EARLY DETECTION OF THE START OF THE WET SEASON IN TROPICAL CLIMATES

IAN T. JOLLIFFE

Department of Mathematical Sciences, University of Aberdeen, AB9 2TY, UK

AND

DORIS E. SARRIA-DODD

Applied Statistics Research Unit, University of Kent, Canterbury, CT2 7NF, UK

Received 10 February 1992
Accepted 9 March 1993

ABSTRACT

In many tropical and sub-tropical climates the year is divided into well-defined wet and dry seasons. To maximize agricultural production it is important that crops are planted as early as possible in a wet season, but deciding exactly when the wet season has started may be a problem. A decision can be made on the basis of previous years' data, or we may try to incorporate information from the current season. The latter approach is discussed in the present paper, and shown to be potentially of greater value than relying solely on historical data.

KEY WORDS Discriminant analysis Early detection Wet season

1. INTRODUCTION

In tropical and subtropical climates with distinct wet and dry seasons, it is important to decide when a wet season has started, especially for farmers who depend on the rains for irrigation purposes. If planting of crops is delayed for too long after the start of the rains then the remaining growing season can be shorter than is ideally required, leading to a substantial reduction in yield of crops. Conversely, if crops are planted too early, and there is a subsequent dry spell, the plants may then die, necessitating a second planting.

In order to decide when a wet season has started, there is clearly a need for a definition of a 'wet season start'. A number of such definitions have been proposed in the literature and will be discussed in Section 3 of this paper. Having defined the 'start of the wet season', we would like to be able to predict or detect when this start occurs. We draw a distinction here between 'prediction' and 'detection' as follows. Prediction is the more commonly used approach and relies mainly on historical data, largely ignoring data from the current season. For example, Stern *et al.* (1981) define the 'start of the rains' as the first occurrence of a specified amount of rain (10 mm, 20 mm, or 30 mm) within two successive days, but this definition is qualified by the requirement that the two days must not be followed by a continuously dry period of specified length (5, 7, or 10 days). With a sufficiently long record of data, it is possible to estimate the probability of the definition being satisfied on or before any chosen date. It would then be feasible to recommend planting on the first date for which this probability is sufficiently large.

This strategy of planting on a fixed date is what might be termed 'pure prediction' of wet season onset and is, of course, not very sensible as it ignores the weather on the chosen date in any given year. Stern *et al.* (1981) advocate an approach that estimates the probability of a dry period of specified length following any given date. The first date for which this probability is sufficiently low is taken to be the earliest possible planting date. The first 2-day spell with more than a specified amount of rain, following this earliest possible date, then

0899–8418/94/010071–06$08.00

defines the start of the rains. This approach is still effectively 'prediction' because the decision on whether a particular 2-day wet spell can be taken as the start of the wet season depends only on probabilities calculated from previous years' data.

It seems reasonable to believe that the current year's pattern of weather has some value in detecting whether an apparent start to a wet season, such as a 2-day wet spell, is a true or false start. We call the use of current season's data in this way 'detection' rather than prediction, and the remainder of the paper is concerned with investigating whether detection is possible. Section 2 describes the data sets that we use in our investigation, and section 3 discusses our methodology, and the software used. Section 4 gives the results, which indicate that use of daily rainfall data from the end of the dry season can be of value in detecting the onset of the wet season. Finally, section 5 discusses the significance of our results and possibilities for further work.

2. THE DATA

Data from three climatological stations were used in this investigation. Details of each data set are summarized in Table I. Two of the stations, Masaka and Kabanyolo, are very close to the Equator, and have two wet seasons, each of which may be looked at separately. The third station, Chingombe, as well as having only a single wet season, has the lowest annual rainfall of the three. All three are at similar altitudes. It is appreciated that results for only three stations cannot be generalized to all tropical rainfall stations (there is further discussion of this point in section 5). We concentrate our discussion on Chingombe, which has the longest record, using the other stations to demonstrate that the rules selected for Chingombe may be of wider value.

Table I. Details of three rainfall stations

Station	Location	Altitude (m)	Mean annual rainfall (mm)	Data availability
Chingombe (Zambia)	14°25'S, 29°57'E	1000	921·5	1961–1983
Kabanyolo (Uganda)	0°28'N, 32°37'E	1312	1336·7	1963–1972[a]
Masaka (Uganda)	0°19'S, 31°44'E	1250	1181·1	1963–1972[b]

[a] 1968 second rainy season was excluded from the analysis due to missing data.
[b] 1966 and 1968 second rainy seasons were excluded from analyses due to missing data.

3. METHODOLOGY

In order to investigate whether it is possible to detect the onset of the wet season using data from the current season, it is necessary to:

(i) define the start of a wet season;
(ii) define a potential or apparent start to a wet season;
(iii) develop a technique for deciding when a potential start to a wet season is a true start and when it is a false start.

Each of these three aspects will now be discussed in turn.

3.1. *Definition of the start of a wet season*

Many attempts have been made in the literature to define when a rainy season has begun. To some extent we must expect different definitions in different geographical areas, or even in different seasons at the same station for stations with two wet seasons in a year. However, similar elements are likely to occur in most definitions.

Two important ingredients which are almost certain to be included are a sufficiently high rainfall in a period of specified length, and an absence of subsequent dry periods exceeding a certain length. The definitions used by Stern *et al.* (1981), which were discussed in section 1, are of this type.

There are also definitions that use relative, rather than absolute, rainfall values. For instance, a wet month should have a certain percentage of the annual rainfall, or the onset of the rains occurs when the total rainfall of two consecutive months differs significantly (Gregory (1969), cited by Hulme (1987), who also reviews a number of other definitions). Nicholls (1984) defines the start of the rainy season in North Australia as the date when 15 per cent of the mean annual station rainfall had accumulated, from a starting day of accumulation of 1 September.

Other, more sophisticated, definitions of wet season onset include concepts such as potential evapotranspiration, soil moisture, water balance, and crop water requirements. Hulme (1987) defines a model for the wet season in central Sudan using daily rainfall and potential evapotranspiration, assuming a minimum 10-day period of positive water balance to open the wet season. Benoit (1977) defines the start of the growing season in northern Nigeria as the date when the accumulated rainfall exceeds and remains greater than one-half of potential evapotranspiration for the remainder of the growing season, provided that no dry spell longer than 5 days occurs immediately after this date.

Jackson (1982) suggests that some traditional indicators of the start of the rainy season, such as high or increasing temperatures, thunder, lightning, arrival or departure of bird and insect species, etc., should be taken into account when trying to understand local climatology, and could improve the forecasting of tropical rainfall. However, it would be difficult to quantify many of these indicators.

Our own definition is based on those of Stern *et al.* (1981). Specifically, the rainy season has started when:

(i) a period of 5 days with at least 25 mm of rainfall occurs;
(ii) the start day and at least two other days in this period are wet (i.e. $\geqslant 0{\cdot}1$ mm rainfall received);
(iii) no dry period of 7 days or more occurs in the following 30 days.

It was decided to use a 5-day period (pentad) with at least 3 wet days, since in the tropics it is common to find heavy rains in a short period of time; to use only two consecutive days as Stern *et al.* (1981) suggest increases the chance of a false start. In some cases two consecutive wet days could correspond to a single rainfall event overlapping two days.

3.2. *Definition of a potential start*

Given our definition of the start of a wet season, there is a fairly obvious definition of a potential start, namely that conditions (i) and (ii) are satisfied.

In order to find the true start, as well as apparent starts, for any particular data set, we have used the EVENTS program (Hopkins and Burt, 1989). EVENTS was originally written at the University of Reading, and later rewritten and enhanced at the University of Kent for the Overseas Development Administration. It is designed to search records of daily rainfall for 'events' of varying complexity. For example, our definition of the onset of the wet season constitutes an 'event' as defined by the program.

3.3. *Discrimination between false and true starts*

Having established what we mean by a potential start to the wet season, we now investigate whether there is any way of discriminating between those potential starts that are true starts and those which turn out to be false starts. The variables that we use as possible predictors or discriminators are listed in Table II. It is seen that they are of two types, numbers of wet days, and total amounts of rain in fixed periods before the potential start.

The four variables listed in Table II were used in a linear discriminant analysis (see Krzanowski, 1988, p. 340), as implemented by the DISCRIM and STEPWISE procedures in SAS, Version 5.18 (SAS Institute Inc., 1985). In DISCRIM, we specify certain subsets of the variables to be included in the analysis, and the procedure then constructs the linear combination of these specified variables that best discriminates between

Table II. Variables used in linear discriminant analysis

Variable	Detail
W-20	Number of wet days in the 20 days preceding the potential start
R-20	Amount of rainfall in the 20 days preceding the potential start
W-10	As for W-20, but for 10 days
R-10	As for R-20, but for 10 days

Table III. Percentage of correctly classified starts for Chingombe (31 observations)

Variables used	Percentage correct classifications	Priors
All four	81	Equal
All four	77	Sample
W-10, W-20	84	Equal
W-10, W-20	77	Sample
W-10, W-20, W15	87	Equal
W-10, W-20, W15	84	Sample

true and false starts. In STEPWISE we do not specify particular subsets of variables. Instead, the procedure carries out a partial search in an attempt to find a subset of the original four variables that best discriminates, using linear combinations of the subset, between true and false starts.

Two choices need to be made in implementing these procedures. First, we need to decide whether or not to assume that true and false starts are equally likely. If this assumption is made, we denote the resulting rule as 'Equal Priors' in Table III. Alternatively, if discriminant rules are based on sample proportions of the true and false starts, they are denoted by 'Sample Priors' in Table III.

Second, we have the option of making an assumption of equal covariances for the variables in the two populations (true and false starts). If the covariances are different in different populations, we end up with quadratic, rather than linear, discrimination (Krzanowski, 1988, p. 340). However, perhaps partly because of our small sample sizes, there was little statistical evidence to suggest inequality of covariances for most analyses, so in the results reported below we assume equality and use linear discriminant analysis.

4. RESULTS

4.1. Chingombe

Table III summarizes the percentages of correctly classified starts for various subsets of variables. With all four variables we achieve an 81 per cent correct classification using equal priors, and 77 per cent correct classification using priors based on sample proportions. STEPWISE selects the subset W-10, W-20, which performs as well as the full set of variables with respect to correct classifications for sample priors and actually produces an improvement for equal priors. The other selection given in Table III, W-10, R-20, W-15 will be explained in Section 5.

4.2. Kabanyolo and Masaka

Kabanyolo and Masaka each has two wet seasons and it might not be surprising if different rules were needed for each season. Unfortunately, because of the small size of the data sets, we cannot draw any firm conclusions regarding the performance of rules derived from these data themselves. However, what we can do is to use these stations as independent data on which to partially confirm or refute the possible value of rules found for Chingombe. What we find is that using all four variables the percentage of correct classification ranges from around 70 per cent (Kabanyolo and Masaka, season 2) to 100 per cent (Kabanyolo, season 1). If we use only W-10 and W-20, the results are rather patchy. Masaka, season 2, stays at 70 per cent, but other results are worse than for the full set of variables. Masaka, season 1, assuming equal priors, has 86 per cent correct classification, and Kabanyolo, season 1, based on sample priors, achieves 92 per cent, but with an equal priors assumption this drops to 75 per cent, and Kabanyolo, season 2, only manages 56 per cent for either assumption.

The overall impression is that these results do not provide convincing evidence of the value of a rule based only on W-10, W-20, but they do at least suggest that such a rule is worthy of further investigation.

5. DISCUSSION

We start with a discussion of the two final lines in Table III. As well as the variables defined in Table II, we also considered similarly defined variables for periods immediately after the potential start. It is clearly preferable to base detection rules on data gathered only before the potential start, but recall our definition of a true start. We need to wait for 30 days (or at least 26) to be certain that a potential start is a true start, so that if information gathered in the first part of the 30 days period can improve our detection performance, then it is worthwhile using it. A stepwise discriminant analysis using 10 variables, the four in Table II plus W-5, W-10, W-15, R-5, R-10, R-15, with fairly obvious notation, selected the three variables W-15, W-10, W-20 and the results for this are given in Table III. There is a clear improvement in the probability of correct classification by including W-15. Similarly, with one small exception, the inclusion of W-15 improves performance for all the cases considered for Masaka and Kabanyolo. Thus incorporation of post-potential-start information may have some merit; whether it is worth including such information depends on the costs of waiting longer to plant compared to the potential gains in correctly detecting true and false starts.

The results obtained so far indicate that numbers of wet days preceding and immediately following a potential start of a wet season may have predictive value in deciding whether the 'start' is a true or a false one. Because of the relatively small sample sizes, and the selection effects inherent in the discriminant analysis, it is unwise to draw definitive conclusions. However, there is a sufficiently consistent picture for us to be cautiously optimistic that our detection rules have real merit. There is certainly a strong indication that it is possible to improve on predictions that do not incorporate the current season's data. Investigation of similar rules on further data sets will be undertaken in the future.

We could also consider including variables in the discriminant analysis other than those based on rainfall, but rainfall data are far more often available than other meteorological variables, so that inclusion of other such variables will make any rules of less practical use. However, one additional variable that may be included routinely is the date of the potential start. This could be simply incorporated in the discriminant analysis using an appropriate variable representing the date. Alternatively, the date could be used indirectly by letting prior probabilities of true and false starts vary with the date, where these prior probabilities could be calculated using similar methodology to Stern *et al.* (1981). It should, however, be remembered that, unless very large data sets are available, increasing the numbers of variables may seriously increase the chances of over-fitting the data.

ACKNOWLEDGEMENTS

The research was partially supported by a grant from the University of Kent Research Committee. We are grateful to John Dancer for originally suggesting the topic to us, and for the Kabanyolo and Masaka data sets. Thanks are also due to Alan Sykes for supplying the Chingombe data set. Finally, we are grateful to two referees, and particularly to the editor, for constructive comments on an earlier version of this paper.

REFERENCES

Benoit, P. 1977. 'The start of the growing season in Northern Nigeria', *Agric. Meteorol.*, **18,** 91–99

Gregory, S. 1969. 'Rainfall reliability', In Thomas, M. F. and Whittington, G. W. (eds.), *Environment and Land Use in Africa*, Methuen, London, pp. 63–82.

Hopkins, T. and Burt, L. 1989. '*User Guide to the EVENTS program*', Applied Statistics Research Unit, University of Kent.
Hulme, M. 1987. 'Secular changes in wet season structure in central Sudan', *J. Arid. Environ.*, **13**, 31–46
Jackson, I. J. 1982. 'Traditional forecasting of tropical rainy seasons', *Agric. Meteorol.*, **26**, 167–178.
Krzanowski, W. J. 1988. *Principles of Multivariate Analysis: A User's Perspective*, Oxford University Press, Oxford, 563 pp.
Nicholls, N. 1984 'A system for predicting the onset of the North Australian wet season' *J. Climatol.*, **4**, 425–435.
SAS Institute Inc. 1985. *SAS User's Guide: Statistics, Version 5 Edition*, SAS Institute Inc., Cary, Nc,956 pp.
Stern, R. D., Dennett, M. D. and Garbutt, D. J. 1981. 'The start of the rains in West Africa', *J. Climatol.*, **1**, 59–68