

The Analysis of Daily Rainfall Data for Agricultural Purposes

SAVITRI ABEYASEKERA, E. K. SENEVIRATNE

Statistical Unit, University of Colombo, Colombo, Sri Lanka.

AND

ANN LEAKER, R. D. STERN

Department of Applied Statistics, University of Reading, Reading, RG6 2AN, U.K.

(Date of receipt : 08 July 1982)

(Date of acceptance : 11 November 1983)

Abstract: It is demonstrated that a comprehensive analysis of rainfall data to give useful agronomic results should use the daily measurements. Two methods of analysis are described and are illustrated using daily rainfall records for 56 years from Maradankadawala in the Anuradhapura district.

1. Introduction

Though rainfall measurements are recorded on a daily basis the daily data have seldom been made use of directly. A standard analysis of rainfall data for agricultural purposes involves first summarising the data to give monthly, or sometimes ten or seven day totals. These totals are then used as the basic units in all subsequent work. This approach has many inherent limitations. For example, it is not possible to obtain useful information on dry spells from this kind of analysis.

The main reason for totalling the daily data has been that the volume of data to be handled subsequently is greatly reduced. However, much larger volumes of data are now collected and analysed routinely in many other areas of study. Another reason for totalling the data was possibly the hope that the rainfall totals could then be assumed to be approximately normally distributed.¹⁰ This assumption is however almost never appropriate for seven or ten day rainfall totals. Both these reasons for avoiding the analysis of the daily records are no longer valid. The relevant statistical methods for handling non-normal data, together with the associated computing facilities needed for handling the larger volumes of data, are now available in Sri Lanka.

Our first aim in writing this paper is to demonstrate that a comprehensive analysis of rainfall data for agricultural purposes should use the daily measurements. At least this level of detail is needed to study many characteristics, such as the risk of

long dry spells or the probability of the occurrence of erosive rainfalls. Daily rainfall records are available from more than 400 sites in Sri Lanka and 110 of these have records for 50 years or more.

The second aim is to compare two methods of analysing daily rainfall data. The two methods which are discussed in the next sections are described in more detail.^{14,15} They are illustrated here using the daily rainfall records for 1923 to 1978 obtained from the meteorological station at Maradankadawala in the Anuradhapura District.

Our final objective is to contribute to the debate on ways in which climatic data in Sri Lanka should be analysed to be of maximum benefit to agriculture. The debate is far wider than the views expressed in this paper, because it relates to a range of other climatic variables besides rainfall, and also to available crop records. However, the dependence of crops on water necessitates the study of rainfall patterns, especially in tropical climates where droughts and crop failure are common. Thus, if it is accepted that much of the year to year variation in crop yields and in agricultural strategy (such as planting dates) may be caused by the variability in rainfall, then many important questions could be studied via a comprehensive analysis of the rainfall data. We would like to introduce such a system here.

2. The Direct Method of Analysis

The distinctive feature of what we have called the direct method of analysis is that, for any event or characteristic of interest, each year of data provides just one number. Thus if N years of data are available, and monthly totals are of interest, N observations would result for each month of the year. These observations are then treated as a simple random sample from a single population. Estimates of the probability of an event can then be obtained, either directly from the relative frequency of occurrence, or by fitting a suitable distribution to the sample values.

For example, if we are interested in total monthly rainfall amounts in November, percentage points can be estimated from a simple ordering of the N observations. Thus for Maradankadawala the 20th percentile for instance, can be estimated by noting that the total rainfall for November was less than 6.34 inches in 20% of the 56 years. The 20th percentile is therefore 6.34 inches and this means that the chance of receiving more than 6.34 inches rain in November is estimated as 0.8. The method is straightforward and makes no assumption about the distributional pattern of the observations. However, the estimates will not be very precise unless data are available for a very large number of years. This is generally true for all the methods described in this section.

An alternative approach is to fit a distribution to the 56 monthly totals for November. For periods of a month or longer, the rainfall totals are often found to be

approximately normally distributed. The histogram of totals for November shown in Figure 1 is approximately symmetrical with a mean of 11.4 inches and a standard deviation of 4.38 inches. So the assumption of a normal distribution for the totals may be reasonable and properties of the normal distribution then provide an estimate of the 20th percentile as 6.52 inches.

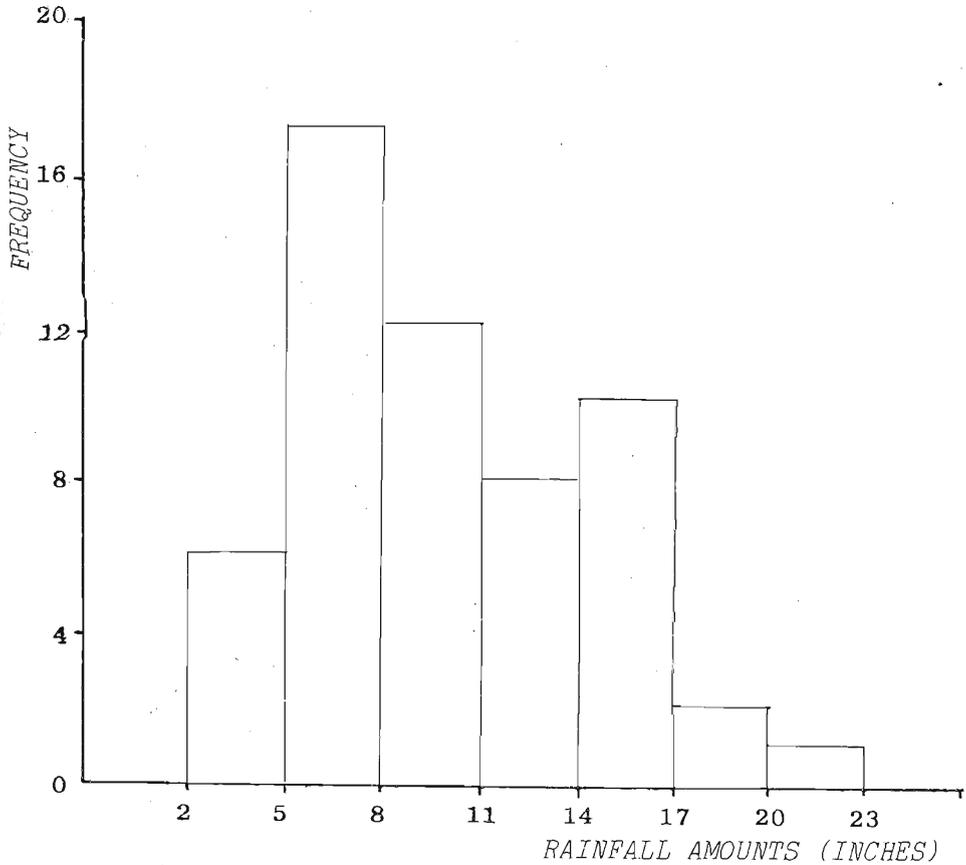


Figure 1. Histogram of total rainfall amounts in November (1923 - 1978) at Maradankadawala.

In the case when 10-day rather than monthly totals are of interest, the distribution-free approach can still be used as before, to estimate percentage points. Some results are shown in Figure 2. However, for 10-day totals, the assumption of normality is rarely appropriate. Transformations to normality can be used,⁹ but

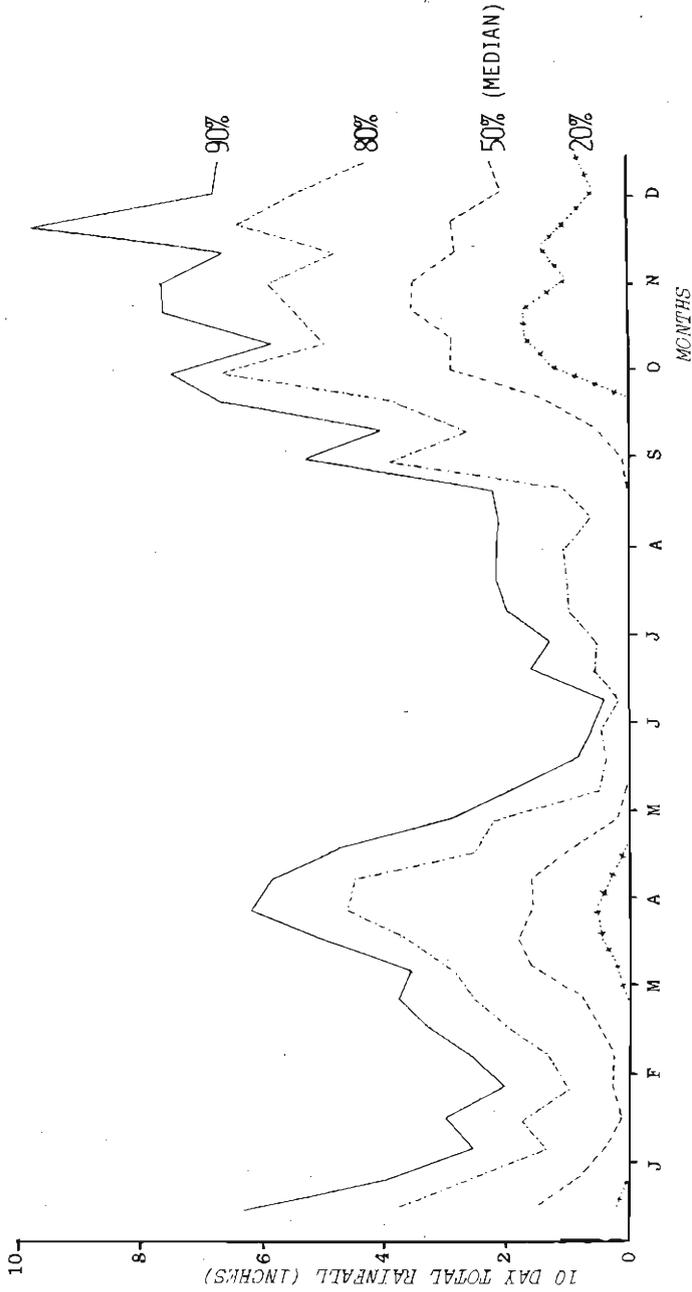


Figure 2. Percentiles for the 10 day total rainfall amounts at Maradankadawala (1923 - 1978) estimated from the raw data.

different transformations are usually needed for different periods of the year. A better approach is to fit a suitable alternative distribution. One possibility is to fit a gamma distribution.¹¹

The type of analysis of amounts described above is useful and does provide valuable information, but is often all that is done. By making use of the daily data, many other characteristics of importance to agriculture can be studied. Here the risk of long dry spells is considered. This risk, defined as the "probability of a dry spell of greater than a specified number of days" within any period of interest, can be calculated for different times of the year by using the following method.

Table 1. Daily rainfall measurements (inches) for 1977 at Maradankadawala together with wet/dry sequences. Positive numbers show sequences of dry days and negative numbers show sequences of wet days.

October		November		December	
Rainfall	dry/wet state	Rainfall	dry/wet state	Rainfall	dry/wet state
0.11	-1	**	4	0.11	-1
**	1	**	5	**	1
**	2	0.85	-1	**	2
0.13	-1	0.29	-2	0.21	-1
1.12	-2	0.07	-3	**	1
0.29	-3	**	1	**	2
**	1	**	2	0.04	-1
**	2	**	3	**	1
2.06	-1	**	4	**	2
1.36	-2	0.72	-1	0.02	-1
**	1	0.34	-2	**	1
0.08	-1	5.31	-3	1.48	-1
**	1	**	1	**	1
1.82	-1	**	2	**	2
0.44	-2	**	3	**	3
0.16	-3	**	4	0.31	-1
2.02	-4	**	5	2.49	-2
0.42	-5	**	6	0.40	-3
0.29	-6	**	7	0.09	-4
1.65	-7	**	8	1.12	-5
2.07	-8	0.08	-1	0.18	-6
1.22	-9	0.06	-2	0.38	-7
0.45	-10	1.12	-3	0.30	-8
0.51	-11	0.10	-4	0.02	-9
**	1	0.13	-5	0.27	-10
**	2	0.19	-6	**	1
**	3	**	1	**	2
0.27	-1	**	2	**	3
**	1	**	3	**	4
**	2	**	4	**	5
**	3	-	-	**	6

First it is necessary to specify what is meant by a day being dry. A threshold value is chosen, and a day is classified as being dry if the amount of rain falling on that day is less than the threshold value. If not it is classified as being wet. The threshold can be chosen to be large enough to exclude days where the rainfall is so minimal as to be agriculturally insignificant. A threshold of 0.005 is often found acceptable although its choice would be dependent on the crop water requirements. The daily observations are then recorded as a sequence of dry and wet days. Table 1 shows the daily records from October to December in 1977, the positive numbers indicating a sequence of dry days and the negative numbers indicating a sequence of wet days.

Having specified what a "dry" day is, we can now define a dry spell. An n -day dry spell is a period of n successive dry days preceded and followed by a wet day. We also say that the period of interest will include an n day dry spell if at least n consecutive dry days are included within the period. It is then simple to calculate the proportion of years which have a dry spell of n or more days within that period.

Some results are given in Figure 3. This shows for example that there was a dry spell of 10 days or more in 96% of years in June compared with 16% in November. This simple analysis can often give useful information which complements the results from an analysis of total amounts. It is also clear that such information could not have been obtained without using the daily data.

Daily rainfall measurements can also be used to obtain results on the distribution of the start and length of rains¹⁴. An event to mark the start of the rains can be defined in different ways depending on the ultimate aims of the analysis. For instance, if 25 mm of rain in one day is needed for planting a particular crop, and if planting before 1st September is never done, the start of the rains from the farmers' viewpoint will coincide with the first day with 25 mm or more after 1st September. Stern¹³ gives a more general definition which specifies an earliest possible starting date D and a potential start date coinciding with the first occurrence of at least y mm totalled over n days. They also consider the potential start as a false start if a dry spell of t or more days occurs in the next m days. Values for D , y , n , t and m are set by considering the purpose for which the start of the rains becomes important.

The end of the rains can also be defined on the basis of rainfall amounts or by using in addition, information on soil water storage. When the start and end of the rains have been satisfactorily established, the length of the rains can be calculated by subtraction and its distributional pattern investigated. This can provide a guide to the selection of crop varieties.

The direct method of analysis described in this section is very straightforward and only few assumptions are needed about the underlying structure of the data. However, the cost that is incurred for this simplicity is that the method can only really be used for long records and even then the estimates obtained have large standard errors. A further problem arising from the lack of precision is that the comparison of

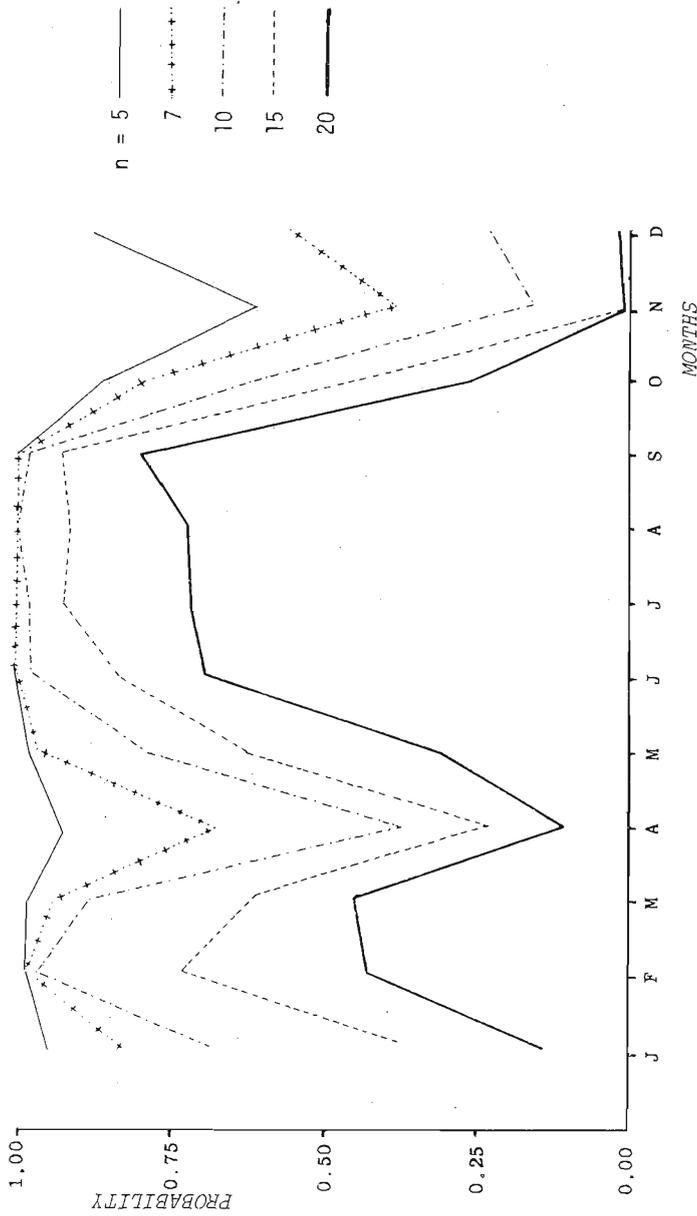


Figure 3. Probability of a dry spell of length $\geq n$ days, for $n = 5, 7, 10, 15, 20$, in each month, estimated using the raw data from 1923-1978 at Maradankadawala.

rainfall from different sites is difficult. Because of the limitation of the direct method, an alternative approach, which involves the fitting of models to the daily rainfall records, is discussed in the next section.

3. Fitting models to the daily data

Many authors have considered the fitting of models to daily rainfall data,^{2,7} However, little practical use has been made of this approach. One reason may be that use of this modelling approach involves terminology that has been daunting to all but statisticians. In addition, the computer packages that permit the routine fitting of suitable models to the daily data from a number of sites have only recently become available. It will be shown in this section that it is possible to use models of the patterns of rainfall on a daily basis to give any of the results that can be derived from the direct analysis. Because this modelling approach makes efficient use of the data, the estimates obtained are more precise. The method can therefore be used for relatively short records, and more accuracy obtained, than if the direct method were used.

When simple descriptions of the rainfall patterns, such as percentage points for the total rainfall amount during a specified period, are obtained using the modelling approach, they give results that are "smoothed" versions of those obtained directly from the data. The probability of long dry spells in a given period⁵ and the probability distribution of the start of the rains¹³ can also be obtained from the model. The modelling approach is described in considerable detail.^{4,15} The methods are illustrated here using the rainfall records at Maradankadawala.

We first describe a very simple model which although inadequate in practice, shows the way in which estimates can be obtained for the same quantities considered in the direct method of analysis described above. Suppose that in the N years of data available, $100p\%$ of days in a given month were rainy. The estimated probability of rain in that month is then p . A histogram of rainfall amounts on the rainy days in the month will indicate a suitable distribution that could be fitted to the data. A gamma distribution is often appropriate.³ The probability density function of the gamma distribution is given by

$$f(x) = \left(\frac{k}{\mu}\right)^k x^{k-1} e^{-kx/\mu} / \Gamma(k), \quad X > 0, k, \mu > 0.$$

The observed rainfall amounts are used to estimate the two parameters of this distribution. One of these parameters μ , is the mean rain per rainy day, while the shape parameter, k , can be interpreted by noting that $1/\sqrt{k}$ is the coefficient of variation of the distribution.

For example a possible model for rainfall in November at Maradankadawala is the following. The probability of rain on any day is estimated as $p = 968 / (30 \times 56) = 0.576$.

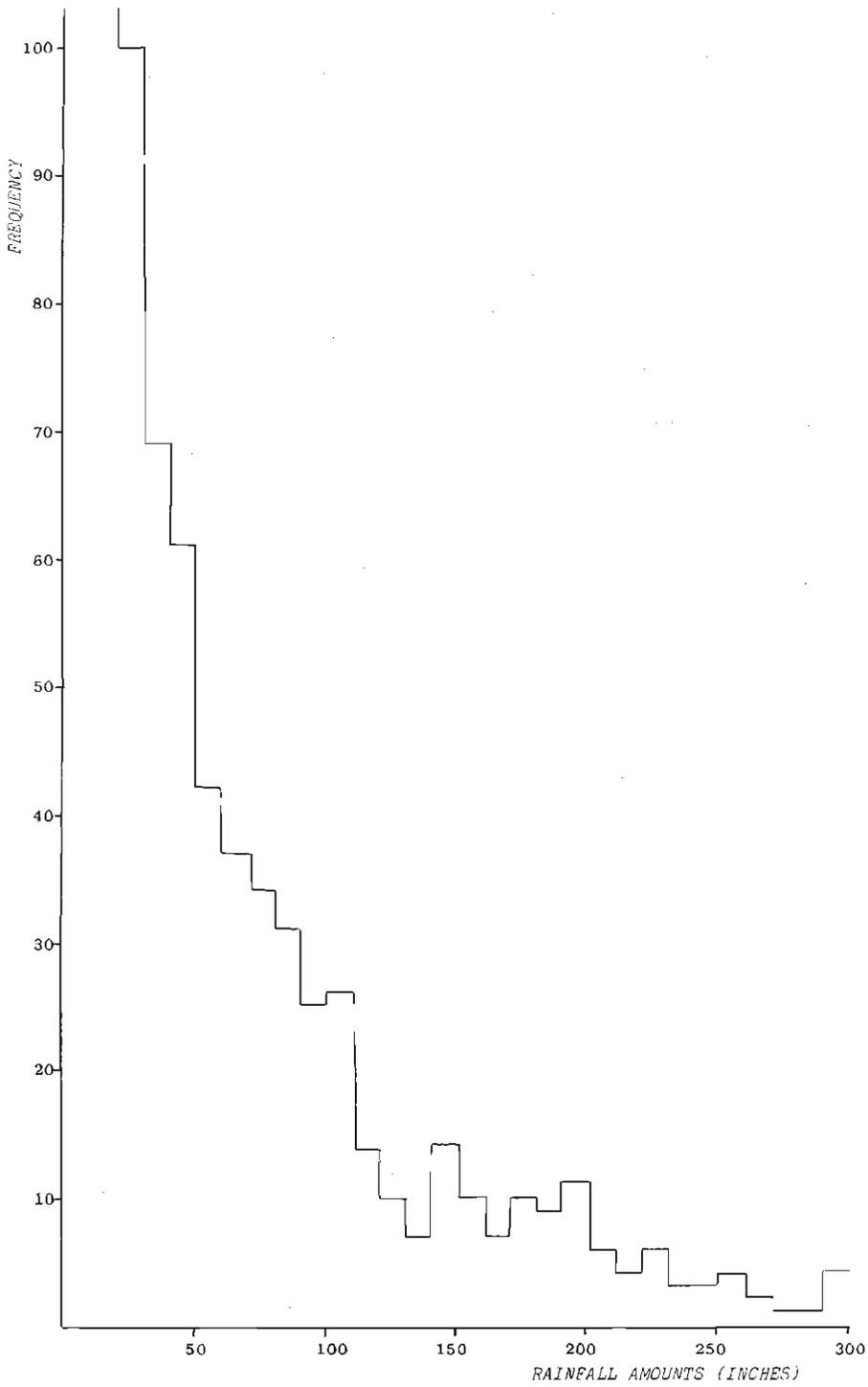


Figure 4. Histogram of daily rainfall amounts in November (1923 - 1978) at Maradankadawala.

This is assumed to be the same irrespective of whether the previous day was dry or wet. The distribution of rainfall amounts is very skew. Figure 4 gives the histogram of this distribution for the 968 rainy days. The rainfall amounts on rainy days is assumed to come from a gamma distribution with mean $\mu = 0.64$ inches and shape parameter $k = 0.74$. It is also assumed that the distribution of rainfall amounts is the same on the first day of a wet spell as on subsequent rainy days.

One result that can be derived using the above model concerns the distribution of the number of rainy days in a given month. With the assumptions of the model this distribution is binomial with $n = 30$ and $p = 0.576$. Tables of the binomial distribution then allow various percentage points to be evaluated. For example, approximately one year in six is estimated to have more than 20 rainy days in November. It is also straightforward to derive results on many other aspects that are of interest such as the distribution of total rainfall in the month, or the chance of a dry spell of a given length or more.¹⁵

Although the model described above is simple, it is too limited to be of real use, particularly because some of the basic assumptions do not hold. The proportion of rainy days through the year, evaluated using 10 day periods, shows that the assumption of a constant chance of rain is not particularly realistic, even for periods as short as a month, (see Figure 5). A curve is therefore fitted to the observed proportions and is also given in Figure 5. This shows clearly the bimodality of the pattern of rainfall occurrence with the chance of rain rising to its highest value of 0.63 in November.

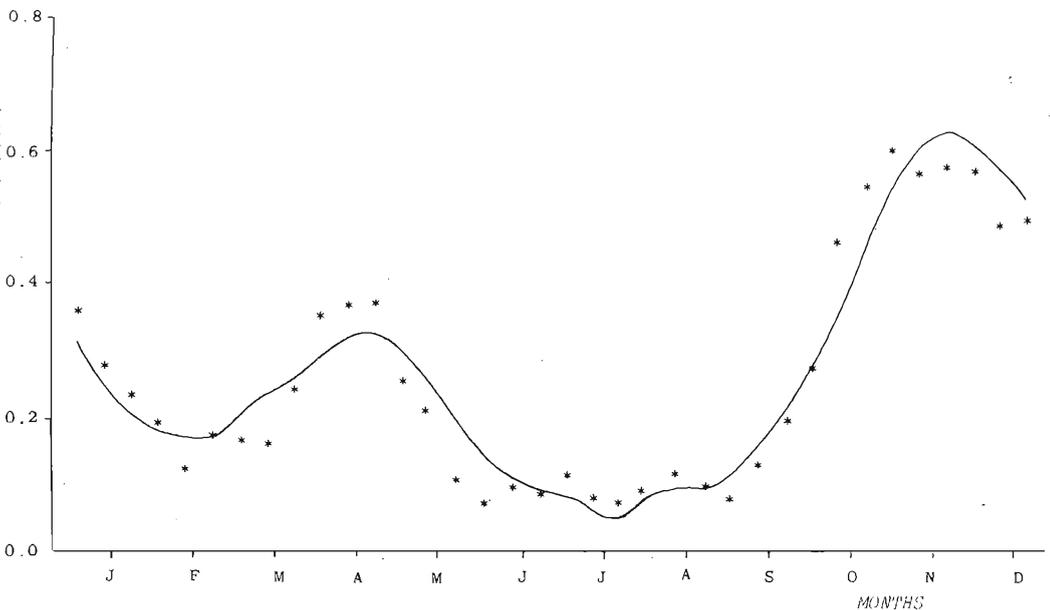


Figure 5. Observed proportion of rainy days together with a fitted curve.

Very frequently, the chance of rain on any day is conditional on the dry or wet state of one or more of the previous days. If it can be assumed that this dependence is limited to the state of the previous day, then two curves are needed to describe the chance of rain. One of these curves gives the probability of rain on a given day conditional on the previous day being dry, while the second curve gives the corresponding probability conditional on the previous day being wet. For Maradankadawala it is found that four curves are required to model the data adequately. The observed data are given in Figure 6 and the corresponding fitted curves in Figure 7. These curves give the estimated chance of rain depending on whether or not either of the two previous days were dry or rainy. For example, in November the chance of rain is estimated to be about 0.3 if the two previous days were dry compared with about 0.8 if both had rain.

In statistical terminology, what has been done in Figure 7 is to fit a Markov Chain to the data on rainfall occurrence. It is a two state (dry or rainy), second order chain, with non-stationary transition probabilities. The model has been fitted using the computer package GLIM¹ and a binomial error structure has been assumed.

The distribution describing the rainfall amounts has also to be generalised. It is often found that rainfall amounts per rainy day tend to vary through the year. For instance at Maradankadawala the average amount of rain per rainy day during the monsoon is greater than in the intermonsoonal period. Figure 8 gives the mean rain per rainy day through the year together with the fitted curve for the rainfall record at Maradankadawala. The curve fitted was a Fourier series with four harmonics and is fitted assuming a gamma error structure. It shows for example that in June, rainy days had an average of 0.18 inches and that this rose to 0.64 inches in November. Thus instead of a single gamma distribution to model rainfall amounts, a gamma distribution with varying mean was used. The second parameter of the gamma distribution, k , was approximately constant through the year and was estimated as 0.74.

A more appropriate model for Maradankadawala now consists of the equations of the curves given in Figures 7 and 8 plus the estimate of k . With this model and employing the method of derivation described in Stern,¹² it is possible to estimate the percentage points of the 10 day totals. These are given in Figure 9 and it can be seen that the results are very similar to the corresponding results derived from the direct analysis, (Figure 2). Figure 10 gives results on the chance of dry spells of different lengths, which is to be compared with the corresponding results from the direct analysis given in Figure 3.

One advantage of using the modelling approach is that results for any characteristic of interest can be derived directly from the model, whereas with the direct method we have to refer repeatedly to the daily data. A second point to be noted is the

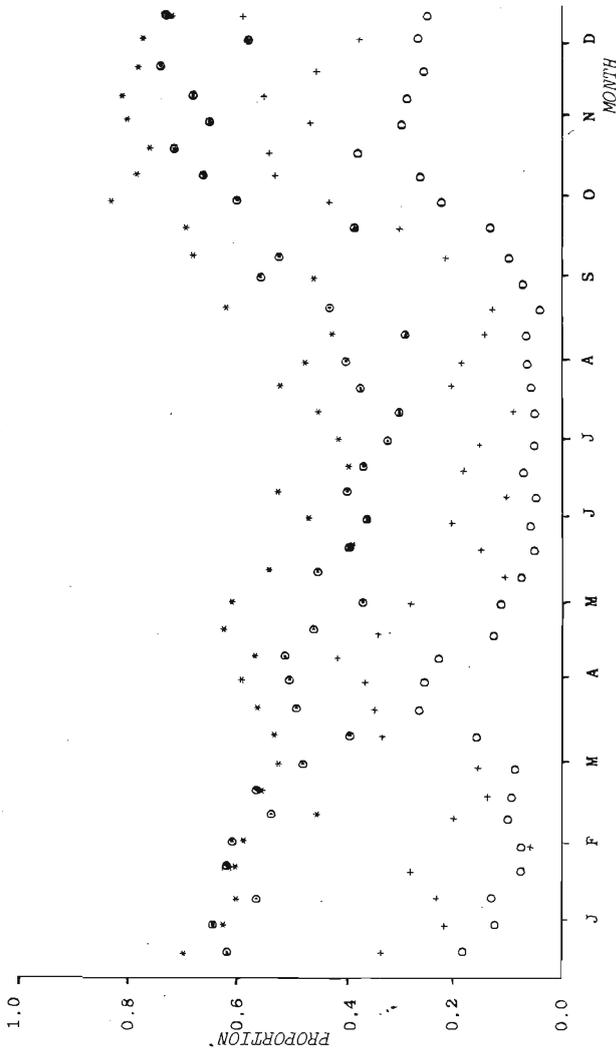


Figure 6. Proportion of rainy days conditional on whether the two previous days were dry or wet.
 (a) Yesterday dry, day before dry is denoted by o ;
 (b) Yesterday dry, day before wet is denoted by + ;
 (c) Yesterday wet, day before dry is denoted by o with a dot ;
 (d) Yesterday wet, day before wet is denoted by * .

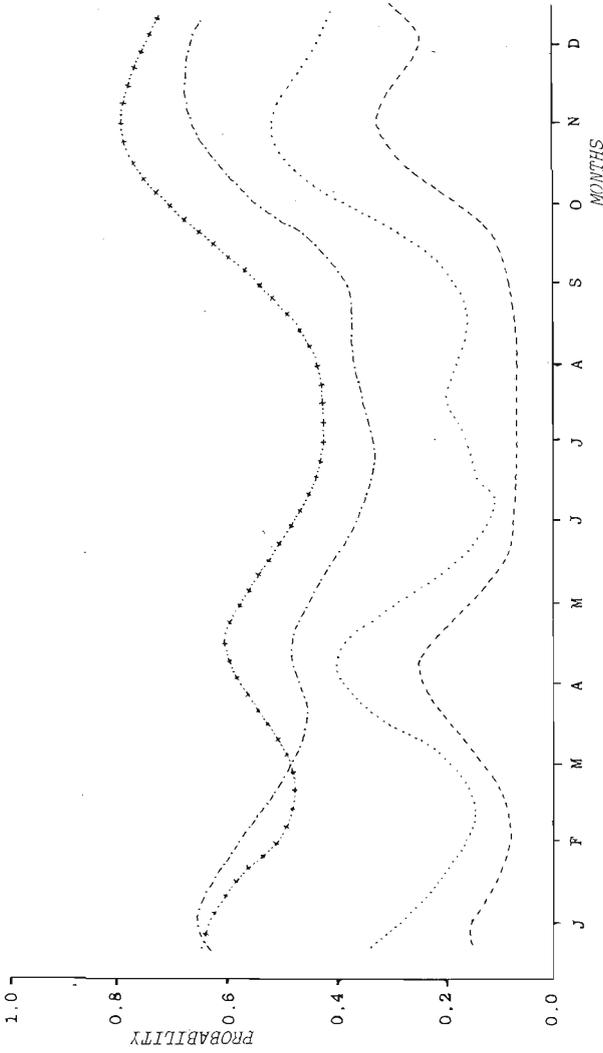


Figure 7. Second order Markov chain analysis for the probability of rain at Maradankadawala. The four curves represent: +,+,+ P(R/RR); - - - - - P(R/DR); P(R/RD); - - - - - P(R/DD); Where for example, P(R/RD) denotes the probability of rain given dry the day before and rain two days earlier.

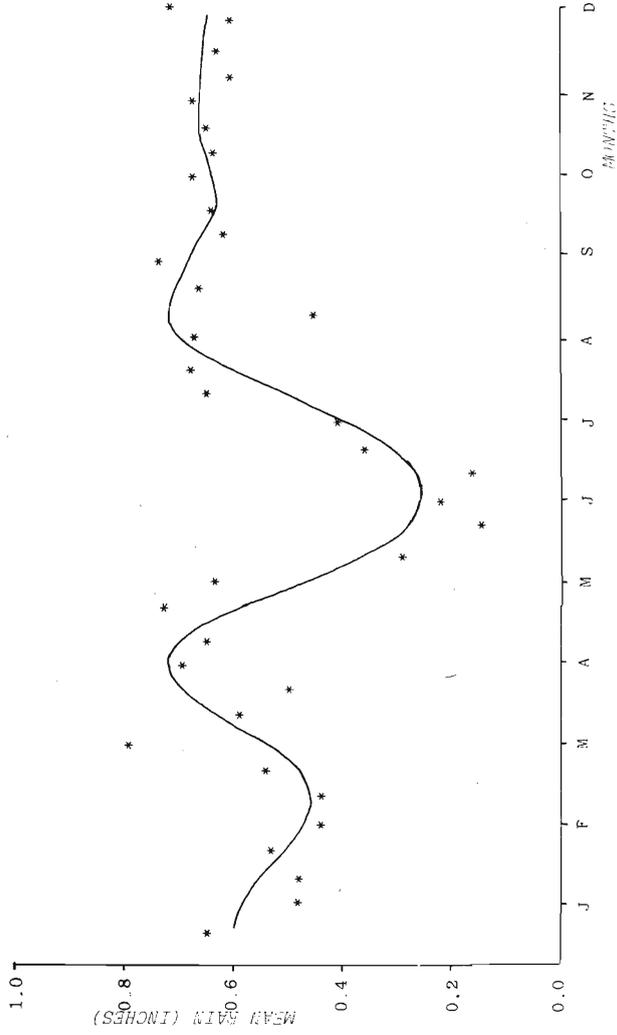


Figure 8. Mean rain per rainy day (in inches) together with a fitted curve.

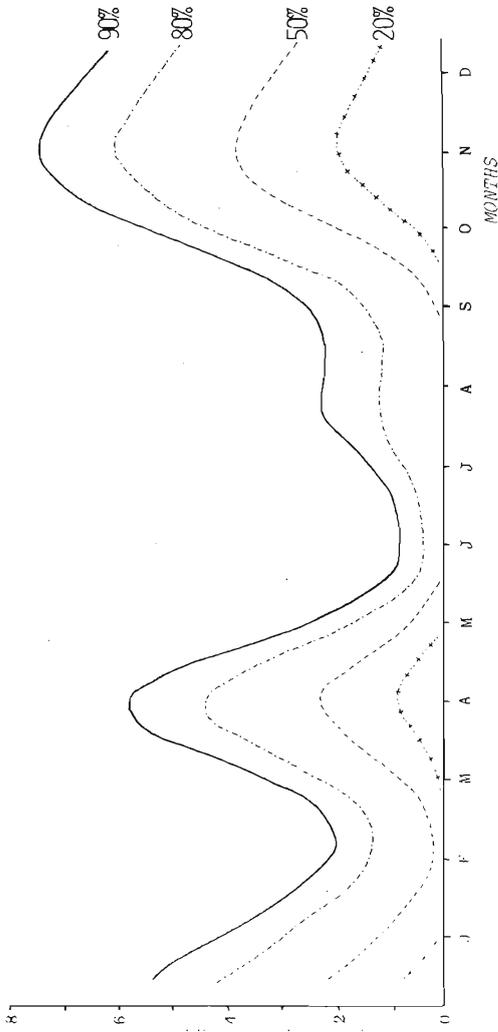


Figure 9. Percentiles for the 10 day total rainfall amounts at Maradankadawala.

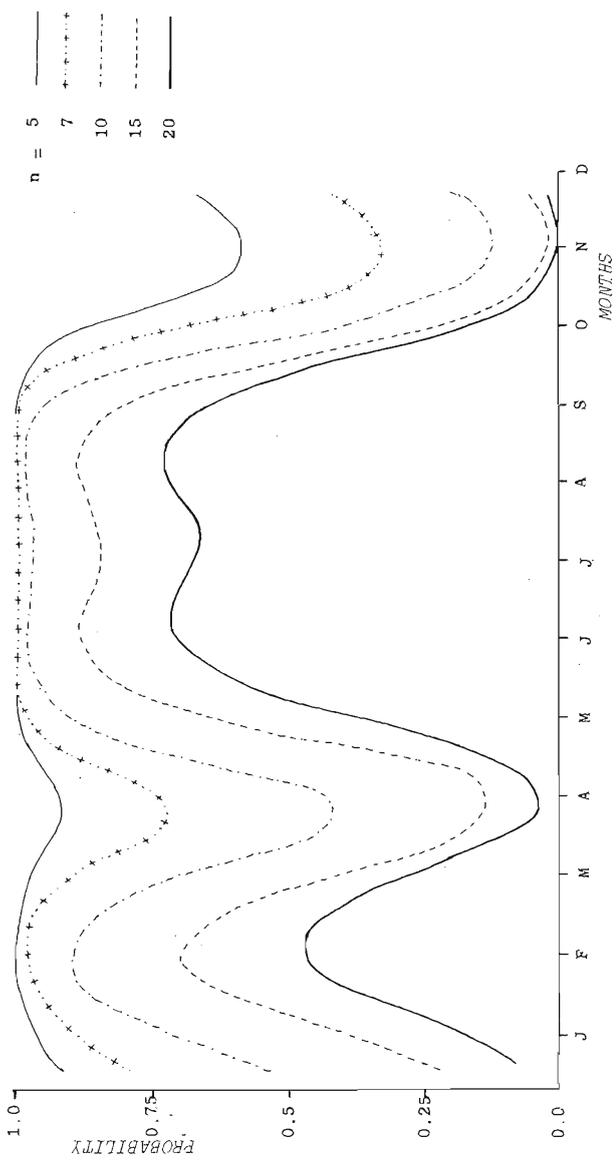


Figure 10. Probability of a dry spell of n or more days ($n = 5, 7, 10, 15, 20$) at Maradankadawala

smoothness of curves obtained from the model; (compare for instance Figures 9 and 10 with Figures 2 and 3). This makes it far easier to compare the pattern of rainfall at different sites. As an example, Garbutt *et al*⁸ discuss the fitting of models similar to the one adopted here for 11 sites in West Africa and make use of these models to compare the distribution of the data of the start of the rains.

4. Discussion

The type of model discussed in the last section seems to model the pattern of rainfall well in many countries^{2,8}. The results obtained from Sri Lankan data indicate that this type of model is also likely to be appropriate here. The necessary computing facilities are now available in Sri Lanka and it is therefore timely to assess the practical use that could be made of such work.

One prerequisite is for agricultural researchers to visualise the wide range of questions that can be answered together with the necessity to phrase these questions precisely. A few examples are given below. These were taken from a recent case study which analysed data from two sites in Kenya,⁶ having a bimodal rainfall pattern with the 'short' rains centered around April and the 'long' rains round November. The questions were posed by agriculturalists and are not entirely satisfactory, but they serve to illustrate some possibilities.

- (i) What is the average date for the beginning of the rainy season? This event is defined as the first day after 10 March that rainfall reached 20 mm, totalled over three consecutive days. What is the probability that this event will occur 5, 10, 15 and 20 days earlier or later than this average date?
- (ii) A dry spell is requested for weeding after crop emergence. What is the chance that there will be a dry spell (defined as no rain or as rain totalling less than 5 mm per day) of at least 5 days duration during the period 15 to 25 days after the onset of the rains?
- (iii) What are the chances of getting 20, 30, 40, 50 mm or more in a day within given 10 day periods? Impressions indicate that heavy showers fall within the first 30 days of the short rains before a crop canopy is established and run-off therefore occurs.
- (iv) What is the average date for the end of the rains? (A range of definitions are given for this event⁶). What is the probability that the end of the rains will occur 5, 10, 15 and 20 days earlier or later than the average date?

Finally, we do not wish to give the impression that the research is over from a statistician's point of view and that all that is required is for agricultural researchers to ask the appropriate questions. Even within the limited field of the analysis of rainfall data, there are many important questions which still remain to be explored on the range of models that can be fitted and the ways in which they can be used. In addition, at present the analysis is limited to single sites. Little work has been done on the way in which the results could be extended to allow maps to be drawn showing, for example, the risk of a long dry spell over a region of the country.

We believe that the application and the extension of the methods of analysis discussed in this paper are rewarding areas for collaborative research in Sri Lanka, involving meteorologists, agriculturists and statisticians and we believe that it is possible for the results to make a significant contribution to agricultural planning in Sri Lanka.

Acknowledgements

This work was supported by the Natural Resources, Energy and Science Authority of Sri Lanka and the Overseas Development Administration of the British Government. We thank colleagues in the Statistical Unit, Colombo and the Agroclimatology team, Reading for their help and support.

References

1. BAKER, R. J. & NELDER, J. A. (1978). *The GLIM Manual - Release 3*. Numerical Algorithms Group, Oxford.
2. BASU, A. N. (1971). Fitting of a Markov chain model to daily rainfall data at Calcutta. *Indian J. Met. Geoph.* **22**, 67-74.
3. BUISHAND, T. A. (1977). Stochastic modelling of daily rainfall sequences. Veenman and Zonen, Wageningen, Netherlands.
4. COE, R. & STERN, R. D. (1982). Fitting models to daily rainfall data. *J. Appl. Meteor.* **12**, No. 7, 1024-1031.
5. DENNETT, M. D. & STERN, R. D. (1979). Incidence of dry spells in the cropping season. Paper presented at the Annual Research Conference, International Institute of Tropical Agriculture, Ibadan, Oct. 15-19.
6. DENNETT, M. D., RODGERS, J. A. & STERN, R. D. (1981). Analysis of daily rainfall data from Machakos, Kenya. *Report No. 3, Trop. Agric. Met. Group*, University of Reading.
7. GABRIEL, K. R. & NEUMANN, J. (1962). A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quart. J. Royal Met. Soc.*, **88**: 90-95.
8. GARBUTT, D. J., STERN, R. D., DENNETT, M. D. & ELSTON, J. (1981). A comparison of the rainfall climate of eleven places in West Africa, using a two-part model for daily rainfall. *Arch. Met. Geoph. Biokl. Series B* **29**, 137-155.
9. HILLS, R. C. & MORGAN, J. H. T. (1981). An interactive approach to the analysis of rainfall records for agricultural purposes: Rainfall Statistics. *Experimental Agriculture*. **17**, 1-16.

10. KOWAL, J. M. & KNABE, D. T. (1972). An agroclimatical atlas of the Northern States of Nigeria. Ahmadu Bello University Press, Zaria, Nigeria.
11. MOOLEY, D. A. (1973). Gamma distribution probability model for Asian summer monsoon monthly rainfall. *Monthly Weather Review* **101**, 160-176.
12. STERN, R. D. (1980). The calculation of probability distributions for models of daily precipitation. *Arch. Met. Geoph. Biokl. Series B*, **28**, 137-147.
13. STERN, R. D., DENNETT, M. D. & GARBUTT, D. J. (1981). The start of rains in West Africa. *J. Climatology*, **1**: 59-68.
14. STERN, R. D. DENNETT, M. D. & DALE, I. C. (1982). Methods for analysing daily rainfall measurements to give useful agronomic results. I. Direct Method. *J. Experimental Agriculture* (In press).
15. STERN, R. D., DENNETT, M. D. & DALE, I. C. (1982). Methods for analysing daily rainfall measurements to give useful agronomic results. II. A modelling approach. *J. Experimental Agriculture* (In press).