HAL open science

Medium Lead-Time Predictability of Intraseasonal Variability of Rainfall in West Africa

Benjamin Sultan, Serge Janicot, Cyrille Correia

To cite this version:

Benjamin Sultan, Serge Janicot, Cyrille Correia. Medium Lead-Time Predictability of Intraseasonal Variability of Rainfall in West Africa. Weather and Forecasting, 2009, 24 (3), pp.767. 10.1175/2008WAF2222155.1 . hal-00763171

HAL Id: hal-00763171

https://hal.science/hal-00763171v1

Submitted on 22 Oct 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

SPECIAL
West African Weather Prediction and Predictability
COLLECTION

Medium Lead-Time Predictability of Intraseasonal Variability of Rainfall in West Africa

BENJAMIN SULTAN, SERGE JANICOT, AND CYRILLE CORREIA

IRD–LOCEAN/IPSL, Universite Pierre et Marie Curie, Paris, France

(Manuscript received 6 May 2008, in final form 24 December 2008)

ABSTRACT

The variability of the West African monsoon on the intraseasonal time scale is a major issue for agricultural strategy, as the occurrence of dry spells can strongly impact yields of rain-fed crops. This study investigates this intraseasonal variability of rainfall over West Africa and gives a first overview of its predictability at a medium lead time.

A statistical method, the singular spectrum analysis, is applied to a ground-based rainfall index in West Africa to describe first temporal patterns of the main leading modes of intraseasonal variability. The results point out the existence of one oscillatory mode of 34 days, one of 20 days, and one of 14 days. The same methodology is applied to rainfall from two reanalysis datasets and to deep convection from satellite data in order to assess the accuracy of the representation of intraseasonal variability in these datasets. It is shown that although the day-to-day variability of rainfall is not well captured in these datasets, intraseasonal features and, in particular, the low-frequency mode are very well reproduced.

The medium lead-time predictability (5–10 days) of the intraseasonal modes is investigated using both the dynamical forecast scheme of the ECMWF and a statistical method, the maximum entropy method. For the latter method, an operational application using unfiltered input data is also considered. The performance of these prediction schemes is compared using a simple reference technique in which forecasts are based entirely on persistence. It is found that statistical predictions are much more promising than the dynamical ones, though they encounter problems when applied operationally. In an operational application, the forecast skill for the 10–90-day intraseasonal band is low but the predictability of individual intraseasonal modes is higher. The stability of the forecast skill levels is influenced by the characteristics of the intraseasonal mode. When the characteristics (i.e., amplitude and period) of the considered intraseasonal mode are well defined, skillful forecasts can be obtained. However, when the characteristics change rapidly, the forecast fails.

1. Introduction

West African societies depend heavily on summer monsoon rainfall, especially in the Sahel where rain-fed crop production is the main source of food and income of one of the world's most rapidly growing populations (Baron et al. 2005). Predicting the fluctuations of the West African monsoon would be greatly beneficial to regional agriculture and water resource management. The distribution of rainfall within the rainy season is of particular importance to agricultural strategy (Ingram et al. 2002), as the occurrence of dry spells can strongly impact yields of rain-fed crops (Sultan et al. 2005). Although there is more and more evidence of specific intraseasonal variability in convective activity during the West African monsoon (Janicot and Sultan 2001; Sultan and Janicot 2003; Matthews 2004; Mounier and Janicot 2004; Mounier et al. 2008), no study has investigated its predictability. Nevertheless, there are many examples of skillful forecasts of intraseasonal variability of convection in other regions of the tropics. Most of these examples concern the prediction of the Madden–Julian oscillation (MJO), which is the dominant oscillatory mode in the tropics (Madden and Julian 1972). Skillful predictions of the MJO have been obtained at a medium lead time (less than 10 days) using either dynamical forecasts or statistical methods. For instance, the National Centers for Environmental Prediction (NCEP) Medium-Range Forecast (MRF) model shows skillful forecasts of convection anomalies associated with the MJO for up to 7 days (Hendon et al. 2000; Waliser et al.

Corresponding author address: B. Sultan, IRD–LOCEAN/IPSL, Universite Pierre et Marie Curie, Boite 100, 4 Place Jussieu 75252, Paris CEDEX 05, France.
E-mail: benjamin.sultan@locean-ipsl.upmc.fr

DOI: 10.1175/2008WAF2222155.1

© 2009 American Meteorological Society

Unauthenticated | Downloaded 10/22/21 12:31 PM UTC

1999). On the other hand, several statistical methods, such as principal oscillation pattern (POP) techniques (von Storch and Xu 1990) or singular value decomposition (SVD) methods (Waliser et al. 1999), have been used to produce skillful forecasts of large-scale intraseasonal anomalies of convection. Until now, the comparisons between dynamical and empirical predictions of the intraseasonal variability of convection indicate modeling progress must be made in achieving the likely potential of dynamic models (Waliser et al. 1999; von Storch and Baumhefner 1991).

The aim of this paper is to give a first overview of the predictability of the intraseasonal variability of rainfall over West Africa at a medium lead time. We use a statistical method, singular spectrum analysis (SSA), which has already provided promising results in filtering and predicting intraseasonal oscillations of convection (Mo 2001). The SSA (Vautard and Ghil 1989; Vautard et al. 1992; Ghil et al. 2002) is related to empirical orthogonal functions (EOFs) but is applied to lagged time series providing SSA modes that correspond to intraseasonal oscillations in a frequency band. We apply SSA to rainfall amounts from three different sources: the Institut de Recherche pour le Developpement (IRD) rainfall dataset and the NCEP–National Center for Atmospheric Research (NCAR) reanalysis and the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40). In addition, we use satellite outgoing longwave radiation (OLR) data to estimate deep convection. The approach is twofold.

We first document the intraseasonal variability in the IRD rainfall dataset, providing the most reliable measurement of the ground-based truth. We apply the SSA to a rainfall index in West Africa computed from the IRD rainfall dataset to statistically extract the main leading modes of intraseasonal variability. The same methodology is applied to rainfall from the two reanalysis datasets and to deep convection from OLR data in order to assess the accuracy of the representation of intraseasonal variability in these datasets.

Second, the medium lead-time predictability (5–10 days) of these intraseasonal modes is documented using both the maximum entropy method (MEM; Burg 1968; Penland et al. 1991) and the dynamical forecast scheme of the ECMWF. The performance of these two prediction schemes is compared using a simple reference technique in which forecasts are based entirely on persistence.

Section 2 introduces the datasets used to describe the intraseasonal variability of the West African monsoon. Section 3 describes the SSA and its application toward extracting and predicting oscillatory modes. Section 4 provides the main results of the study. Section 5 concludes the study and discusses future steps.

2. Datasets

a. The IRD daily rainfall

Daily rainfall amounts at stations located in the West African domain 3°–20°N, 18°W–25°E have been compiled by IRD, the Agence pour la Securite de la Navigation Aerienne en Afrique et a Madagascar (ASECNA), and the Comite Interafricain d'Etudes Hydrauliques (CIEH). These data are available for the period 1968–90, and this dataset includes more than 1300 stations for the period 1968–80 and between 700 and 860 stations from 1981 to 1990. These daily values were interpolated onto the NCEP 2.5° × 2.5° grid by assigning each station daily value to the nearest grid point and averaging all the values related to each grid point. They were also interpolated in time with reference to NCEP daily wind fields since daily rainfall amounts were measured between 0600 local solar time (LST) on the given day and 0600 LST the following day. We applied a time lag of 12 h between the average time of the NCEP daily values at 0900 UTC and an approximated average time of daily precipitation over the West African continent (2100 LST). Duvel (1989) indicates a maximum of high cloud coverage over land between 1800 LST and midnight, and Sow (1997) finds a maximum of half-hourly precipitation over Senegal between 1700 LST and the end of the night, depending on the station. The greatest density of stations is located between 5° and 15°N. Data at 17.5°N can also be taken into account since 30–45 stations are available.

b. The NCEP–DOE and ERA-40 reanalyses

The NCEP and NCAR have completed a reanalysis project with a current version of the MRF model (Kalnay et al. 1996). This dataset consists of a reanalysis of the global observational network of meteorological variables (wind, temperature, geopotential height, humidity on pressure levels, surface variables, and flux variables such as precipitation rate) with a "frozen" state-of-the-art analysis and forecast system at a triangular spectral truncation of T62 to perform data assimilation throughout the period from 1948 to the present. This circumvents problems with previous operational analyses due to changes in techniques, models, and data assimilation. Data are available on a 2.5° × 2.5° grid every 6 h (0000, 0600, 1200, and 1800 UTC) on 17 pressure levels from 1000 to 10 hPa. This study uses the NCEP–Department of Energy (DOE) Atmospheric Model Intercomparison Project (AMIP-II) reanalysis (reanalysis-2),

Unauthenticated | Downloaded 10/22/21 12:31 PM UTC

which is based on the NCEP–NCAR reanalysis but with improvements in the physical parameterizations and error fixes (Kanamitsu et al. 2002).

The ECMWF has released reanalyzed datasets for the time frame 1957–2002 (Uppala et al. 2005). The ERA-40 reanalysis has a finer resolution corresponding to a T159 spectral truncation with 60 vertical levels from 1000 to 0.1 hPa. Data are available on a 1.125° × 1.125° grid every 6 h (0000, 0600, 1200, and 1800 UTC).

In this study, we used rainfall from the two reanalysis datasets covering the period 1 April–31 October 1979–2000 with one value per day obtained by averaging the four outputs of each day.

c. *The OLR/NOAA dataset*

Since its launch in 1974, the National Oceanic and Atmospheric Administration's (NOAA) polar-orbiting Television and Infrared Observation Satellite (TIROS) has enabled the establishment of a quasi-complete series of twice-daily measures of OLR at the top of the atmosphere and at a resolution of 2.5° latitude–longitude (Grueber and Krueger 1984). The Interpolated OLR dataset (Liebmann and Smith 1996) provided by the Climate Diagnostics Center has been used here. In tropical areas, deep convection and rainfall can be estimated through low OLR values. The local time the measurements were taken during the period 1979–2000 varied between 0230 and 0730 LST in the morning and between 1430 and 1930 LST in the afternoon. Since the deep convection over West Africa has a strong diurnal cycle, the sample of daily OLR based on two values separated by 12 h suffices in providing a daily average. Moreover, this dataset has already been widely used for tropical studies. We used data covering the period 1 April–31 October 1979–2000, with one value per day obtained by averaging the two outputs of each day.

3. Methods

From the rainfall datasets described above, we first compute a rainfall index over West Africa by averaging daily rainfall data between 5°N and 17.5°N and between 10°W and 10°E. A similar index for deep convection is computed using OLR data. The spatial domain used for the average corresponds to the area of the maximum variance of convection over West Africa where the intertropical convergence zone (ITCZ) is located from late spring to early autumn. These ITCZ indexes are computed from March to October using each dataset separately. The longest overlapping period used here is 1979–90. We then apply the SSA and MEM to document and predict the intraseasonal variability of rainfall over West Africa.

a. *The SSA*

To document the rainfall variability in the intraseasonal band, data must be filtered to retain fluctuations in the band. Whereas a filter based on the Fourier decomposition is well adapted for time series with fixed amplitudes, phases, and periods, one requires other methods to deal with the noisy and stochastic nature of convection variability (Mo 2001). The SSA is an interesting alternative as it is designed to extract information from noisy time series and can be used to compute an adaptive filter (Ghil et al. 2002). The SSA is a statistical tool based on an EOF analysis applied to time series (Vautard and Ghil 1989; Vautard et al. 1992; Ghil et al. 2002). Given a time series $x(t)$ of length N, the first step is to embed x in a vector space of dimension M to represent the behavior of the system by a succession of overlapping "views" of the series through a sliding M-point window (Ghil et al. 2002). The embedding procedure generates a matrix $\mathbf{D}$ whose dimensions are $(N - M + 1) \times M$:

$$\mathbf{D} = \begin{bmatrix} x(1) & x(2) & \dots & x(M) \\ x(2) & x(3) & \dots & x(M+1) \\ \vdots & \vdots & \ddots & \vdots \\ x(N-M+1) & x(N-M+2) & \dots & x(N) \end{bmatrix}.$$

The procedure is then similar to a principal component analysis since we compute the $M \times M$ time-lagged covariance matrix $\mathbf{C}$ and extract the M eigenvalues and $\mathbf{M}$ eigenvectors from C. By analogy with the meteorological literature, the eigenvectors are called TEOFs (EOFs in the time domain). Quasiperiodic modes appear as pairs of degenerate eigenvalues that are approximately equal and associated with TEOFs in quadrature. The SSA thus allows isolation of quasiperiodic modes from the initial time series.

The projection of the original time series onto the kth TEOF gives the corresponding principal components (TPCs):

$$\mathrm{TPC}_k\left(t + \frac{M}{2}\right) = \sum_{j=1}^{M} x(t + j - 1)\mathrm{TEOF}_k(j). \qquad (1)$$

The length of each TPC is $N - M + 1$. Each TPC isolates an oscillatory component defined in a very short-range window of the spectral domain. This oscillatory component explains a part of the variance of the original time series. This explained variance decreases from the first to the Mth mode. One can reconstruct the part of the original time series (RC_k) associated with the mode k by combining the kth TEOF and the kth TPC (Vautard et al. 1992). The reconstructed components

Unauthenticated | Downloaded 10/22/21 12:31 PM UTC

(RCs) are defined in the time space (the total length is N), which is an advantage over the TPCs, which cannot be directly compared to the original time series. The TPCs thus contain the phase information. The RCs are additive. The entire original time series can be reconstructed by summing up the M reconstructed components, RC_k:

$$x(t) = \sum_{k=1}^{M} \mathrm{RC}_k(t). \tag{2}$$

As the TPCs isolate oscillatory modes in a given intraseasonal band, one can filter the original time series by partially summing up a subset of RCs in the intraseasonal band of interest. The procedures are similar to the Fourier techniques but based on TEOF functions instead of sines and cosines (Mo 2001).

The choice of the window size M is arbitrary. It must be large enough to get as much information as possible and yet small enough to ensure many repetitions of the original signal by maximizing the ratio N/M (Ghil et al. 2002). The M selection should also accommodate the oscillatory modes in the intraseasonal band. As described by previous works addressing the question of the intraseasonal time scale of convective activity in the West African monsoon (Janicot and Sultan 2001; Sultan and Janicot 2003; Matthews 2004; Mounier and Janicot 2004; Mounier et al. 2008), the intraseasonal activity has been found to be in two modes: one of around 15 days and one of around 40 days. Thus, $M = 40$ has been used to accommodate the latter intraseasonal mode at 40 days. Similar analyses have been performed for different values of M (50 and 60) and the results are not very sensitive to the window size M.

b. Intraseasonal predictions

As the SSA allows the capturing of quasiperiodic modes from noisy and stochastic time series, the use of MEM was considered to be an interesting way of predicting intraseasonal variability. Autoregressive (AR) processes can do a good job when applied to narrowband time series such as the ones isolated by the SSA. The combination between SSA and MEM has already found applications in climate science for forecasting the Southern Oscillation index (Keppenne and Ghil 1992) and the intraseasonal variability of convection in California (Mo 2001). The authors used the SSA to filter the original time series and then used an AR process based on the MEM to extrapolate the obtained TPCs (Mo 2001) or RCs (Keppenne and Ghil 1992; Keppenne and Ghil 1993). Given the MEM properties (Childers 1978), a time series $x(t)$ can be extrapolated up to $x(t + L)$ by applying an AR process of order L as follows:

$$x(t + L) = \sum_{k=1}^{L} a(k)x(t + M - k) + \xi, \tag{3}$$

where ξ is a white-noise process and $a(k)$ the kth AR coefficient. The order L must be chosen before applying the AR process. The L AR coefficients a can be determined by several methods: for example, the Yule–Walker approach (Yule 1927; Walker 1931) or the Burg method (Burg 1968). The latter is more fitted to the SSA filter as it is based on the symmetric Toeplitz matrix structure of the autocovariance matrix (Penland et al. 1991). The choice of L can be determined by minimizing the Akaike information criterion (AIC; Akaike 1974). In this study, we consider two applications of the SSA–MEM combination:

First, we examine the predictability of the SSA modes. The ITCZ index is first split into two samples: a training period from 1979 to 1984 and a test period from 1985 to 2000. We compute a TEOF basis over the training period 1979–84 and project the unfiltered but seasonally adjusted ITCZ index (the annual and semiannual cycles defined by the first and second harmonics are removed) over the test period in order to get SSA modes after 1984. We then quantify the predictability of the intraseasonal modes by extrapolating separately each RC using the MEM [Eq. (3)]. Two time lags are explored: 5 and 10 days.

To evaluate the accuracy of the MEM forecasts, we use two skill measures. First, we compute the correlation coefficients between the observed and the predicted intraseasonal modes over the test period 1985–2000. The significance of the correlations is established by setting up a test reproducing our experimental conditions: we generate 10 000 random Gaussian time series with the same length as the observed rainfall time series. We extract the SSA modes from these random time series in the same way as is done with our observed rainfall time series. We then calculate the correlation between each random SSA mode and the observed ones and compute the probability distribution function (PDF) of these correlations. We then take correlation values higher than the 95% random PDF value to be significant at the 95% confidence interval. Second, we generate a skill score (later SS):

$$\mathrm{SS} = \frac{\mathrm{MSE} - \mathrm{MSE}_{\mathrm{benchmark}}}{-\mathrm{MSE}_{\mathrm{benchmark}}}, \tag{4}$$

where MSE is the mean square error and $\mathrm{MSE}_{\mathrm{benchmark}}$ is the MSE of a simple reference technique in which forecasts are based entirely on persistence [the last observation $x(t)$ is used as the predicted value $x(t + p)$ at the lag p]. The SS accounts simultaneously for correlation

Unauthenticated | Downloaded 10/22/21 12:31 PM UTC

and amplitude. Negative (positive) values of SS mean that the forecast method performs lower (better) than persistence. The range of possible positive SS values goes from 0 to 1, where a perfect forecast system has a value of 1 and a forecast system with no information with respect to persistence has an value of 0 (MSE being equal to $MSE_{benchmark}$).

The MEM forecast skill is also compared to the skill of the dynamical forecasts running operationally at ECMWF. We use the medium-range deterministic forecast model rainfall outputs [archived since 1985 into the ECMWF's Meteorological Archival and Retrieval System (MARS)] based on the 1200 UTC high-resolution forecast model. Rainfall forecasts are produced from 3 to 72 h and every 6 h from 72 to 240 h. The 5- and 10-day forecasts are extracted and averaged over the ITCZ area. This predicted index over the 1985–2000 period is then filtered by using SSA and compared to the index predicted using SSA–MEM.

Second, we construct a prediction scheme to address an operational objective. This method differs from the SSA–MEM described below in one way: the projection of the ITCZ index onto the TEOF basis is not just done once during the test period but is applied iteratively throughout the period. To predict the day $T_o + p$, we use data from $(T_o - N + 1)$ to T_o, where N is the length of the training period from 1979 to 1984 and T_o is the last day of the training period, that is, 30 October 1984. The seasonally adjusted ITCZ index is filtered using the SSA (see below) and the RCs are extrapolated from T_o to $T_o + p$. The whole procedure is applied iteratively by incrementally increasing the value of T_o from the last day of the training period to the end of the test period. Such a procedure is adapted to real-time data acquisition systems with new measurements every day, such as OLR data, and thus can easily be used for operational applications. It is important to note that in this prediction scheme we use the SSA as a real-time filter, which implies the presence of edge effects that might impact the prediction skill at any lead time. The skill of this operational method is compared to the persistence-based prediction skill by computing the SS. However, because of inconsistencies between the ERA-40 reanalyses and the ECMWF dynamical forecasts, it is not possible to directly compare the dynamical forecast skill with the skill of the SSA–MEM operational prediction scheme.

Note that for these two SSA–MEM applications, we have checked the robustness of the results by examining their sensitivity to the choice of the training/test periods lengths, the determination of the order of the AR process and the extrapolation of the TPCs instead of the RCs as suggested by Mo (2001). It appears that the results are not very sensitive to these choices.

4. Results

a. *The intraseasonal modes within the ITCZ rainfall index*

Figure 1a shows an example of the intraseasonal variability of IRD rainfall captured by the ITCZ index for 1979. Gray bars represent the average daily rainfall time series in the ITCZ box. Note that the seasonal cycle has been removed. The day-to-day variability of rainfall is very clear, characterizing the influence of synoptic-scale weather systems such as easterly waves (Diedhiou et al. 1999) and mesoscale convective systems (Mathon and Laurent 2001). Moreover, sequences of persistent high or low rainfall amounts can be observed throughout the year. The lengths of such sequences vary from a very long dry spell from mid-June to mid-July to a more rapid alternating pattern of 5-day high and low rainfall sequences starting in mid-September. The thick line in Fig. 1 represents the 10–90-day bandpass-filtered rainfall. It is obvious from this filtered index that rainfall is not modulated at a single intraseasonal time scale but at several different time scales, the influence of which seems to be intermittent throughout the year. Using wavelet analysis (Torrence and Compo 1998) applied to rainfall time series, Sultan et al. (2003) have documented this intraseasonal time-scale variability as intermittent signals with more variance within two period intervals of 10–25 and 25–60 days, respectively. To better characterize the intraseasonal time scale of rainfall fluctuations, we apply the SSA to the 10–90-day ITCZ index based on IRD rainfall over the 1979–90 period. Table 1 shows the explained variance of the first 10 TEOFs. Most of the variance is explained by the first 10 TEOFs (99.9% of the total variance in Table 1). Oscillatory modes can be detected by pairs of eigenvalues that are approximately equal and by TEOFs in quadrature. The first oscillatory mode is captured by the first pair of eigenvalues associated with two TEOFs in quadrature (Fig. 2) and characterized by a period of 34 days. This low-frequency mode explains 38.4% of the 10–90-day variance (Table 2). The time series of this mode (later mode 1) is given by the sum of RC_1 and RC_2 [see Eq. (2)]. A MEM spectrum (Burg 1968) applied to this time series finds the power to be between 27 and 90 days. An example of the time evolution of this first mode is shown for 1979 as a thin line in Fig. 1b. It is compared to the bandpass filter in the corresponding intraseasonal band of 27–90 days. The next pair of eigenvalues represents another oscillatory mode with a higher frequency. It explains 27.1% of the 10–90-day variance (Table 2). TEOFs 3 and 4 show periods of around 20 days. Applying the MEM spectrum to the reconstruction based on these two TEOFs highlights the

Unauthenticated | Downloaded 10/22/21 12:31 PM UTC

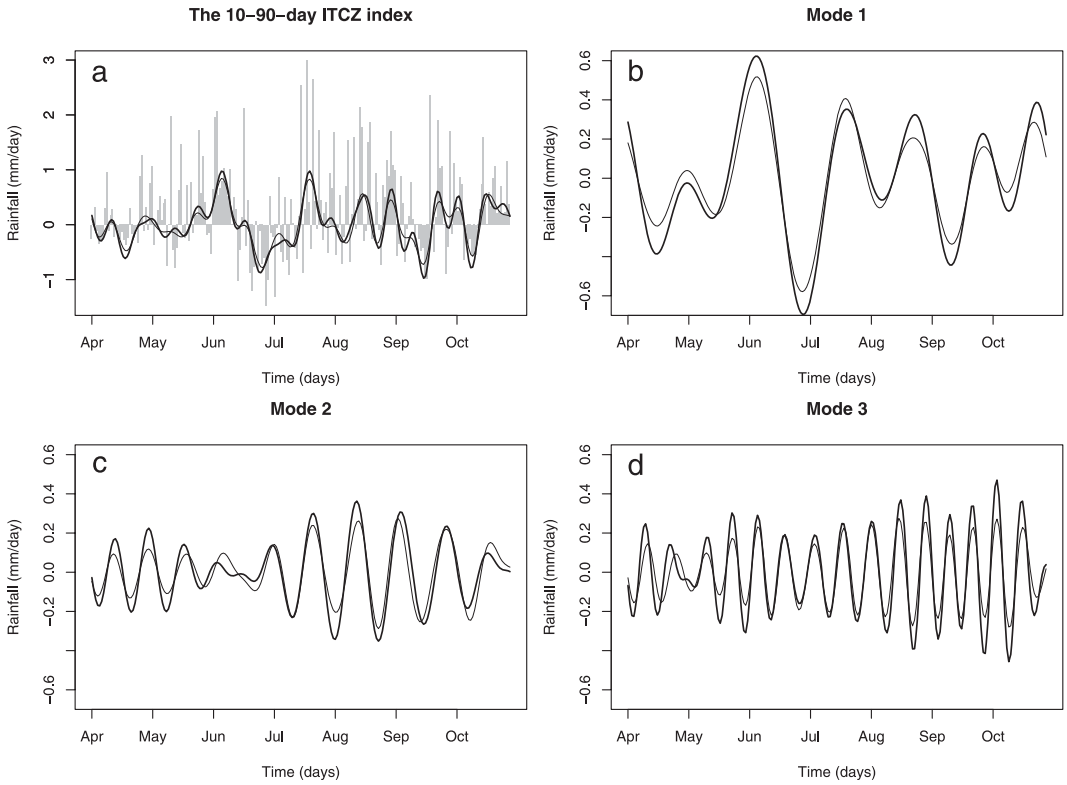

FIG. 1. An example of the SSA intraseasonal modes for the year 1979. (a) The seasonally adjusted unfiltered ITCZ index (gray bars), 10–90-day bandpass-filtered ITCZ index (thick line), and SSA filtered ITCZ index (thin line). The SSA filter time series is the sum of the six first RCs. (b) The 26–90-day bandpass-filtered ITCZ index (thick line) and the first SSA mode (thin line). The mode 1 time series is given by the sum RC_{1+2}. (c) The 16–28-day bandpass-filtered ITCZ index (thick line) and the second SSA mode (thin line). Mode 2 is given by the sum RC_{3+4}. (d) The 11–17-day bandpass-filtered ITCZ index (thick line) and the third SSA mode (thin line). Mode 3 is given by the sum RC_{5+6}.

variance between 16 and 28 days. The time series of this second mode (the sum of RC_3 and RC_4; later mode 2) is very close to that of a 16–28-day bandpass filter (see Fig. 1c). A third quasiperiodic mode is captured by TEOFs 5 and 6, with a period of about 14 days and power between 11 and 17 days. It explains 19.9% of the 10–90-day variance (Table 2). Its time series (the sum of RC_5 and RC_6; later mode 3) is compared to a 11–17-day bandpass filter in Fig. 1d. The periodicity between 11 and 26 days dominates the intraseasonal variability of the ITCZ index. Indeed, the sum of the explained variance of modes 2 and 3 (near 47%) is greater than the explained variance of the low-frequency mode (38.4%). The three oscillatory modes represent more than 85% of the 10–90-day filtered ITCZ index and the time series given by the sum between these three modes (the sum between the six first RCs) is very close to the 10–90-day filtered index with a correlation of up to 0.97 (see Fig. 1a for the comparison between the two time series in 1979). Because of their dominance, we retain only these intraseasonal modes in the following sections of this paper.

To quantify how well the intraseasonal variability of rainfall is represented in other datasets, we perform a similar analysis, applied separately to the 10–90-day ITCZ indexes derived from each dataset described in section 2. We first compute the correlation between the raw IRD rainfall index and the ERA-40, NCEP–DOE, and OLR unfiltered ITCZ indexes (Table 3). Note the indexes have been seasonally adjusted by removing the

TABLE 1. Percentage of explained variance of the 10 first TEOFs.

Mode	Explained variance (%)	Cumulative (%)
1	20.6	20.6
2	17.8	38.4
3	14.1	52.5
4	13.0	65.5
5	10.6	76.1
6	9.3	85.4
7	6.2	91.5
8	5.8	97.3
9	2.1	99.4
10	0.5	99.9

Unauthenticated | Downloaded 10/22/21 12:31 PM UTC

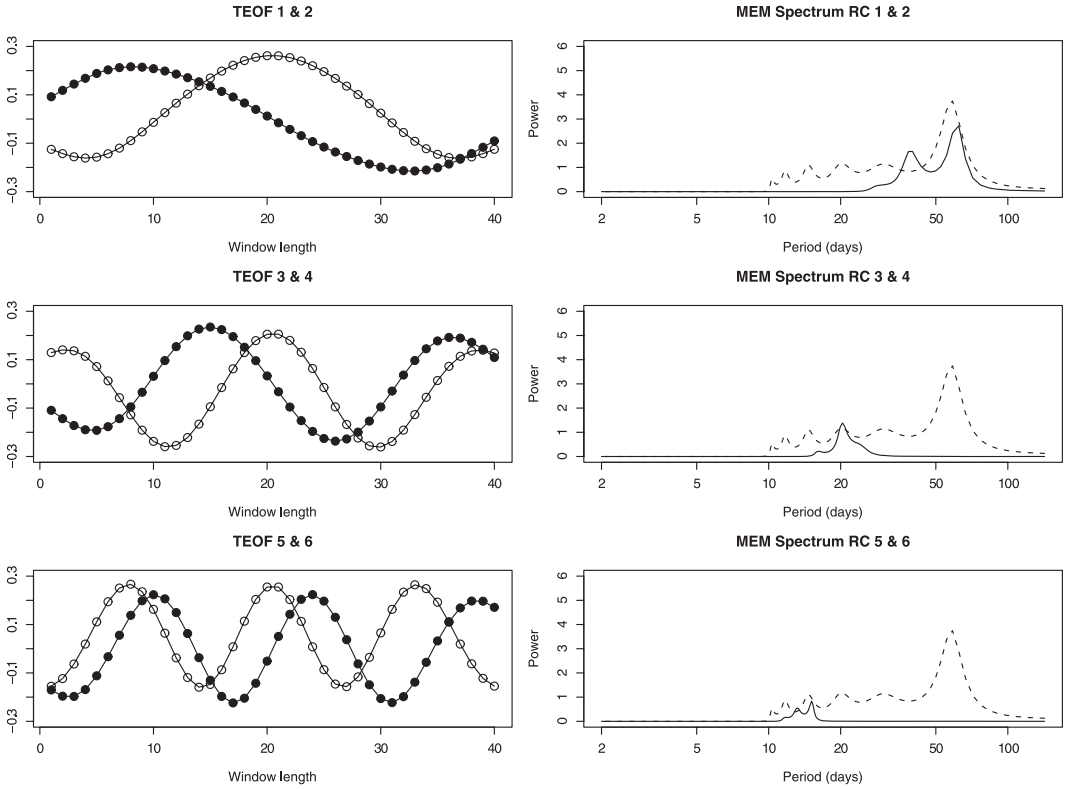

FIG. 2. (left) Pairs of TEOFs corresponding to the three intraseasonal modes captured by the SSA. The open circles represent the TEOFs 1, 3 and 5 and the filled circles, the TEOFs 2, 4 and 6. (right) The MEM spectrum of the ITCZ index (dashed lines) and for each intraseasonal mode (full lines). The SSA–MEM spectrum is applied separately to the sum RC_{1+2}, RC_{3+4}, and RC_{5+6}. The order of the MEM spectrum is 40 days.

first two harmonics. The correlations are weak in particular for the NCEP–DOE and ERA-40 reanalyses and explain no more than 14% and 27% of the rainfall variance, respectively. The inconsistency of the day-to-day rainfall variability is a well-known characteristic of the two reanalysis datasets. At this point, it is important to remember that a reanalysis is a combination of model and measurement, using observations to constrain the dynamical model to optimize between the properties of complete coverage and accuracy (Betts et al. 2006). The relative contribution of model and measurement varies between variables, and rainfall is the most model-dependent variable in the reanalysis production. The fact that such numerical models have difficulties simulating rainfall in convective areas can explain why this is one of the less reliable variables in the reanalysis datasets. Moreover, if deep convection and rainfall can be estimated through low-OLR values in tropical areas, the relatively weak correlation between the OLR and IRD rainfall ITCZ indexes shows that this relation is not strong at the synoptic time scale since the occurrence of convective clouds does not necessarily imply rainfall. In addition, this inconsistency can be partly explained by the fact that the OLR daily mean is an average of two instantaneous measurements during the day while the rainfall data are an integration of measurements from throughout the day. It is interesting to see that despite the fact that the day-to-day rainfall variability is not well reproduced by the different datasets; the intraseasonal variability is much more reliable. This is particularly so with the ERA-40 and OLR datasets, which show respective correlations of 0.74 and 0.70. The decrease in the number of degrees of freedom with the use of the 10–90-day filter is certainly not sufficient to explain the increase in the correlation values. It is more likely that this means intraseasonal rainfall variability is induced by large-scale patterns, which are well reproduced in the

TABLE 2. Percentages of the explained variance of the three first SSA modes and by the sum of these three first modes (mode 1–3).

Mode	IRD	ERA-40	NCEP–DOE	OLR
1 (RC_{1+2})	38.4	38.7	40.3	41.4
2 (RC_{3+4})	27.0	25.0	22.9	24.3
3 (RC_{5+6})	19.9	20.1	19.1	19.0
1–3 ($RC_{1+\ldots+6}$)	85.4	83.8	82.3	84.8

Unauthenticated | Downloaded 10/22/21 12:31 PM UTC

TABLE 3. Correlations between the ITCZ index based on IRD ground-based measurements and the same index based on ERA-40, NCEP–DOE, and OLR data. These correlations are shown for the unfiltered rainfall (first row), for the 10–90-day filtered rainfall (second row), and for the three SSA intraseasonal modes (remaining rows). Note that in the first row, the annual and semiannual cycles defined by the first and second harmonics are removed to compute the correlations. The correlations are computed over the 1979–90 period from 1 April to the end of October. All values are significant at the 95% confidence interval.

	ERA-40	NCEP–DOE	OLR
Raw ITCZ index	0.52	0.37	0.58
10–90-day ITCZ index	0.74	0.58	0.70
Mode 1 (RC_{1+2})	0.76	0.69	0.74
Mode 2 (RC_{3+4})	0.73	0.46	0.68
Mode 3 (RC_{5+6})	0.70	0.39	0.62

reanalysis datasets and more easily estimated by the convection monitoring through OLR than individual rainy events. Table 3 shows the correlations between the three rainfall intraseasonal modes described below and the modes obtained from ERA-40, NCEP–DOE, and OLR. The percentage of the 10–90-day variance explained by the three modes is shown in Table 2. The contribution of each mode to the 10–90-day signal is very close in the different datasets although it seems that the NCEP–DOE and OLR data overestimate the contribution of the low-frequency mode and underestimate (in particular NCEP–DOE) the contribution of the higher-frequency mode. The first mode is well reproduced in all datasets with the highest correlation obtained by ERA-40 ($R = 0.76$). The two other modes are also well reproduced using ERA-40 and OLR. Nevertheless, NCEP–DOE fails to represent these shorter modes with correlation coefficients of 0.46 and 0.39, respectively for the second and the third modes. To try to understand where these differences come from, we look at the spectral characteristics of each intraseasonal mode by computing a MEM spectrum (Fig. 3). Although there are some differences in the power amplitude, the spectral characteristics of the three modes are very close in the IRD and ERA-40 rainfall. The spectrum of the first intraseasonal mode from OLR and NCEP–DOE is close to the IRD one although there is an overestimation in the lower frequencies using OLR data. Concerning the second mode, each dataset shows the same 20-day peak as the one obtained with the IRD rainfall but each differs mainly by an overestimation of the 14–19-day signal. This overestimation is particularly obvious using NCEP–DOE. Additionally, NCEP–DOE underestimates the signal in the frequencies that are lower than 20 days while the OLR data show too much signal in the frequencies that are lower than 20 days. The spectrum of the third mode shows two peaks at 13 and 15 days. These two peaks are reproduced quite well using the ERA-40 and OLR datasets but are lagged using NCEP–DOE to show peaks at 12 and 14 days. These differences in the spectral characteristics of the second and third modes in NCEP–DOE can partly explain the weak correlations between these modes and those obtained with IRD rainfall.

Previous studies have already examined the intraseasonal time-scale variability of convection over western and central Africa. Janicot and Sultan (2001) and Sultan et al. (2003) examined the importance of 10–25- and 25–60-day periodicities in rainfall and the convective activity over the Sahel. Mounier and Janicot (2004) extended this work by carrying out an EOF analysis on convection fields during the northern summer over western and central Africa, and showed evidence of two independent modes of variability in the 10–25-day range. The first one (hereafter the Guinean mode) is characterized by a stationary and uniform modulation of convection within the African ITCZ. It is associated with a modulation of the zonal low-level wind over the equatorial Atlantic and a zonal dipole of convection between Africa and the north equatorial Atlantic off the coast of South America (Mounier et al. 2008). The second mode (hereafter the Sahelian mode) is a westward-propagating signal from eastern Africa to the western tropical Atlantic, consistent with the signal already detected over the Sahel (Sultan et al. 2003). On the 25–90-day range, the dominant mode on the global scale is the MJO (Madden and Julian 1972). Matthews (2004) showed that the remote circulation to the MJO over the warm pool sector offers a plausible explanation for the dominant mode of variability in convection over western and central Africa at these time scales. This mode, the first EOF mode of filtered OLR over western and central Africa (hereafter the MJO mode), consists of an enhancement of convection over most of western and central Africa, whose northern part propagates westward from northeast of Lake Chad to the northwestern part of Africa, while its southern part is stationary and increases and weakens along the Guinean coast and over central Africa (Janicot et al. 2009). Twenty days prior to an enhancement of convection over Africa, convection is reduced over the equatorial warm pool. In response to this change in warm pool convection, an equatorial Kelvin wave propagates eastward and an equatorial Rossby wave response propagates westward. Together they complete a circuit of the equator and meet up 20 days later over Africa, favoring an enhancement of deep convection (Matthews 2004).

To better characterize the three SSA modes, they are now interpreted in terms of Guinean, Sahelian, and MJO modes (the EOF modes). First, the time series of

Unauthenticated | Downloaded 10/22/21 12:31 PM UTC

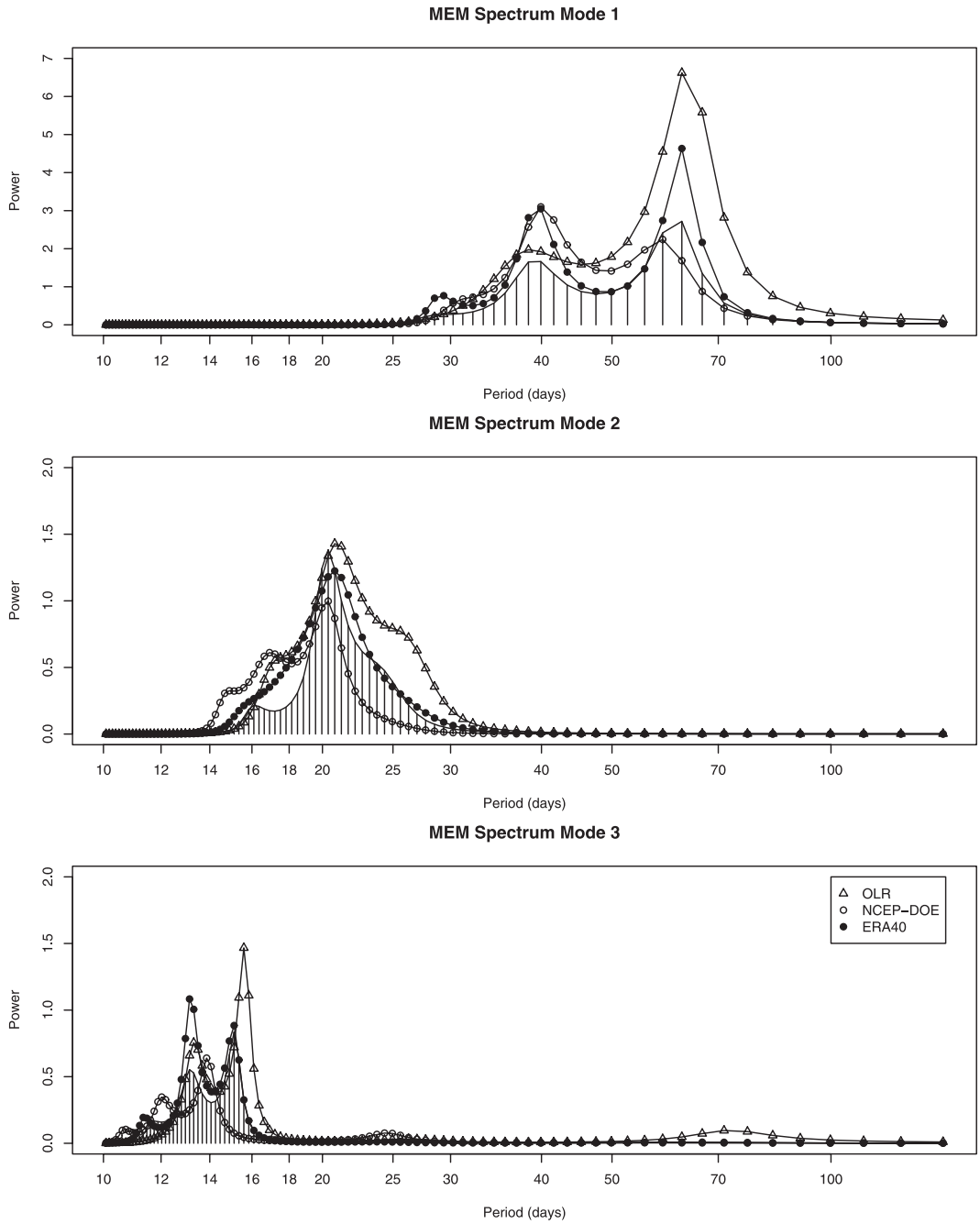

FIG. 3. Spectral characteristics of the (top to bottom) three SSA modes. A MEM spectrum has been computed for each mode (corresponding to the three panels) and for each used dataset. The histogram represents the MEM spectrum of the intraseasonal modes based on the IRD dataset and the different solid lines correspond to the spectrum applied to each of the datasets (see the legend).

the reconstructed OLR signal over the ITCZ area (5°–17.5°N, 10°W–10°E) for each of these modes is computed. Second, correlations between these ITCZ indexes based on the SSA modes and those of the EOF modes are calculated. Third, the composite time sequences of the OLR fields associated with the difference between the highest and the lowest values of the ITCZ indexes, computed from the SSA modes, is drawn and compared to those computed from the EOF modes and previously published (see above). Mode 1 of the SSA is highly correlated with the MJO mode ($R = 0.87$) and its composite time sequence (Fig. 4) is very similar to that

Unauthenticated | Downloaded 10/22/21 12:31 PM UTC

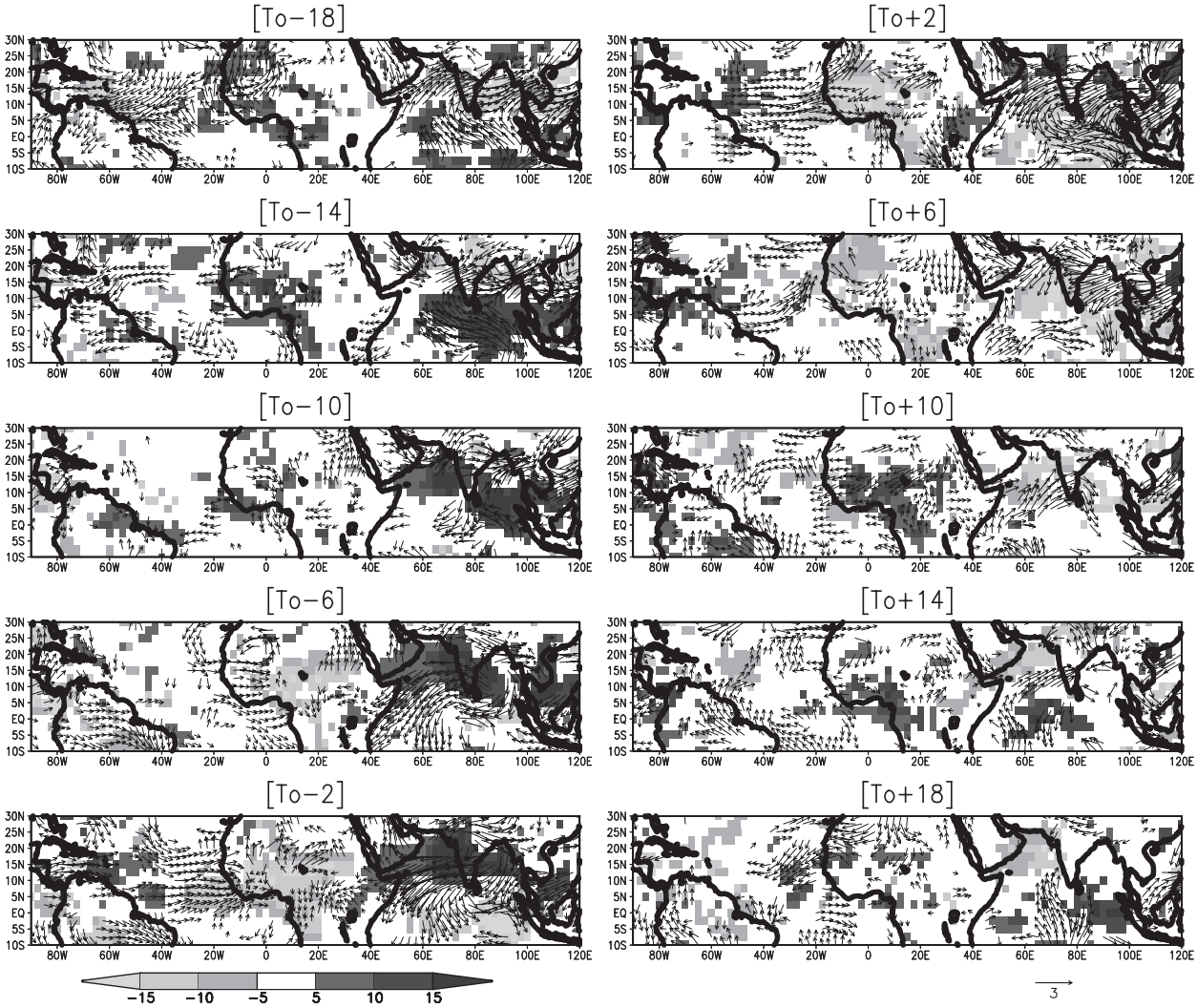

FIG. 4. Composite time sequence based on the SSA mode 1 index. The dates (called t_0) where this index is maximum (minimum) between June and September over the 1979–1990 period are retained to define a weak (high) convective phase. The respective high minus weak composite sequence is shown for the unfiltered OLR and 925-hPa wind fields. This sequence goes from $t_0 - 18$ days to $t_0 + 18$ days with a 4-day lag. Shaded areas represent the OLR anomalies (W m^{-2}). The vector scale (m s^{-1}) is shown in the lower right. Only values significant at 90% are displayed.

associated with the MJO signal shown by Matthews (2004) and Janicot et al. (2009). Figure 4 highlights the negative OLR signal covering the whole region around western and central Africa around t_0, associated with enhanced westerly low-level winds bringing more moisture inland. This OLR signal appears 10 days earlier over the eastern Sahel, then moves and develops widely westward, before disappearing over the western part of Sahara 8 days later, giving way for the occurrence of the reversed phase. The whole cycle visible on this sequence is consistent with the dominant 34-day periodicity identified for this SSA mode 1. Figure 4 also shows the high MJO activity in the Indian–Asian sector, characterized by the northeastward propagation of convective rainbands. As the SSA modes 2 and 3 have closed periodicities (20 and 14 days, respectively) included in the 10–25-day band, they have been combined before being compared to the 10–25-day EOF modes. This combined mode (corresponding to the ITCZ index reconstructed with these two modes) correlates with the Guinean mode at 0.70, with the Sahelian mode at 0.40 and the combined Guinean–Sahelian ITCZ index at 0.79. Moreover, the corresponding composite time sequence (Fig. 5) is very similar to the sequence computed from the combined Guinean–Sahelian ITCZ index (not shown). Its spatial pattern is difficult to interpret as it mixes properties of the Guinean and the Sahelian patterns. This is not surprising as we

Unauthenticated | Downloaded 10/22/21 12:31 PM UTC

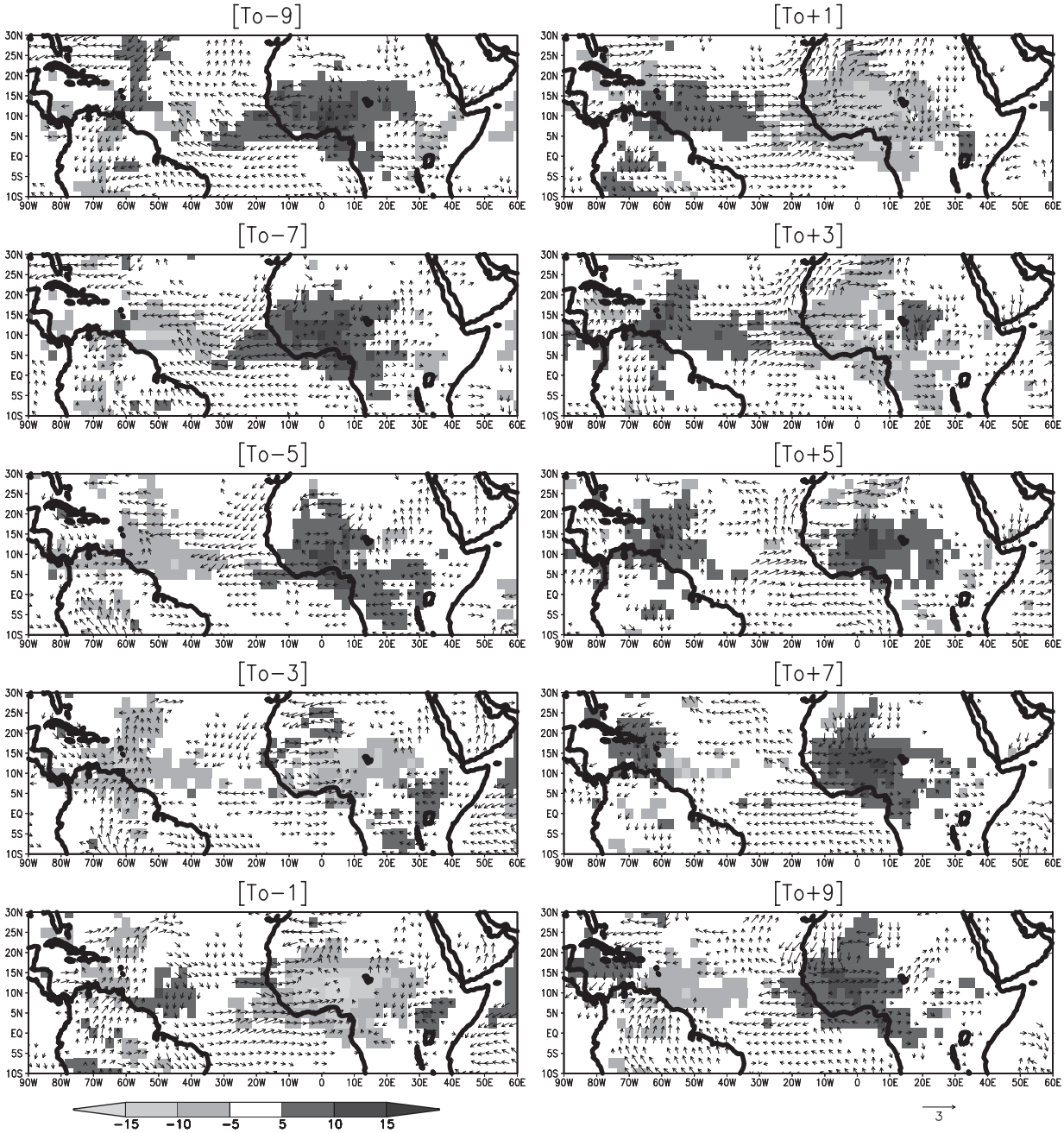

FIG. 5. As in Fig. 4, but for the combination of SSA modes 2 and 3. This sequence goes from $t_0 - 9$ days to $t_0 + 9$ days with a 2-day lag.

use an ITCZ index over the area from 5° to 17.5°N before computing the SSA modes, meaning we cover the areas of influence of both the Guinean and the Sahelian modes. We can conclude from all these results that the SSA mode 1 can be interpreted to be the previously detected MJO mode over Africa (Matthews 2004; Janicot et al. 2009), and that the combination of the SSA modes 2 and 3 well represents the combination of the Guinean and Sahelian modes described in Mounier et al. (2008) and Sultan et al. (2003), respectively.

b. *Predictability of the intraseasonal modes*

We now examine the medium-range predictability of the intraseasonal modes. In the following, we have chosen to work only on the ERA-40 intraseasonal modes in order to use the longest time series (1979–2000) and to

Unauthenticated | Downloaded 10/22/21 12:31 PM UTC

TABLE 4. Correlation coefficients between the observed SSA modes and the 5- and 10-day predicted modes using the SSA–MEM and the dynamical forecasts of ECMWF. The correlations are computed over the 1985–2000 period from 1 April to the end of October. All values have a correlation value significant at the 95% confidence interval.

	SSA-MEM 5 day/10 day	ECMWF 5 day/10 day
10–90 days ($RC_{1+\ldots+6}$)	0.95/0.95	0.36/0.27
Mode 1 (RC_{1+2})	0.99/0.99	0.50/0.35
Mode 2 (RC_{3+4})	0.98/0.98	0.33/0.34
Mode 3 (RC_{5+6})	0.95/0.95	0.23 /0.18

TABLE 5. Skill scores (SSs) of the 5- and 10-day SSA–MEM and ECMWF dynamical forecasts. Negative (positive) values of SS mean that the forecast method performs worse (better) than persistence. The range of possible positive SS values goes from 0 to 1, where a perfect forecast system has a value of 1 and a forecast system with no information has a value of 0. The SS is computed over the 1985–2000 period from 1 April to the end of October.

	SSA–MEM 5 day /10 day	ECMWF 5 day/10 day
10–90 days ($RC_{1+\ldots+6}$)	0.93/0.94	0.19/0.28
Mode 1 (RC_{1+2})	0.99/0.99	−1.63/0.03
Mode 2 (RC_{3+4})	0.99/0.99	0.45/0.65
Mode 3 (RC_{5+6})	0.81/0.60	0.43/−0.24

compare statistical and dynamical predictions obtained by the SSA–MEM and the ECMWF forecast systems. This choice has few consequences on the results since the IRD and ERA-40 are very close (see section 4a) and since the skill of the statistical method does not depend greatly on the input data. We calibrate the SSA–MEM over the 1979–84 period and examine the predictability of the SSA modes from the ERA-40 datasets over the 1985–2000 test period using this statistical method and the ECMWF forecasts. We also attempt to document the predictability of the 10–90-day signal by predicting separately each SSA mode and by summing up the three predicted modes. Table 4 shows the correlations between the observed SSA modes and their predictions using the two methods for two time lags: 5 and 10 days. Table 5 shows the skill score in reference to the persistence-based forecasts. The correlation values and skill scores are quite high, up to 0.95 regardless of the mode or the lag, meaning the MEM is well suited to the oscillatory characteristics of the SSA modes. The ECMWF dynamical forecast, on the other hand, gives lower skill levels although the correlation coefficients are significant at the 95% confidence interval. The highest correlation value is $R = 0.50$ for the 5-day prediction of the first mode, which is lower than the persistence forecast since the skill score value is negative (SS = −1.63). This low skill is quite surprising since the ERA-40 reanalysis dataset can accurately reproduce the rainfall SSA modes. Figure 6 shows the variations of the dynamical forecast skill levels according to the prediction time lag. Although there is a clear decrease in the skill with the time lag increase, the correlations between the observed and predicted modes, although significant at the 90% level, are low even for the 1-day time lag. The discrepancy between the model analyses and forecasts has already been shown by Thorncroft et al. (2003) within the context of the JET2000 experiment. The authors found that although the ECMWF analyses were able to accurately represent the characteristics of the African easterly jet in 2000, despite the absence of upper-air observations at this latitude, it was very difficult to see any African easterly jet at all in the ECMWF 5-day forecasts. Assuming the starting analysis for the forecast was accurate, it is then a concern that in the space of 5 days the model can move so far from the observed states. It clearly indicates some processes are being misrepresented in the region. The examination of the ECMWF-predicted SSA modes reveals the same spectral characteristics as the observed ones but with shifts in the phase. It also reveals a strong intermittency in the quality of the forecasts, with 1–2-month sequences very accurately forecasted while the following months are quite badly predicted. This variance in the forecast skill does not seem to be linked to the characteristics of the intraseasonal mode time series. Regular and strong intraseasonal oscillations can be poorly forecasted while a weaker signal may be much better forecasted.

Since we have attested that the SSA modes are highly predictable using the SSA–MEM approach, we can now look at the application of such a method in an operational way. This application presents a major difficulty, namely the use of raw data to predict filtered data, since most of the typical filtering applications require information beyond the end of the time series. We therefore constructed an iterative prediction scheme adapted from Mo (2001) that is well fitted to real-time data acquisition to predict SSA modes based on unfiltered data (see section 3). The skill of this operational method is shown in Table 6 for the three SSA modes and for the sum of these three modes, which is close to the 10–90-day signal. It is not possible to directly compare the dynamical forecast skill with the skill of the SSA–MEM operational prediction scheme due to inconsistencies between the ERA-40 reanalyses and the ECMWF dynamical forecasts. As expected, the correlation coefficients are weaker than those obtained previously due to edge effects that impact the prediction skill at any lead

Unauthenticated | Downloaded 10/22/21 12:31 PM UTC

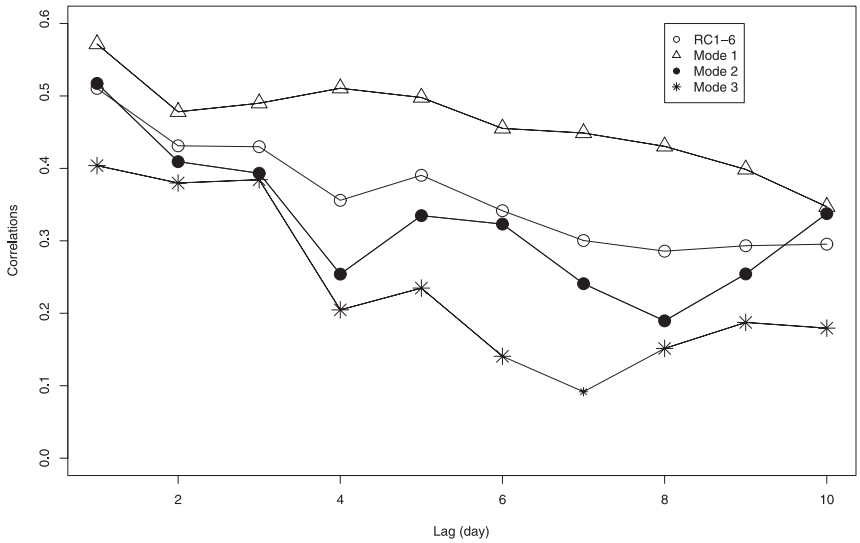

FIG. 6. Correlations between the observed and the predicted SSA modes using the ECMWF forecasts according to the time lag. The correlations are computed over the 1985–2000 period from 1 April to the end of October. A correlation value significant at the 95% confidence interval is plotted using a bigger symbol.

time. The skill score values are also weaker but remain positive, meaning the forecast scheme performs better than the persistence-based forecasts. Although the prediction skill is very low for the 10–90-day intraseasonal band ($R = 0.09$ and $R = 0.06$, respectively, for the 5- and 10-day lags), the individual intraseasonal modes remain predictable. The prediction of the first mode shows the highest skill with a correlation between the observed and the predicted mode of 0.54 at the 5-day lag and 0.36 at the 10-day lag, which is higher than the persistence-based forecasts in particular for the 10-day lag. Note the skill score values increase from the 5- to the 10-day lag. The accuracy of this mode is increased by using pentads instead of daily values. By applying the same method to 5-day means of the ITCZ index (with $M = 40/5$), the correlation between the predicted and observed signal reaches 0.65 at lead times of one pentad and 0.50 for the two-pentad prediction. The correlation values are lower for the predictions of the second and third modes with their respective correlation coefficients of 0.44 and 0.40 at the 5-day lag and a decrease to 0.37 and 0.24 at the 10-day lag. However, the skill score values show that the skill remains much better than that of forecasts obtained with persistence. The computation of the correlation between observed and predicted intraseasonal modes for each year of the test period reveals a strong interannual variability (Fig. 7). The predictions of the first mode are less variable, although the year-to-year variance in the skill increases from the 5- to the 10-day lag. At the 5-day lag, the first mode shows 5 yr of successful forecasts with a correlation coefficient higher than 0.60 and only 4 yr with a nonsignificant correlation value at the 95% confidence interval. The prediction skill levels of the second and third modes are more variable, with some years characterized by a correlation greater than 0.5 and others with correlations near 0 and even negative. Figure 7 also shows the loss of skill between the 5- and 10-day lags is the highest for the first mode and the lowest for the second mode.

The examination of the observed and predicted time series reveals that the amplitude of the intraseasonal signal must be high to be well reproduced by the SSA–MEM prediction scheme. It also reveals that the method better predicts regular intraseasonal oscillations in time and amplitude, as illustrated by Fig. 8 for 1994. The observed time series of the third mode is shown in the top panel of Fig. 8 and the wavelet modulus is shown in the bottom panel. The wavelet analysis points out two sequences characterized by strong and regular oscillations of a 12-day period from May to mid-June and from mid-September to the end of October. The 5-day

TABLE 6. As in Tables 4 and 5 but computed operationally. Italics indicate that the correlation value is significant at the 95% confidence interval.

	SSA–MEM 5 day/10 day	ECMWF 5 day/10 day
10–90 days ($RC_{1+\ldots+6}$)	*0.09*/0.06	0.13/0.07
Mode 1 (RC_{1+2})	*0.54/0.36*	0.15/0.35
Mode 2 (RC_{3+4})	*0.44/0.37*	0.51/0.53
Mode 3 (RC_{5+6})	*0.40/0.24*	0.53/0.14

Unauthenticated | Downloaded 10/22/21 12:31 PM UTC

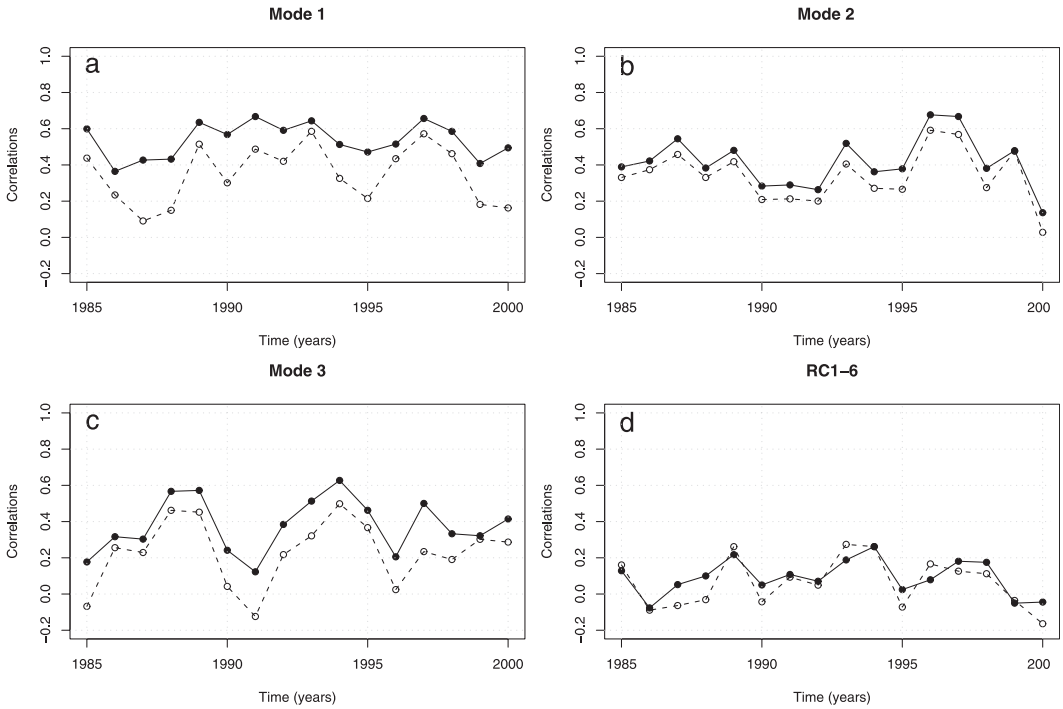

FIG. 7. Interannual variability of the June–September skill of the 5-day (full line) and 10-day (dashed line) forecasts of the intraseasonal modes (a) 1 (RC_{1+2}), (b) 3 (RC_{3+4}), and (c) 3 (RC_{5+6}). (d) The forecast skill of the sum of these three modes ($RC_{1+\ldots+6}$). The skill is determined by the correlation between the predicted SSA modes and the observed SSA modes for each year between April and October.

prediction of this mode (thin line in Fig. 8) is very accurate for these two periods while it fails elsewhere. These are well-known features of the AR forecasts already raised in previous works dealing with long-range forecasting of the break and active summer Indian monsoons (Cadet and Daniel 1988). When the characteristics, that is, the amplitude and period, of the considered intraseasonal mode are well defined, skillful forecasts can be obtained. However, when characteristics change rapidly, the forecasts fail.

5. Conclusions and discussion

In this paper we have investigated the intraseasonal variability of rainfall over West Africa. We used the SSA method, which is an interesting alternative to the Fourier decomposition as it is designed to extract information from noisy time series and can be used to compute an adaptive filter (Ghil et al. 2002). The SSA was first applied to a ground-based rainfall index within the West African ITCZ domain to isolate oscillatory modes in several intraseasonal bands. The results showed the existence of one oscillatory mode of 34 days, one of 20 days, and one of 14 days. This confirms the results of previous studies on the intraseasonal timescale variability of convection over western and central Africa during northern summer (Sultan and Janicot 2003; Mounier and Janicot 2004; Mounier et al. 2008), which used, in particular, spatial EOF analysis and showed from the 10–90-day filtered OLR signal that intraseasonal variability can be split into two periodicity ranges: 10–25 and 25–90 day.

We then used the SSA to intercompare the intraseasonal variability in several widely used datasets. We have shown that although the day-to-day variability of rainfall is not well captured by OLR datasets nor the two reanalysis datasets, the intraseasonal variability is far better reproduced. The intraseasonal features revealed by the SSA are particularly well captured by the rainfall data from ERA-40, while the NCEP–DOE reanalysis fails to accurately reproduce the shorter modes of variability. The discrepancies between the two reanalysis datasets can result from several factors. First, the ERA-40 and the NCEP–DOE have different atmospheric models, with different parameterization schemes and resolutions. Second, the assimilation of satellite data is very different in each reanalysis (see Dell'Aquila et al. 2005). For instance, both ERA-40 and NCEP–DOE

Unauthenticated | Downloaded 10/22/21 12:31 PM UTC

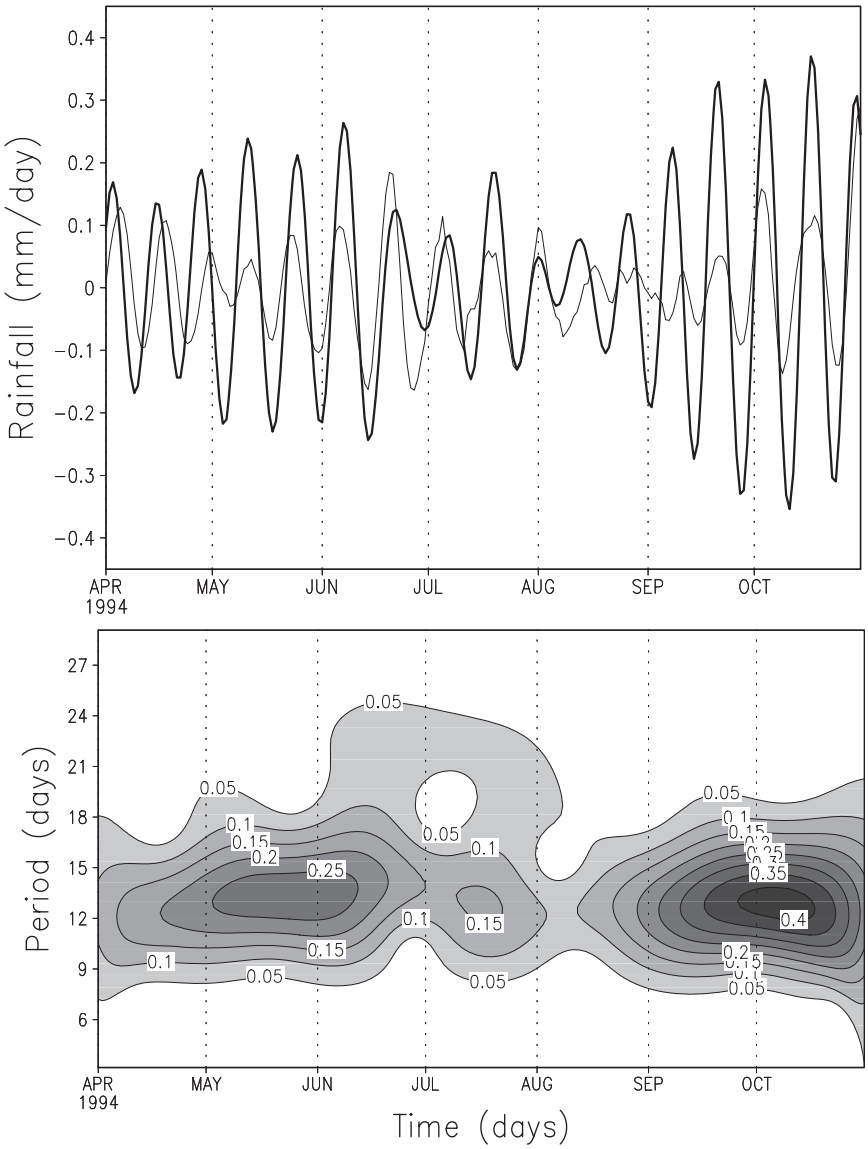

FIG. 8. (top) Time series of the observed (thick line) and predicted (thin line) third intraseasonal modes by using the SSA–MEM method for 1994. (bottom) Corresponding wavelet analysis of the observed mode. Contour lines represent the wavelet modulus depicting the clearest periodic oscillations (in terms of amplitude and regularity) with the highest values.

reanalyses assimilate the TOVS data, but ERA-40 assimilates satellite radiance directly while NCEP–DOE assimilates profiles of the retrieved temperature and humidity. Third, the instrumental data basis has very large spatial and temporal homogeneities.

We then investigated the medium-range predictability of the intraseasonal modes by using both statistical and dynamical forecasts. We have shown that although ERA-40 reanalysis can accurately reproduce the intraseasonal features in rainfall, the dynamical forecasts are far less skillful even at very short time leads. The statistical predictions based on the SSA–MEM are much more promising, though they encounter problems when applied operationally. In an operational application, we found that the forecast skill is very low for the 10–90-day intraseasonal band but the predictability of individual intraseasonal modes is greater. The forecasts skill levels are lower than those from Mo (2001) using the same forecast scheme and the same convection data but for California in winter. These differences may be due to the stronger and more persistent influence of the MJO oscillation in the pan-American convection (Mo 1999),

Unauthenticated | Downloaded 10/22/21 12:31 PM UTC

while the 10–90-day intraseasonal variability in West Africa is much more intermittent and less energetic. We found the year-to-year variability of the forecast skills is influenced by the characteristics of the intraseasonal mode. When the characteristics of the considered intraseasonal mode are well defined, skillful forecasts can be obtained. However, when the characteristics change rapidly, the forecast fails. These conclusions are very close to those of previous works dealing with the long-range forecasting of the break and active summer Indian monsoons (Cadet and Daniel 1988).

Even if the forecast skill is not high, the results of the present paper are important for the study of the West African monsoon. Since it is the first forecast exercise of the intraseasonal fluctuations of convection in West Africa, it can be considered as a skill reference to be improved upon using other prediction systems. The strategy for achieving this improvement is twofold.

First, we can continue to develop empirical forecasts initiated by the present study. The crux of the intraseasonal prediction problem shown in our study using a purely statistical approach with the SSA–MEM is the use of raw data to predict filtered data. As discussed in several studies (Wheeler and Weickmann 2001; Wheeler and Hendon 2004) dealing with MJO monitoring and prediction, the use of a typical bandpass filter to extract its frequency-limited signal is restricted for a real-time task because of its need for information beyond the end of the time series. Alternative approaches must be employed, but these introduce a level of noise that affects the forecast skill. In the present study we used the SSA filtering, which is one of these alternative approaches. The results of this method should be compared with the skills of other methods like the one used by Wheeler and Hendon (2004) to construct a real-time MJO index. This involves the projection of the daily observed data onto the multiple-variable EOFs, with the annual cycle and components of interannual variability removed, to extract principal PC time series that vary primarily only on the intraseasonal time scale of the MJO (Lo and Hendon 2000; Wheeler and Hendon 2004). This projection thus serves as an effective filter for the MJO without the need for conventional time filtering, making the PC time series an effective index for real-time use. Another interesting way to improve the skill of the physically based prediction scheme introduced in this paper is the application of a wavelet banding technique to the predictand and predictors, before performing the linear process, to sort time series into specific spectral bands (Webster and Hoyos 2004). This method, combined with a linear regression model for predicting monsoon rainfall and river discharge on 15–30-day time scales, has shown promising results.

Second, more work must be done to investigate the skill of dynamical forecasts. Several studies (Slingo et al. 1996; Waliser et al. 2003) have recently shown the potential for the numerical predictions of intraseasonal variability in general circulation models in other regions. In addition, we have shown ERA-40 very accurately reproduces intraseasonal variability, although the ECMWF rainfall prediction skill level has been found to be very low and highly variable within 1 yr. A more detailed examination of these ECMWF forecasts to understand, if possible, why they fail or succeed could be one avenue of future exploration. This better understanding will require the examination of other dynamical variables, usually more reliable than rainfall, to identify atmospheric patterns that can be related to rainfall intraseasonal variability. This may show for instance that the predictability of intraseasonal rainfall in the ECMWF model is greater with the occurrence of a specific large-scale atmospheric pattern. One can also imagine building statistical adaptations of the intraseasonal rainfall forecasts by linking statistically observed rainfall and such atmospheric patterns as well as other more reliably predicted dynamical variables.

Even if the intraseasonal forecasts developed in this paper remain far from end users' primary interests, which are mainly the occurrence and length of dry spells at the local scale, we believe our approach can progress toward real applications. For instance, the African Centre of Meteorological Applications for Development (ACMAD), whose mission is the provision of weather and climate information addressed to the end users in the fields of agriculture, water resources, health, public safety, and renewable energy, publishes a decadal climate bulletin with operational analyses and 10-day forecasts based on ECMWF forecasts scheme. The improvement of the forecast skill by using the statistical approach instead of the dynamical approach shown in this paper is thus relevant for African end users with an interest in the publication of this bulletin.

Acknowledgments. We are thankful to F. Mounier for her help in building the operational forecast system during the summer of 2006. We are also thankful to the CIRES Climate Diagnostics Center (Boulder, Colorado) for providing the NCEP–NCAR reanalysis dataset and the interpolated OLR dataset from their Web site (http://www.cdc.noaa.gov/). We also thank the ECMWF for providing the ERA-40 reanalysis dataset and forecasts. We also want to thank the team of the IPSL climate data server (CLIMSERV) and particularly Sophie Clocher, who helped us by providing access to all of these datasets. We are thankful to the Naomi Service that has greatly improved the readability of the paper.

Unauthenticated | Downloaded 10/22/21 12:31 PM UTC

Based on a French initiative, AMMA was built by an international scientific group and is currently funded by a large number of agencies throughout the world, and particularly from France, the United Kingdom, the United States, and Africa. It has been the beneficiary of a major financial contribution from the European Community's Sixth Framework Research Programme. Detailed information on the scientific coordination and funding is available on the AMMA International Web site (http://www.amma-international.org).

REFERENCES

Akaike, H., 1974: A new look at the statistical model identification. *IEEE Trans. Automat. Contr.,* **19,** 716–723.

Baron, C., B. Sultan, M. Balme, B. Sarr, T. Lebel, S. Janicot, and M. Dingkuhn, 2005: From GCM grid cell to agricultural plot: Scale issues affecting modelling of climate impact. *Philos. Trans. Roy. Soc. London,* **360B,** 2095–2108.

Betts, A., J. Ball, H. Barr, T. Black, J. McCaughey, and P. Viterbo, 2006: Assessing land–surface–atmosphere coupling in the ERA-40 reanalysis with boreal forest data. *Agric. For. Meteor.,* **140,** 365–382.

Burg, J., 1968: Maximum entropy spectral analysis. *Modern Spectrum Analysis,* D. G. Childers, Ed., IEEE Press, 34–48.

Cadet, D., and P. Daniel, 1988: Long-range forecast of the break and active summer monsoons. *Tellus,* **40A,** 133–150.

Childers, D.E., Ed., 1978: *Modern Spectrum Analysis*. IEEE Press, 331 pp.

Dell'Aquila, A., V. Lucarini, P. M. Ruti, and S. Calmanti, 2005: Hayashi spectra of the Northern Hemisphere mid-latitude atmospheric variability in the NCEP–NCAR and ECMWF reanalyses. *Climate Dyn.,* **25,** 639–652.

Diedhiou, A., S. Janicot, and H. Laurent, 1999: Easterly wave regimes and associated convection over West Africa and the tropical Atlantic: Results from NCEP/NCAR and ECMWF reanalyses. *Climate Dyn.,* **15,** 795–822.

Duvel, J., 1989: Convection over tropical Africa and the Atlantic Ocean during the northern summer. Part I: Interannual and diurnal variations. *Mon. Wea. Rev.,* **117,** 2782–2799.

Ghil, M., and Coauthors, 2002: Advanced spectral methods for climatic time series. *Rev. Geophys.,* **40,** 1003, doi:10.1029/2000RG000092.

Grueber, A., and A. F. Krueger, 1984: The status of the NOAA outgoing longwave radiation data set. *Bull. Amer. Meteor. Soc.,* **65,** 958–962.

Hendon, B., M. Newman, J. Glick, and J. Schemm, 2000: Medium-range forecast errors associated with active episodes of the Madden–Julian oscillation. *Mon. Wea. Rev.,* **128,** 69–86.

Ingram, K., M. Roncoli, and P. Kirshen, 2002: Opportunities and constraints for farmers of West Africa to use seasonal precipitation forecasts with Burkina Faso as a case study. *Agric. Syst.,* **74,** 331–349.

Janicot, S., and B. Sultan, 2001: Intra-seasonal modulations of convection in the West African monsoon. *Geophys. Res. Lett.,* **28,** 523–526.

——, F. Mounier, N. Hall, S. Leroux, B. Sultan, and G. Kiladis, 2009: Dynamics of the West African monsoon. Part IV: Analysis of 25–90-day variability of convection and the role of Indian monsoon. *J. Climate,* **22,** 1541–1565.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S. Yang, J. Hnilo, M. Fiorino, and G. Potter, 2002: NCEP–DOE AMIP-II reanalysis (R-2). *Bull. Amer. Meteor. Soc.,* **83,** 1631–1643.

Keppenne, C. L., and M. Ghil, 1992: Adaptive filtering and prediction of the Southern Oscillation index. *J. Geophys. Res.,* **97,** 449–454.

——, and ——, 1993: Adaptive filtering and prediction of noisy multivariate signal: An application to sub-annual variability in atmospheric angular momentum. *Int. J. Bifurcat. Chaos,* **3,** 625–634.

Liebmann, B., and C. Smith, 1996: Description of a complete (interpolated) outgoing longwave radiation dataset. *Bull. Amer. Meteor. Soc.,* **77,** 1275–1277.

Lo, F., and H. Hendon, 2000: Empirical extended-range prediction of the Madden–Julian oscillation. *Mon. Wea. Rev.,* **128,** 2528–2543.

Madden, R., and P. Julian, 1972: Description of global scale circulation cells in the tropics with a 40–50-day period. *J. Atmos. Sci.,* **29,** 1109–1123.

Mathon, V., and H. Laurent, 2001: Life cycle of Sahelian mesoscale convective cloud systems. *Quart. J. Roy. Meteor. Soc.,* **127,** 377–406.

Matthews, A., 2004: Intraseasonal variability over tropical Africa during northern summer. *J. Climate,* **17,** 2427–2440.

Mo, K. C., 1999: Alternating wet and dry episodes over California and intraseasonal oscillations. *Mon. Wea. Rev.,* **127,** 2759–2776.

——, 2001: Adaptive filtering and prediction of intraseasonal oscillations. *Mon. Wea. Rev.,* **129,** 802–817.

Mounier, F., and S. Janicot, 2004: Evidence of two independent modes of convection at intraseasonal timescale in the West African summer monsoon. *Geophys. Res. Lett.,* **31,** L16116, doi:10.1029/2004GL020665.

——, ——, and G. Kiladis, 2008: The African monsoon dynamics. Part III: The quasi-biweekly zonal dipole. *J. Climate,* **21,** 1911–1929.

Penland, C., M. Ghil, and K. Weickmann, 1991: Adaptive filtering and maximum entropy spectra with application to changes in atmospheric angular momentum. *J. Geophys. Res.,* **96** (D12), 659–671.

Slingo, J., K. Sperber, and J. Boyle, 1996: Intraseasonal oscillations in 15 atmospheric general circulation models: Results from an AMIP diagnostic subproject. *Climate Dyn.,* **12,** 325–357.

Sow, C., 1997: Diurnal rainfall variations in Senegalin (in French). *Secheresse,* **8,** 157–162.

Sultan, B., S. Janicot, and A. Diedhiou, 2003: The West African monsoon dynamics. Part I: Documentation of intraseasonal variability. *J. Climate,* **16,** 3389–3406.

——, C. Baron, M. Dingkuhn, B. Sarr, and S. Janicot, 2005: Agricultural impacts of large-scale variability of the West African monsoon. *Agric. For. Meteor.,* **128,** 93–110.

Thorncroft, C., and Coauthors, 2003: The JET2000 project: Aircraft observations of the African easterly jet and African easterly waves. *Bull. Amer. Meteor. Soc.,* **84,** 337–351.

Torrence, C., and G. P. Compo, 1998: A practical guide to wavelet analysis. *Bull. Amer. Meteor. Soc.,* **79,** 61–78.

Uppala, S., and Coauthors, 2005: The ERA-40 Re-analysis. *Quart. J. Roy. Meteor. Soc.,* **131,** 2961–3012.

Vautard, R., and M. Ghil, 1989: Singular spectrum analysis in nonlinear dynamics with applications to paleoclimatic time series. *Physica D,* **35,** 392–424.

Unauthenticated | Downloaded 10/22/21 12:31 PM UTC

——, P. You, and M. Ghil, 1992: Singular spectrum analysis: A toolkit for short, noisy chaotic signals. *Physica D,* **58,** 95–126.

von Storch, H., and J. Xu, 1990: Principal oscillation pattern analysis of the 30–60-day oscillation in the tropical troposphere. *Climate Dyn.,* **4,** 175–190.

——, and D. Baumhefner, 1991: Principal oscillation pattern analysis of the 30–60-day oscillation. Part II: The prediction of equatorial velocity potential and its skill. *Climate Dyn.,* **6,** 1–12.

Waliser, D., C. Jones, J. Schemm, and N. Graham, 1999: A statistical extended-range tropical forecast model based on the slow evolution of the Madden–Julian oscillation. *J. Climate,* **12,** 1918–1939.

——, W. Stern, S. Schubert, and K. Lau, 2003: Dynamical predictability of intraseasonal variability associated with the Asian summer monsoon. *Quart. J. Roy. Meteor. Soc.,* **129,** 2897–2925.

Walker, G., 1931: On periodicity in series of related terms. *Philos. Trans. Roy. Soc. London,* **131A,** 518–532.

Webster, P., and C. Hoyos, 2004: Prediction of monsoon rainfall and river discharge on 15–30-day time scales. *Bull. Amer. Meteor. Soc.,* **85,** 1745–1765.

Wheeler, M., and K. Weickmann, 2001: Real-time monitoring and prediction of modes of coherent synoptic to intraseasonal tropical variability. *Mon. Wea. Rev.,* **129,** 2677–2694.

——, and H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.,* **132,** 1917–1932.

Yule, G., 1927: On a method of investigating periodicities in disturbed series. *Philos. Trans. Roy. Soc. London,* **226A,** 267–298.

Unauthenticated | Downloaded 10/22/21 12:31 PM UTC