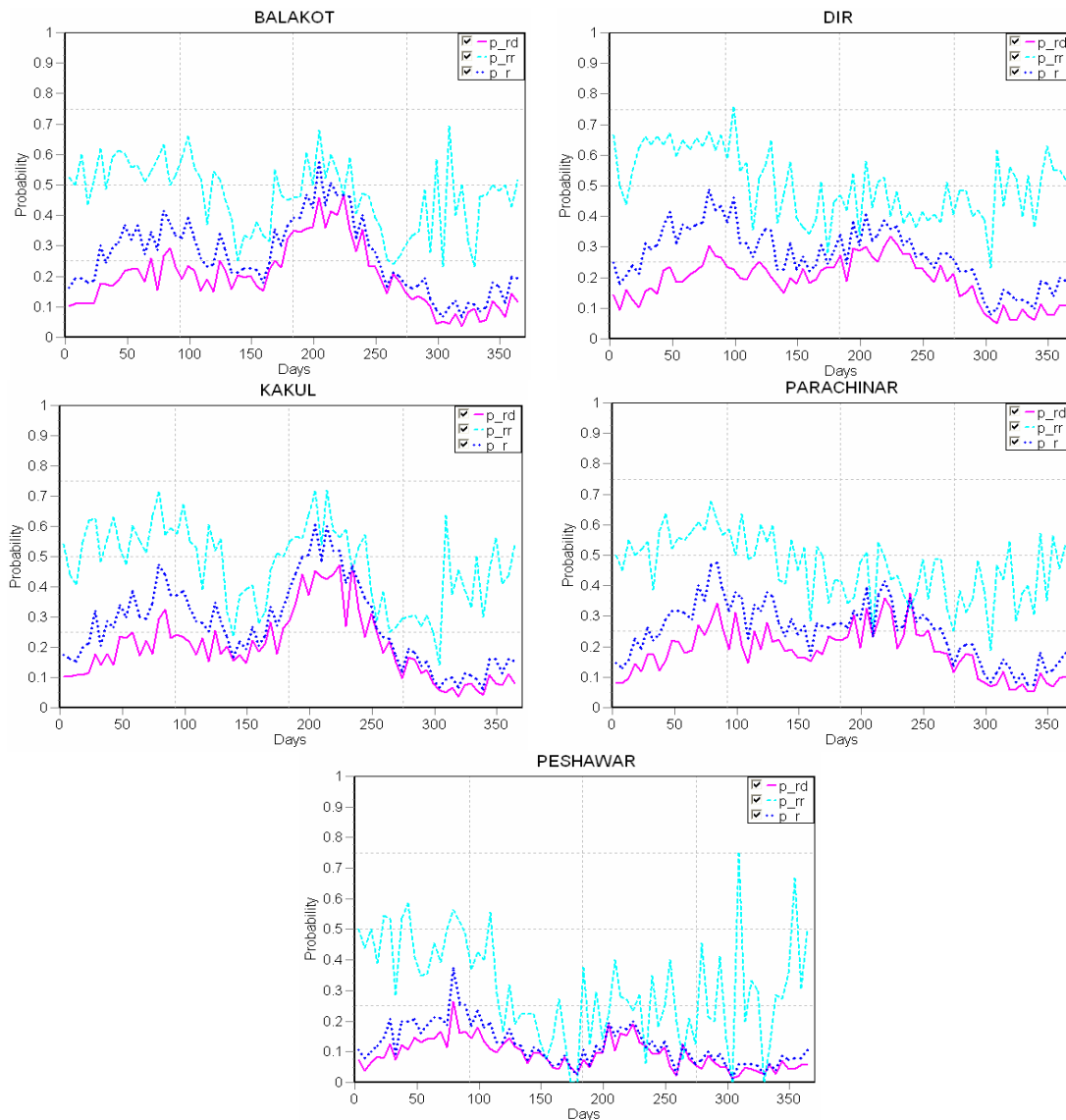


Fig. 3. Overall (p_r) and conditional probabilities (p_{rr} & p_{rd}) for the different cities of Pakistan

of squares) concludes 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, viz. :

$$SS_{Total} = SS_{Explained} + SS_{Unexplained} \quad (10)$$

which may be given in the form,

$$\sum (y_{ij} - \bar{y})^2 = \sum (\hat{fit}_i - \bar{y})^2 + \sum (y_{ij} - \hat{fit}_i)^2 \quad (11)$$

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

$$y_{ij} = (\text{explained by factor}) + (\text{unexplained}) \quad (12)$$

$$\text{or } y_{ij} = \mu_i + \epsilon_{ij} \quad (13)$$

For $i = 1$ to g and $j = 1$ to n_i where $\epsilon_{ij} \sim \text{normal}(0, \sigma)$ and fitted values for categorical, linear and quadratic models, respectively are given by

$$\hat{fit}_i = b_0 + b_1 x_i + b_2 x_i^2 \quad (14)$$

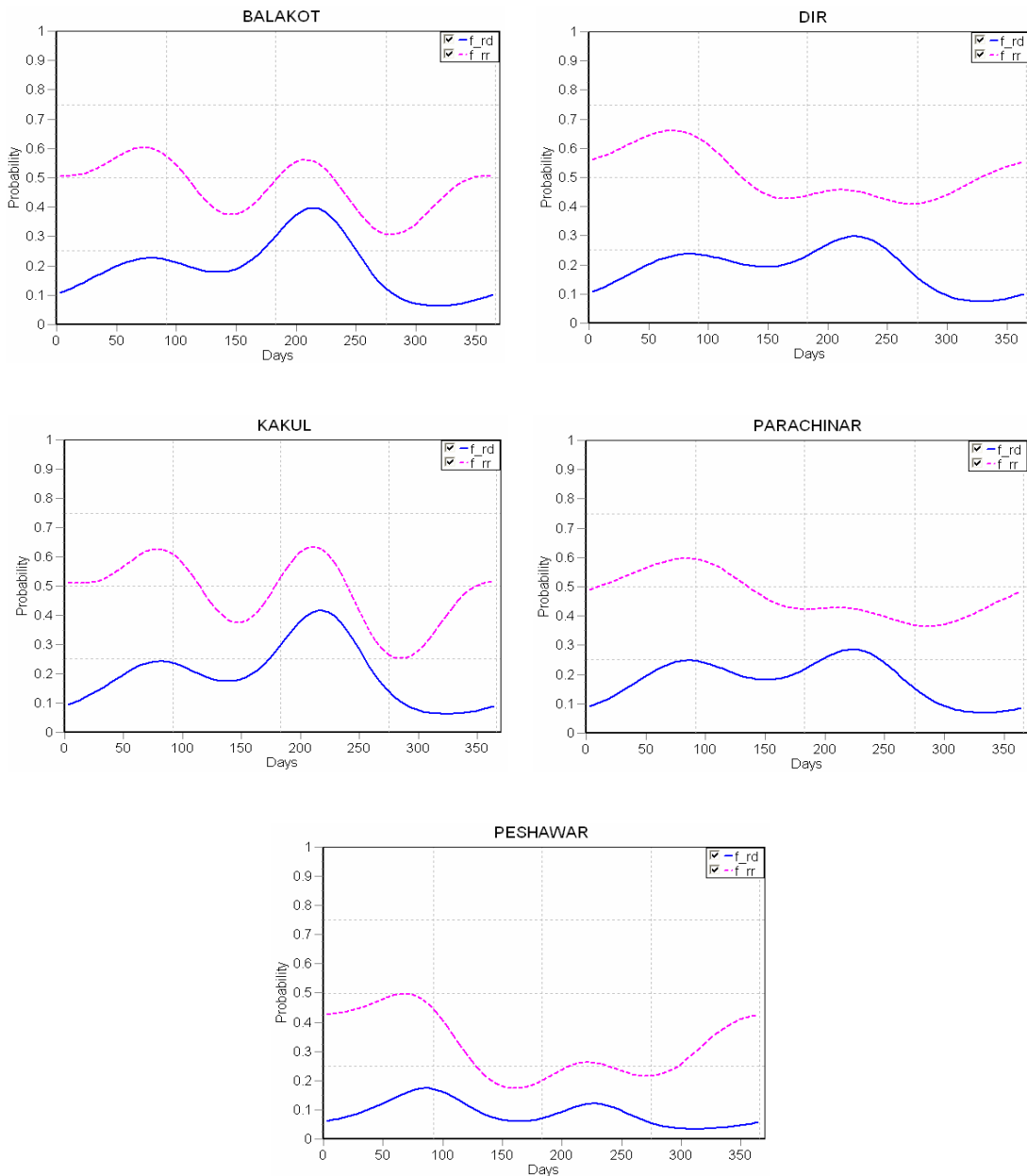


Fig. 4. Fitted means to the conditional probabilities for different cities of Pakistan

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

$$\begin{aligned}
 df_{\text{total}} &= n - 1 \\
 df_{\text{residual}} &= n - k \\
 df_{\text{explained}} &= k - 1
 \end{aligned}
 \tag{15}$$

Mean sums of squares are found by dividing each sum of squares by its degrees of freedom. F ratios are then

calculated by dividing each mean explained sum of squares by the mean residual sum of squares

$$F = \frac{MSS_{\text{explained}}}{MSS_{\text{unexplained}}}
 \tag{16}$$

If the parameter(s) being added 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

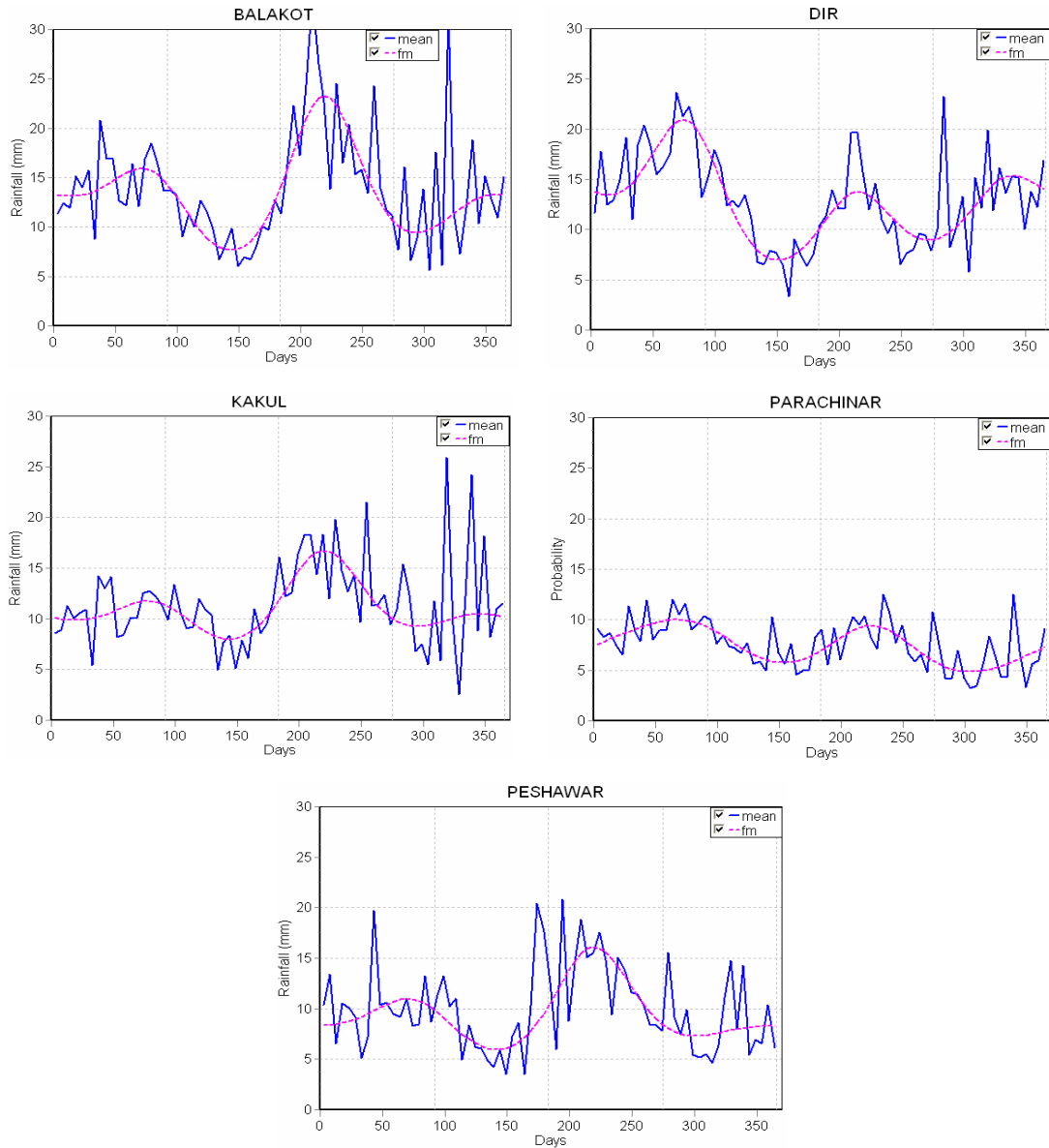


Fig. 5. Overall mean and fitted mean of rainfall amounts for different cities of Pakistan

to zero, the stronger the evidence that the term is needed in the model.

Hence, finally regression coefficients are estimated and obtained the fitted probabilities (f_{r_d} & f_{r_r}) as shown in Fig. 4. For Peshawar chances of rainy day after the previous wet day are less than 50% throughout the year while for monsoon times it is less than or equal to 25%. Parachinar and Dir also have less than 50% probability of rain after rainy day in monsoon season. Chance of rain after a dry day is less than 20% for Peshawar throughout the year. While for Dir and Parachinar it is less than 25% except for very little time in monsoon when it exceeds up to 28%.

4.4. Modelling of rainfall amount

Gamma Distribution is used to describe the amount of rainfall. This distribution comprises of two parameters, mean rain per day (μ) and shape 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 show that mean rain per rain day vary in time and acquire maximum values in the peak months of large (*i.e.*, summer) season with larger values in Balakot (23 mm), Kakul (17 mm) and Peshawar (around 16mm) as these cities are more influenced in monsoon rather than westerlies while Dir, Chitral and Parachinar amount are greater in small rainy season.

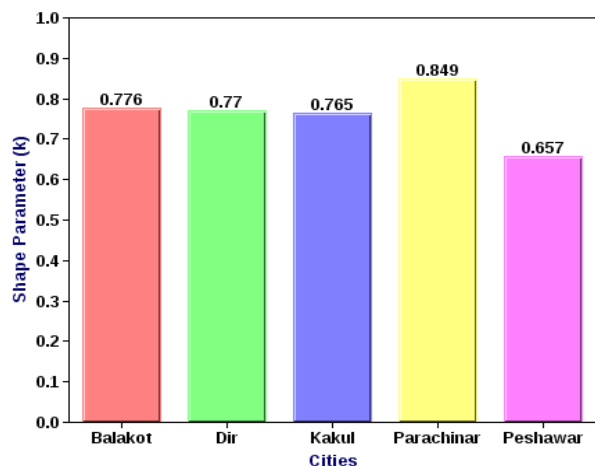


Fig. 6. Annual shape parameter for the respective cities

The maximum likelihood estimated the k-value for each city which is assumed to be constant throughout the year. The shape values of Peshawar and Parachinar are minimum and maximum, respectively (Fig. 6). Higher values of the shape parameter show similar to exponential behaviour of rainfall in the considered cities.

5. Conclusion

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 (Balakot, Dir, Kakul, Parachinar and Peshawar) of the province Khyber Pakhtunkhwa, 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 Balakot and Peshawar are found to be of the highest amount (1540.2 mm) and lowest (502 mm) amount, respectively. In contrast, Peshawar bears the extreme rainfall amount (274 mm) while Parachinar extreme (113 mm) is least. In small rainy season chance of rainfall in Dir, Parachinar and Peshawar is found higher if the previous day was also rainy. Fitted means show that this chance is 60% or more during throughout the small season. Fitted mean of rainfall amount shows that significant amount of rainfall amount is found in Balakot, Kakul and Peshawar in large season. The study may be extended to drought prone areas of the country at large. Further, application of Markov chain to the individual seasons may be more useful for further investigation and microanalysis of each season.

References

- Ahmed, I. and Sadiq, N., 2012, "Spatio-temporal analysis of western disturbances over Pakistan", *The Nucleus*, **49**, 4, 329-338.
- Barkotulla, M. A. B., 2010, "Stochastic generation of the occurrence and amount of daily rainfall", *Pak. J. Stat. Oper. Res.*, **6**, 1, 61-73.
- Chaudhary, J. L., 2006, "Rainfall Analysis using Markov Chain Modeling for Bastar", *Geog. Rev. of Ind.*, **60**, 1, 74-83.
- Deni, S. M. and Jemain, A. A., 2009, "Fitting the distribution of dry and wet spells with alternative probability models", *Met. & Atmosph. Phys.*, **104**, 13-27.
- Faisal, N. and Sadiq, N., 2009, "Climatic Zonation of Pakistan through Precipitation Effectiveness Index", *Pak. J. Met.*, **6**, 11, 51- 60.
- Gabriel, K. R. and Neumann, J., 1962, "A Markov chain model for daily rainfall occurrences at Tel Aviv", *Quart. J. Roy. Met. Soc.*, **88**, 90-95.
- Kottegoda, N. T., Natale, L. and Raiteri, E., 2004, "Some considerations of periodicity and persistence in daily rainfalls", *J. Hydrol.*, **296**, 23-37.
- Richardson, C. W., 1981, "Stochastic simulation of daily precipitation temperature, and solar radiation", *Water Resour. Res.*, **17**, 1, 182-190.
- Sadiq, N. and Qureshi, M. S., 2010, "Climatic Variability and Linear Trend Models for the Five Major Cities of Pakistan", *J. of Geog. & Geol.*, **2**, 1, 83-92.
- Sadiq, N., 2012, "Thunderstorm and Rainfall Activity over Khyber Pukhtunkhwa (KPK), Pakistan", *The Nucleus*, **49**, 3, 231-237.
- Sadiq, N. and Qureshi, M. S., 2014, "Estimating Recurrence Intervals of Extreme Rainfall through a Probabilistic Modeling approach for different urban cities of Pakistan", *Arab. J. for Sci. and Engin.*, **39**, 1, 191-198
- Samuel, S. R. and Tamil, S. S., 2010, "Stochastic modeling of daily rainfall at Aduthurai", *Int. J. of Advan. Comp. and Math. Sci.*, **1**, 1, 52-57.
- Senthilvelan, A., Ganesh, A. and Banukumar, K., 2012, "Markov Chain Model for probability of weekly rainfall in Orathanadu Taluk, Thanjavur District, Tamil Nadu", *Int. J. of Geomat. & Geosci.*, **3**, 1, 191-203.
- Stern, R. D. and Coe, R., 1982, "The use of rainfall models in agricultural planning", *J. Agril. Met.*, **26**, 35-50.
- Wilks, D. S., 1992, "Adapting stochastic weather generation algorithms for climate change studies", *Clim. Cha.*, **22**, 67-84.
- Wilks, D. S., 1999, "Inter annual variability and climate change studies", *Clim. Cha.*, **22**, 67-84.