

# Water Resources Research<sup>®</sup>



## RESEARCH ARTICLE

10.1029/2021WR031480

### Key Points:

- The European Centre for Medium-Range Weather Forecasts is the most skillful model for precipitation in Africa
- Weighted multi-model mean provides the highest performance compared to individual models for precipitation and forecasting droughts
- Forecast skill of monthly meteorological drought events at lead 1-month is modest

### Supporting Information:

Supporting Information may be found in the online version of this article.

### Correspondence to:

S. H. Gebrechorkos,  
S.H.Gebrechorkos@soton.ac.uk

### Citation:

Gebrechorkos, S. H., Pan, M., Beck, H. E., & Sheffield, J. (2022). Performance of state-of-the-art C3S European seasonal climate forecast models for mean and extreme precipitation over Africa. *Water Resources Research*, 58, e2021WR031480. <https://doi.org/10.1029/2021WR031480>

Received 26 OCT 2021

Accepted 27 FEB 2022

## Performance of State-of-the-Art C3S European Seasonal Climate Forecast Models for Mean and Extreme Precipitation Over Africa

S. H. Gebrechorkos<sup>1</sup> , M. Pan<sup>2,3</sup> , H. E. Beck<sup>2,4</sup>, and J. Sheffield<sup>1,2</sup> 

<sup>1</sup>School of Geography and Environmental Sciences, University of Southampton, Southampton, UK, <sup>2</sup>Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA, <sup>3</sup>Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA, <sup>4</sup>European Commission, Joint Research Centre (JRC), Ispra, Italy

**Abstract** Seasonal hydrological forecasts at high spatial and temporal resolution can help manage water resources and mitigate impacts of extreme events but are dependent on skillful and operational seasonal forecasts from climate models. In this study, we evaluate precipitation forecasts from five operational climate models with a potential to drive hydrological forecasts: European Centre for Medium-Range Weather Forecasts (ECMWF), UK Met Office (UK-Met), Météo France, Deutscher Wetterdienst, and Centro Euro-Mediterraneo sui Cambiamenti Climatici. The Multi-Source Weighted-Ensemble Precipitation is used as a reference data set to evaluate the model skill. The performance of individual models is evaluated on daily, weekly, monthly, seasonal, and climatological periods, and for selected target months, lead-times and drought events, and compared to unweighted and skill-weighted multi-model ensemble mean forecast. For all models, the lead 1-month forecast can replicate the climatological mean, monthly mean, and monthly anomaly precipitation, although much of this skill originates from the first week of the forecast. The skill drops rapidly for lead 2-month and longer and is highest in drier regions and seasons. The forecast skill of monthly meteorological drought events at lead 1-month is modest. All models represent the monthly variation in the length of wet and dry spell days at lead 1-month, but the skill is weak for heavy and very heavy precipitation days. ECMWF is found to be the most skillful model, followed by the UK-Met, although the multi-model weighted average provides the highest performance compared to individual models and the un-weighted multi-model mean.

## 1. Introduction

Extreme hydrological events, such as droughts and floods have a large societal impact if mitigation and adaptation measures are not developed and implemented (Kundzewicz & Kaczmarek, 2000). This is particularly the case in many developing regions where there is high dependency of livelihoods and economies on rain-fed agriculture and limitations to adaptive capacity (Niang et al., 2014); for example, in Africa, more than 80% of the population depends on agriculture for their livelihood (FAO, 2014). Early warning systems that combine forecasts of climate and hydrology with risk reduction measures can reduce the social and economic impacts of extreme climate and hydrological events (Wanders & Wood, 2016), and particularly so in at-risk developing countries (Hansen et al., 2011; Pozzi et al., 2013; Sheffield et al., 2014). For example, seasonal forecasts of low precipitation or higher risk of extreme events could enable farmers to make adaptive choices on types of crops, technology investment, and labor usage to minimize the impacts (Hansen et al., 2011).

During the last few years, several national, regional, and global operational and experimental hydrological forecasting systems, which combine short-term to seasonal forecasts of meteorological variables with hydrological models, have been developed to provide risk information and alerts to help reduce the impacts of floods and droughts. Some examples of continental to global and national systems are the Global Flood Awareness System (Alfieri et al., 2013), the African Flood and Drought Monitor (Sheffield et al., 2014), the European Flood Alert System (Thielen et al., 2009), the pre-operational Pan-European Multimodel Seasonal Hydrological Forecasting System (Samaniego et al., 2019; Wanders et al., 2019), and World-wide HYPE (Arheimer et al., 2020).

The utility of such systems relies on the accuracy and appropriateness of the hydrological models used to provide the risk indicators such as drought and flood indices but fundamentally on the quality and availability of operational meteorological forecasts to drive these models (Wanders & Wood, 2016). There has been continued

© 2022. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

development and production of climate forecasts from national and international agencies at ever higher resolution and with incremental improvements in skill (Kirtman et al., 2013; Wanders et al., 2019). Of note is the push for extended skill at sub-seasonal time scales (beyond 2 weeks but less than seasonal) (Lang et al., 2020) that has important implications for a range of applications, including disaster risk reduction and for different economic sectors (Morss et al., 2008). Criteria for choosing a set of seasonal climate models are based on their resolution (space and time), availability of hindcasts (historic forecasts) to evaluate their skill, the skill of the models for the particular application, and availability of forecasts operationally. Ideally, a climate model will provide data for precipitation and temperature at daily resolution at the least to capture extreme meteorological events and to drive hydrological models, have several decades of hindcast data to allow the skill to be evaluated over a reasonable number of years and to capture some extreme events, and crucially be available operationally to allow the development of early warning in practice. Higher spatial resolution does not necessarily guarantee better forecast skill, but can potentially better resolve some scale-dependent processes (e.g., orographic effects on precipitation) that are important locally. Higher resolution also reduces the need for spatial and temporal downscaling for input into hydrological models that can be used to provide local-scale and more relevant hydrological information for water management and other water-dependent sectors (Samaniego et al., 2017). This may further reduce the uncertainties in the forecasts (Thober et al., 2015; Wanders et al., 2019; Yuan et al., 2015).

Currently, there are several international initiatives that provide multi-model seasonal forecasts at a different spatial and temporal resolution that can be utilized for early warning. The availability of multiple models allows for model inter-comparison and account of the model structural uncertainties that are inherent in forecasting. These initiatives include the North American Multi-Model Ensemble (NMME, Kirtman et al., 2013; <https://www.cpc.ncep.noaa.gov/products/NMME/>) and the Copernicus Climate Change Service (C3S) multi-system seasonal forecast (<https://cds.climate.copernicus.eu/>). The first phase of the NMME provides monthly data, and so is less suitable for forecasts of sub-seasonal metrics (e.g., probability of dry spells, the start of the rainy season) and for hydrological forecasting as hydrological models generally require daily or finer climate data. The latest Phase 2 provides sub-daily information that can resolve the key intra-seasonal events and be used to drive hydrological models without any downscaling but currently does not do so operationally. On the other hand, the European C3S provides sub-daily information for a hindcast period (1993–2016) and operational forecasts out to 180–215 days depending on the model. Other initiatives such as the Subseasonal-to-Seasonal prediction project (F. Vitart et al., 2017; Frédéric Vitart & Robertson, 2018) provide access to forecasts from a range of climate models, but currently do not provide a consistent set of operational forecasts.

The objective of this study is to evaluate the performance of the C3S seasonal forecast models over Africa for a range of climate indices and to understand their potential for driving seasonal hydrological forecasts. We used daily precipitation and evaluated these against a high-resolution observational data product. The C3S models evaluated are the European Centre for Medium-Range Weather Forecasts (ECMWF: version 5), UK Met Office (UK-Met: version 14), Météo France (Meteo-France: version 6), Deutscher Wetterdienst (DWD: version 2), and Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC: version 3). The evaluation was done on multiple time scales and different lead times, with a focus on differences between the wet and dry seasons, and at different spatial scales, from the pixel level for the whole continent to regional scales. We evaluated monthly and weekly averages as well as a set of climate extreme indices using daily totals, which provides a novel and holistic view of the performance of the models across time scales. In addition, we evaluated the ensemble mean of individual models, the un-weighted multi-model mean, and the weighted mean of all models using different weighting schemes. Finally, the most skillful models are identified and a weighted average of all models developed with potential for driving hydrological models for early warning purposes and for analysis of meteorological droughts.

## 2. Material and Methods

### 2.1. Reference Datasets

To evaluate the skill of the seasonal forecast models, we used an observation-based gridded reference data set: the Multi-Source Weighted-Ensemble Precipitation (MSWEP). MSWEP (version 2.2) is a high-resolution global precipitation data developed by merging multiple datasets based on gauge, reanalysis, and satellite estimates (Beck et al., 2019; Beck, van Dijk et al., 2017). The second version of MSWEP (<http://www.gloh2o.org/mswep/>) is a unique precipitation data set in terms of its spatial (0.1°) and temporal resolution (3-hourly), and the inclusion

**Table 1**  
*Description of the Copernicus Climate Change Service (C3S) Seasonal Forecast Models and Their Attributes*

Model	Time (days)	Forecast initial condition	Hindcast initial condition	Forecast ensemble size	Hindcast ensemble size	Hindcast period	Spatial resolution
ECMWF	215	1st of month	1st of month	51	25	1981–2016	T <sub>CO</sub> 319/L91 (~0.32°–0.72°)
UK-Met	215	Each day of month	1st, 9th, 17th, 25th	2	7	1993–2016	N216/L85 0.83° × 0.56°
Meteo-France	215	1st of month	1st of month	25	24	1993–2016	TL359/L91 (~0.5°)
DWD	180	1st of month	1st of month	50	30	1993–2017	T127 (~100 km)
CMCC	180	1st of month	1st of month	50	40	1993–2016	~1°

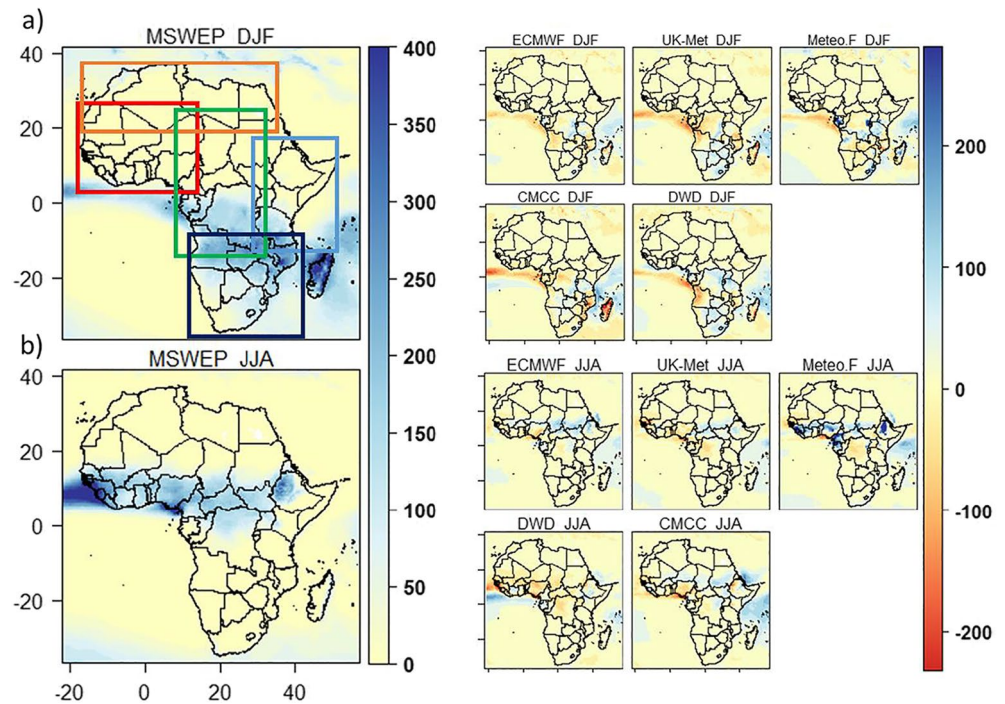
of multiple datasets; station data such as from Global Historical Climatology Network-Daily, WorldClim, Global Precipitation Climatology Centre (GPCC), and Global Summary of the Day; satellite-based products including the Tropical Rainfall Measuring Mission Multisatellite Precipitation Analysis (3B42RT), Climate Prediction Center morphing technique, Gridded Satellite, Global Satellite Mapping of Precipitation; and reanalysis data of the European Centre for Medium-Range Weather Forecasts interim reanalysis (ECMWF ERA-Interim) and Japanese 55-year Reanalysis. In addition to the inclusion of more than 77,000 daily observations from field-based meteorological stations, MSWEP also applies different bias correction techniques to improve the frequency of precipitation and uses river discharge observations (from about 14,000 stations globally) to bias correct terrestrial precipitation, which makes it unique compared with other precipitation products (Beck et al., 2019). MSWEP has been widely used in global and regional scale hydroclimate studies such as the analysis of water resources, rainfall variability, lake dynamics, soil moisture and evaporation, and ecohydrological modeling (Chen & Dirmeyer, 2016; Chen et al., 2017; Liu et al., 2016; Martens et al., 2017; Schellekens et al., 2017). In a recent study (Beck, Vergopolan, et al., 2017) comparing 22 global and semi-global precipitation products against station data, MSWEP was found to be one of the best performing datasets. For Africa, Awange et al. (2019) show that MSWEP has better performance for water storage flux, climate impacts, and river discharge assessment compared to other precipitation products such as GPCC and shows a similar skill for other aspects such as extremes.

## 2.2. Seasonal Forecast Models

The five seasonal forecast models (Table 1) were evaluated for the hindcast period of 1993–2016 using data downloaded from the C3S website (<https://cds.climate.copernicus.eu/cdsapp#!/dataset/seasonal-original-single-levels?tab=form>). All models provide global, seasonal (180–215 days; 6–7 months) forecasts of precipitation on a daily time scale, at a spatial resolution of 1.0°, and initialized on the first day of the month. The native spatial resolution of some of the models is higher (e.g., Meteo-France is 0.5°), but the publicly available data provided by the CDS is on a reduced, but consistent, 1.0-degree grid. To evaluate the skill of the models on a grid-scale, the climate model data are bilinearly interpolated to match the resolution of the reference data set (MSWEP, 0.1°).

## 2.3. Evaluation Metrics

The models were evaluated for precipitation at the grid cell level and summarized at regional and at multiple time scales (daily to climatology). We evaluated model performance for the ensemble mean for weekly, monthly and seasonal means to provide an overview of model performance for mean precipitation. For the drought and daily extremes, we evaluated the ensemble of forecasts to provide a probabilistic assessment and avoid smoothing of the extreme values from ensemble averaging. The models were evaluated for different lead-times (months and weeks since initialization), different target months, and separately for the wet and dry seasons, because of the strong seasonality over the continent. First, the mean monthly seasonal cycle (climatology) was computed for each model to assess the overall forecast bias at different lead times. Second, how well the forecasts follow the observed weekly and monthly variability in terms of the absolute and anomaly (after subtraction of the seasonal climatology) values was quantified.



**Figure 1.** Climatological mean (1993–2016) of precipitation (mm/month) for December–February (DJF) (a) and June–August (JJA) (b) seasons from Multi-Source Weighted-Ensemble Precipitation (MSWEP). For precipitation from European Centre for Medium-Range Weather Forecasts (ECMWF), UK Met Office (UK-Met), Meteo-France (Meteo-F), Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), and Deutscher Wetterdienst (DWD), the absolute difference of the lead-1 month forecast from MSWEP is displayed. The blue, green, red, orange, and dark polygons (a) represent the East, Central, West, North, and Southern Africa, respectively.

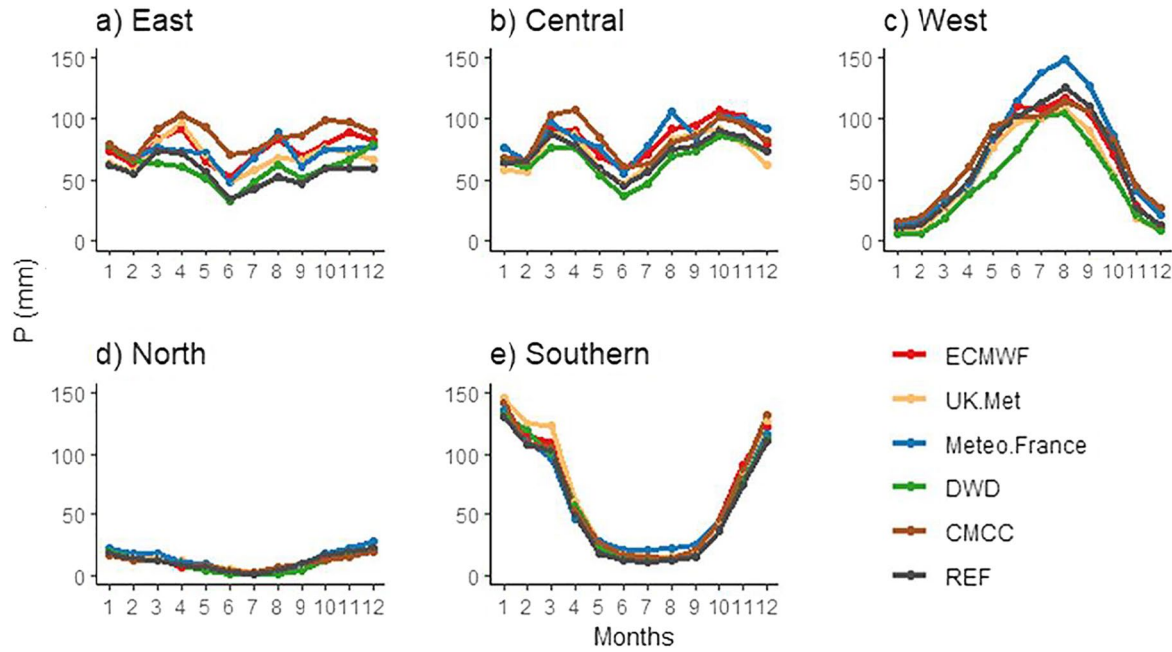
This was complemented by an analysis of the performance of the models in predicting meteorological drought and pluvial events at the monthly scale based on the Standardised Precipitation Index (SPI). The SPI (McKee et al., 1993) is a transformation of the monthly precipitation time series into a standardized form that is often used to identify relatively wet or dry meteorological conditions. Finally, the performance of the models for a set of daily precipitation metrics was evaluated based on the Expert Team on Climate Change Detection and Indices (ETCCDI) climate change indices (Karl et al., 1999): length of dry spell (CDD; numbers of consecutive days with daily precipitation <1.0 mm) and wet spell (CWD; numbers of consecutive days with daily precipitation >1.0 mm) days and numbers of heavy (>10 mm) and very heavy (>20 mm) precipitation days. The ETCCDI indices were chosen as they are a standardized set, which allows direct comparison spatially and with other studies, with the trade-off that the specific thresholds (e.g., for heavy precipitation days) may not be as relevant for very dry or very wet regions. The extreme indices are computed for each ensemble instead of the ensemble mean as averaging will tend to smooth out the extremes. These indices are aggregated to a monthly time scale for the evaluation, given the difficulty in predicting individual extreme events at time scales beyond a few days, but assuming that there is some potential to forecast the probability of occurrence of extremes.

To assess the agreement of individual models with the reference data, we used different statistical methods: such as the Pearson correlation coefficient (CC), bias, and Root Mean Square Error (RMSE) for the monthly and weekly absolute and anomaly time-series and the daily extremes.

$$CC = \frac{\sum_{i=1}^N (O_i - \bar{O}) \cdot (M_i - \bar{M})}{\sqrt{\sum_{i=1}^N (O_i - \bar{O})^2} \cdot \sqrt{\sum_{i=1}^N (M_i - \bar{M})^2}} \quad (1)$$

$$\text{Bias} = \frac{\sum (M_i - O_i)}{N} \quad (2)$$





**Figure 2.** Mean seasonal cycle of precipitation from Multi-Source Weighted-Ensemble Precipitation (MSWEP) (REF) and the lead-1 forecasts of the five models averaged over (a) East, (b) Central, (c) West, (d) North, and (e) Southern Africa.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (O_i - M_i)^2}{N}} \quad (3)$$

To assess the skill of individual models in forecasting meteorological droughts computed using the SPI, we used the Probability Of Detection (POD, Equation 4), False Alarm Ratio (FAR, Equation 5) and Brier Score (BS, Equation 6).

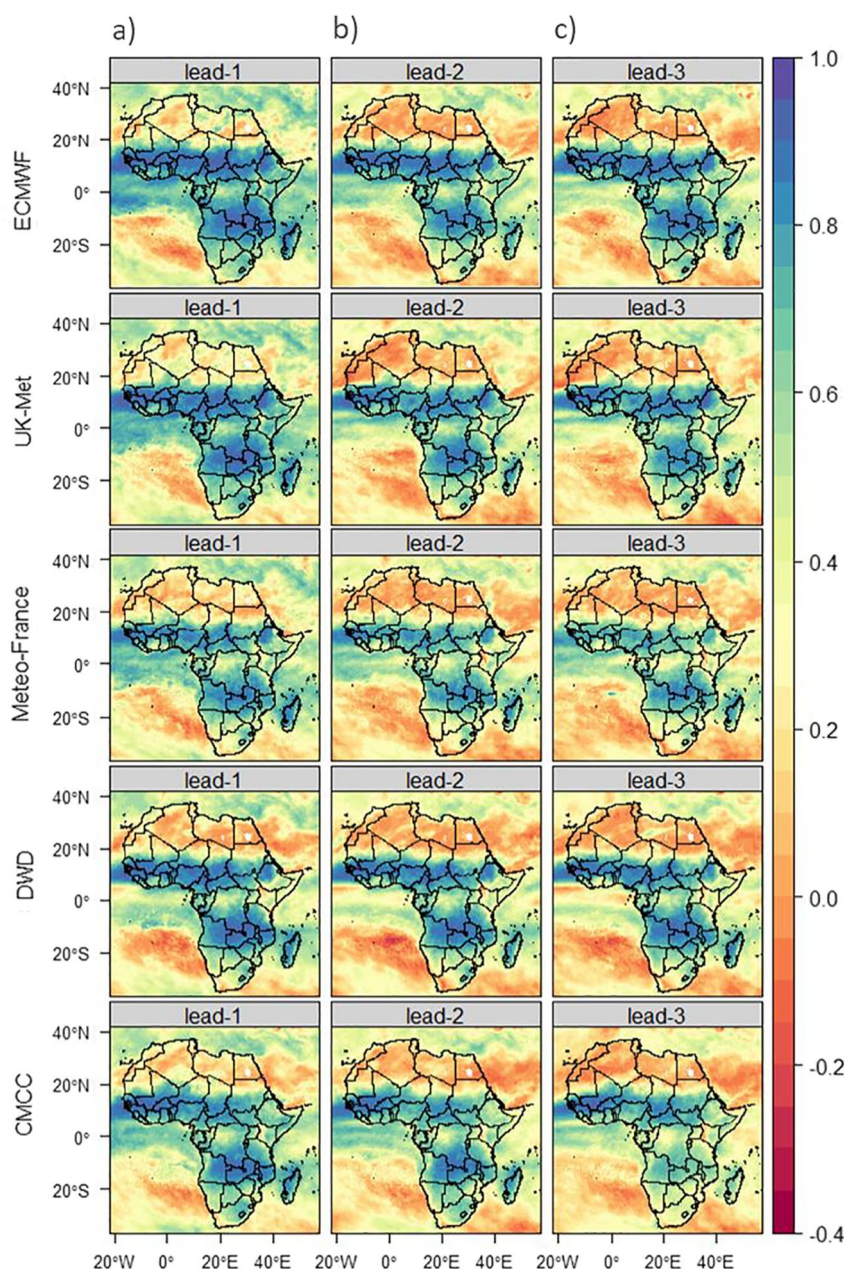
$$POD = \frac{\text{Hits}}{\text{Hits} + \text{Misses}} \quad (4)$$

$$FAR = \frac{\text{False Alarms}}{\text{Hits} + \text{False Alarms}} \quad (5)$$

$$BS = \frac{1}{n} \sum_{i=1}^n (M_i - O_i)^2 \quad (6)$$

where  $O$  is the observation,  $M$  is the model and  $N$  is the number of data pairs (e.g., months). *Hits* are the number of drought events that are correctly identified by the model. *Misses* are the number of drought events that were observed but not forecast by the model. *False Alarms* are the number of drought events that are forecast but were not observed.

We also assessed the unweighted and weighted multi-model means. The weighted multi-model mean (WMMM) is computed based on the performance of individual models, where performance is represented by either CC (Equation 1) or partial correlation (PC), at each grid cell. PC accounts for the correlation between models and therefore the shared or redundant information that is otherwise over-emphasized (Glowienka-Hense et al., 2020). As the results using the CC and PC to weight the models are very similar (except in East Africa, where the PC based results are better), we only show the WMMM results based on the PC. The PC is computed using the ppcor package in R (Kim, 2015).



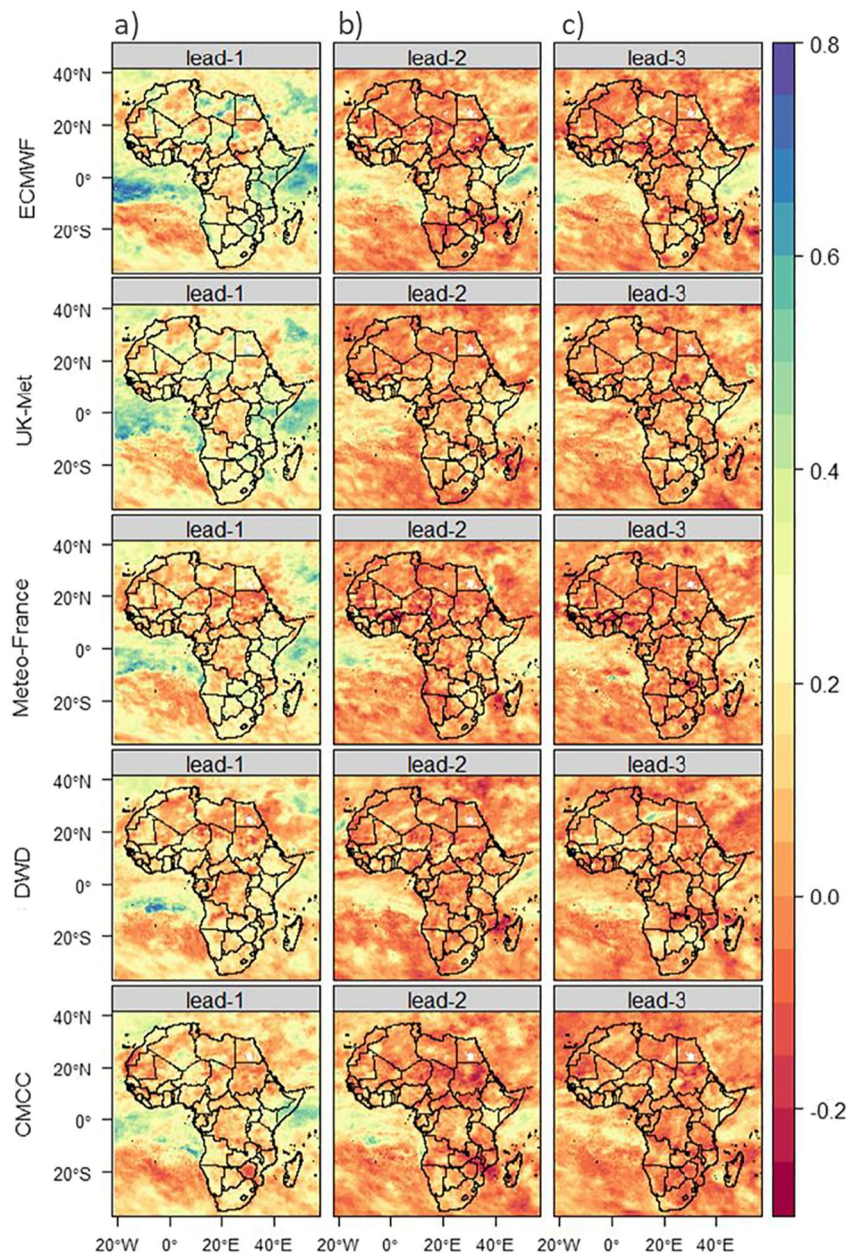
**Figure 3.** Temporal correlation of monthly precipitation between Multi-Source Weighted-Ensemble Precipitation and (a) lead-1, (b) lead-2, and (c) lead-3 month forecasts for the five climate models.

### 3. Results

#### 3.1. Seasonal Climatology of Precipitation

Figure 1 shows maps of observed mean seasonal precipitation and the lead-1 forecasts for December-February (DJF) and June-August (JJA). Figure 1a also shows the broad regions that are referred below and are used to summarize the results spatially. In East Africa, the lead-1 forecast from Meteo-France, DWD, and CMCC show a positive bias (up to 160 mm/month) during the DJF season in large parts of Tanzania and southwest parts of Kenya. On the other hand, the climatological precipitation of the southern and south-eastern parts of Tanzania is underestimated (up to 150 mm/month) by all models. In Ethiopia, Sudan, Eritrea, and Somalia all models except Meteo-France are within about 50 mm/month of the observed climatology, which is expected as DJF is the dry season of these relatively dry countries.

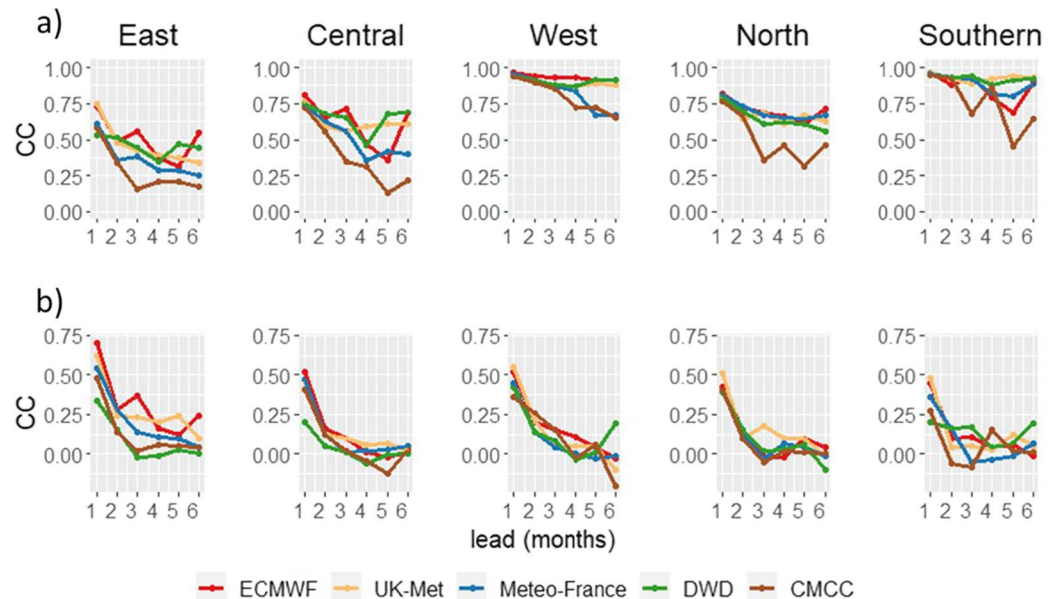




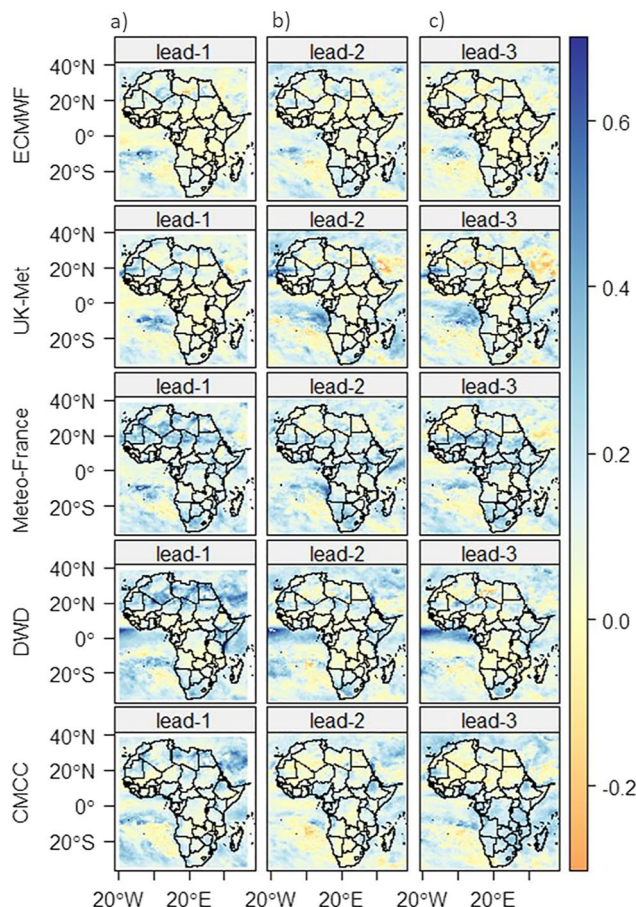
**Figure 4.** Temporal correlation of monthly precipitation anomalies between Multi-Source Weighted-Ensemble Precipitation and (a) lead-1, (b) lead-2, and (c) lead-3 month forecasts for the five climate models.

During JJA, ECMWF, UK-Met, and Meteo-France overestimate (up to 240 mm/month) and DWD and CMCC underestimate (up to 130 mm/month) the monthly average precipitation in the west part of Ethiopia, where JJA is one of the most important seasons in terms of water resources. All models except DWD show an overestimation of up to 160 mm/month in the southern part of the South-Sudan but are close to the observations in a large part of Kenya and Tanzania and the eastern part of Ethiopia.

Across most of West Africa, the seasonal precipitation is well represented (within 50 mm/month) by all models during DJF, except for underestimation (up to 80 mm/month) by UK-Met, Meteo-France, and DWD in Côte D'Ivoire and Ghana. During JJA, the precipitation climatology in West Africa is well represented by ECMWF, UK-Met, and CMCC but Meteo-France and DWD show higher biases (up to 210 mm/month), particularly in Nigeria and Guinea. In Central Africa, ECMWF, Meteo-France, and CMCC overestimate the DJF precipitation



**Figure 5.** Regional summary of correlation of monthly (a) and monthly anomaly (b) precipitation for lead-1 to lead-6 months forecasts for the five models.



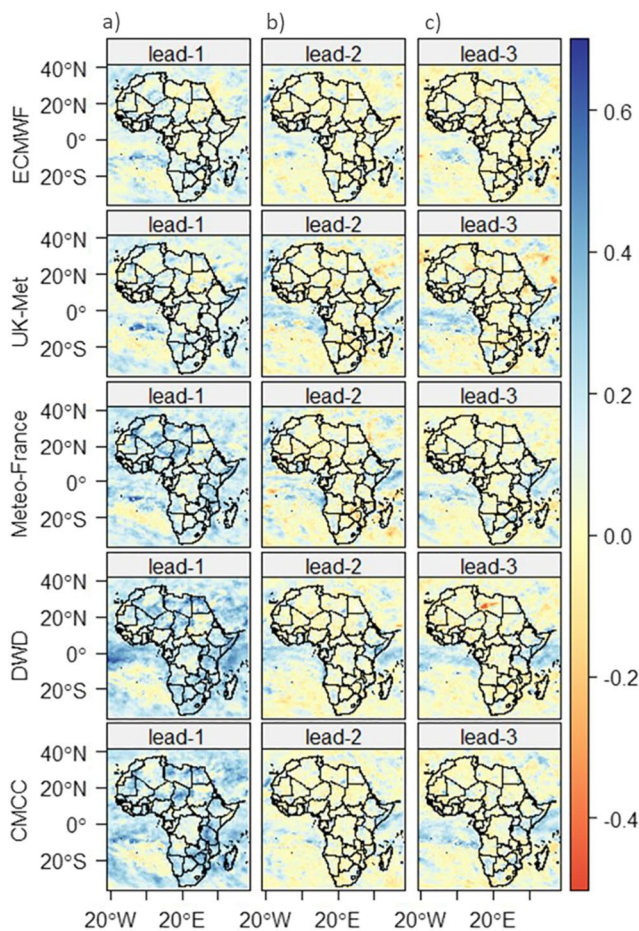
**Figure 6.** Monthly correlation coefficient difference between weighted multi-model mean and individual models for (a) lead-1, (b) lead-2, and (c) lead-3.

in Gabon and Congo, whilst UK-Met and DWD underestimate. However, the climatology is well represented by all models in Cameroon, Chad, and the Central African Republic. ECMWF, UK-Met, Meteo-France, and CMCC overestimate the JJA precipitation in parts of southern Chad and northern parts of the Central Africa Republic and North-eastern Congo. Meteo-France shows an overestimation in Cameroon and Gabon up to 235 mm/month.

In Southern Africa, all models show biases of more than 50 mm/month precipitation during the DJF wet season. For example, all models underestimate (up to 220 mm/month) the monthly precipitation in Mozambique and overestimate in the eastern part of South Africa (around KwaZulu-Natal provinces). They also overestimate precipitation during JJA in this region. Unlike DJF, all models, except DWD, show a JJA overestimation up to 60 mm/month in large parts of Mozambique. In general, in Southern Africa, ECMWF, and DWD show a better representation of the climatology with a lower bias and Meteo-France is the least performing model with high bias up to 60 mm/month. Compared to the other regions, the North African climatology is very well represented by all models during both seasons, again partly because the climatology is relatively dry. However, DWD and CMCC, and UK-Met show an overestimation up to 40 mm/month in Morocco (DJF) and Southern Algeria (JJA), respectively.

Regional averages of the observation and model lead-1 climatologies are shown in Figure 2, with regions defined as East, Central, and West Africa compared to North and Southern Africa (Figure 1a). All models have a good representation of the seasonal cycle of precipitation with biases generally low in East, Central, and West Africa compared to North and Southern Africa (Figure 2). In East and Central Africa, all models tend to overestimate the monthly precipitation, but their performance is mixed in West Africa, reflecting the variable biases at national scales as noted previously. In East Africa, DWD and UK-Met show the least bias (0.6–9 mm) in May–November, and January, February and December, respectively. CMCC shows the highest bias (11–41 mm) across all





**Figure 7.** Monthly anomaly correlation coefficient difference between weighted multi-model mean and individual models for (a) lead-1, (b) lead-2, and (c) lead-3.

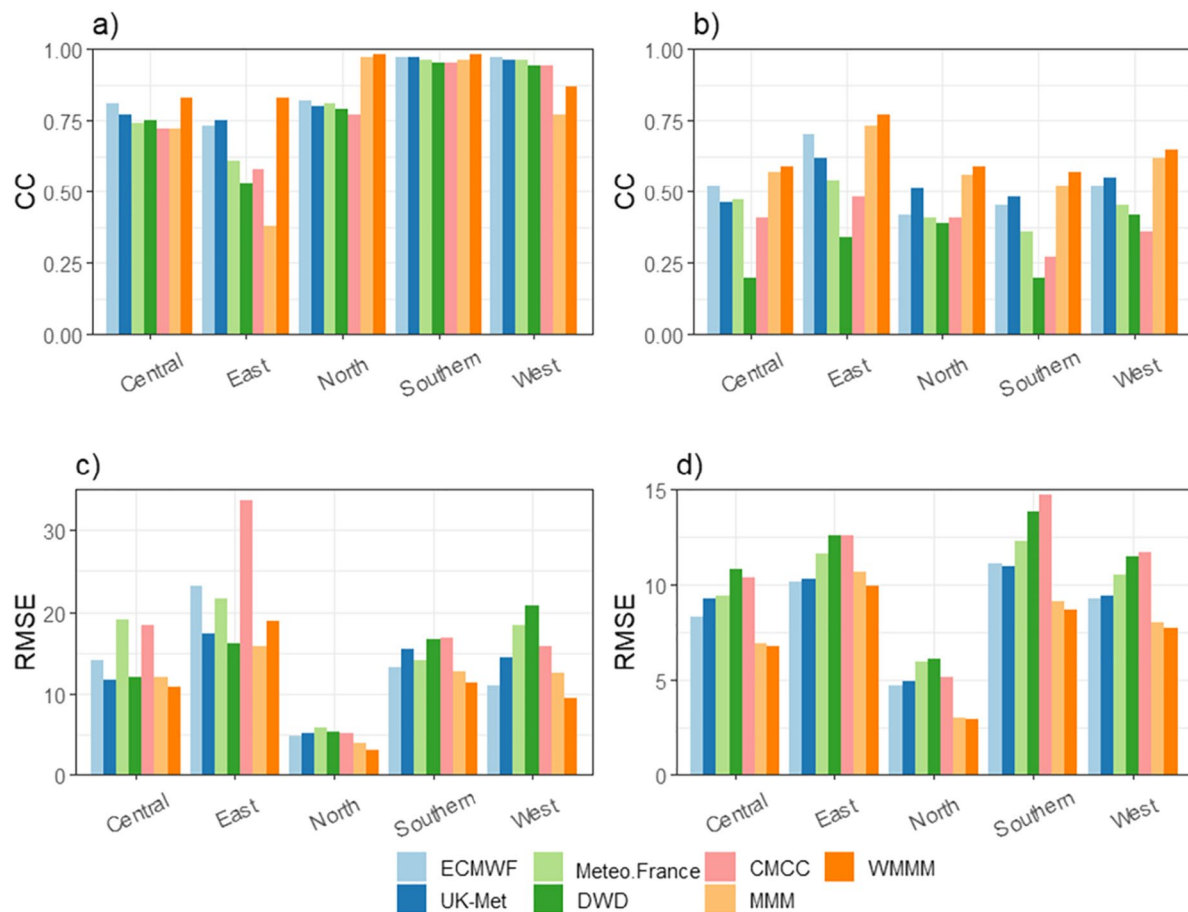
months. In Central Africa, August–December and March–June precipitation is well represented by DWD and UK-Met, respectively, and the other models show a higher monthly bias up to 31 mm. ECMWF is the least biased model in West Africa with a bias from 0.2 to 9 mm, and DWD and CMCC show a high positive bias (up to 30 mm). Unlike the other regions, CMCC shows the lowest bias in North Africa during January, March, May, and September followed by ECMWF and UK-Met, with DWD and Meteo-France having the highest biases. Similar to East and Central Africa, in Southern Africa, DWD represents the monthly climatology, except in February and April, very well. Meteo-France shows a higher bias (8–10 mm) during May–October.

### 3.2. Monthly and Monthly Anomaly Precipitation

The performance of the individual models is first evaluated on a grid cell basis and then summarized over regions. Monthly and monthly anomaly correlation values for 1–3 month lead-times are summarized in Figure 3 and Figure 4, respectively. For each model, anomalies were computed as the departure from the seasonal climatological mean. The performance of all models is highest for lead-1 and decreases with increasing lead time for monthly (Figure 3) and monthly anomalies (Figure 4). This decrease is significantly more pronounced when considering the anomaly correlations. The lead-1 monthly correlation of all models is higher in the drier tropics, particularly in the semi-arid regions where there is strong seasonality (Figure 3), and lower in arid and semi-arid regions. In East Africa, the monthly precipitation is well represented (CC up to 0.95) by the lead-1 forecast from ECMWF and UK-Met, and in comparison to the other models. The other models show lower correlation across the Horn of Africa, especially in Kenya and southeast Ethiopia. Similar to the monthly correlation, ECMWF and UK-Met show the highest anomaly correlation (Figure 4) in East Africa (lead-1 CC up to 0.65) compared to the other models, which show lower correlation, especially in countries such as in Tanzania and Sudan. In West Africa, ECMWF and UK-Met again show the highest monthly correlations and anomaly correlations overall. However, the anomaly correlations for all models are generally lower in this region compared to East Africa.

In central Africa, particularly in the Central Africa Republic, Cameroon, and Western parts of Chad and Congo, the monthly variability is well represented by all models with a correlation between 0.8 and 0.96. All models have lower correlation in the wet tropical regions of the Congo and out to the Atlantic coast. Again, ECMWF and UK-Met are, in general, the best performing models. In this region, the anomaly correlation values are relatively lower compared to other regions, although ECMWF and UK-Met show smaller areas of higher correlation (e.g., up to 0.52 in Chad). In Southern Africa, all models show a monthly correlation between 0.8 and 0.95, but the anomaly correlation is again low when compared to East Africa. The monthly correlation of all models is low in much of South Africa and southern Botswana and Namibia. Again, ECMWF and UK-Met show the highest monthly correlations and anomaly correlations regionally. In North Africa, the monthly and monthly anomaly correlations of all models are overall quite low. ECMWF and UK-Met show slightly higher values than the other models, but the differences are small and spatially variable. The highest anomaly correlations (CC up to 0.65) are in Libya and Northern Morocco for ECMWF.

The monthly and monthly anomaly correlations for East, Central, West, North, and Southern Africa are summarized in Figure 5. This highlights the higher correlations in predominantly semi-arid regions which have distinct seasonality (West and Southern Africa) that is somewhat easier to predict, versus regions that are predominantly arid (North Africa) with sporadic rainfall, regions that are relatively wet (Central Africa) or those with more complex seasonality (e.g., two rainy season peaks such as in East Africa). The area average correlation is highest at lead-1 and drops with increasing lead time, particularly the anomaly correlation, which falls to around zero

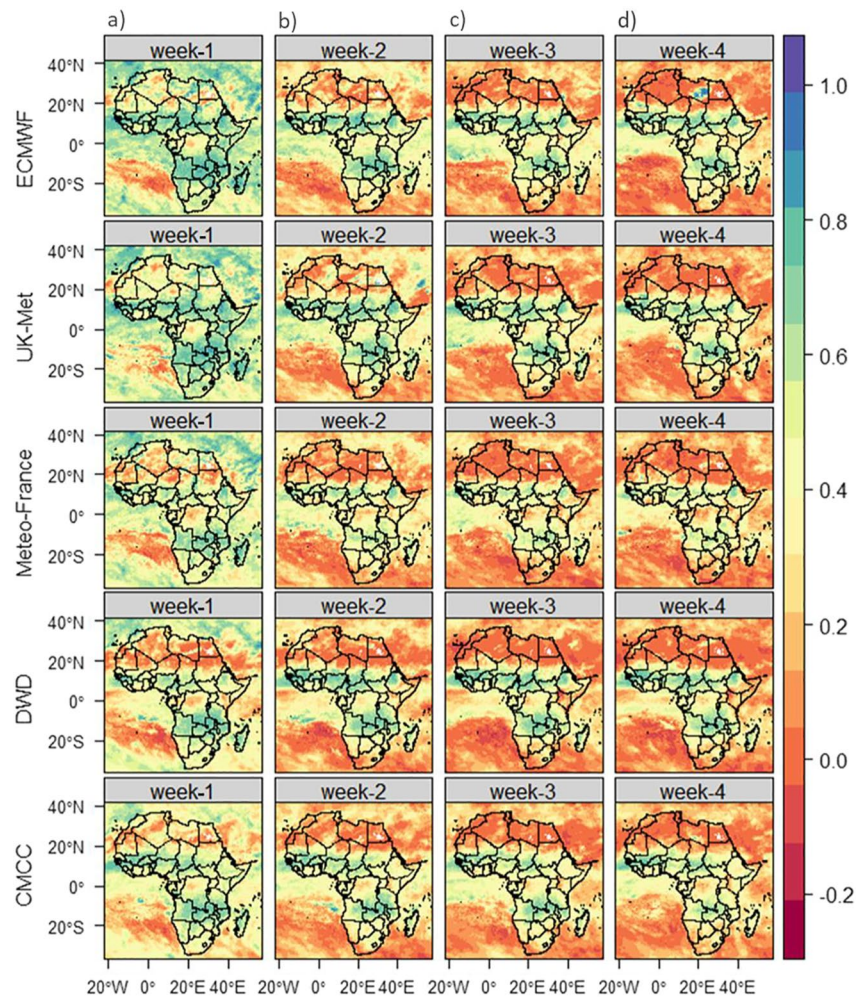


**Figure 8.** Regional lead-1 (a) monthly and (b) anomaly CC, and (c) monthly and (d) anomaly Root Mean Square Error (RMSE) for all models, multi-model mean (MMM) and weighted multi-model mean (WMMM) averaged over Central, East, North, Southern and West Africa.

by lead-3 to 4 for all regions. The performance of all models is generally similar across lead times in West and North Africa, and more diverse in Central, East and Southern Africa. ECMWF and UK-Met outperform the other models for almost all regions at lead-1 and generally for longer lead times.

Compared to the individual models, however, the weighted multi-model mean (WMMM) generally show higher monthly (Figure 6) and anomaly correlations (Figure 7). The WMMM shows an improved correlation (up to 0.6), particularly compared to CMCC, DWD and Meteo-France, for leads 1–3 (Figure 6). The skill difference between WMMM and ECMWF and UK-Met is lower in Central and Northern Africa. However, the WMMM does show a significant improvement in parts of East (e.g., Tanzania), West (e.g., Mauritania and Niger) and Southern (e.g., South Africa, Namibia and Botswana) Africa. Similarly, the MMM shows a higher correlation compared to CMCC, DWD and Meteo-France for lead-1 but did not improve on the skill at lead-2 and lead-3 months (Figure S1 in Supporting Information S1). In general, WMMM outperforms the individual models in large parts of the continent and outperforms the MMM in North and East Africa (Figure S2 in Supporting Information S1).

The anomaly correlation also shows a clear improvement when using the WMMM compared to the individual models (Figure 7). In addition, the WMMM outperforms the MMM in the East region (e.g., lead-3) and other smaller parts of the continent at leads 1, 2 and 3, respectively (Figure S3 in Supporting Information S1). However, the MMM shows a higher correlation, particularly at lead-1, compared to the individual models (Figure S4 in Supporting Information S1). Overall, the WMMM significantly improves the lead-1 monthly and monthly anomaly CC of the WMMM is higher than the individual models and the MMM (Figure 8). The WMMM also shows a



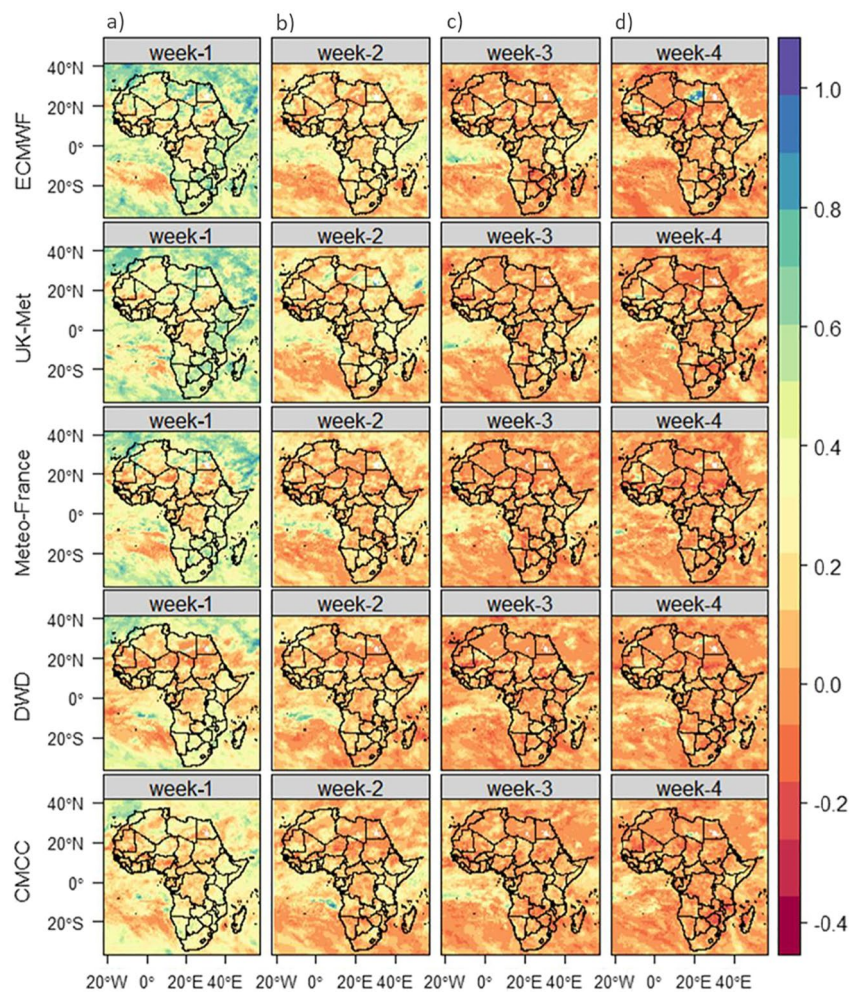
**Figure 9.** Temporal correlation of weekly (7-days) precipitation between Multi-Source Weighted-Ensemble Precipitation and (a) week-1, (b) week-2, (c) week-3, and (d) week-4 forecast from all models.

reduced RMSE compared to the individual models and MMM. In East Africa, however, the monthly RMSE from the WMMM is much higher than UK-Met, DWD and the MMM. Compared to Central, North, Southern, and West Africa, the monthly CC difference between WMMM and MMM is about 0.4 and CMCC shows the highest RMSE (5–14 mm), compared to the other models and the MMM and WMMM, in East Africa.

### 3.3. Weekly and Weekly Anomaly Precipitation

We further analyze the lead-1 results at weekly time scale to understand how the skill evolves within the first month for weekly values and weekly anomalies (Figures 9 and 10). For all models, the weekly correlations are highest for week-1 and these persist through weeks 2–4 in the seasonally wet regions of the west and South African monsoonal regions, but drop rapidly in drier regions (Figure 9). Again, ECMWF and UK-Met tend to outperform the other models at lead-1 week, and longer leads for some regions such as the Horn of Africa. The anomaly correlation values generally follow the same pattern of highest values at lead-1 week and a rapid drop through weeks 2–4, although the spatial patterns differ substantially (Figure 10). The highest correlation values are in the east from the Horn of Africa down to southern Africa with highest skill for ECMWF, UK-Met and Meteo-France. There is some persistence of skill in the Horn of Africa for ECMWF and UK-Met, but otherwise, anomaly correlation values drop to below about 0.3 and close to zero by week 2 for much of the continent.

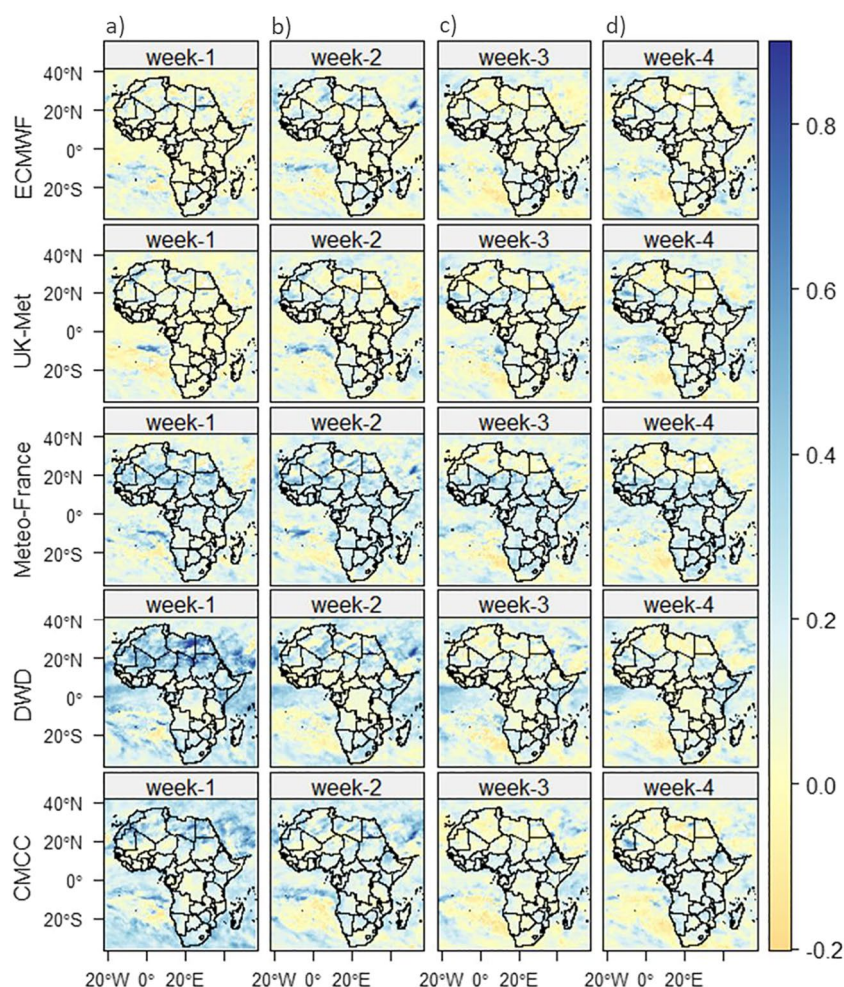




**Figure 10.** Temporal correlation of weekly anomaly (7-days) precipitation between Multi-Source Weighted-Ensemble Precipitation and (a) week-1, (b) week-2, (c) week-3, and (d) week-4 forecast from all models.

The WMMM weekly precipitation shows a better performance in large parts of the continent (Figure 11), compared to the individual models and is similar to the monthly evaluation. The WMMM improves the correlation in all weeks, particularly compared to Meteo-France, and less so for DWD and CMCC. The difference between WMMM and ECMWF and UK-Met is very small in Central (e.g., Congo) and East (Ethiopia and Tanzania) Africa with slight improvements elsewhere. Similar to the WMMM, the MMM also shows a higher CC compared to Meteo-France, CMCC, and DWD for most regions for all weeks (Figure S5 in Supporting Information S1). However, compared to the MMM, the WMMM provide better performance (CC up to 0.4) in all weeks, particularly in North and West (Mali and Mauritania) Africa (Figure S6 in Supporting Information S1).

The WMMM also shows a higher performance for the weekly anomaly compared to all models, particularly Meteo-France, DWD, and CMCC (Figure 12). The CC of ECMWF and UK-Met is lower in parts of North and Southern Africa in weeks 1 and 2 compared to WMMM but the difference, particularly with ECMWF, is lower in East, West and Central Africa. In general, the WMMM shows a higher CC in large parts of the region in week-1 and week-2 and some parts of East and Southern Africa in week-2 and week-4 compared to Meteo-France, CMCC and DWD. Similarly, the MMM shows a higher performance compared to individual models particularly in week-1 and week-2 (Figure S7 in Supporting Information S1). Overall, the performance of the WMMM, compared to MMM is higher (CC up to 0.4) in smaller parts of North, East, West, and Southern Africa (Figure S8 in Supporting Information S1).



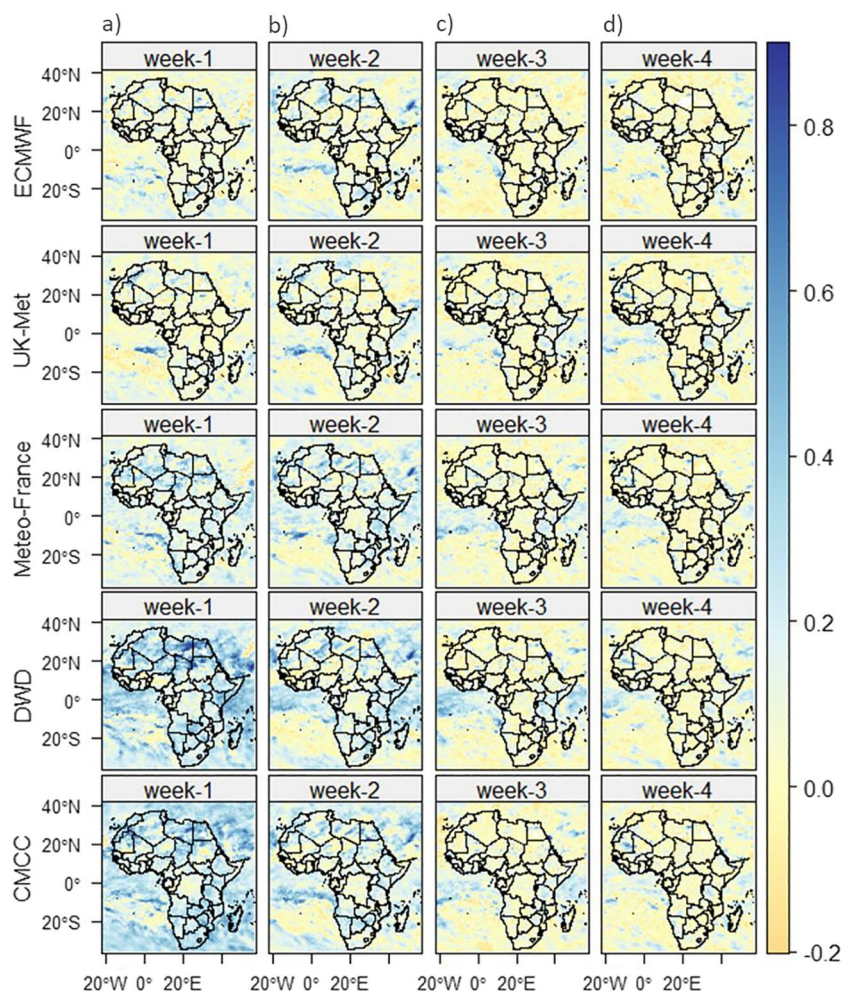
**Figure 11.** Temporal correlation difference between weighted multi-model mean and individual models for (a) week-1, (b) week-2, (c) week-3, and (d) week-4.

### 3.4. Daily Precipitation Indices

All models show a good performance in representing the monthly variations in CDD and CWD at lead-1 but the skill is very low for heavy and very heavy precipitation days (Figure 13). In East Africa, UK-Met and Meteo-France in Ethiopia, Kenya, and Tanzania show a higher CC compared to ECMWF, which showed a higher CC in Sudan, and DWD is the least performing model for CDD. In large parts of Central Africa, ECMWF followed by Meteo-France shows a higher CC whereas UK-Met and CMCC appeared to be the best models in Southern Africa. For CWD, in general, ECMWF in East and West Africa, Meteo-France in Central Africa, and UK-Met in Southern Africa are the best performing models (CC up to 0.8).

All models, however, are very poor in forecasting heavy and very heavy precipitation days compared to CDD and CWD. In East Africa, UK-Met in Kenya and Ethiopia, and ECMWF in Tanzania show a higher CC (up to 0.78) for heavy precipitation compared to DWD, CMCC, and Meteo-France. Similarly, in Central Africa, ECMWF is the best model and DWD and CMCC are the least performing models. However, ECMWF, except in Malawi and Mozambique, is one of the least performing models in Southern Africa and UK-Met shows the highest CC for heavy precipitation days. Compared to the other parts of Africa, all models show a higher CC for heavy precipitation days in West Africa. In comparison to UK-Met, DWD, and CMCC, Meteo-France and ECMWF are the best models with the highest CC for heavy precipitation days.



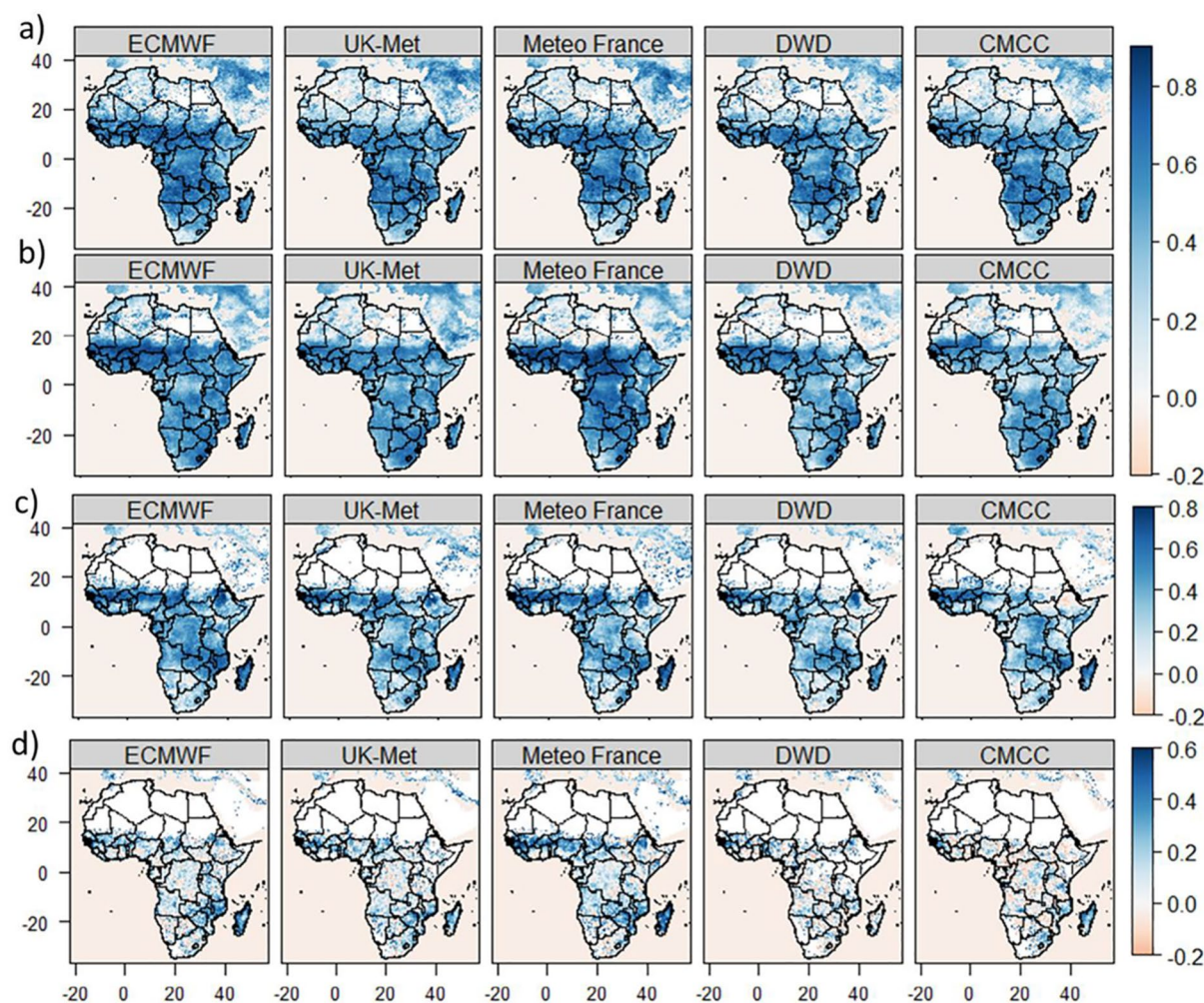


**Figure 12.** Temporal anomaly correlation difference between weighted multi-model mean and individual models for (a) week-1, (b) week-2, (c) week-3, and (d) week.

### 3.5. Application to Seasonal Forecasting of Meteorological Droughts

The monthly precipitation forecasts are used to compute the SPI drought index at each grid cell. Given the rapid drop in skill for monthly precipitation at lead-2, we focus on the lead-1 forecast of the 1-month SPI (SPI-1). Figure 14 shows the FAR, POD, and BS for SPI-1 averaged over the five regions for all models and the MMM and WMMM values. For the individual models, the SPI-1 is computed for each ensemble member, and the ensemble average FAR and POD is used for analysis. Overall, the POD ranges from about 0.25 to 0.6 with a mean of about 0.4 across models and regions of East, Central, West, North and Southern Africa, and the FAR from about 0.4 to 0.8, with a mean of 0.66. This indicates that the probability of detecting drought events is modest with a likelihood of false detection. No model stands out as having the highest POD or lowest FAR scores across all regions, although the range across models can be quite large (e.g., West and Central Africa). This contrasts with the results from the monthly analysis which highlighted the ECMWF and UK-Met models as the best performing models across almost all sets of results. The ensemble means show better values than the individual models overall, with a few exceptions, and distinctly better values for some regions (e.g., Southern and East for POD). Interestingly, the MMM tends to have slightly better values than the WMMM.



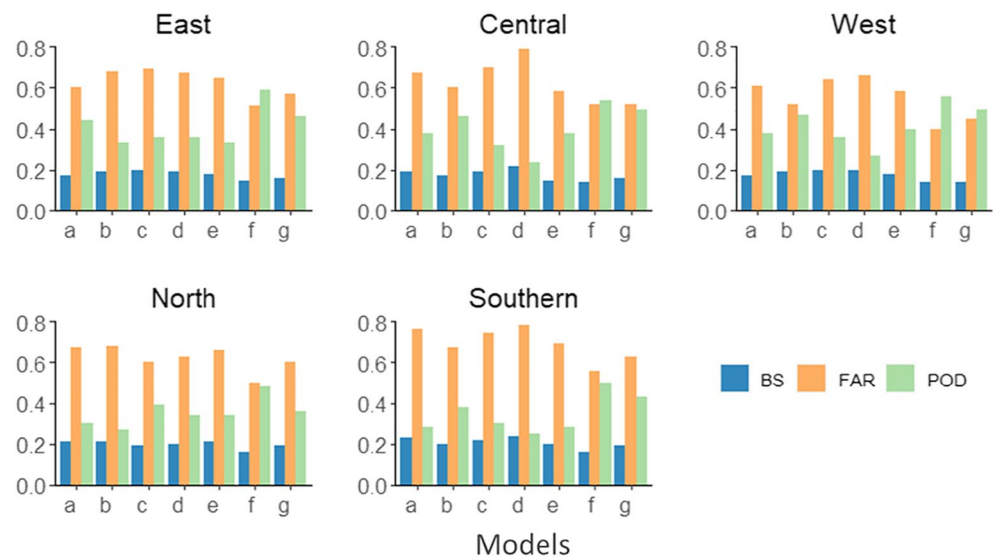


**Figure 13.** Correlation of monthly (a) consecutive dry days, (b) consecutive wet days, (c) heavy precipitation days, and (d) very heavy precipitation days between Multi-Source Weighted-Ensemble Precipitation and lead-1 forecast from European Centre for Medium-Range Weather Forecasts (ECMWF), UK Met Office (UK-Met), Meteo-France, Deutscher Wetterdienst (DWD), and Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC).

#### 4. Summary and Discussion

Providing timely and reliable climate information is essential to monitor extreme events and develop risk reduction strategies, which is highly dependent on the skill of the seasonal forecast models. Here, we evaluated the skill of five state-of-the-art European seasonal forecast models, as well as their unweighted and weighted multi-model means for their hindcast period (1993–2016) over Africa. The models considered here are selected based on the availability on daily or higher temporal resolution and for both hindcast and real-time data for potentially driving hydrological forecasts at regional and basin scales. We focused on the ensemble mean of models to assess the skill on weekly, monthly and seasonal time scales, and the full ensemble was assessed for daily extremes and meteorological droughts.

The ensemble mean of all models show relatively high skill at lead-1 with the highest correlations over East Africa, particularly for ECMWF and UK-Met. However, the skill drops with an increase in lead-time, which is in agreement with other studies that have focused on the same models but other regions (e.g., Nobakht et al., 2021; Wang et al., 2019) or on other seasonal forecast models (e.g., Barnston et al., 2011; King et al., 2020; Wanders & Wood, 2016). The limitation of skill to a few weeks is expected and due to the chaotic nature of the atmosphere (e.g., Lavers et al., 2009; Yuan et al., 2011). In large parts of Africa, the ECMWF and UK-Met models show a higher performance compared to Meteo-France, DWD, and CMCC for monthly and monthly anomaly precipitation.



**Figure 14.** Probability of Detection (POD), False Alarm Ratio (FAR) and Brier Score (BS) values of 1-month SPI (SPI-1) for the lead-1 forecast for (a) European Centre for Medium-Range Weather Forecasts, (b) UK Met Office, (c) Meteo-France, (d) Deutscher Wetterdienst, (e) Centro Euro-Mediterraneo sui Cambiamenti Climatici, (f) multi-model mean, and (g) weighted multi-model mean averaged over East, Central, West, North, and Southern Africa.

However, there are areas where the latter three models show relatively higher performance across the range of indices and time scales. Hence, we developed a weighted average based on the correlation values of individual models on a grid cell basis. Weighting models based on their skill improved the skill compare to averaging all models, as less skilful models can strongly degrade the performance (Casanova & Ahrens, 2009; Xu et al., 2020).

The unweighted (MMM) and weighted (WMMM) multi-model mean improved the overall skill in large parts of the region for monthly and anomaly correlation and forecasting droughts. Compared to the MMM, WMMM improves the monthly and anomaly correlation and reduces model errors. In agreement with our findings, many studies (e.g., Batté & DéQué, 2011; Casanova & Ahrens, 2009; Dash et al., 2019; Pan et al., 2016; Pepler et al., 2015; Rozante et al., 2014; Wanders & Wood, 2016; Weigel et al., 2008) have concluded that the MMM is consistently better than individual models for seasonal mean climate. Furthermore, the agreement of the weighted multi-model mean with the observational data is higher than the individual models and the MMM, and provides improved performance for other climate variables and extreme events (Casanova & Ahrens, 2009; Ratnam et al., 2019; Wanders & Wood, 2016), although it might underestimate short-term peak values because of model averaging.

Concerning daily extremes, taking all ensembles of the lead-1 forecast, all models represent the monthly variation in the length of dry (CDD) and wet (CWD) spell days and heavy precipitation days well, particularly in the tropics. However, the correlation is weak for very heavy precipitation days and this agrees with the findings of Pepler et al. (2015) and King et al. (2020), which highlight the low skill of climate models for prediction of extreme precipitation indices. Forecasting precipitation and its extremes using climate models is challenging due to the complex and intermittent nature of precipitation (Han et al., 2019; Pepler et al., 2015), and is relatively more difficult compared to forecasting temperature. In general, ECMWF and UK-Met followed by Meteo-France provide a higher correlation for CDD, CWD, and heavy and very heavy precipitation days at lead-1.

Application of the lead-1 forecast for drought monitoring shows modest skill, but improvements when using the MMM or WMMM. In East, Central, West, North and Southern Africa up to 44% (ECMWF), 46% (UK-Met), 58% (CMCC), 47% (UK-Met), 39% (Meteo-France), and 38% (UK-Met) of droughts events are forecasted by the models, respectively. Similar to our findings, in Latin America, ECMWF showed a POD of 30% for drought events with FAR less than 70% (Carrão et al., 2018). The drought forecast performance is improved when using the MMM and WMMM. For example, in East Africa, the POD of the MMM and WMMM is 0.59 and 0.46, respectively and the FAR and BS is lower than individual models. According to Hao et al. (2018), applying unequal weights to multi-model ensembles improves the prediction skill for droughts, particularly in tropical and

subtropical regions. The higher prediction skill for precipitation and droughts in these regions is due to the strong teleconnection with El Niño–Southern Oscillation (Hao et al., 2018; Pozzi et al., 2013). Overall, the performance of the weighted multi-model mean forecasts is promising compared to individual models and could provide the most accurate information on regional and basin levels than individual models. However, improving the skill for leads beyond 1 month remains a challenge.

## Data Availability Statement

All seasonal forecast data used in this work can be found at the following sites: The Copernicus Climate Change Service (C3S) (months) forecasts of precipitation) and the Multi-Source Weighted-Ensemble Precipitation Can be found at GloH2O (<http://www.gloh2o.org/mswep/>).

## Acknowledgments

This work was supported by “FutureDams” (grant number ES/P011373/1) project and the “Building REsearch Capacity for sustainable water and food security In drylands of sub-saharan Africa” (BRECcIA) (grant number NE/P021093/1) project, both of which are supported by UK Research and Innovation as part of the Global Challenges Research Fund.

## References

- Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., & Pappenberger, F. (2013). GloFAS – Global ensemble streamflow forecasting and flood early warning. *Hydrology and Earth System Sciences*, 17(3), 1161–1175. <https://doi.org/10.5194/hess-17-1161-2013>
- Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., Andersson, J. C. M., Hasan, A., & Pineda, L. (2020). Global catchment modelling using World-Wide HYPE (WWH), open data, and stepwise parameter estimation. *Hydrology and Earth System Sciences*, 24(2), 535–559. <https://doi.org/10.5194/hess-24-535-2020>
- Awange, J. L., Hu, K. X., & Khaki, M. (2019). The newly merged satellite remotely sensed, gauge and reanalysis-based multi-source weighted-ensemble precipitation: Evaluation over Australia and Africa (1981–2016). *The Science of the Total Environment*, 670, 448–465. <https://doi.org/10.1016/j.scitotenv.2019.03.148>
- Barnston, A. G., Tippett, M. K., L’Heureux, M. L., Li, S., & DeWitt, D. G. (2011). Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? *Bulletin of the American Meteorological Society*, 93(5), 631–651. <https://doi.org/10.1175/BAMS-D-11-00111.1>
- Batté, L., & DeÉQueÉ, M. (2011). Seasonal predictions of precipitation over Africa using coupled ocean-atmosphere general circulation models: Skill of the ENSEMBLES project multimodel ensemble forecasts. *Tellus A: Dynamic Meteorology and Oceanography*, 63(2), 283–299. <https://doi.org/10.1111/j.1600-0870.2010.00493.x>
- Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., & de Roo, A. (2017). MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences*, 21(1), 589–615. <https://doi.org/10.5194/hess-21-589-2017>
- Beck, H. E., Vergopolan, N., Pan, M., Levizzani, V., Van Dijk, A. I. J. M., Weedon, G. P., et al. (2017). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*, 21(12), 6201–6217. <https://doi.org/10.5194/hess-21-6201-2017>
- Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., van Dijk, A. I. J. M., et al. (2019). MSWEP V2 global 3-hourly 0.1° precipitation: Methodology and quantitative assessment. *Bulletin of the American Meteorological Society*, 100(3), 473–500. <https://doi.org/10.1175/BAMS-D-17-0138.1>
- Carrão, H., Naumann, G., Dutra, E., Lavaysse, C., & Barbosa, P. (2018). Seasonal drought forecasting for Latin America using the ECMWF S4 forecast system. *Climate*, 6(2), 48. <https://doi.org/10.3390/cli6020048>
- Casanova, S., & Ahrens, B. (2009). On the weighting of multimodel ensembles in seasonal and short-range weather forecasting. *Monthly Weather Review*, 137(11), 3811–3822. <https://doi.org/10.1175/2009MWR2893.1>
- Chen, L., & Dirmeyer, P. A. (2016). Impacts of land-use/land-cover change on afternoon precipitation over North America. *Journal of Climate*, 30(6), 2121–2140. <https://doi.org/10.1175/JCLI-D-16-0589.1>
- Chen, L., Dirmeyer, P. A., Tawfik, A., & Lawrence, D. M. (2017). Sensitivities of land cover–precipitation feedback to convective triggering. *Journal of Hydrometeorology*, 18(8), 2265–2283. <https://doi.org/10.1175/JHM-D-17-0011.1>
- Dash, I., Nagai, M., & Pal, I. (2019). Forecast Customization System (FOCUS): A multimodel ensemble-based seasonal climate forecasting tool for the homogeneous climate zones of Myanmar. *Advances in Meteorology*, 2019, 495712. <https://doi.org/10.1155/2019/495712>
- FAO. (2014). *Food and Agricultural Organization of the United Nations and World Bank. The State of Food Insecurity in the World (SOFI)*. Retrieved from <http://www.fao.org/3/a-i4030e.pdf>
- Glowienka-Hense, R., Hense, A., Brune, S., & Baehr, J. (2020). Comparing forecast systems with multiple correlation decomposition based on partial correlation. *Advances in Statistical Climatology, Meteorology and Oceanography*, 6, 103–113. <https://doi.org/10.5194/ascmo-6-103-2020>
- Han, T., Wang, H., Hao, X., & Li, S. (2019). Seasonal prediction of midsummer extreme precipitation days over Northeast China. *Journal of Applied Meteorology and Climatology*, 58(9), 2033–2048. <https://doi.org/10.1175/JAMC-D-18-0253.1>
- Hansen, J. W., Mason, S. J., Sun, L., & Tall, A. (2011). Review of seasonal climate forecasting for agriculture in sub-Saharan Africa. *Experimental Agriculture*, 47(2), 205–240. <https://doi.org/10.1017/S0014479710000876>
- Hao, Z., Singh, V. P., & Xia, Y. (2018). Seasonal drought prediction: Advances, challenges, and future prospects. *Reviews of Geophysics*, 56(1), 108–141. <https://doi.org/10.1002/2016RG000549>
- Karl, T. R., Nicholls, N., & Ghazi, A. (1999). CLIVAR/GCOS/WMO workshop on indices and indicators for climate extremes workshop summary. In T. R. Karl, N. Nicholls, & A. Ghazi (Eds.), *Weather and climate extremes: Changes, variations and a perspective from the insurance industry* (pp. 3–7). Springer Netherlands. [https://doi.org/10.1007/978-94-015-9265-9\\_2](https://doi.org/10.1007/978-94-015-9265-9_2)
- Kim, S. (2015). Partial and semi-partial (part) correlation. Version 1.1.
- King, A. D., Hudson, D., Lim, E.-P., Marshall, A. G., Hendon, H. H., Lane, T. P., & Alves, O. (2020). Sub-seasonal to seasonal prediction of rainfall extremes in Australia. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 2228–2249. <https://doi.org/10.1002/qj.3789>
- Kirtman, B. P., Min, D., Infanti, J. M., Kinter, J. L., Paolino, D. A., Zhang, Q., et al. (2013). The North American multimodel ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bulletin of the American Meteorological Society*, 95(4), 585–601. <https://doi.org/10.1175/BAMS-D-12-00050.1>



- Kundzewicz, Z. W., & Kaczmarek, Z. (2000). Coping with hydrological extremes. *Water International*, 25(1), 66–75.
- Lang, A. L., Pegion, K., & Barnes, E. A. (2020). Introduction to special collection: Bridging weather and climate: Subseasonal-to-Seasonal (S2S) prediction. *Journal of Geophysical Research: Atmospheres*, 125(4), e2019JD031833. <https://doi.org/10.1029/2019JD031833>
- Lavers, D., Luo, L., & Wood, E. F. (2009). A multiple model assessment of seasonal climate forecast skill for applications. *Geophysical Research Letters*, 36(23). <https://doi.org/10.1029/2009GL041365>
- Liu, Q., McVicar, T. R., Yang, Z., Donohue, R. J., Liang, L., & Yang, Y. (2016). The hydrological effects of varying vegetation characteristics in a temperate water-limited basin: Development of the dynamic Budyko-Choudhury-Porporato (dBCP) model. *Journal of Hydrology*, 543, 595–611. <https://doi.org/10.1016/j.jhydrol.2016.10.035>
- Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., et al. (2017). GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, 10(5), 1903–1925. <https://doi.org/10.5194/gmd-10-1903-2017>
- McKee, T. B., Doesken, N. J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. 8th Conference on Applied Climatology, Anaheim, 17–22 January 1993, 179–184. [https://www.droughtmanagement.info/literature/AMS\\_Relationship\\_Drought\\_Frequency\\_Duration\\_Time\\_Scales\\_1993.pdf](https://www.droughtmanagement.info/literature/AMS_Relationship_Drought_Frequency_Duration_Time_Scales_1993.pdf)
- Morss, R. E., Lazo, J. K., Brown, B. G., Brooks, H. E., Ganderton, P. T., & Mills, B. N. (2008). Societal and economic research and applications for weather forecasts: Priorities for the North American THORPEX program. *Bulletin of the American Meteorological Society*, 89(3), 335–346. <https://doi.org/10.1175/BAMS-89-3-335>
- Niang, I., Ruppel, O. C., Abdrabo, M. A., Essel, A., Lennard, C., Padgham, J., & Urquhart, P. (2014). Africa. In V. R. Barros, et al. (eds.), *Climate change 2014: Impacts, adaptation, and vulnerability. Part B: Regional aspects* (pp. 1199–1265). Cambridge University Press. Retrieved from [http://www.ipcc.ch/pdf/assessment-report/ar5/wg2/WGIIAR5-Chap22\\_FINAL.pdf](http://www.ipcc.ch/pdf/assessment-report/ar5/wg2/WGIIAR5-Chap22_FINAL.pdf)
- Nobakht, M., Saghaian, B., & Aminyavari, S. (2021). Skill assessment of Copernicus climate change Service seasonal ensemble precipitation forecasts over Iran. *Advances in Atmospheric Sciences*, 38(3), 504–521. <https://doi.org/10.1007/s00376-020-0025-7>
- Pan, S., Fan, Y., Chen, J., & Kao, C. (2016). Optimization of multi-model ensemble forecasting of typhoon waves. *Water Science and Engineering*, 9(1), 52–57. <https://doi.org/10.1016/j.wse.2016.02.001>
- Pepler, A. S., Díaz, L. B., Prodhomme, C., Doblas-Reyes, F. J., & Kumar, A. (2015). The ability of a multi-model seasonal forecasting ensemble to forecast the frequency of warm, cold and wet extremes. *Weather and Climate Extremes*, 9, 68–77. <https://doi.org/10.1016/j.wace.2015.06.005>
- Pozzi, W., Sheffield, J., Stefanski, R., Cripe, D., Pulwarty, R., Vogt, J. V., et al. (2013). Toward global drought early warning capability: Expanding international cooperation for the development of a framework for monitoring and forecasting. *Bulletin of the American Meteorological Society*, 94(6), 776–785. <https://doi.org/10.1175/BAMS-D-11-00176.1>
- Ratnam, J. V., Dijkstra, H. A., Doi, T., Morioka, Y., Nonaka, M., & Behera, S. K. (2019). Improving seasonal forecasts of air temperature using a genetic algorithm. *Scientific Reports*, 9(1), 12781. <https://doi.org/10.1038/s41598-019-49281-z>
- Rozante, J. R., Moreira, D. S., Godoy, R. C. M., & Fernandes, A. A. (2014). Multi-model ensemble: Technique and validation. *Geoscientific Model Development*, 7(5), 2333–2343. <https://doi.org/10.5194/gmd-7-2333-2014>
- Samaniego, L., Kumar, R., Thober, S., Rakovec, O., Zink, M., Wanders, N., et al. (2017). Toward seamless hydrologic predictions across spatial scales. *Hydrology and Earth System Sciences*, 21(9), 4323–4346. <https://doi.org/10.5194/hess-21-4323-2017>
- Samaniego, L., Thober, S., Wanders, N., Pan, M., Rakovec, O., Sheffield, J., et al. (2019). Hydrological forecasts and projections for improved decision-making in the water sector in Europe. *Bulletin of the American Meteorological Society*, 100(12), 2451–2472. <https://doi.org/10.1175/BAMS-D-17-0274.1>
- Schellekens, J., Dutra, E., Martínez-de la Torre, A., Balsamo, G., van Dijk, A., Sperna Weiland, F., et al. (2017). A global water resources ensemble of hydrological models: The earth2Observe tier-1 dataset. *Earth System Science Data*, 9(2), 389–413. <https://doi.org/10.5194/essd-9-389-2017>
- Sheffield, J., Wood, E. F., Chaney, N., Guan, K., Sadri, S., Yuan, X., et al. (2014). A drought monitoring and forecasting system for sub-Saharan African water resources and food security. *Bulletin of the American Meteorological Society*, 95(6), 861–882. <https://doi.org/10.1175/BAMS-D-12-00124.1>
- Thielen, J., Bartholmes, J., Ramos, M.-H., & de Roo, A. (2009). The European flood alert system – Part 1: Concept and development. *Hydrology and Earth System Sciences*, 13(2), 125–140. <https://doi.org/10.5194/hess-13-125-2009>
- Thober, S., Kumar, R., Sheffield, J., Mai, J., Schäfer, D., & Samaniego, L. (2015). Seasonal soil moisture drought prediction over Europe using the North American multi-model ensemble (NMME). *Journal of Hydrometeorology*, 16(6), 2329–2344. <https://doi.org/10.1175/JHM-D-15-0053.1>
- Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., et al. (2017). The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, 98(1), 163–173. <https://doi.org/10.1175/BAMS-D-16-0017.1>
- Vitart, F., & Robertson, A. W. (2018). The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *npj Climate and Atmospheric Science*, 1(1), 1–7. <https://doi.org/10.1038/s41612-018-0013-0>
- Wanders, N., Thober, S., Kumar, R., Pan, M., Sheffield, J., Samaniego, L., & Wood, E. F. (2019). Development and evaluation of a pan-European multimodel seasonal hydrological forecasting system. *Journal of Hydrometeorology*, 20(1), 99–115. <https://doi.org/10.1175/JHM-D-18-0040.1>
- Wanders, N., & Wood, E. F. (2016). Improved sub-seasonal meteorological forecast skill using weighted multi-model ensemble simulations. *Environmental Research Letters*, 11(9), 094007. <https://doi.org/10.1088/1748-9326/11/9/094007>
- Wang, Q. J., Shao, Y., Song, Y., Schepen, A., Robertson, D. E., Ryu, D., & Pappenberger, F. (2019). An evaluation of ECMWF SEAS5 seasonal climate forecasts for Australia using a new forecast calibration algorithm. *Environmental Modelling & Software*, 122, 104550. <https://doi.org/10.1016/j.envsoft.2019.104550>
- Weigel, A. P., Liniger, M. A., & Appenzeller, C. (2008). Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Quarterly Journal of the Royal Meteorological Society*, 134(630), 241–260. <https://doi.org/10.1002/qj.210>
- Xu, L., Chen, N., Zhang, X., & Chen, Z. (2020). A data-driven multi-model ensemble for deterministic and probabilistic precipitation forecasting at seasonal scale. *Climate Dynamics*, 54(7), 3355–3374. <https://doi.org/10.1007/s00382-020-05173-x>
- Yuan, X., Wood, E. F., Luo, L., & Pan, M. (2011). A first look at Climate Forecast System version 2 (CFSv2) for hydrological seasonal prediction. *Geophysical Research Letters*, 38(13). <https://doi.org/10.1029/2011GL047792>
- Yuan, X., Wood, E. F., & Ma, Z. (2015). A review on climate-model-based seasonal hydrologic forecasting: Physical understanding and system development. *WIREs Water*, 2(5), 523–536. <https://doi.org/10.1002/wat2.1088>