Proceedings of the Technical Sessions, 16 (2000) 42-49
Institute of Physics – Sri Lanka

Preliminary Results of Long-Term and Short-Term Rainfall Forecasting

H.K.W.I. Perera and D.U.J. Sonnadara
Department of Physics, University of Colombo

Possibility of using statistical methods for long-term and short-term rainfall forecasting was investigated. Daily rainfall data from 8 meteorology stations namely, Colombo, Ratnapura, Kandy, Galle, Hambanthota, Batticoloa, Anuradhapura and Trincomalee were utilised in this study. A time series model was used for long-term forecasting and a Markov Chain model was used for short-term forecasting. The preliminary results show that the time series model with exponential smoothing fitted the data best and seasonal variations can be predicted with this model from weekly and monthly averages. The Markov chain model, applied by considering only two states, wet or dry, was successful to the level of 70% in predicting the status of a given day.

1. INTRODUCTION

Sri Lanka being an agricultural country and its main energy source is hydropower; the daily rainfall plays a dominant role in its economy. In some areas, rainfall is markedly seasonal in character, greatly limiting water availability for certain periods of the year. At other times, the same areas may receive excessive rainfall leading to a different set of problems. There is also considerable variation in rainfall from season to season and year to year. These temporal variations have a direct influence on water availability for agricultural, industrial and domestic requirements [1].

Rainfall forecasting is carried out at the Department of Meteorology for only short periods (generally for 24-hour periods). They use synoptic methods rather than models. Although accurate predictions of rainfall cannot be made due to high fluctuations in daily amounts and unpredictable nature of the weather parameters, reasonable level of predictions can be made on weekly and monthly variations by relating to past observations. Both long term as well as short term forecasting is important in designing water storage's, drainage channels for flood mitigation, and estimate crop growth and so on.

This study explores the possibility of rainfall forecasting using statistical techniques. Two different approaches were utilised, namely, the time series model for long-term forecasts and Markov chain model for short-term forecasts. In time series approach, exponential smoothing model was used to forecast for weeks and months ahead. Since persistence is a possible feature of daily rainfall, Markov chain model was used to forecast for few days ahead.

2. DATA

The data used in this study were amount of daily rainfall measurements from 8 meteorology stations (see Figure 1) maintained by the Department of Meteorology, namely, Colombo, Rathnapura, Kandy, Galle, Hambanthota, Batticaloa, Anuradhapura and Trincomalee for a period of 6 years (1992 to 1997).

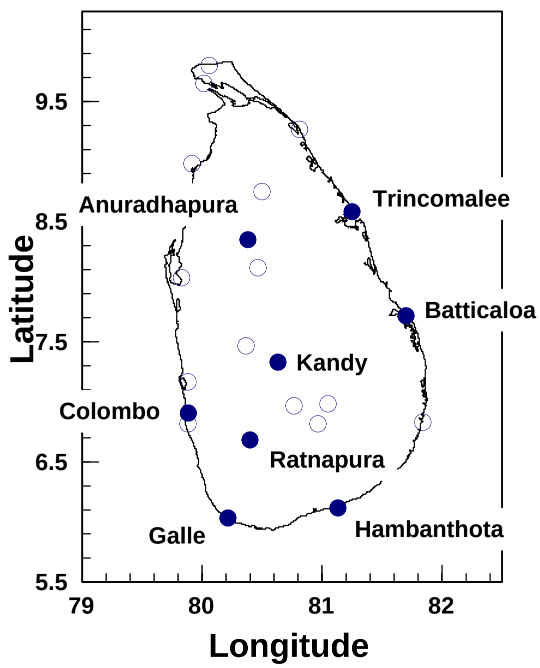

Figure 1: The 24 hour weather stations maintained by the Department of Meteorology. The selected stations for this work is marked as closed circles.

3. METHODOLOGY

3.1 Time series model

Time series is a collection of observations made sequentially in time, which represent the behaviour of the system in the past. If the past data are indicative of what we can expect in the future, we can postulate an underlying mathematical model that is representative of the process. The model can then be used to generate forecasts.

Various times series models were examined including ARIMA (Auto-regressive integrating moving average model) and finally the exponential smoothing model was chosen as the most appropriate model for the present work [2]. The smoothing "bring out" the major patterns or trends in a time series, while de-emphasising minor fluctuations such as random noise. In this transformation, each point is computed as a weighted sum of all proceeding observations.

For non seasonal time series with no systematic trend, T_{N+1} is taken as an estimate of weighted sum of past observations T_1, T_2, T_3, …T_N, where the weights lie on an exponential curve as given in Equation (1).

$$est(T_{N+1}) = C_0T_N + C_1T_{N-1} + \ldots\ldots\ldots \qquad (1)$$

Here $C_i = \alpha(1-\alpha)^i$, where i=0,1,2,… and α is a constant. Equation (1) can be rewritten in the recurrence form for finite number of past observations as shown below.

$$est(T_{N+1}) = \alpha T_N + (1-\alpha)est(T_N) \qquad (2)$$

$$est(T_{N+1}) = \alpha[T_N - est(T_N)] + est(T_N) \qquad (3)$$

$$S_T = \alpha\, E_N + S_{T-1} \qquad (4)$$

Here S_T is the transformed series at time T and S_{T-1} is the transformed series at time T-1. The error in prediction is $E_N = T_N - est(T_N)$.

Exponential smoothing can be generalised to deal with time series containing trend and seasonal variation, which can be additive or multiplicative depending on the series. The time series of rainfall data follows recurring seasonal variations with no trend. Thus, it is useful to smooth the seasonal components independently with an extra parameter δ. Seasonal components are additive in time series of rainfall data as shown in Equation (5).

$$I_T = I_{T-P} + \delta(1-\alpha)E_T \qquad (5)$$

Here I_T represents the predicted seasonal component at time T and I_{T-P} stands for the smoothed seasonal factor at time T minus the length of the season (P) (i.e. respective seasonal component in the last seasonal cycle). The parameter δ lies between 0 and 1. In general the one step ahead forecasts can be computed as shown in Equation (6) for additive seasonal model.

$$Forecast_T = S_T + I_T \qquad (6)$$

The value of the smoothing parameter α and seasonal smoothing parameter δ were estimated from past data to produce the smallest sums of squares of the residuals.

3.2 Markov chain model

Daily rainfall data generally show persistence, since weather systems producing rain last for more than one day. One way of exploiting this persistence in forecasting is through Markov chain method. Markov chain has special property that the future probability behaviour of the process is uniquely determined by the present state of the system.

Weather in a certain place could be one of two possible states, Dry or Wet. The definition of a "wet day" is a 24 hours period from 8.30 a.m. with a rainfall exceeding some threshold amount, normally, taken as 0.25 mm for Sri Lanka [2].

Under this technique, the transition from one state to another is taken as not predetermined, but rather be determined in terms of certain probabilities, which depend on the history of the system. Also it was assumed that these transition probabilities depend only on its state at the immediately proceeding observation. This process is called a Markov process with two states.

The states of the system are either "dry" or "wet" and they are defined as "0" and "1" respectively. The transition probability Pij (i, j = 0, 1) was defined as the probability that the system in state i at any one observation, it will be in state j at the next observation. For example, if state "0" correspond to a dry day at Colombo, and state "1" correspond to a wet day, then P_{01} is the probability that the weather in Colombo changes from dry to wet in two consecutive days. The transition probabilities, P_{00} P_{01} P_{10} P_{11} were expressed in terms P(D/D), P(R/D), P (D/R), P(R/R) respectively.

The transition probabilities make a 2×2 matrix and can be written as $\mathbf{P} = \{P_{i,j}\}^T$. Initial probabilities of the Markov process can be defined as $P^{(0)}$. Hence, $P^{(n)}$ for n = 1,2,3 … represent the final state vectors of a Markov process.

Consider the probability $P_i^{(0)}$ that the system is in state i initially and the probability of transitions leading to state j is $P_{i,j}$. To calculate the probability of the system in state j, we must sum over all transitions leading to state j. In matrix form, we can express the above idea as follows.

$$P^{(1)} = \mathbf{P}\,P^{(0)}$$

Similarly, $P^{(2)} = \mathbf{P}\,P^{(1)} = \mathbf{P}^2\,P^{(0)}$

In general, $P^{(n)} = \mathbf{P}^n\,P^{(0)}$ (7)

The matrix $\mathbf{P}^n$ therefore gives the required set of n–step transition probabilities $\{ P_{ij}^{(n)} \}$.

$$P^{(n)} = \mathbf{P}^n\,P^{(0)} \qquad \begin{pmatrix} P_0^{(n)} \\ P_1^{(n)} \end{pmatrix} = \begin{pmatrix} P_{00} & P_{01} \\ P_{10} & P_{11} \end{pmatrix}^n \begin{pmatrix} P_0^{(0)} \\ P_1^{(0)} \end{pmatrix}$$

Equation (7) shows how to calculate the absolute probabilities at any stage, (i.e. after n days) in terms of the initial probability distribution $P^{(0)}$ and the transition matrix $\mathbf{P}$.

4. RESULTS

4.1 Time series model

The daily rainfall data were first categorised into two sets; weekly and monthly averages from 1992 to 1996. Seasonal variations from weekly and monthly rainfall at each station were extracted via the seasonal decomposition method described in Section 3.1. The results are shown in Figure 2.

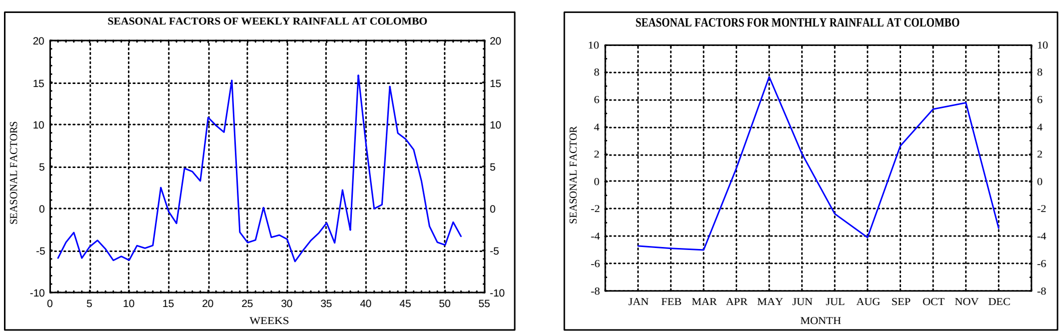

Figure 2: (a) seasonal variations in rainfall extracted by weekly average (b) same using monthly average

The comparison between the forecasted values for 1997 extracted from the past observations with the measured values is shown in Figure 3. It was found that the additive seasonal and no trend type model would be adequate when computing forecasts for monthly and weekly amount of rain. The reason may be the amount of rain is stable from year to year and change very slowly. At the same time, there can be seasonal changes (rainy seasons) which again may change slowly from year to year. Due to the limited number of data available it was not possible to look for long-term or short-term trends of this type.

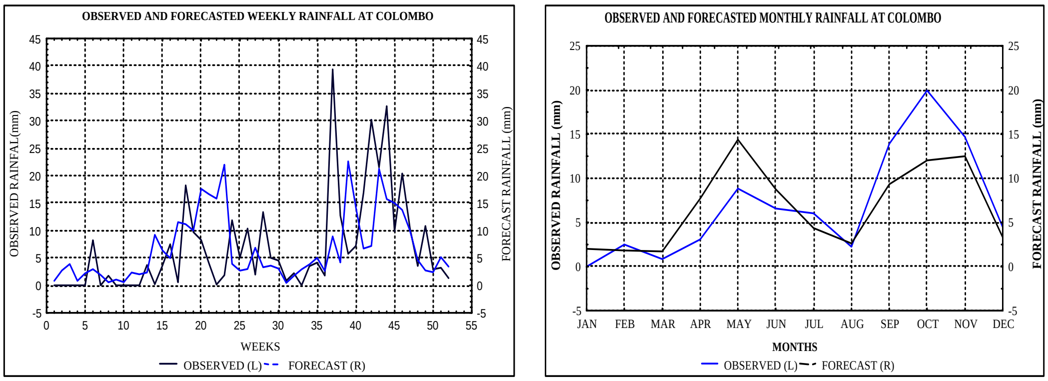

Figure 3: The comparison between the forecasted values for 1997 (solid line) with the measured values (dotted line) (a) weekly data (b) Monthly data.

4.2 Markov chain model

The definition of a wet day was used in preparing the rainfall data as records of states of the system for the period from 1992 to 1996. To simplify the calculations, it was assumed that the transition probabilities are the same during a given month. The numbers of wet and dry days were counted for 5 years (1992-1996) for each month also with the two day relationships having dry-dry, dry-wet, wet-dry and wet-wet states. These numbers are shown in Table 1 for the Colombo station.

Table 1: The number of wet and dry days for the 5-year period for the Colombo station.

Month	Dry	Wet	Dry-Dry	Dry-Wet	Wet-Dry	Wet-Wet
Jan	124	31	106	19	18	12
Feb	118	24	102	15	15	9
Mar	132	23	117	15	15	8
Apr	82	68	51	29	31	39
May	60	95	37	21	23	74
Jun	64	86	43	24	21	62
Jul	85	70	50	36	35	34
Aug	93	62	72	19	21	43
Sep	55	95	25	27	30	68
Oct	42	113	21	23	21	90
Nov	62	88	33	31	29	57
Dec	110	45	84	26	26	19

These numbers were used in calculating the transition probabilities for each month. The initial probability was calculated by taking average of the 3 proceeding days. By combining all transitions from initial states leading to a given final state, possibility of rain in any given day can be calculated.

In Figure 4 (a) we show the measured probability for rain during each month against the predicted probability by the model. A good linear correlation is seen. In Figure 4(b) we show the agreement between the actual rainy days and the model predicted rainy days for each month. A wide fluctuation in accuracy can be seen. In general Northeast monsoon period and Southwest monsoon months efficiency is high as expected.

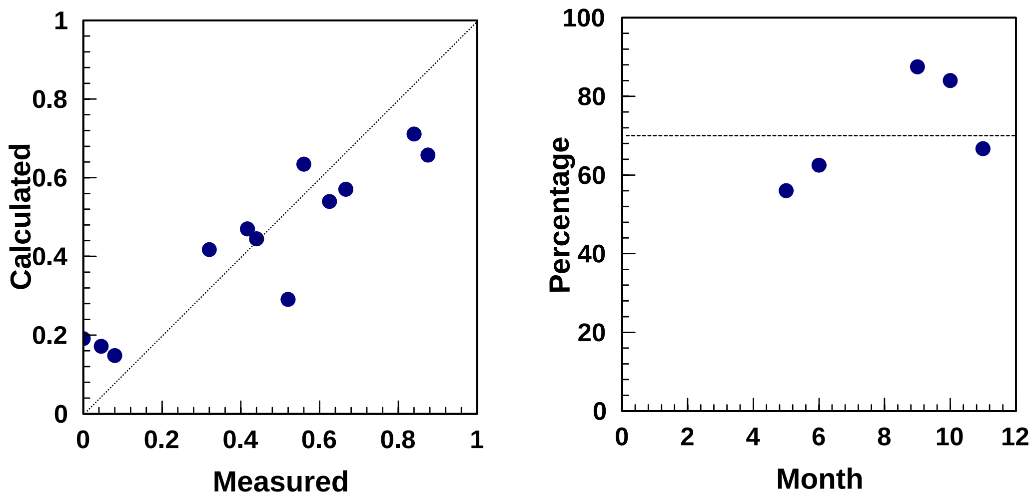

Figure 4: (a) Measured probability vs. predicted probability (b) agreement between the measured and predicted rainy days

5. CONCLUSIONS

The time series model for weekly and monthly levels can be used to predict the behaviour of a week or month, whether one can expect rain or not and to some extent even the amount of rainfall. Monthly predictions are better than weekly predictions when compared with actual values and show a high correlation.

Forecasts from Markov chain model can also be used to predict a given day to be rainy or not. Average efficiency observed in the present analysis is limited to 70%. The comparison between observed and predicted values shows (daily comparisons are not shown in this paper) the suitability of this model for short term forecasting from 1-7 days.

The size of the rainfall data used in the present work was limited to 6 years, and hence variations such as El–Nino Southern oscillation events, sun spot variations was not investigated.

ACKNOWLEDGEMENTS

Authors wish to thank Mr. M. Wimalasuriya, Mr. L. Chandrapala and Mr. A.B. Bandara at the Department of Meteorology for their support and useful discussions.

REFERENCES

1. Estimation of probable rainfall for irrigation and water management in the tropics, M.R. Sooriyarachchi, M.Sc. Thesis, University of Reading (1989)
2. Long and short term rainfall forecasting, H.K.W.I. Perera, B.Sc. Thesis, University of Colombo (1999)