Regionalizing Africa: Patterns of Precipitation Variability in Observations and Global Climate Models

HAMADA S. BADR
Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, Maryland

AMIN K. DEZFULI
Climate and Radiation Laboratory, NASA Goddard Space Flight Center, Greenbelt, and Universities Space Research Association, Columbia, Maryland

BENJAMIN F. ZAITCHIK
Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, Maryland

CHRISTA D. PETERS-LIDARD
Earth Sciences Division, NASA Goddard Space Flight Center, Greenbelt, Maryland

(Manuscript received 1 March 2016, in final form 28 July 2016)

ABSTRACT

Many studies have documented dramatic climatic and environmental changes that have affected Africa over different time scales. These studies often raise questions regarding the spatial extent and regional connectivity of changes inferred from observations and proxies and/or derived from climate models. Objective regionalization offers a tool for addressing these questions. To demonstrate this potential, applications of hierarchical climate regionalizations of Africa using observations and GCM historical simulations and future projections are presented. First, Africa is regionalized based on interannual precipitation variability using Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) data for the period 1981–2014. A number of data processing techniques and clustering algorithms are tested to ensure a robust definition of climate regions. These regionalization results highlight the seasonal and even month-to-month specificity of regional climate associations across the continent, emphasizing the need to consider time of year as well as research question when defining a coherent region for climate analysis. CHIRPS regions are then compared to those of five GCMs for the historic period, with a focus on boreal summer. Results show that some GCMs capture the climatic coherence of the Sahel and associated teleconnections in a manner that is similar to observations, while other models break the Sahel into uncorrelated subregions or produce a Sahel-like region of variability that is spatially displaced from observations. Finally, shifts in climate regions under projected twenty-first-century climate change for different GCMs and emissions pathways are examined. A projected change is found in the coherence of the Sahel, in which the western and eastern Sahel become distinct regions with different teleconnections. This pattern is most pronounced in high-emissions scenarios.

1. Introduction

a. Climate regionalization

Climate regionalization divides a region into homogeneous subregions based on one or more climatic variables. It is a very important step in climate studies.
because it helps in identifying the drivers of climate variability specific to each region (e.g., Dezfuli and Nicholson 2013; Nicholson and Dezfuli 2013). Applying conventional geographic boundaries for climate studies is often problematic because climate conditions and sensitivities can vary widely within a single study area (e.g., a country or a river basin). Standard climate classification systems (e.g., Köppen climate classification) also have limitations: 1) they represent mean climatic conditions rather than temporal (e.g., interannual) variability, 2) they are not informative for identifying drivers of variability, and 3) as prescribed classification systems, rather than tools, they specify which variables and variable relationships are employed in the classification, rather than allowing the investigator to define characteristics of interest. Climate regionalization provides a useful alternative method for defining regions when we want to 1) unravel drivers of climate variability specific to different regions and seasons (e.g., Badr et al. 2014b) and explore potential changes in the future, 2) understand the spatial distribution of climate sensitivities, or 3) employ a flexible system to understand how spatial patterns of variability differ for different climate variables (e.g., air temperatures as opposed to precipitation). These features make climate regionalization particularly valuable for applications that rely on identifying areas of common variability in a parameter of interest; for example, managing a climatically diverse hydrologic unit in the context of climate variability or change, filling in data gaps in the historic climate record, or optimizing seasonal forecast systems.

Climate regionalization applies an objective single- or multivariate statistical technique such as clustering (e.g., Burn 1989; Dezfuli 2011; Gong and Richman 1995; Isik and Singh 2008; Rao and Srinivas 2006). Different clustering techniques have been applied in the literature (Djomou et al. 2015; Herrmann and Mohr 2011; Janicot 1992; Mahe et al. 2001; Nicholson 1986; Ogallo 1989), which are sensitive to conceptual approach, clustering algorithm, data processing, and validation criteria. It is difficult to compare the regionalization results of previous studies due to the lack of software tools that are technically designed for climate studies and meet the preprocessing and postprocessing requirements. This has motivated us to develop an open-source R package (HiClimR). Badr et al. (2015) describe the methodology and technical details of HiClimR.

The criteria used to interpret and validate climate regionalization may vary depending on the study objectives. Here, our regionalization criteria are to find homogeneous regions that are geographically contiguous, provided that the minimum region size is reasonable with respect to the nature of the problem (i.e., it is consistent with general size constraints such as landscape structure, data coverage and density, and known climate phenomena) and that the total number of regions matches the inherent physical properties of interest (e.g., the number of regions for identifying large-scale driver variability is different from dividing a country or area into regions of coherent climate variability for national development). The optimum regionalization will always have to involve some subjective decisions (e.g., contiguity checks and geographical characteristics) since the problem is a combination of applications and statistics. However, the objective criteria aim to maximize intraregional correlations (i.e., the average correlations between the region mean and all of its members) and minimize interregional correlations (i.e., the correlations between region means).

b. Variability of African precipitation

Africa is a continent of climate contrasts. The general spatial pattern of variability in mean annual precipitation is widely familiar: a humid equatorial zone that includes the Congo forest, transitional savannah zones as the tropics grade into the subtropics, tropical deserts in the northern (Sahara) and southern (Kalahari) portions of the continent, and midlatitude influences at northern and southern extremities. Temporal precipitation variability in portions of Africa can be large and is widely reported, in large part because of the social and economic impacts that hydroclimatic extremes have had on the Sahel, the Horn of Africa, and several other regions.

From a climate dynamics perspective, the spatial and temporal variability of precipitation present diverse challenges for process understanding, event prediction, and climate change projection, and the climate vulnerability of many communities across Africa makes the problem particularly urgent. In general, we understand variability to be a function of synoptic-scale to mesoscale atmospheric phenomena, including migration of the intertropical convergence zone (ITCZ; Barry and Chorley 2009), the strength and location of atmospheric jets (e.g., the African easterly jet and the tropical equatorial jet; Flohn 1964), monsoon circulations in West and East Africa, and significant land–atmosphere interactions (Dickinson 1995), particularly in semiarid zones. These phenomena, in turn, are influenced by the sea surface temperatures (SST) in neighboring oceans and by remote climate drivers that include El Niño–Southern Oscillation (ENSO), the South Asian monsoon, the Indian Ocean dipole, the Atlantic multidecadal oscillation (AMO) circulation, and many others. The relative influence of these dynamics and drivers varies in space and time, and even the sign of influence of important drivers
such as ENSO can flip from season to season or between subregions of a single country or river basin. This complexity can make it difficult to characterize drivers of variability in a systematic way, or to track spatial changes in their influence over time.

The purpose of this study is to identify regions within Africa that are coherent with respect to interannual precipitation variability. This exercise offers an example of how regionalization can be applied to characterize and study patterns of climate variability. It also provides a set of regionalization results that can be applied to future studies of African climate. The regionalization is performed using monthly precipitation estimates from observations and outputs from global climate models (GCMs). Importantly, although perhaps not surprisingly, we find that regionalization of Africa is a seasonally and even monthly specific problem. For this reason we define regions for each season separately, where the season is defined as a combination of months for which regions are spatially stable. We also find that regions differ when using different data sources (e.g., observations vs GCMs). This paper presents the results of these seasonal and dataset-specific regionalizations and explores implications for understanding drivers of interannual precipitation variability and for projecting climate change using different GCMs.

2. Data

The Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) dataset version 2.0 (Funk et al. 2015) was used to represent the observed precipitation in the period 1981–2014. The data are distributed on a 0.05° grid and a primary pentad temporal resolution with the availability of aggregates (decadal and monthly) or disaggregates (daily). In this study, we used CHIRPS data for Africa at the monthly temporal resolution. The data were resampled to 0.25° resolution due to the computational and memory requirements for regionalization of the entire continent of Africa. In comparison with the observational datasets available, CHIRPS data provide higher resolution, better station coverage over Africa, improved statistical approaches, and updated temporal coverage.

Extended Reconstructed Sea Surface Temperature (ERSST) version 4 (Huang et al. 2015; Liu et al. 2015) monthly data (2° × 2° grid resolution; 1854–present) were used to test the correlation patterns of each region’s mean time series with global SST.

The outputs from five different GCMs (CCSM4, CNRM, GFDL, HadGEM2, and MIROC5; expansions of acronyms available online at http://www.ametsoc.org/PubsAcronymList) were used to demonstrate the effectiveness of regionalization in the evaluation of GCMs in terms of capturing the spatial patterns of precipitation variability and the evolution of regions in response to greenhouse gas concentration as represented by coherent regions of the four representative concentration pathways (RCPs): RCP2.6, RCP4.5, RCP6, and RCP8.5 (Moss et al. 2010).

3. Methods

Climate regionalization was performed to divide Africa into smaller regions that are homogeneous with respect to interannual variability of precipitation for all months and 3-month running average seasons. The HiClimR package (Badr et al. 2015) was used. The most appropriate clustering method was used for each case based on three main criteria: homogeneity, separability, and contiguity. The homogeneity criterion minimizes the within-region variability and is measured by the average correlation between the region mean and its members (intraregional correlation). The separability criterion maximizes the differences between regions and is measured by the maximum correlation between the different means (interregional correlation). The contiguity criterion visually identifies the member of each region in geographical proximity not to have a region divided into distant subregions or members. The method that provides higher overall homogeneity, lower interregional correlations, and contiguous regions is referred as “better,” even if the differences are small. In general, the different methods will likely generate similar regions with slightly different statistics besides the visual contiguity of the regions.

We choose the method with better statistics and better contiguity. Specifically, results from two clustering methods are presented: Ward’s method (Murtagh 1983; Ward 1963), which minimizes the error sum of squares between all members within a region after merging, and regional linkage (Badr et al. 2015), which minimizes the interregional correlation between region means at each merging step. Ward’s method tends to generate well-proportioned regions with high homogeneity regardless of the interregional correlations, while regional linkage emphasizes separation of systematically dissimilar regions. An added advantage of regional linkage is that separating systematically dissimilar records isolates noise in the dataset (i.e., stations or very small clusters with completely different variability that cannot be merged into any of the regions and are not correlated with other stations or clusters). The noisy stations or clusters can be the result of bad data or may represent a phenomenon at a different scale (e.g., local effect). For regionalization at continental scale, both forms of noise
are undesirable as they are not representative of the broad regions that are being defined. The isolation of that noise can also help in the quality control of the data.

Several preprocessing options in HiClimR have been utilized to find the optimal regions and to test the sensitivity of regionalization. Geographic masking was used to mask all stations outside the continent of Africa. Grid cells with near-zero precipitation variability and/or very low mean precipitation are masked out to avoid any negative impacts on the quality of regionalization. The final regions are generated from detrended and standardized data to account only for the interannual precipitation variability without any possible effects of the linear trend or precipitation totals. The entire CHIRPS record available at the time of analysis (1981–2014) was used for all regionalizations.

For GCM regionalization we focus on boreal summer [July–September (JAS)]. We present only one season in order to make the results digestible, and we choose JAS because of its importance as the primary rainy season in the Sahel. Results are presented for the historical simulations of five GCMs (CCSM4, CNRM, GFDL, HadGEM2, and MIROC5; 1960–90). A unified time period of 30 years (1960–90) was used for all GCMs; this period is used as a baseline to compare the historical and future simulations of GCMs such as the reports of the Intergovernmental Panel on Climate Change (IPCC). The historical simulations include historic observations of greenhouse gases and other external forcings but are fully coupled to the ocean and are not initialized from observations; as such, climatological patterns and long-term trends are expected to match historical patterns but specific year-to-year and even decade-to-decade variability does not align with the observed climate record.

The possible effects of greenhouse gas concentration on spatial patterns of interannual precipitation variability are examined by performing regionalization for the four RCPs (RCP2.6, RCP4.5, RCP6, and RCP8.5) simulated by CCSM4 for different 30-yr periods within the 2006–2100 projection, and for the entire 2006–2100 time period. Additionally, to compare between the regions generated from observations (1981–2014) and CCSM4 using the same time period, we combined the CCSM4 historical simulation (which runs through 2005) with the RCP4.5 projection (2006 forward) to create a 1981–2014 CCSM4 record. The RCP4.5 emissions are reasonably consistent with observation for this period, and the impact of emissions trajectory on climate response on such a short time scale is small relative to internal variability. CHIRPS data were regridded to match the coarse model resolution.

4. Results and discussion

a. Data preprocessing

HiClimR implements several features to facilitate spatiotemporal analysis applications, including data filtering with geographic masking and/or mean–variance thresholds, data preprocessing via detrending, and standardization. These were applied as described below.

1) DATA FILTERING

Figure 1 shows the effect of masking options on regionalization results for the interannual variability of
precipitation over Africa in January using Ward’s clustering method, and Fig. 2 shows the associated clustering dendrograms for no masking (Fig. 1a), geographic masking of Africa (Fig. 1b), and masking with filtering to remove all noncontiguous subregions, including northern Africa (Fig. 1c). Filtering in this application included excluding the stations above 10°N, in order to focus on sub-Saharan Africa, and removing small spatially

FIG. 2. Clustering dendogram associated with Fig. 1: (a) no masking, (b) geographic masking of Africa, and (c) geographic masking and data filtering. The colors are unitless and represent regions identified in Fig. 1 to distinguish between the different clusters.
discontinuous regions, such as Ethiopia in the result shown in Fig. 1c. This filtering is additional to the automatic filtering techniques in HiClimR that remove stations with near-zero variance and/or very low mean defined by a mean threshold (the mean threshold is typically selected as a very small fraction of the average monthly total precipitation where reasonable changes in its value do not affect the regionalization results; for CHIRPS data, a mean threshold of 12 mm month$^{-1}$ was selected). The geographic masking improves the results because it excludes the artifacts introduced by Europe and small islands around Africa while filtering cleans up the regions and increases overall homogeneity.

Figure 2 clarifies the regionalization quality in each case. The vertical axis represents the clustering height. The units of this axis depend on the clustering method, but the metric generally needs to be minimized to achieve maximum intraregional homogeneity. For example, region 4, which is mainly in Europe, has an artifact subregion on the border between Angola and Namibia that disappears after geographic masking. For Ward’s clustering, the $y$ axis on the dendrogram is the sum of squared distances within all regions and is a measure of intraregional variance. This is different from the regional linkage method that minimizes the maximum interregional correlation between regions as a measure of region separation. To get the optimal number of regions for final regionalization, we need to minimize the interregional correlations (region separation) and maximize the intraregional correlation (region homogeneity). Hence, we add a horizontal axis on the left of the dendrogram plots in Fig. 2 to show the maximum interregional correlation at each dendrogram cut that defines the number of regions. It is clear that the overall clustering height decreases when applying the geographic masking from Figs. 2a to 2b and further with filtering the data from Figs. 2b to 2c, which is consistent with the regionalization qualities in Figs. 1a–c.

We emphasize that our selection of geographic extent and masking procedure is specific to the objectives of this study—to provide a stable and informative regionalization of Africa at continental scale. Regionalization could just as easily be applied to a global domain, including land and ocean, in order to study global scale response to major modes of climate variability, for example. This would change the regionalization of Africa but might yield other insights on climate variability.

2) DETRENDING AND STANDARDIZATION

Figure 3 shows the effects of data detrending and standardization on regionalization results for the interannual variability of precipitation over Africa in January using Ward’s clustering method. Figure 4 shows the associated clustering dendrograms. We found that standardization had no visible effect on regionalization results in this month (not shown; this may be different when using a different dataset or at another temporal scale, when there is large spatial
variability in precipitation magnitudes), while de-trending affects regions that have strong similarities or differences in linear trends. In this application, de-trending tends to shift the borders between regions without fundamentally altering the character of the map (Fig. 3). But these shifts are systematic: for example, regions 3 and 4 share a positive linear trend that tends to increase the correlation between regions if detrending is not applied. These regions also show the highest interregional correlation in the analysis, so the maximum interregional correlation figures shown on the left side of Fig. 4 reflect higher, trend-influenced correlations between these regions in the raw data (0.58) and a lower correlation when data are detrended (0.48). Note that the region merging in Fig. 4 is based on Ward’s method, which minimizes the variance within regions.

b. Dissimilarity measures

The dissimilarity measure—or the nature of the temporal dimension used to regionalize spatial data in HiClimR—is a crucial decision, and the choice depends on the specific application. For example, regions can be generated based on interannual, intraseasonal, or daily variability or on seasonal cycle. Figure 5 shows 12 regions generated for precipitation over Africa using Ward’s clustering method based on the interannual variability (annual mean; Fig. 5a) and seasonality (annual cycle; Fig. 5b). It is clear that the differences in seasonality do not...
c. Monthly specific regions

Figure 6 shows results of regionalization on interannual precipitation variability performed separately for each month of the year. This is an optimized version of regionalization that applies masking, filtering, standardization, and detrending, as described above. These options were selected for this application based on the sensitivity analysis described in section 4a. Ward’s method provided the best regionalization results in winter months (December–March) and very similar regions in May. However, it was sensitive to region size, which sometimes results in dividing a large but highly homogeneous region into two or more regions. In contrast, the regional linkage method is able to identify large homogenous regions like the Sahel in summer and to filter out noisy data. In winter months–seasons, the regions have relatively similar size and Ward’s method performs well: even though the method optimizes for intraregional homogeneity rather than interregional separability, the correlation between regions was reasonably low, indicating that the separation of regions was meaningful. In summer and most transitional months, however, regional linkage provides more climatically meaningful regions, while Ward’s method tends to split regions based on size, producing multiple regions that have high interregional correlation. In May, both methods generated very similar regions (not shown) and Ward’s method was selected for the slightly better overall homogeneity. Note that the numbering of regions is arbitrary; the regionalization process simply distinguishes between regions based on dissimilarity measure, and the analyst must interpret association of regions across datasets or regionalization methods. It is clear that the spatial patterns of precipitation variability over Africa (as represented by homogeneous regions in Fig. 6) are specific to calendar month. This can be used as a guideline for researchers studying climate processes, as one would not want to average across months with significantly different regionalization patterns in the area of interest, just as one would not want to average across two poorly correlated regions in any given season. For example, for a large-scale study these results suggest that January and February can be treated as a coherent season, with only small changes in region boundaries in central Africa. Moving back to December or forward to March, however, regions in southern Africa begin to split and shift relative to January–February in ways that merit attention before attempting to generalize across December–February or January–March in those areas. This could have implications for forecasting.

Table 1 lists the average intraregional correlations (between region mean and all members) for all regions.
and the maximum interregional between region means. The winter months (December–March) and May utilized Ward’s method and their rows are highlighted in bold font. All other months, especially the summer months (JAS) and months of complicated variability in transition between seasons, utilize the regional linkage method and its ability to isolate noisy areas for removal. In other applications, where optimal quality control of the data is desired, statistically isolated regions (or weather stations) could be treated in a statistical or dynamical analysis to understand reasons for isolation and fill their gaps. The values of intraregional correlations in Table 1 are affected by the very high resolution and continental scale (using coarser data would increase the overall homogeneity of the regions since larger grid cells are smoothed averages of the included finer grid cells), which indicates that each of the regions can be divided into smaller regions for finer-scale applications such as hydrological analysis over one country or a smaller region of interest. However, the current results target the association of subcontinental regions with large-scale drivers of variability (teleconnections). The maximum interregional correlation indicates the correlation between the most similar regions in the regionalizations. All other regions have smaller or negative correlations between means.

d. Historical versus future (GCMs)

Figure 7 shows the regions of Africa based on interannual variability of summer (JAS) precipitation using observations from CHIRPS (OBS; 1981–2014) and different GCM historical simulations (CCM4, CNRM, GFDL, HadGEM2, and MIROC5; 1960–90). A unified time period of 30 yr (1960–90) was used for all GCMs. To avoid regionalization sensitivity based on clustering approach, all models are treated similarly regarding the number of regions (cutoff level of the dendrogram), data thresholds (variance and mean

![Fig. 6. Regionalization of Africa based on interannual variability of monthly precipitation.](image-url)
thresholds for data filtering), and preprocessing options (e.g., detrending and standardization). The regional linkage method is appropriate for summer precipitation when using CHIRPS observations, the outputs from the two models that exhibit similar spatial variability to observations (CCSM4 and MIROC5), and CNRM. In contrast, Ward’s method was relatively better for GFDL and HadGEM2 as it provides slightly better overall

<table>
<thead>
<tr>
<th>MM</th>
<th>$R_{01}$</th>
<th>$R_{02}$</th>
<th>$R_{03}$</th>
<th>$R_{04}$</th>
<th>$R_{05}$</th>
<th>$R_{06}$</th>
<th>$R_{07}$</th>
<th>$R_{08}$</th>
<th>$R_{09}$</th>
<th>$R_{10}$</th>
<th>$R_{11}$</th>
<th>$R_{12}$</th>
<th>$R_{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>0.71</td>
<td>0.49</td>
<td>0.66</td>
<td>0.61</td>
<td>0.61</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>0.48</td>
</tr>
<tr>
<td>02</td>
<td>0.72</td>
<td>0.51</td>
<td>0.68</td>
<td>0.64</td>
<td>0.53</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>0.43</td>
</tr>
<tr>
<td>03</td>
<td>0.59</td>
<td>0.49</td>
<td>0.74</td>
<td>0.60</td>
<td>0.63</td>
<td>0.62</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>0.42</td>
</tr>
<tr>
<td>04</td>
<td>0.57</td>
<td>0.72</td>
<td>0.66</td>
<td>0.67</td>
<td>0.73</td>
<td>0.87</td>
<td>0.70</td>
<td>0.74</td>
<td>0.57</td>
<td>0.72</td>
<td>0.67</td>
<td>0.73</td>
<td>0.49</td>
</tr>
<tr>
<td>05</td>
<td>0.54</td>
<td>0.67</td>
<td>0.46</td>
<td>0.64</td>
<td>0.63</td>
<td>0.57</td>
<td>0.69</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>0.49</td>
</tr>
<tr>
<td>06</td>
<td>0.54</td>
<td>0.46</td>
<td>0.50</td>
<td>0.57</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>0.35</td>
</tr>
<tr>
<td>07</td>
<td>0.60</td>
<td>0.58</td>
<td>0.54</td>
<td>0.66</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>0.39</td>
</tr>
<tr>
<td>08</td>
<td>0.58</td>
<td>0.61</td>
<td>0.62</td>
<td>0.60</td>
<td>0.59</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>0.42</td>
</tr>
<tr>
<td>09</td>
<td>0.52</td>
<td>0.56</td>
<td>0.54</td>
<td>0.59</td>
<td>0.61</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>0.36</td>
</tr>
<tr>
<td>10</td>
<td>0.69</td>
<td>0.67</td>
<td>0.64</td>
<td>0.68</td>
<td>0.56</td>
<td>0.82</td>
<td>0.72</td>
<td>0.64</td>
<td>0.73</td>
<td>0.70</td>
<td>0.80</td>
<td>----</td>
<td>0.51</td>
</tr>
<tr>
<td>11</td>
<td>0.54</td>
<td>0.68</td>
<td>0.70</td>
<td>0.60</td>
<td>0.63</td>
<td>0.67</td>
<td>0.68</td>
<td>0.70</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>0.44</td>
</tr>
<tr>
<td>12</td>
<td>0.58</td>
<td>0.47</td>
<td>0.73</td>
<td>0.57</td>
<td>0.67</td>
<td>0.68</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>0.33</td>
</tr>
</tbody>
</table>

**TABLE 1.** Intraregional and maximum interregional correlations for each month: $R_{nn}$ is the average intraregional correlation for region number $nn$ between the region mean and all members within the region, and $R_{\text{max}}$ is the maximum interregional correlation between region means. Boldface values indicate using Ward’s clustering rather than the regional linkage method.

**FIG. 7.** Regionalization of Africa based on interannual variability of summer (JAS) precipitation using CHIRPS observations (1981–2014) and different GCMs (1960–90).
homogeneity and contiguity, and none of the clustering methods or preprocessing options had a significant impact on the dominant patterns. The observed spatial patterns show the Sahel as one homogeneous region with strong agreement between the regions generated from observations: the Sahel region is homogeneous (intraregional correlation of \(-0.65\)) and independent from the other three regions. It is found that CCSM4 and MIROC5 have good skill in capturing the coherence of the Sahel precipitation signal in summer (strong signal of unique variability over the big green region that represents the Sahel, which cannot be divided into smaller subregions or the mean time series of subregions will be strongly correlated), while the other models miss this coherence and divide the Sahel into smaller regions with dissimilar interannual variability. CNRM generates random spatial patterns, GFDL simulates precipitation in summer shifted in the northwestern direction, and HadGEM2 divides the Sahel region into eastern and western subregions. The observational analysis was repeated using CRU in place of CHIRPS; results were similar and are not shown.

The robustness of Sahel region in summer (JAS) was examined using regionalization of different observational data sources (CHIRPS, CRU, and GPCC) and a variety of homogeneity checks such as correlation patterns between the region mean and precipitation over Africa. All data sources yield similar spatial patterns and identify the Sahel as a single coherent region. Increasing the number of clusters or changing the clustering algorithm divides the Sahel into two or more “very similar” regions (i.e., the interregional correlations between the subregions are high, meaning that we need to merge them into one region).

Figure 8 shows the regionalizations of Africa based on interannual variability of JAS precipitation in 1960–90 for six different ensemble members of CCSM4. All ensemble members identify the Sahel and East Africa as homogeneous regions, consistent with the observations. The known West African dipole mode is also detected in all cases. The differences primarily appear in western equatorial Africa, which has a strong intrinsic heterogeneity with respect to interannual variability (Balas et al. 2007; Dezfuli and Nicholson 2013).
To compare between the regions generated from observations (1981–2014) and CCSM4 using the same time period, we combine the CCSM4 historical simulation (which runs through 2005) with the RCP4.5 projection (2006 forward) to create a 1981–2014 CCSM4 record. The RCP4.5 emissions are reasonably consistent with observation for this period, and the impact of emissions trajectory on climate response on such a short time scale is small relative to internal variability. CHIRPS data were regridded to match the coarse model resolution. It is found the spatial patterns of interannual precipitation variability captured by CCSM4 in 1981–2014 (Fig. 9), especially over the Sahel, are very similar to the unified period (1960–90) in Fig. 7. In the next sections, we focus on CCSM4 for the GCM and use the unified period to facilitate comparison with CHIRPS.

Figure 10 shows the correlation patterns of the four regions generated based on the interannual variability of JAS precipitation using CHIRPS observations with global ERSST for the mean time series of the region, and Fig. 11 shows the correlations patterns from CCSM4 with its own SST (both 1981–2014). Similarities and differences between observed regions and CCSM4 regions are strongly related to SST teleconnections. Region 3 of CCSM4, which is the largest difference between the observational and GCM-based regionalization, shows a positive correlation with the SSTs over the ENSO region. However, such an association does not exist for any of the observation-based regions; it appears to be a false ENSO teleconnection that exists only in the model. For region 1, which corresponds to the Sahel in both model and simulations, the correlation with North Atlantic SST (i.e., the AMO) is markedly different and even slightly reversed for CCSM4 (Fig. 11) relative to observations (Fig. 10). In addition, the model shows significant region 1 correlations with Gulf of Guinea SSTs that are not seen in observations. Together these differences suggest that Sahel sensitivity to Atlantic Ocean variability is substantially different in CCSM4 than is in observation. Region 2 of CCSM4 in Fig. 11, along the Guinean coast, shows weaker correlation with SSTs of adjacent waters than the observations. These agreements or disagreements between the regions of the observed and simulated precipitation raise many interesting questions about the underlying mechanisms and skills of the model dynamics and need further investigation.

The regions do not respond to unique climate drivers, but they have different variability in response to a set of large-scale and local drivers. Regionalization helps in identifying distinct regions with coherent variability that differ from each other. The climate drivers of a specific region can then be identified, which are greatly improved over the commonly used geographic boxes that could mix inhomogeneous regions and create unrealistic variability and associated drivers (e.g., the mean of an area defined by a box without performing regionalization to test its homogeneity may
include different, or perhaps opposite, variability that can ruin the analysis).

e. Climate projections

Figure 12 shows the possible effect of greenhouse gas concentrations on spatial patterns of interannual precipitation variability as represented by coherent regions simulated by CCSM4 model for the period 2006–2100 for each of the four RCPs (RCP2.6, RCP4.5, RCP6, and RCP8.5). In these simulations the Sahel falls into two regions: regions 1 and 2 in Fig. 12. As radiative forcing increases (RCP2.6 through RCP8.5), region 2 shrinks and region 1 grows: essentially, there is a shift from a coherent Sahel band that stretches

---

**Fig. 10.** Correlation patterns of CHIRPS precipitation with global ERSST for the mean time series of the four regions of interannual variability of JAS precipitation for the period 1981–2014. All correlations are significant at 90% confidence level.

**Fig. 11.** As in Fig. 10, but of CCSM4 precipitation and the period 1960–90.
almost across the continent (region 2 in RCP2.6) to two distinct eastern and western Sahel regions in higher-emissions scenarios. The changes in equatorial West Africa and East Africa (regions 3 and 4) are small. We note that these 2006–2100 regionalization results average across variability in regions that is observed in shorter periods of analysis. The evolution of regions for the four climate projections examined at three different 30-yr simulation periods (2010–40, 2040–70, and 2070–2100) suggest that the spatial patterns vary throughout the twenty-first century, perhaps as a response to changes in global SST–rainfall relationships, but that there is a gradual trend toward a split between the eastern and western Sahel (not shown).

Figures 13 and 14 respectively show the correlation patterns of CCSM4 JAS region 1 (western Sahel) and region 2 (eastern Sahel) precipitation from all RCPs with the corresponding SST field for the entire simulation period (2006–2100). Teleconnections in low-emissions scenarios (RCP2.6 and RCP4.5) are, as expected, more similar to Sahel teleconnections in historical simulations. This is particularly evident for the eastern Sahel (region 2; Fig. 14), which approximately corresponds to the unified Sahel in historical simulations (Fig. 9). For all emissions scenarios, region 1 shows weaker correlation with global SST patterns than region 2 does, but the strength of these connections increases as the region expands under higher emissions. Region 2 shows strong correlations with the ENSO region and with the Indian Ocean under all RCPs, although the relationship is strongest for low emissions.
Interestingly, the influence of tropical South Atlantic and Gulf of Guinea SSTs on both regions changes as a function of emissions, shifting from a positive correlation at low emissions to a neutral or negative correlation at high emissions. This suggests changing dynamical interactions between remote forcings like ENSO and the more local influence of SSTs in the neighboring Atlantic Ocean. It is also noteworthy that for region 2 there is an Atlantic meridional mode (AMM)-type signal in the historical simulations and RCP2.6 and RCP4.5, but this signal disappears in RCP8.5, where the cross-equatorial SST gradient is absent. This again points to changing relationships between the tropical Atlantic Ocean and Sahel precipitation. We do note that all results shown here are for a single GCM ensemble member, which is expected to impact results for shorter
time scales and might also impact the details of these 2006–2100 results.

5. Conclusions

The different clustering methods available in HiClimR can be useful for different regionalization problems. Ward’s method is sensitive to region size and tends to divide a large homogeneous region into multiple separate regions. In contrast, the regional linkage method can identify a large coherent region and filter out noisy data. For example, Ward’s method was effective in winter months, when regions have relatively similar size, while regional linkage yielded cleanly separated and contiguous homogeneous regions in summer, when the interannually variability of Sahel precipitation is dominant. Regional linkage also provides more climatically relevant results in the transition seasons.

In the historical observational record, the Sahel expands from west to east and dominates the interannual variability of African precipitation in summer. This is confirmed with the spatial correlation patterns of the region mean and precipitation over Africa using CHIRPS data. The overall homogeneity of the region is only moderately high, but it is spread across the region with highly significant correlation between any potentially identifiable subregions. This suggests that interannual variability is characterized by one dominant mode of variability for the entire region plus local modes of variability in the subregions. The winter months (December–March) are relatively stable in their regionalization, and with a few exceptions can be treated as a coherent season for climate analysis. As expected, the transitional months have complicated spatial variability.

We tested the potential of climate regionalization for model intercomparison and assessment. CCSM4 and MIROC5 showed good skill in capturing the precipitation signal over the Sahel in summer, while the other models miss the spatial patterns by dividing the Sahel into smaller regions with dissimilar interannual variability. CNRM generates random spatial patterns, GFDL simulates Sahel-like precipitation variability in summer shifted to the northwest, and HadGEM2 divides the Sahel region into eastern and western sub-regions. This does not mean that models such as CNRM, GFDL, or HadGEM2 are not useful for projecting climate change in Africa. On the contrary, the regionalization analysis presented in this paper might allow us to interpret these simulations properly for studies of climate process and projections of climate change impacts. If GFDL systematically shifts the Sahel climate pattern to the northwest, then analysis of Sahel sensitivity in this model should be performed with a corresponding shift. This opens the possibility of a more meaningful, regionalization-based multimodel ensemble of climate projections for the Sahel—or for any region in which models differ in the spatial representation of variability. In the long term, these studies can also inform model development to correct spatial biases and displacements in the representation of climate variability.

Acknowledgments. We acknowledge the World Climate Research Programme’s Working Group on Coupled Modeling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. For CMIP the U.S. Department of Energy’s Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. Work for this paper was supported in part by NASA Applied Sciences Grant 13-WATER13-0010 and NSF’s Dynamics of Coupled Natural and Human Systems (CNH) Program Award GEO-1211235.

REFERENCES


Djomou, Z. Y., D. Monkam, and R. Chamani, 2015: Characterization of climatic zones, variability and trend in


