Assessing the Skill and Reliability of Seasonal Climate Forecasts in Sahelian West Africa

JENNIFER S. R. PIRRET, JOSEPH D. DARON, PHILIP E. BETT, AND NICOLAS FOURNIER

Met Office, Exeter, United Kingdom

ANDRE KAMGA FOAMOUHOU

African Centre for Meteorological Applications for Development, Niamey, Niger

(Manuscript received 12 August 2019, in final form 10 March 2020)

ABSTRACT

Seasonal climate forecasts have the potential to support planning decisions and provide advanced warning to government, industry, and communities to help reduce the impacts of adverse climatic conditions. Assessing the reliability of seasonal forecasts, generated using different models and methods, is essential to ensure their appropriate interpretation and use. Here we assess the reliability of forecasts for seasonal total precipitation in Sahelian West Africa, a region of high year-to-year climate variability. Through digitizing forecasts issued from the regional climate outlook forum in West Africa known as *Prévisions Climatiques Saisonnières en Afrique Soudano-Sahélienne* (PRESASS), we assess their reliability by comparing them to the Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) project observational data over the past 20 years. The PRESASS forecasts show positive skill and reliability, but a bias toward lower forecast probabilities in the below-normal precipitation category. In addition, we assess the reliability of seasonal precipitation forecasts for the same region using available global dynamical forecast models. We find all models have positive skill and reliability, but this varies geographically. On average, NCEP's CFS and ECMWF's SEAS5 systems show greater skill and reliability than the Met Office's GloSea5, and in turn than Météo-France's Sys5, but one key caveat is that model performance might depend on the meteorological situation. We discuss the potential for improving use of dynamical model forecasts in the regional climate outlook forums, to improve the reliability of seasonal forecasts in the region and the objectivity of the seasonal forecasting process used in the PRESASS regional climate outlook forum.

1. Introduction

The Sahel region of West Africa exhibits high year-to-year variability in seasonal precipitation amounts (Nicholson 2013). While many factors combine to determine the vulnerability of communities in the region (Tschakert 2007; Ado et al. 2018; Eze 2018), in years with anomalously low or high precipitation there can be widespread negative impacts on the predominant

livelihoods of pastoralism and rain-fed agriculture (Boyd et al. 2013), alongside other negative impacts of water stress (Chinwendu et al. 2017) or flooding (Cissé et al. 2016). Weather and climate forecasts can therefore help countries and communities to prepare in advance, capitalizing on favorable conditions and reducing the impacts of extreme weather and climate events (Tall et al. 2012; Braman et al. 2013). Seasonal climate forecasts in particular have the potential to inform planning decisions at different scales, from national contingency planning to local-scale farming decisions, provided that the information is timely, reliable and at a scale relevant for decision-making (Sultan et al. 2005). However, there has been limited work to assess the reliability of seasonal forecasts produced for the region, which is an essential step to ensure their appropriate interpretation and use.

Despite continued improvements in numerical weather prediction, the chaotic nature of the atmosphere

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/WAF-D-19-0168.s1>.

Corresponding author: Jennifer S. R. Pirret, jennifer.pirret@metoffice.gov.uk

DOI: 10.1175/WAF-D-19-0168.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

(i.e., high sensitivity to initial conditions) precludes skillful weather forecasts for lead times greater than two to three weeks (Buizza and Leutbecher 2015). To provide forecast information about the atmosphere a month or season ahead, methods that capture important boundary conditions (e.g., ocean processes, sea ice) are the key sources of long-term predictability. The methods used to generate seasonal forecasts vary from dynamical methods using complex numerical models that generate ensemble-based predictions on large computers (Stockdale et al. 1998; Becker and Van Den Dool 2016; Scaife et al. 2019a), to simpler statistical methods relating observations of the atmosphere and/or ocean to predicted variables of interest based on empirically derived statistical relationships (Van den Dool 2007). Dynamical forecasting models have developed considerably in recent years, building on advances in weather and climate modeling capabilities. Recently, some modeling centers have favored a “seamless” approach, taking both initial and boundary conditions into account. This has improved forecasts on intermediate time scales, including seasonal time scales, where the sensitivity to both initial and boundary conditions is pronounced (Doblas-Reyes et al. 2013). Forecast quality also varies spatially, due to different regions’ response to seasonal climate influences such as El Niño–Southern Oscillation (ENSO). Of the 20 regions investigated by Landman et al. (2019), the Sahel’s predictability of seasonal total precipitation ranks near the middle.

In Sahelian West Africa, the variation in year-to-year rains is heavily influenced by the slowly varying components of the climate system, particularly the neighboring oceans and land surface conditions (Rodríguez-Fonseca et al. 2015; Druyan and Fulakeza 2018). The onset, duration, cessation and amount of rains is associated with the poleward migration of the West African monsoon (WAM) and affected by the large-scale flow, including changes in the African easterly jet (AEJ), tropical equatorial jet (TEJ), and Saharan heat low (Vellinga et al. 2013; Okoro et al. 2017). Topography has also been argued to have an important influence on the WAM migration (Sultan and Janicot 2003), as well as inertial instability advancing moisture migration northward (Cook and Vizy 2006). Capturing these important processes, and initializing forecasts using observations of sea surface temperatures (SSTs) in the tropical Atlantic and other relevant ocean basins, is essential to enable skillful seasonal forecasts in the Sahel (Vellinga et al. 2013; Sheen et al. 2017).

At lead times of a month or more, it is not expected that weather events can be forecast reliably. However, it can be possible to forecast overall statistics of a season, such as seasonal mean temperature or seasonal

precipitation totals. It is therefore typical to communicate seasonal forecasts as departures from a climatological “normal” or average (Troccoli et al. 2008). Therefore, seasonal forecasts should always be presented probabilistically (Troccoli et al. 2008; White et al. 2017). Assessing seasonal forecast skill and reliability requires a long time series of observations; we cannot meaningfully assess the reliability of a seasonal prediction system using observations of a single or small number of seasons.

In many regions of the world, seasonal forecasts are produced and disseminated by National Meteorological and Hydrological Services (NMHSs) at the national level and Regional Climate Centres (RCCs) at the multinational level, through Regional Climate Outlook Forums (RCOFs; Semazzi 2011). Through the support of the World Meteorological Organization (WMO), RCOFs cover defined geographical regions and act to convene regional meteorological experts and stakeholders to generate a forecast for the coming season. The outputs of different forecasting techniques and models are discussed, including statistical methods, dynamical models, and expert judgments. A consensus forecast for the region is developed and communicated to key users. The forum that covers the Sahel is currently known as *Prévisions Climatiques Saisonnières en Afrique Soudano-Sahélienne* (PRESASS), previously *Prévisions Climatiques Saisonnières en Afrique de l’Ouest* (PRESAO), and is convened by African Centre of Meteorological Application for Development (ACMAD) and Centre Regional de Formation et d’Application en Agrométéorologie et Hydrologie Opérationnelle (AGRYMET). The PRESASS RCOF typically takes place in April or May to generate forecasts for the rainy season in the region (June–September). The forecasts contain information in the form of maps and advisory documents, focused on either June–August or July–September.

An example forecast map is shown in Fig. 1. Total precipitation for the season is divided into terciles of above, near, or below normal, compared to the long-term climatological average (typically 1981–2010). The areas highlighted indicate where the seasonal total precipitation is expected to differ from climatology, expressed in terms of a percentage chance of each tercile. The baseline forecast is a 33.3% chance for each tercile (i.e., climatology). The orange forecast area that covers Sierra Leone, Liberia, southern Côte d’Ivoire, and southwestern Ghana highlights that there is a 50% chance for the lowest tercile, indicating a forecast tendency toward dry conditions, with an explanatory statement of “well below average precipitation very likely.” However, there is a 20% probability of above average precipitation and a 30% chance of normal precipitation,

FIG. 1. An example PRESASS forecast, issued by ACMAD for the July–September season in 2017.

meaning a 50% chance of normal to above-normal precipitation. Thus there is a 50% chance of below-normal precipitation and 50% chance of normal to above normal associated with the interpretation into “well below average precipitation very likely,” which demonstrates the difficulty and challenges in using textual summaries of the quantitative information.

The RCOF process was first introduced in the late 1990s, and was most recently reviewed by the WMO (2017). The WMO’s recommendations include a call for increased objectivity in the forecasting process, primarily through using the results of dynamical model forecasts and a routine evaluation of forecast skill. Guidance is available to support the verification of operational seasonal forecasts (WMO 2018), which contributes to the overall aim of the WMO to develop objective approaches to seasonal forecasting (Dilley and Kolli 2017). Mason and Chidzambwa (2008) assessed skill for three African RCOFs and found evidence of positive skill in all cases, showing confidence in the RCOF process. More recently, Bliefernicht et al. (2019) assessed the performance of the PRESASS forecasts in West Africa, finding that these forecasts have skill in the above- and below-normal rainfall categories.

However, both Mason and Chidzambwa (2008) and Bliefernicht et al. (2019) uncovered issues in the forecasting process. Mason and Chidzambwa (2008) revealed a tendency to “hedge” with higher probabilities in the near-normal category, while the below-normal tercile was forecast less frequently than observed.

Bliefernicht et al. showed that forecast categories are skewed away from the 1/3 probability category with overforecasting of the near-normal category. Bliefernicht et al. (2019) also identify that slight changes in stated probability can give large changes in forecast value. These findings corroborate the need for RCOFs to use more objective methods to mitigate these outcomes.

The current study builds on the findings of these studies to assess the quality of PRESASS forecasts since 1998 and compare the findings with the performance of several dynamical models providing forecasts for the region. The aim is to provide new scientific evidence to help improve the existing seasonal forecasting processes, in line with the WMO’s recommendations. The work has been conducted as part of the Adaptive Social Protection: Information for enhanced Resilience (ASPIRE) project, funded by the U.K. Department for International Development (DFID), which ultimately aims to improve the use of climate information and forecasts, to enable the scale-up of social protection systems in Sahelian West Africa in advance of weather and climate shocks.

In section 2, we outline the different sources of data including observational data, PRESASS forecasts and Global Producing Centre (GPC) forecasts from four different models. Section 3 describes the methods used to assess forecast quality in terms of skill (section 3b), reliability (section 3c), and bias (section 3d). The results comparing PRESASS forecasts to one GPC model can be found in section 4a, with the comparison between

GPC models in [section 4b](#). These results are discussed in [section 5](#) before conclusions are drawn in [section 6](#).

2. Data

a. Observational data

There are multiple observational datasets that cover the region of interest. For this study, the chosen observations are from the Climate Hazards Group Infrared Precipitation with Stations (CHIRPS) project. CHIRPS uses satellite estimations of precipitation, corrected by gauge and station data, to produce precipitation data for the tropics and subtropics (50°N–50°S) on a 0.05° grid for the period from 1981 to present ([Funk et al. 2015](#)). CHIRPS data were chosen because they are daily, available since 1981, at high resolution compared to similar products, and are well respected and used in Africa and the focal region ([Poméon et al. 2017](#); [Dinku et al. 2018](#)). In this study, CHIRPS data are used in the forecast verification process, though at the lower resolution of 1.0° given that similar results (not shown) were obtained at the two resolutions. Results obtained using CHIRPS data were compared to those with GPCC ([Becker et al. 2013](#)) and Tropical Applications of Meteorology Using Satellite Data and Ground-Based Observations (TAMSAT; [Maidment et al. 2017](#)) data, with reliability diagrams available in (Figs. S1 and S2 in the online supplemental material). The diagrams based on CHIRPS and GPCC data are very similar, with that of TAMSAT showing underestimation of the forecast probabilities in the “above-normal” category. We suggest that this difference demonstrates observational uncertainty. In the rest of this work, we use CHIRPS because it combines use of satellites and stations, whereas TAMSAT uses only satellites and GPCC uses only stations/gauges.

b. RCOF data

As discussed in [section 1](#), we focus on season total precipitation forecasts for the main rainy season in Sahelian West Africa. PRESASS map forecasts are made for either June–August or July–September (JAS), and ACMAD provided the full archive (1998–2017) of images for the JAS season. The forecast images (Figs. S3 and S4) were digitized using QGIS ([QGIS Development Team 2018](#)) onto the same grid as the CHIRPS observation data. This resulted in gridded forecast probabilities of each tercile of precipitation, which were then compared to observations. More detail on the digitization process can be found in [section 3a](#).

c. Dynamical model data

The observational data were also compared to seasonal forecasts produced using dynamical models from

the Met Office, the National Center for Environmental Prediction (NCEP), the European Centre for Medium-Range Weather Forecasts (ECMWF), and Météo-France. The forecast models were chosen due to the availability of data, longevity of hindcast records, and their use in the region. NCEP data was downloaded from the National Center for Atmospheric Research (NCAR) Research Data Archive ([Saha et al. 2011](#)). Data from the other centers was downloaded from the Copernicus Climate Data Store (CDS), part of the Copernicus Climate Change Service (C3S; [Raoult et al. 2017](#)). The figures containing data from CDS have a caption to note this.

The Met Office’s operational ensemble model for seasonal forecasting is the Global Seasonal Forecast System version 5 (GloSea5; [MacLachlan et al. 2015](#)). The operational forecasts are initialized daily and are run for up to 6 months, calibrated using a set of hindcasts (i.e., initialized forecasts using historical reanalysis data) that are run for the period 1993–2015. GloSea5 uses ERA-Interim reanalysis data from the ECMWF ([Dee et al. 2011](#)). Hindcasts consist of seven members, initialized on the 1st, 9th, 17th, and 25th of each calendar month, and so operational bias correction uses a weighted average of the four nearest dates. In this work, we concentrate on the data that would have been available at the PRESASS meeting; that is, the four hindcasts centered around the start of May for the JAS forecast (17 April, 25 April, 1 May, 9 May). We combined these forecasts to give a 28-member ensemble.

The Climate Forecast System, version 2 (CFSv2), has been the operational seasonal model at NCEP since March 2011 ([Saha et al. 2014](#)). The hindcasts took initial conditions from the Climate Forecast System Reanalysis ([Saha et al. 2010](#)) and cover the period 1982–2010. Operational forecasts and hindcasts are/were made four times daily, but different runs have different maximum lead times. In this work, only the 9-month hindcasts are used, because lead times of longer than one season are required. The 9-month runs are initialized every 5 days, but four times on that day. We combine the forecasts initialized on 21 April, 26 April, 1 May, 6 May and 11 May to give a 20-member ensemble.

Version 5 of the seasonal prediction system of the ECMWF (SEAS5) is underpinned by their medium-range atmospheric model, with coupling to ocean, sea ice, land, and wave models. SEAS5 has been operational since November 2017 producing a 51-member ensemble, with hindcasts covering the period 1981–2016 consisting of 25 ensemble members initialized on the first of each month. We use all hindcast members from 1 May.

Météo-France System 5 (Sys5) has been operational since March 2017, consisting of atmospheric, land

surface, ocean, and sea ice models coupled together. Every month, 51 forecasts are initialized: 25 on the first Wednesday after the 12th and 26 on the first Wednesday after the 19th. The hindcast consists of 15 ensemble members initialized on the first Wednesday after the 19th, again covering the period 1981–2016. We use all hindcast members for May, but in the Copernicus CDS these are labeled as 1 May.

3. Methods

In this section, we outline the methods used to assess seasonal forecasts, consistent with the concept of forecast “goodness” as discussed by [Murphy \(1993\)](#). Various metrics exist to measure the quality or performance of a forecast, and we focus on skill, reliability and bias. Skill considers how the forecast compares to a reference forecast, for example a forecast of no skill (“random chance”) or a persistence forecast. Reliability quantifies the probabilistic agreement between the forecasts and the observations. Bias quantifies the extent of any systematic differences, on average, between the forecasts and the observations.

In our analysis we concentrate on the Sahel area, defined as 10°–20°N, 20°W–30°E (marked as a green box in [Fig. 3a](#)). This area is similar to that used by [Vellinga et al. \(2016\)](#) but extended farther west to include all of Senegal. We evaluate the forecasts for different time periods, owing to differences in data availability, ensuring we use concurrent data for comparisons between the dynamical model forecasts and PRESASS forecasts. In [section 4a](#), we consider the period of overlap between GloSea5 hindcasts and the PRESASS forecasts (1998–2015), and in [section 4b](#) we use the period of overlap between the different models hindcasts (1993–2010).

a. Digitization of PRESASS forecasts

The digitization was completed using QGIS ([QGIS Development Team 2018](#)), with more detailed instructions available from the ASPIRE website ([Met Office 2019](#)). When the 20-yr forecasts are examined, a natural development of the forecasting process is apparent, with two key implications. First, the forecast area is not consistent, so we only use forecasts from the 18 countries shown in [Fig. 1](#) excluding the Cape Verde Islands. This represents the main core of the countries that appears on most of the forecasts throughout the 20 years.

Second, there is not a consistent way to interpret the area shown in gray in [Fig. 1](#). Analyzing the 20 forecasts Figs. S3 and S4 shows two different approaches, which are sometimes colored differently (e.g., 2012) but sometimes the same (e.g., 2017). In the north, the climate is arid so in most years, no forecast is made.

Conversely, farther south and more recently, gray becomes a forecast of “near-average precipitation,” but there is no boundary between the regions for no forecast due to aridity and forecasts of near-normal rainfall (except in 2012 and 2013). However, the boundary between the regions of no-forecast and implicit near-normal forecast varies across the 20 forecasts, so it is not possible to identify an area where null forecasts can be consistently assumed. This raises the question of what percentage probability should be assigned to the gray or white areas. An even split across all three terciles (33–33–33) would represent a null forecast and is a reasonable assumption made by [Walker et al. \(2019\)](#), but this would not be a forecast of near-normal precipitation and is forecast explicitly in 2006. A near-normal precipitation forecast could be represented by percentages that are skewed toward the middle tercile such as 30–40–30, but this is forecast explicitly in 2004. These conflicting motivations for having regions with no explicit forecast and different interpretations means that in this study, we use only the regions where the PRESASS process explicitly gives a percentage forecast.

b. Skill assessment

As an initial assessment, correlation maps illustrate the spatial characteristics of deterministic skill. At each point, the Pearson correlation coefficient is calculated between every year’s ensemble mean forecast and the observations ([Fig. 3](#)). Where the correlation is statistically significant (at the 95% confidence level), the map is hatched. The Pearson correlation is chosen for its simplicity and common usage, but relies on normally distributed data. While this is not the case for daily precipitation, calculating season total precipitation means the values tend to become normally distributed but does not guarantee that the correlation is unaffected by outliers. For this reason, we also calculated the Kendall tau correlation, based on ranking and not affected by outliers, and the results are qualitatively similar (not shown) to those for Pearson correlation.

Taking a probabilistic approach, relative operating characteristic (ROC) scores assess the skill of forecasts in terms of whether a forecast “event” occurs. Events are expressed as precipitation falling into each tercile (above, near, and below normal) in turn. Thresholds are applied to the forecast probability for each event, with the event being forecast if the probability exceeds the threshold. The corresponding point in the observational data is identified, with four possible outcomes:

- hit: the event is forecast and occurs;
- false alarm: the event is forecast and does not occur;

- miss: the event is not forecast and does occur; and
- correct negative: the event is not forecast and does not occur.

In this paper, ROC diagrams are used to illustrate the skill of seasonal forecasts produced using different models over the Sahel. The vertical axis of a ROC diagram shows the “hit rate,” which is the number of hits compared to the total number of observed occurrences (number of hits plus the number of misses). The horizontal axis shows the “false alarm rate,” which is the number of false alarms compared to the total number of nonoccurrences (the number of false alarms plus the number of correct negatives). For each tercile of precipitation, these two rates are plotted against each other for each forecast probability threshold. The resulting curve shows whether or not the forecast system is skillful. A skillful forecast system will maximize the hit rate and minimize the false alarm rate, so the ROC curve would bow toward the top left of the plot. If the forecast system has no skill, then the false alarm rate and hit rate would be equal, with the ROC curve along the straight diagonal line. As such, ROC diagrams indicate the forecast skill compared to a “random chance” forecast. A ROC skill score (ROCSS) is given in the plot legend (Fig. 4), calculated as per Wilks [2011, Eq. (8.46)]. ROCSS is 1 for perfect forecasts and 0 for “random guess” forecasts.

In addition, ROC maps can be used to provide spatial illustrations of forecast skill. The ROC maps used in this study show the area under a ROC curve at each grid point, for each precipitation tercile separately. A forecast with no skill, and ROC curve that falls along the diagonal, equates to an area under the curve of 0.5. Therefore, where a forecast has skill, the ROC map shows a value of greater than 0.5; the higher the value, the more skillful the forecast. Note that the area is calculated using the trapezoidal rule, which will slightly underestimate the area under the curve so should be used primarily to compare different areas of the forecast and the different forecast systems.

c. Reliability assessment

Using the example PRESASS forecast in Fig. 1, and focusing on the orange forecast area, there is a 50% probability of the area receiving below-average precipitation, a 30% probability of receiving near-average conditions and a 20% probability of above-average conditions. In the scenario that this region received a perfectly reliable forecast, we would expect 20% of the points in this region to receive precipitation in the above-normal tercile, 30% near normal, and 50% below normal. This is repeated over the 20 years, in order to fairly assess the reliability.

To assess the reliability, we collected forecast probabilities for each tercile over the 20 years of available forecasts. For example, we collected the cases where the forecast of above-average precipitation was 20%, and count how many times that the precipitation was in the above average tercile; if this occurred 20% of the time, then those forecasts would be reliable. We assess the forecasts in intervals of 10% probability.

Reliability diagrams illustrate how reliable forecasts are by comparing forecast probabilities with the frequencies of actual events. The vertical axis of a reliability diagram shows the observed frequency, and the horizontal axis shows the forecast probability. In this study we group the observed frequency and forecast probabilities into bins with 10% intervals. The lower half of Fig. 2 shows two example diagrams. As in the example described above, reliable forecasts give similar values for forecast probability and observed frequency, so perfect reliability would result in a line on the main diagonal ($y = x$).

d. Bias assessment

Frequency diagrams (or sharpness diagrams) illustrate the forecast bias. Examples are shown at the bottom of Fig. 2, giving the number of forecasts within each bin over all years and for the Sahel, our region of interest. Frequency diagrams are used in two ways: first, to ensure there are a sufficient number of forecasts in each percentage category to allow statistically significant interpretation; and second, to assess probabilistic bias.

Over a long period of time, the frequency of forecast probabilities in each tercile will converge to the climatological frequency of one-third (0.33). If on average a tercile is forecast with a higher (or lower) percentage chance than the climatological average, this will be represented as a shift in the histogram away from a center point of 0.33 and indicate a probabilistic bias in the forecast. Note that because the model forecasts are assessed against terciles calculated from model climatology, they will not exhibit probabilistic bias, and we do not assess absolute bias (i.e., the long-term average difference between model hindcast mean and observed mean).

4. Results

Concentrating on the JAS season, we first assess the reliability of the PRESASS forecasts by generating ROC, reliability and frequency diagrams for both the PRESASS forecasts and the hindcasts of GloSea5 over a concurrent period (1998–2015). We then compare the performance of GloSea5 with three other dynamical models (section 4b) using all of the approaches discussed in section 3, also for the JAS season. For the dynamical

FIG. 2. (top) ROC diagrams and (bottom) reliability diagrams for (left) PRESASS forecasts and (right) GloSea5 hindcasts of July–September season total precipitation for 1998–2015.

model comparisons, we choose a period of 1993–2010, which is the longest concurrent period available to compare model results. However, that overlapping available forecasts cover only 18 years limits our assessment, discussed further in [section 5](#).

It is also important to highlight that there are relatively few occurrences of forecasts with very high

probabilities for a particular tercile outcome at a given location (e.g., higher than 0.7 or 70% chance of above-normal precipitation). For such small sample sizes, it is challenging to meaningfully and significantly assess forecast quality and reliability and the results often lack significance. Therefore, reliability diagrams and frequency diagrams must be used together to provide a

measure of confidence in the results. For the correlation maps, a significance test is performed [Rees 2001, Eq. (14.2)] and the areas with statistically significant correlation are hatched, supporting our interpretation of the results.

a. PRESASS forecasts

The top part of Fig. 2 shows the ROC curves for each tercile of the PRESASS forecasts and GloSea5 hindcasts, covering the period 1998–2015. In both PRESASS and GloSea5, for the lower and upper terciles the curves bow slightly toward the top-left corner, indicating that there is some skill in forecasts of below and above-normal precipitation. However, for the middle tercile the curves are close to the diagonal line, showing almost zero skill for forecasts of near-normal precipitation, as also found in Bliefernicht et al. (2019). Overall when comparing the PRESASS and GloSea5 approaches, they give similar levels of skill using this measure.

The lower part of Fig. 2 shows the reliability diagram and frequency diagrams for each forecast tercile for the GloSea5 model hindcasts for the years 1998–2015. It shows good reliability for all three terciles when the forecast probabilities are less than around 30% (around the climatological value), with the reliability curves close to the diagonal. However, reliability gets worse for higher forecast probabilities, particularly for the middle (near-normal precipitation) and lower (below-normal precipitation) terciles. The frequency diagrams show that there are few forecasts for these higher probabilities, particularly when the forecasts are higher than 60%, so it is difficult to draw firm conclusions. Assessing reliability over a longer period of time than 1998–2015 would allow more data to increase the robustness of results, but in this case we are limited to the time period available for the PRESASS forecasts. Examining the PRESASS forecasts' reliability diagram shows the upper and middle tercile forecast percentage chances are similar to the percentage of observations that have near-normal and above-normal precipitation, indicating good reliability. However, the middle tercile is only forecast across three probability bins, limiting the interpretation of forecast reliability. For the lower tercile, most of the reliability points are above the diagonal indicating an underestimation of forecast probabilities. This is compounded by the frequency diagrams underneath the reliability plot. For the lower tercile, the frequency diagram is skewed toward values below the climatological rate of 0.33. This means that there is a bias toward forecasting lower probabilities in the below-normal precipitation category.

Comparing the reliability diagrams for PRESASS and GloSea5 indicates that, at typical forecast probabilities (less than 50%), the PRESASS process has marginally better reliability for the above-normal precipitation category because the plotted line is closer to the diagonal. GloSea5 tends toward higher reliability in the near- and below-normal precipitation categories, compared to the PRESASS forecasts. It is important to note from the frequency diagrams that for higher forecast probabilities, there are too few forecasts to robustly assess reliability.

b. Comparing GPC forecasts

In this section, we compare hindcasts from GloSea5, CFSv2, SEAS5, and Sys5 (section 2c), considering only the overlapping hindcast period of 1993–2010. First, we examine maps of Pearson correlation (Fig. 3), providing spatial information on forecast performance. There are regional variations for each model but in general the models show lower skill in the north toward the Sahara, and in the southeast around Cameroon, Central African Republic, and the Congo basin. Focusing on the domain of interest used for the subsequent ROC curves and reliability diagrams (green box in Fig. 3a), forecasts generally show lower skill in the west compared to the east, with consistent areas of significant correlation in southern Niger and Chad. Sys5 exhibits overall lower correlations compared to the other models in our domain of interest, but better performance farther south (Côte d'Ivoire, Ghana, Togo, Benin, Nigeria) along with ECMWF Seas5.

The ROC curves for each model are shown in Fig. 4. In all four hindcasts, the upper and lower terciles show positive forecast skill as their ROC curves bow toward the top-left corner. Skill is lower for the middle tercile in all models. Considering the ROC skill scores for the upper and lower terciles gives some indication of which models show greater skill on average for season total precipitation over the Sahel. NCEP CFS and ECMWF SEAS5 perform better than GloSea5, which in turn performs better than Météo-France's Sys5. Otherwise, the models only show subtle differences. Some models show better skill in the upper tercile compared to the lower tercile while others perform better at higher or lower forecast probabilities.

Maps showing the area under the ROC curve for each tercile of precipitation at each grid point are shown in Fig. 5. For all models and terciles, skill scores are absent in the northeast of the mapped domain because it is arid and receives almost no precipitation, thereby limiting the relevance of any results. Skill is higher in the lower and upper terciles than the middle tercile, corroborating the results from Fig. 4. Different models show higher skill in different areas: GloSea5 and CFS perform well in the east of the region, SEAS5 performs well in the

FIG. 3. Maps of the correlation coefficient between CHRIPS observations and hindcasts of JAS season total precipitation for 1993–2010. Statistically significant correlation is hatched. The green box indicates the area used throughout to draw ROC curves and reliability diagrams. Contains modified information from [Copernicus Climate Change Service \(2018\)](#).

central regions, and Sys5 shows less skill overall. None of the models show clear skill in the west of the region.

In comparing the reliability diagrams between the four models (Fig. 6), the reliability varies between models and chosen tercile. All four show good reliability at lower forecast probabilities, albeit with some underestimation of forecast probabilities, particularly for GloSea5 and Sys5. For higher probabilities reliability is poor, again likely due to the small number of high probability forecasts as seen on the frequency diagrams. Therefore, only the part of the reliability diagrams where there is a large enough number of forecasts to draw useful conclusions will be considered; that is where forecast probabilities for each tercile are below around 60%. All models show a tendency toward lines with gradients less than the diagonal, indicating that the forecasts are overconfident. Thus, events forecast with low probabilities are observed more frequently than they are forecast, whereas events that are forecast with higher probabilities occur less frequently than that probability. On average, the reliability is slightly better for the above- and below-normal categories with more

of the points nearer the diagonal, compared to the near-normal category, but this varies between models and with forecast probability.

5. Discussion

Using different metrics to assess forecast performance, we have demonstrated that the PRESASS consensus forecasts and dynamical model forecasts of JAS seasonal total precipitation in Sahelian West Africa both show positive skill and reliability. In all forecasts, the near-normal precipitation category shows less skill than the above- and below-normal categories; the near-normal forecasts also have a tendency toward lower reliability, albeit less pronounced. This finding is in agreement with the results of [Mason and Chidzambwa \(2008\)](#) and [Bliefernicht et al. \(2019\)](#). This can be physically explained because in an above- or below-normal year, the physical climate drivers and processes (e.g., El Niño; [Joly and Voldoire 2009](#)) are likely to be working together to give a clear forecast signal. However, in a near-normal year, different processes

FIG. 4. ROC curves for hindcasts of July–September season total precipitation for 1993–2010 from the dynamical models. Contains modified information from [Copernicus Climate Change Service \(2018\)](#).

can be conflicting, resulting in a mixed forecast signal and lower skill in the precipitation outlook. [Kharin and Zwiers \(2003\)](#) also discuss that the near-normal category is less sensitive to perturbations in the signal (i.e., factors that influence the seasonal forecast), reducing forecast skill in this category; this is also seen in the analog forecasting method ([Toth 1989](#)). Furthermore, there are mathematical reasons for reduced skill in the near-normal category. As discussed by [Van Den Dool and Toth \(1991\)](#) and [Kharin and Zwiers \(2003\)](#), since the upper and lower bounds of this middle category are explicitly defined from climatology, whereas the upper (for above normal) and lower (for below normal) bounds of the other categories are unbounded (i.e., a year drier than any previously observed will still be considered as simply below normal).

The frequency diagrams presented ([Fig. 2](#)) show “hedging” away from the below-normal category and

toward the near-normal category, as discussed by [Mason and Chidzambwa \(2008\)](#), who link it to a typically deterministic mindset of forecasters. The RCOF forecasts typically combine multiple dynamical models with the alternative methods. ACMAD adopt a nine-step process ([Foamouhoue 2017](#)) to combine forecasts generated using dynamical models and empirical methods such as variability and trend analysis, persistence composites, and analog techniques. However, there is no objective process to assess the relative skill of these forecast methods, nor to combine the different outputs; the process relies on forecaster expertise and the consensus process. We therefore conclude the need for forecast verification and more objective methods to be used in the RCOF process, consistent with the recommendation made by Mason and Chidzambwa and the [WMO \(2017\)](#). This could include further use of

FIG. 5. Maps showing the area under the ROC curve for each tercile (columns) and each GPC model (rows). Values of greater than 0.5 indicate a skillful forecast. Contains modified information from [Copernicus Climate Change Service \(2018\)](#).

dynamical seasonal forecast models, which have been assessed here and are shown to perform well. Indeed, when comparing one dynamical model with the PRESASS forecasts, the dynamical model shows higher reliability in the below-normal category; otherwise, the two approaches produce similar skill and reliability, with PRESASS showing marginal improvements over the

dynamical model for reliability of above-normal precipitation. It should be noted that, according to the frequency diagrams, some probabilities are rarely given in a forecast, and therefore robustly assessing their reliability is not possible.

The skill and reliability of dynamical model forecasts are positive overall but all four models assessed

FIG. 6. Reliability diagram for hindcasts of July–September season total precipitation for 1993–2010. Contains modified information from Copernicus Climate Change Service (2018).

give too-flat reliability curves, meaning that events forecast with low probabilities are observed to occur at a greater frequency than the forecast would suggest, and when events are forecast with higher probabilities, they occur less frequently than that probability. This tendency to be overconfident may be a consequence of erroneous observations as well as due to model error. There is a low temporal and spatial density of observations in the Sahel, meaning that low likelihood (i.e., rare) events are unlikely to be adequately sampled by the observational record. Improvements to the number and availability of observations might ameliorate these apparent systematic errors in the models, and improvements to weather stations would also facilitate early warnings to avert weather-related disasters (Eze 2018). It may be worth considering postprocessing the model output using statistical techniques, but so far this technique has given mixed results (Barnston and Tippett 2017).

Over the Sahel, the SEAS5 and CFSv2 models perform slightly better in terms of both skill and reliability compared to the GloSea5 model, which in turn performs better than Sys5. However, there is variation in the region and some models show marginal gains over others for one tercile or for a range of forecast probabilities. These models are also regularly updated and model development may lead to changes in the performance. Furthermore, this ranking may depend on the meteorological situation, with some models performing better when particular dynamical drivers prevail as has been found recently in Southern Africa (Hoell and Eischeid 2019). Further investigation might include identification of situations where each model performs well or not, and this can be used to direct model development into the relevant processes. Although increases in horizontal model resolution may not improve seasonal forecast skill (Scaife et al. 2019b), larger ensembles or longer hindcast periods may also be beneficial.

The SEAS5 and CFSv2 models also show greater skill when the correlation coefficient is mapped. However, all models show a tendency to perform less well north of the Sahel (likely due to its aridity) and in the south around the Congo basin (Washington et al. 2013) discussed some possible reasons for this). Furthermore, all models perform less well in the western Sahel and better in the east. This result is also visible to a lesser extent in the ROC maps. The reason that forecast skill is lower in the west compared to the east is unlikely to be related to precipitation amount, but rather be due to the stronger link to large-scale dominant climate drivers and processes (such as ENSO) in the east compared to the west (Joly and Voldoire 2009). Further study of the processes driving precipitation variability on seasonal time scales

in the west would allow identification of future model improvements. However, assessing the quality of the forecast systems will always be limited by the number of years available in the datasets. The frequency diagrams show that some forecast probabilities contain few forecasts, meaning that the results at these probabilities lack robustness.

In addition to the quantitative skill and reliability comparisons assessed here, the changing visual display of PRESASS forecasts over the past 20 years, discussed in section 3a, also impacts this assessment of forecast performance. While refining the methods used to generate forecasts is essential to improve seasonal forecast utility, is also critical to ensure clear and consistent communication. Hence, standardizing the ways in which the forecasts are visually presented would help improve understanding of and confidence in the forecasts, resulting in more consistent interpretation and improved legitimacy (as defined by Hansen et al. 2011). Developing the visual presentation formats in collaboration with forecast users, using a coproduction approach as used in the SCIEPA Project (2018), would also increase their salience (Hansen et al. 2011).

6. Conclusions

This study compares the performance of seasonal forecasts from four dynamical models and from the RCOF forecasts over Sahelian West Africa. Tercile forecasts of above-, near- and below-normal seasonal total precipitation for the July–September season are assessed. The months chosen cover the peak of the rainy season and adequate precipitation is critical for the success of rain-fed agriculture, which is a dominant livelihood in the region. The forecasts are assessed for both reliability and skill (Murphy 1993).

The forecasts generally show positive skill and reliability. In all forecasts the skill is higher in the above- and below-normal terciles compared to the near normal, due partly to inherent mathematical reasons (Van Den Dool and Toth 1991) but also since the dynamical drivers of precipitation may conflict for near-normal years with forecast signals being less clear than in seasons that have typically above- or below-normal precipitation. In the PRESASS forecasts, there is a bias toward low forecast probabilities for the below-normal precipitation category. This may be due to the “hedging” toward the normal category, also identified by Mason and Chidzambwa (2008).

Comparison of four different dynamical models shows that over the Sahel, the CFS and SEAS5 models show greater skill and reliability than the GloSea5 model, with the lowest skill and reliability results for Sys5. However,

it may be that some models perform better in certain meteorological situations or in certain areas, which is a source of potential further investigation. In general, all models perform better in the eastern Sahel compared to the western Sahel, due in part to the reduced influence of ENSO in the west (Joly and Voldoire 2009). Factors other than ENSO may be influencing rainfall variability in the west, such as westerly winds from the Atlantic or mesoscale convective systems, and these may be less well represented. Future work might consider the relative influence of such phenomena.

Consistent with recommendations of previous studies, the results presented here suggest a need for more objective methods to improve the forecasts generated through the RCOF process, including further use of dynamical models and a consistent representation of the forecast products. The skill of RCOF forecasts could be improved by routinely and rigorously assessing the skill and reliability of different forecast systems and techniques, using the approaches adopted in this paper, as a precursor to including the information in the RCOF consensus forecasts. While all dynamical models show positive and significant skill and reliability in the Sahel, there are clear model limitations and their output should be used alongside information regarding forecast skill and reliability. Information regarding forecast uncertainty can be determined from the study of ensemble spread in each model and by utilizing different models. This could be used by the RCOF process to ameliorate the bias toward forecasting too-low probabilities in the below-normal category.

Seasonal forecasting is made complicated due to the probabilistic nature of forecasts and the different dynamical processes that influence the climate on such time scales. Expertise is required to distill and harmonize information from a range of models and methods and create audience-appropriate forecast products. Building on the work in this study, there is a need for further and more regular assessments of the quality of seasonal forecasts in the region. It is recommended that the verification procedure becomes a regular and routine exercise to better inform future PRESASS forums and better understand the performance of seasonal forecasts in the region. Future work should also focus on how users interpret and apply forecasts to illuminate where improvements have greatest societal impact. This would also allow identification of challenges to the uptake of seasonal forecasts (Ouedraogo et al. 2018). Ensuring that forecasts are user relevant might lead to alternative quantities being forecast, although they would need to be robustly assessed for forecast quality to ensure that any information provided is credible.

Acknowledgments. The authors would like to acknowledge funding from the U.K. Department for International Development (DFID), as part of the Adaptive Social Protection: Information for enhanced Resilience (ASPIRE) Project under Weather and Climate Information Services for Africa (WISER). We thank Issa Lele, Richard Graham, and Andrew Colman for useful discussions during the drafting of this paper. We are grateful to the two anonymous reviewers for their insight and comments.

REFERENCES

- Ado, M., J. Leshan, P. Savadogo, S. Kollin Koivogui, and J. Chrisostom Pesha, 2018: Households' vulnerability to climate change: Insights from a farming community in Aguié district of Niger. *J. Environ. Earth Sci.*, **8**, 2224–3216.
- Barnston, A., and M. Tippett, 2017: Do statistical pattern corrections improve seasonal climate predictions in the North American Multimodel Ensemble models? *J. Climate*, **30**, 8335–8355, <https://doi.org/10.1175/JCLI-D-17-0054.1>.
- Becker, A., P. Finger, A. Meyer-Christoffer, B. Rudolf, K. Schamm, U. Schneider, and M. Ziese, 2013: A description of the global land-surface precipitation data products of the global precipitation climatology centre with sample applications including centennial (trend) analysis from 1901-present. *Earth Syst. Sci. Data*, **5**, 71–99, <https://doi.org/10.5194/essd-5-71-2013>.
- Becker, E., and H. Van Den Dool, 2016: Probabilistic seasonal forecasts in the North American Multimodel Ensemble: A baseline skill assessment. *J. Climate*, **29**, 3015–3026, <https://doi.org/10.1175/JCLI-D-14-00862.1>.
- Bliefernicht, J., M. Waongo, S. Salack, J. Seidel, P. Laux, and H. Kunstmann, 2019: Quality and value of seasonal precipitation forecasts issued by the West African regional climate outlook forum. *J. Appl. Meteor. Climatol.*, **58**, 621–642, <https://doi.org/10.1175/JAMC-D-18-0066.1>.
- Boyd, E., R. J. Cornforth, P. J. Lamb, A. Tarhule, M. I. Lélé, and A. Brouder, 2013: Building resilience to face recurring environmental crisis in African Sahel. *Nat. Climate Change*, **3**, 631–637, <https://doi.org/10.1038/nclimate1856>.
- Braman, L. M., M. K. van Aalst, S. J. Mason, P. Suarez, Y. Ait-Chellouche, and A. Tall, 2013: Climate forecasts in disaster management: Red Cross flood operations in West Africa, 2008. *Disasters*, **37**, 144–164, <https://doi.org/10.1111/j.1467-7717.2012.01297.x>.
- Buizza, R., and M. Leutbecher, 2015: The forecast skill horizon. *Quart. J. Roy. Meteor. Soc.*, **141**, 3366–3382, <https://doi.org/10.1002/qj.2619>.
- Chinwendu, O. G., S. O. E. Sadiku, A. O. Okhimamhe, and J. Eichie, 2017: Households vulnerability and adaptation to climate variability induced water stress on downstream Kaduna River Basin. *Amer. J. Climate Change*, **6**, 247–267, <https://doi.org/10.4236/ajcc.2017.62013>.
- Cissé, G., and Coauthors, 2016: Vulnerabilities of water and sanitation at households and community levels in face of climate variability and change: Trends from historical climate time series in a West African medium-sized town. *Int. J. Global Environ. Issues*, **15**, 81, <https://doi.org/10.1504/IJGENVI.2016.074360>.
- Cook, K. H., and E. K. Vizy, 2006: Coupled model simulations of the West African monsoon system: Twentieth-and twenty-first-century simulations. *J. Climate*, **19**, 3681–3703, <https://doi.org/10.1175/JCLI3814.1>.

- Copernicus Climate Change Service, 2018: Climate data store. Accessed 15 January 2019, <https://cds.climate.copernicus.eu/#/home>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- Dilley, M., and R. Kolli, 2017: Draft discussion paper on the development of objective regional sub-seasonal to seasonal forecasts in Africa, Asia-Pacific and South America. WMO, accessed 2 March 2020, <http://www.wmo.int/pages/prog/wcp/wcasp/linkedfiles/Draftdiscussionpaperobjectivesseasonalforecastsawayforward20180831.docx>.
- Dinku, T., C. Funk, P. Peterson, R. Maidment, T. Tadesse, H. Gadain, and P. Ceccato, 2018: Validation of the CHIRPS satellite rainfall estimates over eastern Africa. *Quart. J. Roy. Meteor. Soc.*, **144**, 292–312, <https://doi.org/10.1002/qj.3244>.
- Doblas-Reyes, F. J., J. García-Serrano, F. Lienert, A. P. Biescas, and L. R. L. Rodrigues, 2013: Seasonal climate predictability and forecasting: Status and prospects. *Wiley Interdiscip. Rev.: Climate Change*, **4**, 245–268, <https://doi.org/10.1002/wcc.217>.
- Druyan, L., and M. Fulakeza, 2018: Downscaling atmosphere-ocean global climate model precipitation simulations over Africa using bias-corrected lateral and lower boundary conditions. *Atmosphere*, **9**, 493, <https://doi.org/10.3390/atmos9120493>.
- Eze, B. U., 2018: Climate change, population pressure and agricultural livelihoods in the West African Sahel (special reference to northern Nigeria): A review. *Pyrex J. Ecol. Nat. Environ.*, **3**, 1–7.
- Foamouhoue, A. K., 2017: ACMAD: Current status of operations of PRESASS, PRESAGG & PRESAC. *WMO International Workshop on Global Review of Regional Climate Outlook Forums*, Ecuador, WMO, 16 pp., http://www.wmo.int/pages/prog/wcp/wcasp/meetings/documents/rcofs2017/presentations/7_PRESASS_PRESAGG_PRESAC_Presentations.pdf.
- Funk, C., and Coauthors, 2015: The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Sci. Data*, **2**, 150066, <https://doi.org/10.1038/sdata.2015.66>.
- Hansen, J., S. Mason, L. Sun, and A. Tall, 2011: Review of seasonal climate forecasting for agriculture in sub-Saharan Africa. *Exp. Agric.*, **47**, 205–240, <https://doi.org/10.1017/S0014479710000876>.
- Hoell, A., and J. Eischeid, 2019: On the interpretation of seasonal Southern Africa precipitation prediction skill estimates during austral summer. *Climate Dyn.*, **53**, 6769–6783, <https://doi.org/10.1007/s00382-019-04960-5>.
- Joly, M., and A. Voldoire, 2009: Influence of ENSO on the West African monsoon: Temporal aspects and atmospheric processes. *J. Climate*, **22**, 3193–3210, <https://doi.org/10.1175/2008JCLI2450.1>.
- Kharin, V., and F. Zwiers, 2003: Improved seasonal probability forecasts. *J. Climate*, **16**, 1684–1701, [https://doi.org/10.1175/1520-0442\(2003\)016<1684:ISPF>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<1684:ISPF>2.0.CO;2).
- Landman, W., A. G. Barnston, C. Vogel, and J. Savy, 2019: Use of El Niño–Southern Oscillation related seasonal precipitation predictability in developing regions for potential societal benefit. *Int. J. Climatol.*, **39**, 5327–5337, <https://doi.org/10.1002/joc.6157>.
- MacLachlan, C., and Coauthors, 2015: Global Seasonal forecast system version 5 (GloSea5): A high-resolution seasonal forecast system. *Quart. J. Roy. Meteor. Soc.*, **141**, 1072–1084, <https://doi.org/10.1002/qj.2396>.
- Maidment, R. I., and Coauthors, 2017: A new, long-term daily satellite-based rainfall dataset for operational monitoring in Africa. *Sci. Data*, **4**, 170063, <https://doi.org/10.1038/sdata.2017.63>.
- Mason, S., and S. Chidzambwa, 2008: Verification of African RCOF forecasts. World Meteorological Organization RCOF Review Tech. Rep. 09-02, 26 pp., <https://doi.org/10.7916/D85T3SB0>.
- Met Office, 2019: Adaptive social protection: Information for enhanced resilience (ASPIRE). Met Office, accessed 22 July 2019, <https://www.metoffice.gov.uk/about-us/what/working-with-other-organisations/international/projects/wiser/aspire>.
- Murphy, A., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281–293, [https://doi.org/10.1175/1520-0434\(1993\)008<0281:WIAGFA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2).
- Nicholson, S. E., 2013: The West African Sahel: A review of recent studies on the rainfall regime and its interannual variability. *ISRN Meteor.*, **2013**, 1–32, <https://doi.org/10.1155/2013/453521>.
- Okoro, U., W. Chen, T. Chineke, and O. Nwofor, 2017: Anomalous atmospheric circulation associated with recent West African monsoon rainfall variability. *J. Geosci. Environ. Prot.*, **5**, 1–27, <https://doi.org/10.4236/gep.2017.512001>.
- Ouedraogo, I., N. Diouf, M. Ouedraogo, O. Ndiaye, and R. Zougmore, 2018: Closing the gap between climate information producers and users: Assessment of needs and uptake in Senegal. *Climate*, **6**, 13, <https://doi.org/10.3390/cli6010013>.
- Poméon, T., D. Jackisch, and B. Diekkrüger, 2017: Evaluating the performance of remotely sensed and reanalysed precipitation data over West Africa using HBV light. *J. Hydrol.*, **547**, 222–235, <https://doi.org/10.1016/j.jhydrol.2017.01.055>.
- QGIS Development Team, 2018: QGIS geographic information system. Accessed 30 July 2018, <http://qgis.osgeo.org>.
- Raoult, B., C. Bergeron, A. L. Alos, J.-N. Thépaut, and D. Dee, 2017: Climate service develops user-friendly data store. ECMWF Newsletter, No. 151, ECMWF, Reading, United Kingdom, 22–27, <https://doi.org/10.21957/p3c285>.
- Rees, D., 2001: *Essential Statistics*. Chapman & Hall/CRC, 361 pp.
- Rodríguez-Fonseca, B., and Coauthors, 2015: Variability and predictability of West African droughts: A review on the role of sea surface temperature anomalies. *J. Climate*, **28**, 4034–4060, <https://doi.org/10.1175/JCLI-D-14-00130.1>.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, <https://doi.org/10.1175/2010BAMS3001.1>.
- , and Coauthors, 2011: NCEP Climate Forecast System version 2 (CFSv2) 6-hourly products. Accessed 22 November 2018, <https://doi.org/10.5065/D61C1TXF>.
- , and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, <https://doi.org/10.1175/JCLI-D-12-00823.1>.
- Scaife, A. A., and Coauthors, 2019a: Tropical rainfall predictions from multiple seasonal forecast systems. *Int. J. Climatol.*, **39**, 974–988, <https://doi.org/10.1002/joc.5855>.
- , and Coauthors, 2019b: Does increased atmospheric resolution improve seasonal climate predictions? *Atmos. Sci. Lett.*, **20**, e922, <https://doi.org/10.1002/asl.922>.
- SCIPEA Project, 2018: Strengthening Climate Information Partnerships—East Africa (SCIPEA). Accessed 15 March 2019, <https://www.metoffice.gov.uk/about-us/what/working-with-other-organisations/international/projects/wiser/scipea>.
- Semazzi, F., 2011: Framework for climate services in developing countries. *Climate Res.*, **47**, 145–150, <https://doi.org/10.3354/cr00955>.

- Sheen, K. L., D. M. Smith, N. J. Dunstone, R. Eade, D. P. Rowell, and M. Vellinga, 2017: Skilful prediction of Sahel summer rainfall on inter-annual and multi-year timescales. *Nat. Commun.*, **8**, 14966, <https://doi.org/10.1038/ncomms14966>.
- Stockdale, T. N., D. L. Anderson, J. O. S. Alves, and M. A. Balmaseda, 1998: Global seasonal rainfall forecasts using a coupled ocean–atmosphere model. *Nature*, **392**, 370–373, <https://doi.org/10.1038/32861>.
- Sultan, B., and S. Janicot, 2003: The West African monsoon dynamics. Part II: The “preonset” and “onset” of the summer monsoon. *J. Climate*, **16**, 3407–3427, [https://doi.org/10.1175/1520-0442\(2003\)016<3407:TWAMDP>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<3407:TWAMDP>2.0.CO;2).
- , C. Baron, M. Dingkuhn, B. Sarr, and S. Janicot, 2005: Agricultural impacts of large-scale variability of the West African monsoon. *Agric. For. Meteorol.*, **128**, 93–110, <https://doi.org/10.1016/j.agrformet.2004.08.005>.
- Tall, A., S. J. Mason, M. van Aalst, P. Suarez, Y. Ait-Chellouche, A. A. Diallo, and L. Braman, 2012: Using seasonal climate forecasts to guide disaster management: The Red Cross experience during the 2008 West Africa floods. *Int. J. Geophys.*, **2012**, 986016, <https://doi.org/10.1155/2012/986016>.
- Toth, Z., 1989: Long-range weather forecasting using an analog approach. *J. Climate*, **2**, 594–607, [https://doi.org/10.1175/1520-0442\(1989\)002<0594:LRWFUA>2.0.CO;2](https://doi.org/10.1175/1520-0442(1989)002<0594:LRWFUA>2.0.CO;2).
- Troccoli, A., M. Harrison, D. L. Anderson, and S. J. Mason, 2008: *Seasonal Climate: Forecasting and Managing Risk*. Vol. 82. Springer Science & Business Media, 467 pp.
- Tschakert, P., 2007: Views from the vulnerable: Understanding climatic and other stressors in the Sahel. *Global Environ. Change*, **17**, 381–396, <https://doi.org/10.1016/j.gloenvcha.2006.11.008>.
- Van den Dool, H., 2007: *Empirical Methods in Short-Term Climate Prediction*. Oxford University Press, 215 pp.
- , and Z. Toth, 1991: Why do forecasts for “near normal” often fail? *Wea. Forecasting*, **6**, 76–85, [https://doi.org/10.1175/1520-0434\(1991\)006<0076:WDFNO>2.0.CO;2](https://doi.org/10.1175/1520-0434(1991)006<0076:WDFNO>2.0.CO;2).
- Vellinga, M., A. Arribas, and R. Graham, 2013: Seasonal forecasts for regional onset of the West African monsoon. *Climate Dyn.*, **40**, 3047–3070, <https://doi.org/10.1007/s00382-012-1520-z>.
- , M. Roberts, P. L. Vidale, M. S. Mizielski, M.-E. Demory, R. Schiemann, J. Strachan, and C. Bain, 2016: Sahel decadal rainfall variability and the role of model horizontal resolution. *Geophys. Res. Lett.*, **43**, 326–333, <https://doi.org/10.1002/2015GL066690>.
- Walker, D. P., C. E. Birch, J. H. Marsham, A. A. Scaife, R. J. Graham, and Z. T. Segele, 2019: Skill of dynamical and GHACOF consensus seasonal forecasts of East African rainfall. *Climate Dyn.*, **53**, 4911–4935, <https://doi.org/10.1007/s00382-019-04835-9>.
- Washington, R., R. James, H. Pearce, W. M. Pokam, and W. Moufouma-Okia, 2013: Congo Baisin rainfall climatology: Can we believe the climate models? *Philos. Trans. Roy. Soc. London*, **368B**, 20120296, <https://doi.org/10.1098/rstb.2012.0296>.
- White, C. J., and Coauthors, 2017: Potential applications of subseasonal-to-seasonal (s2s) predictions. *Meteor. Appl.*, **24**, 315–325, <https://doi.org/10.1002/met.1654>.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.
- World Meteorological Organisation, 2017: WMO workshop on global review of regional climate outlook forums. Accessed 19 November 2019, http://www.wmo.int/pages/prog/wcp/wcasp/meetings/workshop_rcofs.php.
- , 2018: Guidance on verification of operational seasonal climate forecasts. Accessed 2 March 2020, https://library.wmo.int/doc_num.php?explnum_id=4886.