



INSTITUTE OF PHYSICS - SRI LANKA

Forecasting the Occurrence of Rainfall in Selected Weather Stations in the Wet and Dry Zones of Sri Lanka

H.K.W.I. Perera¹, D.U.J. Sonnadara^{1,*} and D.R. Jayewardene²

¹ *Department of Physics, University of Colombo, Colombo 3, Sri Lanka.*

² *Department of Mathematics, University of Colombo, Colombo 3, Sri Lanka.*

Abstract

Two models were developed based on the Markov method to predict the occurrence of rainfall. Daily rainfall data from 9 meteorology stations in Sri Lanka namely, Colombo, Ratnapura, Kandy, Galle, Hambanthota, Batticaloa, Anuradhapura, Badulla and Trincomalee were analysed using these models. The results show that the models can forecast the status of a given day at an average of 73%. Higher agreement is seen for stations in the dry zone of Sri Lanka when compared with those in the wet zone. No significant difference was observed between the results obtained using the 1st order Markov process and the 2nd order Markov process. The models were not sensitive to the range of data used or the short-term fluctuations.

1. INTRODUCTION

The estimation and the prediction of rainfall amounts are problems of fundamental importance in many applications in agriculture, hydrology and ecology. In a tropical country such as Sri Lanka, rainfall variability is a strong limiting factor in food production and hydropower generation¹.

Over the last 40 years several models have been developed with increasing degrees of sophistication to capture the underlying physical dynamics that govern rainfall¹⁻⁸. All these models can be classified in to three types, namely, empirical statistical models, dynamic models and intermediate stochastic models². These classifications are based on the amount of physical realism incorporated into the model structure.

* Corresponding author (Tel: 584777, E-mail: upul@phys.cmb.ac.lk)

In the empirical case there is no attempt to incorporate the physical aspects of the atmosphere but fit empirical stochastic models to the available data. Models of the second type are purely physically based models. The third type is a combination of both methods and certain physical processes of rainfall structure such as rain cells, rain bands and cell structures can be described with a stochastic approach.

Though there are new approaches, stochastic models of rainfall are still being used for agricultural and meteorological purposes to generate forecasts and, in simulation. In this paper the occurrence of rain in Sri Lanka for stations in the wet and dry zones have been modelled using the first and second order 2-state Markov chains. The prediction skills of the Markov models were evaluated using the observed data that were not used in the estimation of the parameters of the model.

Rainfall forecasting in early works basically concentrated on modeling two distinct components of the rainfall time series. They are the occurrence of rainfall, namely wet and dry sequences, and the amounts on the wet days. The reliability of predicting the amount first depends on the accuracy of the prediction of wet and dry days. However, results of work published so far lack in showing how accurately the models predict wet and dry days for any given area or any given year²⁻⁸. Very often the work is limited to applying the techniques only to data from one station or irrigation site or showing how well it fits by calculating confidence limits for a given year or two.

In addition, the published works concentrate on developing models for monthly or weekly rainfall aggregates or averages. To our knowledge no work has been reported on the accuracy of Markov models on daily precipitation. Although fluctuations in daily precipitation are very high, it is one of the intentions of the present work to investigate the reliability of predicting daily precipitation using Markov models.

The present study is aimed at examining in detail the ability of Markov models to forecast the wet or dry condition of a given day. The main objectives of the present study are,

1. To identify whether there is any significant improvement in the accuracy of forecasts by using higher order Markov chains for local weather stations
2. To investigate the accuracy of the models in modelling stations in different weather zones
3. To investigate the effect of the range of the historical rainfall series on the final result
4. To identify whether there is any advantage in fitting functions such as polynomials to transitional probabilities
5. To estimate the predictive power and the stability of the developed models using efficiencies and errors of forecasts.

2. METHODOLOGY

2.1 Data Sample:

In Sri Lanka, the measurement of daily rainfall is mainly carried out at weather stations maintained by the Department of Meteorology. They currently maintain 22 main meteorological stations, 42 agricultural stations and over 350 rainfall stations throughout the country. They have access to satellite images transmitted by NOAA 12, 14, 15 and GMS 5 weather satellites. In addition, they receive weather reports from ships with the help of the telecommunication department. The history of meteorological observations in the form of rainfall measurements dates back to year 1850.

The data used in this study are daily rainfall measurements from 9 meteorology stations maintained by the Department of Meteorology. The geographical locations of these stations are shown in figure 1.

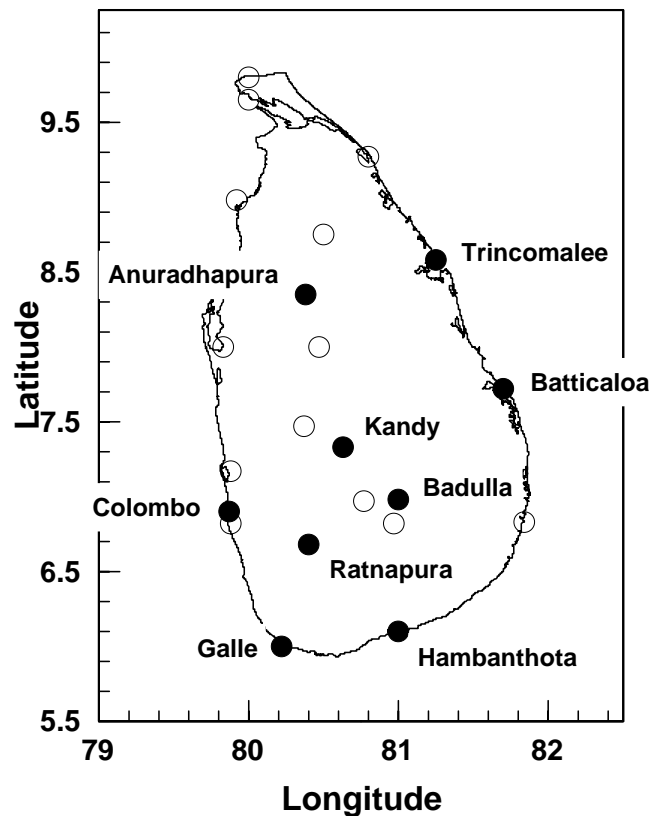


Fig. 1: The 24-hour weather stations maintained by the Department of Meteorology in Sri Lanka. The stations selected for this work are marked as closed circles.

In table 1, the exact locations of these stations together with the altitude and the range of data used in the study are given. The stations Colombo, Galle, Hambanthota, Ratnapura and Kandy belong to the wet zone and, Anuradhapura,

Trincomalee and Batticaloa belong to the dry zone. The station in Badulla belongs to the inter-mediate zone.

Table 1. Data used in this work and information of the meteorological stations.

Weather Station	Range	Longitude	Latitude	Altitude
Colombo	1900-1999	79.87 E	6.09 N	7 m
Badulla	1900-1999	81.05 E	6.98 N	670 m
Anuradhapura	1970-1999	80.38 E	8.35 N	93 m
Trincomalee	1970-1999	81.25 E	8.75 N	3 m
Rathnapura	1990-1999	80.04 E	6.68 N	34 m
Batticaloa	1990-1999	81.07 E	7.72 N	3 m
Galle	1990-1999	80.22 E	6.03 N	13 m
Hambanthota	1990-1999	81.13 E	6.12 N	16 m
Kandy	1990-1999	80.63 E	7.33 N	477 m

Sri Lanka is an island in the Indian Ocean at the southern most tip of India. It lies between 5° and 10° north latitude and has a tropical climate. Sri Lanka can basically be divided into two climatic divisions. On the west is the wet zone where the annual rainfall is over 2000mm. The other area is the dry zone. These regions can be further divided into four climatic zones, namely, wet zone low land, dry zone low land, hill country wet zone and hill country dry zone.

Table 2. Rainfall received during the 4 seasons over a 10-year period (1990-1999) as a percentage of the total annual rainfall

Weather Station	1st inter-monsoon Mar-Apr	South-West Monsoon May-Sep	2nd inter-monsoon Oct-Nov	North-East Monsoon Dec-Feb
Colombo	13.0	45.7	29.9	11.4
Galle	11.9	49.2	26.1	12.8
Rathnapura	14.1	49.9	24.2	11.7
Kandy	11.6	34.4	32.9	21.2
Anuradhapura	17.1	17.1	41.3	24.6
Trincomalee	3.9	25.2	35.9	35.0
Batticaloa	5.0	17.7	34.4	42.9
Hambanthota	10.3	30.4	35.9	23.4
Badulla	12.9	24.8	29.2	33.2

In Sri Lanka, a year can be divided into four rainfall seasons. They are, the 1st inter-monsoon from March to April, the Southwest monsoon from May to September, the 2nd inter-monsoon from October to November and the Northeast monsoon from December to February. During inter-monsoons, rainfall is generally distributed throughout the island with the heaviest rainfall occurring in the central part of the island and diminishing as the coastal belts are reached. In table 2, the percentages of

rainfall received in each season (during 1990 to 1999) in each of the weather stations are shown. The same data is graphically represented in figure 2.

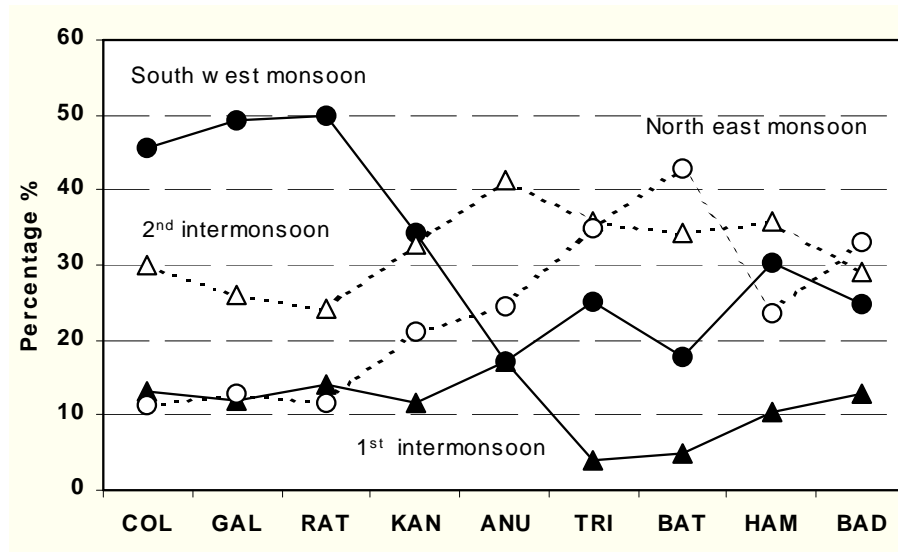


Fig. 2: Rainfall received during the 4 seasons over a 10-year period (1990-1999) as a percentage of the total annual rainfall.

2.2 Review of Markov Chains:

A Markov chain can be defined as a type of time ordered probabilistic process that goes from one state to another according to the probabilistic transition rules that are determined by the current state only. That is, the probability of a day being in a certain state (either wet or dry) is conditioned on the states of the previous periods, where the number of previous periods is termed as the order of the chain.

In the first-order, two-state Markov model, the current state is dependent solely on the state of the previous period, while in the second-order, two-state Markov chains the current state is determined by the states of the two previous periods⁶.

When recording whether a measurable amount of rain has occurred over time at a particular location (2-state time series), data will be in the form of a discrete state series. The most commonly used model for discrete state series is a low order stationary Markov chain. Such models are used to predict the occurrence of certain sequences.

The states of the system are dry and wet. A wet state is defined as a 24 hour period from 8.30 a.m. with the total amount of rainfall exceeding some threshold amount. Otherwise the state is taken as dry. The discrete state series of rainfall can be represented as $X_1, X_2, X_3, \dots, X_t$ for a t length sequence, where

$$\begin{aligned}
 X_t &= 0, \text{ if day } t \text{ is dry} \\
 X_t &= 1, \text{ if day } t \text{ is wet}
 \end{aligned}$$

2.2.1 First order Markov model:

The probabilities of a first order Markov chain are defined as $P\{X_t = j \mid X_{t-1} = i\}$ $i, j = 0, 1$

The *transition probability* P_{ij} , (where $i, j = 0, 1$) is the probability that if the system is in state i at any one observation, it will be in state j at the next observation. As probabilities, the numbers P_{ij} must all lie in the interval $[0, 1]$. Transitional probabilities for the 1st order can be expressed as follows.

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

For any fixed i , $\{i = 0, 1\}$ we must have, $P_{i0} + P_{i1} = 1$

This expresses the fact that if the system is in one of the states at one observation, it will with certainty be in one of the two states at the next observation. With these transition probabilities, a 2×2 matrix $P = \{P_{ij}\}^T$, called the transition matrix of the Markov process can be formed where the sum of the entries in each column of P is one.

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

The probability vectors $p^{(n)}$ for $n = 0, 1, 2, \dots$ are said to be the *state vectors* of a Markov process, where $p_i^{(n)}$ is the probability that the system is in the i^{th} state at the n^{th} observation. In particular, the state vector $p^{(0)}$ is called the *initial probability* or *initial state vector* of the Markov process. If P is the transition matrix and $p^{(n)}$ is the state vector at the n^{th} observation, one can write

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

where $p^{(n+1)}$ is the state vector at the $n+1^{\text{th}}$ observation⁹⁻¹². From this it follows that

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

i.e., the initial state vector $p^{(0)}$ and the transition matrix P determine the state vector $p^{(n)}$ at the n^{th} day. The n^{th} step transition probabilities are called *conditional probabilities* and are denoted by $P_{ij}^{(n)}$, (where $i, j = 0, 1$) with $P_{ij}^{(n)} \geq 0$, for $n=0, 1, 2, 3, \dots$ and

$$\sum_{j=0}^1 P_{ij}^{(n)} = 1$$

$p^{(n)} = p^{(0)} P$ in matrix notation for the 1st order Markov chain can be written as;

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

2.2.2 Second order Markov Model:

The transitional probabilities of the 2nd order Markov model can be expressed as

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

The transitional probabilities for the 2nd order model can be calculated by counting the wet days-satisfying dry or wet states during the previous 2 days. In other words a two-step or a second order transitional probability is the probability of going from state i to state k in two transitions. As in the 1st order process, P represents the 2nd order transition matrix¹¹.

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

The state vector $p^{(n)}$ can be defined by a process similar to the 1st order process. The initial probability matrix $p^{(0)}$ is a 1×4 row matrix.

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

Therefore, the future state vector $p^{(n)}$ can be evaluated if the initial state vector and the transition matrix are known.

2.3 Development of models:

Daily rainfall data from 9 meteorology stations were used in this modelling study. The period 1990-1999 was chosen in the evaluation of the model, but 100-year and 30-year periods were used to investigate the various aspects of the developed model.

A wet state is defined as a 24 hours period from 8.30 a.m. with the total rainfall exceeding 0.25 mm. One of the main assumptions in Markov chains is stationarity. That is $P_{ij}(t) = P_{ij}$ for $t_k < t < t_{k+T}$ where T is the time interval taken to calculate the transitional probabilities. This assumes that the chance of rain remains constant within T day periods. The data was grouped into 5 day, 7 day and 10-day groups and studied separately.

The transitional probabilities P_{ij} of the first order were calculated as

$$\begin{aligned} P_{00} &= P(D/D) \\ P_{01} &= P(W/D) = 1 - P_{00} \\ P_{10} &= P(D/W) \\ P_{11} &= P(W/W) = 1 - P_{10} \end{aligned}$$

For the second order model, the transition probability P_{ijk} were calculated as;

$$\begin{aligned} P_{000} &= P(D/DD) \\ P_{001} &= P(W/DD) = 1 - P_{000} \\ P_{010} &= P(D/WD) \\ P_{011} &= P(W/WD) = 1 - P_{010} \\ P_{100} &= P(D/DW) \\ P_{101} &= P(W/DW) = 1 - P_{100} \\ P_{110} &= P(D/WW) \\ P_{111} &= P(W/WW) = 1 - P_{110} \end{aligned}$$

where $P(W/D)$ represents the transitional probability of a day being dry and the following day being wet and $P(D/WD)$ represents the transitional probability of a day being dry, given that the previous two days are dry day followed by a wet day.

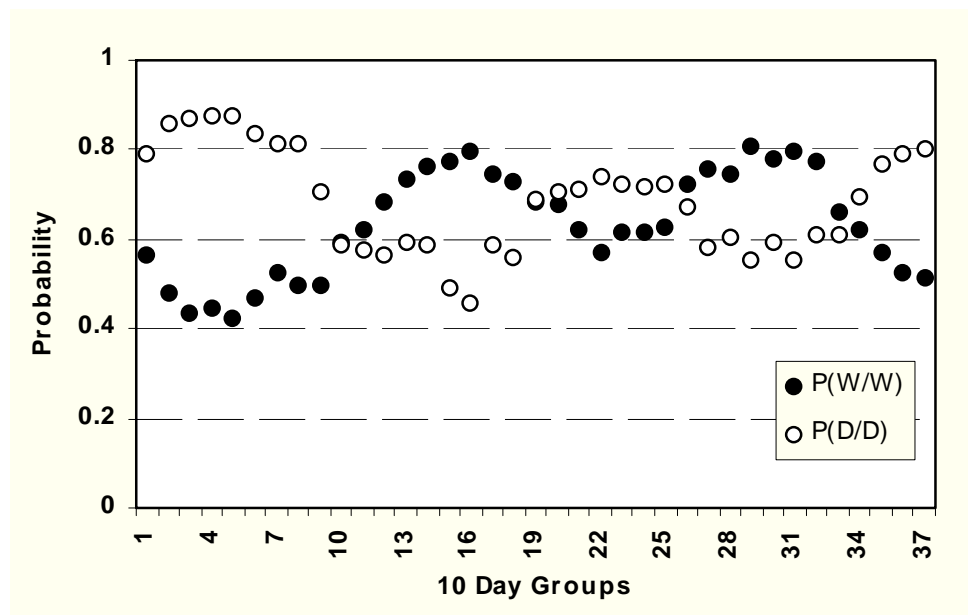


Fig. 3: Variation in the transition probabilities for the Colombo station with 100 years rainfall data.

As a result of seasonal variation in rainfall, the elements of the transitional matrices vary through out the year, the pattern of which depends on the geographical region of that particular station. Figure 3 shows the transition probabilities $P(W/W)$ and $P(D/D)$ calculated using 10 day periods with 100 years of rainfall records for the Colombo station. Since the other two probabilities $P(W/D)$ and $P(D/W)$ can be obtained from these, they are not shown in the figure to maintain clarity. The two distinct peaks in the figure for $P(W/W)$ correspond to the well-known Northeast and

Southwest monsoon rains in Sri Lanka in March-May and September-October respectively. Some authors have discussed fitting models to transition probability distributions to reduce the statistical fluctuations. This topic is addressed later on this paper.

In the present model, to reduce the statistical fluctuations, initial probability values were computed by taking the average of the past m days for $m=3,5,7$. In the present approach out of the 10 years data, 9 years data were used to estimate the transitional probabilities. The remaining year was considered as the observed year and it was used to compute the initial probabilities as well for forecast verifications.

3. RESULTS

In applying the Markov model, one has to determine the best combination of the period (T) and the number of days used to calculate the initial probability (m) for each station of interest, since the performance of the model for each station may vary depending on the values of these two parameters. In this analysis, for all stations, varying T among 5,7,10 and m among 3,5,7 and directly comparing the agreement between the prediction and the actual status first investigated the possible combinations of T and m .

Table 3 shows the best combinations obtained for T and m for both the first and the second order models for each station. In general, the preferred combination for most stations was $m=3$ and $T=10$. In the analysis that follows, the combinations given in table 3 were adopted for each station. Unless stated, the results obtained are for forecasting for the next day.

Table 3 Best combinations of the period (T) and the number of days used to calculate the initial probability (m)

Meteorology Station	1 st Order		2 nd Order	
	m	T	m	T
Colombo	3	10	3	10
Galle	3	10	3	10
Ratnapura	3	7	3	10
Kandy	3	7	3	10
Anuradhapura	3	5	3	10
Trincomalee	5	10	3	10
Batticaloa	3	10	3	10
Hambanthota	3	10	3	10
Badulla	3	10	5	7

Figure 4 shows the agreement between the forecast and the actual data for the 1st order model. The solid line represents the agreement for both wet and dry days. The dashed line shows the agreement for only wet days. The plotted values are the means and standard deviations of 10 years predictions for the period 1990-1999 for each station. In general, the overall prediction was better for stations in the dry zone.

As expected poor agreement is seen for stations in the dry zone when predicting only wet days. The overall agreement of the model is 73%. The highest agreement of 79% is seen for Anuradhapura and lowest figure of 63% is seen for Galle.

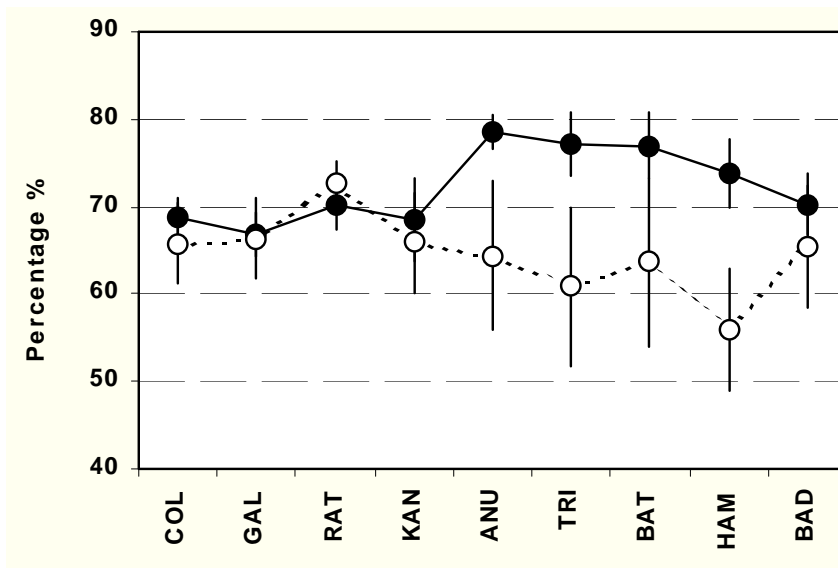


Fig. 4 The agreement between the forecast and the actual data for the 1st order model. (a) Solid line: agreement for both wet and dry days (b) Dashed line: agreement for wet days

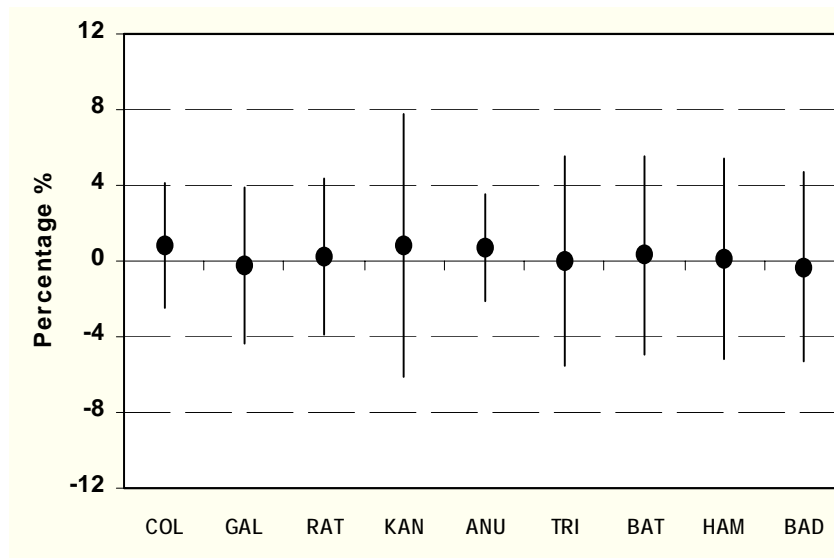


Fig. 5 Difference in agreement between the 1st and the 2nd order models for both wet and dry days.

In figure 5, the difference between the 1st and 2nd order models are shown. It can be seen that the overall difference in the prediction differs within 1%, which is well within the error of predictions. This result confirms that for local data there is no advantage in developing Markov models of higher order. Others have reported similar observations recently for stations in the dry zone ⁶.

In figure 6, the agreement for the number of days forecast is shown for the Colombo station. It can be seen that the efficiencies decrease rapidly up to the third

day (about a 5% drop) and then remains more or less constant. The same result was seen for all stations. The reason is the steady state achieved by the transition probabilities after the third day.

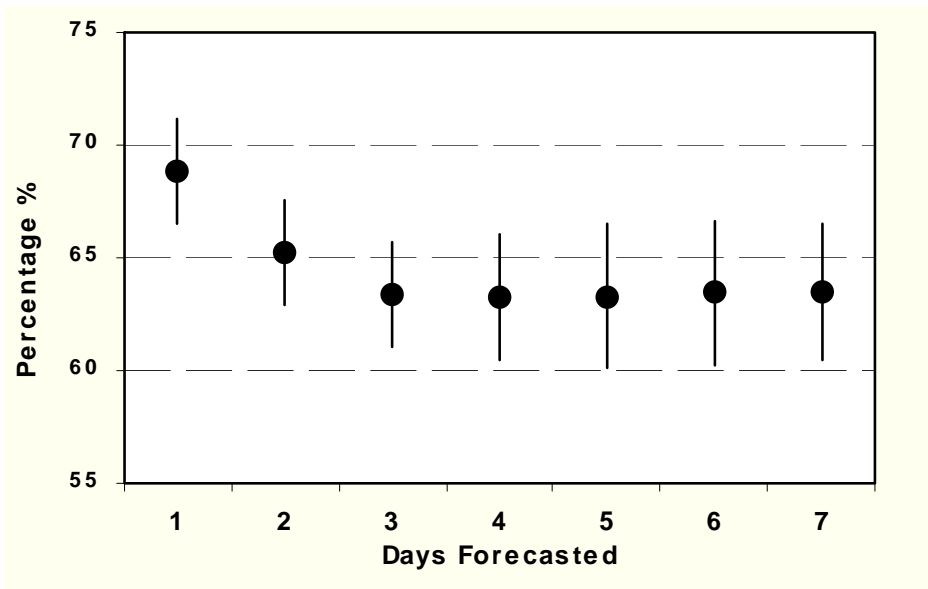


Fig. 6: Forecasting up to 7 days ahead.

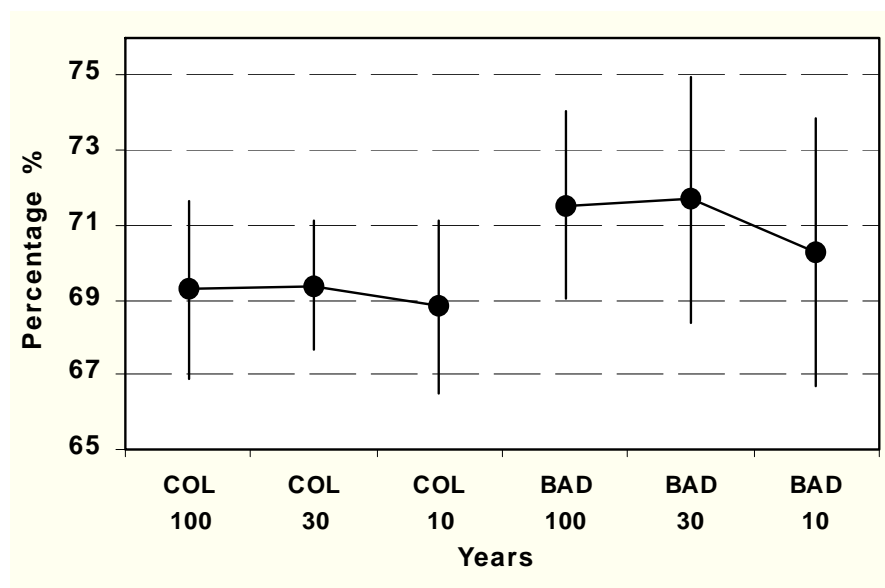


Fig. 7. Effect of data range.

One interesting question is whether the range of data has any effect on the model developed here. To study this effect, for the Colombo and Badulla stations, the data set was divided into 3 groups of 100 years, 30 years and 10 years and the model analysis was carried out in an identical manner. The results of this analysis shows (see Figure 7) that there is no evidence of any significant effect on the forecasts due to the data range.

Some authors³⁻⁵ has discussed fitting models to the transition probabilities to smoothen the statistical fluctuations and using the fitted values in the model. This was

carried out by fitting 2 separate 12 degree polynomials to $P(W/W)$ and $P(D/D)$ (see figure 8) and using the fitted results in the model. Figure 9 shows the difference in both methods for all stations. No significant difference was observed for any of the stations.

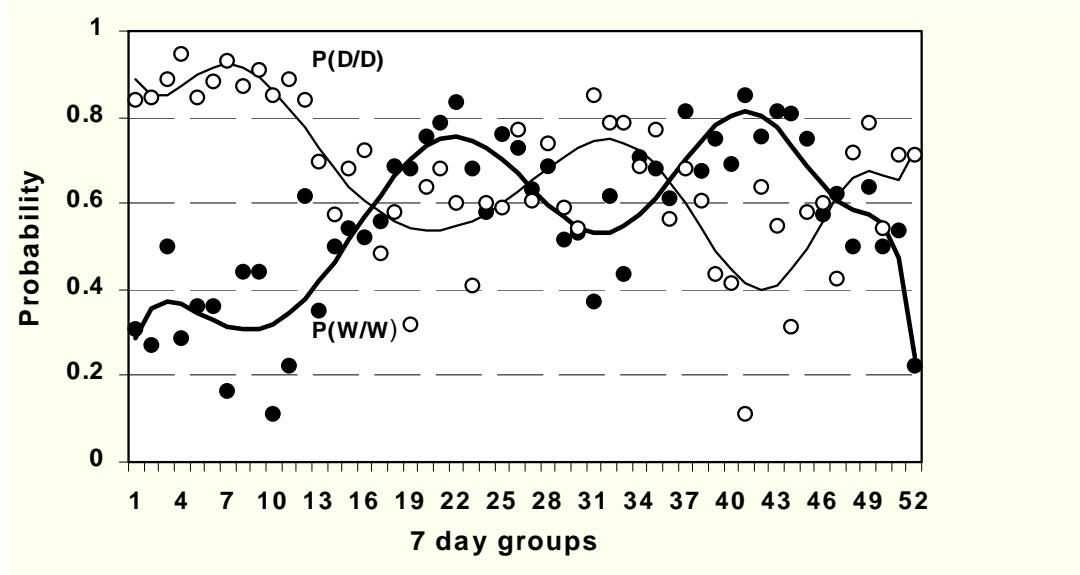


Fig. 8. Polynomial fits to the transition probabilities $P(W/W)$ and $P(D/D)$.

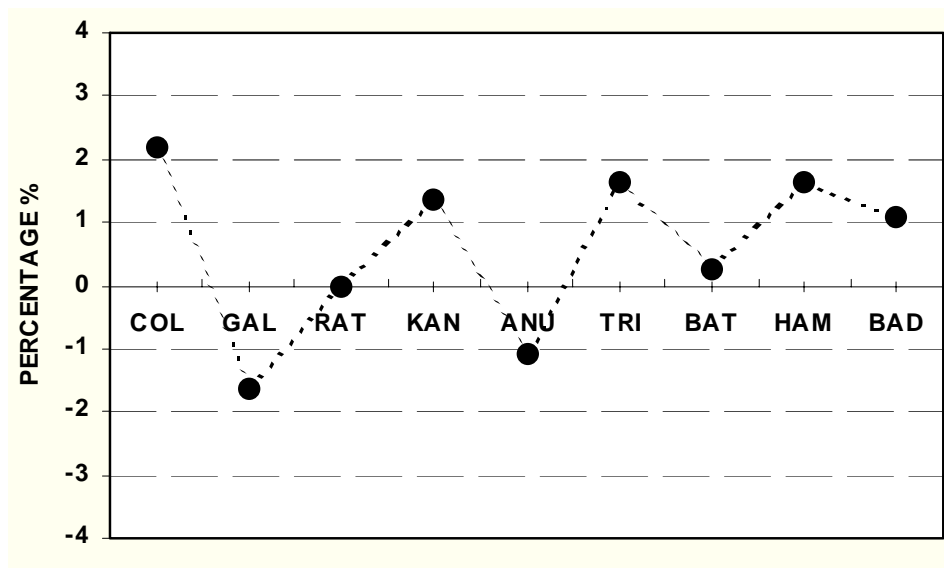


Fig. 9. Difference between the results obtained using the polynomial fitted probabilities and the direct results.

In table 4, a summary of the extreme efficiency figures is given for both order 1 and order 2. The highest value for wet and dry days was observed at Anuradhapura and the lowest values correspond to Galle for both orders. Anuradhapura is a station in the dry zone. The number of wet days in the dry zone is comparatively less than in a station in the wet zone as there is a tendency to rain throughout most of the year in the wet zone. Thus less random variation in rainfall occurrence can be observed in a station in the dry zone than in one in the wet zone. The highest value obtained for

only wet days corresponds to Ratnapura and the lowest is for Hambanthota. This means more wet days with considerable persistence, dominates rainfall records in Ratnapura, which also receives the highest annual rainfall among all the stations considered here. On the other hand Hambanthota is a station, which has the least wet days and the lowest annual rainfall, which reflects less persistence between weather conditions.

Table 4 Extremes in forecasts

	Status	Highest efficiency	Weather Station	Lowest efficiency	Weather Station
Order 1	Wet&Dry	78.57%	Anuradhapura	66.89%	Galle
	Wet Only	72.69%	Rathnapura	55.90%	Hambanthota
Order 2	Wet&Dry	77.91%	Anuradhapura	67.10%	Galle
	Wet Only	72.91%	Rathnapura	54.93%	Hambanthota

In table 5, a summary of the extreme values observed between the forecasts for the 1st day and the 7th day is shown. The reduction for wet and dry days for the first order model is less than 9% and it is less than 8% in second order model for all stations. It is less than 16% for only wet days for the first order model and it reduces to 11% for second order model.

Table 5. Differences between agreements in the first and seventh day.

Weather Station	Order 1		Order 2	
	Wet&dry %	Wet only %	Wet&dry %	Wet only %
Colombo	5.3	7.4	4.8	4.9
Galle	5.3	5.7	4.6	5.1
Rathnapura	1.8	1.6	0.9	1.2
Kandy	8.8	9.3	7.3	8.5
Anuradhapura	4.1	8.6	2.2	9.5
Trincomalee	0.7	1.9	1.9	5.3
Batticaloa	1.7	5.4	0.7	2.3
Hambanthota	1.9	16.0	1.5	10.5
Badulla	3.6	4.2	5.3	9.0

4. DISCUSSION

In forecasting daily rainfall occurrences, the first order Markov model is accurate up to 73% and no improvements were observed by developing higher order models. This observation is supported by a study conducted recently for one station in the dry zone of Sri Lanka⁶. Hence, considering the complexity involved, developing higher order Markov models to forecast rainfall occurrence in local weather stations cannot be justified. The average forecast efficiency for stations in the dry zones is

77% while it is 69% for stations in the wet zone for the 1st order model. The difference between the results obtained for the 1st order and the 2nd order Markov models is less than 1% for all stations.

The models indicate that there is no significant improvement in using data sets with long ranges for local weather stations. Also, no improvements were observed with data smoothing by fitting models to transition probabilities as suggested by some authors³. The developed models can be used to forecast the occurrence of rain over short periods not exceeding 7 days.

The approach adopted in this work for modeling rainfall will be useful in areas or seasons where precipitation is driven by synoptic scale systems. It is unlikely that these models will be successful when rainfall is driven primarily by convective activity (e.g. isolated thunderstorms). These processes evolve on relatively small scales and may not be predictable using synoptic scale circulation data. This is supported by the results given here for the months (October-November and March-April) where convective rainfall is dominant, since the percentages of agreement have low values.

Although the model described here is used only to predict rainfall occurrence, the same Markov chain framework can be further developed to forecast amounts of rainfall.

ACKNOWLEDGEMENTS

Financial assistance by IPPS, Uppsala University, Sweden (research grant number SRI:01/1), and the National Science Foundation, Sri Lanka (grant number RG/2000/P/01) are acknowledged.

REFERENCES

1. Sooriyarachchi M.R. (1989). Estimation of probable rainfall for irrigation and water management in the tropics. M.Sc. Thesis. University of Reading, UK.
2. Sanso B. and Guenni L. (1997). A stochastic Model for Tropical rainfall at a Single Location. <http://www.cesma.usb.ve/reportese.html#1997>
3. Stern R.D. and Coe R. (1984). A model fitting analysis of daily rainfall data. *Journal of Royal Statistics Society A*, **147**(1): 1-34
4. Stern R.D. and Coe R. (1982). The use of rainfall models in agricultural planning. *Agricultural Meteorology*, **26**: 35-50
5. Abeyasekara S., Senevirathne E. K. (1983). The analysis of daily rainfall data for agricultural purposes. *Journal of the National science council of Sri Lanka* **11**:165-183
6. Punyawardena B.V.R., Kulasiri D. (1996). On development and comparative study of two Markov models of rainfall in the dry zone of Sri Lanka. <http://www.lincoln.ac.nz/acms/reports/9611.html>
7. Grunwald G. A. and Jones R. H. (1998). Markov models for time series with mixed distribution. http://www.cgd.ucar.edu/stats/papers/gkg_rhj_env_1999.html
8. Huges J. P., Guthorp P. & Charles S. P. (1999). A Nonhomogeneous Hidden Markov model for precipitation occurrence. *Journal of Royal Statistical society A*, **48**(1), pp15-30
9. Feller W. (1951). An introduction to probabilistic theory and it applications. Chapman & Hall Ltd.
10. C. Roes & H. Ariton (1984). Application of Linear Algebra. John Willey & sons.
11. Kemeny J. G, Snell J. L. (1960). Finite Markov chains. D. Van Nostrand Company Ltd.
12. D.R. Cox & H.D. Miller (1977). Theory of Stochastic Process, Chapman & Hall.