4th International Conference on Water Resources and Arid Environments (ICWRAE 4): 141-146
5-8 December 2010, Riyadh, Saudi Arabia

Modeling Monthly Rainfall Records in Arid Zones Using Markov Chains: Saudi Arabia Case Study

Amro Elfeki and Nassir Al-Amri

Department of Hydrology and Water Resources Management, Faculty of Meteorology, Environment and Arid Land Agriculture, King Abdulaziz University, Jeddah, Saudi Arabia

Abstract: Water scarcity is a major problem in many countries especially in arid and semi-arid areas. The scarcity of water is further stressed by the growing demand due to increase in economic activities and population growth in developing countries. Added to these stresses, is the uncertain climate change and its consequences on water resources, especially in Arid and Semi-arid zones. Attention has to focus on developing a strategic management plan to preserve the resource from abusing, contamination and trying to maximize the resource. Nowadays, numerous governmental organizations are seeking evaluation of the existing water resources and estimation of future water resources in the Kingdom of Saudi Arabia. This knowledge is vital for decision makers to achieve the strategic plans of the country. Hydrological records (such as time series of daily rainfall, evaporation, runoff from intermittent streams etc.) in arid and semi-arid regions have extensive periods of zero values. Traditional approaches such as moving averaging techniques (MA), autoregressive models (AR), combined autoregressive moving average models (ARMA) and autoregressive-integrated-moving average models (ARIMA) could only be used when the time series is Gaussian. However, if the time series is not Gaussian, a transformation has to be applied before these models could be used, however, such tranformation does not always work. The dry-wet days' models are available in the literature. These models are mainly formulated in a context of Markov chain theory. The formal way of application of Markov model is that, the transition probability between dry and wet days are assumed stationary. In this paper an attempt is made to address such records in arid and semi-arid regions. The use of Markov chain seems promising in arid regions. The model is conceptually simple yet mathematically sound it is capable of addressing descrite data (intermittent series). Two spreadsheet models have been developed to check the probability distribution of the rainfall data at some stations and the other model is to calculate the Markov chain parameters form the data series and generate some future values of the rainfall and checking the model performance in a global sense.

Key words: Markov chains • Transition probability • Time series • Rainfall modelling • Statistical analysis

INTRODUCTION

Water scarcity is a major problem in many countries especially in arid and semi-arid areas. The scarcity of water is further stressed by the growing demand due to increase in economic activities and population growth in developing countries. Added to these stresses, is the uncertain climate change and its consequences on water resources, especially in arid and Semi-arid zones. Attention has to focus on developing a strategic management plan to preserve the resource from abusing, contamination and trying to maximize the resource. Nowadays, numerous governmental organizations are seeking evaluation of the existing water resources and estimation of future water resources in the Kingdom of Saudi Arabia. This knowledge is vital for decision makers to achieve the strategic plans of the country.

Hydrological records (such as time series of daily rainfall, evaporation, runoff from intermittent streams etc.) in arid and semi-arid regions have extensive periods of zero values. These series sometimes called intermittent time series. Intermittency is found in many physical phenomena as well [1]. Traditional approaches to handle such data are based on methods that have been developed in wet regions which do not suit the conditions in Saudi Arabia. An attempt should be made to address such records. Models should be developed for such arid and semi-arid regions.

Corresponding Author: Amro Elfeki, Department of Hydrology and Water Resources Management, Faculty of Meteorology, Environment and Arid Land Agriculture, King Abdulaziz University, Jeddah, Saudi Arabia

Time series of daily rainfall in arid and semi-arid regions have special features (e.g. intermentent series with many zeros, presistence of dry periods, short duration storms, highly intensive rainfall, non-Gaussianity of the data) that challenge modelling by traditional approaches such as moving averaging techniques (MA), autoregressive models (AR), combined autoregressive moving average models (ARMA) and autoregressive-integrated-moving average models (ARIMA), for details a reference is made to Box and Jenkins [2] and Yevjevich [3]. These models are discussed in details by Salas *et al*. [4] and Fernandez and Salas [5]. These models could only be used when the time series is Gaussian. However, if the time series is not Gaussian, a transformation has to be applied before these models could be used, Yevjevich [3].

The above mentioned models cannot be applied in arid and semi-arid regions on daily or monthly bases because of the persistent of long dry periods. So, we suggest developing models that are suited for application in arid and semi-arid conditions based on monthly information and taking into account specific features of Saudi Arabic aridity conditions. The suggested models could be integrated to longer time scales e.g. years and decades.

The dry-wet days' models are available in the literature [6, 7]. These models are mainly formulated in a context of Markov chain theory [8]. The formal way of application of Markov model is that, the transition probability between dry and wet days is assumed stationary (i.e. it does not change with time) Gabriel and Neumann, [6] and Sharma, [7].

The use of Markov chain seems promising in arid regions. The model is conceptually simple yet mathematically sound it is capable to address descrite data (intermittent series). It has been applied in many fields in the literature [9, 10] and proved to be successful.

MATERIALS AND METHODS

Modelling Monthly Rainfall as a Markov Chain.

In this research, the monthly rainfall is modeled as a Markov chain. We are going to distinguish two sates: the wet and dry months. The change the two states is modeled with a transition probability matrix in the form:

$$p_{ij} = \begin{matrix} & d & w \\ \begin{matrix} d \\ w \end{matrix} & \multicolumn{2}{c}{\begin{bmatrix} 1-a & a \\ b & 1-b \end{bmatrix}} \end{matrix}, \; i = d, w \text{ and } j = d, w \qquad (1)$$

Where p_{ij} is the transition probability to change between state *i* (dry *d* or wet *w*) and state *j* (dry or wet).

The values of *a* and *b* (the off-diagonal elements of the chain) show how strong or weak the change from wet to dry tor dry to wet. The stationary probability of the chain is given by:

$$\begin{bmatrix} \pi_w \\ \pi_d \end{bmatrix} = \begin{bmatrix} a/(a+b) \\ b/(a+b) \end{bmatrix} \qquad (2)$$

π_w is the proportion of the wet months in the time series and π_d is the proportion of the dry months in the time series.

Data Collection and Inference of Model Parameters: Three stations have been considered, Namely, Khules (J212), Amlog and Tabouk (TB001). The data of these stations are monthly data. Figure 1 shows the monthly data time series of station Khules. Statistical analysis has been performed on the data using an Excel sheet developed in this study. The code is designed to calculate the descriptive statistics for grouped and ungrouped data

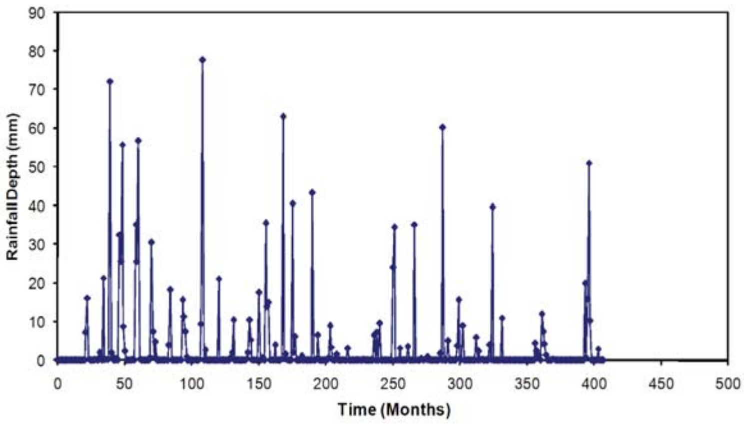

Fig. 1: A Typical monthly rainfall pattern at station J212 (records from 1965 to 1998)

Table 1: Statistical measures and hypothesis testing results

Station	Arith. Mean (mm)	SD (mm)	CV	Geo. Mean (mm)	Skew	Kurt.	χ^2 (α=0.01)	K-S Test (α=0.05)
Khules	9.9	16	1.6	2.48	2.3	5	-------	Log-normal
Amlog	14.1	17.6	1.2	7.8	2.0	4.5	-------	ExponentialLog-normal
Tabouk	7.5	9.75	1.3	3.7	2.5	7.1	-------	Exponential Log-normal

and also perform hypothesis testing to fit a probability distribution function to the data. Some of the descriptive statistics of the data (non-zero values) are presented in Table 1. These statistics provides the arithmetic mean, standard deviation, coefficient of variation, geometric mean, skewness coefficient and kurtosis coefficient respectively as sorted in the table. It is obvious from the table that the data are highly skewed since the skewness coefficient is larger than zero and it is positively skewed. The kurtosis indicates peaked distribution w.r.t. the normal distribution, since the values are larger than 3.

Another Excel sheet has also been developed in this study to estimate the Markov chain model parameters and to perform numerical simulation of the time series based to the statistics developed in the first Excel sheet. Table 2 shows the monthly transition probabilities for station Khules, Amlog and Tabouk calculated by the Excel model. It is obvious from the table that the transition probability from dry to dry is highest which indicates long dry seasons in the series which is common of arid zones.

RESULTS AND DISCUSSION

Some results of the models are presented in Table 1 and in Figures 2, 3, 4, 5, 6 and 7. Table 1 last two columns show the results of X^2 test and Kolmogorov Smirnov (K-S) test. The results show that the X^2 test (applied on grouped data) rejects all the probability distributions considered in this study, however, K-S test (applied on ungrouped data) accepts some distributions as shown in

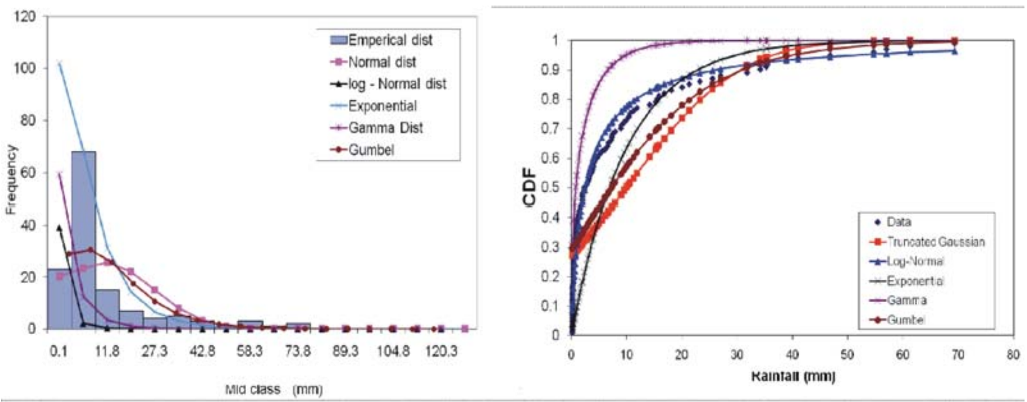

Fig. 2: Fitting a histogram (left) and probability distribution function (right) to the non-zero monthly rainfall depth at station Khules J212

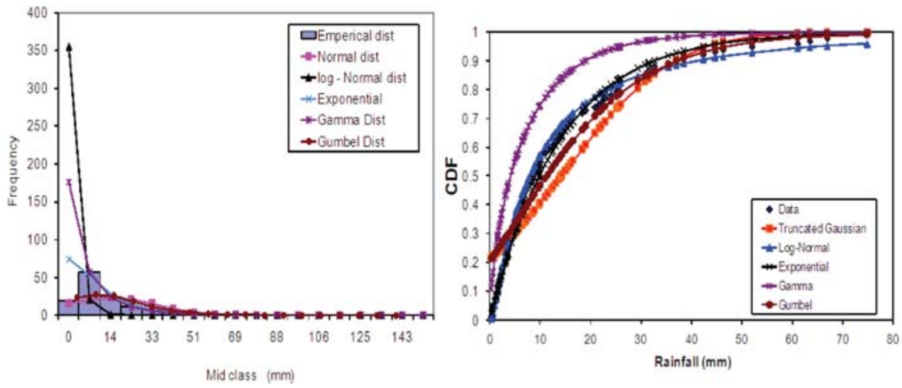

Fig. 3: Fitting a histogram (Left) and probability distribution function (right) to the non-zero monthly rainfall depth at station Amlog

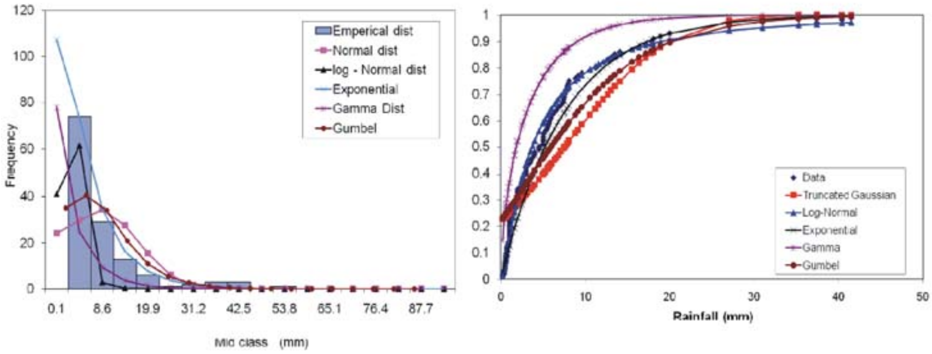

Fig. 4: Fitting a histogram (left) and probability distribution function (right) to the non-zero monthly rainfall depth at station Tabouk TB001.

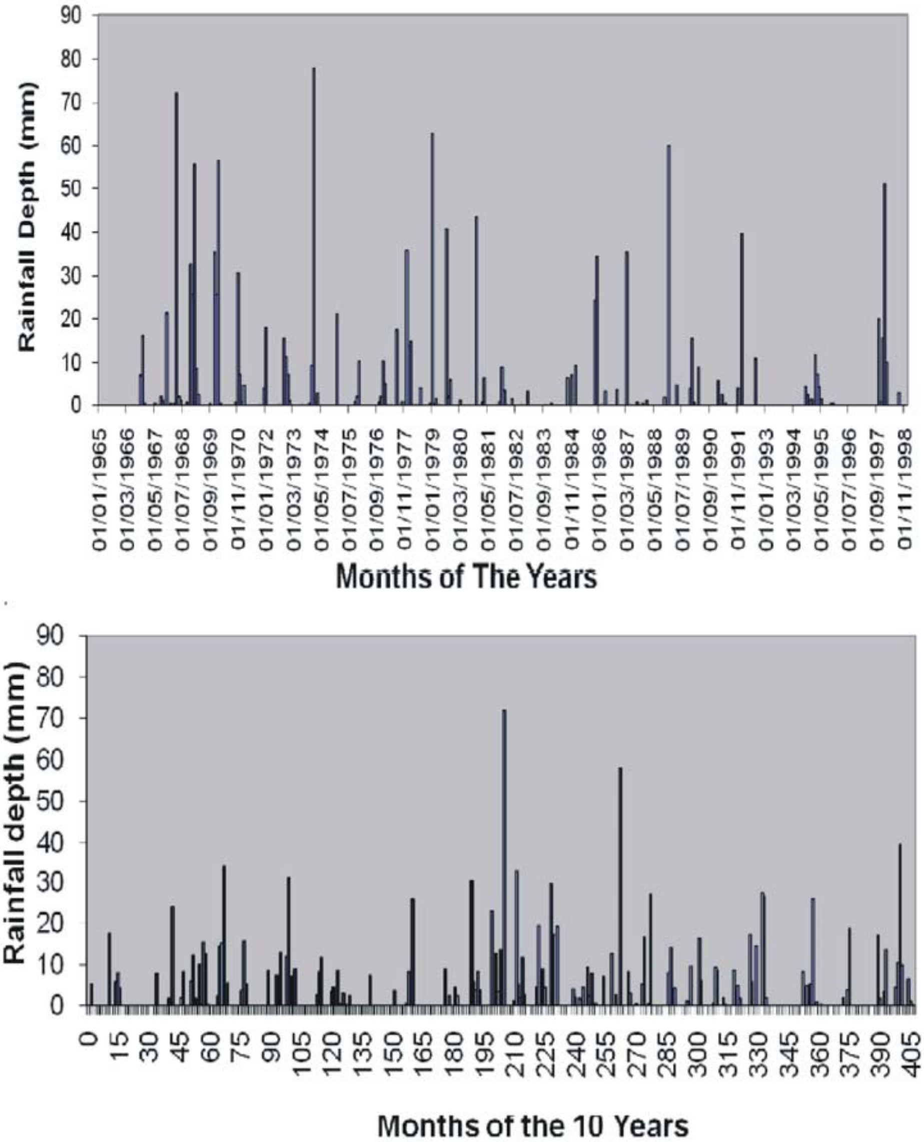

Fig. 5: Comparison between data (top) and simulation (bottom) of monthly rainfall depth at station Khules.

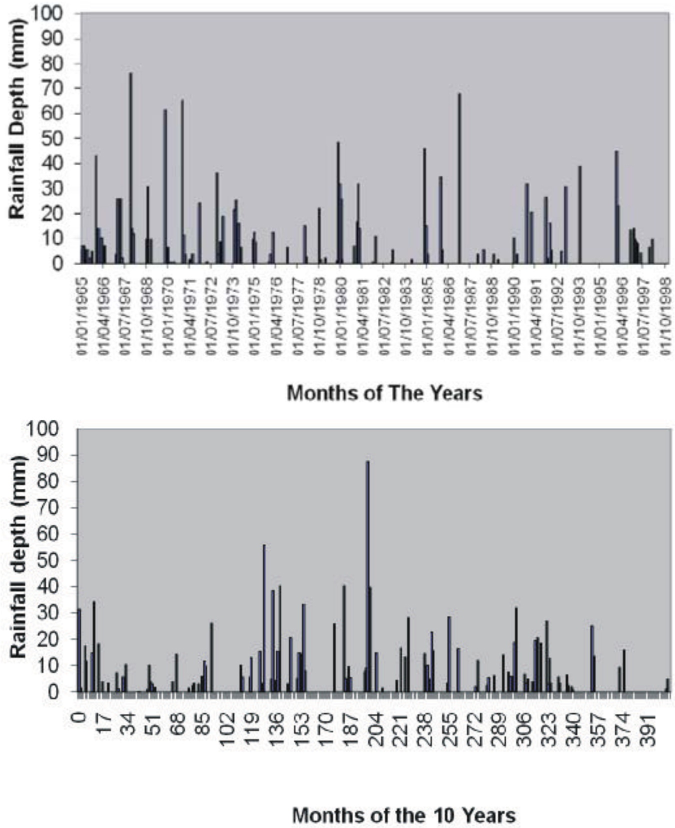

Fig. 6: Comparison between data (top) and simulation (bottom) of monthly rainfall depth at station Amlog

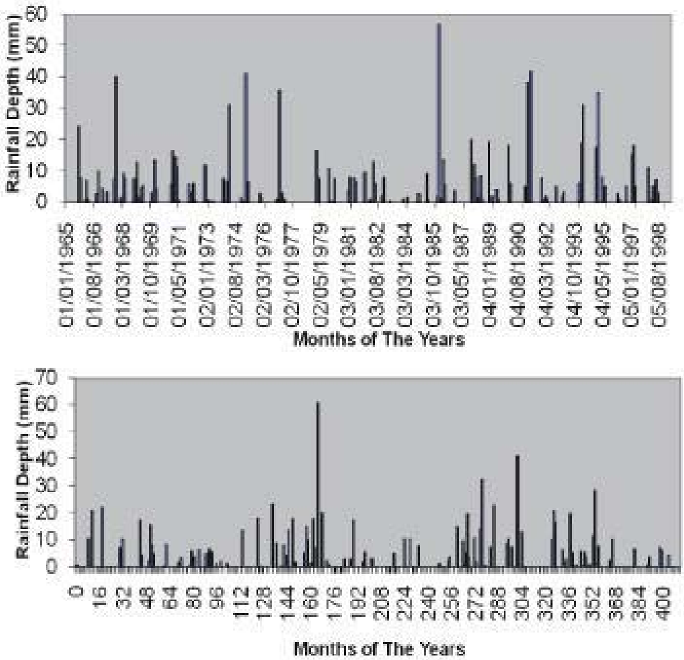

Fig. 7: Comparison between data (top) and simulation (bottom) of monthly rainfall depth at station Tabouk

Table 2: Monthly transition probabilities for station Khules, Amlog and Tabouk

Khules Station	Wet	Dry
Wet	0.54	0.46
Dry	0.22	0.78
Amolg Station	Wet	Dry
Wet	0.43	0.57
Dry	0.18	0.82
Tabouk Station	Wet	Dry
Wet	0.46	0.54
Dry	0.26	0.74

Table 1. Fig 2, 3 and 4 show both the absolute frequency histogram for both grouped data and the corresponding expected frequency of the probability density function considered in the analysis in the left part of the figure. While, in the right part of the figure, the empirical probability distribution function of the ungrouped non-zero monthly rainfall depth and its theoretical ones are presented. It is concluded that the Log-normal distribution seems to fit the three stations and also exponential distribution is suitable for Amolg and Tabouk stations.

Figure 5, 6 and 7 show comparison of the time series data and the simulations for next 10 years based on the statistics presented in Table 1 and 2 for assuming exponential distribution. The results show agreement in the global sense from visual detection of the figures and from the range of values generated and observed in the data, however, a quantitative measure is need in the future to make the final judgement on the model.

CONCLUSIONS

The models developed in this study seem promising in the analysis of rainfall in arid and semi arid zones. A rigorous data collection scheme is need to get reliable data for such analysis. The models need some improvements to model seasonality in the data, improved parameter estimation techniques, etc. The authors are considering these issues in their future studies.

REFERENCES

1. Plate, E.J., 1977. Intermetant processes and conditional sampling. Stochastic processes in water resources engneering, proc. of the second international IAHR Symposioum on stochastic hydraulics, Edited by Gottschalk, L., Lindh, G. and De Mare, L., August, 1976. Water Resources publications, pp. 1-27.
2. Box, G.E. and G. Jenkins, 1970. Time Series Analysis, Forecasting and Control, SF, Holden-Day.
3. Yevjevich, V., 1987. Stochastic Models in Hydrology, Stochastic Hydrology and Hydraulics. Springer-Verlag, 1: 1.
4. Salas, *et al*. 1980. Applied Modelling of Hydrological Time Series, Water Resources Publ., Litleton, Colorado.
5. Fernandez, B. and J.D. Salas, 1986. Periodic Gamma Autoregressive Processes for Operational Hydrology, Water Resources Res., 22: 1385-1396.
6. Gabriel, K.R. and Neumann, 1962. A Markov chain model for daily rainfall occurrence at Tel Aviv, Royal Metrological Society, Great Britain, 88: 90-95.
7. Sharma, T.C., 1996. Simulation of the Kenyan longest dry and wet spells and the largest rain-sums using Markov model. J. Hydrology, 178: 55-67.
8. Haan, C., 1977. Statistical Methods in Hydrology, The Iowa State University Press, pp: 378.
9. Krumbein, W.C., 1967. Fortran Computer Programme for Markov Chain Experiments in Geology. Computer Contribution 13, Kansas Geological Survey, Lawrence, Kansas, USA.
10. Elfeki, A.M.M. and F.M. Dekking, 2001. A Markov Chain Model for Subsurface Characterization: Theory and Applications. Mathematical Geol., 33(5): 569-589.