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

Hamada S. Badr *

Department of Earth and Planetary Sciences (EPS)
Johns Hopkins University (JHU)
3400 N. Charles Street, Olin Hall, Baltimore, MD, 21218, USA
badr@jhu.edu

Amin K. Dezfuli

Climate and Radiation Laboratory (Code 613)
NASA Goddard Space Flight Center (NASA/GSFC)
8800 Greenbelt Road, Greenbelt, MD 20771, USA
and Universities Space Research Association
amin.dezfuli@nasa.gov

Benjamin F. Zaitchik

Department of Earth and Planetary Sciences (EPS)
Johns Hopkins University (JHU)
3400 N. Charles Street, Olin Hall, Baltimore, MD, 21218, USA
zaitchik@jhu.edu

Christa D. Peters-Lidard

Earth Sciences Division (Code 610)
NASA Goddard Space Flight Center (NASA/GSFC)
8800 Greenbelt Road, Greenbelt, MD 20771, USA
christa.d.peters-lidard@nasa.gov

* Corresponding author at: JHU, 3400 N. Charles Street, Olin Hall, Baltimore, MD, 21218, USA.
E-mail address: badr@jhu.edu (Hamada S. Badr).

Key Words: Climate Regionalization, Spatial Analysis, Spatio-temporal, Variability, Africa, Precipitation, Global Climate Models, Climate Change, GCMs, RCPs, GHGs

ABSTRACT

Many studies have documented dramatic climatic and environmental changes that have affected Africa over different timescales. These studies often raise questions regarding the spatial extent and regional connectivity of changes inferred from observations, proxies, and/or derived from climate models. Objective regionalization offers a tool for addressing these questions. To demonstrate this potential, we present applications of hierarchical climate regionalizations of Africa using observations and GCM historical simulations and future projections. First, we regionalize Africa based on interannual precipitation variability using 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, we examine shifts in climate regions under projected 21st century climate change for different GCMs and emissions pathways. We find a projected change 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

1.1 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, a river basin). Standard climate classification systems (e.g., Köppen Climate Classification) also have limitations: (i) they represent mean climatic conditions rather than temporal variability (e.g., interannual), (ii) they are not informative for identifying drivers of variability, and (iii) 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: (i) unravel drivers of climate variability specific to different regions and seasons (e.g., Badr et al. 2014a), and explore potential changes in the future, (ii) understand the spatial distribution of climate sensitivities, or (iii) 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 multi-variate statistical technique such as clustering (e.g., Burn 1989; Dezfuli 2011; Gong and Richman 1995; Isik and Singh 2008; Ramachandra 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 (Badr et al. 2014b) for Hierarchical Climate Regionalization (called, "HiClimR"). Badr et al. (2015) describes 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 intra-regional correlations (i.e., the average correlations between the region mean and all of its members) and minimize inter-regional correlations (i.e., the correlations between region means).

1.2 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, subtropical deserts in the north (Sahara) and southern (Kalahari) portions of the continent, and mid-latitude 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 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 semi-arid zones. These phenomena, in turn, are influenced by the sea surface temperatures (SST) in neighboring oceans and by remote climate drivers that include the 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 like 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, though 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 versus GCM). This paper presents the results of these seasonally 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 Station data (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 (dekadal 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 provides higher resolution, better station coverage over Africa, improved statistical approaches, and updated temporal coverage.

Extended Reconstructed Sea Surface Temperature (ERSST) v4 (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 timeseries with global SST.

The outputs from five different GCMs (CCSM4, CNRM, GFDL, HadGEM2, and MIROC5) 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): 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 Hierarchical Climate Regionalization (HiClimR) R 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 (intra-regional correlation). The separability criterion maximizes the differences between regions and is measured by the maximum correlation between the different means (inter-regional 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 inter-regional 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 Jr 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 inter-regional correlation between region means at each merging step. Ward’s method tends to generate well-proportioned regions with high homogeneity regardless the inter-regional 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; 1960-1990). A unified time period of 30 years (1960-1990) 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 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 Representative Concentration Pathways (RCPs: RCP2.6, RCP4.5, RCP6, and RCP8.5) simulated by CCSM4 model for different 30-year 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 RCP 4.5 projection (2006-forward) to create a 1981-2014 CCSM4 record. The RCP 4.5 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 & Discussion

4.1 Data Preprocessing

HiClimR implements several features to facilitate spatio-temporal 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.

4.1.1 Data Filtering

Fig. 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 10N, in order to focus on sub-Saharan Africa, and removing small spatially 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 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.

Fig. 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 intra-regional homogeneity. For example, Region 4, which is mainly in Europe, has an artifact subregion in 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 intra-regional variance. This is different from the regional-linkage method that minimizes the maximum inter-regional correlation between regions as a measure of region separation. To get the "optimal" number of regions for final regionalization, we need to minimize the inter-regional correlations (region separation) and maximize the intra-regional correlation (region homogeneity). Hence, we add a horizontal axis on the left of the dendrogram plots in Fig. 2 to show the maximum inter-regional 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 Fig. 2A to Fig. 2B and further with filtering the data from Fig. 2B to Fig. 2C, which is consistent with the regionalization qualities in Fig. 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.

4.1.2 Detrending and Standardization

Fig. 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. Fig. 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 data set or at another temporal scale, when there is large spatial variability in precipitation magnitudes), while detrending affects regions that have strong similarities or differences in linear trends. In this application, detrending 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 inter-regional correlation in the analysis, so the maximum inter-regional 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.

4.2 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. Fig. 5 shows 12 regions generated for precipitation over Africa using Ward's clustering method based on the interannual variability (Annual Mean; left) and seasonality (Annual Cycle; right). It is clear that the differences in seasonality don't always align with differences in interannual variability: regions based on seasonality tend to align zonally, following the seasonal migration of the intertropical convergence zone within tropical Africa, while those based on the annual mean show both zonal and meridional structure, reflecting differing influences of remote climate drivers.

4.3 Monthly-Specific Regions

Fig. 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 4.1. 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 big 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 intra-regional homogeneity rather than inter-regional separability, the correlation between regions was reasonably low, indicating the separation of regions was meaningful. In summer and most transitional months, however, regional linkage provides more climatically meaningful regions, while Ward's tends to split regions based on size, producing multiple regions that have high inter-regional 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 DJF or JFM in those areas. This could have implications for forecasting.

Table 1 lists the average intra-regional correlations (between region mean and all members) for all regions and the maximum inter-regional 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 (July-September) 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 intra-regional 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 sub-continental regions with large-scale drivers of variability (teleconnections). The maximum inter-regional correlation indicates the correlation between the *most similar* regions in the regionalizations. All other regions have smaller or negative correlations between means.

4.4 Historical vs Future (GCMs)

Fig. 7 shows the regions of Africa based on interannual variability of summer (July-September, JAS) precipitation using observations from CHIRPS (OBS; 1981-2014) and different GCM historical simulations (CCM4, CNRM, GFDL, HadGEM2, and MIROC5; 1960-1990). A unified time period of 30 years (1960-1990) was used for all GCMs. To avoid regionalization sensitivity based on clustering approach, all models are treated similarly regarding the number of regions (cut-off level of the dendrogram), data thresholds (variance and mean 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 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 (intra-regional correlation = ~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 timeseries 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 Sahel into two or more “very similar” regions (i.e., the inter-regional correlations between the subregions are high meaning that we need to merge them into one region).

Fig. 8 shows the regionalizations of Africa based on interannual variability of JAS precipitation in 1960-1990 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 West 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 RCP 4.5 projection (2006-forward) to create a 1981-2014 CCSM4 record. The RCP 4.5 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 (1960-1990) in Fig. 7. In the next sections, we focus on CCSM4 for the GCM and use the unified period to facilitate comparison with CHIRPS.

Fig. 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 timeseries 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., Atlantic Multidecadal Oscillation or 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 it 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 regions with coherent variability that are different from each other's. 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).

4.5 Climate Projections

Fig. 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 Representative Concentration Pathways (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 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 30-years different simulation periods (2010-2040, 2040-2070, and 2070-2100) suggest that the spatial patterns vary throughout the 21st century, perhaps as a response to changes in global SST-rainfall relationships, but that there is a gradual trend towards a split between the eastern and western Sahel (not shown).

Fig. 13 and Fig. 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, 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, though 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 RCPs 2.6 and 4.5, but this signal disappears in RCP 8.5, 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 big 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 interannual 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 intercomparision 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 subregions. This does not mean that models like 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 multi-model 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.

Acknowledgements

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

Badr, H. S., B. F. Zaitchik, and S. D. Guikema, 2014a: Application of Statistical Models to the Prediction of Seasonal Rainfall Anomalies over the Sahel. *Journal of Applied Meteorology and Climatology*, **53,** 614-636.

Badr, H. S., B. F. Zaitchik, and A. K. Dezfuli, 2014b: HiClimR: Hierarchical Climate Regionalization. *Comprehensive R Archive Network (CRAN)***,** [Available online at http://cran.r-project.org/package=HiClimR].

——, 2015: A tool for hierarchical climate regionalization. *Earth Science Informatics***,** 1-10.

Balas, N., S. Nicholson, and D. Klotter, 2007: The relationship of rainfall variability in West Central Africa to sea-surface temperature fluctuations. *International journal of climatology*, **27,** 1335-1349.

Barry, R. G., and R. J. Chorley, 2009: *Atmosphere, weather and climate.* Routledge.

Burn, D. H., 1989: Cluster analysis as applied to regional flood frequency. *Journal of Water Resources Planning and Management*, **115,** 567-582.

Dezfuli, A. K., 2011: Spatio-temporal variability of seasonal rainfall in western equatorial Africa. *Theoretical and applied climatology*, **104,** 57-69.

Dezfuli, A. K., and S. E. Nicholson, 2013: The Relationship of Rainfall Variability in Western Equatorial Africa to the Tropical Oceans and Atmospheric Circulation. Part II: The Boreal Autumn. *Journal of Climate*, **26**.

Dickinson, R. E., 1995: Land-atmosphere interaction. *Reviews of Geophysics*, **33,** 917-922.

Djomou, Z. Y., D. Monkam, and R. Chamani, 2015: Characterization of climatic zones, variability and trend in northern Africa. *Climate Dynamics*, **44,** 3481-3491.

Flohn, H., 1964: *Investigations on the tropical easterly jet.* Dümmlers Vlg.

Funk, C., and Coauthors, 2015: The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific data*, **2**.

Gong, X., and M. B. Richman, 1995: On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. *Journal of Climate*, **8,** 897-931.

Herrmann, S. M., and K. I. Mohr, 2011: A continental-scale classification of rainfall seasonality regimes in Africa based on gridded precipitation and land surface temperature products. *Journal of Applied Meteorology and Climatology*, **50,** 2504-2513.

Huang, B., and Coauthors, 2015: Extended Reconstructed Sea Surface Temperature Version 4 (ERSST. v4). Part I: Upgrades and Intercomparisons. *Journal of Climate*, **28,** 911-930.

Isik, S., and V. P. Singh, 2008: Hydrologic regionalization of watersheds in Turkey. *Journal Of Hydrologic Engineering*, **13,** 824-834.

Janicot, S., 1992: Spatiotemporal variability of West African rainfall. Part I: Regionalizations and typings. *Journal of Climate*, **5,** 489-497.

Liu, W., and Coauthors, 2015: Extended Reconstructed Sea Surface Temperature Version 4 (ERSST. v4): Part II. Parametric and Structural Uncertainty Estimations. *Journal of Climate*, **28,** 931-951.

Mahe, G., Y. L'hote, J. C. Olivry, and G. Wotling, 2001: Trends and discontinuities in regional rainfall of West and Central Africa: 1951–1989. *Hydrological Sciences Journal*, **46,** 211-226.

Moss, R. H., and Coauthors, 2010: The next generation of scenarios for climate change research and assessment. *Nature*, **463,** 747-756.

Murtagh, F., 1983: A survey of recent advances in hierarchical clustering algorithms. *The Computer Journal*, **26,** 354-359.

Nicholson, S. E., 1986: The spatial coherence of African rainfall anomalies: interhemispheric teleconnections. *Journal of climate and applied meteorology*, **25,** 1365-1381.

Nicholson, S. E., and A. K. Dezfuli, 2013: The Relationship of Rainfall Variability in Western Equatorial Africa to the Tropical Oceans and Atmospheric Circulation. Part I: The Boreal Spring. *Journal of Climate*, **26**.

Ogallo, L., 1989: The spatial and temporal patterns of the East African seasonal rainfall derived from principal component analysis. *International Journal of Climatology*, **9,** 145-167.

Ramachandra Rao, A., and V. Srinivas, 2006: Regionalization of watersheds by hybrid-cluster analysis. *Journal of Hydrology*, **318,** 37-56.

Ward Jr, J. H., 1963: Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, **58,** 236-244.

Tables

Table 1 Intra-regional and maximum inter-regional correlations for each month.

MM	Intra-regional Correlations (R_{nn}[1])												R_{mx}[2]
	R_{01}	R_{02}	R_{03}	R_{04}	R_{05}	R_{06}	R_{07}	R_{08}	R_{09}	R_{10}	R_{11}	R_{12}	
01	**0.71**	**0.49**	**0.66**	**0.61**	**0.61**								**0.48**
02	**0.72**	**0.51**	**0.68**	**0.64**	**0.53**								**0.43**
03	**0.59**	**0.49**	**0.74**	**0.60**	**0.63**	**0.62**							**0.42**
04	0.57	0.72	0.66	0.67	0.73	0.87	0.70	0.74	0.57	0.72	0.67	0.73	0.49
05	**0.54**	**0.67**	**0.46**	**0.64**	**0.63**	**0.57**	**0.69**						**0.49**
06	0.54	0.46	0.50	0.57									0.35
07	0.60	0.58	0.54	0.66									0.39
08	0.58	0.61	0.62	0.60	0.59								0.42
09	0.52	0.56	0.54	0.59	0.61								0.36
10	0.69	0.67	0.64	0.68	0.56	0.82	0.72	0.64	0.73	0.70	0.80		0.51
11	0.54	0.68	0.70	0.60	0.63	0.67	0.68	0.70					0.44
12	**0.58**	**0.47**	**0.73**	**0.57**	**0.67**	**0.68**							**0.33**

[1] *R_{nn} is the average intra-regional correlation for region number nn, between the region mean and all members within the region.*
[2] *R_{mx} is the maximum inter-regional correlation between region means.*

Figures

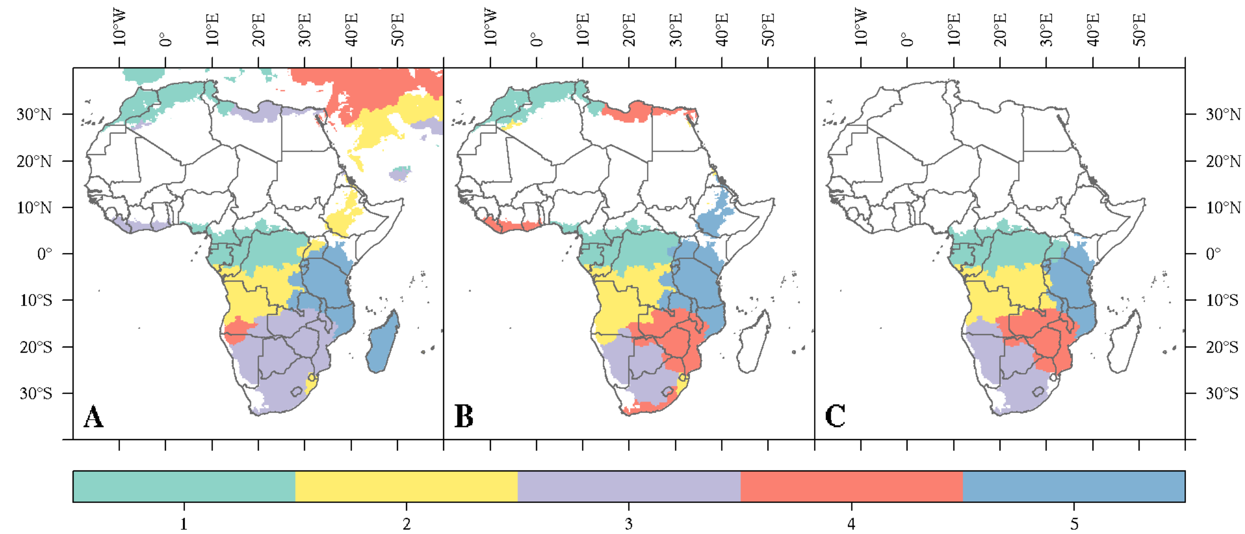

Fig. 1 The effect of geographic masking and data filtering on the quality of regionalization: A) no masking, B) geographic masking of Africa, and C) geographic masking and data filtering.

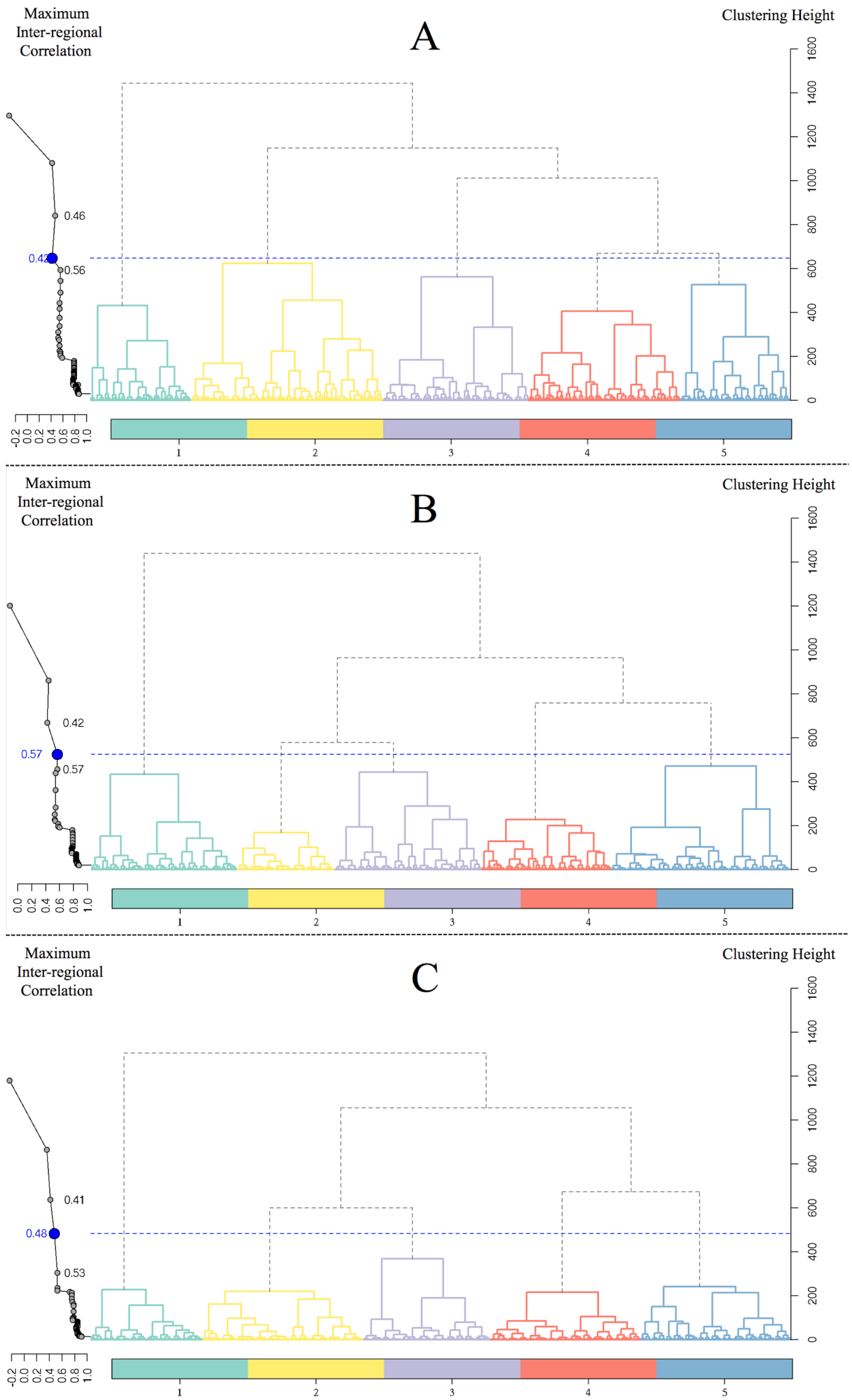

Fig. 2 The effect of geographic masking and data filtering on the clustering dendrogram: A) no masking, B) geographic masking of Africa, and C) geographic masking and data filtering.

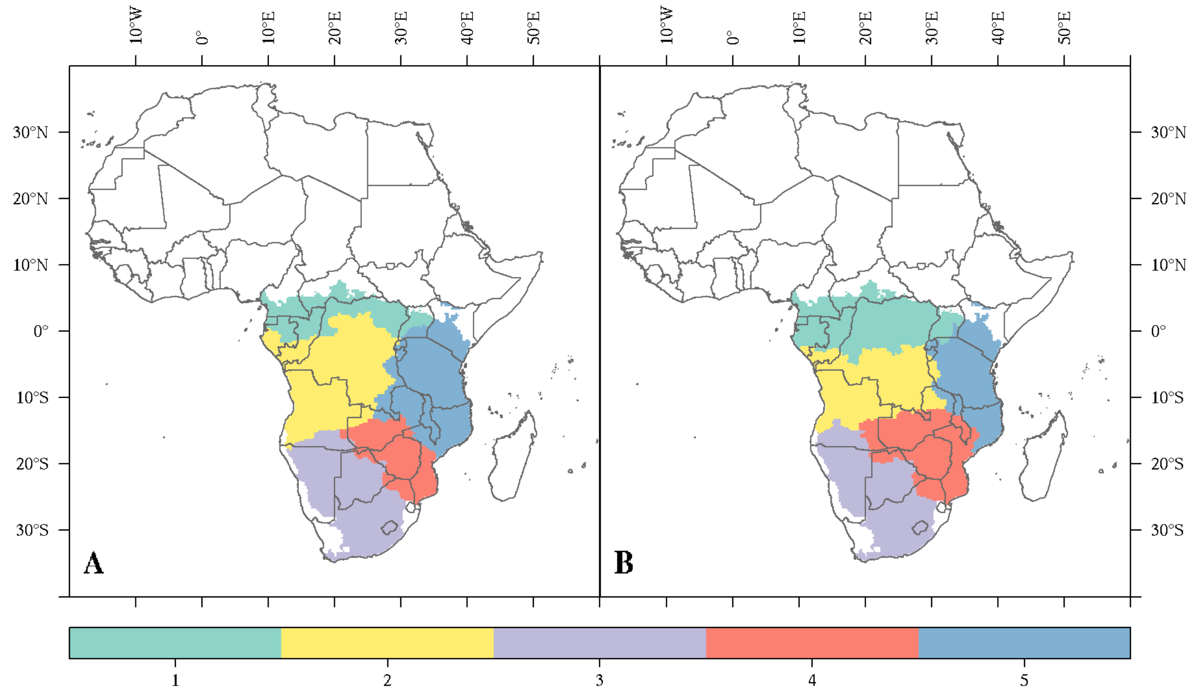

Fig. 3 The effect of detrending and standardization on regionalization quality: A) raw data and B) detrended and standardized data.

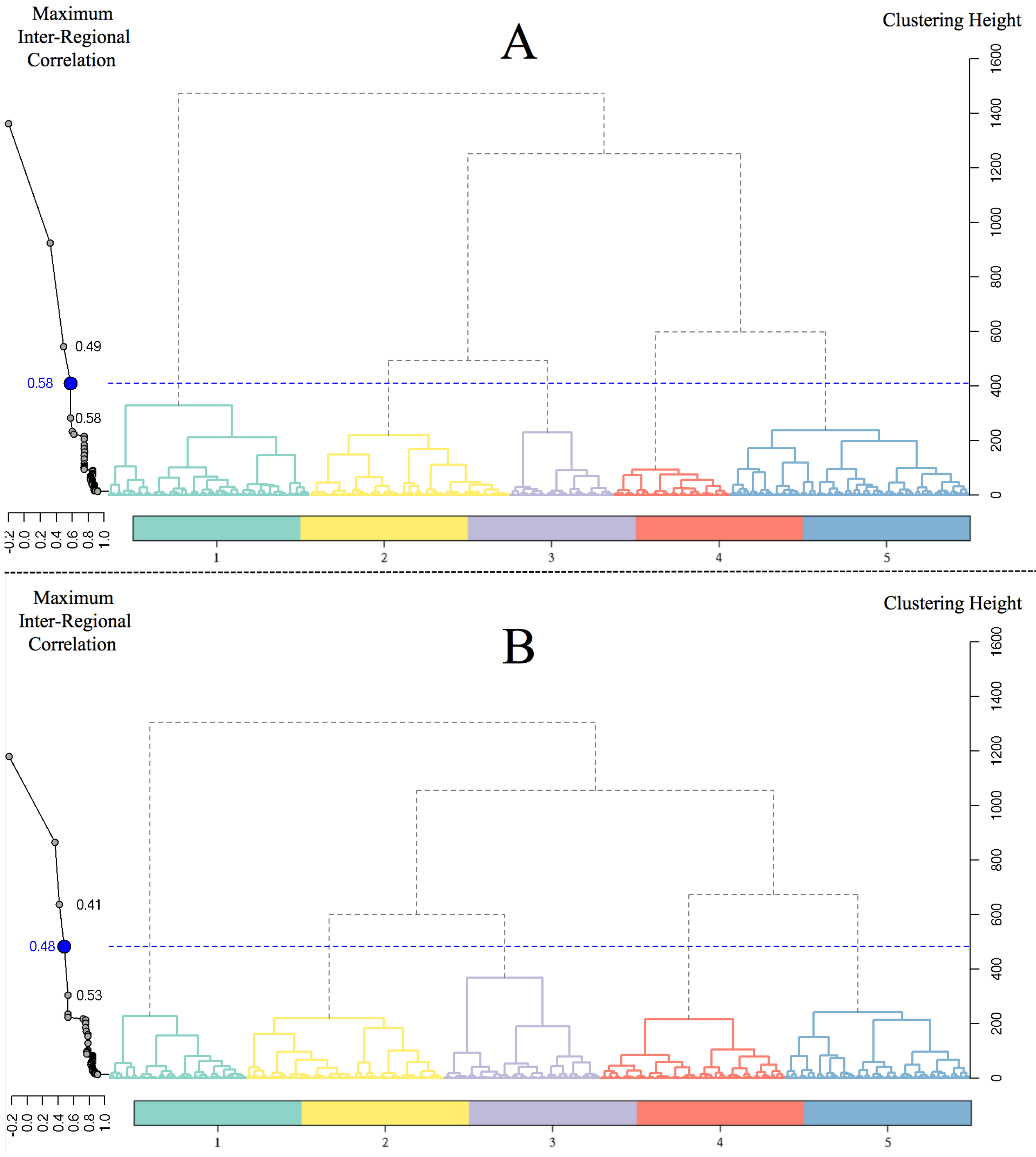

Fig. 4 The effect of detrending and standardization on the clustering dendrogram: A) raw data and B) detrended and standardized data.

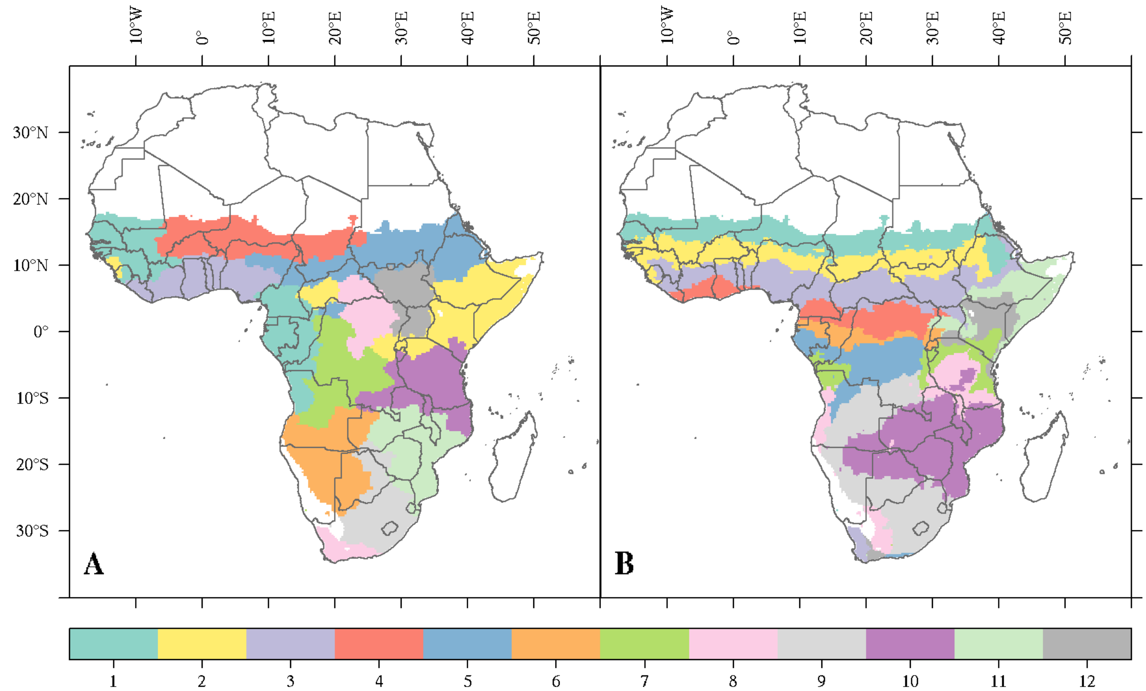

Fig. 5 Regionalization based on: A) interannual variability of annual totals of precipitation and B) annual cycle of precipitation over Africa using CHIRPS data v2.0 (1981-2014).

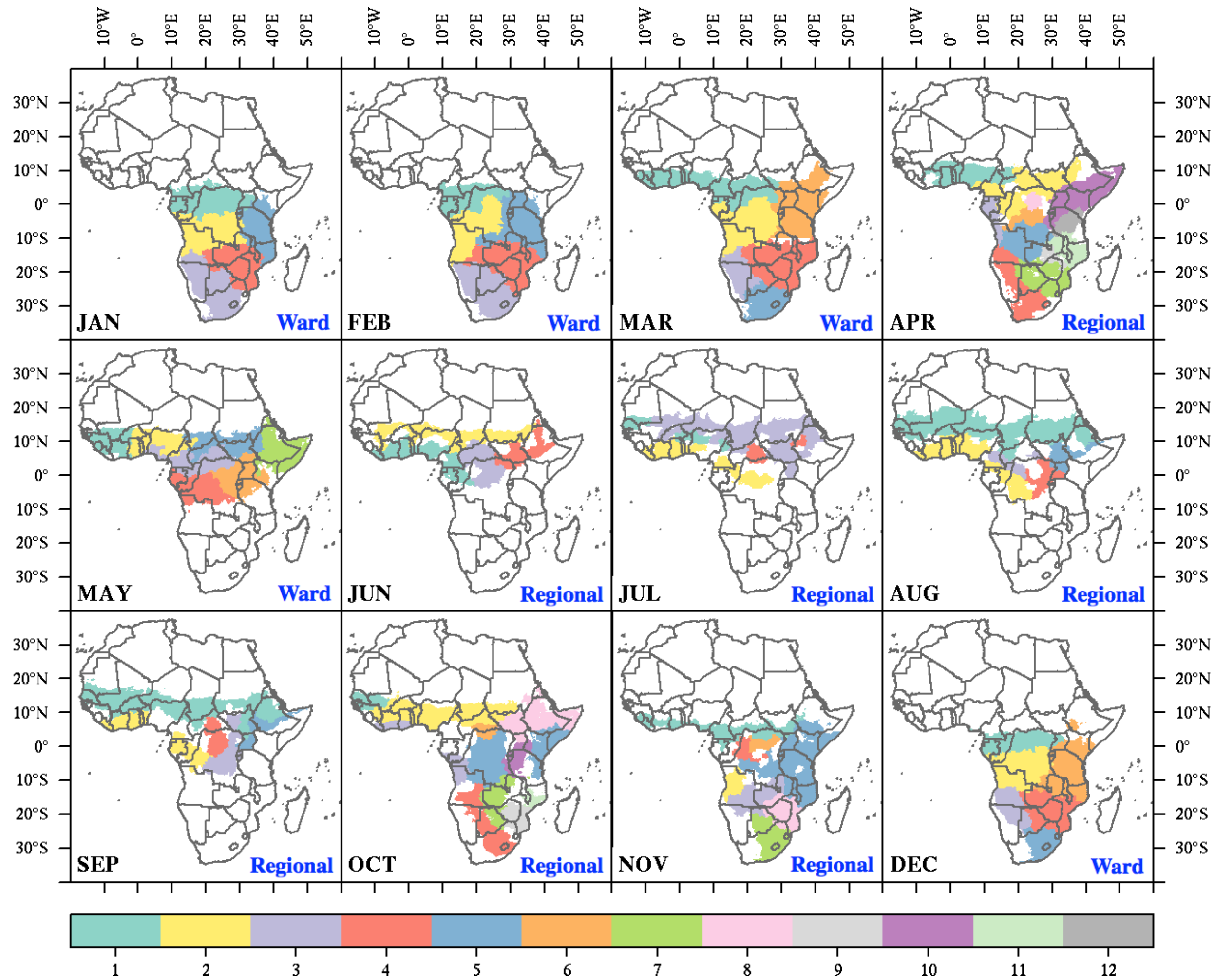

Fig. 6 Regionalization of Africa based on interannual variability of monthly precipitation.

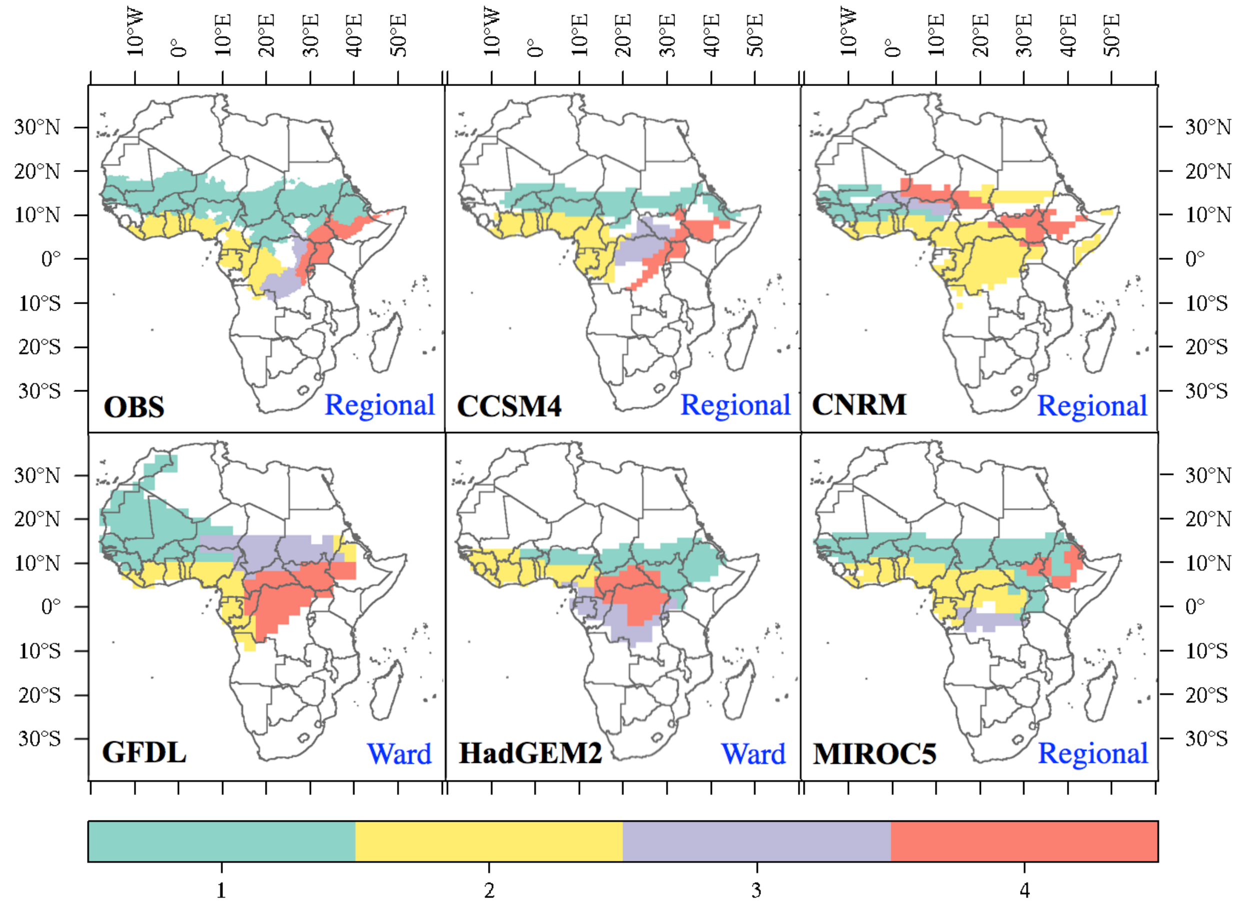

Fig. 7 Regionalization of Africa based on interannual variability of Summer (JAS) precipitation using CHIRPS observations (1981-2014) and different GCMs (1960-1990).

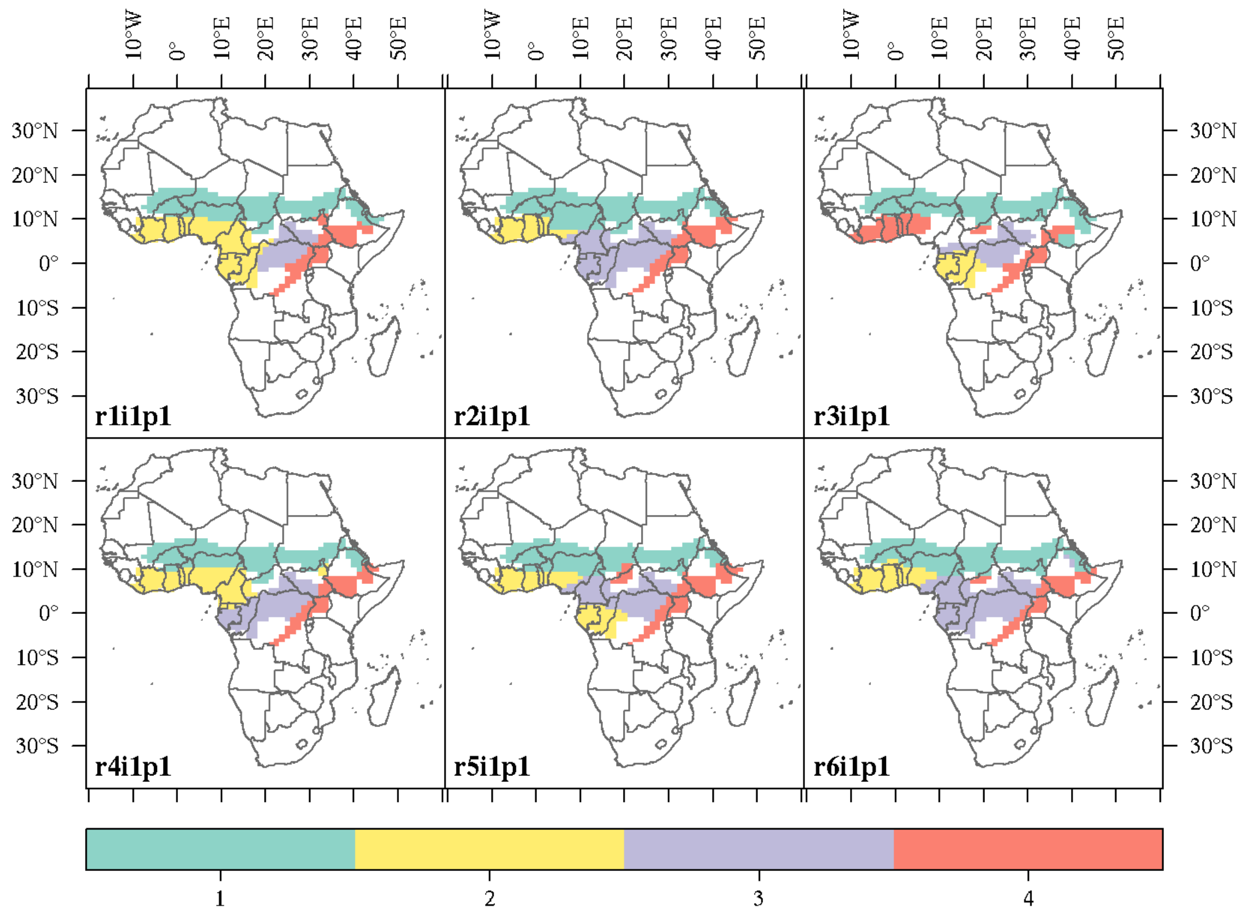

Fig. 8 Regionalization of Africa based on interannual variability of JAS precipitation in 1960-1990 for different ensemble members of CCSM4.

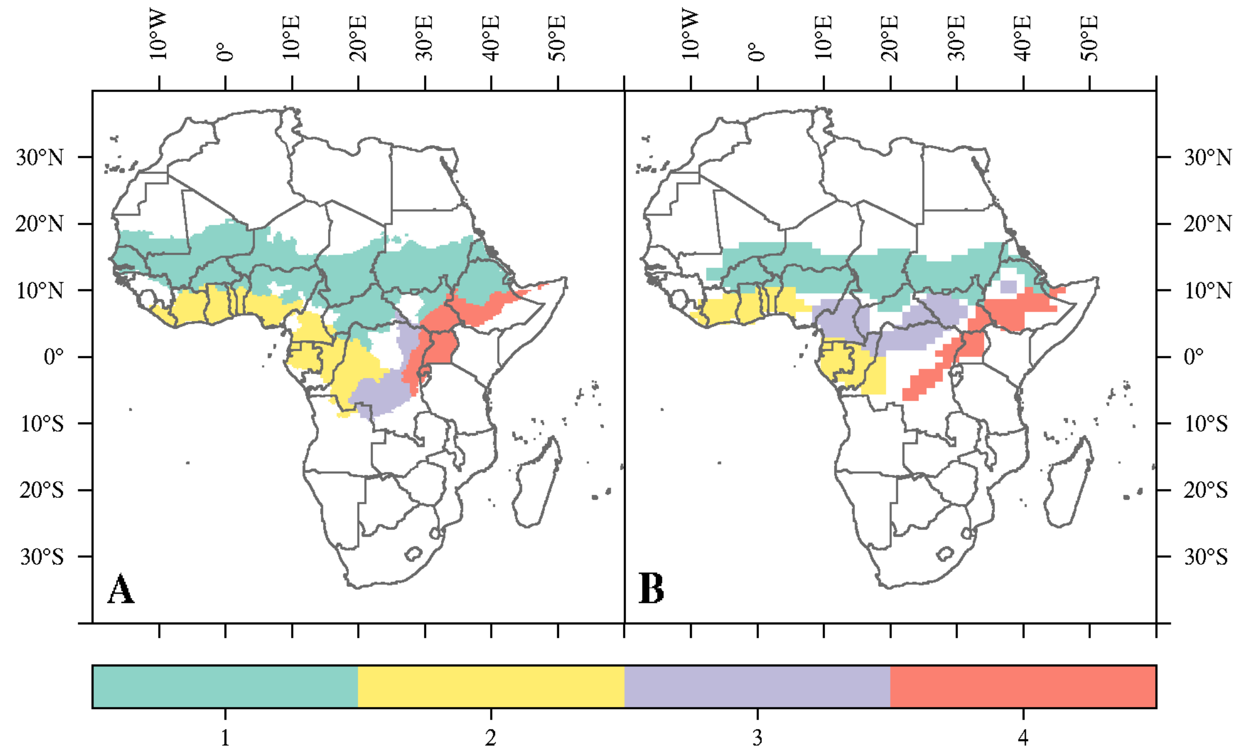

Fig. 9 Regionalization of Africa based on interannual variability of JAS precipitation in 1981-2014 using regional-linkage method: A) CHIRPS and B) CCSM4.

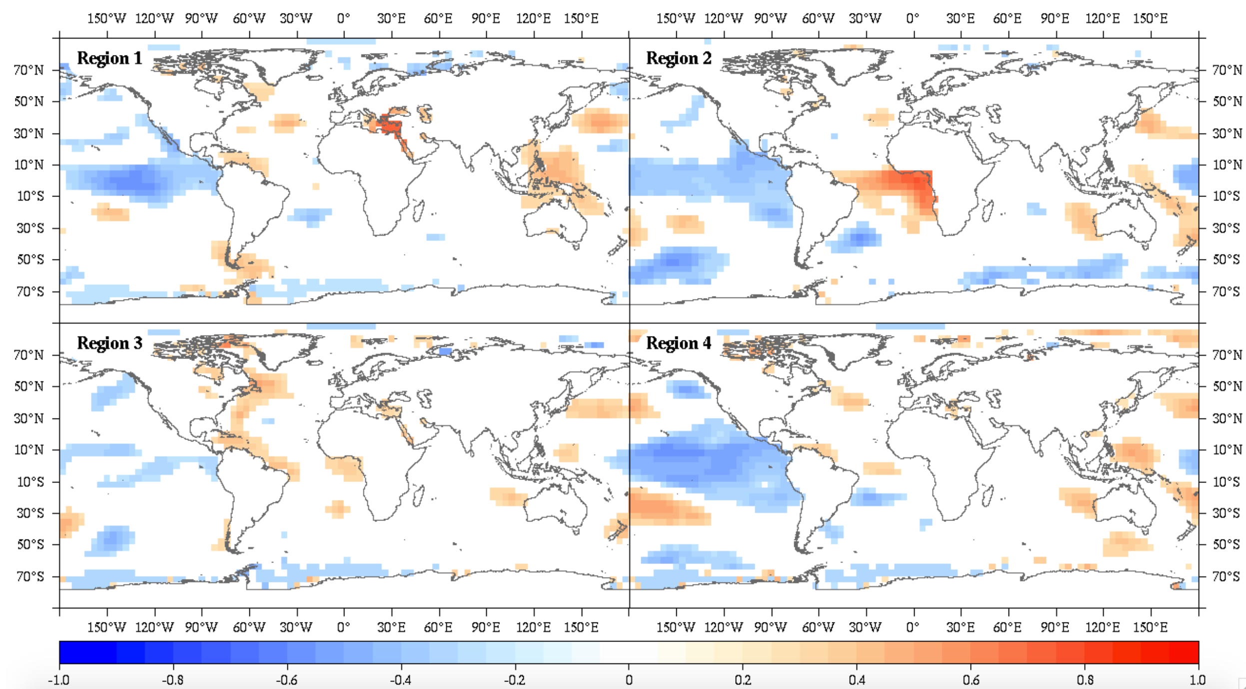

Fig. 10 Correlation patterns of CHIRPS precipitation with global ERSST for the mean timeseries of the four regions of interannual variability of JAS precipitation at the period (1981-2014). All correlations are significant at 90% confidence level.

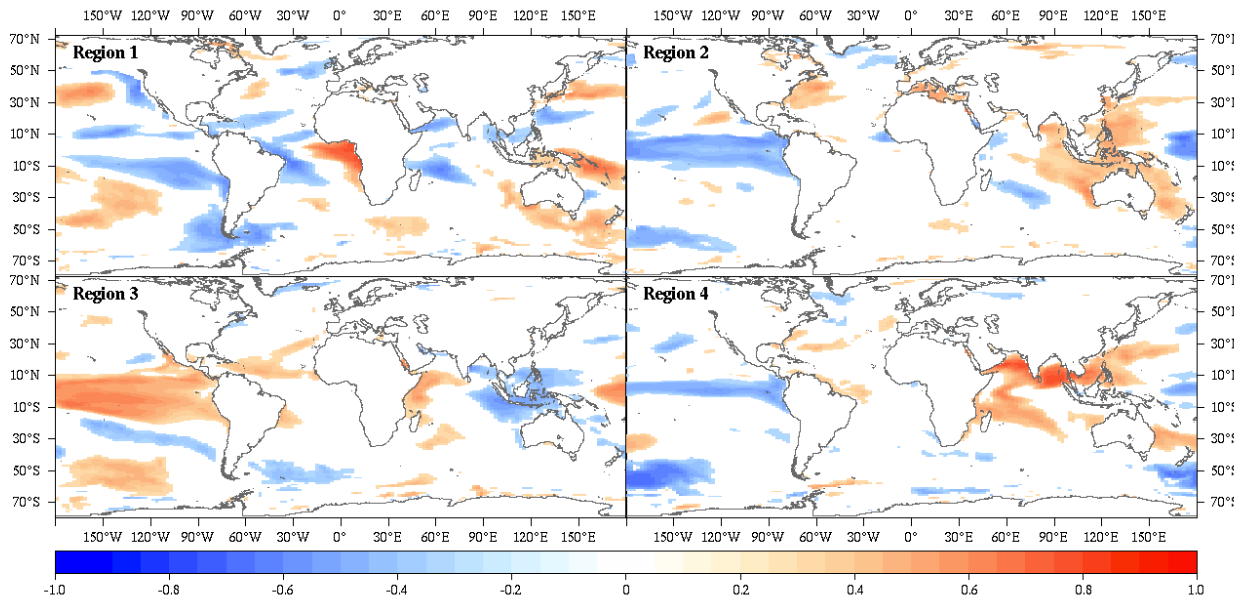

Fig. 11 Correlation patterns of CCSM4 precipitation with the model SST for the mean timeseries of the four regions of interannual variability of JAS precipitation at the period (1960-1990). All correlations are significant at 90% confidence level.

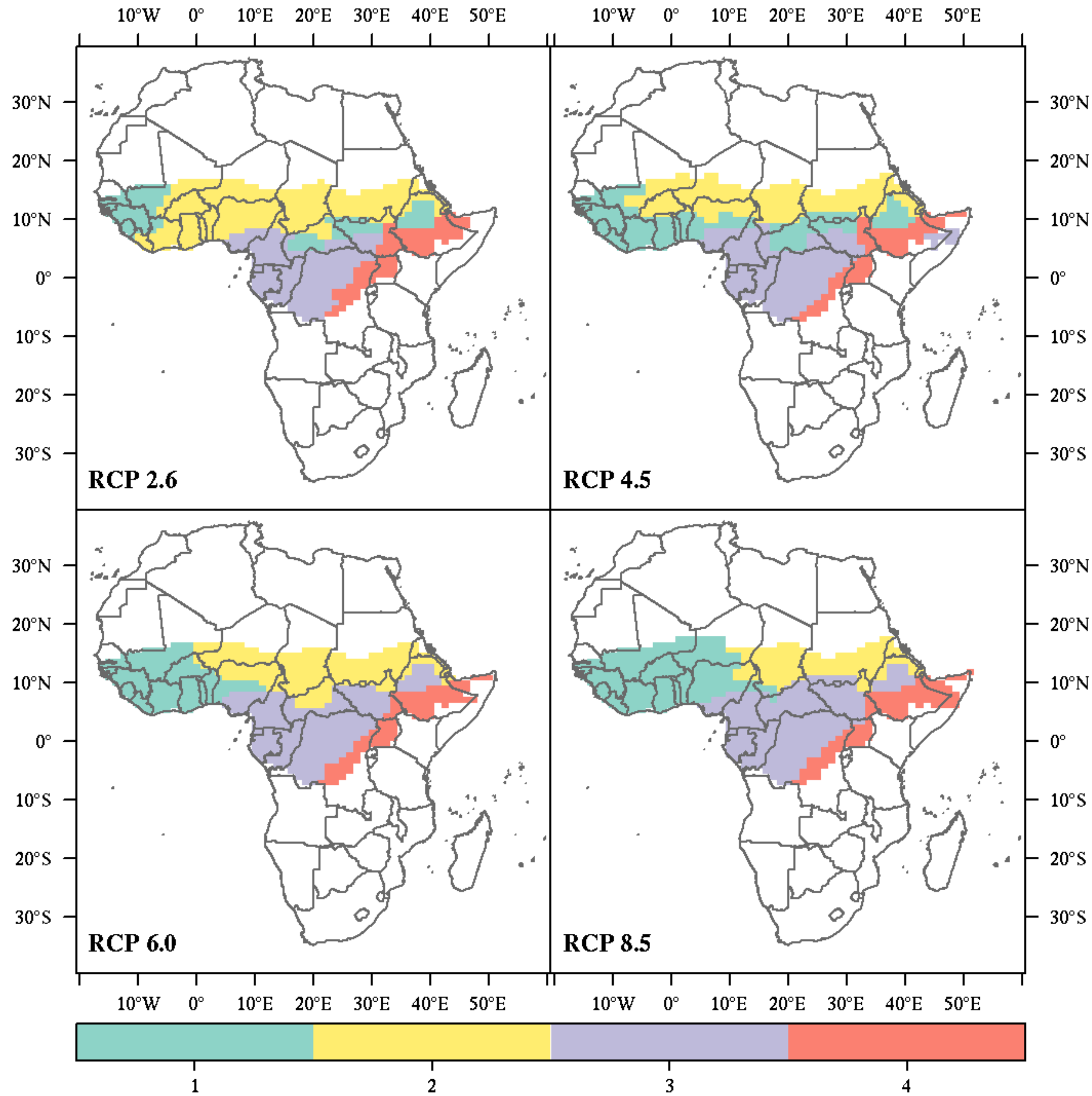

Fig. 12 Changes in the regions of JAS precipitation over Africa using regional-linkage method for different CCSM4 climate projections at the entire simulation period (2006-2100).

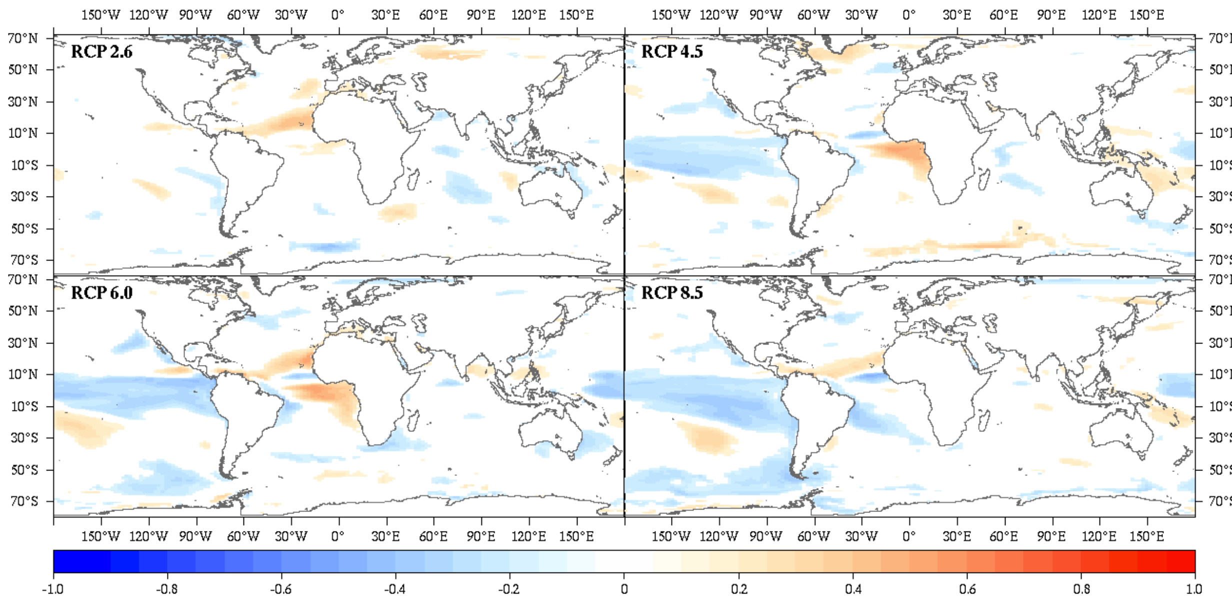

Fig. 13 Correlation patterns of CCSM4 JAS precipitation from the four RCPs with the corresponding SST for region 1 (Western Sahel in **Fig. 12**) at the period (2006-2100). All correlations are significant at 90% confidence level.

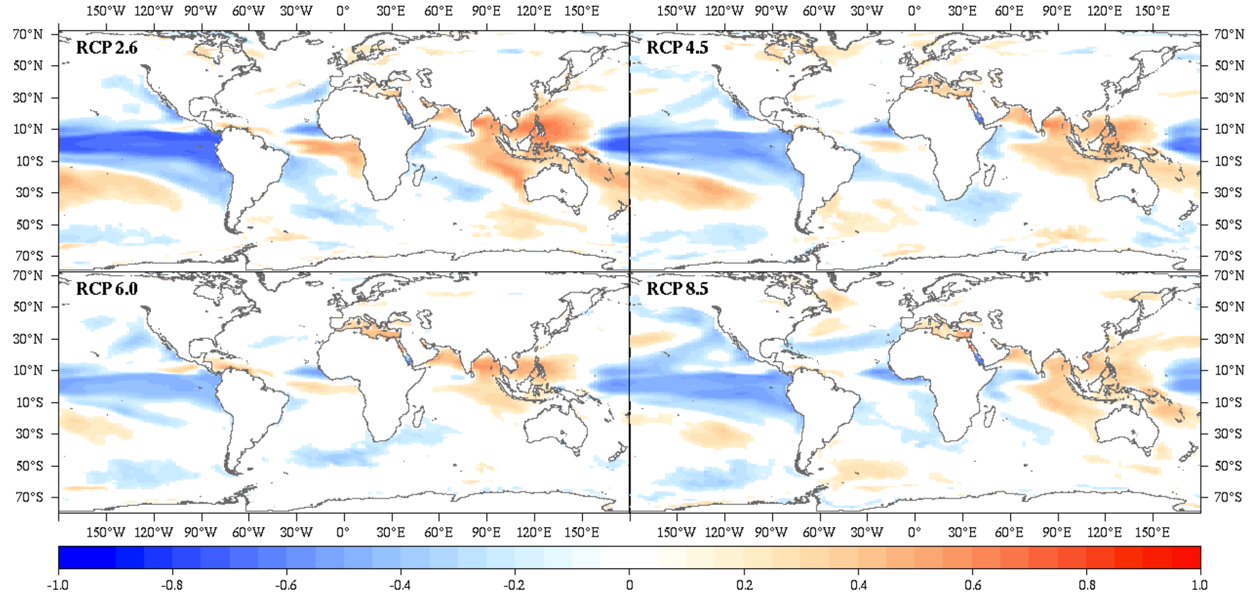

Fig. 14 Correlation patterns of CCSM4 JAS precipitation from the four RCPs with the corresponding SST for region 2 (Eastern Sahel in **Fig. 12**) at the period (2006-2100). All correlations are significant at 90% confidence level.