Probabilistic rainy season onset prediction over the greater horn of africa based on long-range multi-model ensemble forecasts

Michael Scheuerer[1] · Titike K. Bahaga[2] · Zewdu T. Segele[3] · Thordis L. Thorarinsdottir[1]

Abstract
This works proposes a probabilistic framework for rainy season onset forecasts over Greater Horn of Africa derived from bias-corrected, long range, multi-model ensemble precipitation forecasts. A careful analysis of the contribution of the different forecast systems to the overall multi-model skill shows that the improvement over the best performing individual model can largely be explained by the increased ensemble size. An alternative way of increasing ensemble size by blending a single model ensemble with climatology is explored and demonstrated to yield better probabilistic forecasts than the multi-model ensemble. Both reliability and skill of the probabilistic forecasts are better for OND onset than for MAM and JJAS onset where forecasts are found to be late biased and have only minimal skill relative to climatology. The insights gained in this study will help enhance operational subseasonal-to-seasonal forecasting in the GHA region.

Keywords Rainy season onset · Seasonal forecasts · Multi-model ensemble

1 Introduction

Greater Horn of Africa (GHA) is one of the regions most vulnerable to recurrent extreme climate events globally, and the risks are becoming increasingly complex as these hazards are compounded by local and remote political and economic instability (Pörtner et al. 2022). This is mainly attributed to the fact that most of the socioeconomic activities are rain dependent. Intraseasonal to decadal rainfall variability has a large impact on the livelihoods of

✉ Michael Scheuerer
scheuerer@nr.no

Titike K. Bahaga
titike.bahaga@igad.int

Zewdu T. Segele
zewdu.segele@noaa.gov

Thordis L. Thorarinsdottir
thordis@nr.no

1 Norwegian Computing Center, Postboks 114, 0314 Oslo, Norway

2 IGAD Climate Prediction and Applications Centre (ICPAC), Ngong Town Kibiku A Road, P.O. Box 10304-00100, Nairobi, Kenya

3 ERT/NOAA NCEP Climate Prediction Center (CPC), 5830 University Research Court, College Park, MD 20740, USA

millions with roughly 75% of the labor force in the region involved in a smallholder, rain-fed agriculture (Salami et al. 2010). For instance, agriculture and hydroelectric power generation are typically rainfed and thus are heavily reliant on rainfall characteristics such as onset, cessation, intensity, and frequency of rainfall. Therefore, forecasting the GHA seasonal rainfall during their different phases are essential for safeguarding the lives and livelihoods of hundreds of millions of people in the region.

The climate of the GHA region exhibits considerable spatial heterogeneity due to complex topographic variations across the region (e.g., Sepulchre et al. 2006; Hession and Moore 2011). In addition to orographic effects and other factors such as strong land-sea gradients (e.g Anyah et al. 2006; Thiery et al. 2015), the seasonal rainfall climatology is highly influenced by the north–south displacement of the tropical rainbelt, referred to as the Intertropical Convergence Zone (ITCZ) (Nicholson 2018). Figure 1 shows the climatology of seasonal mean rainfall derived from CHIRPSv2 data (see Sect. 2) for the period from 1981 to 2017. Throughout January–February (JF; Fig. 1a) most parts of the GHA are dry, and the northeast monsoon brings dry continental air into the region. Consequently, the rainfall during these months is low (Yang et al. 2015), except over the southern parts of the region. Much of the northern and northwestern parts of the GHA have boreal summer

Fig. 1 Seasonal mean CHRIPSv2 rainfall distribution over GHA averaged from 1981 to 2017. **a** January–February (JF), **b** March–May (MAM), **c** June–September (JJAS) and **d** October–December (OND). Unit is given in $mm(day)^{-1}$
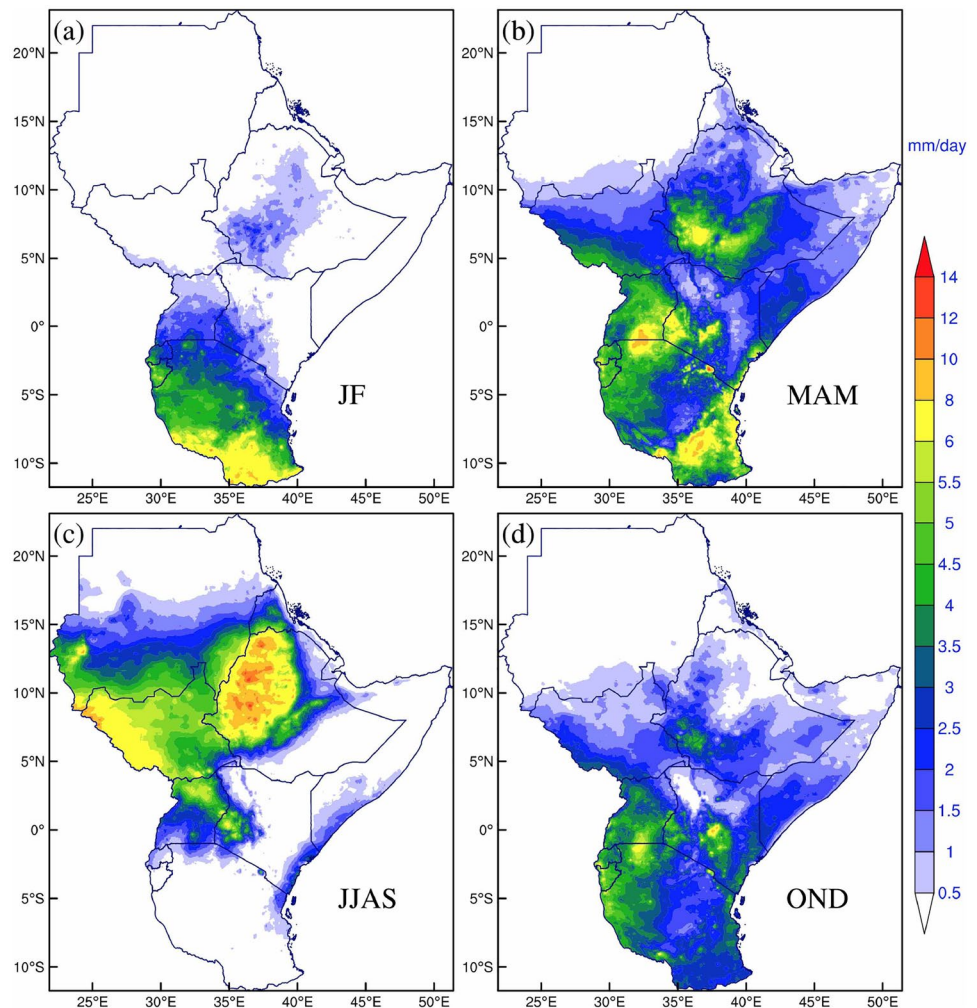
monsoon regimes with maximum rainfall during the period of June–September (JJAS; Fig. 1c), known in Ethiopia as 'Kiremt Rains'. This season accounts for some 50–80% of the rainfall over northern and northwestern agricultural regions (Korecha and Barnston 2007). The equatorial part of GHA has two rainy seasons; the bimodal patterns are experienced over many parts of GHA during the 'Long Rains' of March–May (MAM; Fig. 1b) and the 'Short Rains' of October–December (OND, Fig. 1d). During the Long Rains, the precipitation rate over the coastal areas to the east of the highlands peaks and in general exceeds 1 $mm(day)^{-1}$ except in the northeastern tip of the Horn. The Short Rains show a similar precipitation pattern with a slightly weaker magnitude.

The IGAD Climate Predictions and Applications Centre (ICPAC), as a World Meteorological Organization designated regional climate center, is responsible for providing climate services over the GHA region. Sub-seasonal and seasonal forecasts are produced and disseminated on rolling basis in addition to numerous sector-specific products (https://www.icpac.net/). Additionally,

seasonal climate information for the three main rainfall seasons are co-produced by climate experts from ICPAC, National Meteorological and Hydrological Services, and climate information users, and co-delivered through the Greater Horn of Africa Climate Outlook Forum (GHACOF), a regional user interface platform. The GHACOF platform brings together stakeholders from academic, climate, media, government, private sector and weather forecasters from the 11 countries in the region, to co-produce reliable seasonal forecast information that can facilitate various sectors with their strategic plannings (Walker et al. 2019). GHACOFs are held three times per year, typically in mid-February, the second half of May, and late August, and seasonal forecast products for the MAM, JJAS, and OND seasons are issued, respectively.

In addition to seasonal outlooks for below/above normal rainfall amounts, reliable and actionable forecast information on rainfall's intraseasonal variability can support agricultural decision-making. For instance, knowledge of the beginning of the rainy season (onset) is key for making strategic decisions, such as the time of planting and the

choice of crop to plant (Black et al. 2023). Besides seasonal forecasts, ICPAC provides forecast of intraseasonal rainfall characteristics, which include onset/cessation (dates and anomalies), maximum length of dry and wet spells and start date of wet/dry spells. They are utilised by various sectors including the agriculture and food security sectors. These products, in the past, were generated through dynamical downscaling of ensemble members of the NCEP-CFSv2 model using the Weather and Research Forecasting (WRF) model (Skamarock et al. 2008). Starting with the MAM 2022 GHACOF the prediction of rainy season onset is based on multi-model daily rainfall forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF), Météo-France, UK Met Office (UKMO), Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), and Deutscher Wetterdienst (DWD).

Previous studies have shown the skill of operationally coupled ocean–atmosphere long-range ensemble forecasting systems for predicting seasonal rainfall over East Africa (Bahaga et al. 2016; Diro et al. 2014; Dutra et al. 2013; Gleixner et al. 2017; Young and Klingaman 2020; Ehsan et al. 2021). Forecast systems like ECMWF's SEAS5 system have been demonstrated to have significant skill in predicting the variability of seasonal precipitation amounts over GHA, especially for the 'Short Rains' of the OND season (Mwangi et al. 2014; MacLeod 2019). The scientific basis for such skillful dynamical seasonal forecasting is attributed to the teleconnections between seasonal rainfall over the East African region and remote drivers such as the El Niño-Southern Oscillation (ENSO), the Indian Ocean Dipole (IOD), and the Madden-Julian oscillation (MJO) (Hastenrath et al. 1993; Berhane et al. 2014; Ummenhofer et al. 2009; Bahaga et al. 2015; Nicholson 2017; Bahaga et al. 2019). Possible benefits of multi-model ensembles (MMEs) of seasonal forecasts have been studied, for example, by Batté and Déqué (2011) who report improved probabilistic skill compared to single model ensembles over several regions in Africa, including GHA (their Table 3). Bahaga et al. (2016) also evaluated a multi-model ensemble over equatorial East Africa and found that the coupled climate model ensemble reproduces the spatial distribution of mean September–November rainfall and seasonal climate variations, with further improvement in MME mean.

Existing climate studies have extensively focused on understanding the interannual variability and predictability of seasonal rainfall as contrasted to less attention paid to intraseasonal characteristics, despite the possible use of daily seasonal forecast system output for predicting rain season onset and cessation. While daily precipitation amounts themselves are not predictable at seasonal lead times, an ensemble of model simulations may well represent a tendency towards an earlier or later occurrence of days with increased (for onset) or decreased (for cessation)

precipitation, and thus provide guidance for these intraseasonal characteristics. For SEAS5 based predictions of anomalous rainy season onset and cessation over East Africa, MacLeod (2018) reported statistically significant correlations of predicted and observed anomalies during the OND season over large parts of the domain. Their study used the onset definition introduced by Liebmann and Marengo (2001), which avoids the need for bias-correcting the predicted daily precipitation amounts. Here, we also study the skill of rainy season onset predictions over East Africa, but our approach differs from MacLeod (2018) in that we:

- use the agronomic onset definition employed by ICPAC, which requires bias-correction of the long-range daily precipitation forecasts.
- focus on generating and evaluating *probabilistic* onset forecasts.
- assess the benefits of MMEs for probabilistic rainy season onset prediction.

After presenting the data in Sect. 2, the definition of rainy season onset, methods for bias-correcting the long-range forecast of daily precipitation amounts, the probabilistic forecasting framework, and verification metrics employed in this study are detailed in Sect. 3. Results are presented in Sect. 4, and Sect. 5 concludes with a discussion of the main insights gained in this study.

2 Data

We use the Climate Hazards group InfraRed Precipitation with Station dataset version 2 (CHIRPSv2) as a proxy for local precipitation amounts. This dataset is a quasi-global (50° S–50° N) daily, pentad, and monthly precipitation product from a combination of in-situ station observations and satellite precipitation estimates based on Cold Cloud Duration (CCD) observations to represent sparsely gauged regions. CHIRPSv2 was developed by the Climate Hazards Group at the University of California at Santa Barbara in collaboration with the US Geological Survey Earth Resources Observation and Science center (Funk et al. 2015). The high-resolution dataset was downloaded for the 1981–2017 period at 0.05°× 0.05° horizontal resolution and upscaled to the 0.25°× 0.25° resolution typically used for ICPAC forecast products.

For constructing the rainy season onset forecasts in Sect. 3 we use long-range ensemble precipitation hindcasts from three forecast centers: ECMWF, Météo France, and UKMO. These forecast systems were chosen because a) they use common initialisation dates (1st of each month) and b) they provide hindcasts with lead times up to 201 days. Criterion a) permits a clean comparison of the skill

Table 1 Hindcast data used in this study

Center	Forecast system name	Hindcast years	#Members	Max. lead time
ECMWF	SEAS5	1993–2016	25	215 days
Météo-France	System 8	1993–2016	25	211–212 days
UKMO	GloSea6	1993–2016	7	215 days

of the different systems while criterion b) is needed due to the length of the onset search period, the number of days after the candidate onset date required to decide whether an onset has indeed occurred, and the additional days needed to gather training data for bias-correction (see Sect. 3). All three models are among the ensemble forecast systems operationalized at ICPAC and have shown significant skill in predicting seasonal rainfall over the GHA region (Gebrechorkos et al. 2022). The ensemble hindcasts were retrieved from the Copernicus Climate Data Store at a horizontal resolution of 1°×1° and bilinearly interpolated to the to 0.25°×0.25° resolution of the target grid. For this study we use the February, May, and August initialisations during the years 1993–2016 where hindcast ensembles from the most recent respective forecast systems are available. Table 1 provides a summary of the hindcast data used in this study.

3 Methods

3.1 Identifying onset of the rain season

Various techniques to calculate onset have been proposed, including thresholds on accumulated rainfall, anomalies, and percentage cutoff (Segele and Lamb 2005; Fitzpatrick et al. 2015; Dunning et al. 2016; Gudoshava et al. 2022; Zampieri et al. 2023). In the ICPAC operational seasonal rainfall onset forecasts, the threshold on accumulated rainfall technique is used: any day with less than $1\ mm(day)^{-1}$ rainfall is considered a dry day, days with at least $1\ mm(day)^{-1}$ rainfall are considered wet days. The rainy season onset date is then defined as the first day of a wet spell if the total rainfall accumulation within 3 consecutive days exceeds 20 *mm* and there is no dry spell of 7 or more days within the next 21 days. The main advantage of such a local, agro-meteorological onset definition is that it is intuitive and can provide relevant information for local and national planners. The second part of the criterion above is important to identify false starts and avoid damages to crops due to a dry spell. A potential disadvantage is that the use of fixed thresholds entails that wet/dry biases in the precipitation

forecasts translate into early/late biases of the onset date obtained with that definition. Bias-correction of forecasts (discussed in Sect. 3.2 below), especially when they are re-gridded by simple bilinear interpolation as described in Sect. 2, is therefore crucial for this onset definition to yield consistent results.

Patterns of rainy season onset and cessation within GHA differ across different regions (e.g., Fig. 5 in Dunning et al. 2016), associated with north–south migration of the intertropical convergence zone (Camberlin et al. 2009). The Long Rains in the central part of GHA typically begin between March and May and Short Rains typically begin between September and November. The rainy season in the southern part of GHA starts in November/December, and the Kiremt Rains in the north-western part begin in May/June/July. ICPAC provides a rainfall onset forecast across the region, so the search for onset needs to cover the range of potential onset dates across GHA. In order to avoid missing onset dates over the different locations and to capture late and early onsets during extreme years, they add an additional margin of ± 15 days, and define the search window for rainy season onset as follows:

MAM onset: Based on seasonal forecasts initialized on February 1, search for an onset date between February 15 and June 15

OND onset: Based on seasonal forecasts initialized on August 1, search for an onset date between August 15 and January 15

JJAS onset: Based on seasonal forecasts initialized on May 1, search for an onset date between May 15 and October 15

To mask the spatial domain in which we search for MAM/OND/JJAS onset dates we use a criterion similar to that employed by Yang et al. (2015) to identify areas with one/two rain seasons. Specifically, we calculate the average daily CHIRPSv2 precipitation amounts at each grid point over all days within the respective season and all years between 1993 and 2016. Then, for MAM onset, we only consider grid points where the MAM average daily precipitation amount is larger than both JF and JJAS average daily precipitation amounts. For OND onset we only consider grid points where the OND average daily precipitation amount is larger than the JJAS average daily precipitation amount. And for JJAS onset we only consider grid points where the JJAS average daily precipitation amount is larger than both MAM and OND average daily precipitation amounts. The precipitation rate in OND is not compared against the JF precipitation rate
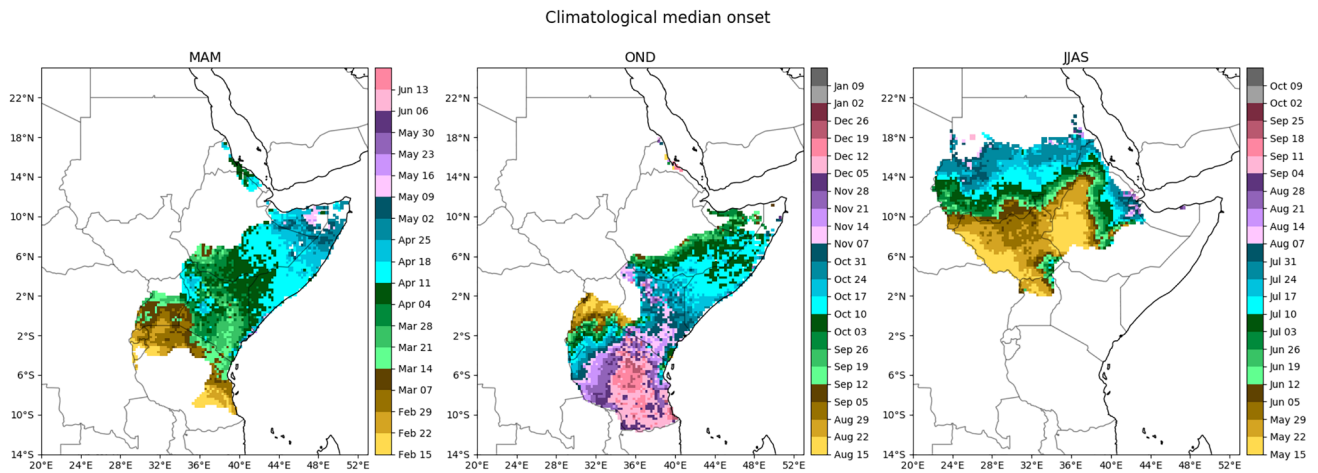
Climatological median onset

Fig. 2 Median of 1993–2016 rainy season onset dates calculated from CHIRPSv2 data

because we do not make a separate JF onset prediction. The long OND search window then also covers areas in Tanzania with a single rainy season and a typical onset in December or early January.

Figure 2 shows the median of the MAM, OND, and JJAS onset dates over the 1993–2016 period with the spatial masks defined as explained above. In general, the spatial distribution of the rainy season onset dates over the GHA region is consistent with the seasonal migration of the ITCZ. Accordingly, during MAM and JJAS, rainy season onset dates move from southwest to northeast, while during OND, the rainy season onset band moves northeast to southwest. Note that during the study period, the above onset criteria are never met for some areas in the northern part of GHA. In large parts of Kenya, Somalia, and southeastern Ethiopia the rains may be too variable to meet all criteria in a given year, i.e., wet spells are frequently interspersed with dry spells, failing the onset criteria. Even in the median, the spatial patterns of local onset dates are somewhat 'noisy' due to the spatial variability in rainfall intensities and the high sensitivity of a threshold based onset definition to unpredictable fluctuations in the time series of daily precipitation amounts. This is one of the reasons why we feel that a probabilistic forecasting paradigm, where forecast products include a quantification of uncertainty, is most appropriate in this setup.

3.2 Bias-correcting long-range ensemble forecasts

Since the onset definition stated above is sensitive to biases in the forecasts of daily precipitation amounts, these biases have to be removed before the onset search is performed with forecast precipitation time series. While the literature on statistical post-processing (which includes bias-correction) of medium-range ensemble forecasts

has seen a trend towards more and more sophisticated, machine learning based methodology (Taillardat et al. 2016; Messner et al. 2017; Rasp and Lerch 2018), a recent study by Hemri et al. (2020) suggests that rather basic bias-correction methods with a limited number of trainable parameters are preferable in a seasonal forecasting context where small training sample sizes and low predictability are common and increase the risk of overfitting. As elaborated by Manzanas et al. (2019), quantile mapping is the most appropriate bias-correction method for variables like daily precipitation amounts which follow highly non-Gaussian distributions and cannot be corrected through shifting and scaling.

In the setup studied here, it suffices to adjust the 1 *mm* and 20 *mm* thresholds used in the onset definition stated above. To see this, we use F_x and F_y to denote the cumulative distribution functions (CDFs) of the climatology of daily precipitation amounts of the forecast model and CHIRPSv2, respectively, at a given location and day of the year. A bias-corrected version $\tilde{x}$ of a new forecast x is then obtained via

$$\tilde{x} = F_y^{-1}(F_x(x)). \tag{1}$$

The condition $\tilde{x} > 1\ mm$ is equivalent to the condition $x > F_x^{-1}(F_y(1\ mm))$. The quantile mapping correction is thus equivalent to mapping the 1 *mm* threshold to a different threshold which depends on the location, day of the year, and forecast model. $F_y(1\ mm)$ is just the climatological probability of no more than 1 *mm* rain per day and can be estimated from CHIRPSv2 data. The inverse CDF F_x^{-1} is estimated from empirical quantiles of the model climatology via linear interpolation. The respective training data sets are composed of the years 1993–2016 except the one for which the forecast is made, i.e. we perform leave-one-year-out cross validation. In

addition to the data from the same day of the year, we use data from a time window of $n_w = 25$ days around this day, e.g., the bias-correction of a forecast valid date of 13 May 2016 uses data from between 1 and 25 May during the years 1993–2015.

The other condition that needs to be checked in order to identify whether rainy season onset has occurred is the exceedance of 20 *mm* within 3 days. To obtain a bias-corrected version of this condition, we aggregate both daily CHRIPSv2 data and predicted daily precipitation amounts to 3-day accumulations and map the 20 *mm* threshold in the same way as described above for the 1 *mm* threshold. Together, this yields all the information required to determine the onset date based on bias-corrected long-range precipitation forecasts.

3.3 Probabilistic onset forecasts

By applying the onset criteria with the adjusted thresholds separately to each ensemble member, an ensemble of onset dates is obtained at each grid point. Locally, the probabilistic information in this ensemble can be presented through the empirical CDF of the onset date, i.e., the probability that rainy season onset occurs no later than a given date is depicted as a curve. Examples of such CDFs for Nairobi and Kigali are shown in Fig. 3. A reference CDF ('climatology') is constructed from an ensemble of the CHIRPSv2 onset dates from all years in the data set except the forecast year as described in more detail in Sect. 3.4. While this reference ensemble should cover the range of possible outcomes and be perfectly reliable, a skillful model-based ensemble allows one to narrow the range of outcomes based on the predicted state of the earth system. In the year 2010 chosen for illustration, the multi-model forecast CDF in Kigali is very similar to the climatological CDF. If we only presented a (deterministic) forecast anomaly, this would communicate that 'normal onset' is expected. By looking at the full CDF, however, it becomes clear that the forecast is as uncertain

as climatology and unable to narrow the range of possible outcomes. For Nairobi, on the contrary, we see a clear shift towards a delayed onset and an increased probability of a 'failed onset' under the criteria employed here, while a 'normal or early onset' in this year is seen as very unlikely.

In order to communicate probabilistic forecast information for many locations simultaneously, one can focus on specific *events* that are relevant for decision making. For rainy season onset, such events could be, for example, 'onset occurs 2 or more weeks earlier than usual' or 'onset fails or is at least 3 weeks delayed'. At each grid point, the forecast then takes the form of a probability for this event to occur, and the set of probabilities can be depicted in a map as shown in for the case studies in Sect. 4.5.

3.4 Verification metrics and statistical testing

In order to base our conclusions on as many verification cases as possible, we perform leave-one-year-out cross-validation, i.e. 1 of the 24 years for which we have data is left out at a time, the statistics needed to perform bias-correction are calculated based on data from the remaining 23 years, and onset forecasts are calculated and verified for the left-out year. This entire process is repeated for each year, and thus all 24 years can be used for verification while the bias-correction is based on independent data. An ensemble representing climatology is constructed in the same way. For each year 1993–2016 we calculate the actual onset date based on the CHIRPSv2 data set, and we build a climatological ensemble of onset dates from all years except the one for which the forecast is evaluated.

3.4.1 Rank histograms

A key requirement for a probabilistic forecast is that it is *reliable*, i.e. it is unbiased and adequately represents forecast uncertainty. The goal of the bias-correction step described
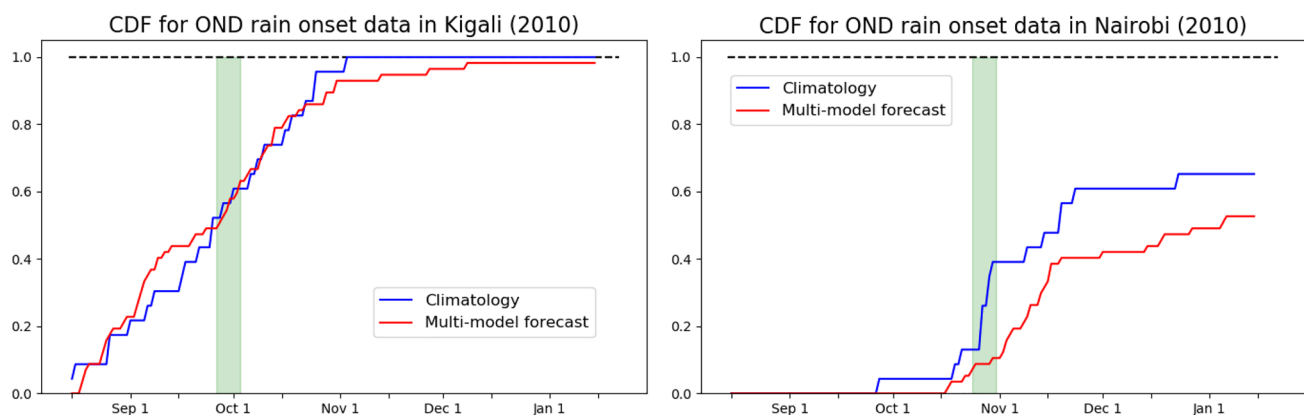

Fig. 3 CDFs visualizing probabilistic forecasts of the OND onset date in 2010. Highlighted by the green shaded area is the climatological median onset week. The probability of a 'failed onset' manifests as a gap between the CDF at the end of the search period and 1.0

in Sect. 3.2 is to remove systematic wet or dry biases from the precipitation forecasts, but this does not guarantee that the resulting ensemble of onset dates is free from early or late biases and that the range of this ensemble adequately represents forecast uncertainty. To assess these features, we use rank histograms (Anderson 1996; Hamill and Colucci 1997; Talagrand et al. 1997), treating predictions of a failed onset as an onset on the first day *after* the search period and resolving tied ranks at random.

3.4.2 The Brier score and its 'fair' variant

When onset dates are communicated through maps, they are often coarsened to a weekly temporal resolution.[1] At this resolution, some of the daily variability is filtered out, while significant shifts towards an early or late onset are still clearly discernible. It then becomes more natural to represent the onset forecasts through a vector of probabilities for each of k different onset categories: the first category is 'failed onset', and the remaining $k-1$ categories correspond to the week after the start of the search period (e.g., for the MAM onset, the second category represents an onset date between 15 and 21 February). Uncertainty about the timing of the onset is thereby encoded in the probability vector $\mathbf{p} = (p_1, \ldots, p_k)'$, while the outcome can be represented by a vector $\mathbf{o} = (o_1, \ldots, o_k)'$ where $o_i = 1$ for the category i in which the onset occurred and $o_j = 0$ for all $j \neq i$.

With this representation of the ensemble onset forecast we perform probabilistic forecast verification which simultaneously evaluates the forecasts' ability to predict a 'failed onset' and their ability to predict shifts of the expected onset date relative to climatology. A common choice for evaluating the performance of such categorical probability forecasts is the (multi-category) Brier score (BS, Brier 1950), which for a single forecast instance is defined as the sum of squared errors between the predicted probabilities and the binary outcome:

$$\text{bs}(\mathbf{p}, \mathbf{o}) = \sum_{i=1}^{k} (p_i - o_i)^2. \tag{2}$$

A skill score (BSS) is obtained by calculating the average score $\overline{\text{bs}_f}$ over all forecast instances in the test set and relating it to the average score $\overline{\text{bs}_{cl}}$ of the reference forecast (here: climatology) via

$$\text{BSS} = \frac{\overline{\text{bs}_{cl}} - \overline{\text{bs}_f}}{\overline{\text{bs}_{cl}}}. \tag{3}$$

A value larger than 0 then implies that the ensemble based forecast provides more information than the climatological

reference forecast; a perfect forecast with 100% accuracy and no uncertainty would score a BSS of 1.

The BS is a 'proper' score (Gneiting and Raftery 2007), i.e. it encourages the forecaster to quote their true belief regarding $\mathbf{p}$. However, Ferro et al. (2008) showed that $\overline{\text{bs}_f}$ decreases as the size m of an otherwise unchanged ensemble forecast system increases. They therefore suggested a 'fair' BS (fBS) which accounts for that effect and yields a score that is (in expectation) independent of ensemble size:

$$\text{fbs}(\mathbf{p}, \mathbf{o}) = \text{bs}(\mathbf{p}, \mathbf{o}) - \sum_{i=1}^{k} \frac{p_i(1 - p_i)}{m - 1}. \tag{4}$$

While the original, proper BS may be more appropriate from the perspective of a forecast user interested in receiving the best possible probability forecast $\mathbf{p}$ regardless of how this forecast was obtained, it is instructive to compare the BSS and its 'fair' variant, the fBSS, to investigate to what extent possible score differences between a single and a multi-model ensemble can be explained by the mere increase in ensemble size.

3.4.3 Testing for significant skill

Seasonal forecast skill is often very limited due to the much lower predictability of seasonal forecast quantities compared to the medium range, and so we expect the BSS and fBSS values to be close to zero. In order to rule out that positive skill reported for a particular region is merely a result of sampling variability, we test the null hypothesis $\overline{\text{bs}_f} = \overline{\text{bs}_{cl}}$ at each grid point using paired t-tests on the scores obtained for the 24 cross-validated verification years. Interpreting the outcome of these tests simultaneously, e.g., by studying spatial patterns of grid points at which significant skill is reported, requires accounting for test multiplicity. Following recommendations by Wilks (2016), we use the framework suggested by Benjamini and Hochberg (1995) and control the false discovery rate (FDR) at the control level $\alpha = 0.2$.

4 Results

4.1 Accuracy of predicted onset dates

Before presenting results based on the probabilistic metrics detailed in Sect. 3.4, we assess whether there are systematic differences between the median onset dates obtained with the CHIRPSv2 data (see Fig. 2) and those calculated from the (single and multi-model) ensemble forecasts. Due to the limited sample size of only 24 verification years one has to be cautious not to over-interpret the local bias patterns shown in Fig. 4. However, some general trends can be observed. All three forecast systems tend to have a late bias

---

[1] See e.g., https://www.icpac.net/seasonal-forecast/?resource_type=9#results.

for the MAM onset in northern Tanzania and Uganda, and an early bias in western Kenya and northern Somalia. For JJAS onset, the ECMWF and Météo France forecasts are late-biased over Sudan and South-Sudan. Finally, for OND onset, all three systems have a tendency towards a late bias over Uganda, Burundi, and Rwanda and the UKMO ensemble additionally has an early bias over Tanzania.

4.2 Reliability

The reliability of the probabilistic onset forecasts is empirically assessed via rank histograms, shown in Fig. 5. The rank histograms are calculated from data across all grid points where an onset forecast is produced. While this impedes a differentiation between regions, it leads to more robust conclusions and allows us to note some general trends. Note that the number of possible ranks differs across the different forecast systems due to the different number of ensemble members. For example, for the 7-member UKMO ensemble there are 8 possible ranks of the verifying observation. The frequency of cases required for the ranks to have a perfectly uniform distribution differs both across seasons and forecast systems. The former is due to a different number of grid points, and hence cases, within each season mask, the latter is a result of dividing the overall number of verification cases by a different number of possible observation ranks.

The histograms confirm the tendency seen in Fig. 4 of the ECMWF and Météo France ensembles towards a late bias for JJAS onset. For the MAM season we also see an overall tendency towards a late bias for all three forecast systems, while for OND there is no apparent bias. The histograms also suggest that ensemble spread is appropriate for all forecast systems and all seasons, i.e. the actual onset date falls within the range of predicted onset dates within the ensemble about as often as one would expect, even for the UKMO ensemble with only seven members.

4.3 Brier skill scores

Figure 6 shows BSSs for the ECMWF, Météo France, and UKMO ensembles individually and their combination to a multi-model ensemble. For OND, the ECMWF model yields positive skill scores at most grid points, especially in northern Kenya, southern Ethiopia, and Somalia, which is in line with the results reported by MacLeod (2018). Statistical significance of these scores is however limited to just a few scattered grid points. For MAM and JJAS results are somewhat mixed with skill scores around zero and perhaps a slight tendency towards positive (though not statistically significant) MAM skill over most of Kenya, south-western Uganda, and northern Tanzania. The skill scores of the Météo France model based predictions are

also around—overall perhaps slightly below—zero for MAM and JJAS, and slightly above zero but lower than ECMWF scores for OND. For the UKMO ensemble we obtain mostly negative skill for all seasons and most regions except Somalia. In those regions, using the climatological ensemble of onset dates would yield better probabilistic forecasts than the bias-corrected UKMO ensemble. We will demonstrate below that this can largely be explained by the smaller ensemble size (see Table 1).

Despite the disappointing results for the Météo France and UKMO ensemble, their combination with the ECMWF ensemble yields onset predictions that are more skillful than those of the ECMWF ensemble alone. BSSs of this multi-model ensemble are positive at the vast majority of grid points in all three seasons, and while skill is not statistically significant in MAM and JJAS except for a few scattered grid points, the area of significant, positive skill for OND is a lot larger than for the ECMWF ensemble.

At first glance it seems paradoxical that the Météo France and UKMO ensemble are able to add skill to the ECMWF ensemble even though the quality of their single-model predictions is at many grid points lower than that of a climatological ensemble. A comparison of Figs. 6 and 7 shows that ensemble size may explain this paradox. While results for the ECMWF and Météo France ensemble are largely unchanged, the fBSSs for the UKMO ensemble are much better than the BSSs reported in Fig. 6 due to the elimination of ensemble size as a factor that influences the scores. The multi-model ensemble, on the contrary, looses its advantage over the ECMWF ensemble under a 'fair' score. Since the rank histograms in Fig. 5 suggest that the ensemble spread of the single model onset date predictions is adequate, the better BSSs of the multi-model ensemble compared to the ECMWF ensemble cannot be explained by an inflation of ensemble spread, suggesting that it is due to a better representation of the onset date distribution by a larger ensemble. Indeed, the onset definition used here (Sect. 3.1) translates a time series of daily precipitation values into an onset date, and due to the non-linearity of that mapping, small changes in the time series (e.g., 7 instead of 6 dry days after an initial wet spell) can cause substantial shifts in that date. In a forecast system with just a few ensemble members, such spurious shifts are more likely to entail an onset date ensemble that misrepresents the distribution of plausible outcomes. A large ensemble, on the contrary, contains sufficiently many precipitation time series that the occasional 'jump' in the onset date has only limited impact on the overall forecast distribution. More evidence for the important role of ensemble size in the present setting, where the forecast quantity is a highly non-linear function of GCM output, is given in Appendix 1. This appendix shows additional results obtained with a 21-member UKMO lagged
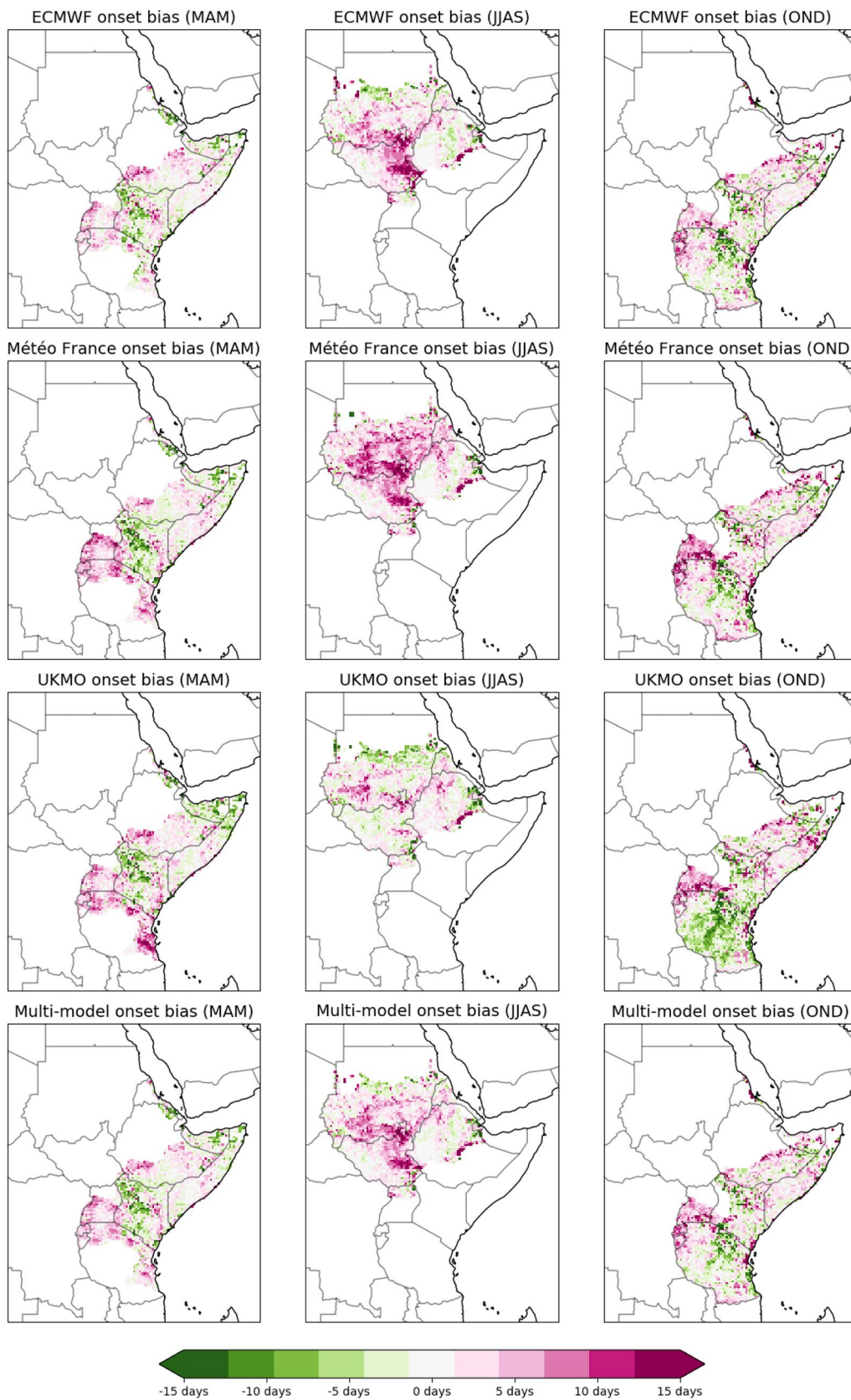
Fig. 4 Median onset date predicted by the different forecast systems minus median onset date calculated from the CHIRPSv2 data, both calculated over the period 1993–2016

ensemble, which is found to be substantially better than the 7-member UKMO ensemble used for the predictions discussed above.

4.4 Climatology-augmented ensembles

The results above highlight the role of ensemble size when it comes to representing the onset probability distribution and show that even the inclusion of sub-ensembles with near or below zero skill can be beneficial. This motivates an additional experiment where we use climatology itself, i.e. ensemble members constructed from historical, CHIRPSv2-based onset dates, to augment a GCM-based onset date ensemble. We repeat the above evaluation with a 48-member ensemble consisting of 25 ECMWF members and 23 (cross-validated) climatology members, and check whether this improves the representation of the forecast distribution compared to the ECMWF ensemble.

The rank histograms in Fig. 8 confirm that the cross-validated, climatological ensemble used as a reference forecast in our BSS calculations is perfectly reliable. The climatology-augmented ECMWF ensemble is somewhat over-dispersed, but the late biases for the MAM and JJAS

onset are reduced compared to both pure ECMWF and multi-model ensemble. Figure 9 shows BSSs for the climatology-augmented ECMWF ensemble, and a comparison with Fig. 6 suggests that its skill is somewhat better than that of the 57-member multi-model ensemble, despite its smaller size. Even though it consists of $\approx 50\%$ climatological ensemble members it outperforms the climatological reference ensemble more consistently than the multi-model ensemble, and OND skill over large parts of GHA is flagged as significant. These results underscore the importance of ensemble size on the one hand, but they also make it clear that there can be better alternatives than including forecast systems with low skill in a multi-model ensemble.

4.5 Case study

In order to illustrate how a probabilistic forecast map could look in practice and to check how well the methodology discussed in this paper works in extreme seasons outside the 1993–2016 hindcast period, we recalculate the bias-adjusted thresholds (see Sect. 3.2) for the ECMWF ensemble, now with all years from the hindcast period, and apply them to ECMWF ensemble forecasts for the 2019 and 2021 OND
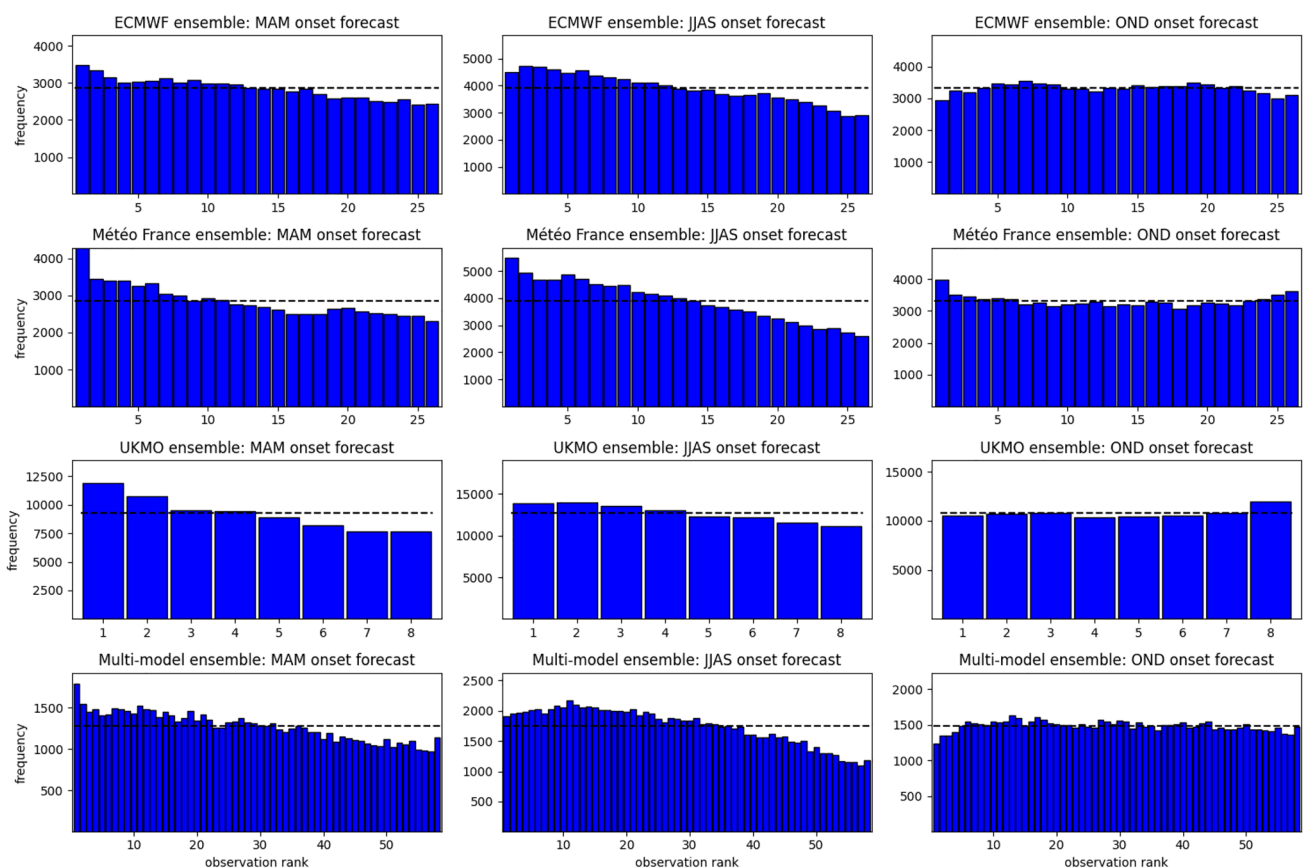

Fig. 5 Rank histograms for the single-model and multi-model ensemble onset date forecasts, computed with cases pooled across all grid points within the respective season mask and all verification years. The dotted lines indicate the frequency required for the ranks to have a perfectly uniform distribution
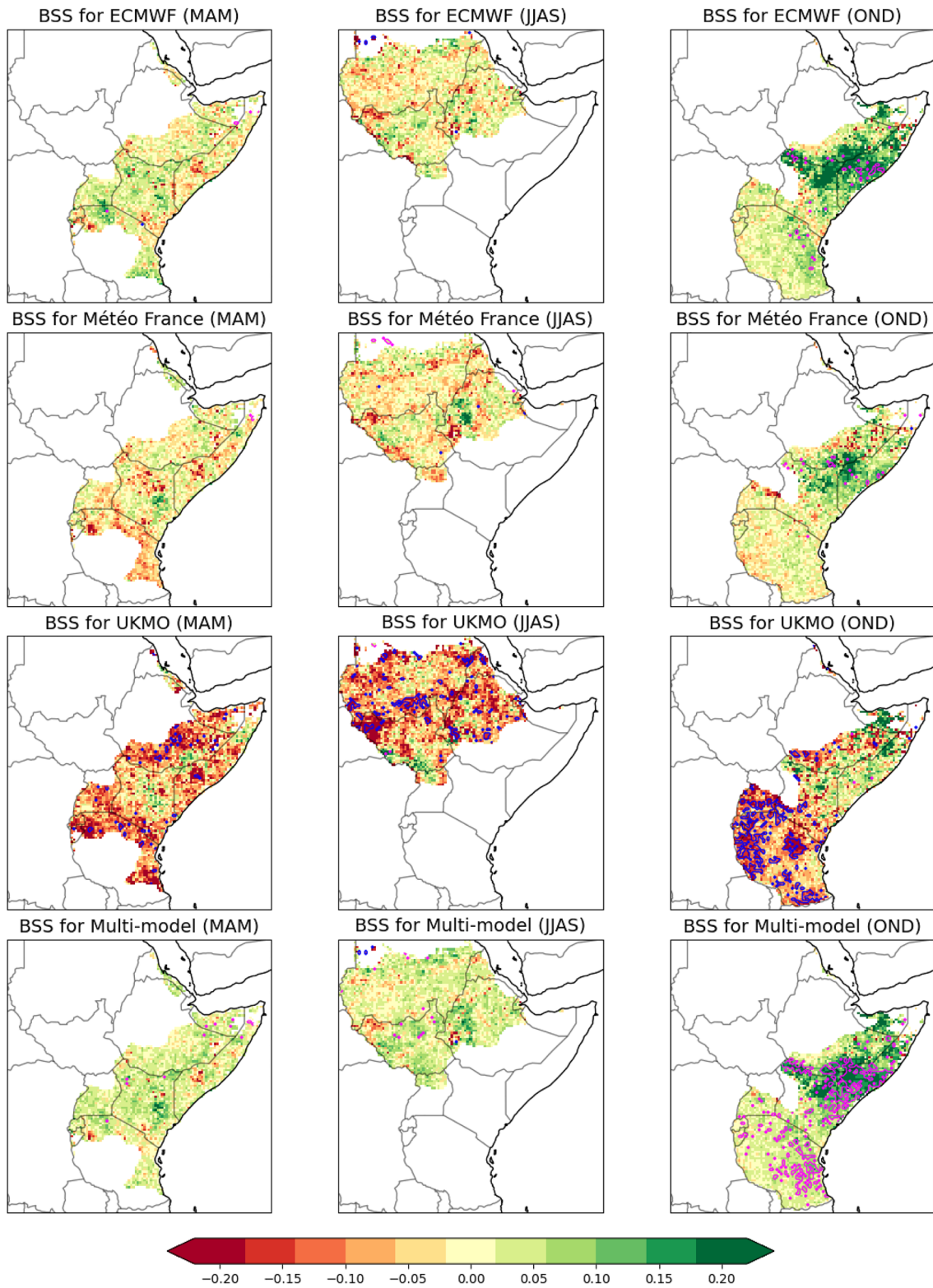
Fig. 6 BSS for different seasons and forecast systems. Grid points where the difference between $\overline{bs_f}$ and $\overline{bs_{cl}}$ is statistically significant when controlling the FDR at $\alpha = 0.2$ are marked with blue contours for negative skill and magenta contours for positive skill

season. The 2019 OND season was characterized by ENSO-positive conditions and an extreme positive Indian Ocean Dipole (IOD), and brought severe flooding to many parts of the GHA region. The 2021 OND season, on the contrary, saw a negative IOD concurrent with a La Niña event which resulted in one of the worst dry periods over almost the entire region. These drivers and anomalies in seasonal rainfall amounts tend to go along with an early (for 2019) and a late or failed (for 2021) rainy season onset (Roy et al. 2023, Table 1). For 2019 and 2021, 51 ECMWF ensemble members are available, which were further augmented with the 1993–2016 climatology (see Sect. 4.4) to a 75-member ensemble from which the probabilities depicted in Fig. 10 were derived.

The upper leftmost panel shows climatological probabilities for the OND onset to occur two or more weeks earlier than usual (i.e., than the 1993–2016 median onset week), the panel below shows climatological probabilities for a failed or at least 3 weeks delayed OND onset. Note that over large parts of Kenya, Somalia, and southern Ethiopia the latter probabilities are quite high due to a high probability of a failed onset (see also Fig. 3 for the specific example of Nairobi). The center-left panels show the probabilities derived from the climatology-augmented ECMWF ensemble forecasts issued in August 2019 and 2021, respectively. In 2019, the forecast probabilities for an *early* onset are higher than the climatological probabilities over most of GHA, while in 2021 the forecast probabilities for a *late or failed* onset were higher than the reference. This can be seen even better in the center-right panels, which depict the difference between the respective probabilities in the left panels. We use an inverted color scale for the upper panel so that green colors always represent an increased probability of an earlier than usual onset. Darker colors in this difference plot indicate that the forecast is rather confident that a departure from climatology (green for 'early', pink for 'late') will occur. The rightmost panels depict the outcome that materialized in those two years in the form of weeks before/after the climatological median onset week. Grid points where the onset failed in the respective years are colored red.

In both years, the forecast captures the general tendency and spatial pattern pretty well. The substantially increased likelihood of an early onset in 2019, especially over Tanzania, verified well, and so did the increased likelihood of a failed or late onset in 2021. There are of course local discrepancies, which is in line with the often limited skill we saw in (Sects. 4.3 and 4.4) at the level of individual grid points. The maps in Fig. 10 suggest, however, that the probabilistic OND forecasts shown here can provide useful and actionable information at least at a regional level.

5 Summary and discussion

This paper proposes to utilize a multi-model forecast ensemble for a probabilistic long-range forecast of rainy season onset over the Greater Horn of Africa. While daily precipitation amounts at individual locations are not predictable at seasonal lead times, an ensemble of model simulations may skillfully present a tendency towards an earlier or later occurrence of days with increased precipitation. Our analysis shows that the long-range forecasts may provide up to 10% improvement over a climatological onset forecast as measured by the Brier skill score, with a larger ensemble yielding better skill. The forecasts show greatest skill in the OND season over northern Kenya, southern Ethiopia, and Somalia.

More specifically, the plots in Sect. 4.2 suggest that the simple bias-correction of the (spatially interpolated) ensemble precipitation forecasts via quantile mapping yields reliable, probabilistic OND onset forecasts. For MAM and JJAS, some biases seem to be present in the ensemble forecasts that cannot be fully removed by an amplitude correction of predicted daily precipitation amounts. The results in Sect. 4.3 suggest that the forecast skill in MAM and JJAS relative to climatology is limited, mostly less than 5% in terms of the Brier skill score. That the results are better for OND onset is in line with the conclusions of MacLeod (2018). While a direct comparison with their results is difficult due to the probabilistic forecast perspective and the different metrics considered here, it still appears that the levels of skill reported in their paper are higher than those reported in our Fig. 6. Additional experiments (results not shown) suggest that at least part of this discrepancy can be explained by the higher horizontal resolution considered here, and the associated 'noise' in the local onset dates.

Our study also explored the utility of a multi-model ensemble relative to a single forecast system. While the results in Fig. 6 confirm the general benefit of such an approach, a closer look revealed that in our setting this benefit is primarily due to the increased number of ensemble members. An alternative strategy in which the ECMWF ensemble was combined with an ensemble representing climatology outperformed the multi-model ensemble. This underscores the importance of a sufficiently large ensemble to represent uncertainty, but it also shows that in a situation where one or more forecast systems are less skillful than climatology, climatology itself can be a more suitable way of augmenting an ensemble. The best configuration for a multi-model ensemble depends on the particular forecast systems considered and their skill over the region of interest. The results in Appendix 1 show that with a different configuration of the UKMO ensemble, conclusions about the relative performance of different
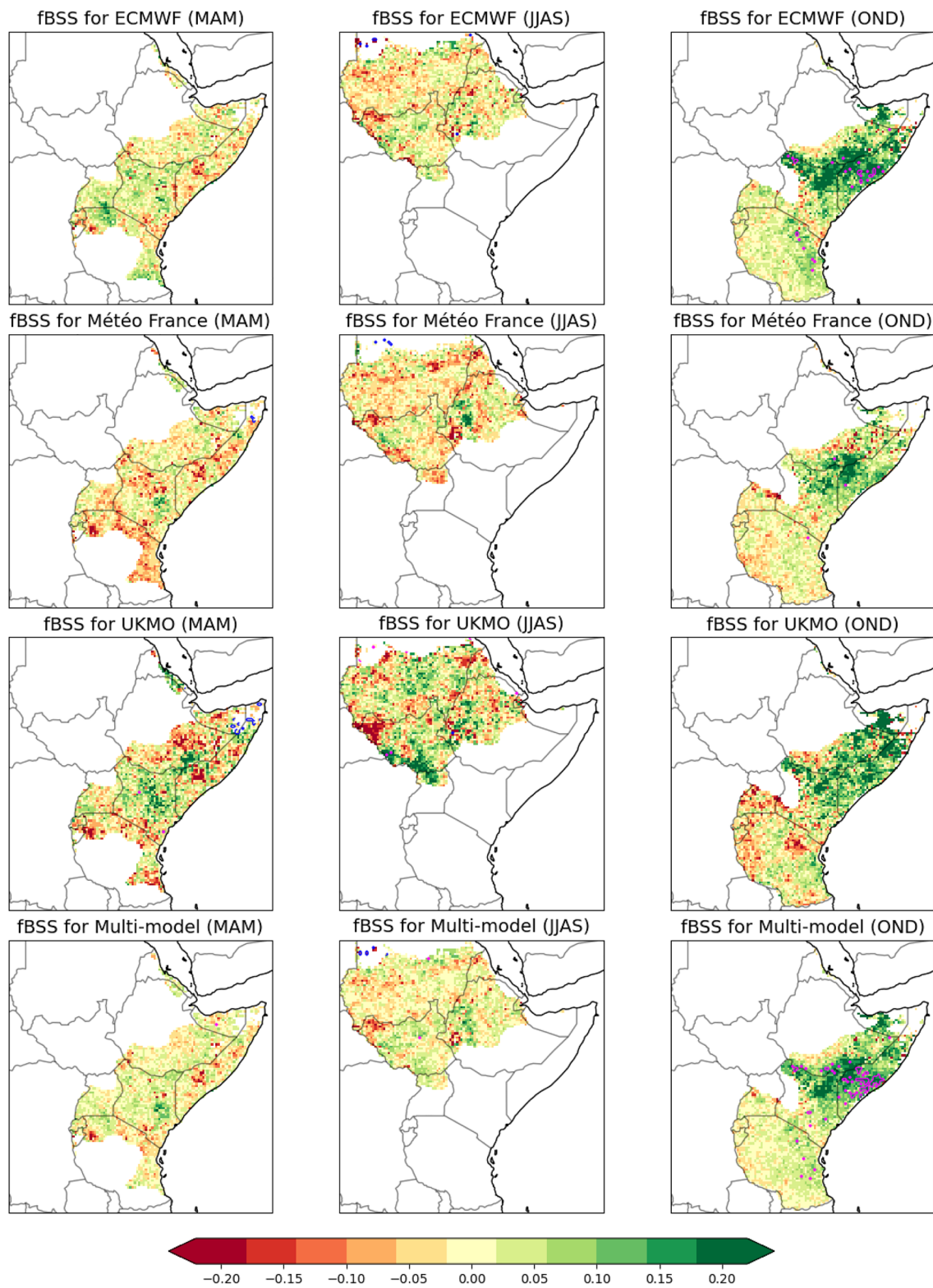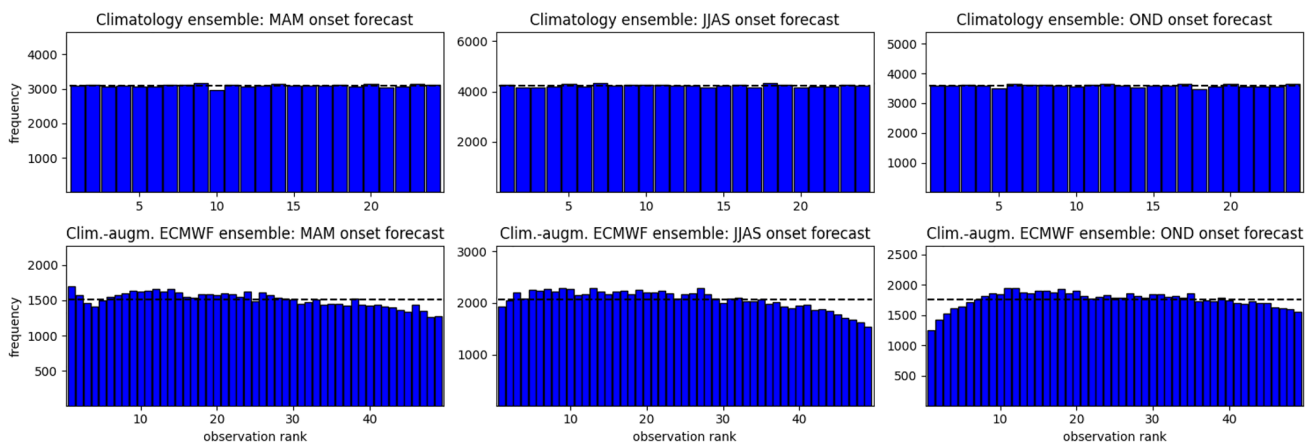
Fig. 7 Same as Fig. 6 but for fBSS

Fig. 8 Same as Fig. 5 but for a climatological and a climatology-augmented ECMWF ensemble
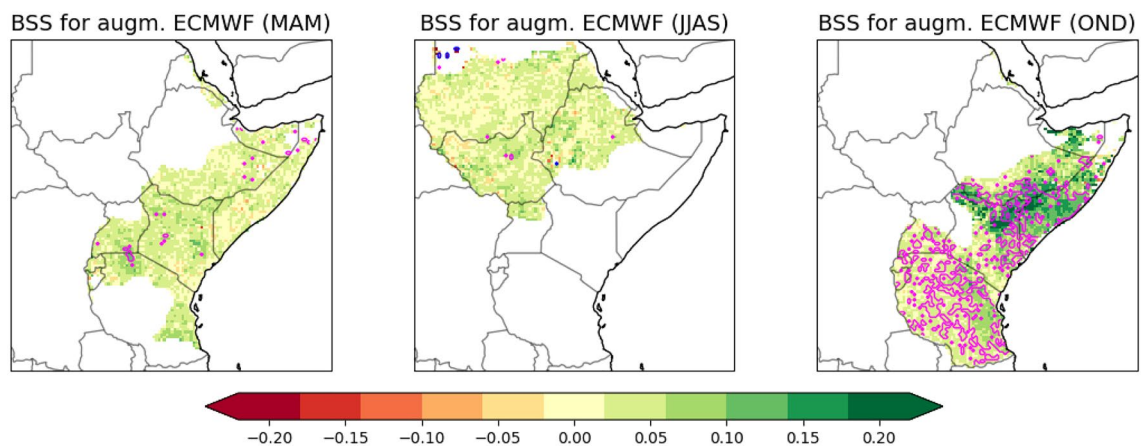

Fig. 9 Same as Fig. 6 but for the climatology-augmented ECMWF ensemble

multi-model combinations can change. The goal of this study is thus not to make a general recommendation as to which systems should be used, but to highlight the aspects that are important for this decision.

Ideally, the implicit weighting of different contributing systems (which may include climatology) resulting from differences in their respective ensemble sizes should be replaced by an explicit weighting scheme which emphasizes or de-emphasizes different systems according to their skill. Different approaches for combining probabilistic forecasts have been proposed in the literature, see e.g. Baran and Lerch (2018) and Zamo et al. (2021) for examples and comparisons. However, identifying approaches that are suitable in a seasonal forecasting context where weights need to be trained based on a limited training sample, or ensemble subselection is

required for computational reasons is still an active area of research (e.g. Hemri et al. 2020; Heinrich-Mertsching et al. 2023).

The rainy season onset predictions studied here were derived directly from long-range ensemble forecasts of daily precipitation amounts without considering additional information like the state of ENSO or IOD that is known to affect the hydroclimate in the Horn of Africa (Anderson et al. 2022). As we gain a better understanding of the drivers behind an early or delayed onset (e.g. Gudoshava et al. 2022; Roy et al. 2023), an alternative approach is the use of machine learning methods to link rainy season onset dates to these remote drivers and test whether this can tap into additional sources of predictability. Such an approach will be explored in future work.
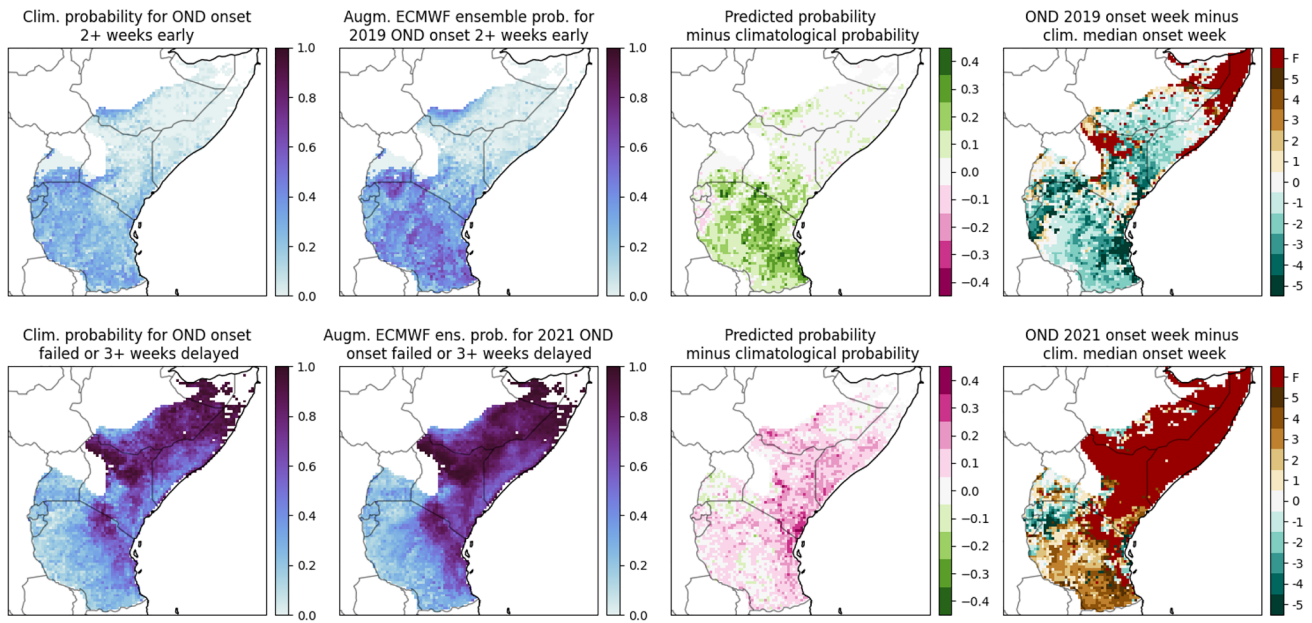
Fig. 10 Probabilities for '2019 OND onset is 2 or more weeks early' (top row) and '2021 OND onset fails or is at least 3 weeks delayed' (bottom row). The observed shifts of the respective onset week relative to the climatological median are depicted in the leftmost panels, more detailed explanations are given in the main text
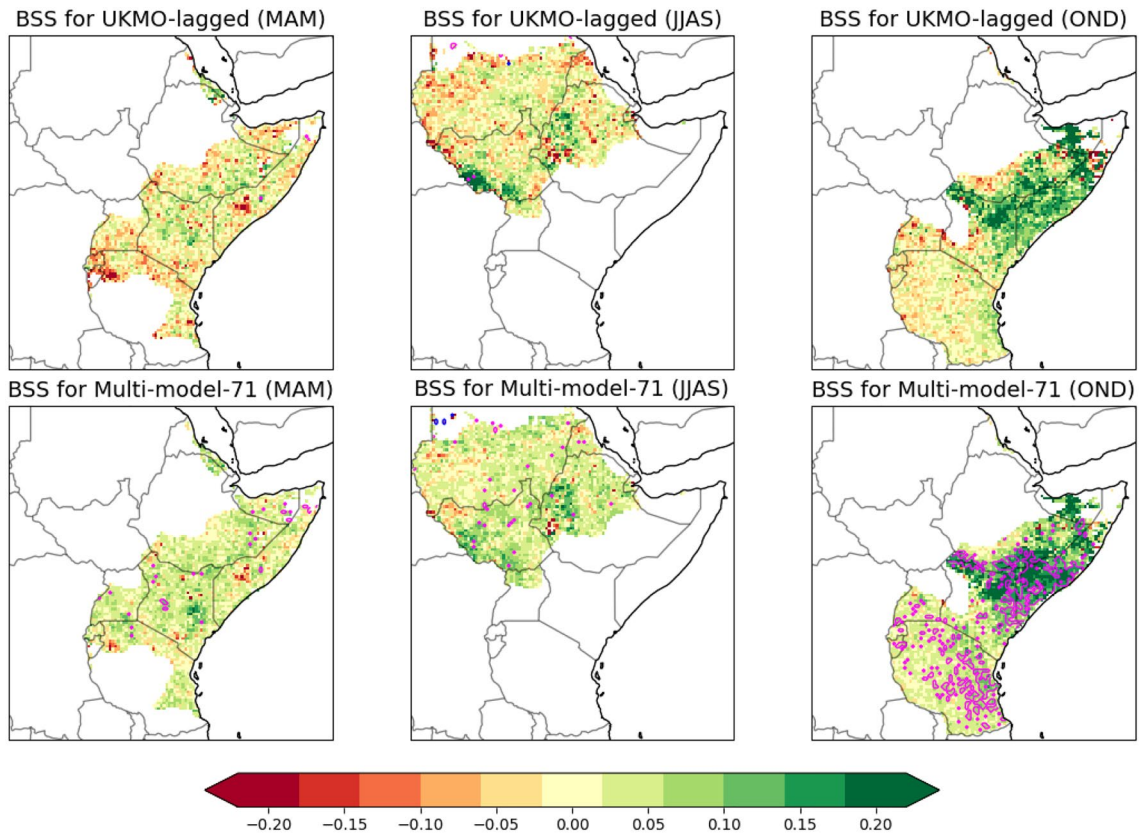

Fig. 11 Same as Fig. 6 but for a 21-member UKMO lagged ensemble (top row) and a 71-member multi-model ensemble consisting of this lagged ensemble and the ECMWF and Météo France ensemble (bottom row)

Appendix A: BSS for UKMO lagged ensemble

For the main results discussed in Sect. 4 we have used a 7-member UKMO ensemble for which the hindcast initialization dates coincide with those of the ECMWF and Météo France ensemble (1st of each month). The negative effects of this small ensemble size were discussed in Sect. 4. Unlike ECMWF and Météo France, however, UKMO runs their long-range forecast system with four different initialization dates each month: the 1st, 9th, 17th, and 25th. We can therefore construct a larger UKMO ensemble by

- considering, in addition to the UKMO hindcasts from the 1st of each month, the hindcasts from the 9th of the same and the 25th of the previous month.
- matching the forecast valid dates (i.e., using shorter forecast lead times for the 9th and longer lead times for the 25th) such that we obtain a 21-member lagged ensemble combining members from three different hindcast runs.

In this appendix, we present results of onset forecasts obtained with this UKMO lagged ensemble and with a 71-member multi-model ensemble that combines it with the ECMWF and Météo France ensemble described in Sect. 2. Figure 11 confirms the conclusions from Sect. 4 regarding the role of ensemble size: the performance of the 21-member UKMO lagged ensemble is comparable to the ECMWF ensemble and substantially better than the 7-member UKMO ensemble studied in Fig. 6. This improvement carries over to the 71-member multi-model ensemble, which attains BSSs of 0.15 and above over large parts of Kenya, Ethiopia, and Somalia for the OND onset.

Availability of data CHIRPSv2 data can be downloaded from ftp://ftp.chg.ucsb.edu/. The hindcasts data used here can be downloaded from Copernicus Climate Data Store (CDS) at https://confluence.ecmwf.int/display/CKB/C3S+Seasonal+Forecasts.

Declarations

References

Anderson JL (1996) A method for producing and evaluating probabilistic forecasts from ensemble model integrations. J Clim 9:1518–1530. https://doi.org/10.1175/1520-0442(1996)009<1518:AMFPAE>2.0.CO;2

Anderson W, Cook BI, Slinski K et al (2022) Multi-year La Niña events and multi-season drought in the Horn of Africa. J Hydrometeorol. https://doi.org/10.1175/JHM-D-22-0043.1

Anyah RO, Semazzi FH, Xie L (2006) Simulated physical mechanisms associated with climate variability over Lake Victoria basin in East Africa. Mon Weather Rev 134(12):3588–3609

Bahaga TK, Mengistu Tsidu G, Kucharski F et al (2015) Potential predictability of the sea-surface temperature forced equatorial East African short rains interannual variability in the 20th century. Quart J R Meteorol Soc 141(686):16–26

Bahaga TK, Kucharski F, Mengistu Tsidu G et al (2016) Assessment of prediction and predictability of short rains over equatorial East Africa using a multi-model ensemble. Theor Appl Climatol 123:637–649

Bahaga TK, Fink AH, Knippertz P (2019) Revisiting interannual to decadal teleconnections influencing seasonal rainfall in the Greater Horn of Africa during the 20th century. Int J Climatol 39(5):2765–2785. https://doi.org/10.1002/joc.5986

Baran S, Lerch S (2018) Combining predictive distributions for the statistical post-processing of ensemble forecasts. Int J Forecast 34(3):477–496. https://doi.org/10.1016/j.ijforecast.2018.01.005

Batté L, Déqué 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: Dyn Meteorol Oceanogr 63(2):283–299. https://doi.org/10.1111/j.1600-0870.2010.00493.x

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300

Berhane F, Zaitchik B, Dezfuli A (2014) Subseasonal analysis of precipitation variability in the Blue Nile River Basin. J Clim 27:325–344

Black E, Asfaw DT, Sananka A et al (2023) Application of TAMSAT-ALERT soil moisture forecasts for planting date decision support in Africa. Front Clim. https://doi.org/10.3389/fclim.2022.993511

Brier GW (1950) Verification of forecasts expressed in terms of probability. Mon Weather Rev 78(1):1–3

Camberlin P, Moron V, Okoola R et al (2009) Components of rainy seasons' variability in equatorial East Africa: onset, cessation, rainfall frequency and intensity. Theoret Appl Climatol 98(3):237–249. https://doi.org/10.1007/s00704-009-0113-1

Diro GT, Tompkins AM, Bi X (2014) Dynamical downscaling of ECMWF ensemble seasonal forecasts over East Africa with

RegCM3. J Geophys Res Atmos. https://doi.org/10.1029/2011JD016997

Dunning CM, Black ECL, Allan RP (2016) The onset and cessation of seasonal rainfall over Africa. J Geophys Res Atmos 121(19):11405–11424. https://doi.org/10.1002/2016JD025428

Dutra E, Magnusson L, Wetterhall F et al (2013) The 2010–2011 drought in the Horn of Africa in ECMWF reanalysis and seasonal forecast products. Int J Climatol 33:1720–1729

Ehsan M, Tippett M, Robertson A et al (2021) Seasonal predictability of Ethiopian Kiremt rainfall and forecast skill of ECMWF's SEAS5 model. Clim Dyn 57:3075–3091. https://doi.org/10.1007/s00382-021-05855-0

Ferro CAT, Richardson DS, Weigel AP (2008) On the effect of ensemble size on the discrete and continuous ranked probability scores. Meteorol Appl 15:19–24

Fitzpatrick RGJ, Bain CL, Knippertz P et al (2015) The West African monsoon onset: a concise comparison of definitions. J Clim 28:8673–8694

Funk C, Peterson P, Landsfeld M et al (2015) The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Sci Data 2(150):066. https://doi.org/10.1038/sdata.2015.66

Gebrechorkos SH, Pan M, Beck HE et al (2022) Performance of state-of-the-art C3S European seasonal climate forecast models for mean and extreme precipitation over Africa. Water Resour Res. https://doi.org/10.1029/2021WR031480

Gleixner S, Keenlyside NS, Demissie TD et al (2017) Seasonal predictability of Kiremt rainfall in coupled general circulation models. Environ Res Lett 12(11):114016. https://doi.org/10.1088/1748-9326/aa8cfa

Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc 102(477):359–378

Gudoshava M, Wainwright C, Hirons L et al (2022) Atmospheric and oceanic conditions associated with early and late onset for Eastern Africa short rains. Int J Climatol 42(12):6562–6578. https://doi.org/10.1002/joc.7627

Hamill TM, Colucci SJ (1997) Verification of eta-rsm short-range ensemble forecasts. Mon Wea Rev 125:1312–1327. https://doi.org/10.1175/1520-0493(1997)125<3C1312:VOERSR>3E2.0.CO;2

Hastenrath S, Nicklis A, Greischar L (1993) Atmospheric—hydrospheric mechanisms of climate anomalies in the western equatorial Indian Ocean. Geophys Res Lett 98:20219–20235

Heinrich-Mertsching C, Sørland SL, Gudoshava M et al (2023) Subselection of seasonal ensemble precipitation predictions for East Africa. Q J R Meteorol Soc 149(755):2634–2653. https://doi.org/10.1002/qj.4525

Hemri S, Bhend J, Liniger MA et al (2020) How to create an operational multi-model of seasonal forecasts? Clim Dyn 55(5):1141–1157

Hession SL, Moore N (2011) A spatial regression analysis of the influence of topography on monthly rainfall in East Africa. Int J Climatol 31(10):1440–1456

Korecha D, Barnston A (2007) Predictability of June-September Rainfall in Ethiopia. Mon Weather Rev 135:628–650

Liebmann B, Marengo JA (2001) Interannual variability of the rainy season and rainfall in the Brazilian Amazon Basin. J Clim 14(22):4308–4318. https://doi.org/10.1175/1520-0442(2001)014<4308:IVOTRS>2.0.CO;2

MacLeod D (2018) Seasonal predictability of onset and cessation of the East African rains. Weather Clim Extremes 21:27–35. https://doi.org/10.1016/j.wace.2018.05.003

MacLeod D (2019) Seasonal forecast skill over the Greater Horn of Africa: a verification atlas of System 4 and SEAS5—part1/2. Tech. rep., ECMWF, https://www.ecmwf.int/node/18906

Manzanas R, Gutiérrez JM, Bhend J et al (2019) Bias adjustment and ensemble recalibration methods for seasonal forecasting: a comprehensive intercomparison using the C3S dataset. Clim Dyn 53:1287–1305

Messner JW, Mayr GJ, Zeileis A (2017) Nonhomogeneous boosting for predictor selection in ensemble postprocessing. Mon Weather Rev 145(1):137–147. https://doi.org/10.1175/MWR-D-16-0088.1

Mwangi E, Wetterhall F, Dutra E et al (2014) Forecasting droughts in East Africa. Hydrol Earth Syst Sci 18(2):611–620. https://doi.org/10.5194/hess-18-611-2014

Nicholson SE (2017) Climate and climatic variability of rainfall over eastern Africa. Rev Geophys 55:590–635. https://doi.org/10.1002/2016RG000544

Nicholson SE (2018) The ITCZ and the seasonal cycle over equatorial Africa. Bull Am Meteorol Soc 99(2):337–348

Pörtner HO, Roberts DC, Adams H, et al. (2022) Climate change 2022: Impacts, adaptation and vulnerability. IPCC Sixth Assessment Report

Rasp S, Lerch S (2018) Neural networks for postprocessing ensemble weather forecasts. Mon Weather Rev 146(11):3885–3900

Roy I, Mliwa M, Troccoli A (2023) Important drivers of East African Monsoon variability and improving rainy season onset prediction. Nat Hazards. https://doi.org/10.1007/s11069-023-06223-3

Salami A, Kamara AB, Brixiova Z (2010) Smallholder agriculture in East Africa: Trends, constraints and opportunities. Tech. rep, African Development Bank Tunis, Tunisia

Segele ZT, Lamb PJ (2005) Characterization and variability of Kiremt rainy season over Ethiopia. Meteorol Atmos Phys 89:153–180. https://doi.org/10.1007/s00703-005-0127-x

Sepulchre P, Ramstein G, Fluteau F et al (2006) Tectonic uplift and Eastern Africa aridification. Science 313(5792):1419–1423

Skamarock WC, Klemp JB, Dudhia J, et al (2008) A Description of the Advanced Research WRF Version 3. University Corporation for Atmospheric Research, No NCAR/TN-475+STR https://doi.org/10.5065/D68S4MVH

Taillardat M, Mestre O, Zamo M et al (2016) Calibrated ensemble forecasts using quantile regression forests and ensemble model output statistics. Mon Weather Rev 144:2375–2393

Talagrand O, Vautard R, Strauss B (1997) Evaluation of probabilistic prediction systems. In: ECMWF Workshop on Predictability. ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, United Kingdom., p 1–25

Thiery W, Davin EL, Panitz HJ et al (2015) The impact of the African Great Lakes on the regional climate. J Clim 28(10):4061–4085

Ummenhofer CC, Gupta AS, England MH et al (2009) Contributions of Indian Ocean Sea Surface Temperatures to Enhanced East African Rainfall. J Clim 22:993–1013

Walker D, Birch C, Marsham J et al (2019) Skill of dynamical and GHACOF consensus seasonal forecasts of East African rainfall. Clim Dyn 53:4911–4935. https://doi.org/10.1007/s00382-019-04835-9

Wilks DS (2016) The stippling shows statistically significant grid points: How research results are routinely overstated and overinterpreted, and what to do about it. Bull Am Meteorol Soc 97:2263–2273. https://doi.org/10.1175/BAMS-D-15-00267.1

Yang W, Seager R, Cane MA et al (2015) The annual cycle of East African precipitation. J Clim 28:2385–2404. https://doi.org/10.1175/JCLI-D-14-00484.1

Young HR, Klingaman NP (2020) Skill of seasonal rainfall and temperature forecasts for East Africa. Weather Forecast 35(5):1783–1800. https://doi.org/10.1175/WAF-D-19-0061.1

Zamo M, Bel L, Mestre O (2021) Sequential aggregation of probabilistic forecasts - Application to wind speed ensemble forecasts. J R Stat Soc C 70(1):202–225. https://doi.org/10.1111/rssc.12455

Zampieri M, Toreti A, Meroni M et al (2023) Seasonal forecasts of the rainy season onset over Africa: preliminary results from the FOCUS-Africa project. Clim Serv 32(100):417. https://doi.org/10.1016/j.cliser.2023.100417