



HAL
open science

Regional-scale rainy season onset detection: a new approach based on multivariate analysis.

Joseph Boyard-Micheau, Pierre Camberlin, Nathalie Philippon, Vincent Moron

► **To cite this version:**

Joseph Boyard-Micheau, Pierre Camberlin, Nathalie Philippon, Vincent Moron. Regional-scale rainy season onset detection: a new approach based on multivariate analysis.. *Journal of Climate*, 2013, 26 (22), pp.8916-8928. 10.1175/JCLI-D-12-00730.1 . hal-00908347

HAL Id: hal-00908347

<https://hal.science/hal-00908347v1>

Submitted on 30 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Regional-Scale Rainy Season Onset Detection: A New Approach Based on Multivariate Analysis

JOSEPH BOYARD-MICHEAU, PIERRE CAMBERLIN, AND NATHALIE PHILIPPON

Centre de Recherches de Climatologie, Biogéosciences, UMR6282, CNRS/Université de Bourgogne, Dijon, France

VINCENT MORON

CEREGE, UMR7330, Aix-Marseille Université, and CNRS, Aix-en-Provence, France, and International Research Institute for Climate and Society, Columbia University, New York, New York

(Manuscript received 17 October 2012, in final form 10 May 2013)

ABSTRACT

In agroclimatology, the rainy season onset and cessation dates are often defined from a combination of several empirical rainfall thresholds. For example, the onset may be the first wet day of N consecutive days receiving at least P millimeters without a dry spell lasting n days and receiving less than p millimeters in the following C days. These thresholds are parameterized empirically in order to fit the requirements of a given crop and to account for local-scale climatic conditions. Such local-scale agroclimatic definition is rigid because each threshold may not be necessarily transposable to other crops and other climate environments. A new approach is developed to define onset/cessation dates and monitor their interannual variability at the regional scale. This new approach is less sensitive to parameterization and local-scale contingencies but still has some significance at the local scale. The approach considers multiple combinations of rainfall thresholds in a principal component analysis so that a robust signal across space and parameters is extracted. The regional-scale onset/cessation date is unequally influenced by input rainfall parameters used for the definition of the local rainy season onset. It appears that P is a crucial parameter to define onset, C plays a significant role at most stations, and N seems to be of marginal influence.

1. Introduction

Agricultural production in the tropical zone is highly dependent on several environmental factors, especially water availability (Wallace 1991; Meinke and Stone 2005). In small and low income family farms, irrigation systems are generally underdeveloped, and crop yields are fully dependent on the rainfall amount and distribution. The variability in seasonal characteristics such as rainy season onset/cessation (Ati et al. 2002) and/or dry spell frequency (Usman and Reason 2004) is potentially damaging to crop production. In particular, the rainy season onset is long awaited, and farmers perceive its prediction as an invaluable tool to help them in agricultural planning (Jones et al. 2000). The Famine Early

Warning System (FEWS) program of the U.S. Agency for International Development considers the onset date prediction as an essential part of a monitoring tool (Verdin et al. 2000; Tadesse et al. 2008).

The definition of rainy season onset/cessation is versatile (Smith et al. 2008). Climatologists, agronomists, or hydrologists have proposed different definitions. For instance, agroclimatologists usually define the onset at the rain gauge scale, using a variety of empirical thresholds (Stern et al. 1981; Sivakumar 1988; Moron et al. 2009; Marteau et al. 2009). They consider that the rainy season onset is the first wet day of a spell receiving a given rainfall amount and not followed by a long dry spell during the subsequent weeks. The rainfall thresholds are determined empirically in order to fit the requirements of a given crop and are adjusted to account for local-scale climatic conditions (e.g., lower potential evapotranspiration in cool, high-altitude areas than in warm areas). These local-scale “agroclimatic” onset definitions are therefore rigid in so far as each threshold may not be transposable to other crops and other

Corresponding author address: Joseph Boyard-Micheau, Centre de Recherches de Climatologie, Biogéosciences, Faculté des Sciences Gabriel, Université de Bourgogne, 6 Blvd. Gabriel, 21000, Dijon, France.
E-mail: joseph.boyard-micheau@u-bourgogne.fr

climate environments. An adjustment to the climatological mean rainfall amount during wet spells at each station has been proposed for the amount of the initial wet spell (Moron et al. 2010) but has not yet been proposed for the other parameters (i.e., length of the initial wet spell and length/intensity of the post-onset dry spell used to define false starts, which are significant wet spells followed by a long dry spell). Simpler methods, based on the accumulation of rainfall up to a given threshold, either fixed (e.g., 50 mm) or relative (e.g., 15% of a station's climatological mean), have been proposed for Australia (Lo et al. 2007). However, Smith et al. (2008) well emphasized the difficulties attached to the parameterization of the monsoon onset and end dates, even based on these simpler methods.

Seasonal climate prediction relies on the existence of a spatially consistent regional-scale signal (Moron et al. 2007), connected to large-scale atmospheric and oceanic forcings such as sea surface temperature anomalies or any other slowly varying boundary conditions. Any regional-scale signal may be difficult to extract when the target variable is strongly dependent on parameterization and/or is highly variable at local scale. To circumvent these difficulties, Liebmann and Marengo (2001) determined the onset and cessation dates from accumulated rainfall anomalies. This method is simply based on the cumulative sum, year by year, of the difference between raw daily rainfall and its long-term seasonal average. For an anomalously wet (dry) day, relative to the long-term mean, this difference is positive (negative). An anomalously dry (wet) period is then shown as continuously declining (rising) accumulated anomalies. Across each season, the day on which these anomalies reach a minimum (maximum) defines the onset (cessation). This method avoids the use of rigid rainfall thresholds and can be applied at both local and regional scales. At the regional scale, onset and cessation dates can be equally defined from the temporal evolution of the cumulative score of the first principal component of daily rainfall (Camberlin and Diop 2003). However, if a regional-scale onset is defined, it remains to be assessed how it is related to the local-scale onsets, adjusted to the precise crop requirements.

The main objective of this work is therefore to propose a new approach to define onset/cessation dates and monitor their interannual variability, which would be minimally sensitive to parameterization and local-scale contingencies but still has some significance at local scale. This approach is applied on a region covering Kenya and northern Tanzania, which is characterized by strong climatic and geographical heterogeneities. However, the large-scale forcing of interannual and intra-seasonal variability of regional-scale rainfall (Ogallo et al.

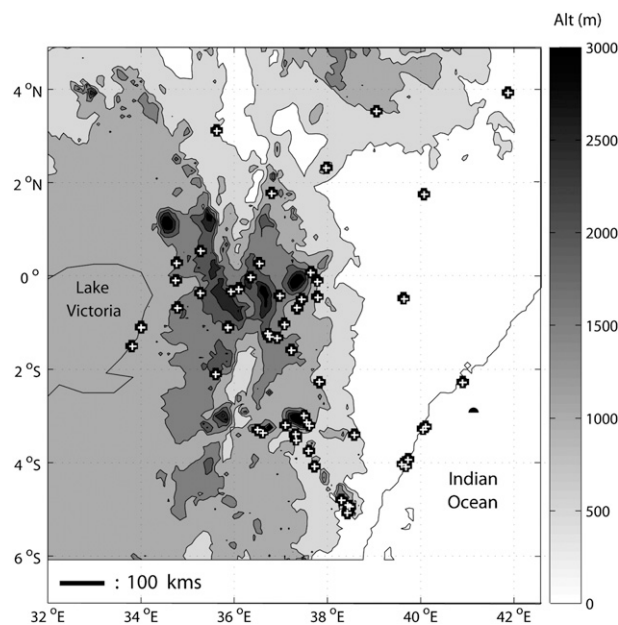


FIG. 1. East African geographical context and rain gauge distribution (black and white crosses). Gray shading show altitude in meters above mean sea level.

1988; Rowell et al. 1994; Nicholson 1996; Camberlin et al. 2001) suggests that the identification of regional-scale onset and cessation dates is possible. This region therefore appears perfectly relevant to estimate the performance of the new approach presented here.

Section 2 presents the station network and the geographical and climatological context. Section 3 is devoted to a preliminary analysis illustrating the sensitivity of the onset determination using the two types of the method presented above: agroclimatological criteria and accumulated rainfall anomalies. Section 4 presents the results obtained from our new approach.

2. Area and dataset

The studied area is characterized by strong geographical and topographical contrasts, in particular an opposition between the semiarid plains of eastern and northern Kenya and the wetter highlands of central and western Kenya and northeastern Tanzania. The highlands include plateaus at 1000–2000 m and several mountain chains ranging from 3000 to 5900 m (Fig. 1). The Indian Ocean coast and Lake Victoria generate mesoscale circulations that bring further complexity to the climate patterns. The Hadley circulation and seasonal migration of the intertropical convergence zone (ITCZ) on both sides of the equator broadly control the alternation between dry and wet seasons. Rainfall is distributed annually in two seasons (Fig. 2a), the “long rains” in boreal spring

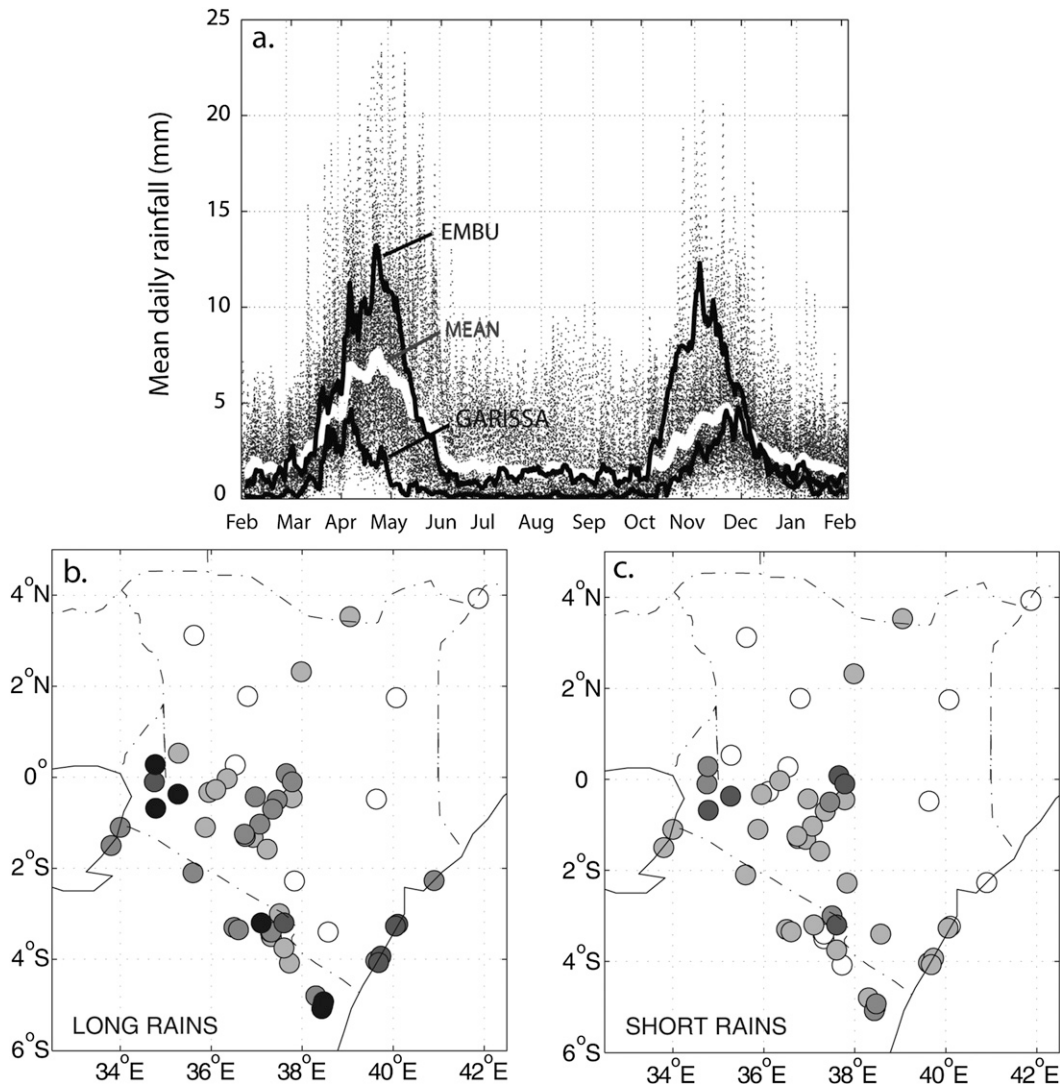


FIG. 2. (a) Daily rainfall mean seasonal cycles (over 1961–2001) at the 53 stations (dots) and the mean of these stations (thick white line). A wet station (Embu) and a dry station (Garissa) are also shown. Climatological maps of the seasonal rainfall amounts for the (b) long rains and (c) the short rains.

(March–May) and the “short rains” in boreal autumn (October–December). The strength of the Walker circulation during the short rains, with strong rising motion over the eastern Indian Ocean and relative subsidence over the western Indian Ocean, explains that the short rains record lower rainfall amounts than the long rains (Figs. 2b,c). Despite local-scale variations, especially toward Lake Victoria, the bimodal regime is more or less valid across the whole region, from northern Tanzania to northern Kenya, justifying the search for regional-scale onset and cessation dates.

The database comprises daily rainfall data recorded at 53 rain gauges over Kenya and northern Tanzania (Fig. 1), collected from 1961 to 2001 by the Kenya Meteorological

Department, the Intergovernmental Authority on Development (IGAD) Climate Prediction and Application Center, and the Tanzania Meteorological Agency. Missing values (10%) are filled using multiple linear regressions (MLR) based on nearby stations. To eliminate biases introduced by the MLR method, that is, the overestimation of the wet days occurrence and underestimation of mean rainfall amounts, a local-scaling correction (Ines and Hansen 2006) has been applied at a monthly time scale. The estimated frequency of wet days (≥ 1 mm) is scaled to match the long-term observed frequency. The estimated daily rainfall amounts are then adjusted by multiplying them by a fixed coefficient in order to match the long-term mean daily rainfall amounts.

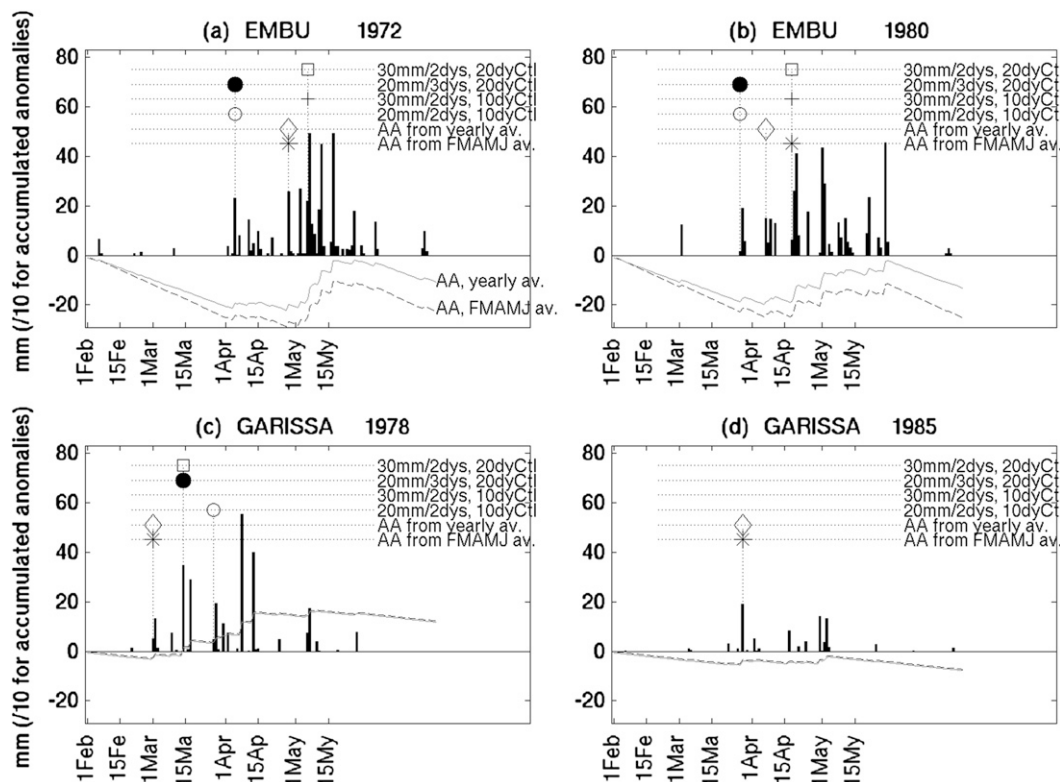


FIG. 3. Onset dates of the long rains for a sample of years at a wet station (Embu) and dry station (Garissa), computed using different methods. Vertical bars indicate daily rainfall (mm). The squares, black circles, crosses, and opened circles show onset date computed from different agronomic definitions, whose thresholds (P , N , and C ; see text) are indicated in the panel top right corner. Diamonds and stars show onset dates computed from the accumulated anomalies (AA) method using as reference either the yearly or the February–June mean rainfall amount, respectively. The solid (dashed) lines at the bottom show the values of the accumulated anomalies, with reference to the yearly (February–June) average.

3. Uncertainties associated with the parameterization of local-scale onset

As noted above, onset (and cessation) dates can be defined in different ways. An illustration of the uncertainties induced by their parameterization is, first, provided for two stations representative of contrasted climatic (i.e., wet versus dry) environments. Then the evaluation is extended to the whole network.

a. Uncertainties in the onset dates determination: A case study

Onset dates at Embu, a wet station receiving a mean 643 mm during the long rains and 544 mm during the short rains, and Garissa, a dry station receiving a mean 150 mm during the long rains and 205 mm during the short rains, are computed from two different methods based on local-scale rainfall. The first method uses accumulated daily rainfall anomalies (Liebmann and Marengo 2001; Camberlin and Diop 2003). The onset

corresponds to a sudden upward variation in accumulated anomalies (AA), computed as departures from the long-term average. This method gives a single onset date for each station. However, this date is dependent on the reference period of time over which the average rainfall is computed. The onset date may be shifted by a few days depending on whether the mean daily rainfall used to obtain the anomalies is computed over the whole year or only part of it (excluding part of the dry season for instance). The second method uses a local-scale agroclimatic definition (Sivakumar 1988; Marteau et al. 2009). The onset is the first wet day (>1 mm) of N consecutive days receiving at least P mm without a dry spell (lasting at least 10 days) receiving less than 5 mm in the following C days (control period). The onset date may therefore vary depending on these thresholds. Four years are selected to illustrate the sensitivity of the long-rains onset to these parameterizations.

At Embu in 1972 (Fig. 3a), the AA method gives an onset date that is different from all those obtained using

the four variants of the agronomic method, based on a few combinations of the P , N , and C thresholds. This difference is either positive (+23 days) or negative (−8 days). A change in P from 20 to 30 mm also leads to a one month delay in the onset date. In 1980 (Fig. 3b), a large difference (22 days) is also found in the onset date, depending on P . The AA method based on the February–June seasonal average (asterisk) gives an onset on 18 April, in agreement with some agronomic definitions ($P = 30$ and $N = 2$). However, for this year, the AA method is not more robust than the agronomic method: when the onset is computed based on deviations from the yearly average, it occurs 11 days earlier. For the agronomic definition, sensitivity to the value of P is also evident.

At Garissa in 1978 (Fig. 3c), the season is anomalously wet (276 mm from March to May instead of 125 on average), but owing to several dry spells lasting 10 days or more after the major wet events, the definition using $C = 10$ fails to detect any onset, or postpones the onset too late in the season. By contrast, the AA method identifies the first significant rainfall event as the rainy season onset, but leads to dates different from the agroclimatic definition.

Finally, at Garissa in 1985 (Fig. 3d) the season is poor (80 mm only from March to May), without any day receiving more than 20 mm of rainfall, and thus all the agroclimatic definitions using $P \geq 20$ mm obviously fail. However, it is not a situation of total failure of the rains, and visually a rainy season can still be identified. The AA method appropriately detects the onset on 29 March.

Although the methods used are not exhaustive, these examples point to uncertainties in the onset date detection associated with the parameterizations. In particular, several cases were found for which the definitions based on a specific rainfall event failed, resulting in undetected onsets, whereas a rainy season was nonetheless apparent. The most adaptive method (AA), though able to detect onsets even in dry climates (Garissa), is still sensitive to parameterization, such as the average daily rainfall taken as reference. Fortunately, such discrepancies do not occur every year, but they still cast doubt on the representativeness of local onset dates when based on a unique definition.

b. Sensitivity analysis

In agreement with Ati et al. (2002), the above case studies confirm that onset dates are relatively sensitive to parameterization. Logically, the robustness of onset dates is weaker for a dry station than for a wet station. The sensitivity to the parameterization is now analyzed over the whole network. Onset dates are computed from the same general local-scale agroclimatic definition as in

section 3a but with a wide range of values for P , N , and C . These thresholds are selected from the distribution of wet spell characteristics (length and rainfall amount) computed for the February–June period, including the long rains, and for the September–January period, including the short rains. The ranges of values for P , N , and C have been determined based on previous studies. For P , most studies consider rainfall thresholds of 20 or 30 mm, but to enable a better adjustment to dry or very wet stations, both lower and higher thresholds (10 to 50 mm) have been used. Mugalavai et al. (2008) indicated that $P = 40$ mm and $N = 4$ days were appropriate to determine the onset in western Kenya, which is the wettest part of the region. Thus, iteratively the onset date is the first rainy day of a sequence of $N = 2, 3, 4, 5$ —consecutive days receiving at least $P = 10, 15, 20, 25, 30, 40,$ or 50 mm—not followed by a long dry spell (at least 10 days) receiving less than 5 mm of precipitation over $C = 20$ or 30 days. To each station is therefore assigned a theoretical set of $4 \times 7 \times 2 = 56$ onset dates for each year. The use of extreme combinations (e.g., $P = 50$) in a context of low and erratic rainfall, as is the case over part of the study area, means that the onset cannot always be defined. These specific cases are first considered as missing values, and two complementary approaches, used to consider these gaps, will be shown in section 4.

The onset date sensitivity to parameterization may be both space and time dependent. It is therefore evaluated for each station using an analysis of variance following Rowell et al. (1995), Rowell (1998), Bouali et al. (2008), and Philippon et al. (2010), among others. The observed total variance of onset dates for each station across the 41 years and 56 combinations is partitioned into the interannual (or external) variance of the mean of the onset dates computed from the whole set of the 56 combinations and the internal variance due to the deviations of each combination relative to the mean. It is somewhat analogous to the analysis of variance in an ensemble of climatic simulations [e.g., SST-forced atmospheric general circulation model (AGCM) runs] between the “signal” conveyed by the interannual variation of the ensemble mean and the “noise,” that is, the deviations between each run and the ensemble mean. In our case, the noisy component is conveyed by the different combinations instead of runs in the AGCM ensemble. A large (respectively small) amount of external variance is thus an indication that the local-scale onset is weakly (highly) sensitive to the parameters used to define it.

In summary, for each of the 53 stations, the total variance of the onset dates (V_{tot}) is the sum of two components: the external variance owing to the interannual variability of the mean onset date over the 1961–2001 period (V_{ext}) and the internal variance owing to the

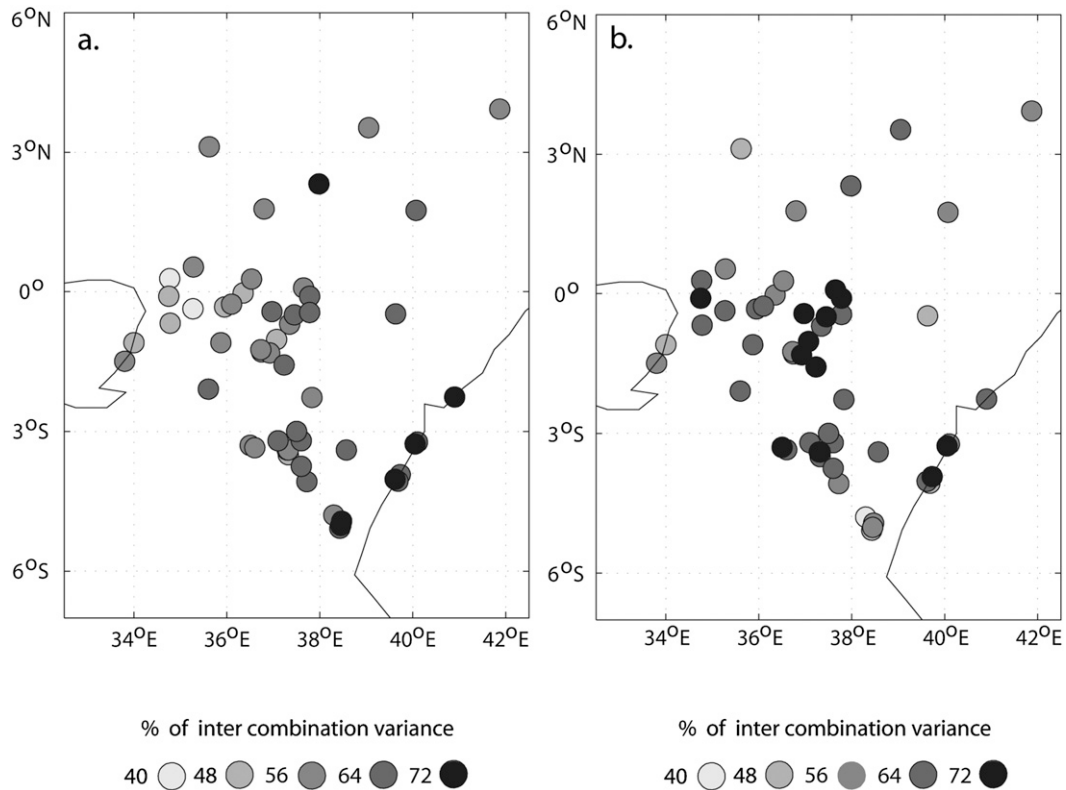


FIG. 4. Percentage of external variance computed separately for the 53 stations for the (a) short rains and the (b) long rains (see text).

deviations among the onset dates from the 56 combinations (V_{int}). For each station, we therefore have

$$V_{tot} = V_{ext} + V_{int},$$

where V_{int} is the internal variability, given by

$$V_{int} = \frac{1}{Nn-1} \sum_{i=1}^N \sum_{j=1}^n (x_{ij} - x_i^2),$$

and V_{ext} the variance of the intercombination means,

$$V_{ext} = \frac{1}{N-1} \sum_{i=1}^N (\bar{x}_i - \bar{\bar{x}}^2),$$

where N is the number of observations (41 yr), n the number of combinations (56 combinations), x_{ij} the onset date computed for observation i and combination j , \bar{x}_i the mean over observation i , and $\bar{\bar{x}}$ is the mean over all observations and combinations.

The intercombination variance is computed separately for the 53 stations for both rainy seasons (Fig. 4). Comparison of the two rainy seasons gives crucial information about onset date sensitivity. The long-rains season (Fig. 4b)

is characterized by a homogeneous spatial pattern over the whole study area. Except for a few dry stations, the external variance is usually greater than 64%. By contrast, the short-rains season shows generally lower and more heterogeneous values. This heterogeneity appears as an east–west gradient with external variance decreasing westward. This is partly related to the fact that western Kenya has no absolute dry season between the long rains and the short rains; therefore, the short-rains onset here is difficult to detect.

On the whole, the local-scale agronomic definition is logically less stable at the driest stations and when the rainy season is less abundant. One of the reasons is that several parameters, used in the local-scale agronomic definition, may be affected by the lower mean rainfall. Indeed, a low rainfall amount can imply fewer rainy days/wet spells and consequently more dry days/dry spells, which may enhance the role of C and also increase the number of undefined onsets (owing to a too large P).

4. Regional onset and cessation determination

Given the uncertainty in onset dates at the local scale, a new approach is developed to be the least sensitive to

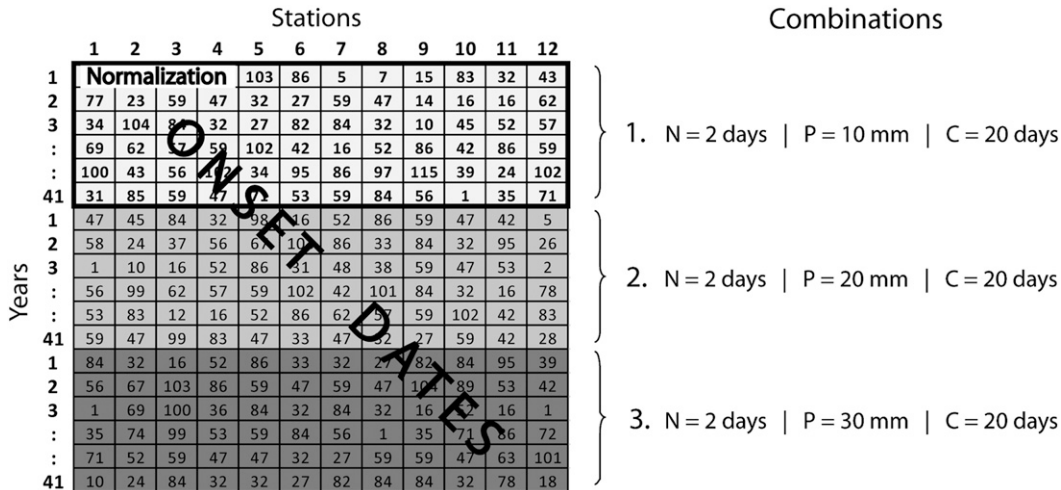


FIG. 5. Multicombination array submitted to PCA with the 41 years and 56 experiments as observations and the 53 stations as variables.

parameterization. It is analogous to ensemble numerical simulations in which a signal is extracted from a number of runs that differ by the initial conditions and/or by the numerical model used. Here, the ensemble is made of 56 onset date time series that differ by the parameter values used to define the onset (i.e., the 56 combinations presented in section 3b and hereafter referred to as experiments). The approach is based on four distinct steps (Fig. 5). First, onset dates using the 56 combinations of parameters are computed from 1961 to 2001 for the whole set of 53 stations. Second, each of these 56 experiments is normalized station by station to zero mean and unit variance in order to remove wet/dry biases associated with the local-scale climatic conditions. Third, the 56 experiments are row concatenated; that is, all (41 yr × 53 stations) arrays, each describing a given combination, are placed beneath each other. Finally, a principal component analysis (PCA) is applied to that matrix of 41 × 56 observations (the years and experiments) and 53 variables (the stations) to extract the leading modes of variability, that is, the regional-scale signals. Each PC is initially 41 yr × 56 experiments long, and the “ensemble mean” is simply obtained by averaging the 56 experiments. Hereafter, this approach is referred to as the “multicombination PCA”. As in Marteau et al. (2009) and Moron et al. (2009), the undefined onset dates were replaced by the latest available onset date observed across the network for the given year combination. This kind of treatment allows keeping the character of the onset season over the whole region for the given year. The number of such cases is fairly high (21% for the long rains and 39% for the short rains). This is because some extreme thresholds (e.g., P = 50 mm) are rarely met at some locations. It is

therefore necessary to check the impact of such missing value replacement on the results. To that end, two complementary approaches are developed to replace the missing onset dates and results compared with the initial approach (S0).

The first one consists of rearranging the onset date array so that experiments are no more considered as observations but as variables: rows describe years and columns describe stations and experiments. Such a rearrangement allows an easy elimination of the combinations that end up with a certain percentage (10%, 20%, 30%, . . . , more than 50%) of undefined onset dates. Then a PCA is applied on this array.

The second approach consists of replacing the undefined onset dates using techniques other than the replacement by the latest available onset date observed across the network for the given year combination. Three strategies have been developed to check sensitivity to the replacement of the undefined onset date. The first (S1) consists of replacing missing values with the latest available onset date observed across the 56 combinations for the given year station. The second (S2) uses the latest available onset date observed across the 41 years for the given combination station. The third strategy (S3) replaces them with the mean onset date computed from the combinations available for the given year station. The effect of the replacement of the undefined onset dates is then estimated by comparing (i) the spatial patterns associated with the first principal components, (ii) the percentage of variance explained by the first PCs, and (iii) the correlation between the first PCs. A similar procedure is implemented for the cessation dates, using the same combinations as for the onset dates but applied to the daily rainfall time series taken

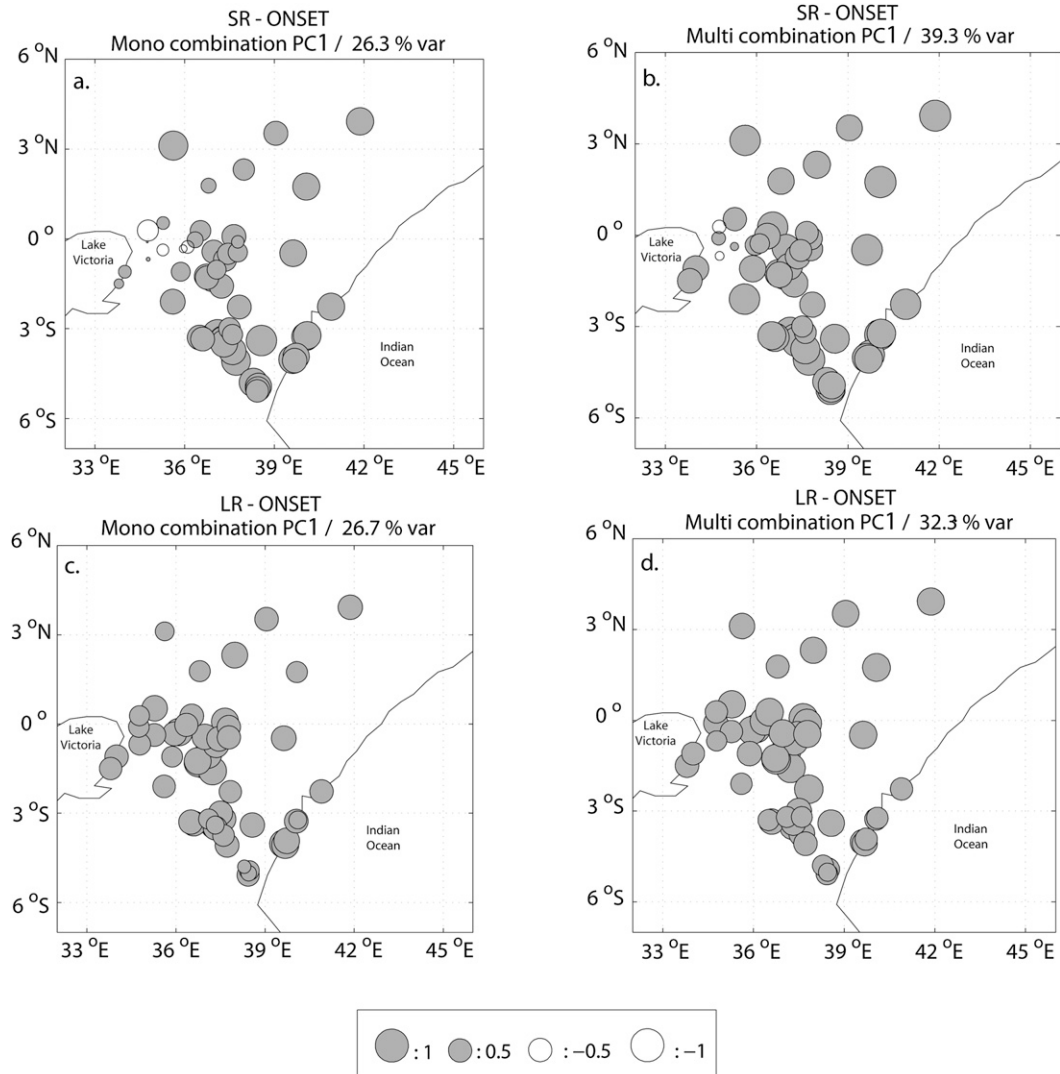


FIG. 6. Spatial patterns of the first principal component of the (left) monocombination PCA and (right) multicomcombination PCA for the (top) short rains and (bottom) long rains. Gray (white) circles denote positive (negative) loadings; size of the circles is proportional to the loading coefficients. The percentage of variance explained is indicated in the title.

backward, that is, from June to February for the long rains and from January to September for the short rains.

Figure 6 (right panels) shows the spatial patterns and the percentage of variance explained associated with the first PC (PC1) of the multicomcombination PCA. Results are compared with those of a “monocombination PCA” (left panels) in which onset dates are defined using a single combination of the N , P , and C parameters. In the example displayed, fixed thresholds of $N = 2$ days, $P = 20$ mm, and $C = 20$ days were retained. The data array subjected to the monocombination PCA has 41 yr \times 1 experiment as observations and 53 stations as variables.

For the short rains onsets (Figs. 6a,b), the spatial structure of PC1 is similar with both methods. Except for a few stations in western Kenya, the signal is quite spatially uniform over the whole study area, with the highest loadings along the Indian Ocean and in northern Tanzania and somewhat lower ones on the eastern slopes of Mount Kenya and near Lake Victoria. Wet stations located in the Rift Valley for the short-rains onset have the weakest loadings on the PC1 using the monocombination. The loadings increase using the multicomcombination approach, which implies that thresholds retained in the monocombination are less suitable to reveal the regional-scale signal present in the data.

TABLE 1. Percentage of variance (% var) carried by the first principal component (PC1) of the monocombination and multicombination PCAs (see text) for the long rains and short rains onset and cessation dates. The last line gives the internal variance computed from the 56 scorings coefficients associated with the PC1 of the multicombination approach used to extract the regional onset/cessation.

	Long rains		Short rains	
	Onset	Cessation	Onset	Cessation
PC1-mono (% var)	26.7	14	26.3	31.5
PC1-multi (% var)	32.3	18.1	39.3	40.5
V_{int}	0.09	0.26	0.18	0.18

There is also a substantial difference in the proportion of variance explained by the two PC1: 39.3% for the multicombination against 26.3% for the monocombination (Table 1). Part of this difference in the explained variance may be related to the filling of undefined onsets. Indeed, Table 2 shows that using other methods to replace the undefined onset leads to lower percentages of variance (note however that these percentages are quite similar whatever the method used). The spatial patterns associated with the first PCs of the monocombination (Fig. 6c) and multicombination approach (Fig. 6d) also look alike for the long-rains onset. Some northern Kenyan dry stations show weaker loadings in the PCA using the monocombination. As for the short rains, the proportion of variance explained by PC1 is higher using the multicombination approach (32.3%) than using only one combination (26.7%). Whatever the technique used to replace the undefined onset, the spatial patterns associated with the first PC are quite unchanged despite that some station loadings decrease (not shown).

These results suggest that the PC1 based on the multicombination is more efficient at extracting the regional-scale component of the onset than when a single, arbitrary combination of thresholds is used. This finding is basically in the same vein as in numerical experiments where an ensemble of multiple runs helps to extract the reproducible signals. Here, any deviations related to some specific parameterizations are partly canceled out by the ensemble, and the redundant variability, already detected in the monocombination, is emphasized since the same regional-scale variability is repeated in each of the experiments while noise conveyed by space and parameterization tends to cancel out.

These findings remain valid whatever the season and the descriptor (onset or cessation) considered (Table 1). Generally, northern stations receiving less rainfall than the rest of the stations have a higher sensitivity to the thresholds used to define onset and cessation dates. Using less stringent thresholds is beneficial to the

TABLE 2. Effect of the strategy of replacement of the undefined onset dates on the percentage of variance (% var) carried by the first principal component of the multicombination PCAs for the short rains onset dates: r is the correlation between the interannual variability of the regional onset date as retained in the study (strategy S0) and the interannual variability of the regional onset date as obtained from the three other strategies.

	S0	S1	S2	S3
PC1 (% var)	39.3	26.4	28.6	26.7
r	1	0.89	0.85	0.88

detection of the onset and cessation dates for these dry stations and contributes to increasing the signal of regional-scale interannual variability and spatial coherence that does exist at these dry stations.

All experiments show a quite similar interannual variability of the regional onset date over the period 1961–2001, as displayed by box plots of the PC1 scores for each year (Fig. 7, with the ensemble mean in bold: Fig. 7a for the short rains and Fig. 7b for the long rains). The example of the short rains shows that in some years, however, there are marked differences in the onset date anomalies between the 56 experiments, as shown by the wider boxes (e.g., in 1971, 1973, 1980, 1993, or 1995). These years often correspond to abnormally dry seasons, for which the influence of the rainfall thresholds used to define the onset is more pronounced. Wet seasons (e.g., in 1961, 1967, 1972, 1977, 1982, and 1997) are characterized by a very small spread between the 56 experiments. Over the whole study area and the entire period 1961–2001, the uncertainty, defined as the internal variance computed from the 56 PC1 scoring coefficients of each year, remains relatively weak (Table 1). Finally, the method used to replace the undefined onset date has a weak influence on the interannual variability of the regional onset date over the period 1961–2001 (Table 2).

It is interesting to note that there is also a significant relationship between the regional-scale onset dates and the regional-scale seasonal amounts. Actually, the correlation between the onset date ensemble means (thick line) and the seasonal rainfall amount (average of the 53 stations, thin line) amounts to -0.69 for the short rains and -0.79 for the long rains, indicating that good (poor) rainy seasons tend to start earlier (later).

It is obvious that the regional-scale signal obtained with a multicombination analysis is valuable for forecasting studies because it emphasizes the maximum covariation between all the local onset dates and partly eliminates variations due to parameterizations. Nevertheless, it is useful to assess which local-scale thresholds maximize the correlation with the regional-scale onset,

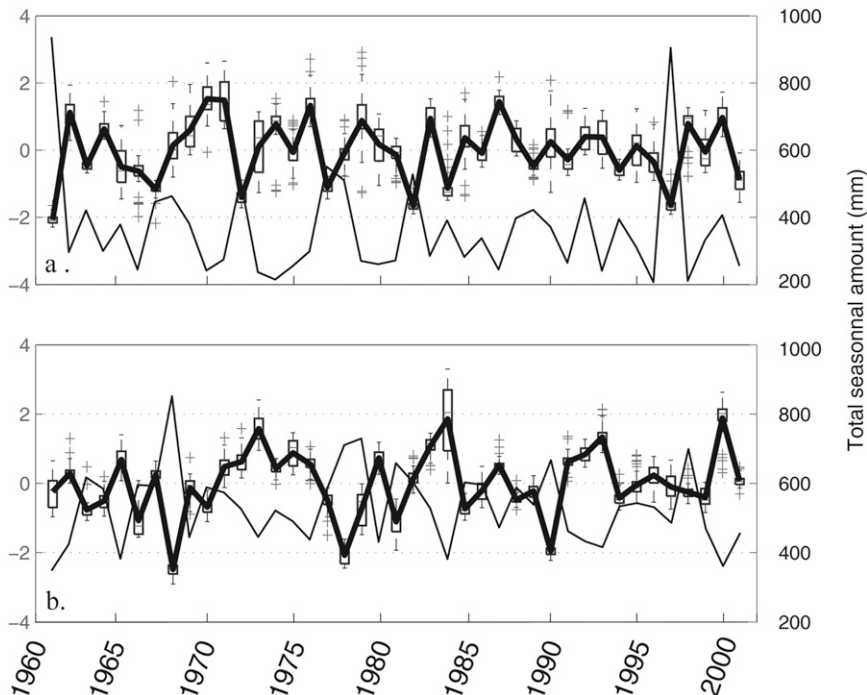


FIG. 7. Interannual variability of onset dates over 1961–2001 for the (a) short rains and (b) long rains, expressed as the standardized PC1 scoring coefficients for the multicomination analysis. Box plots show the spread between the 56 experiments. The thick line is the mean onset date computed from the 56 experiments, and the thin line shows the area-averaged seasonal rainfall amount (mm).

represented by the PC1 ensemble mean. Correlations between each experiment (which represent the local-scale onset dates), and PC1 ensemble mean (which represents the regional-scale onset dates) are computed for each station. The thresholds of the parameters P , N , and C , defined in section 3b, that maximize the correlation are plotted in Fig. 8. Additionally, an analysis of variance of the correlations is carried out separately for each station and each parameter. The aim is to test if a significant part of the variance in the correlations is explained by the threshold values. It is an indication as to whether the given criterion is instrumental in the onset determination.

Figure 8 shows that the interannual variance of local-scale onset date is unequally influenced by the three parameters P , N , and C . For both long rains and short rains, the analysis of variance shows that P controls significantly the interannual variability of onset for most of the stations. This suggests that P is a crucial parameter to define onset. While C also plays a significant role at most stations (Figs. 8c,f), N seems to be of marginal influence (Figs. 8a,d).

The spatial patterns of the thresholds maximizing the correlation between local-scale and regional-scale onset dates are now examined. Parameter N , which varies

between 2 and 5 days, is characterized by a random distribution over the whole study area (Figs. 8a,d). This is consistent with the fact that N is generally not instrumental in the interannual variability of the onset, as shown above by the analysis of variance. By contrast, P , which varies from 10 to 50 mm, shows a much more distinctive geography (Figs. 8b,e). Although there are some exceptions, high thresholds (30–50 mm) tend to prevail in the wet areas in the southwest and low thresholds (10–20 mm) in the dry areas (lowland stations of the north and east). This spatial pattern is particularly obvious for the short rains. Using high P thresholds in dry areas leads to a large proportion of undefined onsets, while too low thresholds in wet areas make any wet spell a candidate onset, which is unrealistic when compared to local knowledge, and results in a very small interannual variability of the onset. This means that the multicomination approach intuitively takes into account the climatological local-scale conditions. As regards C which ranges from 20 to 30 days (Figs. 8c,f), in a majority of cases, especially for the long rains, the use of a longer control period (30 days) is more efficient in reflecting the regional-scale onset. It is likely that the use of too short control periods may in some cases fail to detect false starts. False starts are generally local in character, being

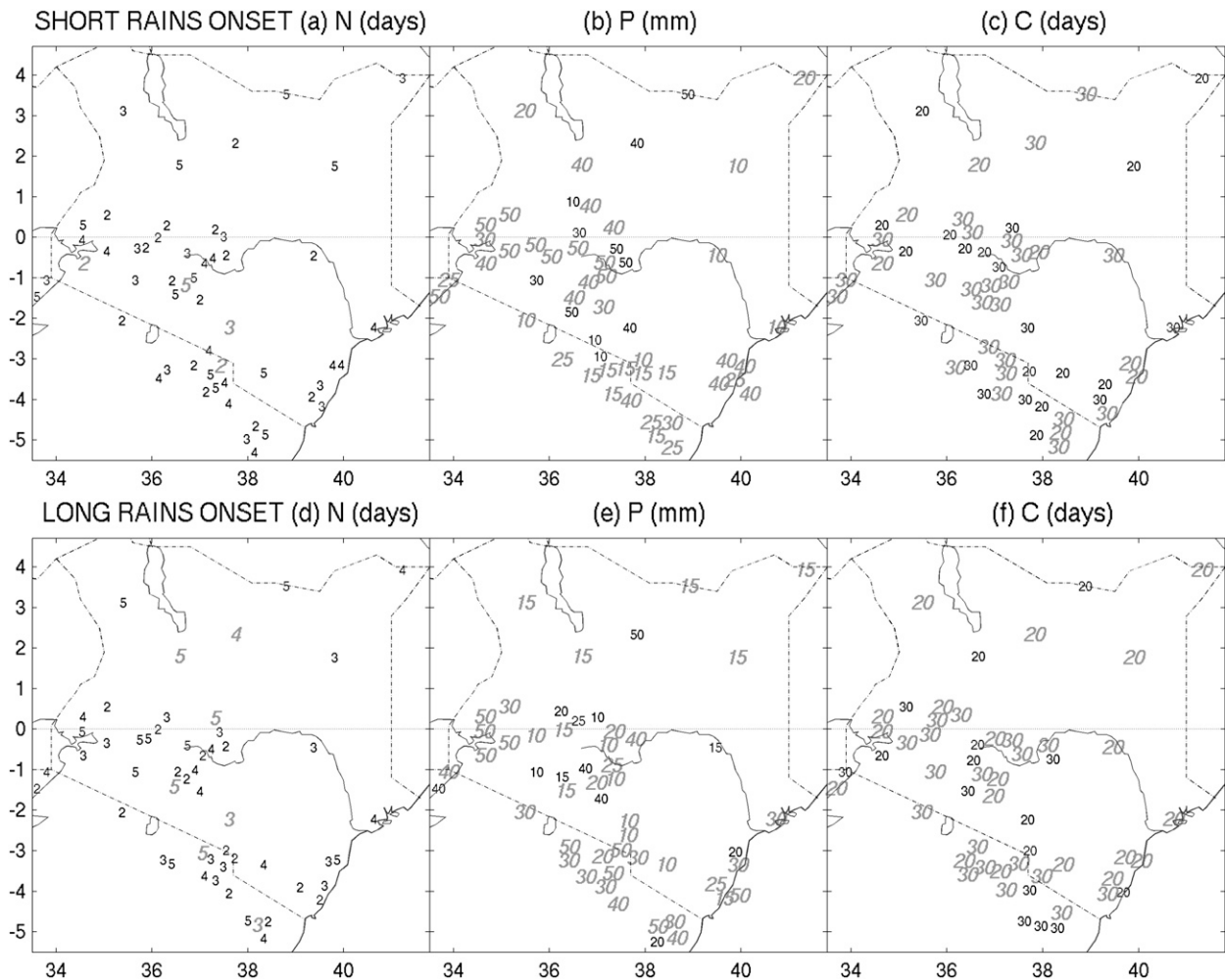


FIG. 8. Spatial distribution of the rainfall parameters that significantly contribute to the variability of the local onset dates and threshold values at which the local onset date is best correlated to the regional-scale onset date (PC1, ensemble mean) for (top) short-rain and (bottom) long-rain onset. Large italic numbers indicate stations at which a given parameter has a significant effect (95% confidence level) on the interannual variability of the onset date, according to the F statistic obtained from a one-way analysis of variance of the correlations with regional onset dates. The values themselves show the threshold value that maximizes the correlation between local and regional onset dates ensemble mean.

related to unseasonable early (local) rains instead of a general shift of the rain belt. However, there are some locations at which a 20-day control period is more in phase with the regional-scale onset. This is particularly so for the short rains and at dry locations for which rainy seasons tend to be short, rainy events scattered, and the onset is undetected when the control period is too long.

5. Conclusions

The main objective of this paper is to present a new approach of extracting the regional-scale signal contained in rainy season onset and cessation dates defined at the local scale and based on multiple combinations of

rainfall parameters. The basic idea behind this new approach is similar to the multiple experiments run with a general circulation model, where the different initial conditions provide the noisy component, while the identical boundary forcing provides the signal. By definition, the noise partly cancels out in the ensemble, allowing the reproducible signal to emerge. In our case, part of the dispersion in the rainy season onset attached to the different combinations of the rainfall parameters is a priori considered as noise, and the ensemble approach helps to detect the reproducible component mostly independent of subjective thresholds and local-scale contingencies. In other words, stations and different parameters could be considered as experiments and

combining them would help to accurately define the signal despite the shortness of the available period. This reproducible signal should fit also more or less with the predictable part of the variability at a seasonal or medium-term time scale.

An elementary comparison of the onset dates computed from two different commonly used definitions, both based on local rainfall distribution, shows significant differences in the rainy season onset dates. The first definition, referred to as agroclimatic, requires a parameterization to detect the rainy season onset or cessation relevant to crop growth and often creates high uncertainties. This definition is characterized by a strong sensitivity to mean rainfall abundance. The alternative method based on accumulated rainfall anomalies is less sensitive to dry conditions but is still sensitive to the choice of the reference period used to compute anomalies (season or year). Overall, even if discrepancies do not occur every year, a unique definition of onset or cessation is always opened to discussion. The determination of the rainy season onset using this type of method is too dependent on the local geographic and climatological context. The new approach used is based on a set of several combinations of three rainfall parameters (P , N , and C) on which we apply a principal component analysis to extract a regional-scale signal. Such a method allows using less stringent thresholds, which is beneficial to the detection of the onset and cessation dates for specific conditions (much drier or wetter than normal) and contributes to increasing the signal of regional interannual variability and spatial coherence. Moreover, the first PC obtained from such a PCA takes into consideration the maximum covariation among all local-scale onset dates. It is therefore relevant for use in predictability studies, for which a prerequisite is that the predictand (here the onset or the cessation dates) shows enough spatial coherence since purely local-scale variability is virtually unpredictable. The regional-scale onset date depicted by the first PC is unequally influenced by the input rainfall parameters used in the raw definition of the local rainy season onsets. Whatever the season, the most critical parameter is P , the amount received during the initial wet spell. Moreover, P shows a distinctive geography that partly reflects the mean seasonal rainfall gradients. Higher thresholds (30–50 mm) tend to prevail in wetter areas and lower thresholds (10–20 mm) in drier areas. While the length of the control period C also plays a significant role at most stations, the number N of consecutive days forming the initial wet spell (from 2 to 5) seems less influential.

The region over which the multicomination approach has been tested is a challenging one, with a wide range of mean precipitation amounts and partially

distinct seasonal regimes. This suggests a good adaptability to other monsoonal areas. However, the approach does not completely alleviate the need for a calibration by forecasters. Forecasters may choose the range of the threshold panel that will be used for their study area according to the range of mean climatic conditions within the station network used and, possibly, the targeted application (pasture growth, crop establishment, etc.). As shown in section 4, a strong relationship exists between the level of onset date uncertainty and the local climate dryness. For a wet region, the decision makers can use a relatively small range of thresholds for P and N for example. For the drier areas, high thresholds may result in a number of undetected onsets and are to be avoided as much as possible, although further work may be useful for an automated selection of the optimal range of thresholds. In East Africa there is no official meteorological definition of the rainy season onset, contrary to what is found in India for instance (e.g., Joseph et al. 2006). Although it may still be possible to define the onset of the rains over East Africa based on meteorological or dynamical criteria, it would require a separate study. The use of the multicomination approach to determine the onset/cessation date of the rainy season can be a good alternative to the use of a meteorological-only definition (based on phenomenon change in the wind flow pattern for instance), which is of regional-scale, and an agroclimatic-only definition, which is based on local-scale conditions. This approach enables one to consider these two scales using the local rainfall information to build an index depicting a regional behavior in the onset of the rains.

Acknowledgments. This research is a contribution to the PICREVAL project, supported by the French National Research Agency (ANR 08-VULN-01-008). Calculations were performed using HPC resources from DSI-CCUB (Université de Bourgogne). We thank the Kenya Meteorological Department, the IGAD Climate Prediction and Application Center, and the Tanzania Meteorological Agency, for providing part of the daily rainfall data.

REFERENCES

- Ati, O. F., C. J. Stigter, and E. O. Oladipo, 2002: A comparison of methods to determine the onset of the growing season in northern Nigeria. *Int. J. Climatol.*, **22**, 731–742.
- Bouali, L., N. Philippon, B. Fontaine, and J. Lemond, 2008: Performance of DEMETER calibration for rainfall forecasting purpose: Application to the July–August Sahelian rainfall. *J. Geophys. Res.*, **113**, D15111, doi:10.1029/2007JD009403.
- Camberlin, P., and M. Diop, 2003: Application of daily rainfall principal component analysis to the assessment of the rainy season characteristics in Senegal. *Climate Res.*, **23**, 159–169.

- , S. Janicot, and I. Pocard, 2001: Seasonality and atmospheric dynamics of the teleconnection between African rainfall and tropical ocean surface temperature: Atlantic vs. ENSO. *Int. J. Climatol.*, **21**, 973–1005.
- Ines, A. V. M., and J. W. Hansen, 2006: Bias correction of daily GCM rainfall for crop simulation studies. *Agric. For. Meteorol.*, **138**, 44–53.
- Jones, J. W., J. W. Hansen, F. S. Royce, and C. D. Messina, 2000: Potential benefits of climate forecasting to agriculture. *Agric. Ecosyst. Environ.*, **82**, 169–184.
- Joseph, P. V., K. P. Sooraj, and C. K. Rajan, 2006: The summer monsoon onset over South Asia and an objective method for the date of monsoon onset over Kerala. *Int. J. Climatol.*, **26**, 1871–1893.
- Liebmann, B., and J. A. Marengo, 2001: Interannual variability of the rainy season and rainfall in the Brazilian Amazon basin. *J. Climate*, **14**, 4308–4318.
- Lo, F., M. C. Wheeler, H. Meinke, and A. Donald, 2007: Probabilistic forecasts of the onset of the north Australian wet season. *Mon. Wea. Rev.*, **135**, 3506–3520.
- Marteau, R., V. Moron, and N. Philippon, 2009: Spatial coherence of monsoon onset over western and central Sahel (1950–2000). *J. Climate*, **22**, 1313–1324.
- Meinke, H., and R. C. Stone, 2005: Seasonal and inter-annual climate forecasting: The new tool for increasing preparedness to climate variability and change in agricultural planning and operations. *Climatic Change*, **70**, 221–253.
- Moron, V., A. W. Robertson, M. N. Ward, and P. Camberlin, 2007: Spatial coherence of tropical rainfall at the regional scale. *J. Climate*, **20**, 5244–5263.
- , A. Lucero, F. Hilario, B. Lyon, A. W. Robertson, and E. D. Dewitt, 2009: Spatio-temporal variability and predictability of summer monsoon onset over the Philippines. *Climate Dyn.*, **33**, 1159–1177.
- , A. W. Robertson, and J. H. Qian, 2010: Local versus regional-scale characteristics of monsoon onset and post-onset rainfall over Indonesia. *Climate Dyn.*, **34**, 281–299.
- Mugalavai, E. M., E. C. Kipkorir, D. Raes, and M. S. Rao, 2008: Analysis of rainfall onset, cessation and length of growing season for western Kenya. *Agric. For. Meteorol.*, **148**, 1123–1135.
- Nicholson, S. E., 1996: A review of climate dynamics and climate variability in eastern Africa. *The Limnology, Climatology and Paleoclimatology of the East African Lakes*, T. C. Johnson and E. O. Odada, Eds., Gordon and Breach, 25–56.
- Ogallal, L. J., J. E. Janowiak, and M. S. Halpert, 1988: Teleconnection between seasonal rainfall over East Africa and global sea-surface temperature anomalies. *J. Meteor. Soc. Japan*, **66**, 807–822.
- Philippon, N., F. J. Doblas-Reyes, and P. M. Ruti, 2010: Skill, reproducibility and potential predictability of the West African monsoon in coupled GCMs. *Climate Dyn.*, **35**, 53–74.
- Rowell, D. P., 1998: Assessing potential seasonal predictability with an ensemble of multi-decadal GCM simulations. *J. Climate*, **11**, 109–120.
- , J. M. Ininda, and M. N. Ward, 1994: The impact of global sea surface temperature patterns on seasonal rainfall in East Africa. *Proc. Int. Conf. on Monsoon Variability and Prediction*, Trieste, Italy, WMO, 666–672.
- , C. K. Folland, K. Maskell, and M. N. Ward, 1995: Variability of summer rainfall over tropical North Africa (1906–1992): Observations and modeling. *Quart. J. Roy. Meteor. Soc.*, **121**, 669–704.
- Sivakumar, M. V. K., 1988: Predicting rainy season potential from the onset of rains in southern Sahelian and Sudanian climatic zones of West Africa. *Agric. For. Meteorol.*, **42**, 295–305.
- Smith, I. N., L. Wilson, and R. Suppiah, 2008: Characteristics of the northern Australian rainy season. *J. Climate*, **21**, 4298–4311.
- Stern, R. D., M. D. Dennett, and D. J. Garbutt, 1981: The start of the rains in West Africa. *Int. J. Climatol.*, **1**, 59–68.
- Tadesse, T., M. Haile, G. Senay, B. D. Wardlow, and C. L. Knutson, 2008: The need for integration of drought monitoring tools for proactive food security management in sub-Saharan Africa. *Nat. Resour. Forum*, **32**, 265–279.
- Usman, M. T., and C. J. C. Reason, 2004: Dry spell frequencies and their variability over southern Africa. *Climate Res.*, **26**, 199–211.
- Verdin, J., R. Lietzow, J. Rowland, R. Klaver, B. Reed, and V. French, 2000: A comparison of methods for estimating start-of-season from operational remote sensing products: First results. USGS-EROS Data Center/Sioux Falls College Rep., 13 pp.
- Wallace, J. S., 1991: The measurement and modelling of evaporation from semi-arid land. Soil water balance in the Sudano-Sahelian zone, IAHS Publ. 199, 131–148.