Potentially predictable components of African summer rainfall in a SST-forced GCM simulation

MICHAEL K. TIPPETT*AND ALESSANDRA GIANNINI

International Research Institute for Climate and Society,

Columbia University, Palisades, New York

October 13, 2005

*M. K. Tippett, International Research Institute for Climate and Society, The Earth Institute of Columbia University, Lamont Campus / 61 Route 9W, Palisades New York 10964, USA. (tippett@iri.columbia.edu)

Abstract

An ensemble of general circulation model (GCM) integrations forced by observed sea surface temperature (SST) represents the climate response to SST forcing as well as internal variability or "noise." Signal-to-noise analysis is used to identify the most reproducible GCM patterns of African summer precipitation related to the SST forcing. Two of these potentially *predictable components* are associated with the precipitation of the Guinea Coast and Sahel regions and correlate well with observations. The GCM predictable component associated with rainfall in the Sahel region reproduces observed temporal variability on both interannual and decadal time-scales, though with reduced amplitude.

1. Introduction

African rainfall during northern hemisphere summer presents variability on interannual and decadal time-scales with the Sahel region experiencing dramatic long-term changes in precipitation and substantial societal impacts in recent decades (Nicholson 1980). Land surface condition changes (Xue and Shukla 1993) and global sea surface temperature variations (SST; Folland et al. 1986) have been proposed as factors that modulate rainfall variability in the Sahel.

Atmospheric general circulation models (GCMs) provide a useful tool to investigate the atmosphere-ocean and land-atmosphere interactions that influence African rainfall and to understand better past climate and future climate scenarios. The weight given to the results of model-based studies depends on the realism of the model used and the fidelity of its representation of physical processes. Results that hold across a variety of models are desirable given the imperfection of the models. An apparent disparity among results from different GCMs in the Sahel region is one of the motivations for this study. Recent modeling studies examining the impact of SST on Sahelian rainfall found that many GCMs forced by observed sea surface temperature (SST) are able to reproduce aspects of the low-frequency variability, but only a single GCM reproduced the interannual component (Giannini et al. 2003; Moron et al. 2003; Bader and Latif 2003; Lu and Delworth 2005). GCMs did reproduce interannual circulation variability more successfully (Moron et al. 2004).

Some studies have investigated the ability of GCMs to reproduce rainfall indices based on spatial averages (Moron et al. 2003). Here, instead, we decompose the GCM output into modes that either maximize variance or potential predictability, and compare their time-series with observations. This approach of separating the model output into modes, some of which may represent observations more skillfully than others, has the potential to treat systematic GCM errors by retaining

skillful modes and discarding unskillful ones. Empirical orthogonal function (EOF) analysis finds the modes that maximize variance. Signal-to-noise (S/N) analysis finds the modes that maximize perfect model potential predictability.

A single GCM simulation forced by prescribed SST contains both the GCM response to the SST boundary condition as well as chaotic internal variability unrelated to the boundary conditions. Averaging an ensemble of GCM integrations forced with identical SST conditions reduces the magnitude of those components associated with internal variability and isolates the SST-forced GCM response, and previous analysis of African summer rainfall has focused on properties of the ensemble mean. Analysis of the ensemble mean is sufficient to determine the modes of SST-forced variability with maximum variance. However, if the goal is to assess the predictability of the single observed realization which necessarily contains contributions from SST forcing and internal variability, the analysis needs to account for internal variability. Signal-to-noise analysis uses the ensemble mean as well as the deviations of the ensemble members from the ensemble mean to identify the modes that are most reproducible or predictable given the SST forcing, that is the *predictable components* of the SST-forced GCM simulation (Hasselmann 1979, 1997; Venzke et al. 1999; Schneider and Griffies 1999; Barreiro et al. 2002; DelSole 2004). The goal of this work is to identify the predictable components of African summer precipitation in a GCM forced with observed SST, and to characterize their relation with observed precipitation on decadal and interannual time-scales.

In this study, we focus specifically on the role of SST forcing on the July - September seasonal rainfall totals in the region 2°N to 20°N and 20°W to 35°E. West African rainfall variability in northern hemisphere summer can be separated into two modes, one associated with the Gulf

of Guinea region and the other with the Sahel region. Rainfall in the Gulf of Guinea region is associated with nearby Atlantic SST, and Sahel rainfall is associated with Indo-Pacific SST. The mechanisms affecting the two regions appear distinct (Giannini et al. 2003; Gu and Adler 2004; Giannini et al. 2005). In addition to identifying the predictable components of the GCM, our goals include quantifying the relation of the predictable components with observed rainfall variability and SST. However, patterns found by predictable component analysis, like EOF analysis, are designed to satisfy statistical criteria and need not isolate a single dynamical mechanism.

Understanding the influence of SST on African rainfall on seasonal time-scales also has implications for seasonal forecasting where SST forcing is the primary source of predictability (Goddard et al. 2001). Although the present work is limited to GCM simulations forced by observed SST, the results provide an estimate of potential predictability and identify the GCM components that maximize potential predictability. Imperfectly known SST is another factor impacting the skill of seasonal forecasts (Goddard and Mason 2002).

The paper is organized as follows: section 2 lists the data and methods used, focusing on the S/N analysis; section 3 presents the predictable components of the GCM simulation; section 4 has a summary and conclusions.

2. Data and Methods

The analysis uses July-September seasonal totals from the 55-year period 1950 to 2004. GCM-simulated precipitation is taken from a 24 member ensemble of ECHAM 4.5 GCM integrations with T42 resolution and forced with observed SST (Roeckner et al. 1996). We define the Africa boreal summer monsoon region as 2°N to 20°N and 20°W to 35°E, containing 176 GCM grid

points. Rainfall observations are taken from station records of the NOAA Global Historical Climate Network (GHCN; Vose et al. 1992) in the box from 0° to 20°N and 20°W to 40°E. There are 41 stations with complete data records for the period 1950-2004 (locations shown with results in Fig. 1d). Additional analysis uses the Hulme precipitation data set, based on gauge data gridded at 2.5° latitude by 3.75° longitude resolution for the 49-year period 1950-1998 (Hulme 1992). SST anomalies are computed from the ERSST data set version 2 from the period 1950-2004 (Smith and Reynolds 2004).

We use empirical orthogonal functions (EOFs) to identify modes of variability in the observations and in the ensemble mean. EOF analysis is a widely used method of identifying the patterns that explain the maximum variance (Kutzbach 1970). EOFs are ordered by the variance they explain and have orthogonal spatial patterns and uncorrelated time-series. EOF analysis of the observed station rainfall anomalies identifies the anomaly patterns that appear, in an average sense, most often in the historical record.

Signal-to-noise (S/N) maximizing EOFs identify the most reproducible patterns or predictable components in systems with internal variability (Hasselmann 1979, 1997; Venzke et al. 1999; Barreiro et al. 2002). Here we focus on the interpretation of S/N EOF analysis as providing a "perfect model" characterization of predictability based only on the behavior of the GCM; observations are not used in the S/N EOF analysis. The perfect model assumption, used in many studies of predictability, is that the observed climate can be represented as a particular realization of the GCM simulation (Kumar and Hoerling 1995; Rowell 1998; Sardeshmukh et al. 2000). This is a strong assumption on the model's ability to represent both the response to SST forcing and internal variability unrelated to SST forcing. Perfect model skill is evaluated from an ensemble of simulations

by computing the expected skill of the ensemble mean to predict the behavior of a single ensemble member. For a sufficiently large ensemble, the ensemble mean contains only the predictable SST-forced response, while a single ensemble member contains both the SST-forced response and unpredictable internal variability. When anomaly correlation is the skill metric, the ratio of the SST-forced ensemble-mean response to internal variability determines the perfect model correlation r_{perfect} (Kleeman and Moore 1999; Sardeshmukh et al. 2000),

$$r_{\text{perfect}} = \frac{S/N}{\sqrt{1 + S^2/N^2}}, \tag{1}$$

where S^2/N^2 is the ratio of the ensemble mean variance to the variance of the deviation of the ensemble members about the ensemble mean. The perfect model skill can be computed using grid-point values or time-series associated with spatial patterns. Often, the perfect model skill surpasses the skill of the ensemble mean when real observations are used, suggesting poor representation of the SST forced signal, an underestimate of internal variability or observational error.

S/N EOF analysis identifies the patterns whose time-series maximize the S/N ratio, and hence, by (1) maximize perfect model correlation; other measures of predictability including mutual information, relative entropy and predictive information are maximized as well when the data is Gaussian (Schneider and Griffies 1999; DelSole 2004). We do not expect the seasonal precipitation data to be strictly Gaussian, and so we must restrict our interpretation of the S/N EOFs to maximizing the S/N ratio and perfect model correlation. Sometimes S/N analysis has been employed in preference to EOF analysis when the ensemble size is relatively small, and the ensemble mean still contains internal variability (Venzke et al. 1999; Barreiro et al. 2002). However, filtering of internal variability from the ensemble mean is not our primary motivation for using S/N analysis. Rather, our goal is to identify the components of the ensemble mean simulation that are most

highly correlated with observations under the perfect model assumption; these components are the most reproducible, in a least-squares sense, given the SST. The value of S/N analysis lies in the accounting for internal variability in observations as approximated by a single ensemble member as well as in the reduction of internal variability in the ensemble mean. The S/N analysis is the perfect model version of model output statistics (MOS) based on canonical correlation analysis which finds the patterns in model output and observations whose time-series are most highly correlated.

S/N EOFs differ from usual EOFs because they maximize different criteria and hence characterize different properties of the SST-forced GCM simulation. Since S/N analysis is less commonly used than EOF analysis, we illustrate some of the qualitative differences and similarities between EOF and S/N analysis with a simple system consisting of two scalar components $[x, y]$; x and y measure anomalies and have zero mean. We assume that the observations $[x_o, y_o]$ (a single ensemble member under the perfect model assumption) contain a forced signal $[x_s, y_s]$ (estimated by the ensemble mean) and a random, mean-zero, unpredictable noise component $[x_n, y_n]$ uncorrelated with the signal, that is

$$\begin{bmatrix} x_o \\ y_o \end{bmatrix} = \begin{bmatrix} x_s \\ y_s \end{bmatrix} + \begin{bmatrix} x_n \\ y_n \end{bmatrix} ; \tag{2}$$

the ensemble size is sufficiently large that there is no dependence on ensemble size. Suppose that the two signal components x_s and y_s are uncorrelated in time, as is the case when x_s and y_s are EOF time-series. Then, the 2×2 covariance matrix used to compute the EOFs of the two-component system is diagonal, and the EOFs are simply the x and y components where the ordering of the EOFs depends on relative sizes of the variances $\langle x_s^2 \rangle$ and $\langle y_s^2 \rangle$; the notation $\langle \cdot \rangle$ denotes expectation and is computed by time averaging.

When the two noise components x_n and y_n are also uncorrelated, the two components of the

system are uncoupled, the noise covariance matrix is diagonal, and the EOFs and S/N EOFs only differ in their ordering. The ordering of the EOFs can be different from that of the S/N EOFs, which depends on the relative sizes of the signal-to-noise ratios $\langle x_s^2\rangle/\langle x_n^2\rangle$ and $\langle y_s^2\rangle/\langle y_n^2\rangle$ rather than the size of the variances; this illustrates that the S/N analysis is invariant with respect to rescaling. The signal-to-noise ratios determine the correlation of the observations with the ensemble mean, since, for example, for x,

$$\text{correlation}(x_s, x_o) \equiv \frac{\langle x_o x_s\rangle}{\sqrt{\langle x_s^2\rangle\langle x_o^2\rangle}} = \frac{\langle x_s^2\rangle}{\sqrt{\langle x_s^2\rangle\langle x_s^2 + x_n^2\rangle}} = \frac{\sqrt{\langle\langle x_s^2\rangle/\langle x_n^2\rangle}}{\sqrt{1 + \langle x_s^2\rangle/\langle x_n^2\rangle}}\,. \tag{3}$$

On the other hand, when the noise components x_n and y_n are correlated, the EOFs and S/N EOFs are generally different. The S/N EOFs are the linear combinations of x and y that maximize the S/N ratio or equivalently the correlation between observations and the signal. A simple case to analyze is when the two noise components are identical and $x_n = y_n$. The EOF analysis and the predictability of the EOFs is unchanged. However, S/N analysis identifies the component $x - y$ which is perfectly predictable, that is, the observed component $x_o - y_o$ is perfectly correlated with the signal $x_s - y_s$. The presence of a perfectly predictable component is due to the noise covariance matrix being singular and having an eigenvalue equal to zero. Another simple case is when x_s and y_s have equal variance, and the S/N EOFs are determined completely by the structure of the internal variability. In this case, the S/N EOFs are the eigenvectors of the internal variability covariance matrix ordered from smallest noise variance to largest, and the leading S/N EOF is parallel to the eigenvector associated with the smallest eigenvalue of the internal variability covariance matrix; that is, it is the direction in phase-space with minimal projection onto the internal variability.

Generally, EOFs and S/N EOFs are not simply related. However, one would expect that leading EOFs that project weakly onto internal variability would have similar structure to S/N EOFs since

they already maximize SST-forced variance (signal) and their projection onto internal variability (noise) is small. Conversely, EOFs that project strongly onto internal variability are less likely to have similar structure to S/N EOFs. A notable difference between EOF and S/N EOF analysis is that a linear change of variable, such as normalizing each grid point value by its standard deviation, changes EOF patterns and time-series but does not change the S/N values or time-series since there is a cancellation in the computation of the ratio of forced to internal variability. In fact, S/N EOF analysis is identical to EOF analysis when the data is transformed by a linear "pre-whitening operator" which makes the internal variability spatially uncorrelated (Venzke et al. 1999; Schneider and Griffies 1999).

Here the GCM internal variability or "noise" is estimated from the deviation of the 24 ensemble members from the ensemble mean. The internal variability covariance matrix is nonsingular since the sample size (1320 = 24 member $\times$ 55 years) exceeds the number (176) of degrees of freedom. Projecting the internal variability covariance matrix onto a limited number of eigenmodes does not significantly impact the results.

In the following, low- and high-frequency components of time series are defined as the 21-year running-average and the deviation from that average respectively; fewer years are used at the ends of the time-series. Correlations greater than the 99% level Student t-test values (0.31 and 0.33 for correlations based on 55 and 49 years of data respectively) are considered significant.

3. Results

We compute EOFs of the station rainfall observations and the ensemble mean GCM precipitation, as well as S/N EOFs of the GCM precipitation. The spatial patterns and time series of the first four

S/N EOFs, as well as their correlation with observed SST anomalies and Hulme precipitation, are shown in Figs.1-4, respectively; Figs. 1 and 3, show the S/N EOFs related to Guinea coast and Sahel rainfall, respectively, and include the corresponding station rainfall EOF and time-series, as well as the spatial patterns of related ensemble mean EOFs.

a. Comparison with station rainfall EOFs

The first two station rainfall EOFs describe respectively Sahel (Fig. 3d) and Guinea Coast (Fig. 1d) rainfall and represent 32.2% and 15.8% of the total station rainfall anomaly variance. The ordering of the station EOFs may reflect the nonuniform spatial distribution of station locations, though GCM EOFs (not shown) have the same ordering when only land points are used in the computation. Correlating the time-series of the stations EOFs with those of the GCM EOFs gives a sense of the correspondence between the GCM and the observed modes of variability, and Table 1 shows the correlations of the two station EOF time series with those of GCM EOFs. The first (Sahel) station rainfall EOF correlates with the second and sixth GCM EOFs which explain 27.7% and 1.6% respectively of the total variance. While the second GCM EOF shows broad positive precipitation anomalies across the Sahel region (Fig. 3a) like the station EOF, the sixth GCM EOF shows positive and negative precipitation anomalies (Fig. 3b). Performing multiple linear regression between the time-series of the second and sixth GCM EOFs and that of the first station rainfall EOF gives a time-series whose correlation with the station EOF is 0.72. The second (Guinea Coast) station rainfall EOF correlates well with both the first and third GCM EOFs which explain 43.1% and 13.4% respectively of the total variance. Using multiple linear regression with the first and third GCM EOF time-series gives a time-series whose correlation with the time-series of the second

(Guinea Coast) station rainfall EOF is 0.73. There are no significant correlations between the Sahel and Guinea Coast station rainfall EOFs and untabulated higher order GCM EOFs.

We now compare the time-series of the stations data EOFs with those of the GCM S/N EOFs to determine the correspondence between the observed modes of variability and the GCM predictable components. The percentage of variance (Table 2, top) explained by the first four S/N EOFs decreases with mode number and has values comparable to those of the ensemble mean EOFs, although the S/N analysis does not use variance as a selection criteria. This result indicates that large-scales feature are most predictable. Table 2 shows the correlations of the first two EOFs of station data with the first four S/N EOFs of the ensemble mean precipitation. The first S/N EOF is associated with Guinea Coast rainfall and its time-series is well-correlated (0.68) with the second EOF of the station data, slightly less than that (0.73) obtained by using multiple regression to combine the first and third GCM EOF time-series. The variability associated with the first S/N EOF is smaller than that of the second EOF of the station data by a factor of 1.43[1]. This factor is consistent with the correlation between the two time-series (0.68) if we assume that the observation time-series variance contains signal and noise and the S/N EOF time-series contains only signal. The second and fourth S/N EOFs do not correlate significantly with station data EOFs; the strongest correlation (-0.28) is between the second S/N EOF and the Guinea Coast station data EOF. The third S/N EOF is associated with Sahel rainfall and the correlation of its time-series with that of the first station EOF is 0.78, compared to 0.61 obtained from the second GCM EOF and 0.72 when the multiple regression combination of the second and sixth GCM EOF time-series is used. The correlation level with the second GCM EOF is consistent with Giannini et al. (2003)

[1]This factor is calculated by comparing the standard deviations of the time-series that multiply the station data EOF and S/N EOF spatial patterns where the spatial patterns are normalized to have unit spatial variance.

who found a correlation of 0.73 between a GCM EOF and a Sahel-region station data EOF using a different period and GCM. The variability associated with of the third S/N EOF is smaller than that of the first station data EOF by a factor of 2.74; this value is smaller by roughly a factor of two than that given by the correlation between the two time-series. Later in section 3d we decompose the Sahel time-series into high and low frequency components and compute their simulation skill.

b. Perfect model skill

The S/N analysis identifies those modes that explain the maximum variance while projecting minimally on internal variability. The ratio of ensemble mean variance to internal variance is 6.97, 3.21, 2.74, and 2.13 respectively for the first four S/N EOFs. The calculation of these values is the most sensitive part of the analysis to the specification of the internal variability covariance matrix; projecting the internal variability covariance matrix onto a limited number of leading EOFs leads to lower S/N ratio values. The S/N ratio is related to the perfect model correlation by Eq. 1, and Table 3 shows the perfect model correlation of the EOF and S/N EOF time-series which appears to be an over-estimate of real skill. The perfect model correlation of the first four EOFs is mostly less than that of the S/N EOFs, and the EOFs are not ordered by their perfect model skill. The perfect model correlation skill (0.6) of the second EOF associated with Sahel precipitation is less than that of the first and third EOFs, indicating that this EOF pattern projects more onto the internal variability than do the other leading EOFs. Another indication of the GCM uncertainty associated with the Sahel EOF is the high degree of similarity between the spatial pattern of the Sahel EOF and the first EOF of the internal variability (not shown); the pattern correlation (uncentered) is 0.93. The projection of the Guinea coast GCM EOF onto the leading mode of internal variability

is considerably less; the pattern correlation is 0.22. Giannini et al. (2005) also noted the relatively high level of internal variability in the Sahel EOF in another GCM. The relation between the Sahel EOF and the dominant internal variability suggests that the Sahel mode has potential to benefit from the S/N analysis.

c. Spatial structures

We now examine the predictable components and their correspondence with observations in more detail. The spatial pattern of the first (Guinea Coast) S/N EOF (Fig. 1c) resembles the station pattern (Fig. 1d) with a maximum south of 10N and modest negative precipitation anomalies to the north, consistent with the correlation of its time-series with Hulme precipitation (Fig. 1f). The S/N pattern has similar structure to the first GCM EOF with the precipitation maximum slightly shifted south and more indication of negative precipitation anomalies to the north. The time-series of the first S/N EOF (Fig. 1e) shows interannual variability, no apparent trends and a strong positive correlation with local SST in the Atlantic (Fig. 1e).

The second S/N EOF spatial pattern shows a localized positive precipitation anomaly off of the coasts of Guinea, Guinea Bissau and Senegal (Fig. 2a). While the model is confident of this response to SST-forcing, the lack of significant correlation with station rainfall EOFs makes it unclear whether this mode in the GCM reflects a physical mechanism in nature or is a systematic error of the GCM or its forcing. The time series (Fig. 2b) and spatial structure is similar (time-series correlation of -0.82) to those of the third GCM EOF which is related to rainfall in the Guinea Coast region. Correlation of the time-series with Hulme estimates of precipitation shows few significant values (Fig. 2c); values in the region near the coast where the S/N pattern is strongly

positive are slightly negative, though insignificant. The second S/N EOF is modestly correlated with local SST anomalies (Fig. 2d), with positive rainfall anomalies being associated with above normal SST. Interestingly, the correlation of the Sahel station data EOF time-series with SST anomalies in the index box covering 7N to 13N and 25W to 15W is -0.55, meaning that below-normal Sahel precipitation is associated with above-normal SST. Two possible mechanisms for this relation are: (i) the pattern of reduced Sahel rainfall and clear skies extends westward over the ocean, leading to increased SST, and (ii) decreased monsoon flow associated with reduced Sahel rainfall decreases local oceanic upwelling and leads to increased SST. Whatever the mechanism, there is the suggestion that the GCM might be erroneously responding to the increased SST with locally enhanced precipitation that is not observed.

The third S/N EOF has broad structure across the Sahel region and resembles the first EOF of the station data and the second GCM EOF (Fig. 3). Correlations with SST are the strongest in Pacific and Indian oceans. However, the presence of decadal variability and serial correlation in the time-series reduces the effective number of degrees of freedom and complicates the interpretation of the significance of these correlations. Correlation of the third S/N EOF time series with the NINO 3.4 index is -0.56; the correlation of the first station data EOF with NINO 3.4 is -0.48. The correlation of the second GCM EOF with NINO 3.4 (-0.3) is not significant at the 99% level used here; the correlation of the multiple regression combination of the second and sixth GCM EOFs with NINO 3.4 is -0.41. Moron et al. (2003) looked at simulations of a Sahel rainfall index in four GCMs and found insignificant correlations with ENSO in three models and a significant correlation of the wrong sign in the fourth. Since the fraction of variance explained by the sixth GCM EOF is small (1.6%), its contribution to the precipitation field and to an area averaged index

is small, and so our finding of correlations between ensemble mean quantities and NINO 3.4 is not inconsistent. Related to our discussion of the second S/N EOF, we note the negative correlation between the Sahel S/N EOF time-series (positive Sahel precipitation anomalies) and SST from the coast of Senegal south to the Ivory Coast (Fig. 3g).

The fourth S/N EOF is not significantly correlated with the station EOFs, and the Hulme data shows little correlation with observed precipitation over land (Fig. 4c). The CAMS-OPI precipitation dataset (Janowiak and Xie 1999), which begins in 1979 and includes satellite observations does indicate some positive correlations (not shown), mostly over ocean in the Gulf of Guinea region. The time-series shows both decadal and interannual variability (Fig. 4b). The correlation with SST anomaly suggests an association with North Tropical Atlantic, tropical Pacific and Indian ocean SST (Fig. 4d).

d. Low- and high-frequency components of the Sahelian predictable component

We now look in more detail at the low- and high-frequency components of the time-series associated with Sahelian rainfall. The low-frequency component of the time-series of the first EOF of the station data shows a clear drying signal from 1950 through about 1990 (Fig. 5a) which is reasonably well captured both by the low-frequency components of the multiple regression combination of the second and sixth GCM EOFs and by the third S/N EOF; note the multiple regression is developed using the unfiltered time-series. Correlation of the low-frequency component of the S/N EOF time-series with SST anomaly (Fig. 5c) shows negative correlations, with the strongest in the Indian ocean, consistent with Giannini et al. (2003) but also encompassing the tropical Atlantic. The interpretation of these correlations is problematic due to the serial correlation of the

time-series; the correlation of the low-frequency component of the S/N EOF time-series with any decreasing linear trend is 0.92. The correlation between the high-frequency components of the second GCM EOF and the first station EOF (Fig. 5b) is not significant (0.25); correlation between the high-frequency components of the multiple linear regression combination of the second and sixth GCM EOFs and the first station EOF (Fig. 5b) is 0.48. Correlation of the high frequency components of the first EOF of the station data and of the third S/N EOF time-series is 0.63. This level of correlation between the high-frequency components is consistent with that (0.52) found by Giannini et al. (2003) using a different GCM and period. This correlation appears robust over the historical record. Moron et al. (2003) found the modest skill of GCMs in simulating high frequency variability of a Sahel index declined in the dry period. Due to the low variance and spatial structure of the sixth EOF (negative anomalies in the eastern part of the domain), we would expect rather different results if we analyzed an index based on a spatial average. The time-series of the high-frequency component correlates with SST anomalies in the ENSO region extending into the Pacific warm pool region. The correlation of the high-frequency component of the S/N EOF with NINO 3.4 is -0.58; the correlation of the high-frequency component of the first station data EOF with NINO 3.4 is -0.57.

4. Summary and Discussion

Ensembles of general circulation model (GCM) integrations forced by observed sea surface temperature (SST) provide a useful tool to investigate the relation between African climate and SST forcing. Any particular ensemble member contains the response to SST forcing as well as internal variability or "noise." The "perfect model" assumption supposes that the statistics of observa-

tions are indistinguishable from those of an ensemble member. In this setting, the *predictable components* are the components that maximize the correlation between the ensemble mean and observations and are the most reproducible patterns of the SST-forced GCM simulation. Here we have used signal-to-noise analysis to identify the predictable components, or S/N EOFs, of African summer precipitation in a GCM forced by observed SST.

Observed precipitation was represented by the first two EOFs of station rainfall data in the region, corresponding to rainfall in the Sahel and Guinea Coast regions, respectively. Two GCM predictable components are associated with the precipitation of the Guinea Coast and Sahel regions and correlate well with the station data EOF time-series. The GCM predictable component associated with Guinea Coast precipitation has the largest S/N ratio, and its time-series correlates strongly with tropical Atlantic SST. The GCM predictable component associated with Sahel rainfall reproduces observed variability on both interannual and decadal time-scales; the high frequency component shows a linear relation with ENSO, similar to that seen in observations.

Applying multiple linear regression to combinations of GCM EOF (variance maximizing) time-series yields comparable reproduction of the station data EOF time-series. While S/N analysis found single modes to represent the Guinea Coast and Sahel rainfall, the EOF analysis found two modes for each region. Overall, GCM EOF modes 1,2,3 and 6 correlated with the Guinea Coast and Sahel station data EOFs. The time series of GCM EOF mode 6 represents interannual time-scale variability of rainfall in the Sahel region. However, its low variance (1.6%) means that it contributes little to the ensemble mean. Also the spatial pattern of GCM EOF 6 has the opposite sign anomaly as the station EOF in the eastern part of the Sahel region. The lack of correlation between GCM EOF modes 4 and 5 with observations, and the low variance of mode 6 suggest

that systematic GCM error is overwhelming some meaningful responses to SST forcing. The S/N analysis also suggests the presence of systematic error. In the second S/N EOF, an Atlantic positive SST anomaly associated with decreased Sahel rainfall led to local positive precipitation anomalies which are not seen in observations, a possible indication that the GCM is overly sensitive to SST forcing. Model output statistics (MOS) methods could be used to calibrate GCM output to observations and reduce systematic error (Feddersen et al. 1999; Landman and Goddard 2002; Mo and Straus 2002; Tippett et al. 2005), using either GCM EOFs or S/N EOFs as predictors; DelSole and Shukla (2005) found predictable components to be better predictor than EOFs of wintertime North America surface temperature.

This analysis has practical implications for seasonal forecasting where most of the potential predictability is related to SST forcing (Goddard et al. 2001). First, the results here suggest that more of the interannual rainfall variability in the Sahel region is related to SST-forcing than had been previously supposed, and that GCMs can simulate that variability. However, the skill of two-tier seasonal forecasts is also limited by the skill of SST forecasts (Goddard and Mason 2002). Although we find that Guinea Coast rainfall is highly predictable given SST, the Guinea Coast predictable component time-series is strongly correlated with SST in the Atlantic where the skill of SST forecasts is relatively low at this time of year. Likewise the forecast skill for interannual rainfall variability in the Sahel region is limited by SST forecast skill in the tropical Pacific where the lead-time of useful SST forecasts of boreal summer SST are limited by the "spring barrier." Another caveat with assessing potential seasonal forecast skill from SST-forced simulations is the issue of ocean-atmosphere coupling. The SST-forced simulation assumes that the ocean drives the atmosphere. Although violating this assumption often degrades the skill of SST-forced simulations

(Fu et al. 2002; Krishna Kumar et al. 2005), there is also the possibility of overestimating the potential to forecast atmosphere driven processes (Bretherton and Battisti 2000).

Second, since the comparison with observations shows that simulation skill is enhanced by focusing on S/N EOFs, there is potential to improve seasonal forecasts by basing them on the behavior of the most predictable components of the GCM rather than the ensemble mean. This potential is illustrated in Fig. 6 which shows the correlation skill of the high-frequency components of the GCM ensemble mean and its projection onto the first four S/N EOFs. Most of the high-frequency simulation skill of the ensemble mean is limited to the Guinea Coast region while the projection shows skill in the Sahel region; projecting the ensemble mean onto variance maximizing EOFs did not improve skill. Additionally, the S/N method should be useful in diagnosing physical mechanisms, since it makes a clearer separation between the SST-forced climate response and internal variability. Finally, as one goal is to obtain robust results across different GCMs, we mention that encouraging results regarding the SST-forced variability of Sahel rainfall are obtained with other GCMs being used routinely at IRI (Barnston et al. 2003; Tippett 2005).

Acknowledgements. We thank Benno Blumenthal for the IRI Data Library and Tony Barnston, Ping Chang, Vincent Moron and two anonymous reviewers for their comments and suggestions. IRI is supported by its sponsors and NOAA Office of Global Programs Grant number NA07GP0213.

REFERENCES

Bader, J., and M. Latif, 2003: The impact of decadal-scale Indian Ocean sea surface temperature anomalies on Sahelian rainfall and the North Atlantic Oscillation. *Geophys. Res. Lett.*, **30**, 2169. doi:10.1029/2003GL018426.

Barnston, A. G., S. J. Mason, L. Goddard, D. G. Dewitt, and S. E. Zebiak, 2003: Multimodel Ensembling in Seasonal Climate Forecasting at IRI. *BAMS*, **84**, 1783–1796.

Barreiro, M., P. Chang, and R. Saravanan, 2002: Variability of the South Atlantic convergence zone simulated by an atmospheric general circulation model. *J. Climate*, **15**, 745–763.

Bretherton, C., and D. S. Battisti, 2000: An interpretation of the results from atmospheric general circulation models forced by the time history of the observed sea surface temperature distribution. *Geophys. Res. Lett.*, **27**, 767–770.

DelSole, T., 2004: Predictability and Information Theory Part I: Measures of Predictability. *J. Atmos. Sci.*, **61**, 2425–2440.

DelSole, T., and J. Shukla, 2005: Specification of wintertime North America surface temperature. *J. Climate*. submitted.

Feddersen, H., A. Navarra, and M. N. Ward, 1999: Reduction of model systematic error by statistical correction for dynamical seasonal predictions. *J. Climate*, **12**, 1974–1989.

Folland, C. K., T. N. Palmer, and D. E. Parker, 1986: Sahel rainfall and worldwide sea temperatures, 1901-85. *Nature*, **320**, 602–607. doi:10.1038/320602a0.

Fu, X., B. Wang, and T. Li, 2002: Impacts of air-sea coupling on the simulation of mean Asian Summer Monsoon in the ECHAM4 model. *Mon. Wea. Rev.*, **130**, 2889–2904.

Giannini, A., R. Saravanan, and P. Chang, 2003: Oceanic forcing of Sahel rainfall on interannual to interdecadal timescales. *Science*, **302**, 1027–1030. doi:10.1126/science.1089357.

——, 2005: Dynamics of the boreal summer African monsson in the NSIPP1 atmospheric model. *Climate Dynamics*. in press, DOI: 10.1007/s00382-005-0056-x.

Goddard, L., and S. Mason, 2002: Sensitivity of seasonal climate forecasts to persisted SST anomalies. *Climate Dynamics*, **19**, 619–632.

Goddard, L., S. J. Mason, S. E. Zebiak, C. F. Ropelewski, R. Basher, and M. A. Cane, 2001: Current approaches to seasonal-to-interannual climate prediction. *Int. J. Climatol.*, **21**, 1111–1152.

Gu, G., and R. F. Adler, 2004: Seasonal evolution and variability associated with the West African monsoon system. *J. Climate*, **17**, 3364–3377.

Hasselmann, K., 1979: *Meteorology of tropical oceans*, chapter On the signal-to-noise problem in atmospheric response studies, pp. 251–259. Royal Meteorological Society, London.

——, 1997: Multi-pattern fingerprint method for detection and attribution of climate change. *Clim. Dyn.*, **13**, 601 – 611.

Hulme, M., 1992: Rainfall changes in Africa: 1931-60 to 1961-90. *Int. J. Climatol.*, **12**, 685–699.

Janowiak, J., and P. Xie, 1999: CAMS-OPI: A global satellite-rain gauge merged product for real-time precipitation monitoring applications. *J. Climate*, **12**, 3335–3342.

Kleeman, R., and A. M. Moore, 1999: A new method for determining the reliability of dynamical ENSO predictions. *Mon. Wea. Rev.*, **127**, 694–705.

Krishna Kumar, K., M. Hoerling, and B. Rajagopalan, 2005: Advancing dynamical prediction of Indian monsoon rainfall. *Geophys. Res. Lett.*, **32**, L08 704. doi:10.1029/2004GL021979.

Kumar, A., and M. P. Hoerling, 1995: Prospects and limitations of seasonal atmospheric GCM predictions. *Bull. Amer. Met. Soc.*, **76**, 335–345.

Kutzbach, J. E., 1970: Large-scale features of monthly mean Northern Hemisphere anomaly maps of sea-level pressure. *Mon. Wea. Rev.*, **98**, 708–716.

Landman, W. A., and L. Goddard, 2002: Statistical recalibration of GCM forecasts over Southern Africa using model output statistics. *J. Climate*, **15**, 2038–2055.

Lu, J., and T. L. Delworth, 2005: Oceanic forcing of the late 20th century Sahel drought. *Geophys. Res. Lett.* submitted.

Mo, R., and D. M. Straus, 2002: Statistical-dynamical seasonal prediction based on principal component regression of GCM ensemble integrations. *Mon. Wea. Rev*, **130**, 2167–2187.

Moron, V., N. Philippon, and B. Fontaine, 2003: Skill of Sahel rainfall variability in four atmospheric GCMs forced by prescribed SST. *Geophys. Res. Lett.*, **30**, 2221. doi:10.1029/2003GL018006.

——, 2004: Simulation of West African monsoon circulation in four atmospheric general circulation models forced by prescribed sea surface temperature. *J. Geophys. Res.*, **109**, D24 105. doi:10.1029/2004JD004760.

Nicholson, S. E., 1980: The nature of rainfall fluctuations in subtropical West Africa. *Mon. Wea. Rev.*, **108**, 473–487.

Roeckner, E., K. Arpe, L. Bengtsson, M. Christoph, M. Claussen, L. Dümenil, M. Esch, M. Giorgetta, U. Schlese, and U. Schulzweida, 1996: The atmospheric general circulation model ECHAM-4: Model description and simulation of present-day climate. Technical Report 218, Max-Planck Institute for Meteorology, Hamburg, Germany. 90 pp.

Rowell, D. P., 1998: Assessing potential seasonal predictability with an ensemble of multidecadal GCM simulations. *J. Climate*, **11**, 109–120.

Sardeshmukh, P. D., G. P. Compo, and C. Penland, 2000: Changes of probability associated with El Niño. *J. Climate*, **13**, 4268–4286.

Schneider, T., and S. Griffies, 1999: A conceptual framework for predictability studies. *J. Climate*, **12**, 3133–3155.

Smith, T. M., and R. W. Reynolds, 2004: Improved extended reconstruction of SST (1854-1997). *J. Climate*, **17**, 2466–2477.

Tippett, M. K., 2005: Filtering of GCM simulated Sahel rainfall. *Geophys. Res. Lett.* Submitted.

Tippett, M. K., L. Goddard, and A. G. Barnston, 2005: Statistical-Dynamical Seasonal Forecasts of Central Southwest Asia winter precipitation. *J. Climate*, **18**, 1831–1843.

Venzke, S., M. R. Allen, R. T. Sutton, and D. P. Rowell, 1999: The atmospheric response over the North Atlantic to decadal changes in sea surface temperature. *J. Climate*, **12**, 2562–2584.

Vose, R. S., R. L. Schmoyer, P. M. Steurer, T. C. Peterson, R. Heim, T. R. Karl, and J. Eischeid, 1992: The Global Historical Climatology Network: Long-term monthly temperature, precipitation, sea level pressure, and station pressure data. Technical Report ORNL/CDIAC-53, NDP-041, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Xue, Y., and J. Shukla, 1993: The influence of land surface properties on Sahel climate. Part I: Desertification. *J. Climate*, **6**, 2232–2245.

List of Figures

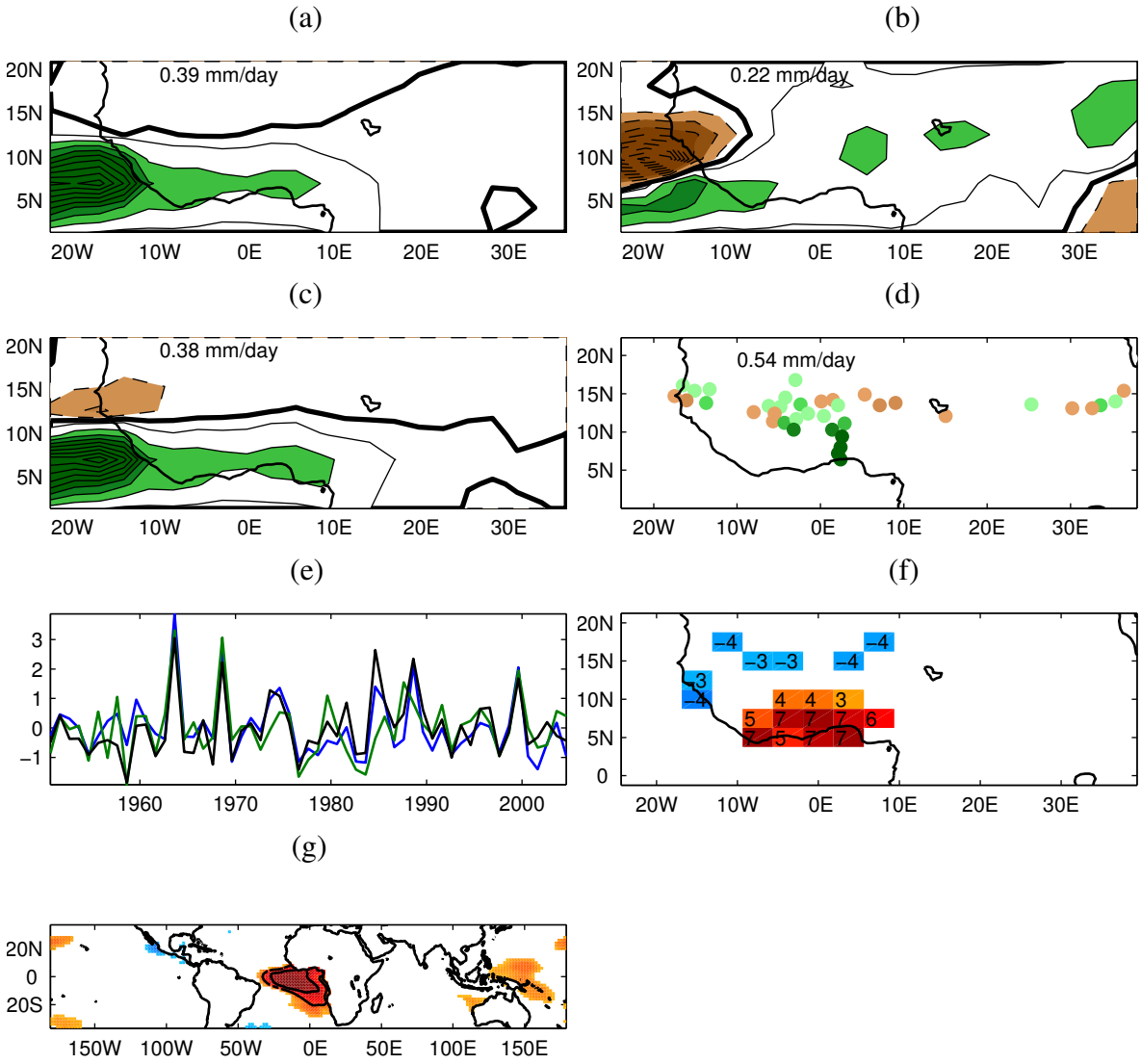

Figure 1. Guinea coast region rainfall variability. Normalized spatial patterns of GCM EOF modes (a) 1 and (b) 3, (c) S/N EOF 1 and (d) Station EOF 2; associated standard deviations are indicated. (e) Normalized time-series of station EOF 2 (black), multiple regression combination of GCM EOFs 1 and 3 (blue) and S/N EOF 1 (green). Correlation of S/N EOF 1 time series with (f) Hulme precipitation anomalies (number in grid box is the correlation multiplied by 10) and (g) SST anomaly (contours at [0.5,0.7]); only significant correlations are shaded. Green (brown) shades represent positive (negative) anomalies, and red (blue) shades represent positive (negative) correlations.

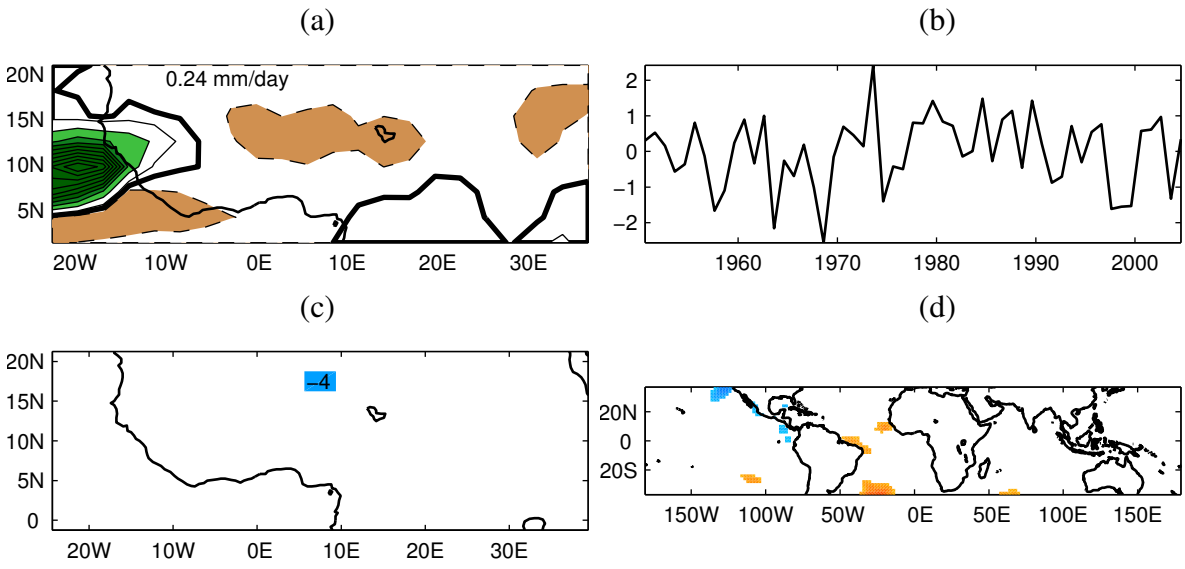

Figure 2. (a) Normalized S/N EOF pattern 2 and its standard deviation. (b) Time-series of S/N EOF 2. (c) Correlation of S/N EOF 2 time series with (c) Hulme anomalies (number in grid box is the correlation level multiplied by 10) and (d) SST anomaly; only significant correlations are shaded.

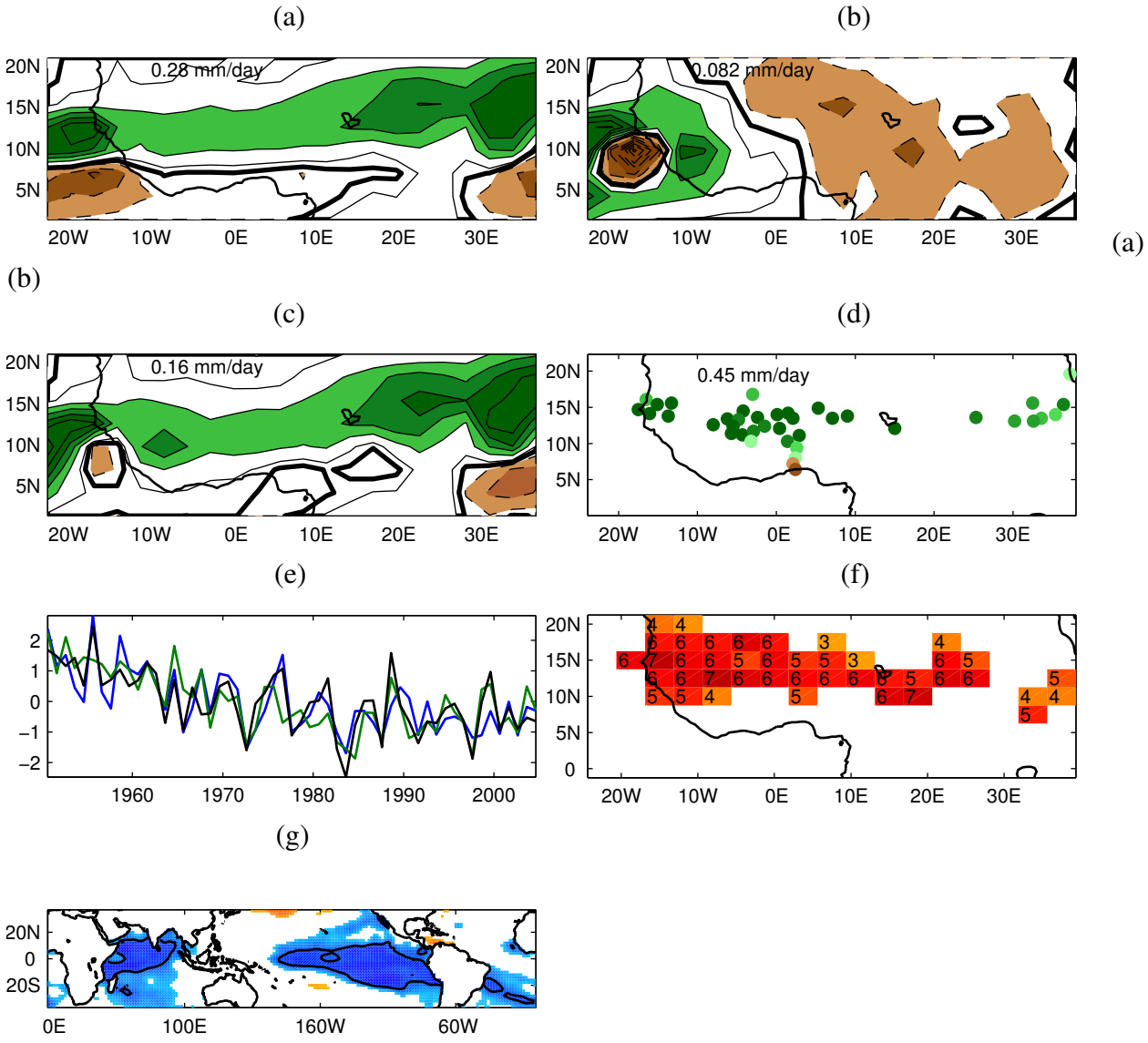

Figure 3. Sahel region rainfall variability. Normalized spatial patterns of GCM EOF modes (a) 2 and (b) 6, (c) S/N EOF 3 and (d) Station EOF 1; associated standard deviations are indicated. (e) Normalized time-series of station EOF 1 (black), multiple regression combination of GCM EOFs 2 and 6 (blue) and S/N EOF 3 (green). Correlation of S/N EOF 3 time series with (f) Hulme precipitation anomalies (number in grid box is the correlation multiplied by 10) and (g) SST anomaly (contour at [-0.5]); only significant correlations are shaded. Green (brown) shades represent positive (negative) anomalies, and red (blue) shades represent positive (negative) correlations.

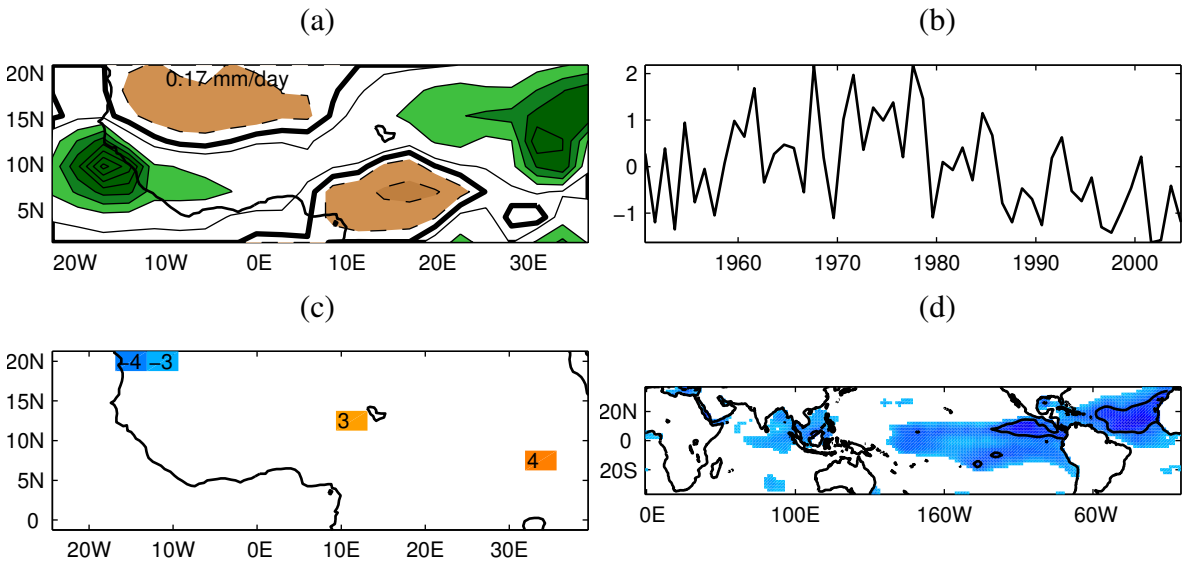

Figure 4. As in Fig. 2 but for S/N EOF 4 and SST correlation contour at -0.5.

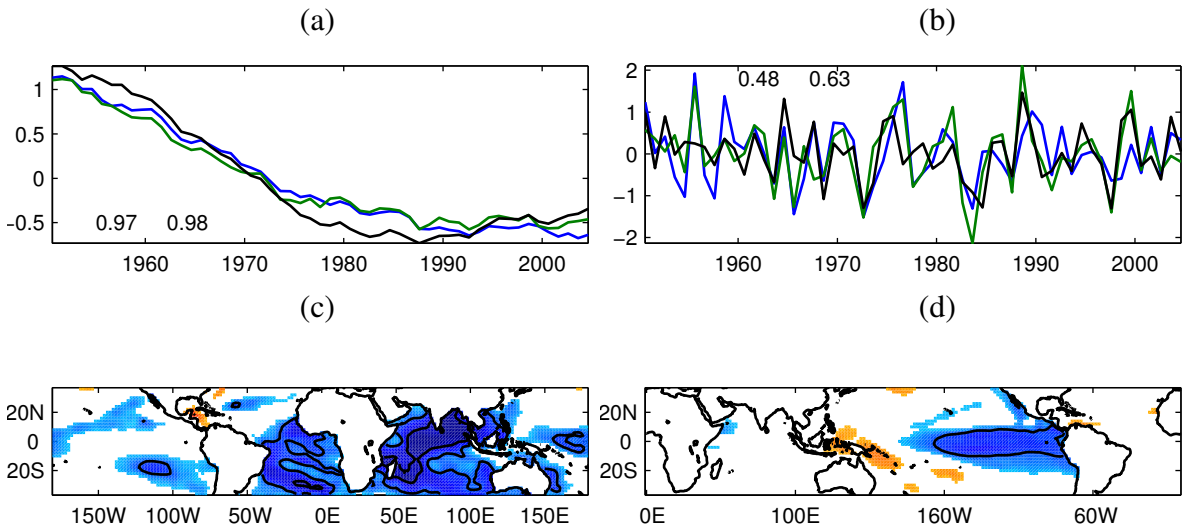

Figure 5. The (a) low-frequency and (b) high-frequency components of the first (Sahel) EOF of station data (black), the multiple linear regression combination of the second and sixth GCM EOFs (blue) and the second S/N EOF (green). The correlation of the high-frequency components of the multiple linear regression combination of the second and sixth GCM EOFs and of the third S/N EOF with the first station data EOF is 0.48 and 0.63 respectively. The correlation of the (c) low-frequency and (d) high-frequency components of the second S/N EOF with SST (contours at [-0.7,-0.5]).

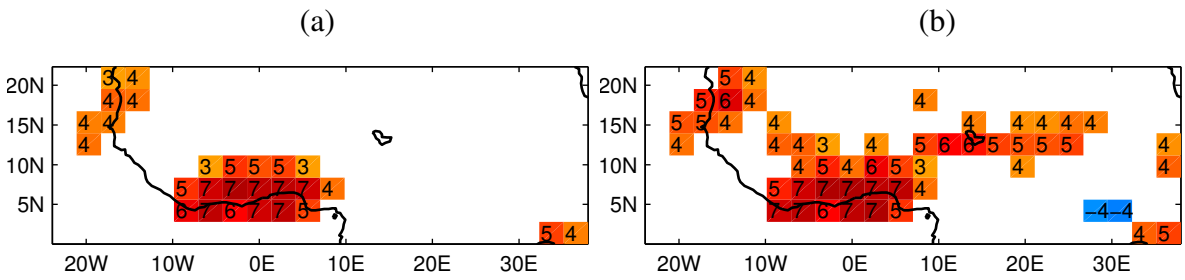

Figure 6. Correlation between the high frequency components of the Hulme dataset precipitation and (a) the GCM ensemble mean and (b) its projection onto the first four S/N EOFs. The number in the grid box is the correlation level multiplied by 10.

List of Tables

	GCM1 (43.1)	GCM2 (27.7)	GCM3 (13.4)	GCM4 (5.8)	GCM5 (2.4)	GCM6 (1.6)
SD 1 (32.2)	-0.10	0.61	0.29	0.02	0.0	0.39
SD 2 (15.8)	0.52	-0.17	0.51	-0.2	0.0	0.06

TABLE 1. Correlation of station data (SD) EOF time series and GCM EOF time series. Numbers in parenthesis are the percentage of total variance.

	S/N 1 (38.1)	S/N 2 (13.9)	S/N 3 (10.9)	S/N 4 (8.9)
SD 1	-0.22	-0.22	0.78	0.17
SD 2	0.68	-0.28	0.06	-0.15

TABLE 2. Correlation of station data (SD) EOF time series and S/N EOF time series. Numbers in parenthesis are the percent of total variance explained by the S/N EOF.

mode	1	2	3	4	5	6
EOF	0.85	0.6	0.71	0.56	0.58	0.52
S/N EOF	0.94	0.87	0.86	0.82	0.58	0.52

TABLE 3. Perfect model correlation skill of EOF and S/N EOF time-series.