Proceedings of the Technical Sessions, 26 (2010) 25-32
Institute of Physics – Sri Lanka

Analysis of wet and dry behavior of weather through Markov models

P.M. Piyadasa and D.U.J. Sonnadara
Department of Physics, University of Colombo, Sri Lanka

ABSTRACT

This study attempts to model the daily rainfall climatology of Sri Lanka using Markov models at three selected weather stations where long term data records were available. Both the first and the second order Markov models were able to forecast the occurrence of daily rainfall to an accuracy of ~72±4%. The accuracy of predicting wet days during the wet season was very much higher than in the dry season, while predicting dry days during the dry season was very much higher than in the wet season. The mean number of wet spells per month and mean length of a wet spell per month has been forecasted using standard probability distributions combined with transition probabilities based on the first order Markov model. The deviation of the simulated values from the observed values was lower than the statistical variations which indicate that the model is suitable for simulating wet and dry spells in Sri Lanka.

1. INTRODUCTION

Sri Lanka is an island in the Indian Ocean, located about 31 km off the southern coast of India. The climate of Sri Lanka is considered to be warm and tropical. Its position between 5 and 10 North latitude has blessed the country with a warm climate moderated by ocean winds and considerable moisture. The mean temperature ranges from about 16 °C in the central highlands, to a maximum of 33 °C in low-altitude areas. There are two main climatic zones in Sri Lanka. The mountains and the south-western part of the country, known as the wet zone, receive ample rainfall (an annual average of 2500 mm), while most of the southeast, east, and northern parts of the country comprising the dry zone, receives less rainfall (between 1200 to 1900 mm annually). When considering the monthly distribution of rainfall, the year can be divided in to four seasons, namely the two monsoons (South-West from May to September and North-East from December to February) and the two inter monsoons. Since Sri Lanka's economy depends on agriculture production, understanding wet and dry behaviour is important to enhance the agricultural productivity and has wide range of applications in many parts of the country.

Markov chains have been used to predict and simulate wet and dry patterns in this study. Some researchers have already studied the applicability of the Markov chain to predict rainfall in Sri Lanka [1-3]. Most of these studies have used probability distributions to estimate rainfall amount. The reliability of predicting the amount first depends on the accuracy of the prediction of wet and dry days. However, results of published work have not focused on showing how accurately the models predict wet and dry days for different geographical regions and the intra-annual variability of predictions. In addition, rather than daily behaviour, the published works have concentrated on developing models for monthly or weekly rainfall averages which are more stable. A recent work has reported on the accuracy and reliability of Markov models on daily rainfall in Sri Lanka [4] with promising results promoting further investigation in this area.

The main objective of this research work is to use the Markov method in order to study the wet and dry behaviour of weather based on daily precipitation. Three weather stations having long series of daily rainfall data records have been selected in this study are Badulla, Colombo and Nuwara Eliya.

2. METHODOLOGY AND IMPLEMENTATION

2.1 Data sample

The data used in this research work are the daily rainfall measurements from three meteorology stations maintained by the Meteorology Department of Sri Lanka. The Meteorology Department of Sri Lanka currently maintains 22 main meteorological stations, 42 agricultural stations and over 350 rainfall stations throughout the country. Their database of daily rainfall data contains over hundred years of data which dates back to year 1870. In order to develop the models, a wet day was defined as a day with the total rainfall exceeding 0.25 mm and a dry day was defined as a day with the total rainfall lower than 0.25 mm.

2.2 First order Markov model

Transition probabilities (P_{ij}) of the first order Markov model is defined as;

$$P_{ij}(t) = P\{X_t = j \mid X_{t-1} = i\} \qquad i, j = 0,1$$

where P_{ij} is the probability of system being in the state j at time t from a previous state i at time t-1. For n state Markov chain n number of different values representing each state is required. Since P_{ij} represent transition probabilities only values in the interval [0, 1] are valid. Because those probabilities are conditional probabilities, for any fixed i value P_{ij} should add up to unity.

With these transition probabilities P_{ij}, a 2×2 matrix $P = \{P_{ij}\}^T$, called the transition matrix of the Markov chain can be formed where the sum of the entries in each column is 1.

$$P = \begin{bmatrix} P_{00} & P_{10} \\ P_{01} & P_{11} \end{bmatrix}$$

Probability vector $p^{(n)}$ represents the probabilities of system being in any state at n^{th} observation and its element $p_i^{(n)}$ is the probability that the system is in the state i at n^{th} observation and, $p^{(0)}$ is the initial probability vector (at t=0). If P is the transition matrix and $p^{(n)}$ is the state vector at the n^{th} observation, one can write p^{n+1} as follows

$$p^{n+1} = p^{(n)}P$$

By considering the initial probability vector, state vector at the n^{th} observation can be written as

$$p^n = p^{(0)}P^n$$

Matrix notation for the above equation for the first order Markov model is as follows.

$$\begin{pmatrix} p_0^{(n)} & p_1^{(n)} \end{pmatrix} = \begin{pmatrix} p_0^{(0)} & p_1^{(0)} \end{pmatrix} \begin{bmatrix} P_{00} & P_{10} \\ P_{01} & P_{11} \end{bmatrix}^n$$

Transition probabilities are considered to be constant when developing the Markov models. Past rainfall data from 1900 to 1979 have been used in constructing the model and data from 1980 to 1999 have been used in testing the model.

2.3 Second order Markov model

In second order Markov models present state of the system depends on the two previous states of the system. Now, instead of using j and i to represent present state and the previous state k, j and i has been used to represent the present state and two previous states respectively. Transition probabilities of the second order Markov model can be written as

$$P_{ijk}(t) = P\{X_t = k \mid X_{t-1} = j, X_{t-2} = i\} \quad \text{i, j, k=0,1}$$

Since present state depends on two previous states, in the second order Markov model transition probability matrix is a 4×4 matrix given by;

$$P = \begin{bmatrix} P_{000} & P_{001} & 0 & 0 \\ 0 & 0 & P_{010} & P_{011} \\ P_{100} & P_{101} & 0 & 0 \\ 0 & 0 & P_{110} & P_{111} \end{bmatrix}$$

Since initial probability matrix (1×4) can be defined as the case of first order Markov model, n^{th} state probability matrix can be evaluated by,

$$\begin{pmatrix} P_{00}^{(n)} & P_{01}^{(n)} & P_{10}^{(n)} & P_{11}^{(n)} \end{pmatrix} = \begin{pmatrix} P_{00}^{(0)} & P_{01}^{(0)} & P_{10}^{(0)} & P_{11}^{(0)} \end{pmatrix} \begin{bmatrix} P_{000} & P_{001} & 0 & 0 \\ 0 & 0 & P_{010} & P_{011} \\ P_{100} & P_{101} & 0 & 0 \\ 0 & 0 & P_{110} & P_{111} \end{bmatrix}$$

Same data ranges have been used in constructing and testing the 2^{nd} order model.

2.4 Simulating wet and dry spells

In order to simulate the number of wet spells and the length of wet spells, a method which was successfully applied in Kenya [5] which was based on the conditional probabilities of the 1^{st} order Markov models was adopted here.

Let pp=P(W/W) be the probability that a wet day followed by a wet day and qq=P(D/D) be the probability that a dry day followed by a dry day. Since we are interested in uninterrupted events of wet and dry days, transition probability matrix T1 for Markov process can be represented by;

$$T1 = \begin{bmatrix} pp & 1-pp \\ 1-qq & qq \end{bmatrix}$$

For completely random rainfall events, transition probability matrix changes to

$$T2 = \begin{bmatrix} p & 1-p \\ 1-q & q \end{bmatrix}$$

where p = P(W) and q = P(D). The equivalence of matrices T1 and the T2 will determine whether the rainfall events are random or not [5].

Let us denote the number of wet spells by N_w and length of such spells in days by L_w. To simulate the expected number of wet spells and mean length of the wet spells, assume that the occurrence of wet spells can be modeled using Poisson distribution and the length of the wet spell can be modeled using geometric distribution. Then, the probability that the number of wet spells has a value *i* is given by [5];

$$P(N_w = i) = \frac{\exp\{-np(1-pp)\}[np(1-pp)]^i}{i!}$$

where *n* is the number of days in the analysis period (for January *n* = 31 days).

The probability that the length of a wet spell is less than or equal to a given value j is represented by;

$$P(L_w \leq j) = 1 - pp^{j-1}$$

Hence;

$$P(L_w > j) = pp^{j-1}$$

The mean number of wet spells and mean length of wet spells for the first order Markov process is given by [5];

$$E(N_w) = np(1-pp)$$

$$E(L_w) = \frac{1}{1-pp}$$

3. RESULTS AND DISCUSSION

3.1 Variation of transition probabilities

Figure 1a shows the transition probabilities P(W/W) and P(D/D) calculated using 80 years of daily rainfall records for the Badulla station. Other two transition probabilities P(W/D) and P(D/W) can be derived from these two probabilities. From the Figure 1a it is clear that there is a considerable variation of transition probabilities throughout the year. This variation is a result of seasonal variation in rainfall. The low probability areas in the middle part of the P(W/W) curve corresponds to the Southwest monsoon period. The line through the data represents the polynomial fitted to actual probability values and the dots represent the five day mean of the probability values. Five day means have been taken in order to reduce the fluctuations in the probability values. The Colombo showed two clear peaks in P(W/W) curve corresponding to the Southwest and Northeast monsoon seasons. Nuwara Eliya showed similar behaviour to Colombo although the peaks were relatively less pronounced than Colombo station.

There is a considerable variation among the transition probabilities in the second order Markov model. These variations are also a result of seasonal variation of rainfall. Figure 1b shows the variation of the transition probabilities P(W/WW), P(W/WD), P(W/DW) and P(W/DD) calculated using 80 years of daily rainfall records obtained from the Badulla station. Other four transition probabilities P(D/WW), P(D/WD), P(D/DW) and P(D/DD) can be derived from the given four transition probabilities. The lines represent the polynomial fitted to actual probability values and dots represent the five day mean of the probability values. Five day mean was taken in order to reduce the fluctuations in the probability values. Similar fluctuations were seen for Colombo and Nuwara Eliya stations. The persistency in rainfall can be clearly seen in the transition probabilities. The occurrence of wet days with high probability can be observed when previous two days are wet days and the occurrence of wet days with low probability can be observed when previous two days are dry days.

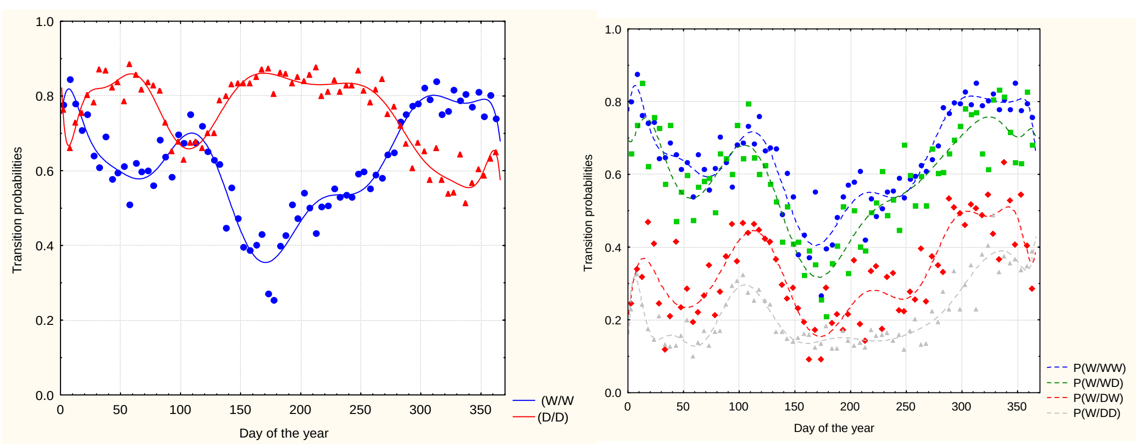

Figure 1: Transition probabilities at Badulla (a) First order (b) Second order

3.2 Variation of model predictions with year

When the accuracy of predictions over 20 year period (1980 - 1999) was compared it was seen that the accuracy between years is quite consistent. In general, for Colombo station, accuracy is lower than the other two stations. This indicates a high variability in daily rainfall patterns for Colombo with no prolonged wet or dry conditions. For the first order model, the maximum accuracy of 78% was seen for Badulla in the year 1982 while the maximum accuracy of 77% was observed for Nuwara Eliya in the year 1992. The accuracy of Markov model applied for Colombo in the year 1994 was rather low due to missing rainfall data during the years 1992 and 1993. For the first order model, with 20 years of data, the mean accuracy values for Colombo, Nuwara Eliya and Badulla were

70±3%, 74±3% and 73±3% respectively. The same for 2nd order model were 68±3%, 73±3% and 73±3% respectively. Since the accuracies are similar, there is no advantage in developing higher order Markov models.

3.3 Accuracy of predicting wet and dry days

In Figure 2a the accuracy of the Markov model is shown separately for wet and dry days for Badulla weather station. The accuracy of predicting wet days are considerably lower than the accuracy of predicting dry days (23% difference). Badulla is situated in an intermediate zone (neither wet nor dry) and it has more dry days than wet days. Since Markov model depends on the probabilities, having long spells of dry days make the model to predict dry days accurately than wet days. Even in Colombo accuracy of predicting dry days are higher than the accuracy of predicting wet days. Although Colombo is located in the wet zone of the country, the number of dry days is usually higher than the number of wet days resulting in higher accuracy of predicting dry days than wet days. However, the difference between two accuracies in Colombo is lower than that of Badulla (12% difference). For Nuwara Eliya the accuracies of predicting both wet and dry days are nearly equal over the past 20 years. Because of the wet weather in Nuwara Eliya, the probability of having wet days and the probability of having dry days are nearly equal. No significant difference was observed between the accuracies between the first and the second order Markov models.

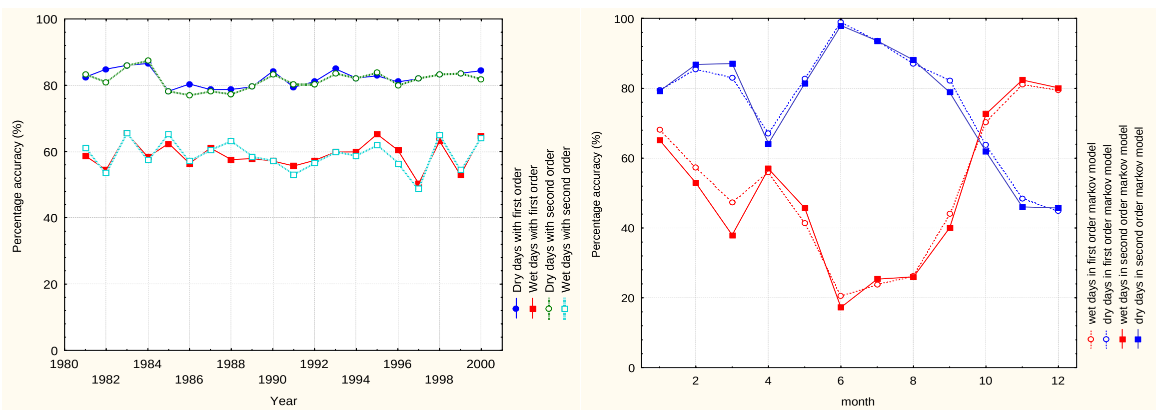

Figure 2: Variation of wet and dry days in Badulla (a) Annual (b) Monthly.

3.4 Variation of the accuracy with month

From the Figure 2b it can be seen that the accuracy of predicting dry days in June at Badulla is almost 100% while the accuracy of predicting wet days in that month is less than 20%. This is a result of having very long dry spells in June. In between those dry spells there are short occasional wet spells which are hard to predict through the Markov model. After June, as the rain begins, the accuracy of predicting wet days has increased gradually while the accuracy of predicting dry days has decreased. In Badulla, the maximum accuracy of predicting wet days have been recorded at the end of the year reflecting the effect of North-East monsoonal rains. For Colombo, the maximum accuracy of predicting dry days was seen at the beginning of the year because of the less influence of the 1st inter-monsoon rainfall. Two peaks were observed representing two major rainy seasons in the Colombo station where the first one was obtained for the April thunder storms combined with South-West monsoon and the second one for the December showers with the North-East monsoon. For Nuwara Eliya station two peaks representing the major rainy seasons were also seen, but the peaks were not significant compared to the Colombo station. This is due to continuous rainfall throughout the year in Nuwara Eliya. The accuracy of the second order Markov model was similar to the accuracy of the first order Markov model in all three weather stations.

3.5 Simulation of the number of wet spells

Figure 3a shows the number of simulated wet spells and observed wet spells per month at Badulla. Values given in the figure are the mean values of simulated and observed wet spells from 1980 to 1999. Two peaks in the figure 3a represent the wet seasons for Badulla. These two peaks were also prominent in Colombo and Nuwara Eliya stations as well. Solid line indicates the observed values with their standard deviation indicated in error bars while points slightly shifted to the right indicate the simulated values with their standard deviations. Since most of the simulated values are within the range of the errors of the observed values it can be concluded that the model can predict the wet spells accurately. The long error bars representing the standard deviation of the observed values indicate that the variance of the number of wet spells in each month is quite large. Since the simulated values were calculated using conditional probabilities for the past 20 years time period which does not fluctuate very much from year to year, simulated values have less fluctuation which is clearly shown by lower standard deviation values in Figure 3a. Compared to other two stations, Colombo showed higher fluctuation in the observed number of wet spells. Simulated wet spells in Nuwara Eliya showed the best agreement with observed values compared to other two stations.

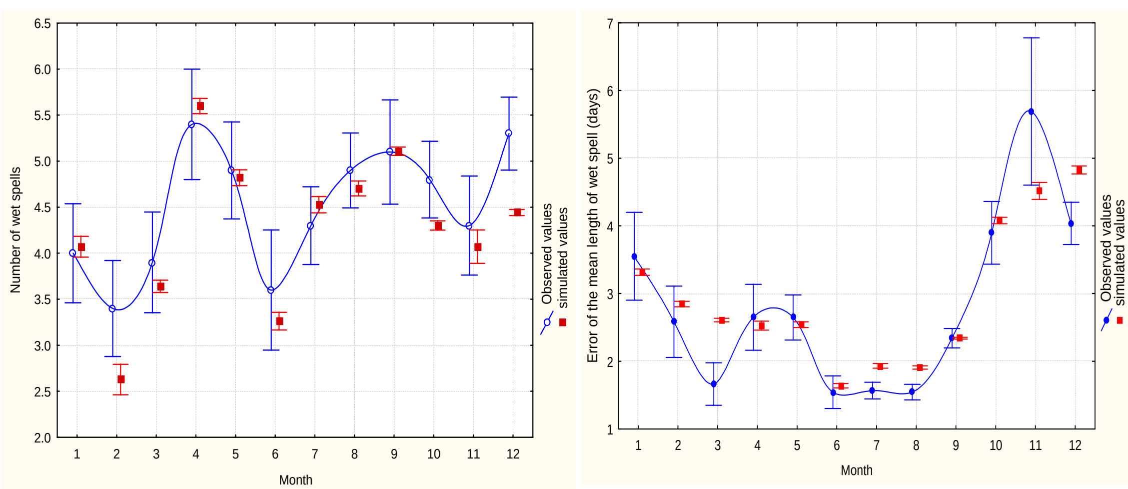

Figure 3: Simulated and observed wet spells in Badulla (a) Number (b) Length

3.6 Simulation of mean length of a wet spell

Figure 3b shows the simulated and observed mean length of wet spells in Badulla. Except for few months, a reasonable conformity was seen between the simulated and observed values. The mean length of wet spells for simulated and observed values were calculated using the data from 1980 to 1999. The data shows observed values having much larger standard deviation compared to simulated values. This is due to the larger variance in the observed mean length of a wet spell during each month. The standard deviations of the simulated values are much smaller since the calculation was based on the parameters of conditional probabilities for last 20 years which are quite stable. Unusually large error was seen in November indicating higher variation in the length of wet spells in Badulla for this month. Colombo had relatively larger fluctuation for mean length of wet spell compared to other two stations. In all stations two peaks were observed representing two wet seasons.

4. CONCLUSIONS

Two types of analysis have been carried out in this study with the aim of modelling wet and dry behaviour of Sri Lanka. In the first part of this study, first and second order Markov models have been developed to simulate the occurrence of the daily rainfall. From the results it can be concluded that the use of higher order Markov models do not increase the accuracy of the predictions. When the model efficiency was compared between the wet days and dry days separately, it was seen that the accuracy of the model in predicting dry days was higher than the accuracy in predicting wet days for Badulla and Colombo stations. For Nuwara Eliya station both accuracies were nearly equal. This is due to less variation in the intra annual rainfall at Nuwara Eliya. By taking accuracy of the Markov model for wet and dry days separately it was seen that the accuracy of predicting wet days in wet season is higher than the accuracy of predicting wet days in dry season. Similarly, the accuracy of predicting dry days in dry season is higher than the accuracy of predicting dry days in wet season.

The second part of this study focused on simulating the number of wet spells per month and the mean length of a wet spell per month using Markov models. Transition probabilities and standard probability distributions have been used in both analyses. The variability of the observed values between years was considerably high making it difficult to simulate. However, the general trends in mean values are well reproduced by the simulated values. The deviation of the simulated values from the observed values was lower than the expected errors. This indicates that the Markov models are suitable for simulating wet and dry spells in Sri Lanka.

Acknowledgements: Financial assistance given by the National Research Council, Sri Lanka for the research grant number NRC 06-18 is greatly acknowledged.

5. REFERENCES

1. Fernando, P.N.P. and Senevirathne, K., 1987: A Markov chain model for daily rainfall data, *Proceedings of the Sri Lanka Association of Advancement of Science*, pp 171.
2. Sooriyarachchi M.R., 1989: Estimation of probable rainfall for irrigation and water management in the tropics, *M.Sc. Thesis*, University of Reading, UK.
3. Punyawardena, B.V.R. and Kulasiri, D., 1998: Spatial interpolation of rainfall in the dry zone of Sri Lanka*, J. Natn. Sci. Coun. Sri Lanka* 26(3): 247-262.
4. Perera H.K.W.I., Sonnadara D.U.J., and Jayewardene, D.R., 2002: Forecasting the Occurrence of Rainfall in Selected Weather Stations in the Wet and Dry Zones of Sri Lanka. *Sri Lankan Journal of Physics, 3*, 39-52.
5. Ochola W.O. and Kerkides P. (2003): A Markov chain simulation model for predicting critical wet and dry spells in Kenya. *Irrigation and Dranage,* 52, 327-342