# Filtering of GCM simulations of Sahel precipitation

Michael K. Tippett

International Research Institute for Climate and Society, Columbia University,

Palisades, New York

---

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

Atmospheric general circulation models (GCMs) forced with observed sea surface temperature (SST) reproduce aspects of observed Sahel rainfall variability, particularly decadal variability. Here a filter based on signal-to-noise (S/N) EOFs is applied to seven GCM simulations of Sahel precipitation to improve representation of interannual variability. Using filter coefficients based on GCM estimates of internal variability has limited, though positive, impact on simulation skill. Additional removal of empirically identified model error improves the representation of both decadal and interannual variability. The spurious GCM response shows coherence across the seven GCMs and correlates with local Atlantic SST. We hypothesize that the model error is related to errors in the represention of ocean-atmosphere interactions in the SST-forced GCM simulations.

## 1. Introduction

Sahelian rainfall shows considerable variability on both interannual and decadal time-scales, including several decades of dry conditions beginning in the late 1960s and continuing into the 1990s. Sea surface temperature (SST) patterns are one of the modulators of Sahelian rainfall, and statistical methods have identified SST patterns related to Sahelian rainfall variability [*Folland et al.*, 1986]. General circulation models (GCMs) provide a dynamical estimate of the atmospheric response to SST forcing [*Rowell et al.*, 1995]. Many GCMs forced by observed SST reproduce aspects of the decadal variability of rainfall in the Sahel, but most have difficulty reproducing the interannual variability [*Giannini et al.*, 2003; *Moron et al.*, 2003].

The deficiencies of GCM simulations could be due to an inherent lack of predictability, in the sense that SST information alone may be insufficient to constrain Sahelian rainfall interannual variability. However, this explanation seems questionable since at least one GCM forced with observed SST does reproduce Sahelian rainfall interannual variability [*Giannini et al.*, 2003, 2005]. Poor representation of physical processes, particularly those related to convection, may be another problem in the GCM simulations. Statistical methods can be used to compensate for such model errors when GCMs successfully simulate part of the climate signal [*Feddersen et al.*, 1999; *Tippett et al.*, 2005], and GCM simulated winds are a useful surrogate for observed Sahelian rainfall [*Moron et al.*, 2004]. Another potential source of error in GCM simulations forced by observed SST is the neglected feedback of the atmosphere to the ocean [*Peña et al.*, 2003] which has been shown to be important in the simulation of the Asian monsoon [*Fu et al.*, 2002; *Wang et al.*, 2004].

Here a filtering approach is used to reduce the effects of atmospheric internal variability and model error in seven GCM simulations of Sahelian precipitation. Using ensemble averages and projecting the GCM output onto signal-to-noise (S/N) EOFs reduces the impact of internal variability. [*Hasselmann*, 1979, 1997; *Venzke et al.*, 1999; *Barreiro et al.*, 2002]. The S/N EOFs maximize the ensemble mean variance relative to the inter-ensemble variance and are the most reproducible modes of the GCM given the SST forcing; S/N analysis of one GCM identified modes whose time-series correlated well with Sahelian rainfall variability on decadal and interannual time-scales [*Tippett and Giannini*, 2005]. A least-squares estimate of the SST-forced signal is constructed by applying a damping factor in conjunction with the projection of the ensemble mean onto S/N EOFs. The damping factor depends on the the ensemble size and the S/N ratio of each mode; larger S/N ratio and ensemble size leads to less damping. The filter mimics the reduction of internal variability obtained by ensemble averaging. However, as shown in Section 3, direct use of this filtering method does not lead to substantial improvement of the interannual skill of the GCM simulations of Sahelian rainfall. Modifying the filter to remove model error improves the representation of Sahelian rainfall variability on both decadal and interannual time-scales. Model error is classified as those simulation components that, in spite of being identified by the S/N analysis as robust responses of the GCM to SST forcing, do not contribute to simulation skill. Some commonality is found in the model error across GCMs.

## 2. Data

All analyses use July-September (JAS) seasonal averages. The seven GCMs analyzed are: ECHAM 4.5, GFDL AM2p12b, NSIPP-1, ECPC, COLA, CCM 3.6, and NCEP/MRF9 [*Roeckner et al.*, 1996; *Anderson and Coauthors*, 2005; *Bacmeister et al.*, 2000; *Kanamitsu and Coau-*

*thors*, 2002; *Kinter et al.*, 1997; *Hack et al.*, 1998; *Livezey et al.*, 1996, respectively]. The GCMs are forced with observed SST. The simulation period, number of ensemble members and horizontal resolution for each GCM are shown in Table 1.

The S/N EOFs are computed using simulated precipitation from the West Africa region 2°N to 20°N and 20°W to 35°E containing 176 GCM grid points for the T42 resolution. The primary rainfall observations used in this analysis come from 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]; additional analysis uses the CPC Merged Analysis of Precipitation (CMAP) [*Xie and Arkin*, 1996]. SST data are taken from the ERSST data set version 2 [*Smith and Reynolds*, 2004].

## 3. Methods

A least-squared error estimate of the SST-forced signal is the starting point for the derivation of the filter. Suppose that the GCM precipitation anomaly field $x$ (an $n$-dimensional column vector) of a particular ensemble member can be decomposed

$$x = x_S + x_N \,, \tag{1}$$

where $x_S$ is the SST-forced signal and $x_N$ represents random internal variability ("noise") unrelated to the SST forcing. The ensemble mean $x_M$ of an $m$-member ensemble has a smaller contribution from internal variability and can be written

$$x_M = x_S + \frac{x_N}{\sqrt{m}} \,. \tag{2}$$

The least-squares estimate $\hat{x}_S$ of the signal based on the ensemble mean $x_M$ is given by linear regression as

$$\hat{x}_S = \left\langle x_S x_M^T \right\rangle \left\langle x_M x_M^T \right\rangle^{-1} x_M \,, \tag{3}$$

where the notation $()^T$ and $\langle\cdot\rangle$ denotes transpose and expectation, respectively. Defining the covariances of the ensemble mean and internal variability $C_M \equiv \langle x_M x_M^T \rangle$ and $C_N \equiv \langle x_N x_N^T \rangle$, respectively, it follows from (3) that $\left\langle x_S x_M^T \right\rangle = C_M - C_N/m$, and that the estimate $\hat{x}_S$ of the signal is

$$\hat{x}_S = \left( I - \frac{1}{m} C_N C_M^{-1} \right) x_M \,, \tag{4}$$

where $I$ is the identity matrix.

S/N analysis is now used to interpret the regression in (4) as a filter, and show that the regression matrix $\left(I - C_N C_M^{-1}/m\right)$ is diagonal in basis of S/N EOFs; this interpretion is not widely known, to the author's knowledge. For each S/N EOF, S/N analysis identifies a vector of coefficients $f$, a pattern $p$ and a S/N ratio $\lambda$. The leading S/N EOF maximizes the S/N ratio

$$\lambda = \frac{f^T C_M f}{f^T C_N f} \tag{5}$$

of the linear combination $(f^T x)$. Subsequent (in order of decreasing S/N ratio) S/N EOFs maximize the S/N ratio under the constraint that their time-series be uncorrelated in time; unlike usual EOFs, S/N EOFs are not orthogonal in space. The coefficient vectors are eigenvectors of the matrix $C_N^{-1} C_M$, whose eigenvalue decomposition is $C_N^{-1} C_M = F \Lambda F^{-1}$ where the columns of $F$ are coefficient vectors and the entries of the diagonal matrix $\Lambda$ are the associated S/N ratios. Associated with the matrix $F$ of coefficients is the matrix $P$ of patterns defined by $P \equiv F^{-T}$. Since $C_N C_M^{-1} = P \Lambda^{-1} F^T$, the optimal estimate in (4) of the SST-forced signal can be expressed

using S/N EOFs as

$$\begin{aligned} \hat{x}_S &= P\left(I - (m\Lambda)^{-1}\right) F^T x_M \\ &= \sum_{i=1}^{n} p_i \left(1 - (m\Lambda_{ii})^{-1}\right) f_i^T x_M \,, \end{aligned} \tag{6}$$

where $f_i$ and $p_i$ are the $i$-th columns of the matrices $F$ and $P$ respectively, and $\Lambda_{ii}$ the diagonal elements of the matrix $\Lambda$. The form of (6) shows the signal estimate is obtained by (i) forming the linear combinations determined by the coefficients $f_i$, (ii) multiplying by a damping factor that depends on the S/N ratio and ensemble size, and (iii) reconstructing the field using the patterns. Small S/N ratio and small ensemble size lead to larger damping. No damping is necessary in the limits of large S/N ratio or large ensemble size, in which case $PF^T = I$ implies $\hat{x}_S = x_M$. Terms can be omitted from the sum or the damping coefficient set to zero when the S/N EOF represents an unphysical or erroneous response of the GCM to SST forcing.

Estimation of the S/N EOFs requires inverting the noise covariance $C_N$ which is invertible for the domain size, ensemble size and number of years used. The inversion of the noise matrix $C_N$ is computed from its eigenvalue decomposition. Since the smallest eigenvalues of the noise covariance are likely underestimated, an approach from ridge regression is used and the leading (in decreasing order) 80% of the eigenvalues are kept at their estimated value and the remaining eigenvalues set to the value of the eigenvalue at the 80% limit. This procedure has the effect of inflating the intra-ensemble spread associated with the smallest eigenvalues of the noise covariance matrix $C_N$.

## 4. Results

The results are summarized using a Sahel rainfall index that is defined as the average of the precipitation in the box 20°W to 17°E and 10°N to 19°N. This index is computed for the

observations and for three sets of GCM data: the raw ensemble mean and two filter experiments. The first filter experiment, denoted F1, uses (6) and keeps the modes in the sum that have a S/N ratio greater than 1.134; this value corresponds to a correlation of 0.75. The second filter experiment, F2, keeps modes that have a S/N ratio greater than 1.134 *and* that improve the correlation (high frequency) with the observed Sahel index. The notation $r_{\text{total}}$ denotes the correlation of the GCM index with the observed index, and $r_{\text{hf}}$ denotes the correlation of the high-frequency components, defined as the deviation from the 21-year running-average.

Correlations between the observations and the GCM indices are shown in Table 2; also included are correlation with the standardized multi-model sum of the GCM time-series. The correlation $r_{\text{total}}$ of the raw GCM output of most of the models is high; $r_{\text{total}} = 0.72$ for the multi-model sum. However, much of that correlation is related to decadal variability, and only one GCM has a significant high-frequency correlation with the observations; $r_{\text{hf}} = 0.31$ for the multi-model sum. Figure 1a shows that the GCM high frequency indices are more similar to each other than to the observations; the correlation of the multi-model sum with the individual models, a perfect model skill measure, is 0.71. The raw indices, for the most part, do not correlate well with NINO 3.4 (see Table 2); the observed correlation between the Sahel index and NINO 3.4 is -0.59.

The F1 filter results show a generally positive impact on the correlation $r_{\text{total}}$ and a more modest positive impact on $r_{\text{hf}}$. In the F2 filter, 3 to 8 modes that improve the correlation $r_{\text{hf}}$ are retained, explaining from 44 to 75 percent of the GCM precipitation variance in the entire West Africa domain (see Table 2). Filter F2 shows increased values of the total correlation ($r_{\text{total}} = 0.86$ for the multi-model sum) as well as increased values of the high-frequency correlation

($r_{\rm hf} = 0.71$ for the multi-model sum; see Fig. 1b). The high-frequency correlation of the multi-model sum of the F2 filter indices with NINO 3.4 is -0.56, similar to the observed value.

Model error is defined for each GCM as the difference between the F1 and F2 filtered simulations and the associated standardized high-frequency Sahel index time-series (see Fig. 2a) is constructed. Although there is considerable variability across models, the correlation of the multi-model sum of the model error indices with that of the individual models is 0.62, and there is consensus in the sign of the model error in some years. Correlation of the SST with the multi-model sum of the model error indices (Fig. 2b) shows small positive (maximum value $\sim 0.4$) correlations off the coast of West Africa. The sign of the correlation means that warm SST near the West African coast is associated with a positive error in the GCM simulated Sahel index. The largest positive value of the mean model error index occurs in JAS 1984, the driest of the drought years, when several of the GCMs overestimated Sahel precipitation (Fig. 2a). The SST of 1984 was anomalously warm in the Gulf of Guinea and off the coast of Senegal, exceeding 27°C (see Figs. 3a,b). CMAP estimates show positive rainfall anomalies in the Guinea coast region (consistent with the warm SST anomalies in the Gulf of Guinea) and negative rainfall anomalies in the Sahel region and off the coast of Senegal (see Fig. 3c). However, the GCMs, while simulating dry conditions in part of the Sahel region, simulated wetter conditions on the coast near the warm SST anomalies; the precipitation anomaly of the ECPC GCM is shown in Fig. 3d. Warm local SST appears to be associated with spurious enhanced GCM simulated precipitation. The negative precipitation anomalies over ocean in the CMAP estimate are based on satellite data and may reflect decreased cloudiness that may be related to the coincident positive SST anomalies. Also reduced Sahel rainfall is associated with weak monsoon flow, driving

less oceanic upwelling in the Guinea Dome region and resulting in warmer local SST [*Fontaine et al.*, 1995; *Signorini et al.*, 1999]. Either mechanism gives SST that is negatively correlated with local preciptation which is difficult to represent in the SST-forced GCM simulations.

## 5. Summary

Many GCM simulations of Sahelian rainfall reproduce aspects of decadal variability observed in the 20th century. Fewer models, however, reproduce the observed interannual variability. Here a signal-to-noise (S/N) filtering technique has been applied to the precipitation simulations of seven GCMs forced by observed SST. There is little improvement when the S/N ratio is used to determine filter coefficients. Significant improvement is seen when modes classified as model error are removed.

Model error, defined as comprising those components that degrade simulation skill in spite of significant S/N ratio, shows some coherence across the seven GCMs. The multi-model model error time-series shows correlation with regional Atlantic SST; the efficacy of using the S/N EOFs as a filter basis may be the result of signal *and* error being related to SST. In 1984 when several GCMs overestimated Sahel rainfall, warm SST conditions were observed near the coast of West Africa. The enhanced SST, possibly due to reduced cloudiness or oceanic upwelling, appears to be associated with spurrious enhanced precipitation the GCMs. We hypothesize that part of the systematic model error observed in GCM simulations of Sahel rainfall is related to error in the representation of ocean-atmosphere interaction in SST-forced simulations.

**Acknowledgments.** I thank Benno Blumenthal for the IRI Data Library, and Suzana Camargo, Tony Barnston, Alessandra Giannini and Vincent Moron for their comments and sug-

gestions. IRI is supported by its sponsors and NOAA Office of Global Programs Grant number NA07GP0213.

## References

Anderson, J. L., and Coauthors (2005), The new GFDL global atmospheric and land model AM2/LM2: Evaluation with prescribed SST simulations, *J. Climate*, in revision.

Bacmeister, J., P. Pegion, S. Schubert, and M. Suarez (2000), Atlas of seasonal means simulated by the NSIPP 1 atmospheric GCM, *Tech. Rep. NASA/TM-2000-104505*, NASA, Vol. 17.

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.

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.

Fontaine, B., S. Janicot, and V. Moron (1995), Rainfall anomaly patterns and wind field signals over West Africa in August (1958-1989), *J. Climate*, *8*, 1503–1510.

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.

Giannini, A., R. Saravanan, and P. Chang (2005), Dynamics of the boreal summer African monsson in the NSIPP1 atmospheric model, *Climate Dynamics*, in press, DOI: 10.1007/s00382-005-0056-x.

Hack, J., J. Kiehl, and J. Hurrell (1998), The hydrologic and thermodynamic characteristics of the NCAR CCM3, *J. Climate*, *11*, 1179–1206.

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

Hasselmann, K. (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.

Kanamitsu, and Coauthors (2002), NCEP Dynamical Seasonal Forecast System 2000, *Bull. Am. Meteor. Soc.*, *83*, 1019–1037.

Kinter, J., D. DeWitt, P. Dirmeyer, M. Fennessy, B. Kirtman, L. Marx, E. Schneider, J. Shukla, and D. Straus (1997), The COLA atmosphere-biosphere general circulation model Volume: 1 Formulation, *Tech. Rep. 51*, COLA, Calverton, MD.

Livezey, R. E., M. Masutani, and M. Ji (1996), SST-forced seasonal simulation and prediction skill for versions of the NCEP/MRF model, *Bull. Am. Meteor. Soc.*, *77*, 507–517.

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.

Moron, V., N. Philippon, and B. Fontaine (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.

Peña, M., E. Kalnay, and M. Cai (2003), Statistics of locally coupled ocean and atmosphere intraseasonal anomalies in Reanalysis and AMIP data, *Nonlinear Process. Geophys.*, *10*, 245–251.

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, *Tech. Rep. 218*, Max-Planck Institute for Meteorology, Hamburg, Germany, 90 pp.

Rowell, D. P., C. K. Folland, K. Maskell, and M. N. Ward (1995), Variability of summer rainfall over Tropical North Africa (1906-92): Observations and modelling, *Quart. J. Roy. Meteor. Soc.*, *121*, 669–704.

Signorini, S. R., R. G. Murtugudde, C. R. McClain, J. R. Christian, J. Picaut, and A. J. Busalacchi (1999), Biological and physical signature in the tropical and subtropical atlantic, *J. Geophys. Res.*, *104*, 18,367–18,382.

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

Tippett, M. K., and A. Giannini (2005), Potentially predictable components of African summer rainfall in a SST-forced GCM simulation, *J. Climate*, 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.

Wang, B., I.-S. Kang, and J.-Y. Lee (2004), Ensemble simulations of Asian-Australian monsoon variability by 11 AGCMs, *J. Climate*, *17*, 803–818.

Xie, P., and P. A. Arkin (1996), Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions, *J. Climate*, *9*, 840–858.

| Model | period | ensemble size | resolution |
|---|---|---|---|
| ECHAM 4.5 | 1950-2004 | 24 | T42 |
| GFDL AM2p12b | 1950-2000 | 10 | $2.5^\circ \times 2^\circ$ |
| NSIPP-1 | 1950-2000 | 9 | $2.5^\circ \times 2.5^\circ$ |
| ECPC | 1950-2001 | 10 | T63 |
| COLA | 1950-2003 | 10 | T63 |
| CCM 3.6 | 1950-2004 | 24 | T42 |
| NCEP MR9 | 1950-1999 | 10 | T42 |

**Table 1.** The simulation period, ensemble size and horizontal resolution of the GCMs used in this study.

| Model | modes (F2) | % variance (F2) | Hulme Sahel index | | | | | | NINO 3.4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $r_{\text{total}}$ | | | $r_{\text{hf}}$ | | | $r_{\text{hf}}$ | | |
| | | | raw | F1 | F2 | raw | F1 | F2 | raw | F1 | F2 |
| ECHAM 4.5 | 1,2,3,4,7,8, | 75 | 0.54 | 0.72 | 0.81 | 0.077 | 0.37 | 0.67 | 0.053 | -0.17 | -0.44 |
| GFDL AM2p12b | 1,2,3,4,5, | 58 | 0.62 | 0.63 | 0.75 | 0.23 | 0.27 | 0.57 | -0.18 | -0.18 | -0.37 |
| NSIPP-1 | 1,2,3,4, | 63 | 0.73 | 0.75 | 0.8 | 0.43 | 0.46 | 0.59 | -0.37 | -0.39 | -0.52 |
| ECPC | 1,2,4,5, | 51 | 0.7 | 0.71 | 0.76 | 0.25 | 0.29 | 0.56 | -0.57 | -0.59 | -0.62 |
| COLA | 1,2,4, | 46 | 0.57 | 0.56 | 0.84 | 0.094 | 0.08 | 0.64 | -0.1 | -0.093 | -0.27 |
| CCM 3.6 | 1,2,4,5,6, | 44 | 0.34 | 0.33 | 0.67 | 0.14 | 0.065 | 0.51 | -0.2 | -0.18 | -0.5 |
| NCEP MR9 | 1,2,4,5,6,7,8,9, | 74 | 0.63 | 0.67 | 0.71 | 0.25 | 0.29 | 0.4 | -0.38 | -0.4 | -0.42 |
| Multi-model sum | | | 0.72 | 0.76 | 0.86 | 0.31 | 0.39 | 0.71 | -0.33 | -0.39 | -0.56 |

**Table 2.** Correlations of the simulation (unfiltered and two types of filters) of each of the seven GCMs with observed precipitation and the ENSO state. See text for detail.

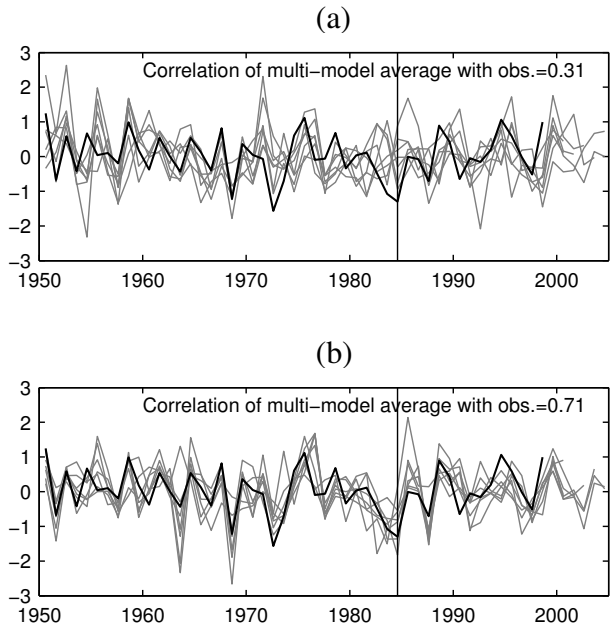


**Figure 1.** Time-series of the high-frequency components of the observed Sahel index (heavy line) and the GCM simulated index computed from the (a) raw ensemble mean and (b) filter F2. The vertical line marks JAS 1984.

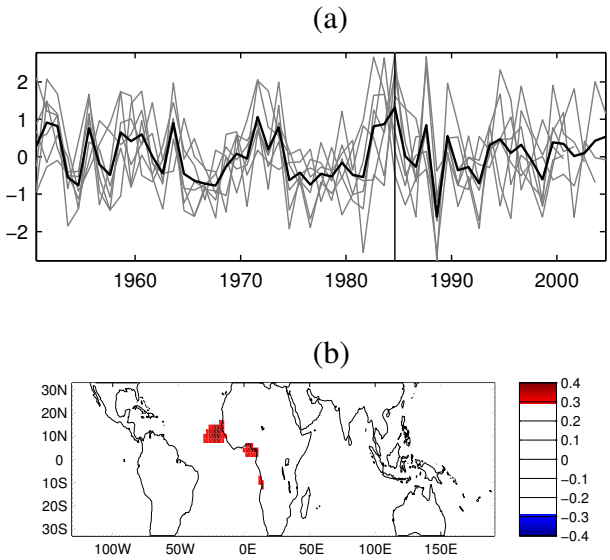


**Figure 2.** (a) The time-series of the high-frequency components of the standardized difference of the F1 and F2 Sahel indices; heavy line is the multi-model mean; the vertical line marks JAS 1984. (b) Correlation of the multi-model mean with SST.

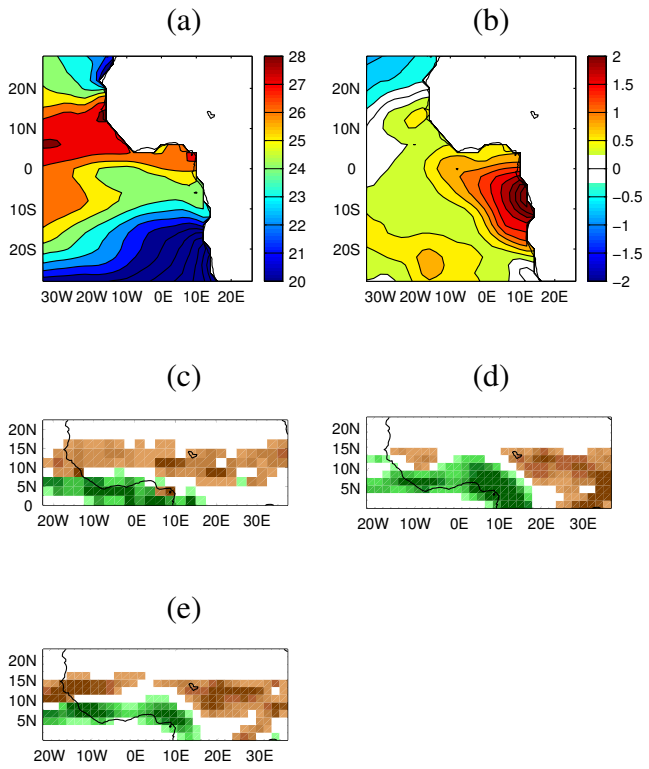


**Figure 3.** JAS 1984 (a) SST (contour interval is 1°C), (b) SST anomaly with respect to 1950-1999 climatology (contour interval is 0.25°C), (c) CMAP, (d) ECPC simulated and (e) F2 filtered precipitation anomalies. Darkest shade of green (brown) indicates an anomaly of 3 (-3) mm/day.