

Simulating SST Teleconnections to Africa: What is the State of the Art?

DAVID P. ROWELL

Met Office Hadley Centre, Exeter, United Kingdom

(Manuscript received 4 October 2012, in final form 17 January 2013)

ABSTRACT

This study provides an overview of the state of the art of modeling SST teleconnections to Africa and begins to investigate the sources of error. Data are obtained from the Coupled Model Intercomparison Project (CMIP) archives, phases 3 and 5 (CMIP3 and CMIP5), using the “20C3M” and “historical” coupled model experiments. A systematic approach is adopted, with the scope narrowed to six large-scale regions of sub-Saharan Africa within which seasonal rainfall anomalies are reasonably coherent, along with six SST modes known to affect these regions. No significant nonstationarity of the strength of these 6×6 teleconnections is found in observations. The capability of models to represent each teleconnection is then assessed (whereby half the teleconnections have observed SST–rainfall correlations that differ significantly from zero). A few of these teleconnections are found to be relatively easy to model, while a few more pose substantial challenges to models and many others exhibit a wide variety of model skill. Furthermore, some models perform consistently better than others, with the best able to at least adequately simulate 80%–85% of the 36 teleconnections. No improvement is found between CMIP3 and CMIP5. Analysis of atmosphere-only simulations suggests that the coupled model teleconnection errors may arise primarily from errors in their SST climatology and variability, although errors in the atmospheric component of teleconnections also play a role. Last, no straightforward relationship is found between the quality of a model’s teleconnection to Africa and its SST or rainfall biases or its resolution. Perhaps not surprisingly, the causes of these errors are complex, and will require considerable further investigation.

1. Introduction

It is well known that variations in sea surface temperatures (SSTs) are partly responsible for large interannual anomalies of seasonal mean rainfall over many areas of Africa. Knowledge of these teleconnections is increasingly used to provide probabilistic predictions of such rainfall anomalies, with the objective of helping vulnerable African communities and assisting national resource management.

These forecasts are founded on one or both of two methodologies. Some employ statistical relationships between SSTs and rainfall derived from several decades of observations. By assuming these relationships are approximately stationary and by exploiting the tendency for SST anomalies to persist, a prediction for the upcoming wet season is produced. On condition that the

empirical model is not overfitted, this approach often provides cheap and reliable predictions. More recently, however, seasonal forecasts have also begun to rely on initialized integrations of general circulation models (GCMs), which are based on a sound physical representation of the real world. These have the potential advantages of being able to respond to more subtle or unusual SST anomaly patterns, of being able to account for changes in teleconnections due to increases in greenhouse gases and aerosols, and of not being susceptible to the sampling errors that occur when evaluating past relationships. In practice, however, GCMs fail to adequately represent all aspects of the real atmosphere. This is due to inevitable limits in their resolution, assumptions in their parameterization of key physical processes, and a partial lack of the knowledge needed to build these parameterizations. Therefore, in the context of providing and improving seasonal predictions of African rainfall anomalies, it is critical that we evaluate and understand the limitations that models have in representing teleconnections from SST anomalies to African rainfall. Furthermore, these teleconnections may also communicate remote effects of anthropogenic change to

 Denotes Open Access content.

Corresponding author address: David P. Rowell, Met Office Hadley Centre, FitzRoy Road, Exeter EX1 3PB, United Kingdom.
E-mail: dave.rowell@metoffice.gov.uk

DOI: 10.1175/JCLI-D-12-00761.1

Africa, and so this too motivates efforts to evaluate and improve their modeling.¹

However, few studies to date have attempted to systematically evaluate the performance of model teleconnections. In the African context, Cook and Vizi (2006) provided a subjective but thorough assessment of the three-way link among Sahelian rainfall, Guinea Coast rainfall, and Gulf of Guinea SSTs in 10 coupled atmosphere–ocean GCMs (AOGCMs). Joly et al. (2007) compute the three dominant covarying patterns of West African rainfall and tropic-wide SSTs in observed data and subjectively compare them with those found in 12 AOGCMs. Others have examined teleconnections from the El Niño–Southern Oscillation (ENSO) to grid-point tropical rainfall, including over Africa; Yang and DelSole (2012) assess the performance of 46 years of seasonal hindcast data and Langenbrunner and Neelin (2013) assess 17 years of atmosphere-only simulation data. Last, some studies have also verified the capability of a single model to represent just one or two teleconnections to Africa, before proceeding to use the model as a tool to enhance understanding of that teleconnection (e.g., Rowell 2001; Vizi and Cook 2002).

In this study, the aim is to systematically assess the capability of a large number of models to simulate teleconnections from SSTs to African rainfall. This will provide a benchmark of the current state of the art. It can also be viewed as a summary of the key challenges that should be addressed on this topic. The foci here are (i) coupled models (AOGCMs), because these are the primary tool used for both for seasonal prediction and climate change projection (noting that uninitialized integrations are employed here because of their large sample of years), and (ii) teleconnections to large-scale regions of Africa, which is partly to reflect the typical horizontal resolution of models considered here and partly to provide a pragmatic limit to the scope of the study. Furthermore, an objective statistical methodology is designed for the assessment, which includes consideration of the role of sampling errors when comparing model and observed teleconnections.

Key questions that are addressed are as follows: 1) Are some teleconnections easier to model than others? 2) Are some models consistently better than others? 3) Can we begin to say anything about the causes of teleconnection errors and their differences between models? These questions are tackled in sections 4 and 5, after describing

the model and observed data (section 2a), the analysis regions and periods (sections 2b and 2c), and the characteristics of the observed teleconnections (section 3).

2. Data and methods

a. Data

The observed rainfall data used in this study derive from the Climatic Research Unit (CRU) Time Series 3.1 (TS3.1) dataset described by Harris et al. (2013) and made available by the British Atmospheric Data Centre. This provides monthly-mean precipitation totals and station counts on a 0.5° grid for the period 1901–2009. Some of the key results of this paper (section 4c) were also verified by instead using the University of Delaware version 2.01 (UDel2.01) dataset described by Johnson et al. (2003). This provides monthly-mean precipitation totals on a 0.5° grid for the period 1900–2008.

The observed SST data derive from the Met Office Hadley Centre Sea Ice and Sea Surface Temperature dataset, version 1.1 (HadISST1.1) reconstruction of monthly-mean SSTs described by Rayner et al. (2003). These are available on a 1° grid from 1870 to the present day.

The modeling focus of this study is on the performance of 44 coupled ocean–atmosphere climate models. Data are sourced from the Coupled Model Intercomparison Project (CMIP), using the “20C3M” experiment from the CMIP phase 3 (CMIP3) database (Meehl et al. 2007a) and the “historical” experiment from the CMIP phase 5 (CMIP5) database (Taylor et al. 2012). These experiments simulate climate variability from the mid to late nineteenth century to the late twentieth or early twenty-first century and are driven by realistic anthropogenic and natural forcings. Where an ensemble of simulations is available, only the first member is utilized, to provide consistent statistics across models. The models used here are listed in Tables 1 and 2 (see Table A1 in the appendix for institution and model name expansions). Data from other available models were not used either because of postprocessing problems or (for some CMIP5 data) because they were not readily available at the time of writing. For all calculations, a 5-yr spinup period was first removed from the model data. Additionally, a linear trend was removed from the remaining data, to negate the impact of drift found in some coupled models.

It will be instructive to also include an analysis of data from atmosphere-only simulations. Such models (AGCMs) are driven by observed global SSTs, allowing us to assess the extent to which erroneous SSTs in the coupled models cause their erroneous teleconnections. Ensembles of simulations from two models are

¹ Note that local mechanisms directly linking changes in atmospheric constituents to regional climate change (e.g., via anomalies in the thermodynamic profile) also play a key role in shaping tropical continental change (Patricola and Cook 2010; Rowell 2012).

TABLE 1. CMIP3 model details. Resolution data (averaged over 30°S–30°N) were obtained from documentation available on the CMIP/Program for Climate Model Diagnosis and Intercomparison (PCMDI) website and converted to degrees.

Institute and country	Model name/version	Average tropical resolution		Simulation period
		Atmosphere	Ocean	
BCCR, Norway	BCCR-BCM2	1.875°	1.25°	1852–1998
CCCma, Canada	CGCM3.1(T47)	3.75°	1.85°	1852–1999
CNRM, France	CNRM-CM3	1.875°	1.625°	1862–1998
CSIRO, Australia	CSIRO Mk3.0	1.875°	1.36°	1873–1999
CSIRO, Australia	CSIRO Mk3.5	1.875°	1.36°	1873–1999
MIUB, Germany, and KMA, South Korea	ECHO-G	3.9°	2.2°	1862–1999
IAP (LASG), China	FGOALS-g1.0	2.8°	1°	1852–1998
NOAA/GFDL, United States	GFDL CM2.0	2.25°	0.83°	1863–1999
NOAA/GFDL, United States	GFDL CM2.1	2.25°	0.83°	1863–1999
NASA GISS, United States	GISS-EH	4.5°	4.5°	1882–1998
NASA GISS, United States	GISS-ER	4.5°	4.5°	1882–2002
NASA GISS, United States	GISS-AOM	3.5°	3.5°	1852–1999
MOHC, United Kingdom	HadCM3	3.125°	1.25°	1871–1998
MOHC, United Kingdom	HadGEM1	1.56°	0.83°	1862–1998
INGV, Italy	INGV-SXG	1.125°	1.25°	1872–1999
INM, Russia	INM-CM3.0	4.5°	2.25°	1873–1999
CCSR/NIES/JAMSTEC, Japan	MIROC3.2(hires)	1.125°	0.23°	1902–99
CCSR/NIES/JAMSTEC, Japan	MIROC3.2(medres)	2.81°	1.175°	1852–1999
MPI, Germany	MPI-ECHAM5-OM	1.875°	1.5°	1862–2005
MRI, Japan	MRI-CGCM2.3.2	2.81°	1.875°	1853–1999
NCAR, United States	PCM	2.81°	0.84°	1892–1998
NCAR, United States	CCSM3	1.4°	0.84°	1872–1998

available, with ensemble members differing only in their initial atmospheric conditions. The first uses the National Atmospheric and Space Administration (NASA) Seasonal-to-Interannual Prediction Project Tier 1 (NSIPP1) model (Schubert et al. 2004) and has 14 members covering the period 1902–2006. The second uses the Hadley Centre Atmosphere Model, version 3 (HadAM3; Pope et al. 2000) and has four members covering 1870–2001. For these simulations, a spinup period of 1 yr was removed, and for consistency with the coupled model analysis a linear trend was also removed.

Last, all data—observed and modeled—were necessarily interpolated to a common grid, which was chosen to be that of HadGEM2 (N96 resolution; i.e., 1.25° latitude by 1.875° longitude).

b. Analysis regions

The focus of this study is on the teleconnections to large-scale regions of sub-Saharan Africa within which interannual rainfall anomalies are reasonably coherent. The regions employed are illustrated in Fig. 1 (shaded in dark gray) and were defined in a quasi-objective manner as follows: First, over the entire continent south of 25°N, a uniform sample of 0.5° grid boxes (every eighth box in each direction) was correlated with all other grid boxes in Africa, using observed seasonal mean rainfall data for the period 1950–99. Then, focusing on areas displaying large-scale coherency, the spatial sampling frequency

was iteratively enhanced and analysis of monthly data was introduced to progressively refine the definitions of all large-scale regions of coherent rainfall anomalies. It was required that within each region almost every grid box at N96 resolution should correlate with the majority of other boxes in that region at $r > 0.6$ and that, where possible, the resulting region should be rectangular (except that marine points are excluded in coastal regions, and this last requirement was found to be too restrictive for the Greater Horn of Africa region).

Monthly-mean regional averages of rainfall were then correlated with SST time series at each grid box, stratified by calendar month, to form global correlation maps for each rainfall region and each month. The month-to-month similarity of these maps, along with the month-to-month similarity of the regional rainfall time series, was then used to define a specific season to be associated with each rainfall region.

Last, two regions were rejected for further analysis because their rainfall teleconnections (i.e., correlations) with SSTs were much weaker and so are less usefully assessed against model data. These were the Greater Horn of Africa for March–April and Madagascar for October. The latter also suffers from limited availability of rain gauge data, reducing confidence in the quality of its long-term observed gridded rainfall estimates.

Hence, in this way, six large-scale regions were defined for sub-Saharan Africa, each with an associated

TABLE 2. CMIP5 model details. Resolution data (averaged over 30°S–30°N) were obtained from the “CMIP5 metadata questionnaire,” except for the IAP models, for which resolution was computed from the model output grid.

Institute and country	Model name/version	Average tropical resolution		Simulation period
		Atmosphere	Ocean	
CSIRO–BOM, Australia	ACCESS1.0	1.56°	0.83°	1852–2004
CSIRO–BOM, Australia	ACCESS1.3	1.56°	1.56°	1852–2004
CCCma, Canada	CanESM2	2.81°	1.17°	1852–2004
CNRM, France	CNRM-CM5	1.40°	0.67°	1892–2004
CSIRO, Australia	CSIRO Mk3.6.0	1.875°	1.41°	1852–2004
IAP (LASG), China	FGOALS-g2	2.93°	0.84°	1902–2008
IAP (LASG), China	FGOALS-s2	2.24°	0.84°	1852–2004
MOHC, United Kingdom	HadGEM2-CC	1.56°	0.83°	1861–2004
MOHC, United Kingdom	HadGEM2-ES	1.56°	0.83°	1861–2004
INM, Russia	INM-CM4	1.75°	0.75°	1852–2004
IPSL, France	IPSL-CM5A-LR	2.82°	1.75°	1852–2004
IPSL, France	IPSL-CM5A-MR	1.875°	1.75°	1852–2004
IPSL, France	IPSL-CM5B-LR	2.82°	1.75°	1852–2004
CCSR/NIES/JAMSTEC, Japan	MIROC5	1.40°	1.1°	1852–2005
CCSR/NIES/JAMSTEC, Japan	MIROC-ESM	2.80°	1.1°	1852–2004
CCSR/NIES/JAMSTEC, Japan	MIROC-ESM-CHEM	2.80°	1.1°	1852–2004
MPI, Germany	MPI-ESM-LR	1.87°	1.5°	1852–2004
MPI, Germany	MPI-ESM-P	1.87°	1.5°	1852–2004
MRI, Japan	MRI-CGCM3	1.12°	0.75°	1852–2004
NCAR, United States	CCSM4	1.10°	0.80°	1852–2004
NCC, Norway	NorESM1-M	2.20°	1.125°	1852–2004
NCC, Norway	NorESM1-ME	2.20°	1.125°	1852–2004

season, within which interannual anomalies of rainfall have relatively high spatiotemporal coherence. These are as follows: the Sahel [12°–18°N, 10°W–35°E; July–September (JAS)], the Guinea Coast (GCoast; 6°–9°N, 7°W–7°E; JAS), the Greater Horn of Africa [GHAfrica; October–December (OND); encompassing much of Kenya and Somalia and also parts of southeastern Ethiopia and northeast Tanzania], central East Africa (CEAfrica; OND; 6°–11°S, 29°–37°E; encompassing parts of southwest and central Tanzania and northeast Zambia), southeast Africa [SEAfrica; November–December (ND); 17°–23°S, 23°–35°E; encompassing much of Zimbabwe and parts of western Mozambique and northeast Botswana], and southwest Africa [SWAfrica; December–February (DJF); 21°–31°S, 16°–23°E; encompassing southeast Namibia, southwest Botswana, and part of the northwest of the republic of South Africa]. Over the remainder of Africa or in other seasons, homogeneity of rainfall anomalies occurs at smaller spatial scales.

We also require indices of the important SST modes that are known to affect rainfall variability over Africa. Six of these are defined and illustrated in Fig. 1 (light gray shading). Four follow standard definitions used by the Ocean Observations Panel for Climate (OOPC) “State of the Ocean” assessment. For the Atlantic, a tropical Atlantic dipole index (TAD) is defined as the

difference between Enfield et al.’s (1999) tropical North and South Atlantic indices (averages of 5°–25°N, 55°–15°W minus 20°S–0°, 30°W–10°E), along with an equatorial east Atlantic (EqEAtl) index, which follows Chang et al.’s (1997) South Atlantic tropical index (average of 5°S–5°N, 15°W–5°E). For the Indian Ocean, we use the standard definition of the Indian Ocean dipole (IOD; averages of 10°S–10°N, 50°–70°E minus 10°S–0°, 90°–110°E; Saji et al. 1999), along with a central Indian Ocean index (CIndO; average of 25°S–10°N, 55°–95°E), defined here because it correlates with the Sahel, SEAfrica, and SWAfrica (see section 3a). In the Pacific, the Niño-3.4 index (average of 5°S–5°N, 170°–120°W) is used to measure ENSO variability. For the Mediterranean (Med), known to affect rainfall variability over the Sahel (see section 3a), we average SSTs over the entire basin. In all cases, seasonal means are produced for each of the four seasons required for the rainfall regions (JAS, OND, ND, and DJF).

c. Analysis periods

One of the critical issues in assessing relationships between climate variables (rainfall and SSTs, in this case) is the choice of time period used for the analysis. Two contradictory requirements affect the magnitude of sampling effects in observed data: first, the analysis period should be as long as possible; second, the quantity

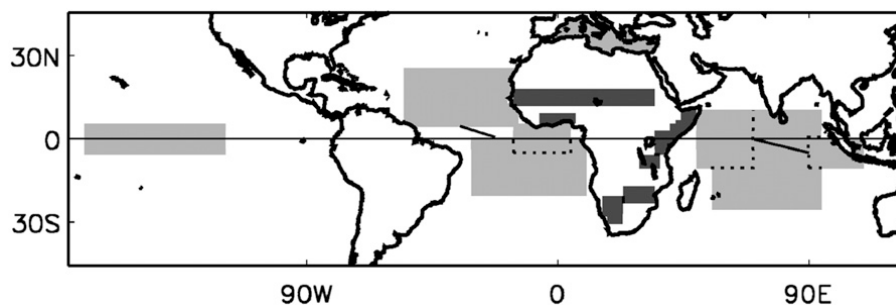


FIG. 1. Definition of the regions used to average rainfall (dark gray), and definition of the SST indices (light gray). For the latter, a dipole index is indicated by a line connecting two boxes.

and quality of the raw data should be as high as possible. Thus, Fig. 2 illustrates the temporal variability of rain gauge density for each of the rainfall regions. Not surprisingly, fewer observations are available early in the twentieth century, but it can also be seen that the maintenance of the gauge network and/or the collection of data substantially declines in recent years. The analysis period for observations is therefore chosen to be 1922–94 (black bars) to provide time series of sufficient length to assess the strength of observed SST–rainfall teleconnections, without including years that have few input data to construct the gridded values in one or more regions.

For model data, we know their quality is stable through each simulation. Here, the requirements are first for a large sample of years and second for intermodel consistency in this sample size to eliminate intermodel variations in sampling error. This latter requirement eliminates intermodel variations in the power of the statistical test that will be used to compare model and observed teleconnections. Thus, all model time series are restricted to 93 years, this being the length of the shortest simulation. In the following sections, the strength of SST teleconnections to African rainfall is measured using the correlation between the rainfall and SST time series. For those models with more than 93 years of data, this correlation will be the average of that computed from the first and last 93-yr periods of model data. This improves the accuracy of estimated model teleconnection strength but assumes stationarity of this teleconnection strength through the full period of model simulation. Sections 4c and 4d will show that this assumption is usually valid and has no impact on our conclusions.

3. Observed teleconnections

a. SST correlation maps

We begin by describing the teleconnections from SSTs to each of the rainfall regions defined above, using observed data for 1922–94. These are shown in Fig. 3 as correlations between regional rainfall and local SST

variability, along with a test against the null hypothesis that the underlying population correlations are zero. Serial correlation is accounted for by using a range of lagged correlations within each time series to adjust the degrees of freedom, following the method of Bartlett (1935) and Folland et al. (1991).

For the Sahel, the well-known relationships are clearly apparent. Drought tends to be associated with warmer than average SSTs in the equatorial and South Atlantic (cf. Lamb 1978; Vizi and Cook 2002), warming of much of the Indian Ocean (cf. Shinoda and Kawamura 1994; Lu and Delworth 2005), ENSO events in the east and central Pacific (cf. Janicot et al. 2001; Rowell 2001), and cooling of the Mediterranean (cf. Rowell 2003; Jung et al. 2006).

Over the Guinea coast, the influence of SSTs is less complex, with droughts being driven only by a cooling

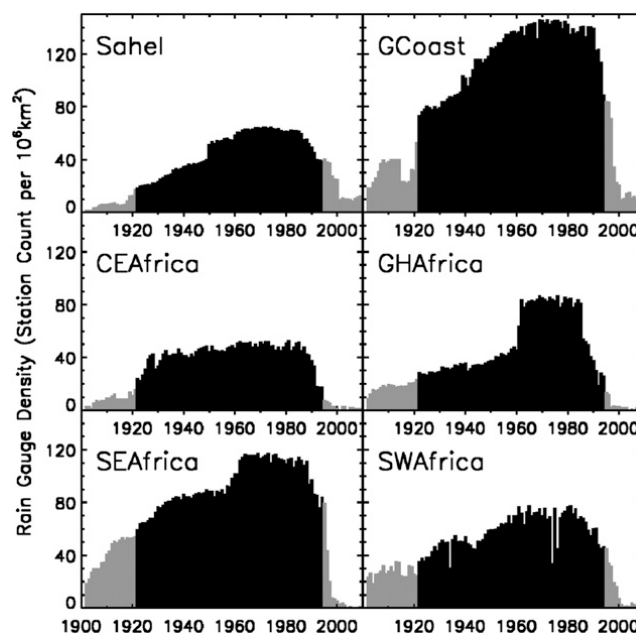


FIG. 2. Time series of the density of rain gauges used by CRU TS3.1, for each of six rainfall regions. Units are station counts per 10^6 km^2 , and black bars denote the chosen analysis period (1922–94).

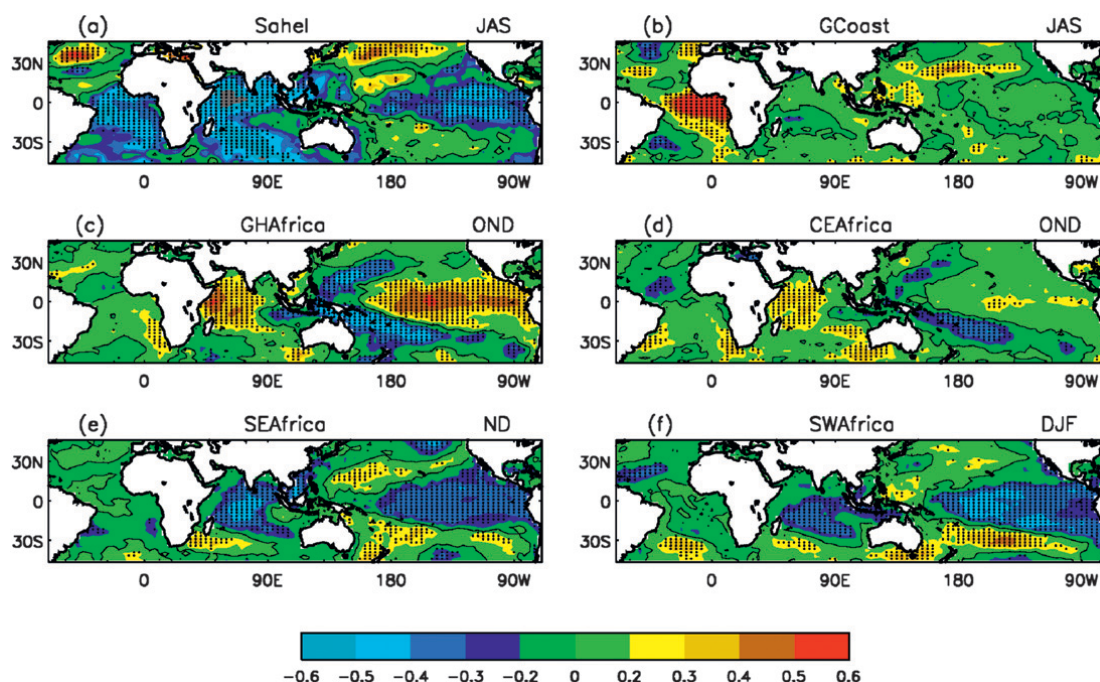


FIG. 3. Correlations of observed rainfall, averaged over each of the African regions, against observed SSTs on an N48 grid, for the period 1922–94. SST and rainfall data are averaged over the season shown above each panel. The local rejection of a null hypothesis of zero correlation, at the 5% significance level, is indicated by a black dot.

of the equatorial Atlantic and near-equatorial South Atlantic (cf. Rowell et al. 1995; Vizy and Cook 2001).

Along the coast of the Greater Horn of Africa, drought tends to be associated with La Niña events and with the negative phase of the IOD: that is, cooling in western Indian Ocean and sometimes a warming in the eastern Indian Ocean (cf. Ogallo et al. 1988; Hastenrath et al. 2004). Farther south, over CEAFrica, similar teleconnections are found, but they are substantially weaker.

Farther south still, over SEAFrica and SWAFrica, there is a broad-scale reversal in sign of the correlation patterns (compared with farther north), with drought being associated with El Niño events and a warming of the central Indian Ocean (cf. Makarau and Jury 1997; Nicholson et al. 2001; Hoerling et al. 2006; and noting that the seasons with strongest SST relationships differ between these two regions).

b. Multidecadal variability of teleconnection strength

The possibility of fluctuations in the strength of SST–rainfall teleconnections is also important to investigate, particularly for its potential to alter the predictability of seasonal rainfall anomalies (cf. Janicot et al. 1996, 2001; Rowell 2001, 2003; who consider this for the Sahel, for example). Such fluctuations may occur because of multidecadal changes in the basic state of the atmosphere or because of changes in SST variability. These in turn could arise either from natural slow variations of the

climate system or from anthropogenic effects. However, the correlation computed between two variables sampled during a stationary period of climate can also be expected to vary merely as an artifact of sampling errors. It is therefore crucial that these variations due to sampling are not confused with genuine variations in teleconnection strength.

Following the methodology of Rowell (2001), the stationarity of the relationships between all SST and rainfall time series is therefore assessed in Fig. 4. This shows the evolution of SST–rainfall correlations in a moving 40-yr window, illustrating apparently large variations for some teleconnections. A null hypothesis that these variations arise only from sampling effects is tested by randomizing the order of the 73 available years. For each of 10 000 resulting pairs of SST–rainfall time series, a new time series of moving 40-yr correlations is computed. The difference between the highest and lowest 40-yr correlations is then used as a metric of the variability of teleconnection strength; this is computed both for the original time series (i.e., those plotted) and for all randomized time series. If, for a particular teleconnection, less than 10% of the randomized data have a correlation range exceeding that of the original time series, then this percentage is recorded in Fig. 4. This only occurs for two teleconnections, a result that is not field significant at the 10% level across the 36 teleconnections analyzed. Similar results are obtained

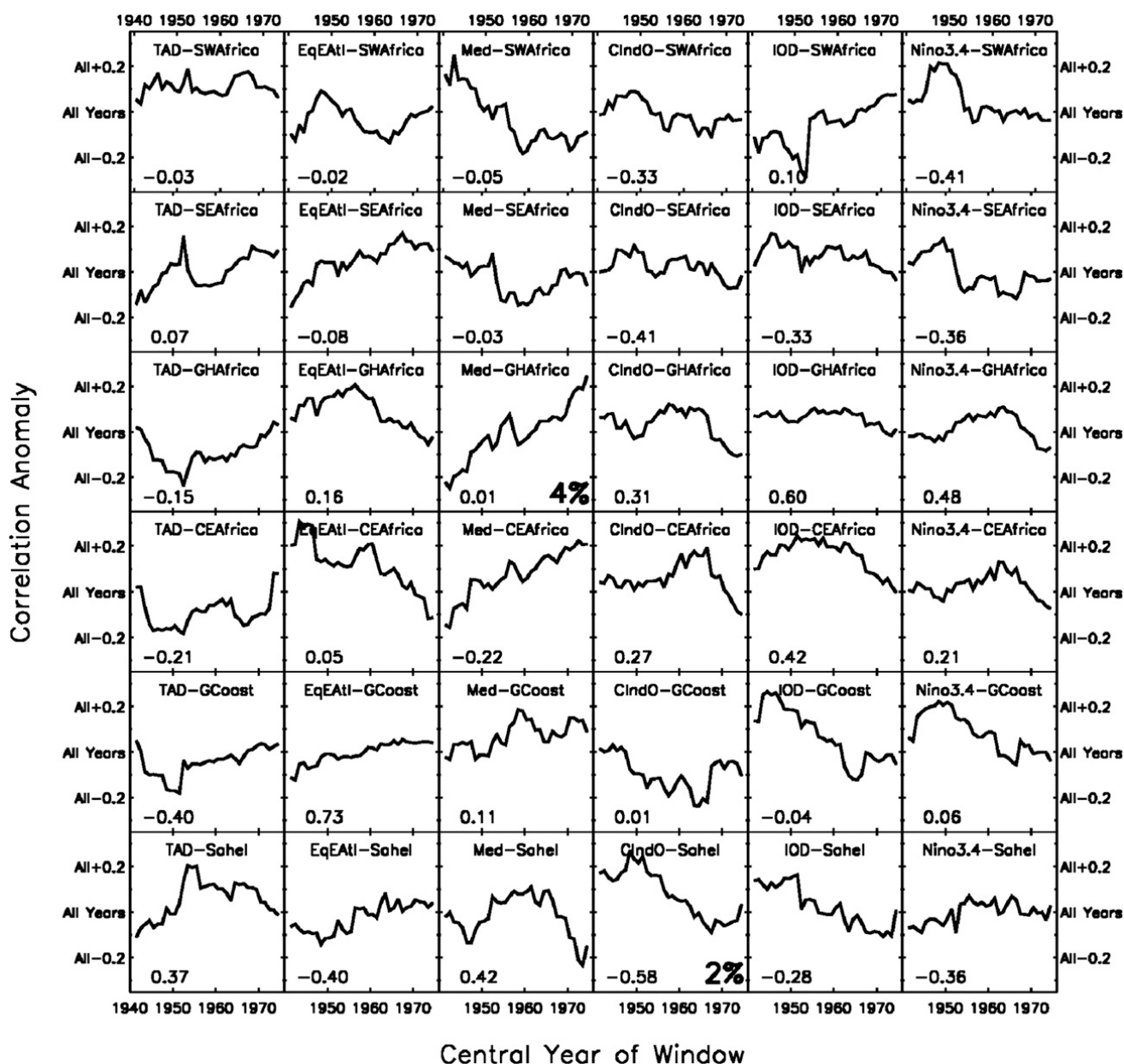


FIG. 4. Time series of correlations between all observed rainfall and SST indices, using a 40-yr moving window from 1922–61 to 1955–94. The x axis shows the central year of each window, and the y axis shows the correlation as an anomaly from the correlation from the full 1922–94 period. The full period correlation is written in the bottom-left corner of each panel. Data in the lower right corner show the significance level, where this is less than 10%, of testing a null hypothesis that the plotted variations in correlation arise only from sampling effects.

using 30-, 50-, and 60-yr moving windows (though with more marginal field significance using the 30-yr window).

Thus, we cannot reject the idea that variations in the strength of twentieth-century SST teleconnections to African rainfall occur primarily because of sampling errors (and therefore that the underlying population correlation remains stationary). This illustrates the need to be aware that apparent multidecadal variability in teleconnection strength may often be imagined rather than genuine. It also demonstrates the

need for long data periods to reliably assess the strength of observed teleconnections and verify their model counterparts.

Last, we note that significant nonstationarity of teleconnections is nevertheless expected over the coming decades, as the magnitude of climate anomalies increases because of anthropogenic changes in atmospheric composition. For the period studied here, these changes have not yet been large enough to have had a detectable impact on correlation statistics.

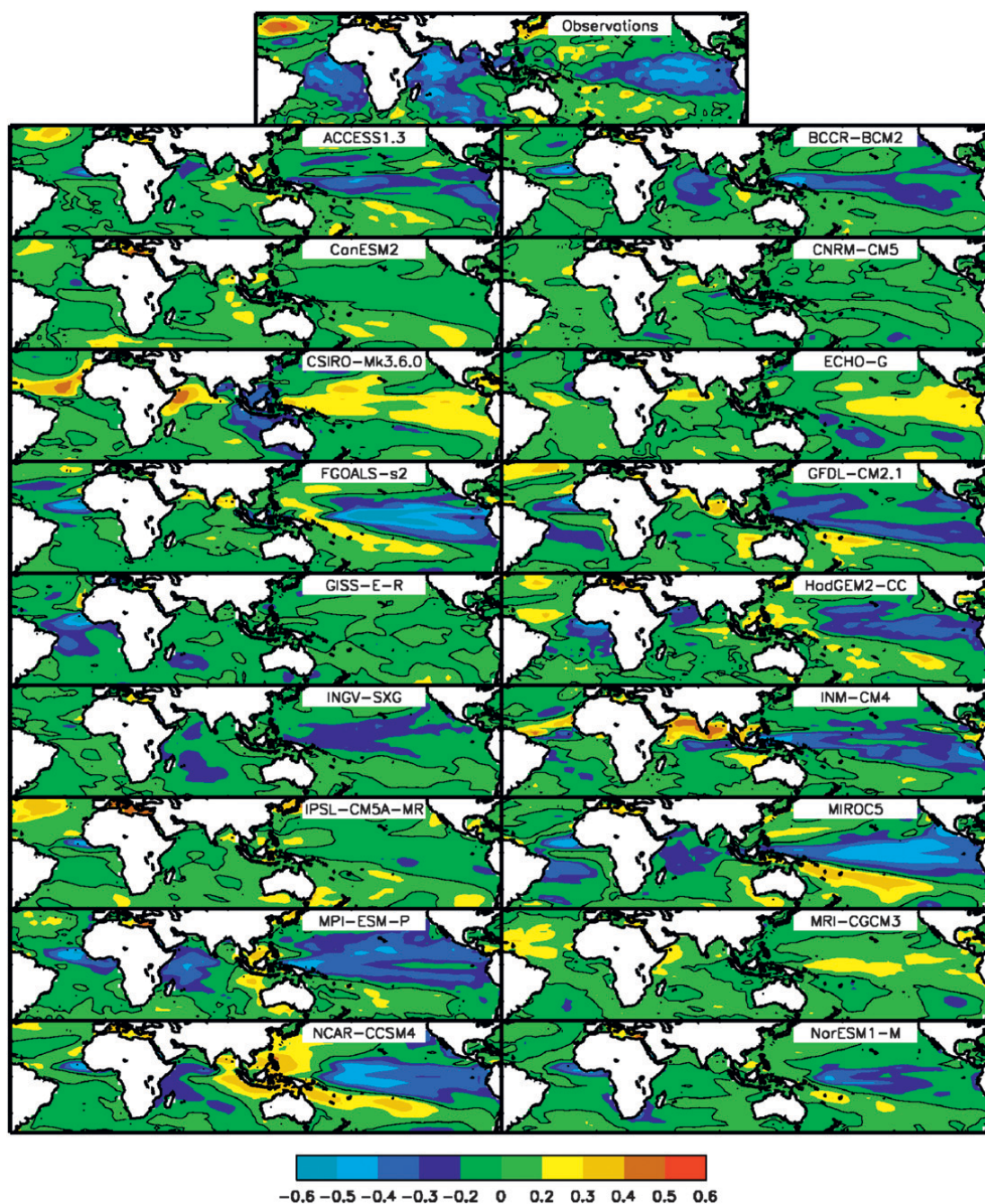


FIG. 5. Correlations between Sahel rainfall and gridbox SSTs (on an N48 grid) for July–September. Observed data use the 1922–94 period, and coupled model data use all available years except the first 5 years. A linear trend is removed from all data. Statistical significance is not assessed.

4. Coupled model teleconnections

a. Example of model SST correlation maps

Ideally, we might like to compute SST teleconnection maps for all the models listed in Tables 1 and 2 and compare these with Fig. 3. However, such a wealth of visual data would not facilitate a useful synthesis of the large number of models and African regions available.

Nevertheless, before providing a more efficient analysis of all teleconnections across all models, it is useful to show an example comparison of model and observed SST correlation maps. Figure 5 therefore shows teleconnections for the Sahel, choosing only the most recent or highest resolution model from each institute. Note that, for consistency with the processing of model data (see section 2a), a linear trend is now also removed from

the observed SST and rainfall data (top panel). A comparison with Fig. 3a demonstrates only a limited impact of this trend removal on the pattern of correlations, although magnitudes are somewhat weaker in the Indian Ocean, tropical northwest Pacific, and subtropical South Atlantic.

In general, the simulated relationship between SSTs and wet-season Sahel rainfall is too weak (Fig. 5). Over the equatorial and South Atlantic, none of the models shown have correlations of similar magnitude and pattern to those observed (although two models not shown, the now superseded FGOALS-g1.0 and MRI-CGCM2.3.2, appear too responsive to SST variability throughout the tropical oceans). Similarly, over the Indian Ocean, correlations are very weak or sometimes of the wrong sign, with the possible exception of the MPI-ESM-P model. Over the Pacific, the simulated impact of ENSO on the Sahel is again often weak, nonexistent, or occasionally of the wrong sign, although here a small minority of these models do exhibit approximately the same teleconnection strength as observed. Finally, in contrast, links between the Mediterranean and Sahel appear to be somewhat easier to simulate, with a notable number of coupled models exhibiting the positive relationship found in the real world.

A further limitation of the comparative approach of Fig. 5 is that it does not address the likelihood that some of the differences found between model and observed correlations must arise from random differences in the strength, number, and combination of anomalous SST events that have been sampled. We can ask whether the subjective assessment described in the previous paragraph is either too optimistic or too pessimistic about model performance, given this context of sampling differences. This too should be addressed in a more objective assessment.

b. Continent-wide assessment of teleconnection errors: Methods

Here we describe an objective and quantitative approach to synthesizing and verifying the SST teleconnections to each African region for all available coupled models. The aim is to provide an overview of the state-of-the-art of modeling African teleconnections.

A key step is to reduce the gridded SST data to one-dimensional time series using the six indices described in section 2b. These are known to be the primary oceanic drivers of interannual variations of African rainfall. Correlations between the SST and rainfall time series are then used to measure the strength of each of the SST teleconnections to African rainfall and the extent to which these differ between model

and observed data. This difference (i.e., an assessment of model skill) is measured by testing the null hypothesis that the observed and modeled correlations (for a particular teleconnection) derive from the same underlying population, using a χ^2 test on Fisher-z-transformed correlations. Again, the calculation accounts for serial correlation, through the use of an “effective number of degrees of freedom” for each time series (see section 3a).

Following this procedure, a three-dimensional matrix of statistical significance (the skill metric) is produced. This is reduced to a two-dimensional diagram in Fig. 6, with the larger boxes showing results for each of the 6×6 teleconnections; then within these, the colored pixels showing results for each model. The color key is as follows:

- Green: This indicates at least reasonable model teleconnection skill. In these cases, the null hypothesis that the observed and modeled correlations derive from the same population is not rejected (at the 10% level): that is, differences could merely be due to sampling effects. (Note that it is not possible to go further and determine whether a model perfectly reproduces the observed teleconnection, because of the unavoidable issue that different phases of natural variability are sampled in the model and real world.)
- Pale brown: This indicates a moderate difference between model and observed teleconnections. Here the model correlation is significantly different from that observed at the 5% level: that is, there is likely a genuine deficit in model skill.
- Dark brown or red: This quantifies poor or very poor skill in a model’s teleconnection, indicated by the rejection of the null hypothesis at a more stringent significance level of 1% or 0.1%, respectively.
- Yellow: This indicates some difference between model and observed teleconnections, which is statistically significant at the 10% level but not at the 5% level.
- White: This is used where results are of little practical interest. It replaces many pixels that would otherwise be green (i.e., where a model has at least reasonable skill and the null hypothesis is not rejected) but is instead used where both the observed and model correlations are not significantly different from zero at the 10% level (this threshold is $r = 0.19$ if serial correlation is low). Thus, part-white boxes indicate teleconnections with little or no relationship between observed SSTs and rainfall (of which there are 18). However, some pixels within these boxes are colored because these models either simulate the existence of a significant teleconnection

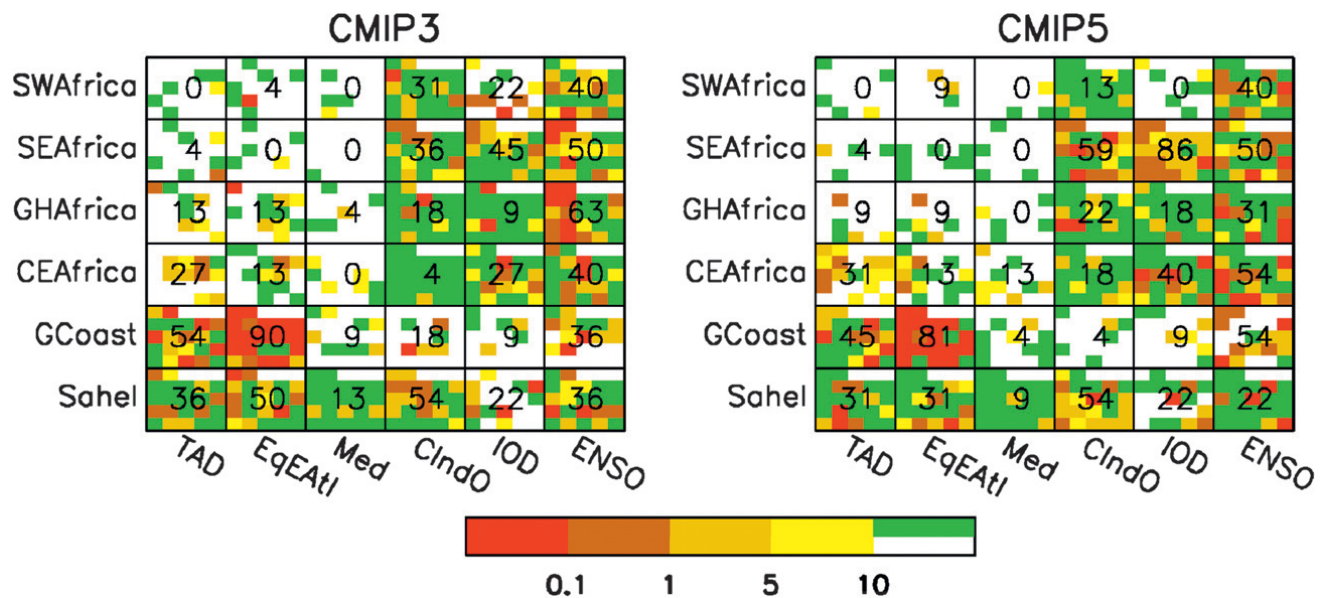


FIG. 6. Model teleconnection skill for CMIP3 and CMIP5 data. Larger boxes show results for each teleconnection, and pixels show results for each model. The color key shows the significance level at which a null hypothesis that the observed and modeled correlations derive from the same population is rejected. The meaning of these colors and skill metric are described fully in section 4b. The models are ordered alphabetically by model name (as in Tables 1 and 2) from the bottom-left pixel row by row to the top pixels, and the last three pixels are not used. The percentage of red + brown pixels is printed within each box: that is, the proportion of model teleconnections that differ significantly from those observed at the 5% level.

and/or differ significantly from the observed teleconnection, so in either case an assessment against observations is provided.

Figure 7 illustrates the effective sample size as a fraction of the actual sample size (N^*/N ; Bartlett 1935; Folland et al. 1991), calculated for each pair of time series, using observed data and CMIP3 data (results from the CMIP5 data are similar; not shown). For observed data (bottom-left pixel in each box), N^* is reduced by serial correlation for only three teleconnections: those from TAD, Med, and CIndO to the Sahel, for which N^* is 63, 37, and 34 years, respectively (N is 73 years). This reduces confidence in our estimates of the observed strength of teleconnections from Med and CIndO to the Sahel (less so for TAD), and so it becomes more difficult to determine whether a model teleconnection is significantly different. In other words, we need to be aware that the green pixels in these two boxes in Fig. 6 represent a less stringent assessment of model skill. Figure 7 also shows that in general the modeled values of N^* are mostly little different to those computed using observed data (noting the nonlinear scale of Fig. 7). The exceptions, however, are the Med and CIndO teleconnections to the Sahel, for which almost all models fail to capture the redness of the observed covariance spectra. This partly arises from a failure of most (but not all) coupled models to capture

the redness of the observed SST spectra and hence also the observed Sahel rainfall spectrum (not shown).

The assumptions and choices made in the methodology presented here have also been verified with the

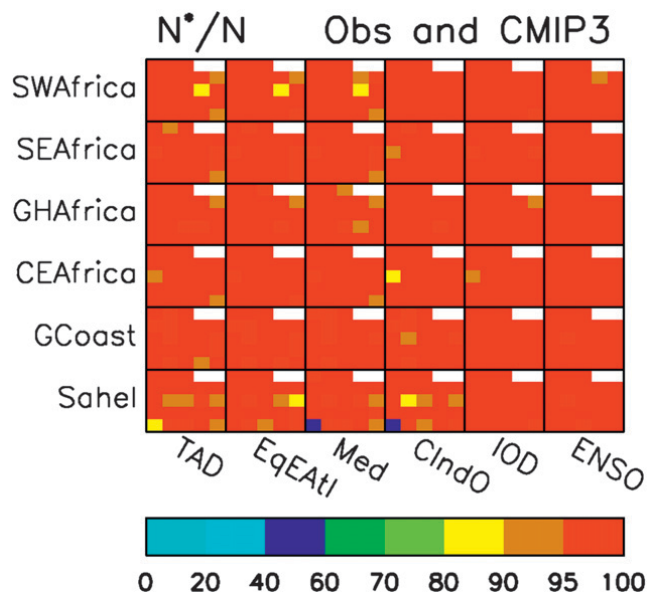


FIG. 7. Effective sample size as a percentage of actual sample size (N^*/N), calculated for each pair of SST/rainfall time series, using observed and CMIP3 data. Larger boxes show results for each teleconnection, and pixels show results for each model. The bottom-left pixel is observed data, and the remaining pixels are CMIP3 data, ordered alphabetically by model name (as in Tables 1 and 2) row by row from bottom left to top right; the last two pixels are not used.

following sensitivity tests: (i) use of UDel2.01 observed precipitation data in place of CRU TS3.1; (ii) use of model data for only the period 1922–94; (iii) use of longer periods of model data (112 or 132 yr) and necessarily excluding models with less data; (iv) use of unforced CMIP3 simulations (i.e., with no variation in atmospheric composition) in place of the 20C3M simulations; and (v) removal of exceptional years (the driest year and wettest year) from the time series for each teleconnection. Only limited sensitivity was found in each case, such that the main conclusions of this paper are unaffected, with the exception that in some instances the CIndO–SWAfrica teleconnection could also be highlighted below as one for which coupled models are generally more capable.

Finally, the possibility of multidecadal variability in the strength of coupled model teleconnections has also been investigated, using the same approach as section 3b. Results (not shown) indicate that for the majority of models the number of teleconnections with apparently significant correlation variability is not field significant across the 36 teleconnections. Exceptions are the CCSM4 and INM-CM4 models, with 9 and 12 (respectively) teleconnections that have significant nonstationarity at the 10% level, and possibly also the CGCM3.1, ECHO-G, and MPI-ESM-LR models, all with 7 teleconnections with significant nonstationarity. Thus, for the majority of models, the correlation statistics analyzed here are approximately stable, and it is also unlikely that different sampling of the phase of natural modes in model and observed data has a notable impact on their estimated teleconnection strengths. Nevertheless, for a small minority of models, the assessments of section 4c are less reliable for those teleconnections with unrealistic multidecadal variations in their strength.

c. *Continent-wide assessment of teleconnection errors: Results*

Figure 6 addresses the key question: Are some SST–Africa teleconnections easier to model than others? In short, the answer is “yes,” with the analysis showing a clear mix of generally well-modeled teleconnections, generally poorly modeled teleconnections, and other teleconnections with a wider range of model skill. Unfortunately, there is no improvement in skill between the CMIP3 and CMIP5 databases. The teleconnections that uninitialized coupled models particularly struggle with are as follows: the link between the equatorial east Atlantic and July–September rainfall in the Guinea coast region, the link between the IOD and southeast Africa in December–February (primarily the CMIP5

models available here), and the impact of ENSO on the short rains (OND) of the Greater Horn of Africa (primarily the CMIP3 models). Furthermore, differences between modeled and observed correlations for the CIndO–Sahel teleconnection are similarly poor but, since this link is exposed to a more lenient assessment (see section 4b), the role of sampling errors in its assessment is less clear in Fig. 6. On the other hand, teleconnections that coupled models are usually much more able to simulate are those from the Mediterranean to the Sahel in July–September (this withstands its more lenient assessment in Fig. 6), from the central Indian Ocean to central East Africa in October–December, and from the IOD to the Greater Horn of Africa in OND. We also note that most models correctly simulate the lack of relationship observed in many SST–rainfall pairings.

To gain further insight, Fig. 8 shows a comparison of the raw correlations assessed and synthesized in Fig. 6. It can be seen that the poorly represented teleconnections are frequently simulated as too weak in their magnitude (and very rarely too strong). This general tendency is quantified in Fig. 9a, which demonstrates that the large majority of coupled models underestimate the strength of 80%–100% of SST–Africa teleconnections (analyzing only the 18 teleconnections that are significant in observed data). Only four models (CNRM-CM3, GFDL CM2.1, MIROC5, and MPI-ESM-LR) underestimate the strength of close to 50% of these teleconnections, as would be expected by chance by a perfect model. Alternatively, Fig. 9b shows the average strength of each model’s teleconnections (measured by the average of its absolute correlations). Many models tend to substantially underestimate teleconnection strength between SSTs and African rainfall, with only a few achieving an average value close to that observed. This may be partly due to a weak response to SST variability in many coupled models and perhaps also excessive chaotic atmospheric variability and feedbacks.

Nevertheless, it is also clear from Fig. 8 that the poor magnitude of teleconnections is not the only problem faced by coupled models. For many teleconnections (and especially those listed above as poorly modeled; marked in Fig. 8), there are a number of models that even fail to simulate the correct sign of relationship. Furthermore, a visual inspection of correlation maps between the SST indices and gridbox rainfall (not shown) demonstrates that modeling errors are often more serious than simply an error in geographic location or extent of anomalous rainfall. There are significant issues to be addressed in many models for at least some SST to Africa teleconnections.

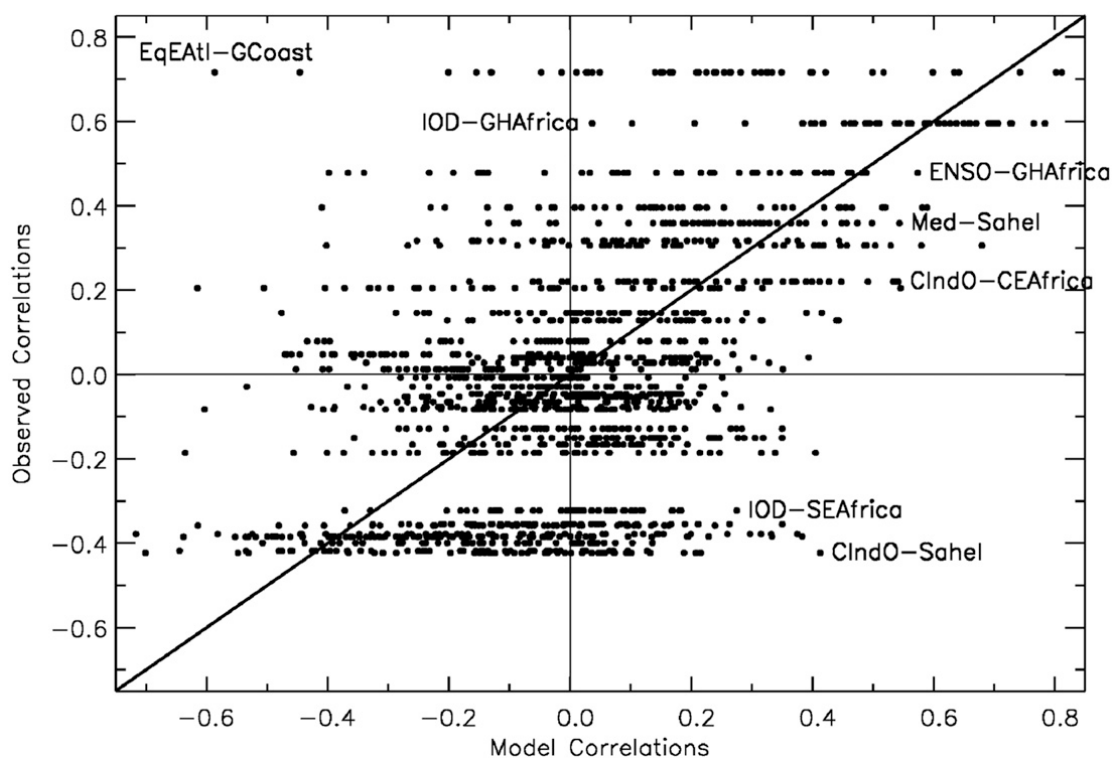


FIG. 8. Scatterplot of the correlations between all combinations of SST and rainfall time series, for all coupled models (x axis) against observations (y axis). Horizontal lines of points therefore illustrate results for a particular teleconnection, for which the observed correlation is fixed. The diagonal line shows where the population correlations of a perfect model would lie. Teleconnections with the highest and lowest multimodel skill are labeled.

A second key question was raised in the introduction: Do some models consistently represent these teleconnections better than others? This is addressed by Fig. 10, which displays the same pixels of the skill metric as Fig. 6

but grouped by model and then sorted by model performance. Clearly, some models do perform consistently better than others in this regard, with the best able to simulate 80%–85% of SST–Africa teleconnections with

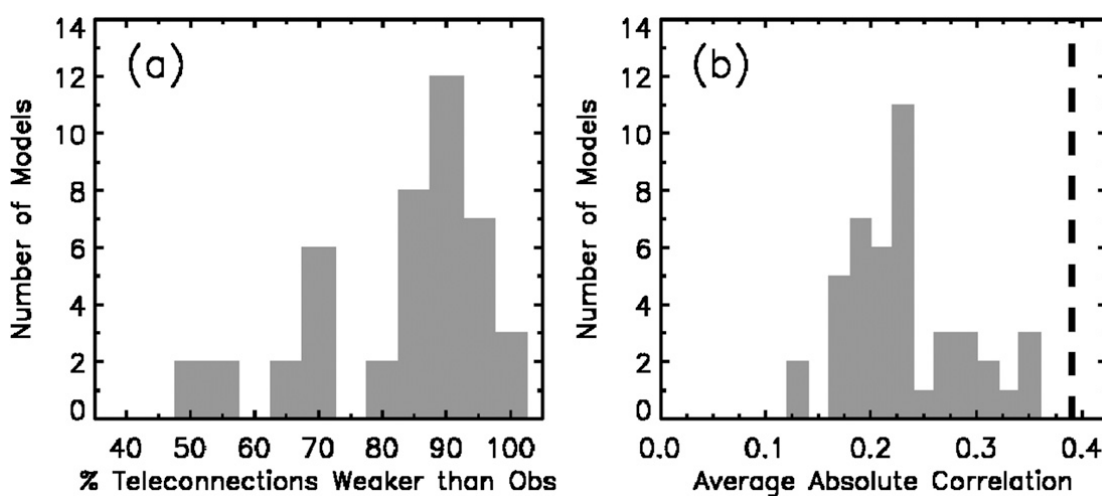


FIG. 9. Histograms of two measures of the aggregated strength of a coupled model's teleconnections. (a) The percentage of a model's SST–rainfall correlations that are weaker than observed (model correlations of the wrong sign but excessive magnitude are counted as too weak, although there are few such cases). (b) The average of a model's absolute correlations. Calculations include only the 18 teleconnections significant in observed data. In (b), the dashed line shows the observed average.

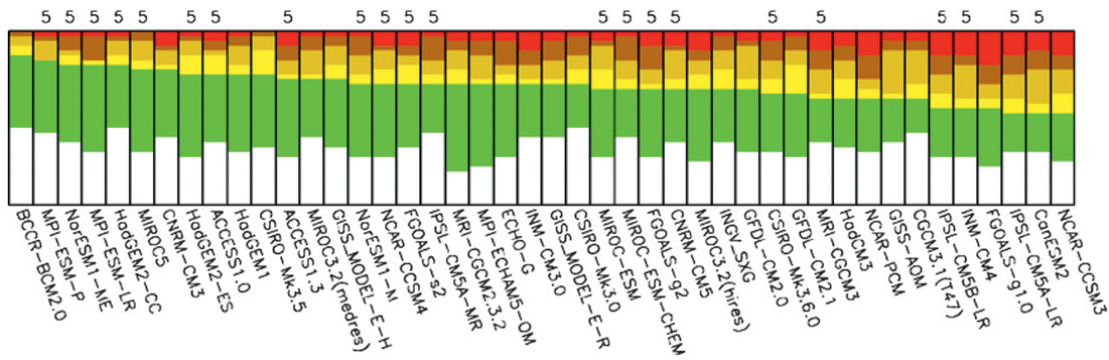


FIG. 10. For each coupled model, the proportion of teleconnections in each of the five skill categories described in section 4b. Models are ranked by the number of teleconnections that do not differ significantly from those observed, at the 10% level (i.e., the boundary between green and yellow). Models from the CMIP5 database are marked “5” above the plot.

at least reasonable skill (i.e., within the range of sampling variability at the 10% significance level) and the poorest able to simulate only 55% of teleconnections. Note that this fraction includes the 18 teleconnections that are insignificant in observations, because for prediction purposes it is equally important that the lack of relationship between rainfall and some SST regions is correctly simulated by models. However, since these “nonteleconnections” tend to be easier to simulate (see above), their inclusion leads to a slightly more optimistic view of modeling capability. Last, Fig. 10 also confirms the statement that there has been no improvement between CMIP3 and CMIP5 in their capability to model SST–Africa teleconnections.

5. Initial assessment of the causes of coupled model teleconnection errors

In this section, we begin to explore why some teleconnections are harder to model than others and why

some models are more skillful than others in this respect. However, it will be seen that this analysis raises many further questions and so should be seen simply as a starting point for further substantive investigation.

a. Impact of SST errors on model teleconnection skill

Inevitably, the patterns and magnitude of SST variability in coupled models differ from those of the real world (e.g., Guilyardi et al. 2009), and so their role in the teleconnection errors described in section 4 is now explored. Figure 11 assesses the representation of SST-to-Africa teleconnections in two ensembles of atmosphere-only simulations (see section 2a), forced by observed SSTs; that is, these are driven by a near-perfect marine surface boundary. This uses an identical assessment methodology to that used above, so that Figs. 6 and 11 can be directly compared. Note also that HadAM3 is the atmospheric component of HadCM3, for which results are shown by the center pixel in each box of the

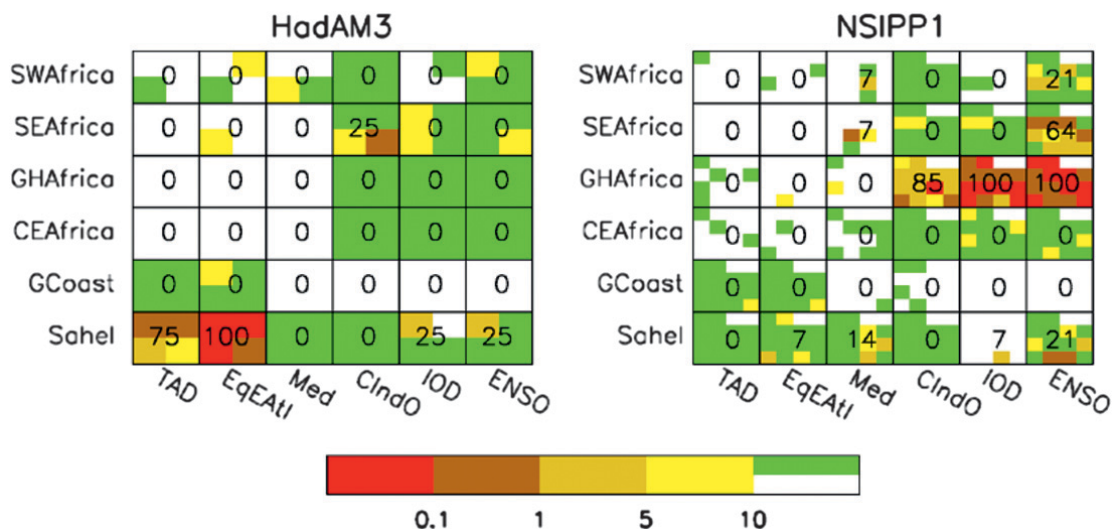


FIG. 11. As in Fig. 6, but for two AGCM ensembles; a 4-member initial condition ensemble of HadAM3 simulations, and a 14-member ensemble of NSIPP1 simulations. The two top-right boxes are not used in the NSIPP1 panel.

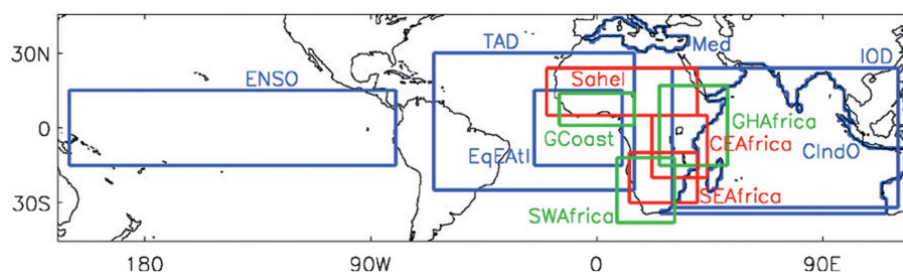


FIG. 12. Definition of the regions used to compute RMSEs of SST (blue) and rainfall (green or red).

CMIP3 analysis in Fig. 6. NSIPP1 has no corresponding coupled model in the CMIP databases.

It is clear that these AGCMs are more capable overall than the coupled models, suggesting that poor SST variability is indeed a notable contributor to poor teleconnection skill in coupled models. In particular, three of the four low-skill teleconnections highlighted above are much improved when perfect SSTs are used: EqEAtl–GCoast, IOD–SEAFRICA, and CIndO–Sahel. Furthermore, for HadCM3, all but one of its poor teleconnections are notably improved in its matching AGCM simulations.

Nevertheless, a small minority of teleconnections are instead less well represented with perfect SST forcing, as seen by comparing HadAM3's assessment of TAD–Sahel and EqEAtl–Sahel with the corresponding HadCM3 assessment. One explanation is that the lack of two-way ocean–atmosphere coupling in AGCMs may degrade their response to SSTs in some regions (cf. Wang et al. 2005; Wu and Kirtman 2007), such as to significantly affect a minority of teleconnections. Alternatively, both HadAM3 and HadCM3 may be unable to properly represent the atmospheric bridge from the Atlantic to Sahel, but there is a compensating effect from SST errors in HadCM3. Either way, this is clearly a model-dependent result, since NSIPP1 is able to reliably simulate both these Atlantic–Sahel teleconnections.

However, NSIPP1 is unable to properly simulate other SST teleconnections to Africa: those from the IOD, ENSO, and CIndO to GHAFRICA and that from ENSO to SEAFRICA. The first two of these are particularly poor. So, not surprisingly, it appears that the difficulty that models have in simulating the atmospheric component of teleconnections to Africa is also strongly model dependent.

Last, from a methodological point of view, Fig. 11 again emphasizes the role of natural variability when assessing teleconnections and the need for long data periods. Here, only the internal component of atmospheric variability differs between the simulations from each model and yet some apparent variation in skill is found,

even using 93 years of data. Hence, in this study, we use the available data only to draw general conclusions across a number of models or teleconnections. If particular teleconnections in individual models are to be assessed, it is wise to attach confidence intervals to the analysis; these might be quite large, particularly for short data periods.

In conclusion, from the point of view of beginning to understand the causes of poor coupled model teleconnections to Africa, the key finding here is the large contribution from their poor representation of SSTs (although this will need to be verified with AOGCM–AGCM pairs from other models). These SST errors can impact teleconnection skill in three distinct ways, via (i) erroneous forcing of the atmosphere overlying the oceanic source of the teleconnection, either due to an incorrect response of the surface fluxes and boundary layer because of bias in the SST mean state or due to incorrect variance of SSTs and their gradients; (ii) an erroneous representation of the atmospheric bridge from the oceanic region to the African region, in this case purely due to SST errors disrupting the atmospheric basic state; and (iii) an erroneous rainfall response over Africa, due to the impact of SST errors on the model's regional climatology over Africa. The next two sections begin to explore the role of error types (i) and (iii), whereas analysis of error type (ii) is beyond the scope of this study since it requires considerably more data and the development of more sophisticated techniques.

b. Model biases

To investigate the role of tropical ocean and African atmospheric biases on the models' representation of African teleconnections, skill metrics of these biases are required.

First, errors in the climatology and gradients of SSTs in the ocean source region are measured by the root-mean-square error (RMSE) of model SSTs, verified against HadISST1.1, and computed using the regions shown in Fig. 12. Analysis of larger regions than the defined indices allows inclusion of SST gradient errors around the margins of the indices and of SST errors in

the surrounding region, both of which may affect the atmospheric response via regional errors in its low-level climatology. Figure 13 (left-hand column) shows a wide range of skill among CMIP3 + CMIP5 coupled models and a tendency for the largest errors to be found in the Mediterranean and Atlantic.

Second, errors in the magnitude of the SSTs that drive the teleconnections are measured by focusing on the standard deviation (SD) of the mode of interest, and computing its ratio to the observed SD (again using HadISST1.1). Figure 13 (middle column) again shows a wide range of model skill, as well as strong tendencies to underestimate variability in the EqEAtl and overestimate variability of the IOD. Note that a wider and more detailed measure of SST variability, the RMSE of SD patterns over the areas shown in Fig. 12, leads to identical conclusions, both here and below.

Third, errors in the atmospheric climatology of relevant areas and seasons of Africa are assessed using the RMSE of rainfall, computed over the regions shown in Fig. 12 and against CRU TS3.1 data. Again, these larger analysis regions allow inclusion of wider errors in the regional climatology that may affect the local response to remote effects. Again, Fig. 13 (right-hand column) illustrates a wide range of model capability. Note that the tendency for errors to be largest in and around GCoast may be because this region has the highest mean rainfall and that to the first-order rainfall errors tend to scale with mean rainfall.

c. Relationship between climatology errors and model teleconnection skill

Figure 14 shows, for some teleconnections, how these metrics of model bias relate to model teleconnection skill. The four chosen examples are teleconnections that the models find hardest to reproduce. The strength of these relationships is measured by the correlation between the bias metric and the skill metric, and this is printed in each panel of Fig. 14. For this calculation, teleconnection skill is measured by the logarithm of the significance level, with data less than $\log(10^{-6}\%)$ reset to this lower limit to avoid occasional extreme negative values. All teleconnections significant in the observed data (of which there are 18) are analyzed in this way, with the exception of the three that are mostly well modeled (section 4c), since their small range of skill is likely to be relatively more affected by sampling effects and so less by model climatological errors. Histograms of these correlations between model bias and teleconnection skill, for these 15 teleconnections, are shown in Fig. 15. Note that the reliability of individual correlations is unlikely to be high because of the sizable impact of outlying models in some cases (Fig. 14), so the interpretation here focuses on the

broad characteristics of the distributions shown in Fig. 15 more than on individual teleconnections.

First, Fig. 15a and the top row of Fig. 14 show that coupled model biases in the mean pattern of SSTs (and so also their gradients) have no obvious and straightforward impact on the reproduction of African teleconnections. Although the EqEAtl–GCoast teleconnection shows the expected relationship—the models with largest errors have least skill—this is not the case for the other teleconnections shown in the top row of Fig. 14 and for many others not shown, which display an apparently random scatter across the domain. Consequently, Fig. 15a shows a distribution of correlations centered approximately on zero. This, suggests that the apparent relationships found for a small minority of teleconnections (the most negative correlations), such as EqEAtl–GCoast, are either abnormal or may occur purely by chance. Similar results are obtained if the bias in the mean of the SST indices or their absolute bias is analyzed instead (not shown). Nevertheless, this does not contradict our expectation that SST errors around the marine source region should contribute to teleconnection errors; the results presented here simply show that either the response to such errors is quite complex or that the mechanisms involved are not consistent across models.

Next, a weak relationship between the variance of the SST indices and the skill of modeled teleconnections is apparent. Figure 15b is skewed toward positive correlations, suggesting that SST anomalies of sufficient (or even excessive) magnitude are more likely to lead to skillful teleconnections than if SST anomalies are too weak. This seems physically plausible and to some extent is illustrated in the scatterplot for the EqEAtl–GCoast teleconnection (Fig. 14, middle row). Nevertheless, this relationship is not strong and is not consistent across all teleconnections. Not surprisingly, other factors also contribute substantially to teleconnection skill.

Last, Fig. 14 (bottom row) and Fig. 15c show that biases in the climatological pattern of African rainfall have no clear and obvious impact on the reproduction of African teleconnections. None of the examples of Fig. 14 demonstrates a relationship between these metrics, and Fig. 15c shows a distribution of correlations approximately centered on zero. Again, use of alternative metrics (the climatology of the rainfall indices, or their absolute climatology, or their coefficient of variation; not shown) leads to the same conclusion. Thus, the extent to which errors in the African regional climatology contribute to the teleconnection errors is unclear. The teleconnections may only be affected by complex responses to African climatological errors that cannot be detected

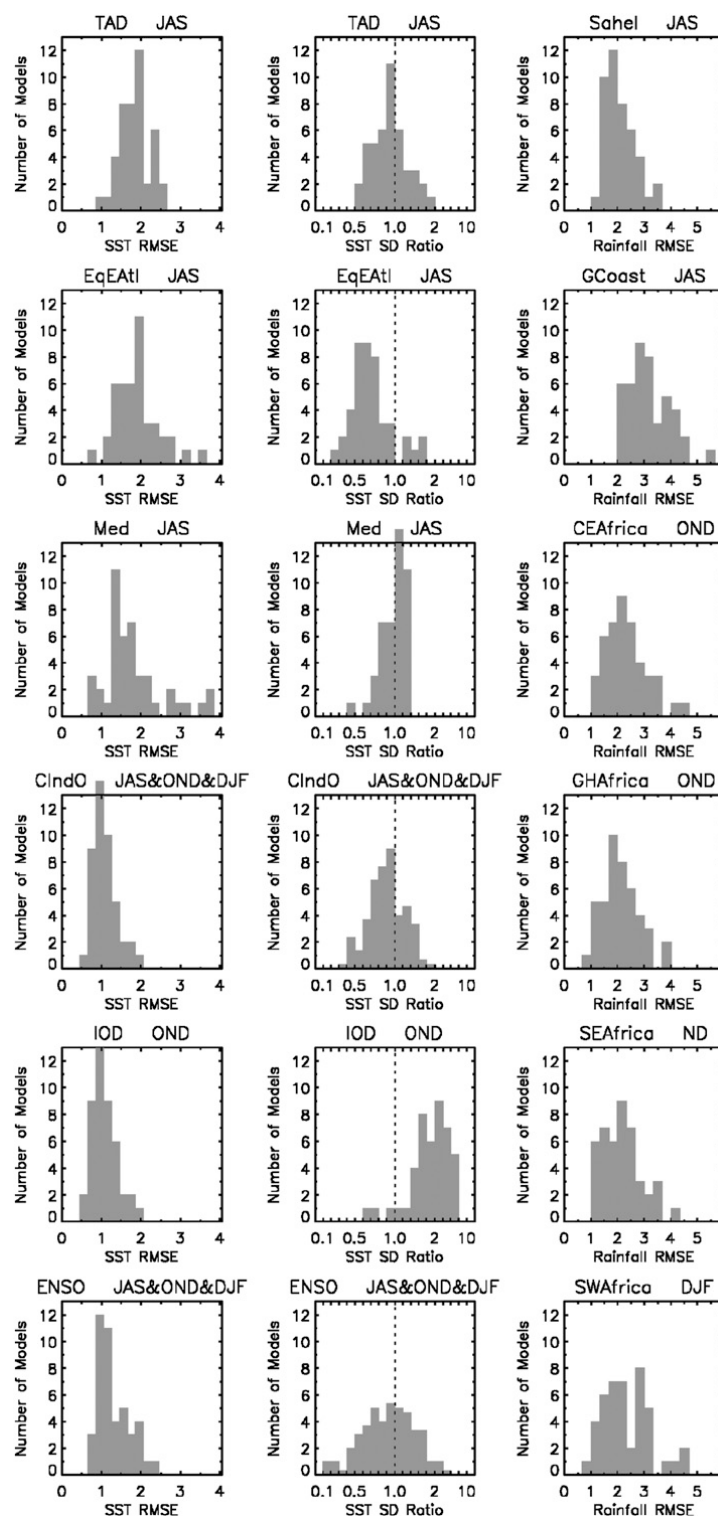


FIG. 13. Histograms of (left) the RMSE of modeled SST, (middle) the ratio of modeled to observed SD of SST indices, and (right) the RMSE of modeled rainfall. Errors from both CMIP3 and CMIP5 models are shown. SST RMSEs are computed over the single or three seasons marked in the panel title. Histograms of SD ratios are computed for individual seasons, and where marked over the panel title these histograms are averaged over three seasons. Agreement between model and observed SDs is shown by the dashed line in the central column.

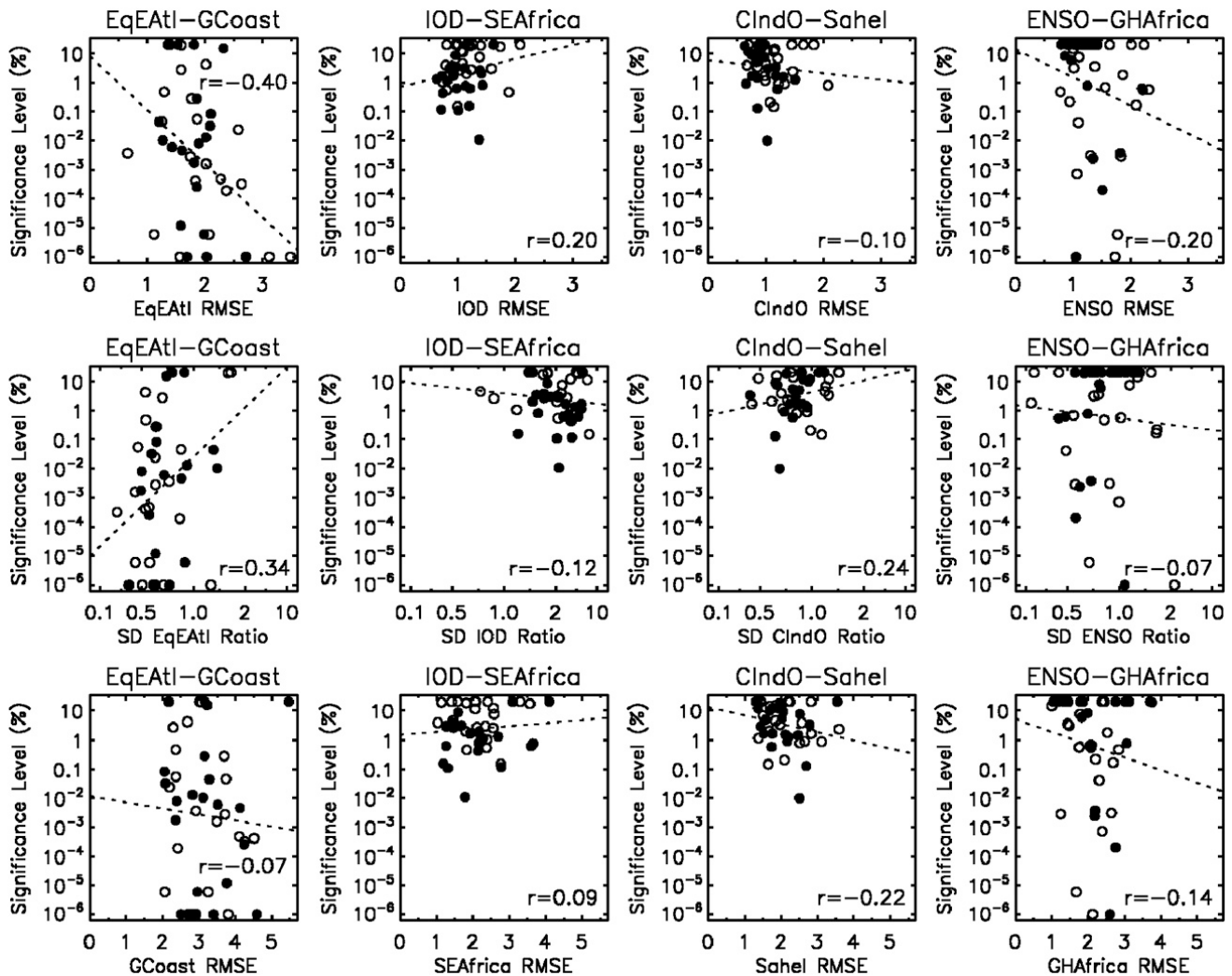


FIG. 14. Scatterplots of model bias metrics (labeled on the x axis) against teleconnection skill (y axis; with a lower limit of $10^{-6}\%$), for the four teleconnections that the models find hardest to reproduce. All CMIP3 (open circles) and CMIP5 (filled circles) coupled models are shown. Correlation (r) between the bias metric and skill metric is shown, along with the linear regression fit between the bias metric and skill metric (dashed line).

by the correlation analysis presented here, or it may be that the roles of error types (i) and (ii), listed at the end of section 5a, dominate that of African climatological errors. These topics will require considerable further investigation.

d. Relationship between model resolution and teleconnection skill

We might also expect and hope that those models with highest resolution will outperform those with lowest resolution. For example, Roberts et al. (2009) show that higher atmospheric and marine resolution lead to an improved mean state in the tropical Pacific and the atmospheric Walker circulation, using a suite of HadGEM1 simulations. The possibility that a similar effect might also be seen in the multimodel context is assessed here by

correlating the logarithmic teleconnection skill metric with the models' atmospheric or marine horizontal resolution, for each of the same 15 teleconnections (those that are significant in observations, with the exception of those that are generally well modeled).

Figure 16 illustrates the histograms of the resulting correlations, showing that both distributions are approximately centered on zero. So, although a relationship between model resolution and teleconnection skill cannot be ruled out for a very small minority of African teleconnections, it is clear that no consistent relationship exists across models for the current range of resolutions. However, there may still be a "threshold resolution"—not yet reached by models—at which a more widespread and obvious improvement is found. Furthermore, this result does not negate the impact of resolution when

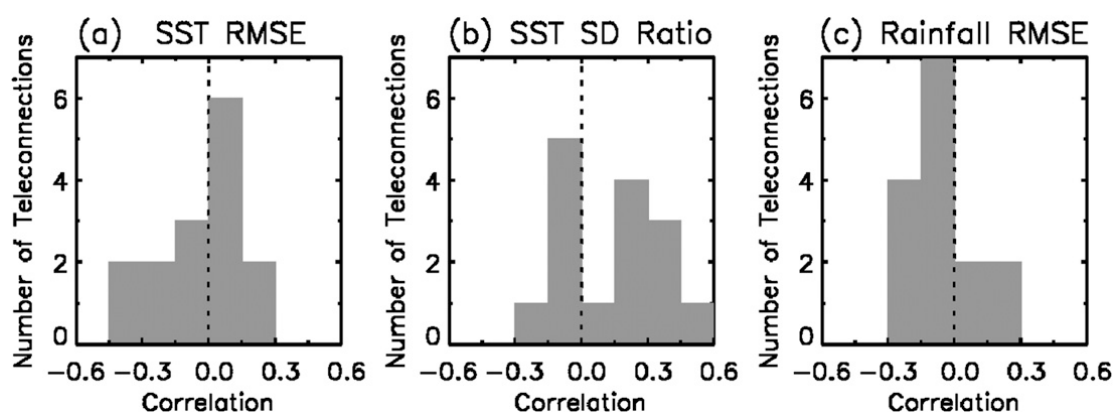


FIG. 15. Histograms of the correlation between model bias metrics and teleconnection skill, calculated separately for each teleconnection. Bias metrics are (a) RMSE of SST, (b) ratio of modeled to observed SD of SST indices, and (c) RMSE of rainfall. For each correlation, the region or index used to compute the bias metric is the same as that of the teleconnection. All teleconnections significant in the observed data are used, except the three that are mostly well modeled. Zero correlation is marked by the dashed line.

isolated in a particular model (e.g., Navarra et al. 2008; Roberts et al. 2009). Nevertheless, it is clear that, at present, other structural and parameterization choices made during the development of a climate model are at least as important as horizontal resolution in achieving reliable simulations of SST-to-Africa teleconnections.

6. Conclusions

This study has provided an overview of the capability of coupled models to represent teleconnections from anomalous SST patterns to African rainfall. The focus has been on six large-scale regions of sub-Saharan Africa, each with an associated season over which rainfall averages have been computed. These were defined such that interannual rainfall anomalies are spatially coherent within each region and temporally coherent (month to month) within the season. Time series of these rainfall indices were then correlated with indices of six SST modes known to affect African rainfall. Of the resulting 36 teleconnections, 18 are statistically significant (i.e., the SST–rainfall correlation is significantly different from zero) in observations. Apparent multidecadal variability in the strength of all 36 teleconnections was also examined, with no evidence found to reject the idea that these might arise simply from sampling variability.

A systematic approach has evaluated the ability of 44 AOGCMs to represent these 36 teleconnections. These are models that contributed to the Intergovernmental Panel on Climate Change (IPCC) Fourth and Fifth Assessments. Their simulations of twentieth-century climate variability were employed, providing the long data periods necessary to distinguish model deficiencies from

the effects of random sampling when compared against observations. Such simulations are continuous with no assimilation of observations. This is appropriate for assessing teleconnection skill for climate change studies, for which such simulations form the historical context. However, the resulting assessment likely provides only a lower bound on the teleconnection skill that might be found in initialized seasonal predictions, where SST errors have yet to reach their equilibrium state. Further work is required to investigate the rate of development of SST errors in initialized predictions, as well as their consequent representation of SST–Africa teleconnections, as far as this is possible given their smaller sample of years.

In section 1, three questions were posed, which have been addressed as follows:

Are some teleconnections easier to model than others? Yes, there are a few teleconnections that the large majority of models are able to capture with at least

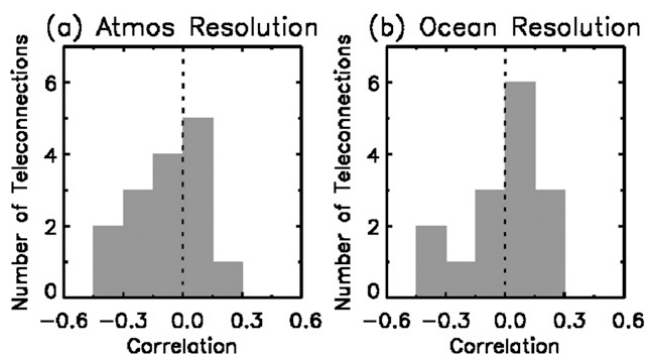


FIG. 16. As in Fig. 15, but for correlations between model tropical horizontal resolution and teleconnection skill, for the (a) atmosphere and (b) ocean components of the models.

reasonable skill: Med–Sahel, CIndO–CEAfrica, IOD–GHAfrica, and perhaps CIndO–SWAfrica. Additionally, there are other teleconnections that the majority of models struggle to simulate: EqEAtl–GCoast, IOD–SEAfrica, ENSO–GHAfrica, and perhaps CIndO–Sahel. This latter group now requires particular attention, taking a detailed mechanistic approach to unravel the problems that models face. The remaining teleconnections exhibit a wider range of skill across models, apart from those that are not significant in observations, which are similarly insignificant in most models (but usually not all models).

Are some models consistently better than others? Yes, the best models are able to at least adequately capture 80%–85% of the 36 SST–Africa teleconnections, whereas the poorest models capture only 55% of these teleconnections. However, there is no improvement between CMIP3 and CMIP5.

Can we begin to say anything about the causes of teleconnection errors and their differences between models? Yes, but further substantive work is required. An analysis of AGCM simulations suggests that coupled model teleconnection errors may arise primarily from errors in the coupled models' SST climatology and variability (although analysis of further AGCM/AOGCM pairs should be undertaken to confirm or refute this). Nevertheless, it is likely that errors in the atmospheric component of teleconnections also play a significant role, so they should not be ignored in future studies. A further finding has been that many models tend to underestimate the strength of SST teleconnections to Africa. In some cases, insufficient variance of tropical SSTs is partly to blame, but there may also be a contribution from excessive chaotic variations in the atmospheric teleconnection chain; this latter idea requires further examination. In contrast, no straightforward link has been found between climatological biases (in SSTs or African rainfall) and teleconnection skill or between model resolution and teleconnection skill for the CMIP3 and CMIP5 models. Although these features may nevertheless play a role, it is clearly not a dominant role, and

the causes of teleconnection errors may be a complex combination of factors, the balance of which varies between models and between teleconnections.

The lack of improvement between CMIP3 and CMIP5 suggests that a “bottom up” approach to model development (i.e., improving the realism of parameterization schemes, with the expectation that improved atmospheric behavior will follow) is not yet providing better modeling of SST teleconnections to Africa. It appears, therefore, that improvements will be hastened by the addition of a “top down” approach: that is, by focusing on the performance of individual teleconnections (noting the need to be aware of sampling effects) and attempting to understand the causes of their errors through idealized experiments and analysis of process-based diagnostics. If such an approach is successful, then improved seasonal predictability will follow, and so this poses an important challenge to the scientific community.

Acknowledgments. This document is an output from a project funded by the U.K. Department for International Development for the benefit of developing countries. The views expressed are not necessarily those of DFID. Discussions with Richard Graham and Cath Senior have been much appreciated. The many modeling groups listed in Tables 1 and 2 of this paper are gratefully acknowledged for producing and making their simulations available, as is the World Climate Research Programme Working Group on Coupled Modelling (WCRP-WGCM), which takes responsibility for the CMIP3 and CMIP5 model archives, and the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison (PCMDI) for archiving the model output, providing coordinating support, and leading the development of software infrastructure with the Global Organization for Earth System Science Portals. Thanks are also due to Philip Pegion and Martin Hoerling for providing the NSIPP1 AGCM data and to David Fereday for running and providing the HadAM3 AGCM data.

APPENDIX

Expansions of Institution and Model Names

Table A1 contains the full definitions of the modeling groups and model names that are referenced in Tables 1 and 2 and the text.

TABLE A1. Expansions of the institution and model names involved with CMIP3 and CMIP5 that are used in this study.

Institution	Model name/version	Model expansion
Bjerknes Centre for Climate Research (BCCR)	BCCR-BCM2	BCCR Bergen Climate Model, version 2
Canadian Centre for Climate Modelling and Analysis (CCCma)	CGCM3.1(T47)	CCCma Coupled Global Climate Model, version 3.1, T47 resolution
	CanESM2	CCCma Second Generation Canadian Earth System Model
Center for Climate System Research (CCSR)/National Institute for Environmental Studies (NIES)/Japan Agency for Marine-Earth Science and Technology (JAMSTEC)	MIROC3.2(hires), MIROC3.2(medres)	Model for Interdisciplinary Research on Climate, version 3.2, with high and medium resolution, respectively
	MIROC5	Model for Interdisciplinary Research on Climate, version 5
	MIROC-ESM	Model for Interdisciplinary Research on Climate, Earth System Model
	MIROC-ESM-CHEM	Model for Interdisciplinary Research on Climate, Earth System Model, Chemistry Coupled
Centre National de Recherches Météorologiques (CNRM)	CNRM-CM3, CNRM-CM5	CNRM Coupled Global Climate Model, versions 3 and 5
Commonwealth Scientific and Industrial Research Organisation (CSIRO)	CSIRO Mk3.0, Mk3.5, Mk3.6.0	CSIRO Model, Marks 3.0, 3.5, 3.6.0
CSIRO–Bureau of Meteorology (BOM)	ACCESS1.0, ACCESS1.3	Australian Community Climate and Earth-System Simulator, versions 1.0 and 1.3
Institute of Numerical Mathematics (INM) Institute Pierre Simon Laplace (IPSL)	INM-CM3.0, INM-CM4	INM Coupled Model, versions 3.0 and 4
	IPSL-CM5A-LR	IPSL Coupled Model, version 5, coupled with Nucleus for European Modelling of the Ocean (NEMO), low resolution
	IPSL-CM5A-MR	IPSL Coupled Model, version 5, coupled with NEMO, midresolution
	IPSL-CM5B-LR	IPSL Coupled Model, version 5, coupled with NEMO, new atmospheric physics at low resolution
Istituto Nazionale di Geofisica e Vulcanologia (INGV)	INGV-SXG	INGV Scale Interaction Experiment (SINTEX)-(version) G model
Max Planck Institute (MPI)	MPI-ECHAM5-OM	MPI ECHAM5 Ocean Model
	MPI-ESM-LR	MPI Earth System Model, low resolution
	MPI-ESM-P	MPI Earth System Model, paleo
	ECHO-G	ECHAM and the global Hamburg Ocean Primitive Equation model
Meteorological Institute of the University of Bonn (MIUB) and the Korea Meteorological Administration (KMA)		
Meteorological Research Institute (MRI)	MRI-CGCM2.3.2, MRI-CGCM3	MRI Coupled General Circulation Model, versions 2.3.2 and 3
Met Office Hadley Centre (MOHC)	HadCM3	Hadley Centre Coupled Model, version 3
	HadGEM1	Hadley Centre Global Environmental Model, version 1
	HadGEM2-CC	Hadley Centre Global Environmental Model, version 2 (Carbon Cycle)
	HadGEM2-ES	Hadley Centre Global Environmental Model, version 2 (Earth System)

TABLE A1. (Continued)

Institution	Model name/version	Model expansion
National Aeronautics and Space Administration (NASA) Goddard Institute for Space Studies (GISS)	GISS-AOM	GISS Atmosphere–Ocean Model
	GISS-EH, GISS-ER	GISS Model E coupled with the Hybrid Coordinate (HYCOM) and Russell ocean models, respectively
National Center for Atmospheric Research (NCAR)	CCSM3, CCSM4	NCAR Community Climate System Model, versions 3 and 4
National Oceanic and Atmospheric Administration (NOAA)/Geophysical Fluid Dynamics Laboratory (GFDL)	PCM	NCAR Parallel Climate Model
Norwegian Climate Centre (NCC), Norway	GFDL CM2.0, CM2.1	GFDL Climate Model, versions 2.0 and 2.1
	NorESM1-M, NORESM1-ME	Norwegian Earth System Model, version 1 (medium resolution and medium resolution including active carbon cycle, respectively)
State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG) Institute of Atmospheric Physics (IAP)	FGOALS-g1.0, FGOALS-g2, FGOALS-s2	Flexible Global Ocean–Atmosphere–Land System Model, gridpoint versions 1.0 and 2 and second spectral version, respectively

REFERENCES

- Bartlett, M. S., 1935: Some aspects of the time correlation problem in regard to tests of significance. *J. Roy. Stat. Soc.*, **98**, 536–543.
- Chang, P., L. Ji, and H. Li, 1997: A decadal climate variation in the tropical Atlantic Ocean from thermodynamic air–sea interactions. *Nature*, **385**, 516–518.
- Cook, K. H., and E. K. Vizy, 2006: Coupled model simulations of the West African monsoon system: Twentieth- and twenty-first-century simulations. *J. Climate*, **19**, 3681–3703.
- Enfield, D. B., A. M. Mestas-Nunez, D. A. Mayer, and L. Cid-Serrano, 1999: How ubiquitous is the dipole relationship in tropical Atlantic sea surface temperatures? *J. Geophys. Res.*, **104** (C4), 7841–7848.
- Folland, C. K., J. Owen, M. N. Ward, and A. Colman, 1991: Prediction of seasonal rainfall in the Sahel region using empirical and dynamical methods. *J. Forecasting*, **10**, 21–56.
- Guilyardi, E., A. Wittenberg, A. Fedorov, M. Collins, C. Wang, A. Capotondi, and G. J. van Oldenborgh, 2009: Understanding El Niño in ocean–atmosphere general circulation models: Progress and challenges. *Bull. Amer. Meteor. Soc.*, **90**, 325–340.
- Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister, 2013: Updated high-resolution grids of monthly climatic observations—The CRU TS 3.1 dataset. *Int. J. Climatol.*, in press.
- Hastenrath, S., D. Polzin, and P. Camberlin, 2004: Exploring the predictability of the “short rains” at the coast of East Africa. *Int. J. Climatol.*, **24**, 1333–1354.
- Hoerling, M., J. Hurrell, J. Eischeid, and A. Phillips, 2006: Detection and attribution of twentieth-century northern and southern African rainfall change. *J. Climate*, **19**, 3989–4008.
- Janicot, S., V. Moron, and B. Fontaine, 1996: Sahel droughts and ENSO dynamics. *Geophys. Res. Lett.*, **23**, 515–518.
- , S. Trzaska, and I. Poccarr, 2001: Summer Sahel-ENSO teleconnection and decadal time scale SST variations. *Climate Dyn.*, **18**, 303–320.
- Johnson, M., K. Matsuura, C. Willmott, and P. Zimmermann, cited 2003: Tropical land-surface precipitation: Gridded monthly and annual time series (1950–1999). [Available online at http://climate.geog.udel.edu/~climate/html_pages/Tropics_files/README.tropic_precip_ts.html.]
- Joly, M., A. Voldoire, H. Douville, P. Terray, and J.-F. Royer, 2007: African monsoon teleconnections with tropical SSTs: Validation and evolution in a set of IPCC4 simulations. *Climate Dyn.*, **29**, 1–20.
- Jung, T., L. Ferranti, and A. M. Tompkins, 2006: Response to the summer of 2003 Mediterranean SST anomalies over Europe and Africa. *J. Climate*, **19**, 5439–5454.
- Lamb, P. J., 1978: Large-scale tropical Atlantic surface circulation patterns associated with sub-Saharan weather anomalies. *Tellus*, **30**, 240–251.
- Langenbrunner, B., and D. J. Neelin, 2013: Analyzing ENSO teleconnections in CMIP models as a measure of model fidelity in simulating precipitation. *J. Climate*, in press.
- Lu, J., and T. L. Delworth, 2005: Oceanic forcing of the late 20th century Sahel drought. *Geophys. Res. Lett.*, **32**, L22706, doi:10.1029/2005GL023316.
- Makarau, A., and M. R. Jury, 1997: Predictability of Zimbabwe summer rainfall. *Int. J. Climatol.*, **17**, 1421–1432.
- Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. E. Taylor, 2007a: The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bull. Amer. Meteor. Soc.*, **88**, 1383–1394.
- Navarra, A., and Coauthors, 2008: Atmospheric horizontal resolution affects tropical climate variability in coupled models. *J. Climate*, **21**, 730–750.
- Nicholson, S. E., D. Leposo, and J. Grist, 2001: The relationship between El Niño and drought over Botswana. *J. Climate*, **14**, 323–335.
- Ogallal, L. J., J. E. Janowiak, and M. S. Halpert, 1988: Teleconnection between seasonal rainfall over East Africa and global sea surface temperature anomalies. *J. Meteor. Soc. Japan*, **66**, 807–822.
- Patricola, C. M., and K. H. Cook, 2010: Northern African climate at the end of the twenty-first century: An integrated application of regional and global climate models. *Climate Dyn.*, **35**, 193–212.

- Pope, V. D., M. L. Gallani, P. R. Rowntree, and R. A. Stratton, 2000: The impact of new physical parametrizations in the Hadley Centre climate model—HadAM3. *Climate Dyn.*, **16**, 123–146.
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, doi:10.1029/2002JD002670.
- Roberts, M. J., and Coauthors, 2009: Impact of resolution on the tropical Pacific circulation in a matrix of coupled models. *J. Climate*, **22**, 2541–2556.
- Rowell, D. P., 2001: Teleconnections between the tropical Pacific and the Sahel. *Quart. J. Roy. Meteor. Soc.*, **127**, 1683–1706.
- , 2003: The impact of Mediterranean SSTs on the Sahelian rainfall season. *J. Climate*, **16**, 849–862.
- , 2012: Sources of uncertainty in future changes in local precipitation. *Climate Dyn.*, **39**, 1929–1950.
- , 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.
- Saji, N. H., B. N. Goswami, P. N. Vinayachandran, and T. Yamagata, 1999: A dipole mode in the tropical Indian Ocean. *Nature*, **401**, 360–363.
- Schubert, S. D., M. J. Suarez, P. J. Pegion, R. D. Koster, and J. T. Bacmeister, 2004: Causes of long-term drought in the U.S. Great Plains. *J. Climate*, **17**, 485–503.
- Shinoda, M., and R. Kawamura, 1994: Tropical rainbelt, circulation, and sea surface temperatures associated with the Sahelian rainfall trend. *J. Meteor. Soc. Japan*, **72**, 341–357.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498.
- Vizy, E. K., and K. H. Cook, 2001: Mechanisms by which Gulf of Guinea and eastern North Atlantic sea surface temperature anomalies can influence African rainfall. *J. Climate*, **14**, 795–821.
- , and —, 2002: Development and application of a mesoscale climate model for the tropics: Influence of sea surface temperature anomalies on the West African monsoon. *J. Geophys. Res.*, **107** (D3), doi:10.1029/2001JD000686.
- Wang, B., Q. Ding, X. Fu, I.-S. Kang, K. Jin, J. Shukla, and F. Doblas-Reyes, 2005: Fundamental challenge in simulation and prediction of summer monsoon rainfall. *Geophys. Res. Lett.*, **32**, L15711, doi:10.1029/2005GL022734.
- Wu, R., and B. P. Kirtman, 2007: Regimes of seasonal air-sea interaction and implications for performance of forced simulations. *Climate Dyn.*, **29**, 393–410.
- Yang, X., and T. DelSole, 2012: Systematic comparison of ENSO teleconnection patterns between models and observations. *J. Climate*, **25**, 425–446.