Skill of seasonal rainfall and temperature forecasts for East Africa

Article

Accepted Version

Young, H. R. ORCID: https://orcid.org/0000-0002-7997-9471 and Klingaman, N. P. ORCID: https://orcid.org/0000-0002-2927-9303 (2020) Skill of seasonal rainfall and temperature forecasts for East Africa. Weather and Forecasting, 35 (5). pp. 1783-1800. ISSN 0882-8156 doi: https://doi.org/10.1175/WAF-D-19-0061.1 Available at https://centaur.reading.ac.uk/91641/

Skill of seasonal rainfall and temperature forecasts for East Africa

Hannah R. Young* and Nicholas P. Klingaman

National Centre for Atmospheric Science–Climate and Department of Meteorology, University of Reading, Reading, UK

*Corresponding author. Email hannah.young@reading.ac.uk

Abstract

Skilful seasonal forecasts can provide useful information for decision makers, particularly in regions heavily dependent on agriculture, such as East Africa. We analyse prediction skill for seasonal East African rainfall and temperature one to four months ahead from two seasonal forecasting systems: the US National Centers for Environmental Prediction (NCEP) Coupled Forecast System Model Version 2 (CFSv2) and the UK Met Office (UKMO) Global Seasonal Forecast System Version 5 (GloSea5). We focus on skill for low or high temperature and rainfall, below the 25th or above the 75th percentile respectively, as these events can have damaging effects in this region. We find skill one month ahead for both low and high rainfall from CFSv2 for December-January-February in Tanzania, and from GloSea5 for September-October-November in Kenya. Both models have higher skill for temperature than for rainfall across Ethiopia, Kenya and Tanzania, two months ahead in some cases. Performance for rainfall and temperature change in the two models during certain El Niño Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD) phases, the impacts of which vary by country, season and sometimes by model. While most changes in performance are within the range of uncertainty due to the relatively small sample size in each phase, they are significant in some cases. For example, La Niña lowers performance for Kenya September-October-November rainfall in CFSv2 but does not affect skill in GloSea5.

1. Introduction

Forecasts of rainfall and temperature for upcoming months can be valuable for decision makers to anticipate and mitigate the effects of unfavourable conditions. Forecast information can guide the decisions of policymakers in agriculture and food security, water management, disaster risk reduction, emergency relief and health (Lemos et al., 2002; Vitart et al., 2012), and at local scales be useful for subsistence farmers (Patt and Gwata, 2002; Hansen et al., 2011). East Africa is particularly vulnerable to the effects of rainfall and temperature extremes, as much of the population depends on rainfed agriculture for their income: 79% in Ethiopia, 77% in Tanzania and 61% in Kenya (FAO, 2018). Crops and livestock can be affected by heat stress, drought conditions (Herrero et al., 2010), and frosts (Kotikot and Onywere, 2015), and the prevalence of disease is also influenced by both rainfall and temperature (Bandyopadhyay et al., 2012).

For forecasts to be successfully interpreted and applied in this region, it is necessary to understand the skill and reliability of ensemble forecasting systems for predicting rainfall and temperature on decision-relevant timescales. Seasonal rainfall variability over the East African region is influenced by local factors such as topography, coastal influences and lakes; regional circulation drivers such as the tropical easterly jet; and remote drivers such as the El Niño–Southern Oscillation (ENSO), the Indian Ocean Dipole (IOD) and the Madden-Julian Oscillation (MJO) (Nicholson, 2017). Successful rainfall predictions rely on the ability of seasonal forecast models to represent a range of drivers and their relationships to regional rainfall. Models also need to capture the seasonal cycle of rainfall. While some parts of the region experience one rainy season per year, such as southern Tanzania and northern Ethiopia, the near-equatorial region including northern Tanzania, Kenya and southern Ethiopia has two rainy seasons (Dunning et al., 2016; Nicholson, 2017). Successful temperature predictions also rely on models representing drivers, both local and remote, and their relationships to regional conditions. In particular, soil moisture conditions and ENSO phase are important for predicting heatwaves (Hirschi et al., 2011; Russo et al., 2016; van den Hurk et al., 2010).

a. Seasonal forecasts for East Africa

Nicholson (2017) provides a comprehensive overview of previous work on the skill of seasonal rainfall forecasts over eastern Africa. The short rains (October to November) are generally more predictable than the long rains (March to May), particularly in dynamical models. In both seasons statistical models are generally more skilful than dynamical ones, however dynamical models do outperform in some cases. For instance, Walker et al. (2019) found a dynamical model had higher skill for predicting East African rainfall than a consensus forecast based on both statistical and dynamical models.

Statistical models identify remote atmospheric and oceanic drivers with teleconnections to the region of interest, to predict conditions in the coming months. These models therefore rely on the availability of information about the relevant drivers ahead of the season of interest to make a prediction. For example, Camberlin and Philippon (2002) used observed ENSO and other predictors identified using principal component analysis to predict March to May rainfall in East Africa. Diro et al. (2008, 2011) developed skilful statistical forecasts for Ethiopia's spring and summer rains based on teleconnections from sea surface temperatures (SSTs) in regions across the globe. Funk et al. (2014) used western central Pacific and central Indian Ocean SSTs to predict East African droughts and Chen and Georgakakos (2015) forecast East African rains using SST dipoles across basins including the Mediterranean Sea, North and South Atlantic, Indian Ocean, and Arabian Sea.

Nicholson (2014) produced regression models based on SST, sea level pressure and vertical and horizontal winds at different heights to predict March-April-May, July-August-September and October-November rains in equatorial and summer rainfall regions. The summer rainfall region was defined by areas of East Africa where the maximum rainfall falls between June and September, and the equatorial rainfall region was defined where the maximum rainfall falls within March-April-May or October-November. Correlations between modelled and observed rainfall were above 0.76 up to five months ahead for October-November rainfall in both regions, two months ahead for July-August-September rainfall in the summer rainfall region, and two months ahead for March-April-May rainfall in the equatorial region. This highlighted how the ENSO spring predictability barrier (Webster and Yang, 1992) can hinder prediction of spring and summer rains at longer lead times (Nicholson, 2017). Along with later work (Nicholson, 2015), the Nicholson (2014) study also found that statistical models that used atmospheric variables on multiple levels as predictors tended to have higher skill than those that used only surface variables.

On the other hand, dynamical forecasts are predictions from models that represent the physical processes underlying weather and climate. While dynamical systems are generally less skillful than statistical models over East Africa (Nicholson, 2017), they are increasingly being developed and used, and there is evidence of skill for predicting seasonal temperature (Weisheimer and Palmer, 2014) and rainfall (Bahaga et al., 2016; Batté and Déqué, 2011; Diro et al., 2012; Dutra et al., 2013; MacLeod, 2018; Mwangi et al., 2014; Walker et al., 2019). In this paper we will contribute to greater understanding of the skill of two contemporary dynamical seasonal forecasting systems, the US National Centers for Environmental Prediction (NCEP) Coupled Forecast System Model Version 2 (CFSv2, Saha et al., 2014) and the UK Met Office (UKMO) Global Seasonal Forecast System Version 5 (GloSea5, MacLachlan et al., 2015).

We focus on the skill of CFSv2 and GloSea5 for high or low rainfall or temperature for countries in East Africa, with lead times of one to four months. While GloSea5 rainfall forecasts have been analysed by Walker et al. (2019) across East Africa at one month lead time for tercile event categories, here we analyse additionally lead times out to four months, to determine how far ahead useful predictions can be made. We also evaluate temperature forecasts, as these may also be relevant for decision-makers in the region, and we provide a comparison with CFSv2. In each case we also analyse the forecasts at country-scale, which has not been done before for these models. Alongside information at both larger regional and smaller sub-national scales, forecasts for individual countries may be useful for decision makers advising at this scale, such as humanitarian agencies, so an understanding of country-level skill is relevant in this context. We compare skill over Ethiopia, Kenya and Tanzania, three relatively large countries in the East Africa region with different seasonal cycles, particularly for rainfall.

In the next section we describe the data and methods used in this study. We present results in section 3 and in section 4 we discuss these in the context of skill of other models and future research opportunities.

2. Data and Methods

a. Re-forecast and observational data

The re-forecast data are from the NCEP CFSv2 and UKMO GloSea5 models. CFSv2 re-forecasts are started every five days starting from the 1st of January, running for nine months, for the period 1982-2011. They have four ensemble members and T126 resolution (approximately 100km). GloSea5 re-forecasts are started on the 1st, 9th, 17th and 25th of each month, with seven ensemble members, and are run for approximately seven months (216 days). We used re-forecasts for 1993-2015 at N216 resolution (approximately 60km at midlatitudes). For each model, all re-forecasts within a calendar month were merged to produce larger ensembles of re-forecasts, and for ease of comparing the models. This resulted in an ensemble size of either 20, 24 or 28 for CFSv2 (four ensemble members started on either 5, 6 or 7 start dates, depending on the month), and an ensemble size of 28 for GloSea5 (7 ensemble members started on 4 start dates). We evaluate both models at lead times of one to four months. The lead time corresponds to the time ahead of the start of the season. For example, re-forecasts for DJF at one month lead are those starting on any dates in November; re-forecasts at four months lead are those starting on any dates in August.

Rainfall data to validate the re-forecasts are taken from the Global Precipitation Climatology Project (GPCP; Adler et al., 2003) at 2.5 x 2.5 degree resolution (approximately 275km at the equator). Temperature data are near-surface (2 metre) air temperatures from the ECMWF Interim reanalysis (ERA-Interim; Dee et al., 2011) at N128 resolution (approximately 80km). Model data were regridded to the grid of the observations for gridpoint scale analyses.

b. Methods

We analysed the re-forecasts over the standard meteorological seasons (DJF, MAM, JJA, SON), with lead times one to four months ahead of the season of interest. Analyses were carried out at gridpoint scale and at country-average scale. Results are shown at country scale as, due to geopolitical considerations, information at this scale is sometimes used to guide decisions in humanitarian or other sectors. Results are shown for DJF and JJA for temperature, and for the main rainy seasons in each country. It is meteorological practice in Ethiopia to partition rainfall in three four-month seasons: FMAM, JJAS (the main rainy season) and ONDJ. For consistency with the analysis for other countries we here retain use of three-month periods and focus on the two seasons where these are a subset of the longer seasons: JJA and MAM.

Biases in the mean of the forecasts compared to the observations were first calculated. Mean biases for each lead time (in weeks) were then removed from the hindcasts before the rest of the analysis was carried out. Anomaly Correlation Coefficients (ACC) were also calculated for the ensemble mean, which is the correlation between the forecasts and observations, each taken as an anomaly from their own climatology. These are temporal correlations, and are calculated both at gridpoint and country scale.

We also calculated the Brier Skill Score (BSS), to compare the skill of the re-forecasts to a climatological forecast. We use a climatological forecast that always forecasts the expected likelihood of an event, e.g. events below the 25th percentile are expected to occur 25% of the time, so are forecast with probability 0.25. The Brier Skill Score is

$$BSS = 1 - \frac{BS}{BS_{ref}}$$

where BS is the Brier Score of the forecast and BS_{ref} is the Brier Score of a reference (climatological) forecast. The Brier Score is defined as

$$BS = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2$$

where N is the number of forecasts, f_t are the forecast probabilities of an event and o_t are the indicators of whether the event was observed (1 if it occurred; 0 if it did not occur), at each forecast instance *t*. A BSS above 0 indicates skill greater than that of a climatological forecast. Finally, reliability diagrams were also produced, comparing the forecast probability of an event with its observed frequency. Using five bins to partition the forecast probabilities, the observed relative frequency for each bin is plotted against the average probability value of all forecasts within the bin (Bröcker and Smith, 2007). Reliability is shown aggregated over leads of one to two months and three to four months. Both BSSs and reliability diagrams were calculated using country-averaged predicted and observed rainfall and temperature.

BSSs and reliability diagrams were calculated for events below the 25th percentile, between the 25th and 75th percentiles, and above the 75th percentile. These were calculated using percentiles based on the re-forecast rainfall and temperature distributions (e.g. f_t= forecast probability of an event above the re-forecasts' 75th percentile, o_t=whether observations were above the observations' 75th percentile), as this calibrates for errors in the models' distributions of rainfall and temperature. Re-forecast percentiles were calculated for each lead time (in months) for the BSS and reliability calculations. To estimate uncertainty in the BSSs, the original ensemble for each year was bootstrapped 1000 times, sampling with replacement to produce ensembles the same size as the original. The re-forecast percentiles were recalculated each time along with the BSS in order to establish the 5th and 95th percentiles of the uncertainty range.

Biases and ACCs were also compared for different El Niño Southern Oscillation (ENSO, Trenberth, 1997) and Indian Ocean Dipole (IOD, Saji et al., 1997) phases, as these large scale circulation changes have teleconnections to East Africa (e.g. Ropelewski and Halpert, 1987; Nicholson and Kim, 1997; Black et al., 2003; Black, 2005) and may affect forecast skill (e.g. Goddard and Dilley, 2005). ENSO indices were retrieved from NOAA and IOD indices from JAMSTEC. Tercile categories were used to determine negative (below 33rd percentile), neutral (between 33rd and 66th percentile) and positive (above 66th percentile) years for each season. Forecasts were then categorised using the observed IOD and ENSO indices for the season. Biases are shown for these subsets, averaging over 11 (La Niña, positive and negative IOD) or 12 (El Niño) years for the observations, 10 years for CFSv2 and 8 years for GloSea5. In order to estimate whether the conditional ACCs for ENSO and IOD phases lie outside the range of sampling uncertainty, the original uncategorised ensemble was also bootstrapped 1000 times. Each time the same number of years were sampled as were in the ENSO or IOD tercile categories (10 for CFSv2 and 8 for GloSea5) and the ACC calculated, in order to establish the 5th and 95th percentiles of the sampling uncertainty range.

3. Results

a. Model biases

Rainfall results for MAM and SON show that both models have clear mean-state rainfall biases over East Africa (Fig. 1a). CFSv2 has a dry bias in MAM in regions of observed high rainfall, and in SON there is a dry bias across the western part of the region. These biases increase slightly from lead-1 to lead-3. CFSv2 also produces wet biases in the northern Ethiopian highlands of up to 3mm/day, suggesting the simulated rainfall is too sensitive to orography. GloSea5 rainfall biases are much smaller in MAM. In SON, GloSea5 has a wet bias of over 3mm/day across the region, corresponding to the wet bias found by Walker et al. (2019) for the OND season. The wet bias increases slightly from lead-1 to lead-3. Both models also show large rainfall biases over the Indian Ocean. CFSv2 interannual rainfall variability is generally too low across East Africa, while GloSea5 has too little variance where observed variance is high, but too much variance in other parts of the region (not shown).

Temperatures vary little throughout the year in Ethiopia, Kenya and Tanzania (Fig. 1b). CFSv2 is too cool over the Horn of Africa, and additionally over Tanzania and the southern part of the region in JJA, but too warm into central Africa. Temperatures over Ethiopia, Kenya and Tanzania are mostly up to 2°C too cool and these biases change little with lead time. GloSea5 is too cool by up to 5°C in the northern part of the region and too warm across the southern part in DJF, and in JJA the warm bias is further north and includes Ethiopia, with biases above 2°C. There is also a strong warm bias over Lake Victoria. The biases over Ethiopia, Kenya and Tanzania change little with lead time but the warm bias over central Africa in JJA decreases from lead-1 to lead-3. Both models also have variability which is too high where observed variance is low, and too low where observed variance is high (not shown).

1) BIASES CONDITIONED ON ENSO AND IOD

Rainfall and temperature in East Africa are affected by SST variations in the ENSO and IOD regions. Fig. 2a shows that observed rainfall in SON over Kenya and Tanzania is generally above-normal in El Niño and positive IOD phases, by up to 0.8 mm/day, while in negative IOD rainfall is up to 0.6 mm/day below-normal. In MAM changes are smaller, but there is slightly above normal rainfall in Ethiopia during El Niño, and across all three countries during negative IOD. During positive IOD events there is below normal rainfall over the region. ENSO and IOD teleconnections to MAM rainfall are generally weak (e.g. Liebmann et al., 2014; Mutai and Ward, 2000; Vellinga and Milton, 2018), however the effects seen here are likely due to the tercile SST categories defined over a relatively short time period. Conditional rainfall biases in CFSv2 (biases in SST phases relative to the bias in all years) are based on only 10 years of data but remain similar in MAM (Fig. 2b) to the unconditional biases (i.e., all years), although the dry bias is increased in negative IOD. In SON, the dry bias over Kenya is amplified during El Niño and positive IOD events, but decreased during the opposite phases. Fig. 2c shows GloSea5 rainfall biases, which are based on only 8 years of data, appear to change relatively little with ENSO and IOD phase in MAM. However in SON the wet bias over the region is increased during La Niña and negative IOD events, but reduced during El Niño and positive IOD phases.

Fig. 3a shows that in El Niño and positive IOD phases, the East African region generally experiences warm anomalies in JJA, particularly over Ethiopia which is up to 0.8°C warmer during El Niño. The region is cooler in JJA during La Niña, while differences in DJF are smaller across the phases. CFSv2 has conditional temperature biases largely similar to the unconditional biases over the East African region (Fig. 3b). Fig. 3c shows conditional biases are also similar to unconditional biases in GloSea5, however the cool bias over Ethiopia and Kenya in DJF is slightly reduced in El Niño and positive IOD years and amplified in the opposite phases. Again these biases are based on small sample sizes.

b. Regional performance

Fig. 4 shows the Anomaly Correlation Coefficients (ACC) for precipitation and temperature at leads 1 and 3 for the ensemble means of the two models. ACCs are lower for rainfall than temperature for both models. For MAM, CFSv2 has rainfall ACC values above 0.4 over northern Ethiopia at lead-1, but by lead-3 these values have decreased, and there is no significant correlation between the model and the observations over Kenya and Tanzania (Fig. 4a). In SON, ACC values are above 0.4 over eastern Ethiopia and Kenya and above 0.2 across Tanzania at lead-1, and these decrease over the eastern part of the region by lead-3. GloSea5 ACCs show similar spatial patterns across the countries of interest in MAM, but with higher magnitudes at lead-1 than CFSv2, particularly over Kenya. However in SON the region over eastern Ethiopia and Kenya has higher ACC values than CFSv2 at lead-1, at over 0.6, and this extends out to lead-3. The region of high correlation over land in GloSea5 is broadly similar to the equatorial region analysed by Nicholson (2014), whose regression model had a correlation of 0.78 with observations at 2 months lead.

ACCs for DJF temperature in CFSv2 are positive over Ethiopia, Tanzania and parts of Kenya, with particularly high values over 0.6 in northern Ethiopia at both lead-1 and lead-3 (Fig. 4b). JJA results are similar, but with higher correlation over Kenya at lead-1 and lower correlation across Ethiopia and Tanzania at lead-3. GloSea5 results for JJA temperature also show positive ACCs across most of the region, with values above 0.6 over Tanzania out to lead-3. Lower ACCs are found in DJF, and particularly over parts of Kenya there is no significant correlation between the model and the observations as is the case in CFSv2.

This analysis is based on different time periods for the two models to make full use of the available data and maximise the sample size. However when using the same time period for both models (1993-2011), the spatial patterns of the ACCs are very similar (Fig. S1 in Supplemental Material). While there are some regions where the magnitude decreases, for example GloSea5 JJA temperature ACCs over Kenya and Tanzania are slightly lower when the shorter time period is used, overall these changes are small.

c. Country-level skill: Ethiopia

For country-level skill, Figures 5-12 show results for leads 1-4, while Tables 1 and 2 provide significance testing for lead-1 and lead-2.

We analyse the skill for forecasts averaged over Ethiopia for key 3-month seasons. The CFSv2 rainfall ACC is 0.48 at lead-1 for JJA, but for MAM ACCs are lower and not significant at lead-1 (Fig. 5a, Table 1). ACCs decrease with lead time out to lead-4 in MAM and lead-3 in JJA. In El Niño and positive IOD phases CFSv2 ACCs are higher than the unconditioned ACC in MAM, and lower in the opposite phases. The conditioned ACC only lies outside the 5th to 95th percentile uncertainty range from bootstrapping the original ensemble for negative IOD years at lead-1 and lead-2, suggesting that the difference is significant at the 90% confidence level at these leads. GloSea5 ACC values are slightly higher than those from CFSv2 in JJA out to lead-3, and are also much lower and not significant in MAM (Fig. 5a, Table 1). ACCs are higher in La Niña and negative IOD phases in JJA, although this lies within the range of uncertainty. CFSv2 rainfall BSS values are around 0 for all event categories in both seasons, except for dry events (below 25th percentile) at lead-2 in JJA where the BSS and its uncertainty range is positive, showing a slight improvement over a climatological forecast (Fig. 6a, Table 1). GloSea5 BSS values are also very close to 0, indicating no improvement over a climatological forecast. Reliability curves from CFSv2 and GloSea5 for rainfall generally show little change with lead time, and are closest to the 1:1 line in both models for wet events in MAM and dry events in JJA, signifying that forecast probabilities correspond well to the observed frequencies of events (Fig. 7a). Reliability for average (25th-75th percentile) events is poor in MAM for both models (Fig. S2a, S3a in Supplemental Material).

ACCs for temperature are higher than for rainfall in both models, and remain around 0.5 for all lead times in DJF but decrease with lead time in JJA, from 0.66 in CFSv2 and 0.88 in GloSea5 (Fig. 9a, Table 2). La Niña and negative IOD phases have higher ACCs in DJF in CFSv2, while positive IOD ACCs are negative and outside the range of uncertainty. El Niño leads to higher ACCs in DJF and JJA in GloSea5, however this conditional ACC does not lie outside the uncertainty range, suggesting the difference is not significant. CFSv2 BSSs and their uncertainty ranges are positive for cool (below 25th percentile) events in DJF at all lead times, and warm events (above 75th percentile) in JJA at lead-1 (Fig. 10a, Table 2). GloSea5 BSSs and their uncertainty ranges are positive at lead-1 and lead-2 for average events in DJF and warm events in JJA, and at lead-1 for average events in JJA. Reliability diagrams for temperature from CFSv2 show slight improvement from lead 3-4 to lead 1-2, and match the 1:1 line reasonably well for cool events in DJF and JJA and warm events in JJA, but they tend to be too shallow (Fig. 11a). Reliability diagrams for GloSea5 (Fig. 12a) show a similar picture. For the average (25th-75th percentile) category, CFSv2 forecast probabilities showing little relationship with the observed frequencies, but curves are close to the 1:1 line in GloSea5, particularly for DJF (Fig. S4a, S5a in Supplemental Material).

d. Country-level skill: Kenya

Kenya receives most rainfall in MAM and SON. CFSv2 has slightly higher correlations in MAM than SON, while GloSea5 ACC values are slightly higher than CFSv2 at lead-1 for MAM, and much higher for SON with ACCs above 0.7 out to lead-2 (Fig. 5b, Table 1). El Niño has a strong forcing on SON rainfall in East Africa, and is associated with above normal conditions (Black, 2005). Separating out years into ENSO and IOD phases suggests that CFSv2 and GloSea5 have consistently greater performance in El Niño years in SON, however this is within the range of uncertainty. While GloSea5 has a slight decrease in performance in La Niña years in SON, CFSv2 exhibits a much greater decrease that consistently lies outside the uncertainty range and the model is strongly anti-correlated with the observations by lead-3 (Fig. 5b). CFSv2 BSS values are very close to 0 for all events and lead times (Fig. 6b, Table 1), showing skill similar to climatological forecast skill. GloSea5 skill is above that of a climatological forecast at lead-1 for wet events in MAM and SON and dry events in SON (Fig. 6b, Table 1). CFSv2 rainfall reliability changes little with lead time and is best for dry events in MAM and wet events in SON (Fig. 7b). GloSea5 reliability curves are particularly close to the 1:1 line for wet events in SON out to lead 3-4 (Fig. 8b). Reliability for average events is better in CFSv2 for MAM and GloSea5 for SON (Fig. S2b, S3b in Supplemental Material).

CFSv2 temperature ACC values remain around 0.5 at all lead times in JJA, and are slightly lower in DJF (Fig. 9b, Table 2). GloSea5 ACCs are lower than CFSv2 in DJF but similar in JJA. ACCs in negative IOD phases are slightly greater than in all years in JJA in both models, and outside the range of uncertainty at longer lead times in both. CFSv2 also has much lower ACCs with negative correlations during positive IOD years in JJA. CFSv2 has positive BSSs, showing an improvement over a climatological forecast, for cool events in DJF at lead-1 and in JJA at lead-1 and lead-2 (Fig. 10b, Table 2). GloSea5 has BSSs above zero only for warm events in JJA at lead-1 and lead-2. Reliability for temperature from CFSv2 is best for cool events in JJA at leads 1-2, whereas curves for warm events are too shallow (Fig.

11b). GloSea5 reliability is poor, particularly for cool events (Fig. 12b). Reliability is generally poor for average events, although the curve at lead 1-2 for JJA in CFSv2 is close to the 1:1 line (Fig. S4b, S5b in Supplemental Material).

e. Country-level skill: Tanzania

In Tanzania in DJF, both models have ACCs for rainfall above 0.5 at lead-1; this decreases with lead time in CFSv2 but remains fairly constant in GloSea5 (Fig. 5c, Table 1). ACCs are much lower in both models in MAM, and are mostly around or below zero. ENSO and IOD phase has little consistent effect on the ACC in DJF in either model, but generally ACCS are lower in negative IOD than the unconditioned case. In MAM, La Niña increases the ACC at all lead times in both models, but this is within the sampling uncertainty range. El Niño decreases skill in MAM in out to lead-3 in both models, and this is outside the range of uncertainty at lead-1. CFSv2 has positive BSSs for dry and wet events at lead-1 in DJF, but other events and leads are close to or below 0 (Fig. 6c, Table 1). The uncertainty range of GloSea5 BSSs at lead-1 and lead-2 does not lie above 0 in either season (Fig. 6c). CFSv2 reliability curves are close to the 1:1 line for dry events in DJF, and wet events in DJF at lead 1-2 (Fig. 7c). Reliability for rainfall is also better in GloSea5 in DJF than MAM, where dry and wet events are both only predicted with low probabilities (Fig. 8c). For average events, forecast probabilities show little relationship with observed frequencies in CFSv2, but the curve is close to the 1:1 line in DJF at lead 1-2 in GloSea5 (Fig. S2c, S3c in Supplemental Material).

Temperature ACCs are around 0.5 at lead-1 for both seasons in CFSv2, and decrease slightly with lead time (Fig. 9c, Table 2). ACCs are lower in DJF in GloSea5 than in CFSv2, but higher in JJA. ENSO and IOD phase influences on the ACCs mostly lie within the range of uncertainty. Both models show slight improvements in negative IOD years in JJA, and there is a decrease in ACC when in El Niño phase for JJA in CFSv2 which does lie outside the range of sampling uncertainty. BSS values and their uncertainty ranges are positive in CFSv2 at lead-1 and lead-2 for cool events in DJF and cool and average events in JJA, and at lead-1 for warm events in DJF (Fig. 10c, Table 2), signifying higher skill than a climatological forecast. GloSea5 has less skill over Tanzania in DJF, with negative BSS values for lead-1 for all temperature categories, but higher skill than a climatological forecast at lead-1 and lead-2 for cool and warm events in JJA. Reliability diagrams for temperature from CFSv2 and GloSea5 show curves are closest to the 1:1 line for warm events in DJF and cool events in JJA at lead 1-2 (Fig. 11c, 12c). Reliability for average events is better for JJA than DJF in both models (Fig. S4c, S5c in Supplemental Material).

4. Discussion and conclusions

a. Rainfall skill

Country ACCs and BSSs for rainfall forecasts at leads 1 and 2 are summarised in Table 1. CFSv2 temperature anomalies significantly correlate with observations only at lead-1 in JJA in Ethiopia, and lead-1 and lead-2 in MAM in Kenya and DJF in Tanzania. Skill is mostly similar to that of a climatological forecast, only showing improvement above this at lead-1 for dry and wet events in DJF in Tanzania, and at lead-2 for dry events in JJA in Ethiopia.

GloSea5 skill is generally similar to CFSv2 for rainfall, also having positive anomaly correlations for JJA in Ethiopia, MAM in Kenya and DJF in Tanzania. However GloSea5 also has strong correlations for SON rainfall in Kenya at both lead-1 and lead-2. Skill is only above that of a climatological forecast for wet events in MAM and wet and dry events in SON in Kenya at lead-1.

Both models have similar skill for average events to a climatological forecast. Average events are mostly forecast by around 50% of ensemble members (Fig. S2, S3 in Supplemental Material), which matches the definition of occurring 50% of the time. This indicates the models have low sharpness - the ability to generate forecast probabilities that are different to the climatological frequency of the event. Where the models do exhibit skill above that of a climatological forecast this is for below normal and above normal rainfall events. This is similar to results from Diro et al. (2008, 2011) for a statistical model based on SST anomalies, and from Walker et al. (2019) for GloSea5 which found higher skill for the outer tercile categories over a large East Africa region. In these skilful cases, the models predict the events with probabilities that are different to the 25% frequency defining dry and wet events, demonstrating sharpness. For example, CFSv2 has skill for dry and wet events in Tanzania in DJF, and GloSea5 has skill for dry and wet events in Kenya in SON. In both of these cases the distribution of forecast probabilities at lead 1-2 peaks at the lowest values, suggesting a high frequency of forecasts giving less than climatological chances of the events, along with some forecasts giving probabilities above 25% (Figure 7, Figure 8).

Walker et al. (2019) found skill for the short rains (OND) over East Africa in GloSea5 at one month ahead, particularly for upper tercile events, along with an increase in skill and reliability of East African rainfall when GloSea5 forecasts an IOD event. ECMWF System 4 also has good rainfall reliability in this region (Weisheimer and Palmer, 2014). The reliability in these analyses appears consistent with the reliability results shown here, in particular for wet events in SON in Kenya. Differences in the reliability from these previous studies are likely to be because here the country-average has been used, rather than an aggregation over the gridpoints in each region. This does mean the sample sizes are smaller, making the reliability estimates less robust and more susceptible to random variations due to undersampling, however the estimates still support those from Walker et al. (2019).

Forecasts for the long rains season (MAM) have been shown to be generally less skilful than forecasts for the short rains (ON, e.g. Nicholson, 2014; Nicholson, 2017; MacLeod, 2018). ACC results in Fig. 4a show that over the East Africa region GloSea5 has greater correlation in SON than in MAM over the eastern part of the region, however CFSv2 at lead-1 has greater performance for MAM than SON over Ethiopia and Kenya. This is reflected in the country scale results, where ACC values for Kenya from CFSv2 are higher in MAM than SON, but values from GloSea5 are higher in SON (Table 1). GloSea5 also has positive BSSs for dry and wet events in

SON and wet events in MAM in Kenya (Table 1), which suggests this model may be able to provide skilful forecast information particularly for the short rains season.

b. Temperature skill

Country ACCs and BSSs for temperature forecasts at leads 1 and 2 are summarised in Table 2. Temperature forecast skill is generally higher than rainfall, with more significant ACCs and BSS uncertainty ranges above zero in both models. CFSv2 ensemble mean temperature anomalies significantly positively correlate with observed anomalies at lead-1 for all countries and seasons, and at lead-2 for all except DJF in Kenya. CFSv2 has skill at predicting cool events in all countries, with the uncertainty above zero indicating skill above that of a climatological forecast in all seasons except JJA in Ethiopia, and mostly out to lead-2. Skill is lower for normal events, where it is only above a climatological forecast for JJA in Tanzania, and for warm events where it is only above for JJA in Ethiopia and DJF in Tanzania at lead-1.

GloSea5 temperature anomaly correlations are higher than CFSv2 for JJA in all countries at lead-1 and lead-2, positive and similar to CFSv2 for DJF in Ethiopia, but not significant for DJF in Kenya and Tanzania. Skill for cool events is lower than CFSv2, with skill above a climatological forecast only for JJA in Tanzania. However, GloSea5 has skill for normal events at lead-1 in DJF and JJA in Ethiopia where CFSv2 does not, and has skill for warm events at lead-1 and lead-2 in JJA in all three countries.

Both CFSv2 and GloSea5 have some skill at predicting temperatures in East Africa at 1 and 2 month lead times, CFSv2 particularly for cool events and GloSea5 particularly for warm events in JJA. Few other studies have analysed seasonal temperature forecast skill over this particular region. However ECMWF System 4 has been shown to have good reliability for warm (upper tercile) and cold (lower tercile) DJF and JJA seasons over a large East African region one month ahead (Weisheimer and Palmer, 2014). Combined with the results shown here this suggests that dynamical models are able to make skilful and reliable temperature predictions for this region in some seasons.

c. Limitations and future opportunities

In this study skill and reliability may be limited by the size of the reforecast datasets (30 years for CFSv2 and 23 years for GloSea5), particularly when the record is split into ENSO and IOD phases; most of the differences between conditional and unconditional performance were not significant. Results may also be sensitive to the time periods which do not match between the models. A much longer re-forecast set would be required to have enough years in each ENSO and IOD phase to substantially reduce uncertainty; around 20 years are currently used to characterise seasonal forecast skill so this may need to be up to three times as long.

The impact of ENSO and IOD phase on prediction performance is often consistent across the two models. For example, increased skill for MAM rainfall in Kenya in positive IOD years and JJA temperature in Tanzania in negative IOD years. However impacts do differ across the two models in some countries and seasons. For example for JJA rainfall in Ethiopia, negative IOD phase improves performance in GloSea5 but has little effect in CFSv2, while for SON rainfall in Kenya, La Niña significantly decreases skill in CFSv2 but has little effect in GloSea5. For JJA temperature in Kenya, La Niña increases skill in CFSv2 but decreases skill in GloSea5, while positive IOD phase significantly decreases skill in CFSv2 but has little impact in GloSea5. Many of the impacts in both models are within the range of sampling uncertainty. With a limited sample size, it is difficult to rule out the effect of other influences and isolate the role of ENSO and IOD. While terciles are used for ENSO and IOD categorisation in each season so there are equal numbers of years in each phase, SST variability is likely to be higher in some seasons, for example El Niño and La Niña events typically peak in DJF, and there may be stronger influences on rainfall in certain seasons. Longer reforecast datasets would aid more robust estimates of changes in performance and skill based on the phase of ENSO and IOD variability.

Bahaga et al. (2016) showed that a realistic representation of the relationship with the Indian Ocean Dipole is important in capturing the variability of the rains in East Africa. Walker et al. (2019) found GloSea5 represents the correlations between East African rainfall and ENSO and IOD region SSTs well, but may have atmosphere-ocean coupling which is too weak. Further work should investigate the teleconnections to East Africa in CFSv2, along with model skill at replicating the SST changes in each year to also help understand why different ENSO and IOD phases may improve skill in each model. There may also be limitations due to the observational datasets used for validation; for example GPCP has been shown to underestimate high rainfall intensities over the region (Dinku et al., 2007; Kimani et al., 2017).

Further work should investigate skill for smaller countries in the East African region, or for sub-national or trans-national regions with consistent meteorological conditions (e.g., timing of the wet season). Our skill results may be artificially low due to variations within our target countries in meteorological regimes, e.g. differing rainy seasons in northern and southern Ethiopia (Dunning et al., 2016). Here we analysed the country scale as these forecasts are applied in humanitarian contexts where decisions are in some cases made at country scales for geopolitical reasons. However, model biases and performance vary across countries (Fig. 1a, Fig. 4a), so the country-scale may reduce skill where it is high only in parts of the countries. Downscaling may also help improve skill in this region (Diro et al., 2012; Kipkogei, 2017).

In general, it still appears that statistical models, for example based on SST anomalies, may provide better skill at forecasting rainfall over the region. An important addition to the literature would be rigorous comparisons of statistical and dynamical model forecasts for the region, which are currently lacking. Models such as those from Chen and Georgakakos (2015) and Nicholson (2014, 2015) have been shown to produce reliable and skilful forecasts out to months ahead using indices such as wind, SST, and SLP. The results here show that forecasts for rainfall at country scale are country, season and event category dependent and in many cases skill is very similar to that of a climatological forecast. However both dynamical models may provide useful information. In particular, CFSv2 has skill for dry and wet events in DJF in Tanzania at lead-1, while GloSea5 has skill for wet events in MAM and dry and wet events in SON in Kenya at lead-1 where CFSv2 does not. Temperature forecast skill is generally higher than rainfall, out to lead-2 in some cases from both CFSv2 and GloSea5. In particular, CFSv2 has skill for cool events in both DJF and JJA while GloSea5 has skill for warm events in JJA. Further understanding of the models analysed here would require investigating the underlying causes of the differences in skill between the CFSv2 and GloSea5 models, for example through identifying common factors between years when one model strongly outperforms the other, and comparing to other dynamical seasonal forecasting systems.

Acknowledgements

Hannah Young was funded by a grant from the UK Department for International Development, via the Science for Humanitarian and Emergency Assistance and Response (SHEAR) programme. Nicholas Klingaman was funded by an Independent Research Fellowship from the Natural Environment Research Council (NE/L010976/1). The authors are grateful for productive discussions with Prof. Andrew Charlton-Perez. The authors thank NCEP and the Met Office for providing re-forecast data. Data storage and analysis facilities were provided by the JASMIN facility (https://www.jasmin.ac.uk).

References

Adler, R. F., and Coauthors, 2003: The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present). *J. Hydrometeor*, **4**, 1147–1167, https://doi.org/10.1175/1525-7541(2003)004%3C1147:TVGPCP%3E2.0.CO;2.

Bahaga, T. K., F. Kucharski, G. M. Tsidu, and H. Yang, 2016: Assessment of prediction and predictability of short rains over equatorial East Africa using a multi-model ensemble. *Theor. Appl. Climatol.*, **123**, 637–649, https://doi.org/10.1007/s00704-014-1370-1.

Bandyopadhyay, S., S. Kanji, and L. Wang, 2012: The impact of rainfall and temperature variation on diarrheal prevalence in Sub-Saharan Africa. *Appl. Geogr*., **33**, 63–72, https://doi.org/10.1016/j.apgeog.2011.07.017.

Batté, L., and M. Déqué, 2011: Seasonal predictions of precipitation over Africa using coupled ocean-atmosphere general circulation models: Skill of the ENSEMBLES project multimodel ensemble forecasts. *Tellus A*, **63**, 283–299, https://doi.org/10.1111/j.1600-0870.2010.00493.x.

Black, E., 2005: The relationship between Indian Ocean sea–surface temperature and East African rainfall. *Philos. T. R. Soc. A.*, **363**, 43–47, https://doi.org/10.1098/rsta.2004.1474.

Black, E., J. Slingo, and K. R. Sperber, 2003: An observational study of the relationship between excessively strong short rains in coastal East Africa and Indian Ocean SST. *Mon. Wea. Rev.*, **131**, 74–94, https://doi.org/10.1175/1520-0493(2003)131%3C0074:AOSOTR%3E2.0.CO;2.

Bröcker, J., and L. A. Smith, 2007: Increasing the reliability of reliability diagrams. *Weather Forecast*., **22**, 651–661, https://doi.org/10.1175/WAF993.1.

Camberlin, P., and N. Philippon, 2002: The East African March–May rainy season: Associated atmospheric dynamics and predictability over the 1968–97 period. *J. Climate*, **15**, 1002–1019, https://doi.org/10.1175/1520-0442(2002)015%3C1002:TEAMMR%3E2.0.CO;2.

Chen, C.-J., and A. P. Georgakakos, 2015: Seasonal prediction of East African rainfall. *Int. J. Climatol.,* **35**, 2698–2723, https://doi.org/10.1002/joc.4165.

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.

Dinku, T., P. Ceccato, E. Grover-Kopec, M. Lemma, S. J. Connor, and C. F. Ropelewski, 2007: Validation of satellite rainfall products over East Africa's complex topography. *Int. J. Remote Sens.*, **28**, 1503–1526, https://doi.org/10.1080/01431160600954688.

Diro, G. T., E. Black, and D. I. F. Grimes, 2008: Seasonal forecasting of Ethiopian spring rains. *Meteor. Appl.*, **15**, 73–83, https://doi.org/10.1002/met.63.

Diro, G. T., D. I. F. Grimes, and E. Black, 2011: Teleconnections between Ethiopian summer rainfall and sea surface temperature: Part II: Seasonal forecasting. *Climate Dyn.*, **37**, 121–131, https://doi.org/10.1007/s00382-010-0896-x.

Diro, G. T., A. M. Tompkins, and X. Bi, 2012: Dynamical downscaling of ECMWF Ensemble seasonal forecasts over East Africa with RegCM3. *J. Geophys. Res.*, **117**, D16103, https://doi.org/10.1029/2011JD016997.

Dunning, C. M., E. C. L. Black, and R. P. Allan, 2016: The onset and cessation of seasonal rainfall over Africa. *Geophys. Res.: Atmos.*, **121**, 11405–11424, https://doi.org/10.1002/2016JD025428.

Dutra, E., L. Magnusson, F. Wetterhall, H. L. Cloke, G. Balsamo, S. Boussetta, and F. Pappenberger, 2013: The 2010–2011 drought in the Horn of Africa in ECMWF reanalysis and seasonal forecast products. *Int. J. Climatol.*, **33**, 1720–1729, https://doi.org/10.1002/joc.3545.

FAO, 2018: FAOSTAT Country Indicators. Accessed 27 June 2018, http://www.fao.org/faostat/en/#country.

Funk, C., A. Hoell, S. Shukla, I. Blade, B. Liebmann, J. B. Roberts, F. R. Robertson, and G. Husak, 2014: Predicting East African spring droughts using Pacific and Indian Ocean sea surface temperature indices. *Hydrol. Earth Syst. Sc.*, **18**, 4965–4978, https://doi.org/10.5194/hess-18-4965-2014.

Goddard, L., and M. Dilley, 2005: El Niño: catastrophe or opportunity. *J. Climate*, **18**, 651–665, https://doi.org/10.1175/JCLI-3277.1.

Hansen, J. W., S. J. Mason, L. Sun, and A. Tall, 2011: Review of seasonal climate forecasting for agriculture in Sub-Saharan Africa. *Exp. Agr.,* **47**, 205–240, https://dx.doi.org/10.1017/S0014479710000876.

Herrero, M., C. Ringler, J. van de Steeg, P. Thornton, T. Zhu, E. Bryan, A. Omolo, J. Koo, and A. Notenbaert, 2010: Climate variability and climate change and their impacts on Kenya's agricultural sector. Nairobi, Kenya. ILRI. https://cgspace.cgiar.org/bitstream/handle/10568/3840/climateVariability.pdf

Hirschi, M., S. I. Seneviratne, V. Alexandrov, F. Boberg, C. Boroneant, O. B. Christensen, H. Formayer, B. Orlowsky, and P. Stepanek, 2011. Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. *Nat. Geosci.*, **4**, 17–21, https://doi.org/10.1038/ngeo1032

Kimani, M. W., J. C. B. Hoedjes, and Z. Su, 2017: An Assessment of Satellite-Derived Rainfall Products Relative to Ground Observations over East Africa. *Remote Sens.*, **9**, 430, https://doi.org/10.3390/rs9050430.

Kipkogei, O., A. M. Mwanthi, J. B. Mwesigwa, Z. K. K. Atheru, M. A. Wanzala, and G. Artan, 2017: Improved Seasonal Prediction of Rainfall over East Africa for Application in Agriculture: Statistical Downscaling of CFSv2 and GFDL-FLOR. *J. Appl. Meteor. Clim.*, **56**, 3229–3243, https://doi.org/10.1175/JAMC-D-16-0365.1.

Kotikot, S. M., and S. M. Onywere, 2015: Application of GIS and remote sensing techniques in frost risk mapping for mitigating agricultural losses in the Aberdare ecosystem, Kenya. *Geocarto International*, **30**, 104–121, https://doi.org/10.1080/10106049.2014.965758.

Lemos, M. C., T. J. Finan, R. W. Fox, D. R. Nelson, and J. Tucker, 2002: The use of seasonal climate forecasting in policymaking: lessons from Northeast Brazil. *Clim. Change*, **55**, 479–507, https://doi.org/10.1023/A:1020785826029.

Liebmann, B., M. P. Hoerling, C. Funk, I. Bladé, R. M. Dole, D. Allured, X. Quan, P. Pegion, and J. K. Eischeid, 2014: Understanding recent eastern Horn of Africa rainfall variability and change. *J. Climate*, **27**, 8630–8645, https://doi.org/10.1175/JCLI-D-13-00714.1

MacLachlan, C., and Coauthors, 2014: 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.

MacLeod, D., 2018: Seasonal predictability of onset and cessation of the east African rains. *Wea. Clim. Extremes*, **21**, 27–35, https://doi.org/10.1016/j.wace.2018.05.003.

Murphy, A. H., and E. S. Epstein, 1989: Skill scores and correlation coefficients in model verification. *Mon. Wea. Rev.*, **117**, 572–582, https://doi.org/10.1175/1520-0493(1989)117%3C0572:SSACCI%3E2.0.CO;2.

Mutai, C. C., and M. N. Ward, 2000: East African rainfall and the tropical circulation/convection on intraseasonal to interannual timescales. *J. Climate*., **13**, 3915–3939, https://doi.org/10.1175/1520-0442(2000)013%3C3915:EARATT%3E2.0.CO;2

Mwangi, E., F. Wetterhall, E. Dutra, F. Di Giuseppe, and F. Pappenberger, 2014: Forecasting droughts in East Africa. *Hydrol. Earth Syst. Sc.*, **18**, 611–620, https://doi.org/10.5194/hess-18-611-2014, 2014.

Nicholson, S. E., 2014: The Predictability of Rainfall over the Greater Horn of Africa. Part I: Prediction of Seasonal Rainfall. *J. Hydrometeorol.*, **15**, 1011–1027, https://doi.org/10.1175/JHM-D-13-062.1.

Nicholson, S. E., 2015: The predictability of rainfall over the Greater Horn of Africa. Part II: Prediction of monthly rainfall during the long rains. *J. Hydrometeorol.*, **16**, 2001–2012, https://doi.org/10.1175/JHM-D-14-0138.1.

Nicholson, S. E., 2017: Climate and climatic variability of rainfall over eastern Africa. *Rev. Geophys.*, **55**, 590–635, https://doi.org/10.1002/2016RG000544.

Nicholson, S. E., and J. Kim, 1997: The relationship of the El Niño–southern oscillation to African rainfall. *Int. J. Climatol.*, **17**, 117–135, https://doi.org/10.1002/(SICI)1097-0088(199702)17:2%3C117::AID-JOC84%3E3.0.CO;2-O.

Patt, A., and C. Gwata, 2002: Effective seasonal climate forecast applications: examining constraints for subsistence farmers in Zimbabwe. *Global Environ. Change*, **12**, 185–195, https://doi.org/10.1016/S0959-3780(02)00013-4.

Ropelewski, C. F., and M. S. Halpert, 1987: Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Mon. Wea. Rev.*, **115**, 1606–1626, https://doi.org/10.1175/1520-0493(1987)115%3C1606:GARSPP%3E2.0.CO;2.

Russo, S., A. F. Marchese, J. Sillmann, and G. Immé, 2016. When will unusual heat waves become normal in a warming Africa? *Environ. Res. Lett.*, **11**, 054016, http://dx.doi.org/10.1088/1748-9326/11/5/054016

Saha, S., 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.

Saji, N. H., B. N. Goswami, P. N. Vinayachandran, and T. Yamagata, 1999: A dipole mode in the tropical Indian Ocean. *Nature*, **401**, 360–363, https://doi.org/10.1038/43854.

Trenberth, K. E., 1997: The Definition of El Niño. *B. Am. Meteorol. Soc.*, **78**, 2771–2777, https://doi.org/10.1175/1520-0477(1997)078%3C2771:TDOENO%3E2.0.CO;2.

van den Hurk, B., F. Doblas-Reyes, G. Balsamo, R. D. Koster, S. I. Seneviratne, and H. Camargo, 2012. Soil moisture effects on seasonal temperature and precipitation forecast scores in Europe. *Clim. Dyn.*, **38**, 349–362, https://doi.org/10.1007/s00382-010-0956-2.

Vellinga, M., and S. F. Milton, 2018: Drivers of interannual variability of the East African “Long Rains”. *Q. J. Roy. Meteor Soc.*, **144**, 861–876. https://doi.org/10.1002/qj.3263

Vitart, F, A. W. Robertson, and D. L. Anderson, 2012: Subseasonal to Seasonal Prediction Project: Bridging the gap between weather and climate. *B. World Meteor. Org.*, **61**, 23–28.

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.,* https://doi.org/10.1007/s00382-019-04835-9.

Webster, P. J., and S. Yang, 1992: Monsoon and ENSO: Selectively interactive systems. *Quart. J. Roy. Meteor. Soc.*, **118**, 825–877, https://doi.org/10.1002/qj.49711850705.

Weisheimer, A., and T. N. Palmer, 2014: On the reliability of seasonal climate forecasts. *J. Roy. Soc. Interface*, **11**, 20131162, https://doi.org/10.1098/rsif.2013.1162.

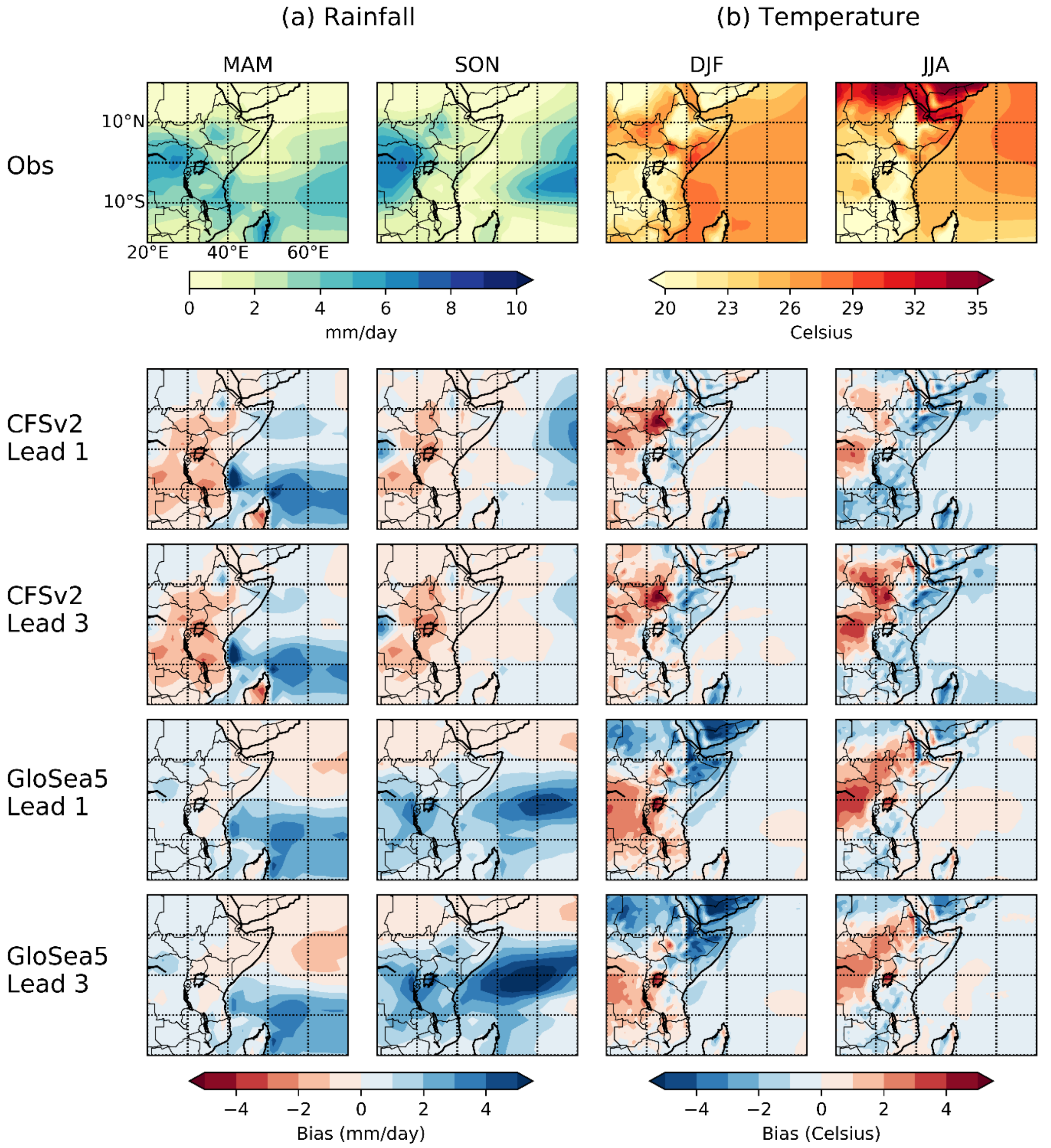

Figure 1. Seasonal mean observations and model biases for (a) rainfall (mm day^{-1}) in MAM and SON (b) 2m air temperature (°C) in DJF and JJA. Observational data are for the period 1982-2015 from (a) GPCP and (b) ERA-Interim; model biases are for their own time periods, lead times are in months ahead of the season.

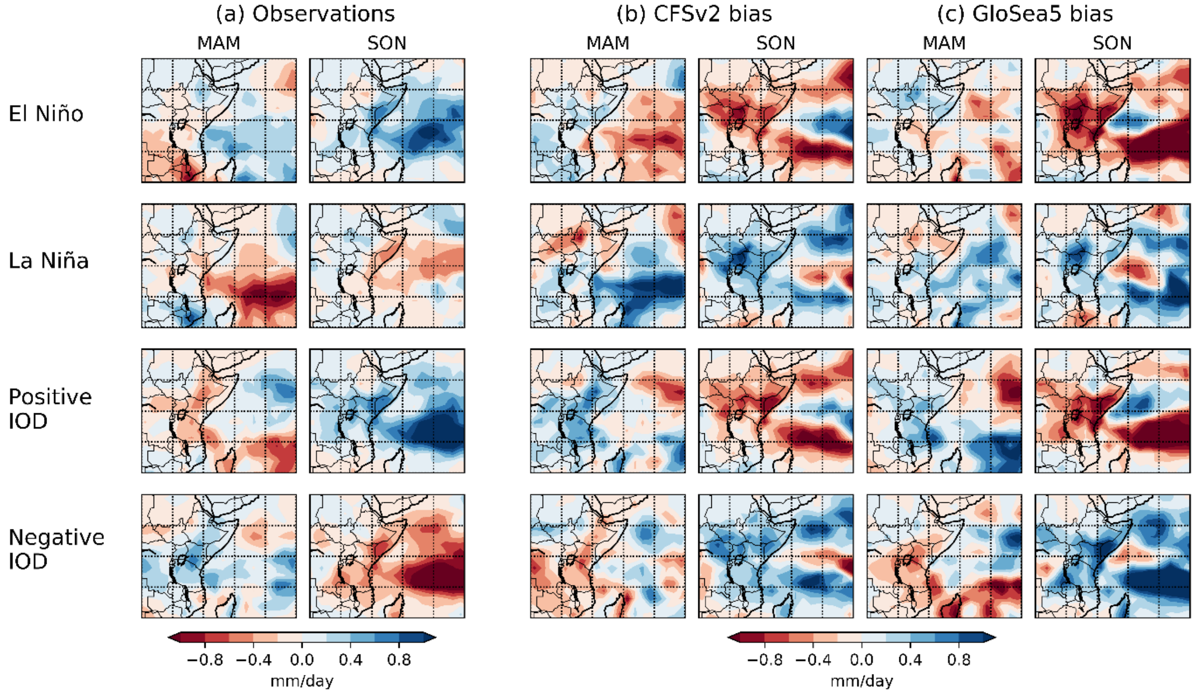

Figure 2. Anomalies in observations during ENSO and IOD phases for rainfall (mm day^{-1}) in MAM and SON, and model biases in those phases relative to the model mean bias at one month lead time. These are composites based on years in which observed SSTs fall into tercile categories (11 or 12 years for observations, 10 years for CFSv2, 8 years for GloSea5).

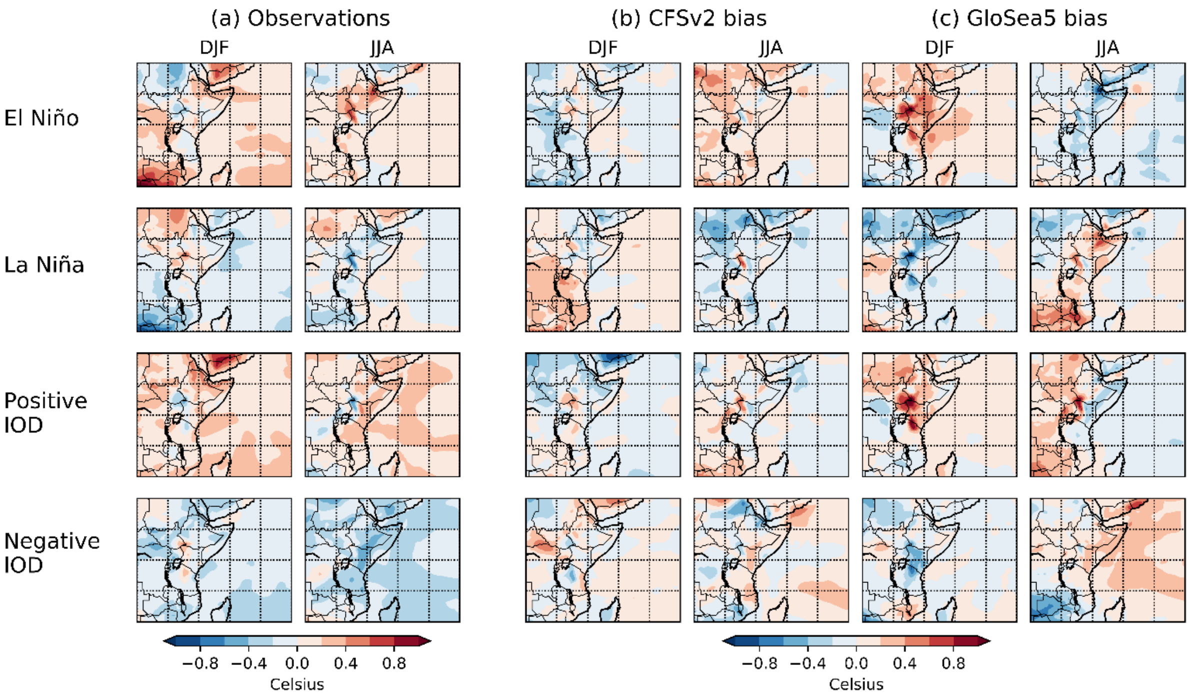

Figure 3. Anomalies in observations during ENSO and IOD phases for 2m air temperature (°C) in DJF and JJA, and model biases in those phases relative to the model mean bias at one month lead time. These are composites based on years in which observed SSTs fall into tercile categories (11 or 12 years for observations, 10 years for CFSv2, 8 years for GloSea5).

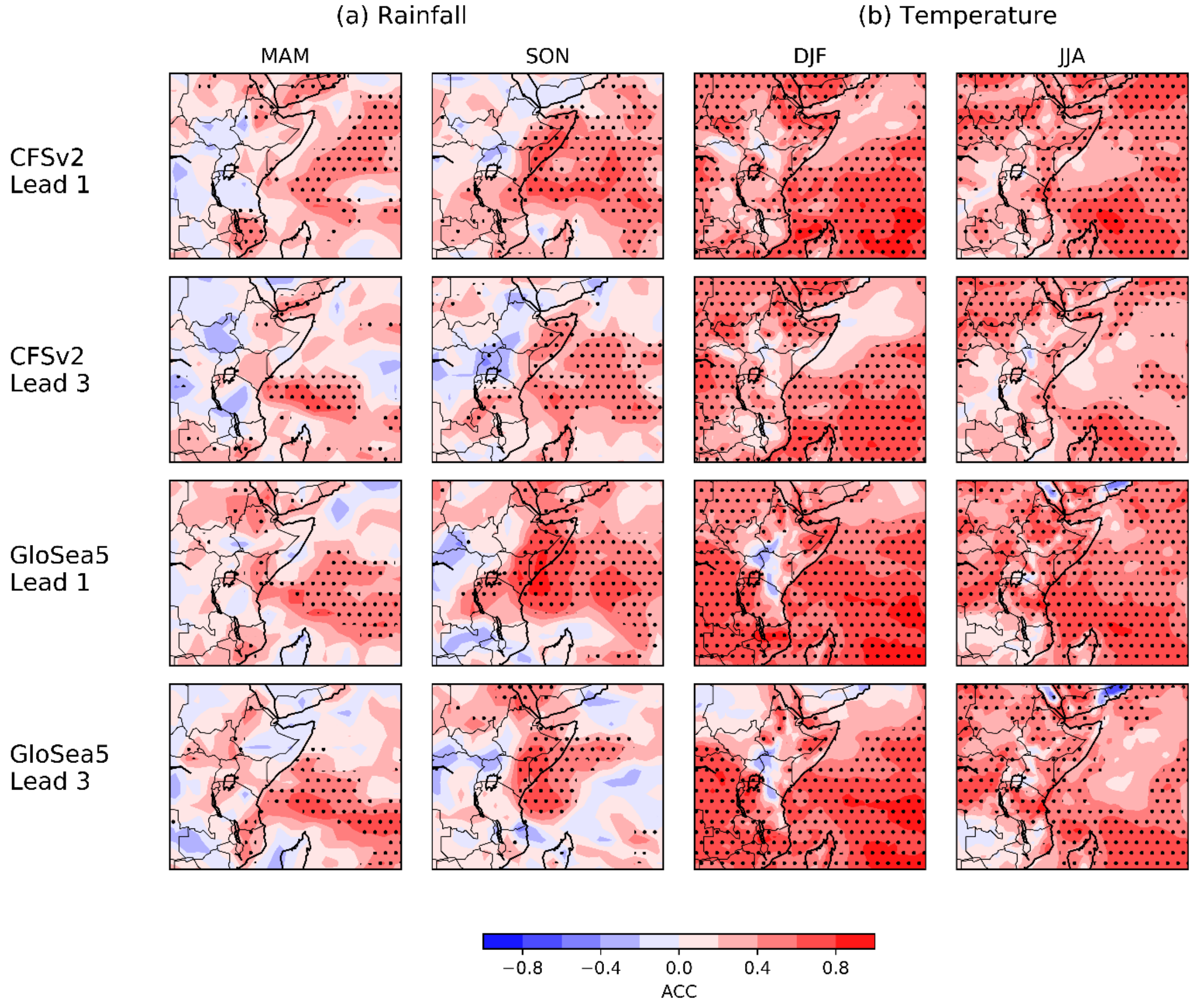

Figure 4. Temporal Anomaly Correlation Coefficients of re-forecasts of (a) rainfall in MAM and SON and (b) 2m air temperature in DJF and JJA. Lead times are in months ahead of the season. Stippling corresponds to areas where the correlation is significant ($p \leq 0.05$).

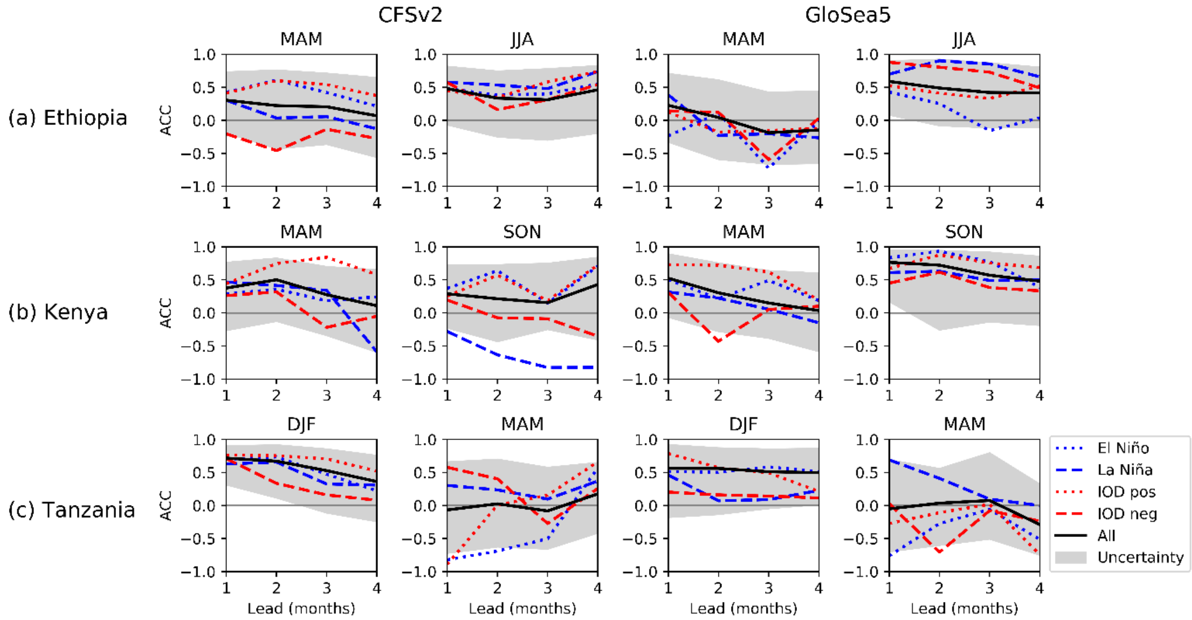

Figure 5. Country-average scale temporal Anomaly Correlation Coefficients (ACCs) of rainfall re-forecasts from (left) CFSv2 and (right) GloSea5 for key rainy seasons in (a) Ethiopia, (b) Kenya and (c) Tanzania. ACCs are shown for all years (black) and years in ENSO and IOD phases (colours); the uncertainty (grey) represents the 5th-95th percentiles from bootstrapping with the number of years in the ENSO and IOD phases.

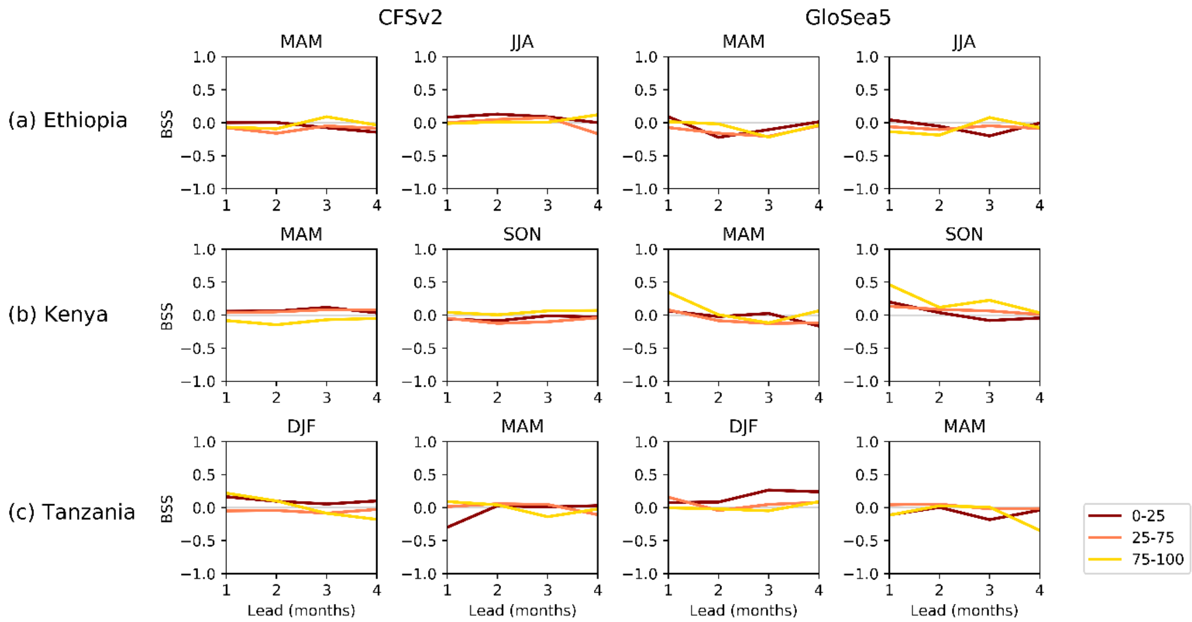

Figure 6. Country-average scale Brier Skill Scores for rainfall re-forecasts using percentiles based on model re-forecast climatologies to define events, from (left) CFSv2 and (right) GloSea5 for key rainy seasons in (a) Ethiopia, (b) Kenya and (c). Colours represent dry (red), normal (orange) and wet (yellow) events.

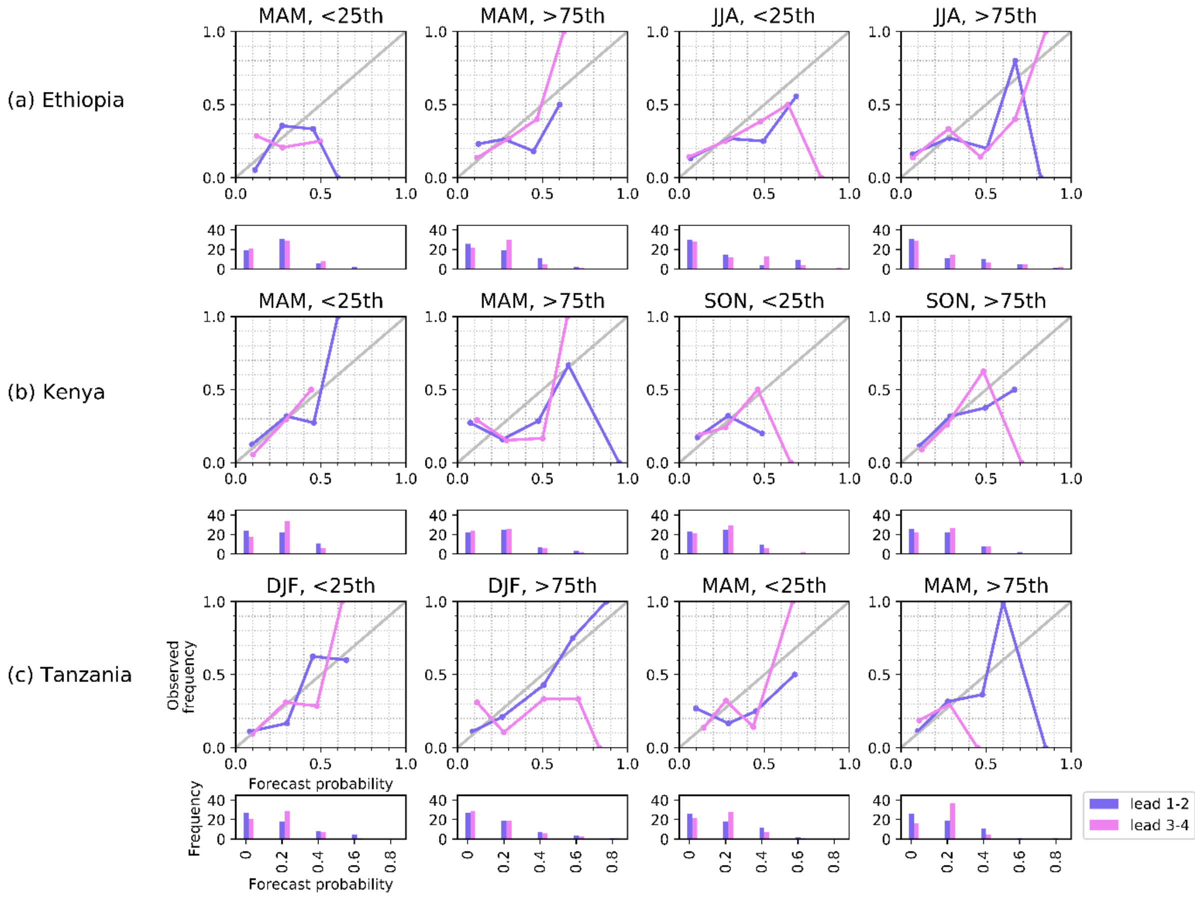

Figure 7. Country-average scale reliability diagrams for rainfall re-forecasts from CFSv2 using percentiles based on model re-forecast climatologies to define events below the 25th and above the 75th percentiles, for key rainy seasons in (a) Ethiopia, (b) Kenya and (c) Tanzania. Colours represent reliability for lead 1-2 (purple) and lead 3-4 (pink) and the 1:1 line (grey).

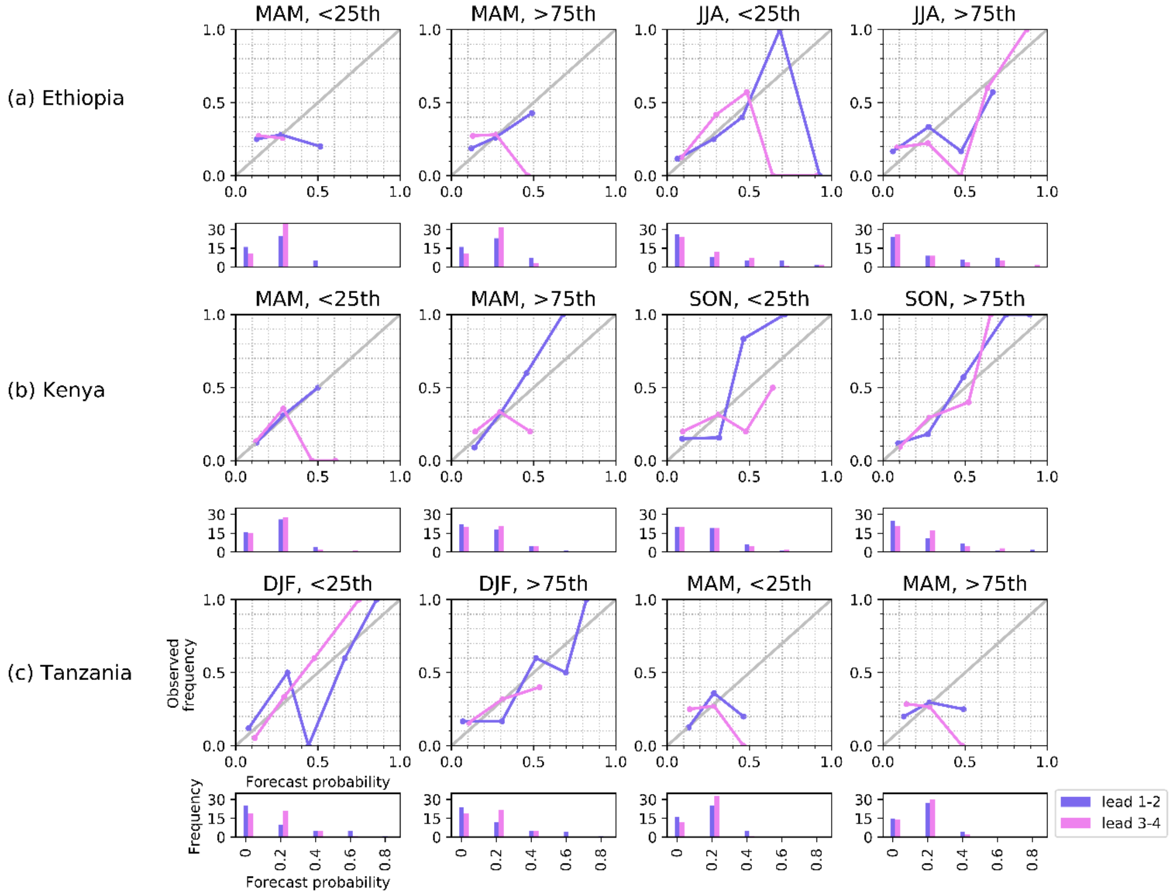

Figure 8. Country-average scale reliability diagrams for rainfall re-forecasts from GloSea5 using percentiles based on model re-forecast climatologies to define events below the 25th and above the 75th percentiles, for key rainy seasons in (a) Ethiopia, (b) Kenya and (c) Tanzania. Colours represent reliability for lead 1-2 (purple) and lead 3-4 (pink) and the 1:1 line (grey).

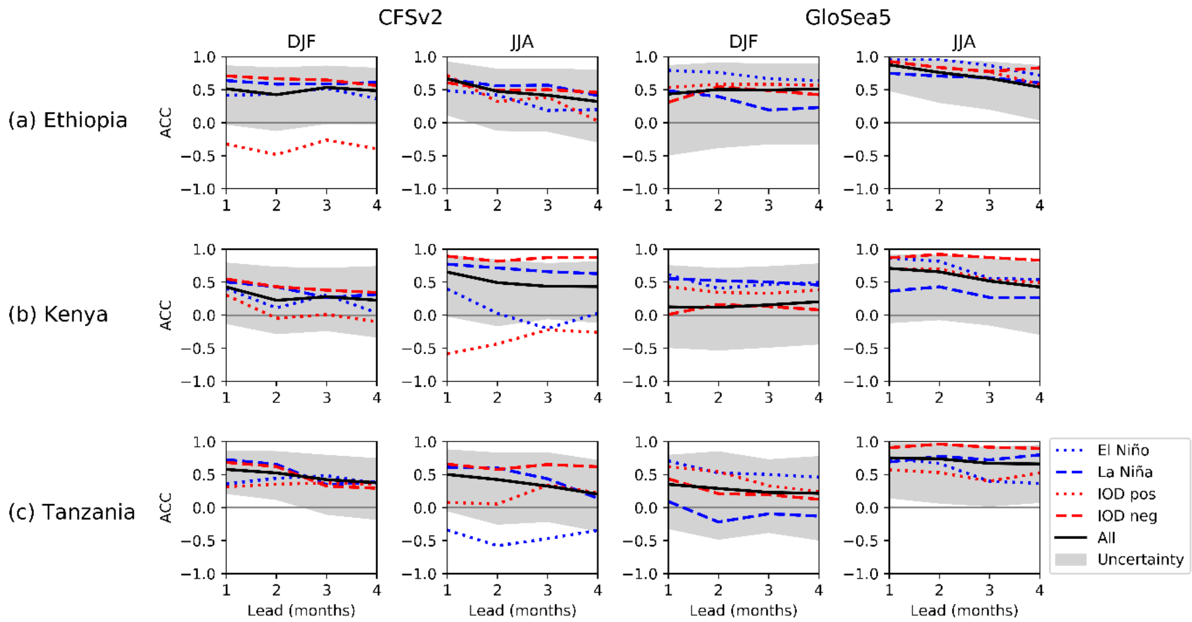

Figure 9. Country-average scale temporal Anomaly Correlation Coefficients (ACCs) of 2m air temperature re-forecasts from (left) CFSv2 and (right) GloSea5 for DJF and JJA in (a) Ethiopia, (b) Kenya and (c) Tanzania. ACCs are shown for all years (black) and years in ENSO and IOD phases (colours); the uncertainty (grey) represents the 5^{th}-95^{th} percentiles from bootstrapping with the number of years in the ENSO and IOD phases.

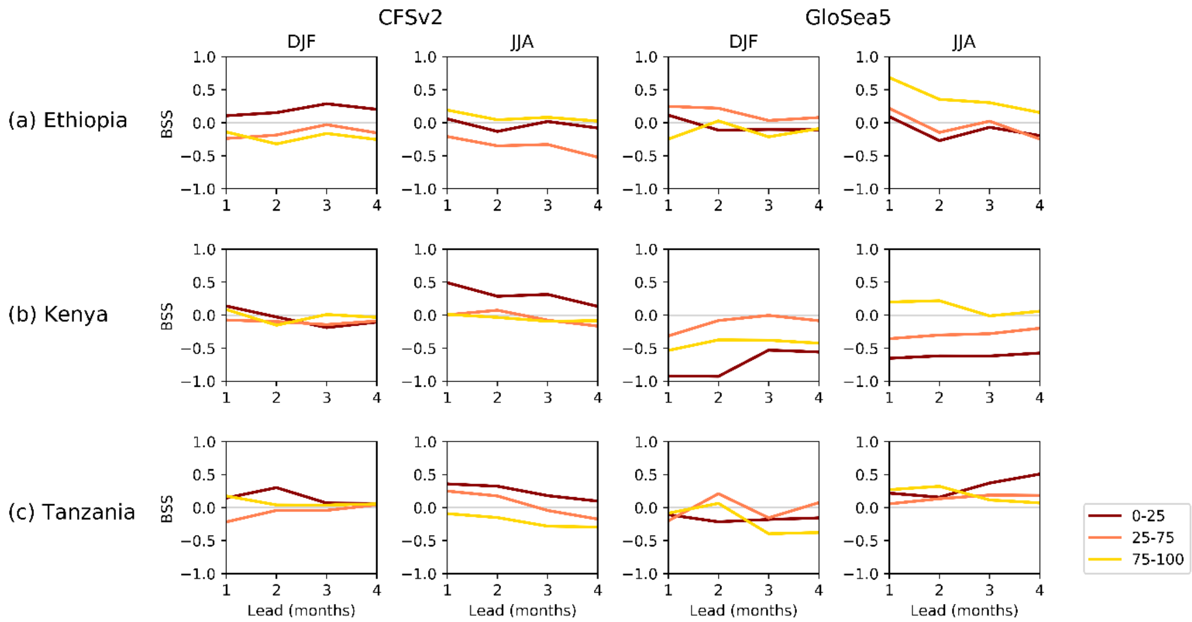

Figure 10. Country-average scale Brier Skill Scores for 2m air temperature re-forecasts using percentiles based on model re-forecast climatologies to define events, from (left) CFSv2 and (right) GloSea5 for DJF and JJA in (a) Ethiopia, (b) Kenya and (c) Tanzania. Colours represent cool (red), normal (orange) and warm (yellow) events.

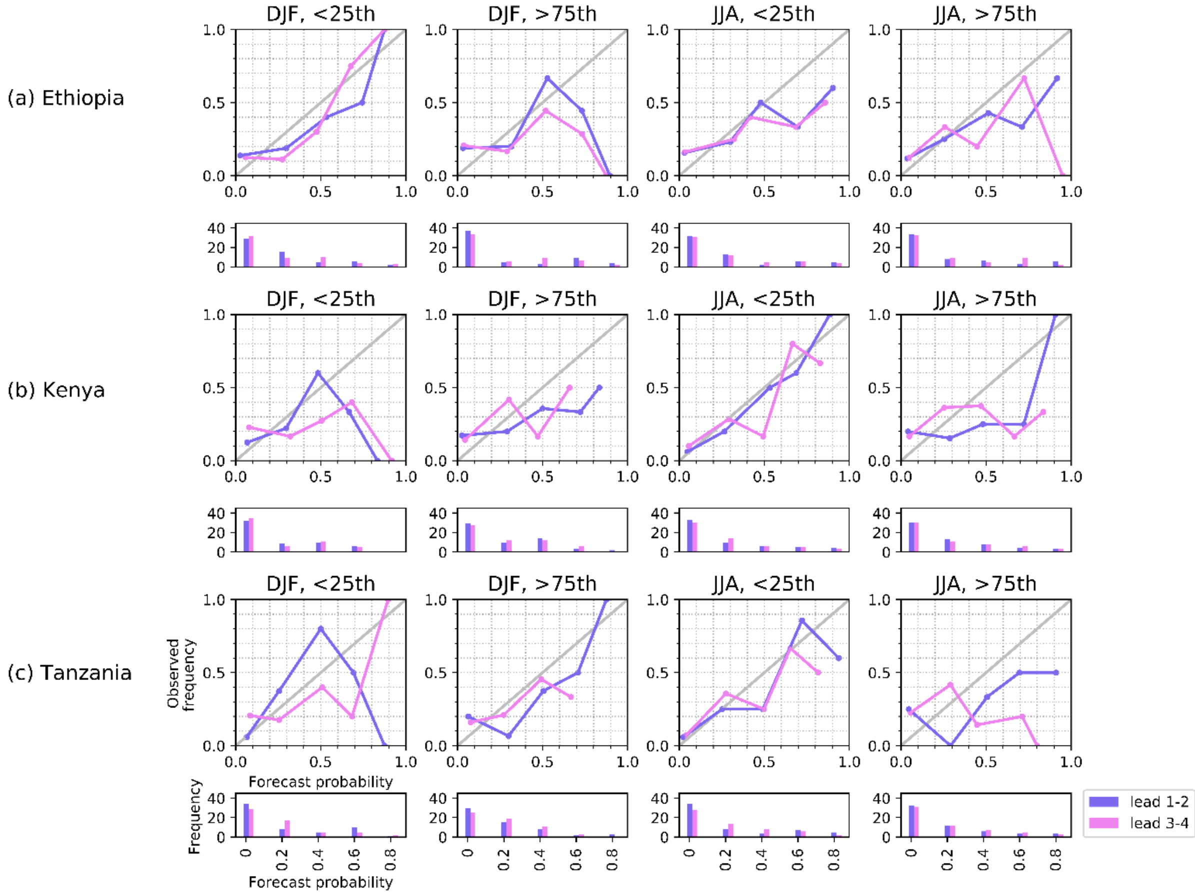

Figure 11. Country-average scale reliability diagrams for 2m air temperature re-forecasts from CFSv2 using percentiles based on model re-forecast climatologies to define events below the 25th and above the 75th percentiles, for DJF and JJA in (a) Ethiopia, (b) Kenya and (c) Tanzania. Colours represent reliability for lead 1-2 (purple) and lead 3-4 (pink) and the 1:1 line (grey).

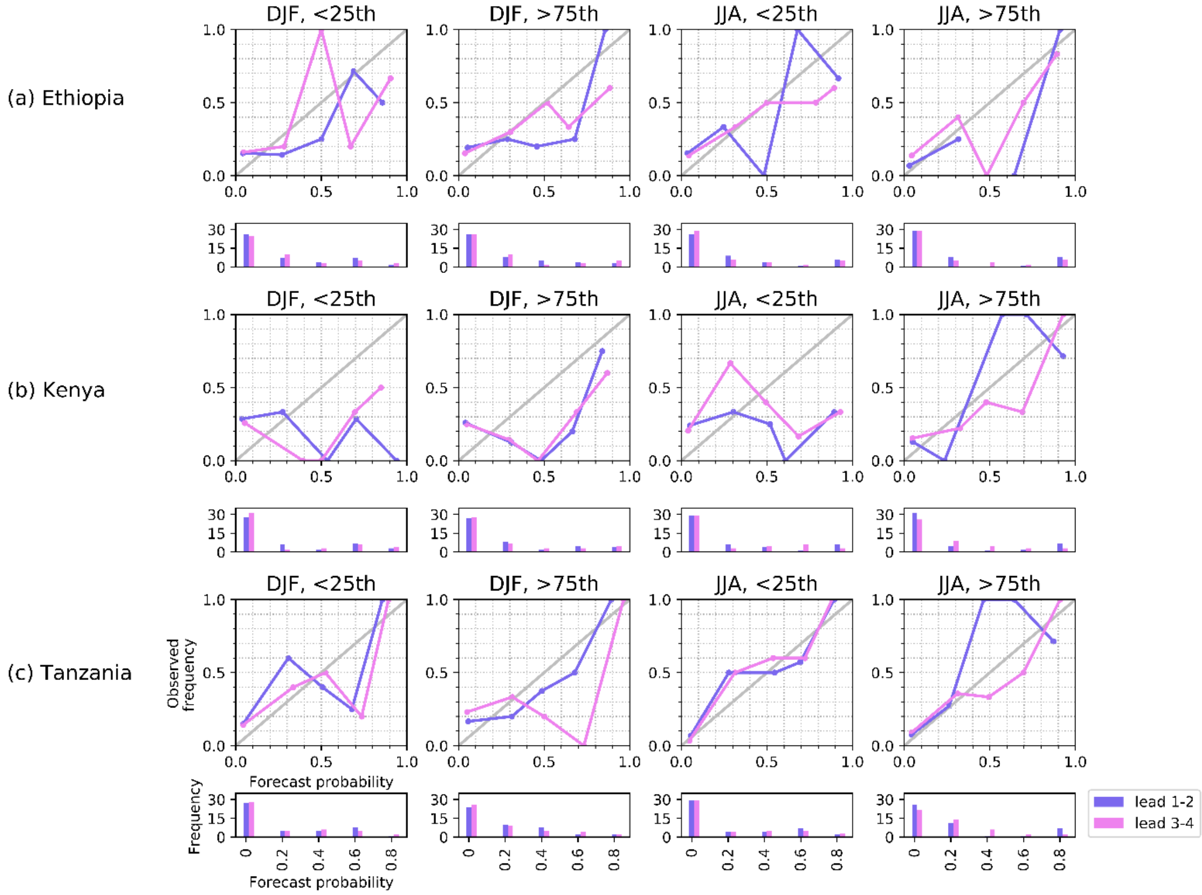

Figure 12. Country-average scale reliability diagrams for 2m air temperature re-forecasts from GloSea5 using percentiles based on model re-forecast climatologies to define events below the 25th and above the 75th percentiles, for DJF and JJA in (a) Ethiopia, (b) Kenya and (c) Tanzania. Colours represent reliability for lead 1-2 (purple) and lead 3-4 (pink) and the 1:1 line (grey).

		Rainfall															
		ACC				BSS 0-25				BSS 25-75				BSS 75-100			
		CFSv2 Lead 1	CFSv2 Lead 2	GloSea5 Lead 1	GloSea5 Lead 2	CFSv2 Lead 1	CFSv2 Lead 2	GloSea5 Lead 1	GloSea5 Lead 2	CFSv2 Lead 1	CFSv2 Lead 2	GloSea5 Lead 1	GloSea5 Lead 2	CFSv2 Lead 1	CFSv2 Lead 2	GloSea5 Lead 1	GloSea5 Lead 2
Ethiopia	MAM	0.30	0.22	0.23	0.04	0.00	0.00	0.09	-0.22	-0.08	-0.16	-0.07	-0.16	-0.07	-0.09	0.02	-0.02
	JJA	**0.48**	0.33	**0.59**	**0.49**	0.08	**0.13**	0.04	-0.05	0.00	0.05	-0.06	-0.10	-0.01	0.01	-0.13	-0.19
Kenya	MAM	**0.38**	**0.50**	**0.53**	0.30	0.06	0.06	0.06	-0.02	0.04	0.05	0.08	-0.08	-0.08	-0.15	**0.35**	0.01
	SON	0.29	0.21	**0.76**	**0.72**	-0.06	-0.09	**0.20**	0.04	-0.05	-0.12	0.13	0.09	0.04	0.01	**0.46**	0.12
Tanzania	DJF	**0.72**	**0.67**	**0.56**	**0.56**	**0.16**	0.10	0.08	0.08	-0.05	-0.04	0.16	-0.04	**0.22**	0.10	0.02	-0.02
	MAM	-0.07	0.03	-0.05	0.04	-0.30	0.03	-0.11	0.00	0.01	0.06	0.04	0.05	0.09	0.04	-0.13	-0.19

Table 1. Summary of ACC and BSS country results for rainfall forecasts at lead 1 and lead 2. Values are shown to 2 decimal places, with bold signifying ACC values that are significant ($p \leq 0.05$) and BSS values where the 5th-95th percentile uncertainty range from bootstrapping lies above 0.0.

		Temperature															
		ACC				BSS 0-25				BSS 25-75				BSS 75-100			
		CFSv2 Lead 1	CFSv2 Lead 2	GloSea5 Lead 1	GloSea5 Lead 2	CFSv2 Lead 1	CFSv2 Lead 2	GloSea5 Lead 1	GloSea5 Lead 2	CFSv2 Lead 1	CFSv2 Lead 2	GloSea5 Lead 1	GloSea5 Lead 2	CFSv2 Lead 1	CFSv2 Lead 2	GloSea5 Lead 1	GloSea5 Lead 2
Ethiopia	DJF	**0.51**	**0.42**	**0.43**	**0.50**	**0.11**	**0.15**	0.11	-0.11	-0.24	-0.19	**0.25**	**0.22**	-0.14	-0.32	-0.25	0.03
	JJA	**0.66**	**0.47**	**0.88**	**0.76**	0.06	-0.13	0.09	-0.27	-0.21	-0.35	**0.22**	-0.15	**0.19**	0.04	**0.68**	**0.35**
Kenya	DJF	**0.43**	0.22	0.12	0.12	**0.14**	-0.03	-0.92	-0.93	-0.07	-0.10	-0.31	-0.08	0.09	-0.15	-0.53	-0.37
	JJA	**0.65**	**0.49**	**0.71**	**0.66**	**0.49**	**0.29**	-0.65	-0.62	0.00	0.07	-0.36	-0.30	0.01	-0.03	**0.20**	**0.22**
Tanzania	DJF	**0.58**	**0.52**	0.35	0.29	**0.15**	**0.30**	-0.12	-0.22	-0.22	-0.04	-0.20	**0.21**	**0.18**	0.04	-0.09	0.06
	JJA	**0.50**	**0.42**	**0.75**	**0.74**	**0.36**	**0.32**	**0.22**	**0.16**	**0.25**	**0.18**	0.06	**0.13**	-0.09	-0.15	**0.27**	**0.32**

Table 2. Summary of ACC and BSS country results for temperature forecasts at lead 1 and lead 2. Values are shown to 2 decimal places, with bold signifying ACC values that are significant ($p \leq 0.05$) and BSS values where the 5th-95th percentile uncertainty range from bootstrapping lies above 0.0.