Pertanika 11(1), 101 – 106 (1988)

A Simple Method for Analysing Daily Rainfall Data – A Case Study for Kota Bharu

MOHAMAD NOOH HASSAN[1] and ROGER D. STERN
Department of Applied Statistics,
University of Reading, England, RG6 2AN

Key words: Gamma distribution; percentage point; dry spell; rainfall totals; threshold.

ABSTRAK

Ditunjukkan bahawa analisis data sukatan hujan yang menyeluruh memerlukan data harian untuk menghasilkan beberapa keputusan penting dalam bidang agronomi. Kaedah-kaedah analisis cara mudah dihurai dan dijelaskan menggunakan rekod sukatan hujan harian selama 34 tahun (1951–84) bagi Kota Bharu, Kelatan.

ABSTRACT

It is demonstrated that a comprehensive analysis of rainfall data to give useful agronomic results should use the daily measurements. Simple methods of analysis are described and are illustrated using daily rainfall records of 34 years (1951–84) for Kota Bharu ini the north-eastern state of Kelantan.

INTRODUCTION

Rainfall is one of the principal controlling elements in agriculture in the tropics. The success of crop establishment and growth depend largely on the availability of adequate rainfall. The amount of rainfall that is received determines the type of agriculture that can be carried out and the crop that can be cultivated in a region.

Various aspects of rainfall pattern in Peninsular Malaysia have been studied by Dale (1959, 1960), Nieuwolt (1966), Wycherley (1966) and more recently by Nieuwolt (1981) and Nieuwolt *et al.* (1982). In addition, specific investigations of the occurrence and pattern of rainfall within a specific area have been carried out by Narayanan (1967), Chia (1968) and Yap (1973).

For agriculture, important aspects which relate to rainfall include the days of the start and the end of the rainy season, and hence its length, plus the rainfall amounts and the risk of dry spells within the season. In this paper, a 34-year record (1951–84) of daily rainfall data for Kota Bharu are used to illustrate a simple method of analysis (Stern *et al.*, 1982) to consider aspects such as these.

MATERIALS AND METHODS

The seasonal distribution of rainfall in Peninsular Malaysia is influenced by the two monsoons. The northeast monsoon brings considerable rain to the eastern part of Peninsular Malaysia in November and December. There is normally little rain during February and March. From about June to September the general circulation is dominated by the southwest monsoon which, despite having crossed the high mountains of Sumatra, brings considerable rain to all parts of the Peninsula.

The Department of Meterological Services Malaysia in Petaling Jaya keeps a data bank on 291 meterological stations. Thirty-five of these stations are operated by the Department with 19 stations having records of over 30 years. The longest record is 56 years and the shortest is just

[1] On study leave from the Department of Mathematics, Universiti Pertanian Malaysia.

Key to authors' name; Hassan, M.N. and Stern, R. D.

one year. There are also 114 'agromet' stations of which all but 38 stations have rainfall records of more than 15 years. Another thirty-one stations are operated by estate managements of which only two have rainfall records as long as 15 years. The rest of the stations are operated by other government agencies.

The daily rainfall data from Kota Bharu were supplied by the Department of Meteorological Services Malaysia. Kota Bharu is located in the northeastern region of Peninsular Malaysia at a latitude of 6 degrees 10 minutes, longitude of 102 degrees 17 minutes. It is 5 meters above sea level. The mean annual rainfall for this period (1951–84) was 2961 mm from an average of 138 rain days.

RESULTS

Rainfall data are often analysed by calculating totals on an annual, monthly, 10- or 7-day basis and using these totals as the 'raw' data for subsequent summaries. We first reviewed methods by which such totals could be processed. These methods are equally applicable to the other characteristics, such as the start and the end of the rains.

Analysis of Rainfall Totals

Initial time series plots are useful to indicate any

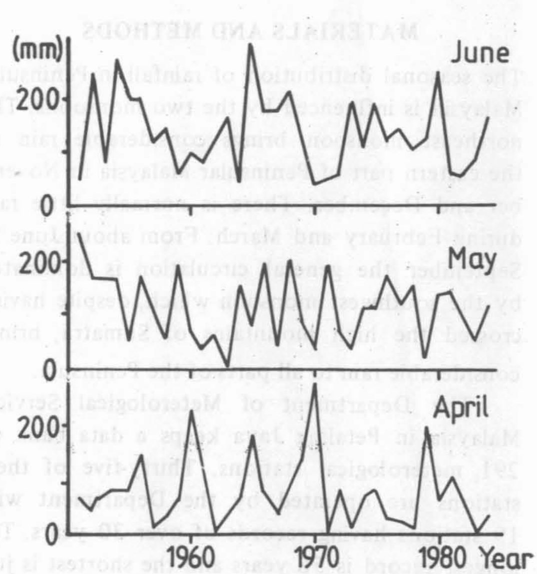

Fig. 1: Monthly totals for April, May and June at Kota Bharu 1951–84 (mm).

trends or obvious serial correlation in the data. *Figure 1* which uses the rainfall totals for April, May and June is one illustration. No structure with time was apparent and hence simple summaries of the data should be adequate.

The boxplots shown in *Figure 2* are a convenient summary of the pattern of rainfall through the year. In *Figure 2*, the main 'box' in each month is from roughly the lower to the upper quartile of the data. The line within the box is for the median. The points marked '+' indicate values that are sufficiently far from the centre that these data should be examined in more detail in case they correspond to errors. Boxplots are simple examples of a move in statistics towards a greater emphasis on the exploration and presentation of data before summary models are fitted. Other examples in climatology are given by Graedel and Kleines (1985).

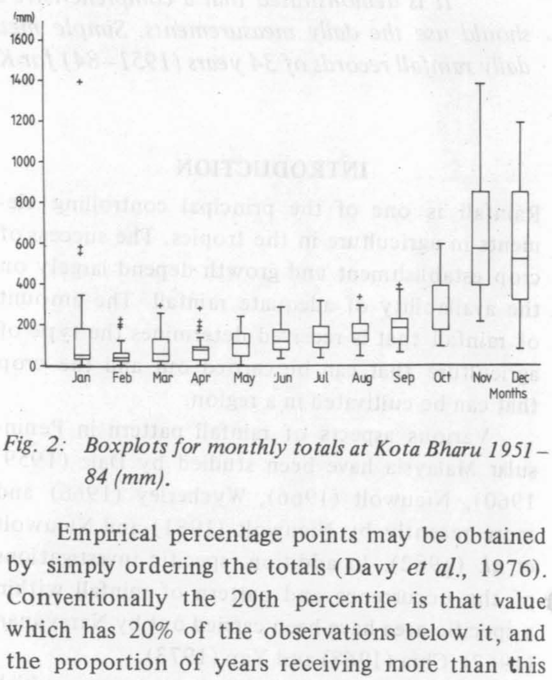

Fig. 2: Boxplots for monthly totals at Kota Bharu 1951–84 (mm).

Empirical percentage points may be obtained by simply ordering the totals (Davy *et al.*, 1976). Conventionally the 20th percentile is that value which has 20% of the observations below it, and the proportion of years receiving more than this value is thus 0.8. From Table 1 it can be seen that the 20% point for April totals at Kota Bharu was 26 mm. The median (i.e. 50% point) was 72 mm and the 80% point was 115 mm. This method of summarising the data is straightforward but, as indicated by the confidence interval in Table 1, the results are not precisely estimated. Confidence limits for these percentage points are only given rarely, although they are straightforward to calculate (Bradley, 1968). As an example, the

TABLE 1

Percentage points and confidence intervals for selected monthly and annual totals at Kota Bharu for 34 years.

Total used	Percentile (%)	Distribution-fitted[1] Amounts (mm)	Distribution-free method Amounts (mm)	95% Conf. Intv. (mm)
April	20	32	26	0 to 42
	50	70	72	38 to 86
	80	129	115	85 to 226
May	20	51	55	25 to 76
	50	97	115	70 to 155
	80	166	169	149 to 200
June	20	73	56	42 to 98
	50	120	121	86 to 167
	80	185	193	162 to 260
Annual	20	2245	2135	1846 to 2460
	50	2691	2627	2406 to 2989
	80	3137	3240	2963 to 3588

1: Gamma distribution fitted to monthly totals and normal distribution fitted to annual totals.

approximate 95% confidence intervals for the 20% point for April totals is 0 mm to 42 mm. The width of such intervals depends on the number of years of data. It is important that this uncertainty is allowed for when interpreting results, particularly when comparing the pattern of rainfall at different sites.

It is alternatively possible to fit a suitable distribution to the rainfall totals and then to estimate percentage points from this distribution Annual rainfall totals are often found to be approximately normally distributed. *Figure 3* shows that the histogram of annual totals for Kota Bharu is fairly symmetrical with a mean of 2691 mm and a standard deviation of 530 mm. Using the properties of normal distribution, selected percentage points are calculated (Table 1). The estimates are similar to those from the distribution-free method. For example, the 20% point is estimated as 2245 mm and 2135 mm for the two methods, respectively.

The monthly data at Kota Bharu based on the 34 observations derived from each year on the

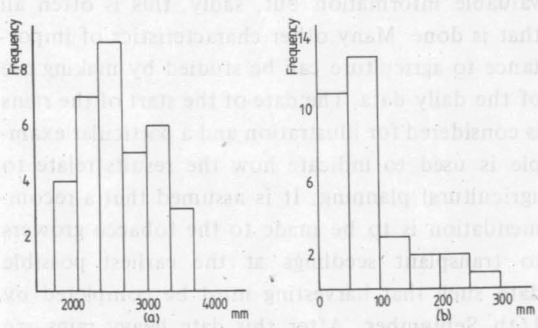

Fig. 3: Histograms of rainfall totals (a) annual and (b) monthly at Kota Bharu 1951–84 (mm).

record, were clearly not symmetrically distributed (e.g *Figure 3*). Gamma distributions are often used to fit such data. The results of fitting a gamma distribution to selected monthly rainfall totals at Kota Bharu are given in Table 1. As an example, the 80% point of April totals is estimated as 129 mm after fitting a gamma distribution with a mean of 85 mm and a *k* value of 1.76.

The assumption of normality is not appropriate for 10-day totals. Transformation to normality can be used (Hill and Morgan, 1981). However, this is rarely statisfactory because different transformations are usually required for different periods of the year. Some results from using the distribution-free approach to estimate various percentage points for 10-day totals are shown in *Figure 4.* An alternative is again to fit a gamma distribution to the non-zero totals.

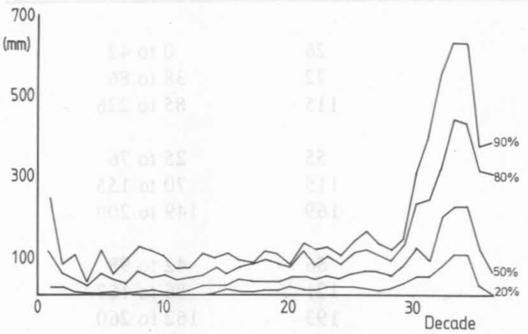

Fig. 4: Percentage points for 10-day rainfall total from Kota Bharu 1951–84.

Start of the Rains

The analysis of rainfall totals is useful and provides valuable information but, sadly, this is often all that is done Many other characteristics of importance to agriculture can be studied by making use of the daily data. The date of the start of the rains is considered for illustration and a particular example is used to indicate how the results relate to agricultural planning. It is assumed that a recommendation is to be made to the tobacco growers to transplant seedlings at the earliest possible date such that harvesting must be completed by 16th September. After this date heavy rains are anticipated, resulting in poor leaf quality.

It is assumed that the growth period of tobacco from transplanting to harvesting is approximately 110 days. Hence the latest date to transplant is 29th May. Different definitions of the start of the rains may be appropriate for different soils or to investigate the sensitivity of the results to the definitions used. Here the first definition defines the date as the first occasion after 1st March with 30 mm or more rain within a 3 day period of which at least 2 days have rain. The second definition is the same but with 1st April as the earliest possible transplanting date.

The results are summarised in Table 2. If transplanting is done as soon as possible after 1st March then, for example, the median transplanting date is 26th March. There is a possible transplanting date before the deadline of 29th May in all but 2 of the years. By comparison, if transplanting is not undertaken until 1st April, then the median transplanting date is 1st May and there are 6 years which do not meet the deadline. These results merely use a 'distribution free' summary of the data. As with the rainfall totals, it is alternatively possible to fit a distribution to the planting dates and then derive results from this distribution. The results above suggest that the 1st March definition is preferred. However in the next section we consider the risk that the transplanting falls, which may change the outcome.

Dry Spells

Dry spells can be considered in their own right (e.g. Stern *et al.*, 1982: Stern and Coe, 1982) but are also useful when considering the start of the rains. For illustration the latter is considered. We assume tobacco seedlings may need replanting if a long dry spell occurs in the first 30 days after transplanting. A 10-day dry spell is chosen and the definitions considered in the previous section are extended. Thus the date of successful transplanting is defined as the first occasion after 1st March (or 1st April) with more than 30 mm of rain in 3 days of which at least 2 have rain and there is no dry spell of 10 or more days in the next 30 days.

The overall results from these definitions are also given in Table 2. For example, with the 1st March definition, there is no successful transplanting date before the 29th May deadline in about a quarter of the years, compared with about one-third if 1st April is the earliest date. However, a comparison of the dates with and without the dry spell condition shows that the first transplanting failed in almost half of the years with the 1st March definition as compared to about one third with the 1st April. Thus there are alternative risks to be compared when trying to decide on a suitable transplanting strategy.

The above analysis can be taken further since

TABLE 2

Summary of results calculated of the distribution of the dates of the start of the rains from the transplanting of tobacco seedlings

	Transplanting done after			
	1 March Conditional		1 April Conditional	
	No	Yes	No	Yes
Earliest date	1 Mar	7 Mar	1 Apr	1 Apr
20th percentile	9 Mar	22 Mar	9 Apr	30 Apr
50th percentile	26 Mar	4 May	1 May	19 May
80th percentile	5 May	1 Jun	25 May	20 Jun
Latest date	29 Jun	7 Sep.	12 Aug	7 Sep
No. of years when transplant done after 29 May	2 (6%)	8 (24%)	6 (18%)	12 (35%)
No. of years needing replanting	16 (47%)		11 (32%)	

the two definitions used became identical in any year when transplanting is not possible until after 1st April. Hence to compare them with the 1st March definition, it is useful to consider that, there was a transplanting date prior to 1st April in 19 of the 34 years and that replanting was required in 10 of these. In this case the risk of failure is not substantially higher for the early transplanting dates (i.e. prior to 1st April) compared to the later dates. The difference in risk is not statistically significant but this may be due to the short length of record. The conclusion from this analysis might be to recommend the strategy which uses the earlier transplanting date.

End of the Growing Season

Definitions for the end of the rainy season normally cannot be based solely on rainfall since a measure of soil water storage is needed to define the end of the period of crop growth more realistically. Examples are given in Stern *et al.* (1982). When that start and end of the rains have been satisfactorily defined, the length of the rains may be obtained by subtraction. The pattern of the distribution of the rainfall lengths can then be investigated to provide a guide to the selection of crop varieties.

DISCUSSION

A comprehensive analysis of rainfall measurement for agricultural purposes should start with the daily data. Many aspects of the rainfall regime such as the risk of dry spells require this level of detail. Some, such as the chance of errosive rainfalls, can benefit from the even greater detail provided by continuously recording gauges. Unfortunately, published reports on rainfall data are often based on five-day, ten-day, or monthly periods. This applies also to the Malaysian rainfall data except for a few studies. Narayanan (1967) conducted an analysis of daily rainfall data to look for pattern in variation of rainfall in the Sungei Buloh area. Daily ranfall data from the same area were used by Yap (1973) to extract spell lengths and fit distributions to spell lengths. Nieuwolt (1981) calculated an Agricultural Rainfall Index (A.R.I.) to classify regions of Peninsular Malaysia as having long or short term drought and a month as dry or wet. His classifications were later used together with the combined effects of soil and landform on agricultural activities to establish agro-ecological regions in Peninsular Malaysia (Nieuwolt *et al.*, 1982). These studies clearly recognised the importance of analysing daily rainfall data from which the results should be useful in agricultural planning.

Our aim in this paper is to demonstrate a simple general method of summarising daily rainfall data which does not involve any new theory. It is merely a recognition that a comprehensive analysis of rainfall data can include the summary of characteristics that are not obtainable from rainfall totals and that the methods for processing these characteristics are identical to those used in the processing of rainfall totals.

Some limitations of this method of analysis must also be recognised. It involves finding the value of the characteristics of interest for each year. Hence, for each characteristic the data for each year provides just one number. Thus, the only way to have a large sample is to have a record that extends over many years. The method is simple, but pays for its simplicity by requiring long records. An example of this limitation was illustrated above, where the comparison between the two criteria for transplanting could only be tentative, from the 34 year record. For further discussion on the direct method of analysis see Stern *et. al.* (1982).

The limitations of the direct method of analysis, particularly given the relatively short records of rainfall data in Malaysia, encourage us to seek an alternative approach. One example of an alternative approach is described by Stern et al. (1982). This alternative approach emphasizes the efficient use of the data, permitting a useful analysis of short records. The approach is to model the daily rainfall itself and, from the model, to derive probabilities of rainfall events. The modelling approach fitted to Malaysian rainfall data should be investigated.

REFERENCES

BRADLEY, J.V. (1968): Distribution-Free Statistical Test. New Jersey: Prentice-Hall.

CHIA, L. S. (1968): An analysis of rainfall patterns in Selangor. *J. Trop. Geogr., Singapore,* **27:** 1–18.

DALE, W. L. (1959): The rainfall of Malaya, Part I, *J. Trop. Geogr., Singapore,* **13:** 30–32.

DALE, W. L. (1960): The rainfall of Malaya, Part II. *J. Trop. Geogr., Singapore,* **14:** 11–28.

DAVY, E. G., F. MATTEI, and S. I. SOLOMAN, (1976): An Evaluation of Climate and Water Resources for Development in Agriculture in the Sudano-Sahelian Zone of West Africa. Special Environmental Report 9, World Meteorological Organisation, Geneva, Switzerland. WMO-No. 459.

GRAEDEL, T.E. and B. KLEINER (1985): Exploratory Analysis of Atmospheric Data. *In* Murphy, A. H. and Katz, R. W. (Ed.): Probability, Statistics and Decision Making in the Atmospheric Sciences. Westview Press, London. p. 1–43.

HILLS, R. C., and J. H. T. MORGAN (1981): An interactive approach to the analysis of rainfall records for agricultural purposes. *Expl. Agric.* **17:** 1–16.

NARAYANAN, R. (1967): Statistical analysis of rainfall records at the PRIM Experimental Station. *J. Rubb. Res. Inst. Malaya, Kuala Lumpur,* **20:** 8–22.

NIEUWOLT, S. (1966): A comparison of rainfall during 1963 and average conditions in Malaya. Erdkunde, **20:** 171–174.

NIEUWOLT, S. (1981): Agricultural Droughts in Peninsular Malaysia. M.A.R..D.I., Serdang, 14 pp.

NIEUWOLT, S., M. ZAKI GHAZALI, and B. GOPINATHAN (1982): Agro-Ecological Regions in Peninsular Malaysia. M.A.R.D.I., Serdang. 20 pp.

STERN, R. D. and R. COE (1982): The use of rainfall models in agricultural planning. *Agric. Meteor.* **26:** 35–50.

STERN, R.D., M.D. DENNETT and I.C. DALE (1982): Analysis of daily ranfall measurements to give agronomically useful results. I. Direct methods. *Expl. Agric.* **18:** 223–236.

WYCHERLEY, P. R. (1966): Rainfall Probability Tables for Malaysia. Rubber Research Institute of Malaya. Planting Manual No. 12.

YAP, W.C. (1973): The persistence of wet and dry spells in Sungei Buloh, Selangor. *Meteorological Magazine,* **102**. 240–245.

(Received 24 April, 1987)