INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **21**: 1251–1262 (2001)
DOI: 10.1002/joc.640

EARLY DETECTION OF THE START OF THE WET SEASON IN SEMIARID TROPICAL CLIMATES OF WESTERN AFRICA

DORIS E.S. DODD[a] and IAN T. JOLLIFFE[b,*]
[a] *ASRU Ltd, SABP, Swalecliffe, Kent, UK*
[b] *Department of Mathematical Sciences, University of Aberdeen, UK*

Received 29 September 1999
Revised 15 December 2000
Accepted 15 December 2000

ABSTRACT

An earlier paper (Jolliffe IT, Sarria-Dodd DE. 1994. *International Journal of Climatology* **14**: 71–76) investigated the problem of deciding when the wet season has started in tropical and sub-tropical climates. In particular, methodology based on linear discriminant analysis was developed for using data from the current season to make the decision, rather than relying only on information from previous seasons. It was shown, for three stations in eastern Africa, that the methodology was potentially valuable. The present study is much larger, using data from 24 stations, covering a range of annual rainfall totals, in western Africa. It is confirmed that linear discriminant analysis can indeed be useful in detecting when the wet season has started, and hence in deciding when to plant crops. As well as being a larger analysis than that reported previously, the present study also extends the previous work by investigating an alternative definition of the start of the wet season and by including 'date' as a potential explanatory variable. Copyright © 2001 Royal Meteorological Society.

KEY WORDS: early detection; linear discriminant analysis; rainy season; western Africa; wet season

1. INTRODUCTION

In semiarid tropical climates with distinct wet and dry seasons, for example in Africa, the date of the start of the wet season is a crucial factor in deciding when to plant crops. Planting too early may lead to crop failure, whereas planting too late may reduce the growing season and hence crop yield. Detecting when the wet season has actually started is therefore important, and any reliable guidance on this is of great value.

One type of strategy examines past climatic records for relevant weather stations and finds the first date by which the wet season has started in a specified high proportion of previous years. More sophisticated variants of this approach have been proposed (for example, Stern *et al*., 1981), but most still rely mainly on previous years' data in making decisions about wet season onset. In an earlier paper (Jolliffe and Sarria-Dodd, 1994, subsequently referred to as JSD) we showed that incorporating information on the current season's pattern of rainfall was potentially useful. JSD described a small study, based on only three stations in eastern Africa. The present work is much more extensive, and uses data from 24 stations in western Africa; it also investigates an alternative definition of the start of the season, and incorporates 'date' as an additional explanatory variable. The results of the present study confirm the earlier promise of the methodology discussed in JSD.

In Section 2 the data are described, and the methodology for detecting when the wet season has begun is discussed in Section 3. This discussion includes definitions of 'wet season start'. The results of our analyses are given in Section 4, and the paper ends with further discussion and conclusions in Section 5.

* Correspondence to: Department of Mathematical Sciences, University of Aberdeen, Aberdeen, AB24 3UE, UK; e-mail: itj@maths.abdn.ac.uk

2. THE DATA

The data analysed in this study consist of daily rainfall records for 24 stations in western Africa. Of these, 14 are in Burkina Faso, eight in Mali, and one each in Niger and Senegal. The data were obtained from Dr Alan Robock of the University of Maryland, and came originally from Dr M.V.K. Sivakumar of ICRISAT in Niamey, Niger.

Figures 1–3 show the location of the 24 stations, and Tables I and II, respectively, give further details of each station in Burkina Faso and Mali. The tables provide information on average annual rainfall and years of data availability. The 22 stations listed have between 40 and 77 years of daily data, with average annual rainfall ranging from more than 1000 mm at five stations in the south to below 700 mm at two stations in the northeast. The annual distribution of rainfall is unimodal throughout the area, with July–September as the wettest months.

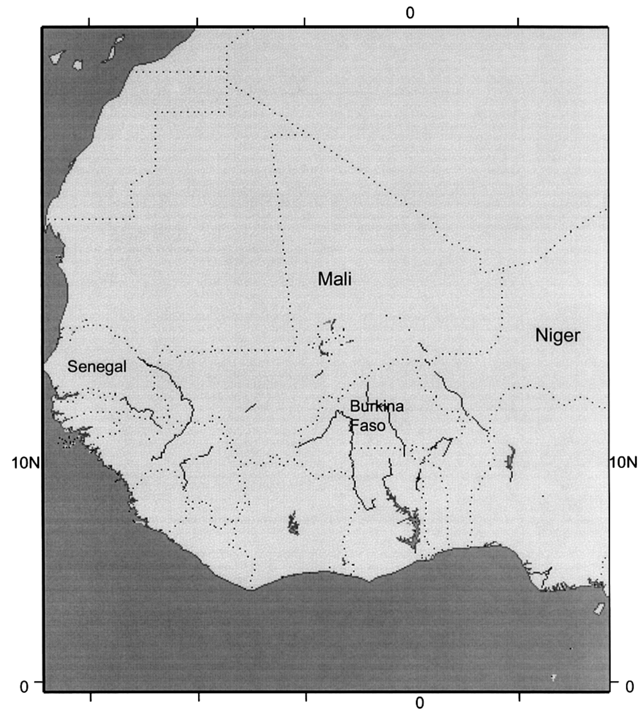

Figure 1. Study area

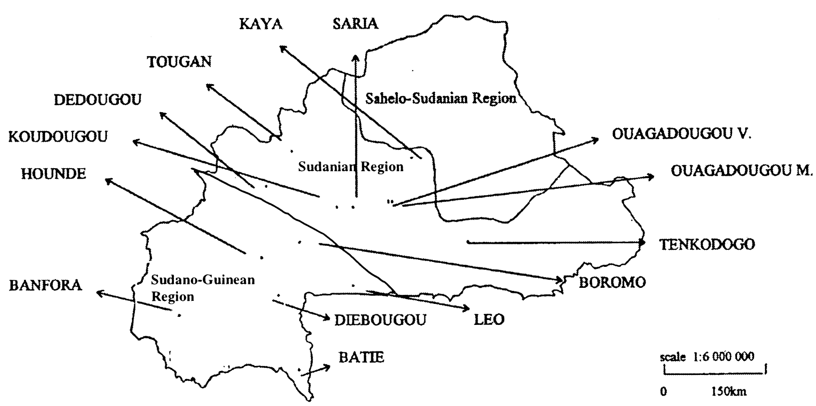

Figure 2. Location of meteorological stations in Burkina Faso

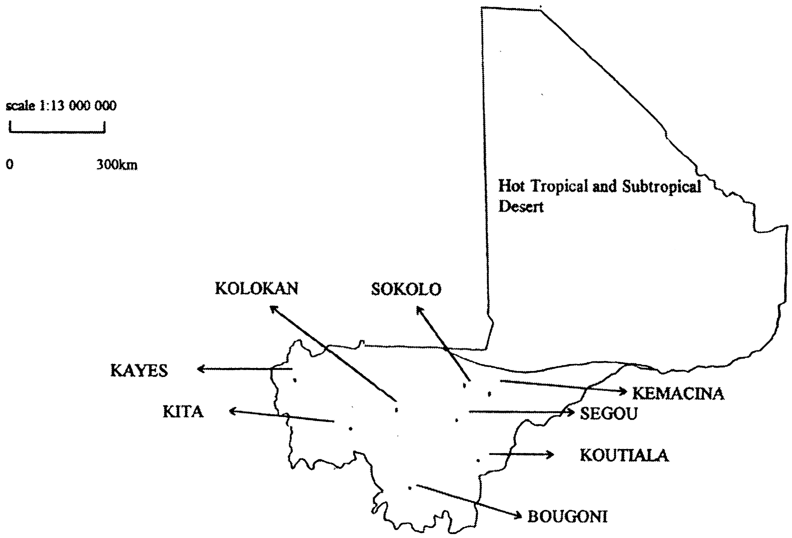

Figure 3. Location of meteorological stations in Mali

3. METHODOLOGY

To determine whether or not it is possible to detect the onset of the wet season using data from the current season, we need to consider three things:

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

Each of these aspects will now be discussed in turn.

Table I. Details of Burkina Faso meteorological stations

Station	Annual rainfall (mm)	Years with available data*	Total no. years
BANFORA	1140.7	1922–1928, 1930–1963, 1965–1967, 1969–1985	61
BATIE	1154.6	1945–1977, 1980–1982, 1984–1987	40
BOROMO	963.7	1923–1930, 1933–1937, 1939–1949, 1951–1976, 1978–1982	55
DEDOUGOU	909.2	1922–1984	63
DIEBOUGOU	1069.0	1923–1953, 1956–1977, 1979–1984, 1986	60
HOUNDE	991.8	1923–1928, 1931–1987	63
KAYA	696.8	1921–1961, 1963–1972, 1974–1984	62
KOUDOUGOU	851.7	1922–1927, 1929–1931, 1933–1977, 1979–1982	58
LEO	987.5	1921, 1923, 1925–1926, 1928–1939, 1941–1964, 1966–1982, 1984–1987	61
OUAGADOUGOU M.	794.8	1902–1948, 1952–1954, 1956, 1958–1972, 1974–1984	77
OUAGADOUGOU V.	858.3	1921–1922, 1924–1925, 1927–1931, 1933–1975, 1977–1979	55
SARIA	806.0	1944–1985	42
TENKODOGO	931.6	1922, 1926, 1928–1983	58
TOUGAN	732.2	1934–1941, 1943–1965, 1967–1977, 1979–1986	50

* Years with missing data around the onset of the rainy season were disregarded.

Table II. Details of Mali meteorological stations

Station	Annual rainfall (mm)	Data availability	Total no. years
BOUGONI	1249.0	1922–1973, 1982, 1984–1986, 1988–1989	58
KAYES	724.6	1926–1930, 1932, 1934–1980, 1984–1985, 1987–1989	58
KEMACINA	559.9	1929–1962, 1964–1980, 1989	52
KITA	1070.7	1931–1980, 1983–1985, 1988–1989	55
KOLOKAN	786.2	1923, 1925–1926, 1928, 1931–1961, 1963–1985, 1988	59
KOUTIALA	966.7	1922–1946, 1948–1980, 1982–1986, 1988	64
SEGOU	701.4	1922–1928, 1936–1973, 1975–1980, 1984–1985, 1988–1989	55
SOKOLO	464.0	1937, 1941–1966, 1974–1982, 1984–1987, 1989	41

3.1. Definition of the onset of the wet season

Over the years a variety of definitions have been suggested and JSD describe a number of these, due to Benoit (1977), Stern *et al.* (1981) and Hulme (1987)—see also Nicholls (1984) and Sivakumar (1988). Here we simply present those definitions implemented in the present study. We use two definitions—the first is the same as that of JSD, which was based on Stern *et al.* (1981), and states that the wet season has started when, for the first time since 1 March:

Definition 1

(a) a period of 5 consecutive days occurs in which at least 25 mm of rain falls;
(b) the start day and at least two other days in the period are wet (at least 0.1 mm rainfall recorded);
(c) no dry period of 7 or more consecutive days occurs in the following 30 days.

It was found that this definition tended to give onset dates in Burkina Faso which were later than seemed intuitively reasonable, relative to the usual time of crop planting, so we also examined a second, less demanding, definition. Condition (b) is unchanged, but (a) and (c) are replaced by:

Definition 2

(a′) a period of 6 consecutive days occurs in which at least 25 mm of rain falls;
(c′) no dry period of 10 or more consecutive days occurs in the following 40 days.

3.2. Definition of a potential onset

Again there are two definitions. We simply remove conditions (c) and (c’) from Definitions 1 and 2 above. For each dataset the true and potential onset dates are found using the EVENTS computer software (Hopkins and Burt, 1989).

3.3. Discrimination between true and false onsets

To decide which potential starts are likely to represent true onsets, and which are false starts, we use linear discriminant analysis with the 15 predictor variables listed in Table III. JSD considered only W10, R10, W20 and R20; other variables, similarly describing the rainfall *before* the potential start have been added here, as have PW5, PR5 denoting rainfall *immediately following* the potential onset. The date of the potential onset is also considered as a potential predictor.

Given a set of potential predictor variables $x_1, x_2, \ldots, x_p$, linear discriminant analysis looks for a linear combination $\mathbf{a}'\mathbf{x} = a_1x_1 + a_2x_2 + \cdots + a_px_p$ which discriminates as well as possible between real and false starts in the following sense. We can compute the variance of $\mathbf{a}'\mathbf{x}$ within groups, where the two groups are real and false starts; we can also define between-group variation of $\mathbf{a}'\mathbf{x}$ as the squared difference between the mean of $\mathbf{a}'\mathbf{x}$ in the two groups. The set of constants $a_1, a_2, \ldots, a_p$ is chosen to maximize the ratio of between- to within-group variation of $\mathbf{a}'\mathbf{x}$, thus defined. For more details, see Krzanowski (1988, p 340).

Table III. Variables used in linear discriminant analyses

Variable	Description
DATE	Day of the year in which the potential onset has started, i.e. For 1 January DATE is 1, and for 31 December DATE is 366
W5	Number of wet days in the **5** days **preceding** the potential onset
R5	Amount of rainfall in the **5** days **preceding** the potential onset
W10	As W5 but referring to **10** days
R10	As R5 but referring to **10** days
W15	As W5 but referring to **15** days
R15	As R5 but referring to **15** days
W20	As W5 but referring to **20** days
R20	As R5 but referring to **20** days
W25	As W5 but referring to **25** days
R25	As R5 but referring to **25** days
W30	As W5 but referring to **30** days
R30	As R5 but referring to **30** days
PW5	Number of wet days in the **5** day-period **following** the potential onset
PR5	Amount of rainfall in the **5** day-period **following** the potential onset

Linear discriminant analysis was implemented using the STEPDISC and DISCRIM procedures within SAS version 6.08 (SAS, 1989). In the DISCRIM procedure, subsets of the 15 variables to be included in the analysis must be specified, and the procedure then constructs the linear combination of these specified variables which best discriminates between true and false onsets. STEPDISC also finds optimally discriminating linear combinations of a subset of the variables, but no prior specification of which variables to include is needed. The procedure carries out a partial search of the subsets of the 15 variables in Table III, in a stepwise manner, and chooses a subset, which gives good discriminatory power.

Two choices must be made in implementing DISCRIM. First, we can either make an implicit assumption that true and false starts are equally likely (Option 'No proportional'), or take the underlying probability of a true start as equal to the proportion of true starts in the data (Option 'Proportional'). Both options were tried in this study. The second choice is whether to assume that the underlying covariances between our predictor variables are the same for true starts as for false starts. If this assumption cannot be made we end up with quadratic, rather than linear, discriminant analysis. There was little statistical evidence to suggest violation of an equal-covariance assumption, so linear discriminant analysis was used.

STEPDISC was implemented separately for each of the 22 stations in Burkina Faso and Mali, and the results examined to identify variables which were selected for many stations. Because 'DATE' is a variable of rather different nature from the remainder the analysis was repeated both including 'DATE' (15 variables) and excluding it (14 variables). For each station 28 DISCRIM analyses were conducted, as follows:

(i) using the variables selected by STEPDISC for that station from the 15 predictor variables;
(ii) as (i), but selecting from the 14 variables excluding DATE;
(iii) using the most commonly chosen single variable across stations in the same country;
(iv) using the most commonly chosen group of variables across stations in the same country;
(v) using variables selected for Batie (Bougoni) by STEPDISC, for all stations in Burkina Faso (Mali);
(vi) using variables selected for Dedougou (Kemacina) by STEPDISC, for all stations in Burkina Faso (Mali);
(vii) as (iv), but excluding PW5, PR5.

Each of these seven possibilities was repeated, using Definitions 1 and 2, and using the 'Proportional' and 'No Proportional' options, giving a total of 28 analyses for each station. Choices (ii) and (vii) were included to investigate the effect of the different types of variables, DATE, PW5 and PR5. Choices (iii), (iv) and (vii) attempt to avoid 'overfitting' and its consequent over-optimism, for individual stations, by looking for common rules across different stations. Analyses (v) and (vi) take this further by examining the stability of performance of a rule derived for a single station, when it is used on other stations in the same country. The stations whose rules are used in (v) have relatively high average annual rainfall, while those in (vi) have low average annual rainfall. Some of the rules were also tried out on single stations from the neighbouring countries of Senegal and Niger.

4. RESULTS

With a total of 24 stations, and 28 separate analyses for most of them, the results are too extensive to be reported in full here. Tables IV, V, VI, VII and VIII give a selection of results for stations in Mali—further details for Mali, together with corresponding information for Burkina Faso, are available in Dodd (1996).

Tables IV and V contain information on the number of real and false starts for each of the Mali stations, using Definitions 1 and 2 respectively. It is seen that using Definition 2 increases the number of real starts, but decreases the number of false starts, relative to Definition 1. On average, both true onset and first potential starts are about 12 days earlier using Definition 2 in Mali than when using Definition 1. The corresponding difference for Burkina Faso is about 20 days.

Table VI and Table VII display, respectively, the STEPDISC analyses (i) and (ii), using Definition 1. It is seen that DATE is often chosen when it is available. In the absence of DATE, W30 is a frequent choice. The only other predictor which is chosen for more than one or two stations is PR5.

Information on the percentage of correct identifications of true and false onsets, using Definition 1, is given for all stations in Mali in Table VIII. In the table, the column headed 'baseline' is the percentage of correct classifications if all potential starts are taken as real starts. As the number of false starts in a dataset decreases, the baseline becomes harder to beat. The next column in Table VIII lists the variables selected in each of the analyses (i)–(vii). There follows a block of three columns, giving for the 'Proportional' option the percentage of correct identifications of real starts, false starts and a combined percentage. The final three columns have equivalent information under the 'No Proportional' option. In our table these options are given the more informative labels 'Sample Probabilities' and 'Equal Probabilities', respectively.

We can summarize the main results in Tables IV, V, VI, VII and VIII, and the corresponding results for Burkina Faso (not shown), and for Definition 2 (not shown) as follows:

Table IV. Proportion of real and false starts derived from daily rainfall datasets in Mali using Definition 1

Station	Number of years	Real starts (%)	False starts (%)	Total no. of starts	Years when Real start failed*
BOUGONI	58	58 (70.73)	24 (29.27)	82	
KAYES	58	56 (71.79)	22 (28.20)	78	1928 1929
KEMACINA	52	45 (67.16)	22 (32.84)	67	1938 1946 1951 1953 1968 1977 1980
KITA	55	55 (84.62)	10 (15.38)	65	
KOLOKAN	59	59 (74.68)	20 (25.32)	79	
KOUTIALA	64	64 (72.73)	24 (27.27)	88	
SEGOU	55	54 (69.23)	24 (30.77)	78	1940
SOKOLO	41	30 (57.69)	22 (42.31)	52	1944 1948 1949 1953 1961 1977 1980 1981 1984 1985 1987

* This column explains the difference between 'Number of years' and 'Real starts'; i.e. in some years the rainy season failed to occur according to Definition 1.

Table V. Proportion of real and false starts derived from daily rainfall datasets in Mali using Definition 2

Station	Number of years	Real starts (%)	False starts (%)	Total no. of starts	Years when Real start failed*
BOUGONI	58	58 (89.23)	7 (10.77)	65	
KAYES	58	57 (86.36)	9 (13.64)	66	
KEMACINA	52	50 (80.65)	12 (19.35)	62	1977 1980
KITA	55	55 (88.71)	7 (11.29)	62	
KOLOKAN	59	59 (80.82)	14 (19.18)	73	
KOUTIALA	64	64 (84.21)	12 (15.79)	76	
SEGOU	55	55 (82.09)	12 (17.91)	67	
SOKOLO	41	38 (77.55)	11 (22.45)	49	1981 1984 1987

* This column explains the difference between 'Number of years' and 'Real starts'; i.e. in some years the rainy season failed to occur according to Definition 2.

Table VI. Variables chosen through STEPDISC using Definition 1 in Mali (significance level to enter and to stay $p \leq 0.15$)

Station	Variables* chosen by STEPDISC		
BOUGONI	**DATE** ($p < 0.0001$)	**W15** ($p = 0.0265$)	**R15** ($p = 0.0157$)
KAYES	**W30** ($p < 0.0001$)	**PR5** ($p = 0.0791$)	**W5** ($p = 0.0704$)
KEMACINA	**PW5** ($p = 0.0994$)	**R5** ($p = 0.1131$)	
KITA	**DATE** ($p < 0.0001$)	**R15** ($p = 0.0328$)	**W10** ($p = 0.0218$)
KOLOKAN	**DATE** ($p < 0.0001$)		
KOUTIALA	**DATE** ($p < 0.0001$)		
SEGOU	**DATE** ($p < 0.0001$)	**PR5** ($p = 0.1354$)	
SOKOLO	**PR5** ($p = 0.0020$)	**DATE** ($p < 0.0345$)	

* Refer to Table III for descriptions of variables.

Table VII. Variables chosen through STEPDISC using Definition 1 in Mali, excluding DATE (significance level to enter and to stay $p \leq 0.15$)

Station	Variables* chosen by STEPDISC			
BOUGONI	**W30** ($p < 0.0001$)			
KAYES	**W30** ($p < 0.0001$)	**PR5** ($p = 0.0791$)	**W5** ($p = 0.0704$)	
KEMACINA	**PW5** ($p = 0.0994$)	**R5** ($p = 0.1131$)		
KITA	**W30** ($p = 0.0008$)	**R15** ($p = 0.0920$)	**R10** ($p = 0.0690$)	
KOLOKAN	**W30** ($p < 0.0001$)			
KOUTIALA	**W30** ($p < 0.0001$)	**R30** ($p = 0.0677$)		
SEGOU	**W25** ($p < 0.0001$)	**PR5** ($p = 0.0455$)	**W20** ($p = 0.1045$)	**R20** ($p = 0.0701$)
SOKOLO	**PR5** ($p = 0.0020$)	**W30** ($p = 0.1286$)	**W10** ($p = 0.1414$)	

* Refer to Table III for descriptions of variables.

- Using option (vii) for all 22 stations, a model containing the variables DATE W30 R30 was derived. Discriminant analyses were performed on all 22 datasets using this combination; on average the percentage of potential starts correctly classified into Real and False starts was 78.21% (5.6% standard deviation (S.D.)), which is 9.29% (5.2% S.D.) higher than the baseline, given by the percentage of Real starts, at each station. Burkina Faso stations showed an average of 9.36% improvement over the baseline; the corresponding figure for Mali stations was very similar, 9.18%. These results were obtained using Definition 1 of the potential start of the wet season. The discriminant analyses were performed using the actual probabilities with which Real and False starts occurred at each station rather than assuming equal probabilities. When Definition 2 of the potential start of the wet season

Table VIII. Percentage of correctly classified potential starts using discriminant analyses in Mali stations—Definition 1

Station	Baseline (Real starts %)	Discriminant variables*	Sample probabilities Real (%)	Sample probabilities False (%)	Sample probabilities Mean (%)	Equal probabilities Real (%)	Equal probabilities False (%)	Equal probabilities Mean (%)
BOUGONI	70.73	i) DATE W15 R15	96.55	58.33	85.37	84.48	83.33	84.15
		ii) W30	91.38	41.67	76.83	75.86	70.83	74.39
		iii) DATE	89.66	54.17	79.27	72.41	79.17	74.39
		iv) DATE PR5	89.66	54.17	79.27	72.41	79.17	74.39
		v) same as i)	–	–	–	–	–	–
		vi) PW5 R5	100.00	0.00	70.73	46.55	75.00	54.88
		vii) DATE W30 R30	94.83	45.83	80.49	77.59	83.33	79.27
KAYES	71.79	i) W30 PR5 W5	92.86	40.91	78.21	71.43	72.73	71.79
		ii) same as i)	–	–	–	–	–	–
		iii) DATE	98.21	18.18	75.64	60.71	81.82	66.67
		iv) DATE PR5	91.07	31.82	74.36	71.43	77.27	73.08
		v) DATE W15 R15	92.86	40.91	78.21	73.21	68.18	71.79
		vi) PW5 R5	100.00	0.00	71.79	55.36	68.18	58.97
		vii) DATE W30 R30	92.86	45.45	79.49	67.86	68.18	67.95
KEMACINA	67.16	i) PW5 R5	100.00	0.00	67.16	42.22	81.82	55.22
		ii) same as i)	–	–	–	–	–	–
		iii) DATE	100.00	9.09	70.15	53.33	59.09	55.22
		iv) DATE PR5	100.00	9.09	70.15	53.33	59.09	55.22
		v) DATE W15 R15	97.78	13.64	70.15	57.78	54.55	56.72
		vi) same as i)	–	–	–	–	–	–
		vii) DATE W30 R30	95.56	22.73	71.64	62.22	59.09	61.19
KITA	84.62	i) DATE R15 W10	96.36	50.00	89.23	87.27	90.00	87.69
		ii) W30 R15 R10	100.00	40.00	90.77	72.73	80.00	73.85
		iii) DATE	98.18	40.00	89.23	78.18	70.00	76.92
		iv) DATE PR5	98.18	40.00	89.23	78.18	70.00	76.92
		v) DATE W15 R15	98.18	50.00	90.77	85.45	80.00	84.62
		vi) PW5 R5	100.00	0.00	84.62	45.45	70.00	49.23
		vii) DATE W30 R30	98.18	50.00	90.77	83.64	70.00	81.54
KOLOKAN	74.68	i) DATE	96.61	50.00	84.81	77.97	70.00	75.95
		ii) W30	93.22	35.00	78.48	69.49	85.00	73.42
		iii) same as i)	–	–	–	–	–	–
		iv) DATE PR5	96.61	50.00	84.81	77.97	70.00	75.95
		v) DATE W15 R15	96.61	50.00	84.81	77.97	70.00	75.95
		vi) PW5 R5	100.00	0.00	74.68	45.76	70.00	51.90
		vii) DATE W30 R30	96.61	55.00	86.08	76.27	65.00	73.42
KOUTIALA	72.73	i) DATE	95.31	50.00	82.95	73.44	70.83	72.73
		ii) W30 R30	90.63	50.00	79.55	71.88	70.83	71.59
		iii) same as i)	–	–	–	–	–	–
		iv) DATE PR5	95.31	50.00	82.95	79.69	70.83	77.27
		v) DATE W15 R15	92.19	54.17	81.82	75.00	79.17	76.14
		vi) PW5 R5	100.00	0.00	72.73	35.94	66.67	44.32
		vii) DATE W30 R30	93.75	54.17	82.95	75.00	79.17	76.14
SEGOU	69.23	i) DATE PR5	94.44	54.17	82.05	81.48	70.83	78.21
		ii) W25 PR5 W20 R20	90.74	62.50	82.05	81.48	75.00	79.49
		iii) DATE	94.44	50.00	80.77	81.48	75.00	79.49
		iv) same as i)	–	–	–	–	–	–
		v) DATE W15 R15	94.44	54.17	82.05	83.33	70.83	79.49
		vi) PW5 R5	100.00	0.00	69.23	37.04	79.17	50.00
		vii) DATE W30 R30	94.44	45.83	79.49	83.33	75.00	80.77
SOKOLO	57.69	i) PR5 DATE	76.67	68.18	73.08	66.67	77.27	71.15
		ii) PR5 W30 W10	73.33	77.27	75.00	66.67	77.27	71.15
		iii) DATE	83.33	54.55	71.15	66.67	63.64	65.38
		iv) same as i)	–	–	–	–	–	–
		v) DATE W15 R15	80.00	59.09	71.15	56.67	68.18	61.54
		vi) PW5 R5	63.33	36.36	51.92	33.33	77.27	51.92
		vii) DATE W30 R30	83.33	32.69	71.15	73.33	63.64	69.23

* Refer to Table III for descriptions of variables.

was used an average 87.78% (4.6% S.D.) of potential starts were correctly classified, implying an improvement of 4.75% (4.35% S.D.) over the baseline. Results were less good than those obtained with Definition 1 when comparing against the baseline, due at least in part to the fact that baselines were greater with Definition 2.

- The combination of variables, specific to each one of the 22 climatic stations analysed, obtained by stepwise analysis (option (i)), also gave better results than the baseline. On average there were 78.57% potential starts correctly classified, an improvement of 9.51% over the baseline, using Definition 1 (9.64% in Burkina Faso and 9.28% in Mali). When Definition 2 was used, an average of 88.3% correct classifications was obtained, with 5.26% improvement over the baseline (3.52% in Burkina Faso and 5.91% in Mali). It is interesting and encouraging that the improvements are almost as good with option (vii) as with option (i).
- In all stations that took part in this investigation Real starts occurred more frequently than False starts, in a proportion of 2:1 for Definition 1 and 5:1 for Definition 2. The rate of success in classifying potential starts correctly was found to be higher when the discriminant analysis used the actual proportion at which Real and False starts occurred rather than assuming that both starts had the same chance to occur. For example, for combination DATE W30 R30 and Definition 1 there was an average 78.21% correct classification using the actual rate of Real and False starts. The figure reached when equal rates were assumed was only 73.88%, a difference of 4.33%.
- The potential start of the rainy season takes place earlier in climatic stations which receive more precipitation. For example, in Burkina Faso, the Real start of the wet season starts around 19 June in Banfora station which receives precipitation of 1141 mm/year, but it is delayed until around 8 July in Kaya station which receives precipitation of only 697 mm/year.
- Of the fifteen variables included in this investigation, the date of the potential start of the wet season (DATE) showed the highest discriminatory power, when a selection was performed using stepwise analyses. When DATE was left out of the analyses, the number of wet days in the 30 days prior the potential start (W30), became the most discriminating variable.
- The combinations of variables obtained through the stepwise analysis when including DATE, obtained, on average, a 1.47% higher rate of success at discriminating potential starts correctly, than their equivalents when DATE was left out.

5. DISCUSSION

This study used extensive daily rainfall datasets (40–77 years), from two countries in western Africa (Burkina Faso and Mali), to derive a linear discriminant function (LDF) that distinguishes between 'Real' and 'False' starts to the onset of the rainy season, in different regions. There were 22 sites studied, located in similar tropical semiarid ecosystems, having annual wet and dry seasons. Farmers here depend entirely on the start of the rains, which is highly variable, to plant crops like millet, sorghum and cowpeas, in their subsistence agriculture.

Several LDFs were tested, from which two were considered to be the most effective at classifying 'Real' and 'False' onsets correctly, in all the regions studied. The first one, LDF (1) used the following variables:

(i) DATE of the potential onset;
(ii) number of wet days (0.1 mm or more) in the 30 days prior to the potential onset;
(iii) precipitation in the 30 days prior to the potential onset.

All three variables can easily be derived, since daily rainfall data are the most common records kept in meteorological stations. Furthermore, all three variables can be obtained soon after the potential onset, 5 days at most, which allows for decisions regarding planting to be taken shortly after the start of the rains.

In general this LDF achieved 4% higher rate of correct classification when using sample probabilities of 'Real' and 'False' starts, rather than equal probabilities. Similar results were found from all LDFs.

Equal probabilities did better at classifying 'False' starts (73% correct with equal probabilities compared to 49% with sample probabilities), but this result is only useful if there is more interest in classifying 'False' than 'Real' starts correctly.

There were two regions studied in Burkina Faso: the Sudanian (650–950 mm/year) and the Sudano-Guinean (> 950 mm/year). The percentages of correctly classified potential onsets were 78% and 76%, respectively; the improvements over baseline were 7% and 12%, respectively. This difference was probably mainly due to the baseline (i.e. percentage of 'Real' starts) being lower in the Sudano-Guinean region, the wetter of the two. The reason why there are fewer 'Real' starts in a wetter area is not very clear, though it can be suggested that as the rains start earlier here, there are more possibilities for 'False' starts. There were also two climatic regions studied in Mali: the Semiarid Tropical and the Hot Tropical, which in that order, correspond ecologically to the Sudanian and Sudano-Guinean regions in Burkina Faso. The percentages of correctly classified potential onsets were 77% and 85%, respectively, and the difference over baseline was 9% for both. The LDF from option (i) was therefore highly consistent when classifying potential onsets in all regions.

Using only DATE of the potential onset in the LDF (choice (iii)) is a good alternative. On average the percentage of potential onsets correctly classified using (iii) was 77%, with 76% in Burkina Faso and 79% in Mali, and an overall improvement over the baseline of 8%. Although we can do almost as well, on average, using DATE alone, in terms of discriminatory power, as we can by adding extra rainfall-related variables, the extra variables show some improvement for 12 of the 22 stations and only make discrimination worse for five stations (four in Burkina Faso, one in Mali). Hence we recommend that if these variables are readily available, it seems worthwhile to include them to obtain even a small improvement in predictive power, given the potential importance of a correct identification.

The use of regional LDFs (rules (v), (vi)) was also tested. There were four climatic regions in the study, two in Burkina Faso and two in Mali. One LDF from an individual station was selected from each region. It was found that as long as the LDF included variables with consistent discriminant power, such as DATE, the rate of correct classification was similar to those already mentioned. In fact there was no tendency for the wetter stations to be better predicted by the rule derived from the wet representative station than by the 'dry rule', or vice versa. In Burkina Faso 'dry rule' (vi) did better than, or the same as, rule (v) for 12 of 14 stations, wet or dry, whereas in Mali 'wet rule' (v) was strictly better than (vi) for all eight stations, regardless of annual rainfall.

The methodology described above to distinguish 'Real' from 'False' starts of the wet season allows some flexibility regarding the definition of the starts of the rains. This is an advantage since the definition can change from one place to another, or even vary depending on the crop to be planted. In this investigation, a second definition, less strict than the original, was implemented. When using this definition, the potential onset of the rains occurred a couple of weeks or so earlier. The percentage of 'Real' starts was higher with this definition, causing baselines to increase from 69% with the original definition to 83% with the second definition. The rate of correct classification was 88% with the second definition, a 5% improvement over the baseline. In general, a higher percentage of success but a lower improvement over the baseline were achieved compared to Definition 1, mainly because the baseline was increased.

The same analyses discussed for Burkina Faso and Mali were repeated in Dagana and Zinder stations, in Senegal and Niger, respectively. Dagana has an average precipitation of 546 mm/year with a Semiarid climate, similar to the Semiarid Tropical climate of Mali and the Sudanian region in Burkina Faso. Zinder was drier, with an annual precipitation of 476 mm and a Sub-tropical Desert climate, drier than the rest of the regions included in the present study. In Dagana, rule (vii) classified 73% potential starts correctly (1% improvement over the baseline), with Definition 1, and 83% with Definition 2 (0% improvement over the baseline). For Zinder station, option (vii) classified 81% potential starts correctly (3% improvement over the baseline), with the first definition, and 89% with the second definition (0% improvement over the baseline). It can be seen that for these two stations the improvements over the baseline are rather small or non-existent. Nevertheless, as only one site from each country was studied, it is not sensible to draw definitive conclusions.

Other investigators have used linear discriminant analysis (LDA) to explain variability in the weather pattern in the Sahel. Janicot *et al.* (1995) cited earlier work by Folland and co-workers which showed that inter-annual Sahel rainfall variability could be explained by global sea-surface temperature (SST) anomalies, using LDA. Folland apparently found that 21 out of 45 quints (i.e. categories: very dry, dry, mean, wet and very wet) of Sahel standardized rainfall were correctly predicted during the period 1901–1945. Janicot *et al.* (1995) have recently done further work in performing LDA for Sahel rainfall using SST predictors. Their approach was more complicated, since it used information derived from analysis of West African rainfall anomaly patterns and from separation between low and high frequency rainfall variability time scales. Preliminary results from the study show up to 69% success at classifying years into the five categories or quints: very dry, dry, mean, wet and very wet.

Upper-wind daily data from West African Sahel have also been used to predict the onset and cessation of the rainy season, more than 2 months ahead. The rainy season is said to start between 49 and 81 days after the first sudden changes in wind direction at specified atmospheric levels. It starts earlier in the most eastern location, which agrees with the climatology of western Africa. The cessation of rainfall is said to occur about 80 days after the destruction of the mid-tropospheric vertical wind shear (Omotosho, 1992). However it is felt that, although the period of time between wind changes and start of rains in a station, like Kano, was not too varied (75 days $\pm$ 7), not enough data were analysed; only results from 8 years data for Kano, 1 for Maiduguri in Niger and 4 for Bamako, in Mali, were presented. The reason for this might be that upper-wind data may not be easily available, which will make this technique of less practical use; it is also of less practical use in the sense that it may be predicted that the season will start 75 days after some upper atmospheric change, but the prediction will be ignored and no crops planted, unless substantial rain actually falls before the predicted date.

An advantage of our methodology over those involving SST or upper winds as predictors is that we require only simple data that could be recorded close to the sites where the crops are planted. Our study extends that of JSD, by looking at a much larger set of rainfall stations in a different geographic region, by introducing an alternative definition of the start of the wet season, and by considering extra variables. The results found, in terms of the potential for correctly identifying the start of the wet season, are similar to those of JSD, suggesting that linear discriminant analysis may be useful in other countries with clear-cut wet and dry seasons. One caveat, however, is that the rules are necessarily based on historical data. There is a strong belief that some degree of climate change may be in progress, not least in the countries studied here. It does not, of course, follow that a change in average rainfall, for example, would automatically lead to a change in the relationships identified in this study. However, one possible way of allowing for change would be to incorporate weighting into the discriminant analysis, with greatest weight being given to the most recent data.

ACKNOWLEDGEMENTS

We are grateful to Dr Alan Robock for sending the data used in this investigation and to Dr M.V.K. Sivakumar for giving permission for its use. Comments from two referees and the Editor have led to improvements in the paper. The paper was revised while the second author was visiting the Bureau of Meteorology Research Centre (BMRC), Melbourne, Australia. He is grateful for the facilities and support provided by BMRC, and to The Leverhulme Trust which partially funded the visit under its Study Abroad Fellowship scheme.

REFERENCES

Benoit P. 1977. The start of the growing season in Northern Nigeria. *Agricultural Meteorology* **18**: 91–99.

Dodd DES. 1996. Early detection of the start of the wet season in semi-arid tropical climates of western Africa, MSc Thesis. University of Kent at Canterbury.

Hopkins T, Burt L. 1989. *User Guide to the Events Program*. Applied Statistics Research Unit, University of Kent at Canterbury.

Hulme M. 1987. Secular changes in wet season structure in central Sudan. *Journal of Arid Environments* **13**: 31–46.

Janicot S, Fontaine B, Moron V. 1995. Global Sea Surface Temperature variability associated with West African rainfall anomaly types. Application to seasonal forecast of Sahel rainfall. *6th International Meeting on Statistical Climatology*. Galway, Ireland; 483–486.

Jolliffe IT, Sarria-Dodd DE. 1994. Early detection of the start of the wet season in tropical climates. *International Journal of Climatology* **14**: 71–76.

Krzanowski WJ. 1988. *Principles of Multivariate Analysis: A User's Perspective*. Oxford University Press: Oxford.

Nicholls N. 1984. A system for predicting the onset of the North Australian wet-season. *Journal of Climatology* **4**: 425–435.

Omotosho JB. 1992. Long-range prediction of the onset and end of the rainy season in the West Africa Sahel. *International Journal of Climatology* **12**: 369–382.

SAS Institute Inc. 1989. *SAS/STAT® User's Guide*, *Version 6* (4th edn, vol. 2). SAS Institute Inc: Cary, NC.

Sivakumar MVK. 1988. Predicting rainy season potential from the onset of rains in the Southern Sahelian and Sudanian climatic zones of West Africa. *Agricultural and Forest Meteorology* **42**: 295–305.

Stern RD, Dennett MD, Garbutt DJ. 1981. The start of the rains in West Africa. *Journal of Climatology* **1**: 59–68.

