



Skilful subseasonal forecasts of dry spells: a case study for Malawi

Erik W. Kolstad^{1,2} · Rondrotiana Barimalala¹ · Douglas J. Parker^{1,3}

Received: 12 March 2025 / Accepted: 8 October 2025
© The Author(s) 2025

Abstract

Smallholder farmers in Sub-Saharan Africa are vulnerable to adverse fluctuations in rainfall, such as dry spells during the critical early stages of the rainy season. In this study, we demonstrate that Malawi is prone to shifts from periods with limited dry spell occurrence to more widespread dry spells later in the season. We develop a predictive model for dry spells, aiming to provide farmers with actionable information to support agricultural decision-making and enhance resilience. The model, based on a dynamical subseasonal prediction system and validated using reanalysis and satellite-based data, focuses on Malawi as a case study. This model has significant skill in predicting the occurrence of at least one dry spell within the three weeks following initialisation, consistently outperforming a climatology-based reference model. Furthermore, we show that the model is applicable beyond Malawi, specifically in East Africa during both the March–May “long rains” and the October–November “short rains”, highlighting its broader relevance for regions where dry spells pose an agricultural risk. The results demonstrate that subseasonal forecasts have the potential to bridge the gap between long-range seasonal outlooks and short-term weather forecasts. Unlike seasonal forecasts, which lack skill at long lead times, subseasonal predictions offer both a longer planning horizon than weather forecasts and greater skill in capturing dry spell risks at actionable lead times. By integrating subseasonal forecasts into national climate services, policymakers and agricultural extension services could provide more timely and targeted advice, potentially helping to mitigate the most severe impacts of dry spells on food production and rural livelihoods.

Keywords Malawi · Africa · Dry Spells · Subseasonal Forecasting · S2S

1 Introduction

A large body of research has highlighted the broad value of climate forecasts in agricultural decision-making (e.g. Hansen et al. 2011; Roudier et al. 2014; White et al. 2017; Streefkerk et al. 2022; White et al. 2022). Subseasonal weather forecasts, which cover lead times up to about six

weeks, bridge the gap between weather and seasonal climate forecasts (Vitart and Robertson 2019) and play a key role in the “ready-set-go” framework (Goddard et al. 2014). These forecasts have a demonstrated impact on agricultural planning, disaster preparedness, and water resource management, particularly in regions where livelihoods are highly sensitive to climatic variability (e.g. White et al. 2017; Hirons et al. 2023).

The potential benefit of climate forecasts is large in areas with rain-fed agricultural systems, like Malawi, where droughts have become more frequent, intense, and widespread over the past two decades, with severe consequences for food and water security, energy resources, and rural livelihoods (Ndlovu et al. 2024). One of the most hazardous phenomena for farmers is the occurrence of erratic rains/dry spells, which place pressure on rain-fed agriculture and food security (Coulibaly et al. 2015; Chimimba et al. 2023). The need for dry spell prediction is particularly acute around the first rains, which are often perceived as the onset of the rainy season but may, in some cases, be followed by

✉ Erik W. Kolstad
ekol@norceresearch.no

Rondrotiana Barimalala
ronb@norceresearch.no

Douglas J. Parker
d.j.parker@leeds.ac.uk

¹ NORCE Research, Bjerknes Centre for Climate Research, Bergen, Norway

² Chr. Michelsen Institute, Bergen, Norway

³ National Centre for Atmospheric Science (NCAS), School of Earth and Environment, University of Leeds, Leeds, UK

a prolonged dry period. Such “false onsets” are so common in Malawi that they have a name: the “fire extinguisher”, or *Chizimalupsya* (Department of Climate 2024). A worst-case scenario arises when farmers plant too early after the perceived onset, only for premature rains to be followed by dry spells that kill the seeds before they can germinate. A particularly damaging example occurred in the 2012–13 season, when farmers in central Malawi had to replant maize six times before the onset of sustained rains (Mittal et al. 2021). The 2023–24 season was not only another instance of delayed onset (Department of Climate 2023); it was also characterised by prolonged dry spells following initial rainfall in some areas. These conditions had severe agricultural impacts: early rains prompted planting, but the subsequent dry spells led to widespread crop failure, contributing to a food security crisis in which millions required humanitarian assistance (ROSEA 2024; see also Fig. 2).

Coulbaly et al. (2015) found that onset predictions were the most sought-after climate service among Malawian farmers, underscoring their perceived importance for agricultural planning despite inherent skill limitations at extended lead times. As part of their response to user-driven needs for information, the Malawian Department of Climate Change and Meteorological Services (DCCMS) issues seasonal forecasts in September (Department of Climate 2024). These forecasts include predictions of rainy season onset and dry spell likelihood, despite being issued with substantial lead time before the typical November start of the rainy season. DCCMS seasonal forecasts rely on analogue years with similar El Niño–Southern Oscillation (ENSO) signatures, with predicted ENSO conditions derived from global seasonal forecasting models. However, Demissie and Gebrechorkos (2024) demonstrated that only southern Malawi exhibits a significant (negative) correlation with ENSO, while northern regions show non-significant (positive) correlations. This spatial heterogeneity reflects Malawi’s position at the boundary between East Africa, where October–December rainfall correlates positively with ENSO (Kolstad and MacLeod 2022), and Southern Africa, where this correlation is negative (Ratnam et al. 2014). Given the limited skill of current seasonal forecasts (Mittal et al. 2021), there would be considerable potential value in developing shorter-lead-time predictions with higher accuracy. While such forecasts may not support long-term decisions like crop variety selection, they could play an important role in guiding time-sensitive choices, most notably when to plant (Streefkerk 2020).

Effectively implementing forecasts for shorter time horizons than seasonal forecasts requires careful consideration of how rainy season onset is conceptualised. A wide range of agronomically defined onset definitions exists, typically incorporating both a rainfall threshold and a criterion

ensuring that no dry spell follows shortly thereafter (Fitzpatrick et al. 2015). Defining rainy season onset by combining a rainfall threshold with a subsequent no-dry-spell criterion presents two major challenges.

First, there is a forecast skill horizon problem: while we found no evaluations of subseasonal rainfall prediction skill for Malawi specifically, studies for East Africa suggest useful skill at lead times of up to 3–4 weeks (de Andrade et al. 2021; Kolstad et al. 2021; MacLeod et al. 2021). Given this limitation, predicting both the initial onset and subsequent dry spells is difficult because much of the skilful forecast horizon is already “used up” in determining the onset itself.

Second, there is a spatial coherence issue: onset dates derived from rainfall-based definitions tend to be highly variable across small spatial scales (Fitzpatrick et al. 2016; Young et al. 2020), often lacking clear correlations with regional climatic drivers. This limits their practical utility for large-scale forecasting and decision-making.

To address these limitations, DCCMS provides daily, five-day, and weekly forecasts, along with ten-day agro-meteorological bulletins throughout the season. However, while these short-term updates are valuable, they do not bridge the gap between seasonal outlooks and real-time weather forecasts. In particular, subseasonal forecasts – providing predictions with lead times of two to four weeks – could significantly enhance early warnings for dry spells and other agricultural risks. By offering a longer planning horizon than weather forecasts while remaining more up-to-date than seasonal outlooks, subseasonal products would better align with farmers’ decision-making needs (Streefkerk 2020). Yet, such forecasts are not currently part of DCCMS’s portfolio.

This gap in forecast provision is a key factor in farmers’ reluctance to rely on predictions. Many cite discrepancies between forecasted and observed weather as a reason for their scepticism, reinforcing fatalistic attitudes towards climate variability (Mkwambisi et al. 2020). Addressing these challenges requires not only improving forecast skill but also enhancing how forecasts are communicated and updated. Subseasonal forecast products focused on key agricultural risks, such as dry spells, could help bridge this gap by offering information that is both actionable and aligned with farmers’ planning horizons.

Acknowledging the differing lead-time requirements of onset and dry spell prediction, we focus here on forecasting dry spells around the time of seasonal onset: arguably the most critical short-term prediction for farmers assessing whether observed rainfall marks a true or false onset. Before planting, these forecasts might help farmers avoid the adverse effects of a false onset by informing decisions on whether to delay planting until sustained rainfall is more likely (Streefkerk 2020). Such a proactive use of

probabilistic forecasts could reduce the likelihood of crop failure due to poor germination or early water stress. After planting, the forecasts would remain valuable by guiding adaptive measures when an increased probability of dry spells is indicated. For example, farmers could implement soil and water conservation practices, such as mulching or tied ridges, to retain moisture (Rockström et al. 2010; Marongwe et al. 2011), or prioritise supplemental irrigation where possible. Such applications highlight the potential of forecast information to support both pre-emptive and reactive decision-making, ultimately enhancing resilience in the face of uncertain rainfall patterns.

Here we present a novel framework for leveraging sub-seasonal forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF), reanalysis data, and satellite-derived rainfall data to develop a dry spell prediction model for the coming three weeks. This framework was developed within the interdisciplinary ARCS (Agricultural Resilience through Climate Services) project, which focuses on tailoring climate forecasts for actionable agricultural advice. Preliminary focus group interviews conducted in Malawi during the 2023/24 and 2024/25 rainy seasons confirmed that dry spells and false onsets pose a major challenge to farmers' livelihoods.

We seek answers to three main questions. First, we ask how dry spells vary within a season and from year to year in Malawi. By taking a country-aggregated view, we investigate whether widespread dry spells commonly occur after periods with limited dry spells. Second, we question whether it is possible to skilfully predict the likelihood of dry spells in Malawi around the critical start of the rainy season. Third, we assess if the findings for Malawi are generalisable to other parts of Sub-Saharan Africa, focusing on East Africa. While Malawi serves as a case study, the framework's use of global data allows for adaptation to other regions facing similar challenges.

2 Data and methods

2.1 Forecast and reforecast data

The analysis is based on subseasonal reforecasts from the ECMWF's Integrated Forecasting System (IFS), which combines a sophisticated data assimilation system and a global numerical model to produce operational forecasts for the extended range (days 1–46). Until 11 November 2024, the IFS forecasts were produced each Monday and Thursday, and after this, every day. Each forecast consists of 100 perturbed ensemble members and one control run. Here we only use the perturbed members.

The bulk of our analysis is based on IFS reforecasts, which provide a consistent baseline for evaluating forecast skill by offering historical forecast data that can be compared against observations to correct biases and calibrate probabilities. Each IFS reforecast consists of 11 ensemble members and spans the 20 years prior to the forecast's initial date. Reforecasts are produced on the fly, at the same time as operational forecasts.

To focus on the part of the season when dry spells are neither very rare nor very common (for reasons explained in Sect. 2.6), we only included reforecasts for reference dates between 15 October and 15 December. We included reforecast initial dates from both 2023 and 2024. While either year provides a full 20-year reforecast set (e.g., 2023 includes 2003–2022), using both years increases the number of reference dates. However, this also means that edge years (2003 and 2023) are sampled only once, resulting in slightly fewer reforecast instances for those years. There is also another source of skewness in the data. After the change to daily forecast production on 11 November 2024, reforecasts were only issued every other day (on odd dates). In our analysis, we included all available reforecast dates during the period 15 October to 15 December, meaning that the period after 11 November 2024 contains more densely sampled data than the preceding period during the same year. This introduces a slight overall weighting toward the later weeks of the season, which we consider negligible in terms of its impact on results. In total, the dataset includes 43 unique forecast initialisation dates across 2023 and 2024. With 20 reforecast years per initial date, this yields a total of 860 reforecast instances, each with 11 ensemble members.

The IFS data were downloaded with a grid spacing of 0.35 degrees, approximating the model's original resolution of about 30 km, and we analysed daily accumulated precipitation from the first three weeks of each reforecast ensemble member. The work of Fitzpatrick et al. (2016) for West Africa indicates that for most locations there is enough spatial coherency that a spatially aggregated measure of interannual variability in onset is meaningful on the 30 km scale.

2.2 Observational data

Observational datasets, while crucial for ground truth, are often hampered by issues such as undercatch (Adam and Lettenmaier 2003), data gaps, varying instrument quality, inconsistencies in measurement techniques, gaps in station coverage, and interpolation uncertainties, particularly in regions with sparse observational networks. These limitations introduce uncertainties that can complicate model evaluation.

To validate and calibrate the model forecasts, we used precipitation data from the ERA5 reanalysis (Hersbach et al. 2020). Reanalysis products offer the advantage of being both temporally and physically consistent. Since ERA5 is produced using the same IFS model as the forecasts, the validation effectively compares the model against a version of itself. This setup provides a controlled environment to isolate and assess the model's intrinsic performance, with minimal influence from observational uncertainties.

In addition to ERA5, we used version 7 (V07) of the Integrated Multi-satellite Retrievals for GPM (IMERG) dataset (Huffman et al. 2023) to validate and bias-correct the forecast model. IMERG provides precipitation estimates at high temporal resolution based on a combination of microwave and infrared satellite data. Unlike similar products like TAMSAT (Maidment et al. 2017) and CHIRPS (Funk et al. 2015), which rely on the disaggregation of multi-day rainfall estimates to produce daily values, IMERG offers native daily precipitation fields. This feature makes it well-suited for assessing dry spell conditions at daily resolution. However, IMERG is not free of uncertainty. IMERG products, including V07, tend to overestimate light rainfall events, which is a common issue across different versions and regions (Yang et al. 2020; Li et al. 2021; Wei et al. 2025). While the Final Run product (which we used) includes a monthly bias correction using gauge data, this adjustment is temporally coarse and often limited in regions with sparse observational coverage, such as Malawi.

To upscale the IMERG data from their native 0.1° resolution to the target 0.35° grid, we employed a Gaussian-weighted interpolation scheme using the nine nearest neighbours of each target grid point. The weights were based on the distances in degree-space. To prevent numerical instability and ensure meaningful contribution from neighbouring points, we applied a minimum threshold $\epsilon = 0.04^\circ$ (about 4 km) to all distances before the weight calculation, so that $d_i = \max(\epsilon, d_i^*)$, where d_i^* is the actual distance to the i -th neighbour, and d_i is the effective distance. The weights for each neighbour were calculated as $w_i = \exp\left(-\frac{d_i^2}{2\sigma^2}\right)$, where $\sigma = 0.07^\circ$ is the bandwidth parameter. This approach ensured that even exact coordinate matches received at most 30% of the total weight, thereby providing spatial smoothing appropriate for upscaling precipitation fields. The values of ϵ and σ were determined through empirical testing of several parameter combinations. A more rigorous optimisation of the interpolation technique would be recommended for operational purposes.

It is important to acknowledge that using ERA5 also entails limitations. Reanalysis products are hampered by scarcity in the observations that drive them, and in Malawi there are relatively few observations available for assimilation.

Studies indicate that ERA5 tends to overestimate the number of wet days (> 1 mm) in Europe (Bandhauer et al. 2022; Gomis-Cebolla et al. 2023), associated with an excessive frequency of very light rain (Ahn et al. 2024). Similar biases are found in Africa. Specifically, Lavers et al. (2021) found a wet bias for the IFS model in dry regions, while Lavers et al. (2022) reported that ERA5 overestimates precipitation on dry days, with biases in southeastern Africa ranging from 0 to 0.5 mm per day in October. To account for this, we adopt a relatively high wet/dry threshold of 2 mm per day, a choice also made in previous studies (e.g. Haghtalab et al. 2019; Streefkerk et al. 2022). Lower thresholds (1 and 1.5 mm) produced qualitatively similar results.

Ultimately, while ERA5 is not a perfect ground truth, it provides a physically consistent and spatially complete dataset for model calibration and validation. For operational applications, local observations should be incorporated to refine forecasts and correct systematic biases. The method presented here should therefore be viewed as a preliminary assessment of subseasonal predictability rather than a substitute for observation-based calibration.

2.3 Bias-correction

Any forecast model, including IFS, has biases that can stem from factors such as inaccurate topography representation, poor soil moisture feedbacks, or circulation biases. Forecast models are also known to experience drift with increasing lead times (Hermanson et al. 2021), which represents another bias with respect to observational data. It was therefore necessary to bias-correct the IFS data and to account for varying lead times within a forecast. We employed a method similar to quantile mapping. Specifically, for each valid time of each reforecast, we pooled reforecast data for the same calendar day across all the years. With 11 ensemble members available for each date, we matched these pooled IFS data with ERA5 data for the same dates. To maintain consistent sample sizes between IFS and ERA5, we used an 11-day window centred on each calendar day for ERA5. We then computed the percentile corresponding to 2 mm of precipitation in the ERA5 data. This percentile was used to identify the equivalent lead time-dependent precipitation threshold in the IFS data, such that IFS precipitation values below this threshold were classified as *no precipitation*.

When investigating the forecast model based on IMERG, we bias-corrected the model with respect to these data, using a no-rain threshold of 1 mm rather than 2 mm. This lower threshold was selected because the IMERG algorithm does not have the same tendency to produce excessive light rain as ERA5. This quantile mapping approach revealed large differences between IMERG and IFS. The IFS threshold values based on the same quantile that corresponded to

1 mm in IMERG were typically in the range between 2.5 and 3 mm.

2.4 Defining dry spells

The dry spell definition used in this study is a period of seven consecutive dry days (CDD), meaning no single day within this period exceeds 2 mm of precipitation. Locally tailored dry spell definitions vary by region, climatological conditions, and time of the year (e.g. Sivakumar 1992; Sharma 1996; Barron et al. 2003; Thoithi et al. 2021). For instance, DCCMS considers period with nine CDDs when they issue dry spell forecasts with a 10-day lead time. Only minor differences were seen when we repeated the analysis herein for nine-day dry periods.

For each dataset (IFS and ERA5), we first calculated a binary variable for each date and grid point, assigning a value of 0 if the daily accumulated total precipitation (bias-corrected for IFS; see previous section) exceeded 2 mm, and

1 otherwise. We then revisited each date to check for the occurrence of any dry spells (seven consecutive days of 1's) during the subsequent three weeks. For each date, a new binary variable, denoted as δ , was assigned a value of 1 if at least one seven-day dry spell occurred during these three weeks, and 0 if there were no dry spells.

We illustrate how δ was calculated in Fig. 1. In the top row, the daily rainfall for a random grid point (marked in the maps in the bottom row of the figure) in Malawi is shown for three 21-day periods in 2009. During the first of these periods, illustrated in Fig. 1a, all the days up to day 13 had less than 2 mm. Consequently, the δ variable for the starting date of this period (18 October) was set to 1. In the second period, there were at most four days in a row with rain below 2 mm, resulting in $\delta = 0$ for the reference date of 2 November. During the final period, starting on 17 November and shown in Fig. 1c, there were seven dry days at the end, yielding a δ value of 1.

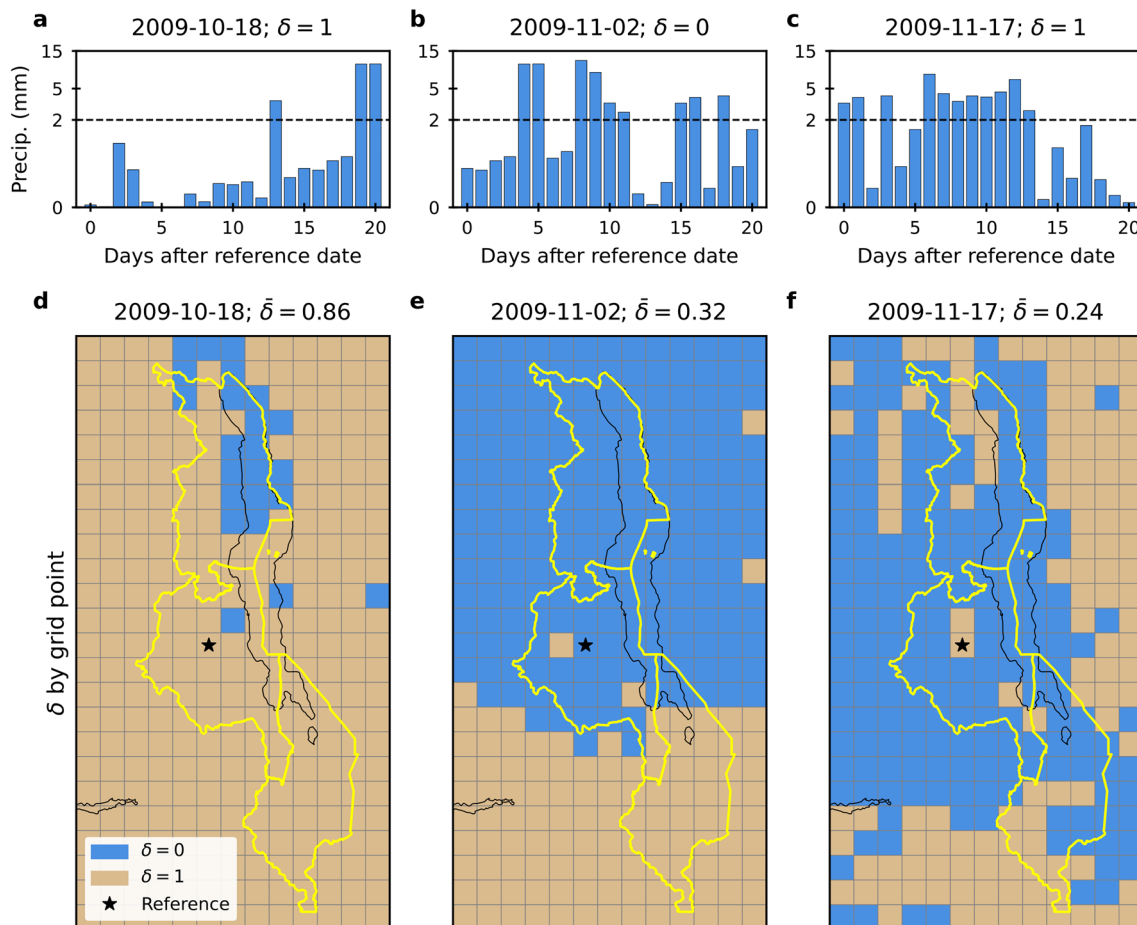


Fig. 1 Illustration of δ (top row) and $\bar{\delta}$ (bottom row). (a–c) Daily accumulated precipitation for a randomly selected grid point within Malawi (marked in d–f) for each of the 21 days following the dates in the captions. The captions also indicate the δ values for that grid point. (d–f) Maps of δ for each grid point for the same dates as in

(a–c), with the captions denoting $\bar{\delta}$, i.e., the mean of δ within Malawi's borders. The yellow lines represent the borders of the three regions of Malawi (Northern, Central, and Southern), and the thin lines show lake boundaries

In Sect. 3.1, we analyse dry spells for Malawi as a whole. To facilitate this analysis, we define $\bar{\delta}$ as the mean δ value across all grid points in the country. This variable is illustrated in the bottom row of Fig. 1, where δ values are mapped for each grid point for the same reference dates as those in the top row. The first panel in Fig. 1d shows that most of the grid points within Malawi's borders had $\delta = 1$ during the first period, including the reference grid point from Fig. 1a, yielding $\bar{\delta} = 0.86$. The map for the second reference date indicates that dry spells during this period were mainly confined to the Southern region, which resulted in $\bar{\delta} = 0.32$. By the final date, only scattered areas of Malawi experienced dry spells, and $\bar{\delta}$ was 0.24.

2.5 Dry spell prediction models

A key objective of this study is to assess the skill of the IFS model in predicting dry spells. To do this, we constructed two probabilistic dry spell prediction models: one based on post-processed IFS data and the other based on ERA5 data. The latter will be referred to as the *climatology-based* model henceforth. For this model a forecast is fixed by location and day of year.

The prediction models were developed using a leave-one-out cross-validation approach. For each year, IFS precipitation data were bias-corrected using reforecast and reanalysis data from the other 19 years, applying the methodology described in Sect. 2.3. We then used these bias-corrected precipitation data to compute δ for the initial date of each grid point and each ensemble member of each reforecast. This binary δ variable indicates whether any dry spells were predicted during the first three weeks of the forecast period.

The probabilistic IFS-based dry spell prediction for each reforecast was calculated as the fraction of ensemble members (out of 11) for which $\delta = 1$, i.e., those that predicted at least one dry spell during the first three weeks. This fraction is equivalent to the mean of the binary δ values across the ensemble members. A value of 1 indicates that all ensemble members predicted a dry spell, while a value of 0 means that none did. For the climatology-based prediction using ERA5, we predicted the chance of dry spells as the climatological frequency (leaving the prediction year out).

To avoid inflating the skill of this model, which has 19 available years compared to the 11 ensemble members of the IFS-based model, we followed Müller et al. (2005) and Weigel (2011) and computed the frequency based on 11 random years (with replacement) out of the 19 available years 1,000 times and used the average frequency as the predicted dry spell probability.

2.6 Forecast performance metrics

The Brier Score (Brier 1950) is a proper scoring rule that quantifies the accuracy of probabilistic predictions. For a binary event (like the occurrence or non-occurrence of dry spells), the Brier Score is defined as:

$$BS = \frac{1}{N} \sum_{i=1}^N (f_i - o_i)^2 \quad (1)$$

where N is the number of predictions, f_i is the forecast probability of the event occurring for the i -th case, and o_i is the observed outcome (1 for occurrence, 0 for non-occurrence). Here “observed” refers to ERA5 data.

The Brier Score ranges from 0 to 1, with 0 indicating perfect forecasts and 1 indicating the worst possible forecasts. A useful benchmark can be derived from a model that always predicts a 50% chance of occurrence: this model gets a Brier Score of 0.25 no matter what happens in reality.

The Brier Skill Score (BSS) (Wilks 2019) is a normalised measure that compares the Brier Score of a forecast to that of a reference forecast (the climatological forecast based on ERA5 in our case) and is defined as:

$$BSS = 1 - \frac{BS_M}{BS_C}, \quad (2)$$

where the “M” and “C” subscripts denote “Model” and “Climatology”, respectively. A BSS of 1 indicates perfect skill, while a BSS of 0 indicates no skill relative to the reference forecast. A negative BSS suggests that the forecast is worse than the reference.

The Brier Score, and consequently the BSS, are known to be less informative for rare events (Lawson et al. 2024). This limitation arises because the Brier Score is sensitive to climatological frequencies: when an event is rare, a naive forecast that simply reflects its low climatological probability can achieve a good score. Since the event rarely occurs, even an unskilled model predicting near-zero probabilities will appear accurate. A similar effect is seen for very frequent events. In such circumstances the BSS struggles to distinguish genuinely skilful models from those merely capturing event frequency (Bröcker and Smith 2007; Wilks 2019).

Additionally, when event occurrence varies seasonally – being rare in some periods and frequent in others – this imbalance further reduces the BSS's effectiveness in assessing forecast skill. To address this, we restricted our analysis to initial dates around the start of the rainy season, from the middle of October to the middle of December. This avoids periods when dry spells are either ubiquitous (before

the rainy season begins) or nearly absent (once it is fully underway), ensuring that the BSS more effectively differentiates meaningful skill from background climatology. Note that we do not investigate the normal cessation period from March to May, even though dry spells may also be relevant for agricultural decisions during this time.

Reliability and sharpness diagrams provide complementary insights into the quality of probabilistic forecasts. A reliability diagram assesses how well forecast probabilities correspond to observed frequencies, with a perfect forecast following the 1:1 diagonal line. Deviations from this line indicate where probabilities systematically overestimate or underestimate the likelihood of an event. Sharpness, on the other hand, measures the concentration of forecasts in extreme probability bins, with a higher sharpness indicating a stronger tendency to issue confident predictions. However, sharpness alone does not imply useful forecasts: a forecast may be sharp but poorly calibrated, or reliable but lack discriminatory power; that is, it fails to distinguish between events and non-events.

Receiver Operating Characteristic (ROC) curves compare the discriminatory skill of the models by plotting the Probability of Detection (or *Hit Rate*) against the False Alarm Rate across various decision thresholds. In the context of forecasting, ROC curves illustrate how well a model distinguishes between different binary outcomes (e.g. occurrence vs. non-occurrence of dry spells). Each point on the ROC curve represents a different threshold; the specific value of forecast probability that serves as a cut-off for making a decision about the occurrence of an event. The ROC curve illustrates the trade-off between correctly identifying positive events (hits) and incorrectly identifying negative events (false alarms). A curve that is closer to the top left corner signifies higher accuracy, as it indicates both high sensitivity (ability to detect true positives) and high specificity (ability to avoid false positives). The Area Under the Curve (AUC) quantifies this performance; a higher AUC indicates a model with better discriminatory skill.

By presenting the BSS alongside reliability, sharpness, and ROC curve diagrams, we provide a comprehensive comparison between the IFS-based and climatology-based prediction models. Together, these metrics highlight the trade-offs between forecast certainty, accuracy, and usefulness for decision-making.

2.7 Significance testing

To assess statistical significance, we applied a bootstrapping approach, generating confidence intervals (CIs) through random resampling with replacement. To preserve the temporal structure and autocorrelation of dry spells within seasons, resampling was performed along the year dimension.

For example, when evaluating a metric such as the Brier Skill Score over a 20-year hindcast period, each iteration involved randomly selecting 20 years (with replacement), retaining the full seasonal cycle within each year. A value was deemed significantly positive or negative if its CI did not contain zero. We conducted 1,000 bootstrap iterations and used a 5% significance level (based on 95% CIs). Additional details are provided where relevant.

2.8 Colour maps

We used colour maps designed by Fabio Crameri (Crameri et al. 2020) to ensure accessibility and reproducibility.

3 Results

3.1 Observed dry spells

To explore the interannual and intraseasonal variability of dry spells in Malawi, we computed $\bar{\delta}$ (see Sect. 2.4) for each day between 1 September and 1 January for the years in the period 1980–2024, using ERA5 data. Note that this period is longer than the reforecast period.

The daily $\bar{\delta}$ values shown in Fig. 2 point to several periods that may have posed risks to agricultural planning. In 2024 and several other years, periods with limited dry spells (blue cells) were followed by widespread dry spells conditions (beige cells) sometime later in the season. The figure also highlights other notable features, such as the persistence of moderate dry spells (green cells) extending to 1 January 2016. The 2015/2016 season was marked by widespread drought conditions in southern Malawi (Mkwambisi et al. 2021). A daily dry spell chart for the Southern Region (not shown) exhibits widespread dry spells (beige cells) during the three-week periods beginning in late December 2015.

Any transition from moderate or limited dry spells to more widespread dry spells – i.e. from blue to green or beige, or from green to beige in Fig. 2 – represents a potential agricultural risk. Such transitions occurred 35 times over the 45 seasons shown in Fig. 2, nearly once per year on average. The high frequency of these shifts likely heightens farmers' receptiveness to subseasonal dry spell forecasts.

A regional breakdown for Malawi's three regions – Northern, Central, and Southern; see borders in Figure 1 – revealed 44, 33, and 56 such transitions, respectively, highlighting that within-season rainfall volatility in recent decades has been most (least) pronounced in the Southern (Central) Region.

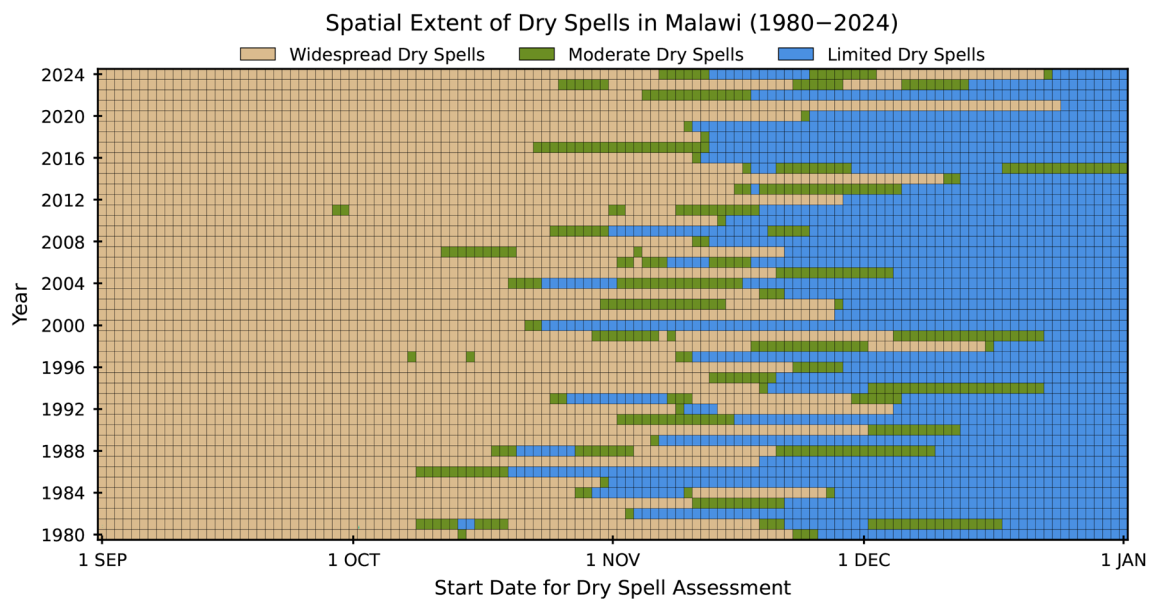


Fig. 2 Fraction of grid points in Malawi experiencing at least one seven-day dry spell during the 21 days following each date on the x-axis, shown for each year from 1980 (bottom row) to 2024 (top row).

The fractions, $\bar{\delta}$, are classified into three categories: Widespread Dry Spells ($\bar{\delta} \geq 2/3$), Moderate Dry Spells ($1/3 \leq \bar{\delta} < 2/3$), and Limited Dry Spells ($\bar{\delta} < 1/3$)

3.2 Forecasting dry spells

3.2.1 Illustration of dry spell forecasts

In this section we provide visual examples of dry spell predictions, occurrence and model accuracy, for both the climatology-based and the IFS-based models. Note that the IFS-based probabilistic forecasts were based on the 100 ensemble members of the operational forecasts, whereas the model performance analysis later in the paper was based on the 11 reforecast ensemble members.

Figure 3 presents the climatology-based forecasts. This model relies on only 20 years of data, giving it an effective ensemble size of 20 – far smaller than the IFS-based model, which has 100 ensemble members. A lower skill is therefore expected. Nevertheless, these examples illustrate a key feature, namely how the climatology-based dry spell probabilities gradually decrease from high at the start of the period (Fig. 3a) to low at the end (Fig. 3e). These changes in probability occur independently of the actual weather conditions, as they reflect only the average evolution across the past 20 years.

The accuracy of the climatology-based predictions fluctuates strongly across the different initial dates. For instance, the mean $(f - o)^2$ value within Malawi for the first initial date is 0.01 (Fig. 3f). This forecast was excellent because the predicted period was as dry as expected given the climatology. In contrast, subsequent forecasts, which reflect the climatological transition from high to low dry spell probabilities across the country, perform increasingly

poorly because the weather deviated from the climatological evolution. The final forecast is particularly inaccurate, with a mean $(f - o)^2$ value of 0.74 across Malawi (Fig. 3j). The reason for this poor performance is the re-emergence of widespread dry spells seen in Fig. 2: dry spells occurred throughout the country, as shown by the filled circles in Fig. 3e. This was unexpected according to the climatology and highlights a key limitation of the climatology-based model: its inability to reflect real-time anomalies makes it unreliable when conditions diverge from the seasonal norm.

The IFS-based forecasts shown in Fig. 4 consistently perform better than the climatological model. Although this model also performs relatively poorly for the last three-week period (Fig. 4j), its mean $(f - o)^2$ value of 0.46 is better than the higher value of 0.74 for the climatological forecast. We note that the model adjusts to higher rather than lower probabilities compared to the previous period (see change from Fig. 4d to Fig. 4e), demonstrating that the IFS-based model predicts actual weather developments.

A key thing to note is that the IFS forecasts are sharper than the forecasts by the climatology-based model, meaning that they produce more probabilities that are either close to 0 or to 1. This is particularly evident when comparing Fig. 4d to Fig. 3d. As mentioned in Sect. 2.6, sharpness is a desirable trait in a forecast model.

To better understand the nuances of the models' skill, reliability, sharpness and discrimination between hits and false alarms, we now perform an investigation of their aggregated performance over the whole reforecast period.

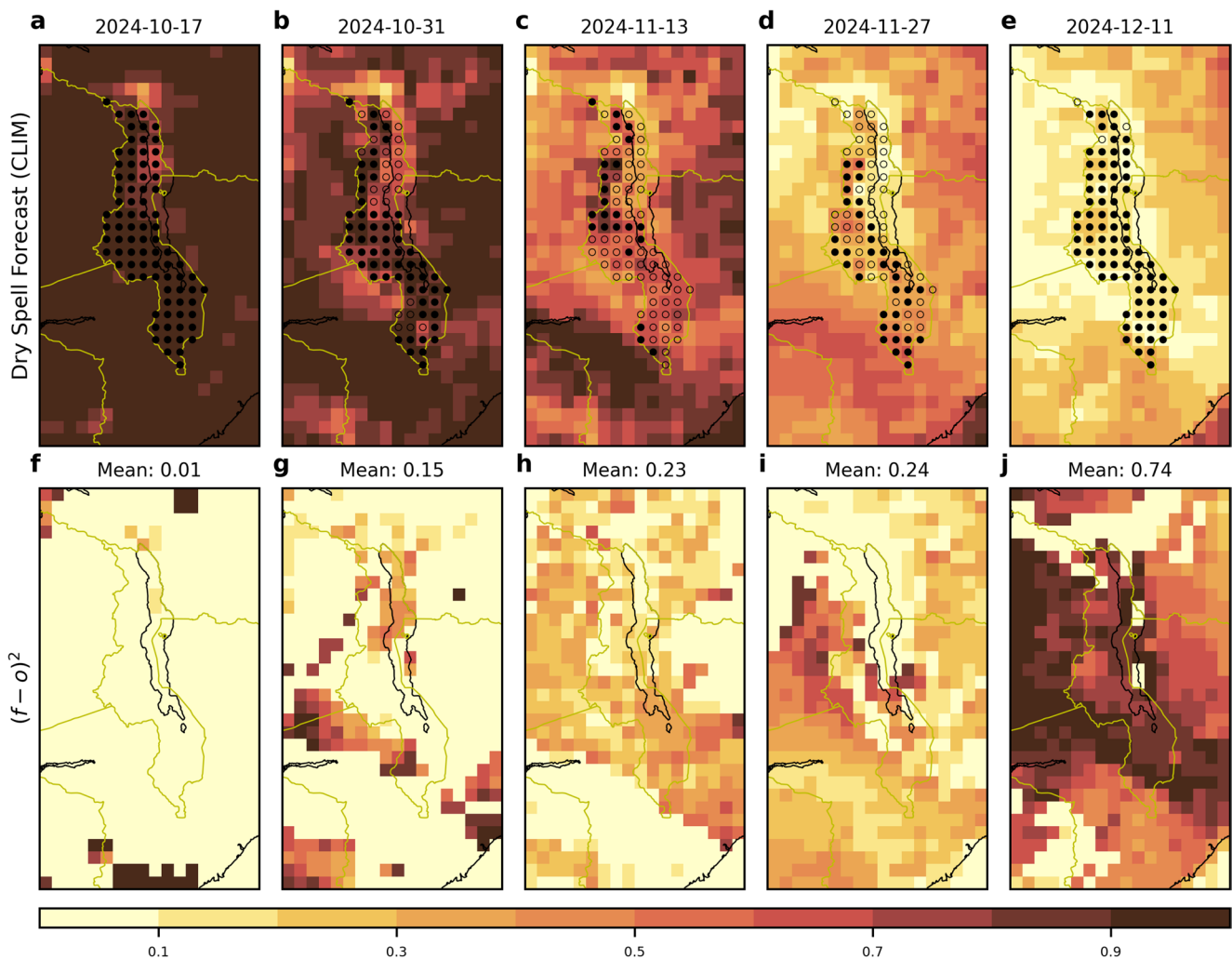


Fig. 3 Top row: A spatial representation of the climatology-based dry spell forecast for initial times during the 2024 season. The colours represent the probability for at least one seven-day dry spell during the three weeks following the initial date. Filled circles indicate grid

points in Malawi where at least one dry spell was observed according to ERA5, and open circles denote locations where there were no dry spells. Bottom row: Grid point values of $(f - o)^2$ for the forecasts in the row above. Yellow lines show country borders

3.2.2 Forecast performance

Figure 5a presents a map of the Brier Skill Score (BSS) at each grid point, derived from the full reforecast dataset. The average BSS within Malawi's borders is 0.26, with a 95% CI of [0.18, 0.35] (estimated through bootstrapping; see Sect. 2.7). As the CI does not include zero, the aggregated BSS is significant at the 5% level. At the grid level, the BSS for most points in Malawi is significant. This confirms that the IFS model's prediction of actual weather conditions, which Sect. 3.2.1 demonstrated anecdotally as favourable compared to the climatology-based model's gradual adaptation to seasonal dry spell patterns, translates into an overall higher skill.

Figure 5a does not yield information about the variability of the BSS within the season. To study this, we show the base rate of dry spells and the BSS for five bins, organised

by the initial date, in Fig. 5b. The first pair of bars show results for all the initial dates, and the bars to the right of the dashed vertical dividing line show data for different segments of the season. The first orange bar shows that the mean base rate is 47%, and the first blue bar repeats the average BSS of 0.26 indicated in the caption of Fig. 5a.

The blue bars to the right of the dashed line reveal fluctuations in the BSS throughout the season, but it remains significantly positive from start to end. Although not shown in the figure, the Brier Scores in the first and last of these bins are very low for both models: around 0.06 for both in the first bin, and 0.10 for IFS versus 0.15 for climatology in the last bin. This supports the point made in Sect. 2.6, namely that the Brier Score is not a useful metric for common or rare events. The low scores arise because dry spells are common at the beginning of the season and rare when the rainy season has taken hold (the base rate is 94% in the

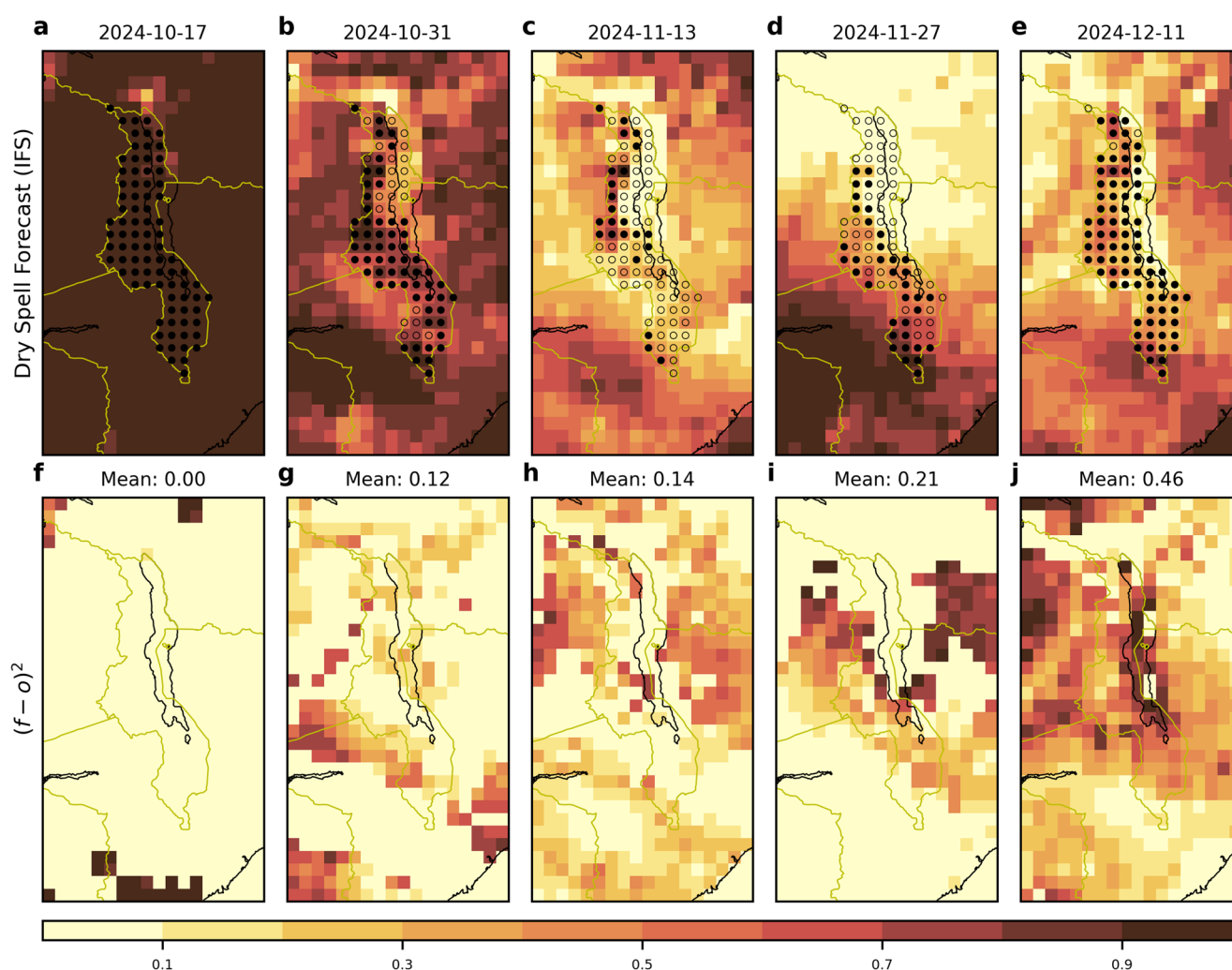


Fig. 4 As Fig. 3, but for the IFS-based prediction model

first bin to the right of the dividing line and 18% in the last bin). During these periods dry spell probabilities are usually easy to predict, barring unseasonal weather fluctuations. The superiority of the IFS-based model lies in its ability to predict many of the departures from the average seasonal cycle.

The two middle bins right of the vertical divider in Fig. 5b correspond to a part of the season where dry spells are neither very frequent nor very rare. These are arguably the most critical periods, as farmers are typically considering whether to plant but remain uncertain about the risk of dry spells. It is encouraging that the IFS-based model outperforms the climatology-based model during these key windows of decision-making.

A point of interest not addressed in Fig. 5 is whether the skill of the model varies with lead time. To investigate this, we evaluated the BSS for dry spell forecasts covering two- and four-week periods after the initial date of the forecasts and compared these scores to the BSS for the standard

three-week window (0.26). To account for the variations in BSS across the rainy season, we ensured that the forecast evaluation period consisted of exactly the same part of the season as the period used to evaluate the three-week window. This was done by including hindcast initial dates up to 22 November for the two-week window and up to 8 November for the four-week window. As expected, the BSS is higher (0.30; 95% CI: [0.23, 0.36]) for the two-week window and lower for the four-week window (0.23; 95% CI: [0.15, 0.32]). These variations are entirely attributable to differences in the Brier Score of the IFS-based model, as the Brier Score of the climatological model does not vary with lead time.

An arguably more robust way to assess how forecast skill varies with lead time is to compare the BSS for two-week windows positioned further from the forecast initial date than those evaluated previously. All earlier windows began at the initial date; the BSS of 0.30 cited above corresponds to the two-week period spanning days 1–14 (weeks

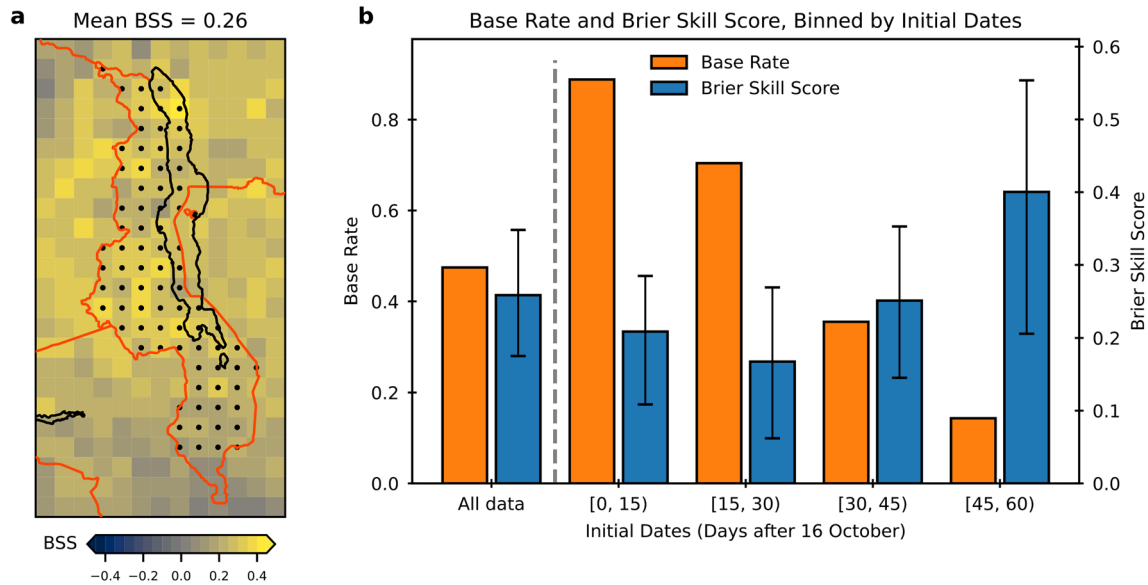
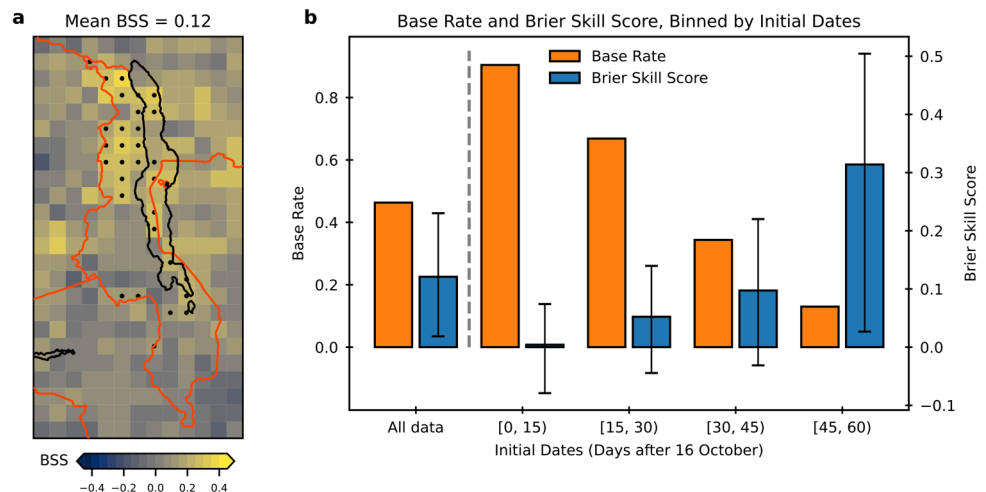


Fig. 5 (a) Brier Skill Score (BSS) for Malawi, aggregated for all initial dates between 15 October and 15 December. Dots indicate BSS values within Malawi that are significantly different to zero at the 5% level according to a bootstrapping test. Yellow lines show country borders. (b) Orange bars: the climatological base rate of dry spells for 15-day

bins relative to the initial dates of the model runs. Blue bars: BSS for IFS-based vs. climatology-based forecasts for the same bins. The first bars show the base rate and BSS for all the data, and whiskers show 95% CIs based on bootstrapping

Fig. 6 As Fig. 5, but for the IFS-based model bias-corrected against IMERG instead of ERA5, using a no-rain threshold of 1 mm. This model was validated against IMERG data



1–2). We now evaluate forecasts for weeks 2–3 and weeks 3–4. For this extended analysis, we also adjusted the set of hindcast initial dates to ensure that the evaluation periods matched the original analysis window exactly. The resulting mean BSS over Malawi for weeks 2–3 was found to be 0.08, with a 95% CI of [0.02, 0.15], demonstrating that the forecast skill for these weeks remains significantly positive. For weeks 3–4, however, the BSS dropped to 0.02 (95% CI: [−0.03, 0.06]), which is not significant. These results confirm that most of the skill in the three-week forecasts stem from the first two weeks after initialisation.

In the remainder of the analysis, we revert to studying three-week forecasts for days 1–21 after the initial date.

Figure 6 shows the BSS for an IFS-based model bias-corrected against IMERG, compared to a model based on IMERG climatology. Recalling from Sect. 2.3 the large corrections needed when adjusting the biases in IFS relative to IMERG, we note that the climatological model has an advantage compared to the IFS-based model in that it is better calibrated to the IMERG data. Still, even with this handicap, the IFS-based model performs better than the climatological model: Figure 6a and the first bin in Fig. 6b show that the average BSS value in Malawi is 0.12 (95% CI: [0.02, 0.23]). This aggregated BSS is significant, but at the grid-point level it is only in the northern part of the country that the BSS is mainly and significantly positive.

The within-season breakdown in Fig. 6b shows that the IFS-based model performs steadily better once the rainy season gets underway and the base rate of dry spells drops. It is only in the last bin that the BSS is significantly positive. These results contrast with those based on ERA5, where the BSS is higher and significant in all the bins (cf. Fig. 5), reflecting the consistency between forecast model and reference dataset. Still, the performance of the IMERG-calibrated model is noteworthy given that the climatological benchmark is derived directly from IMERG and thus perfectly calibrated. We repeat here that we did not perform a thorough optimisation of the interpolation of the IMERG data to the forecast model's grid, nor did we apply a sophisticated bias-correction; further optimisations might yield better skill than the model presented in Fig. 6.

We now return to the IFS-based model that was bias-corrected using ERA5 and the ERA5-based climatological model. Figure 7a displays a reliability diagram for these models. This diagram illustrates the relationship between forecast probabilities and observed frequencies. The climatology-based model is generally well-calibrated, which is a trivial result. It is more interesting that the IFS-based model also exhibits good calibration, with points generally closely aligned to the diagonal. However, forecast probabilities up to 0.6 tend to slightly but systematically underestimate the observed frequency: a sign of underconfidence. Conversely, the model overestimates frequencies above 0.6; in this segment the model is mildly overconfident.

In the forecast examples shown in Fig. 3 and 4, the IFS-based model was noticeably sharper than the climatology-based model. Figure 7b confirms that this holds in the aggregate, highlighting pronounced differences between the models. The climatology-based model exhibits poor sharpness, assigning the lowest probabilities (0–0.1) only

about 10% of the time – similar to what a randomised model would do over time. Although it uses the highest probability bin marginally more frequently, its overall use remains limited. This conservative tendency stems from the model's reliance on historical averages. In contrast, the IFS-based model demonstrates markedly greater sharpness; it issues forecasts in one of the two most extreme probability bins about 60% of the time. Compared to the climatology-based model, these high-stakes forecasts are more likely to spur action, increasing their potential value in decision-making contexts.

Figure 7c shows that the ROC curve for the IFS-based model consistently lies above that of the climatology-based model. The AUC of the former is 0.91 (95% CI: [0.89, 0.92]) compared to 0.84 for the latter model. In other words, the AUC of the IFS-based model is significantly higher than the AUC of the climatological model. This superior performance of the IFS-based model reflects another desirable forecasting property: the ability to achieve high detection rates while maintaining low false alarm rates, thereby reducing both missed events and unnecessary warnings.

3.3 Applicability to other regions

This section includes a short, preliminary analysis of the usefulness of the dry spell forecast model in other regions than Malawi. We choose to focus on East Africa and the March–May “long rains” (Camberlin and Philippon 2002), and the October–December “short rains” (Palmer et al. 2023). For the long rains we used all initial dates between 1 March and 10 May 2024, and for the short rains we used all initial dates between 1 October and 10 December 2024. Figure 8 presents the BSS for these seasons.

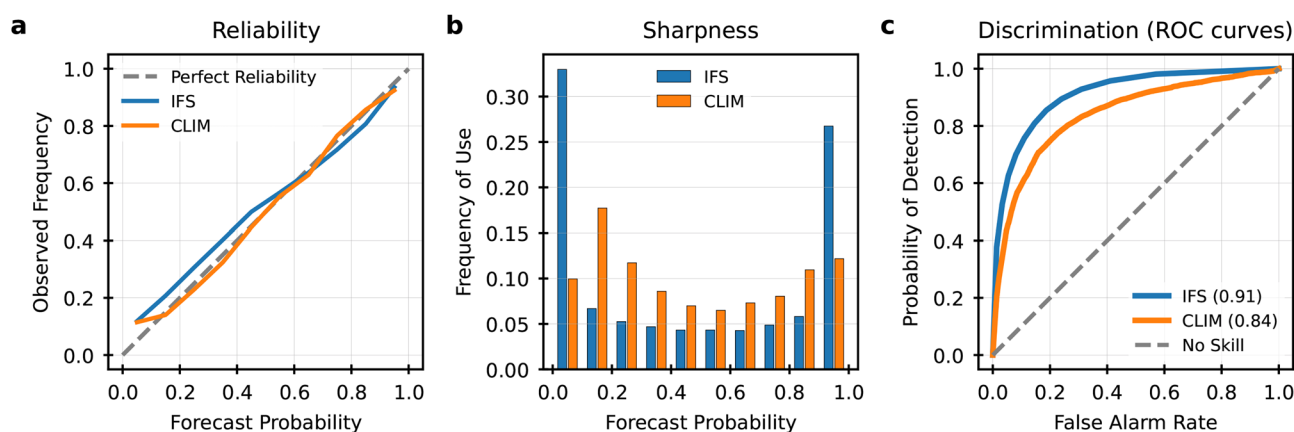


Fig. 7 (a) Reliability diagram comparing the IFS-based and climatology-based forecast models, showing observed frequency versus forecast probability across bins of [0, 0.1), [0.1, 0.2), and so on. (b) Sharpness diagram illustrating the distribution of forecast probabilities for both models within the same probability bins, highlighting the

tendency of each model to issue forecasts across the probability range. (c) ROC curves for the two models, depicting the trade-off between hit rate and false alarm rate, with the area under the curve (AUC; indicated in parentheses in the legend) indicating overall discriminatory skill

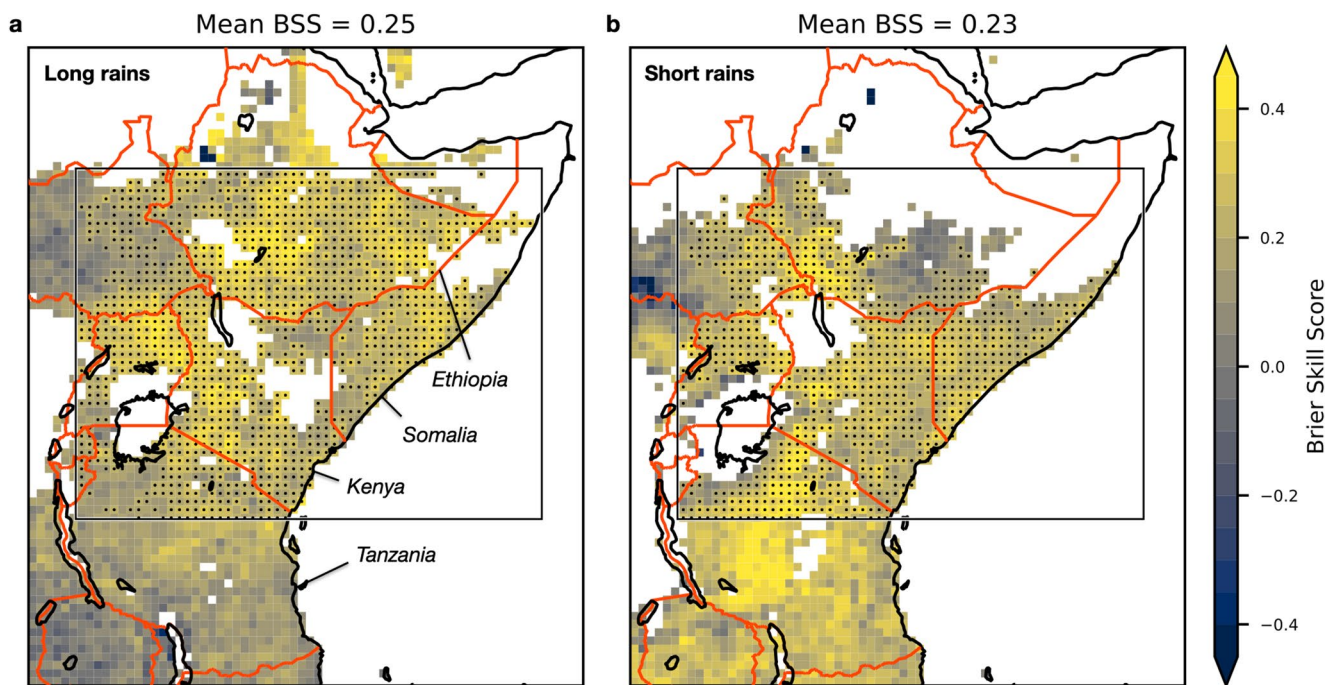


Fig. 8 As Fig. 5a, but for the East African long rains (a) and short rains (b). Grid points with a base rate of less than 0.15 or greater than 0.85 were masked in both maps. The mean scores in the captions were

As discussed previously, the Brier Score is unsuitable for events that are either very rare or frequent. Consequently, in Fig. 8 we masked grid points where the average climatological base rate was lower than 15% or higher than 85%. For comparison, the base rate for Malawi, averaged across all initial dates and grid points, was 47% (see first orange bar in Fig. 5b).

Figure 8 demonstrates that the significant skill of the dry spell forecast model extends beyond Malawi, covering large parts of East Africa during both the long and short rains. The mean BSS values for non-masked grid points within the East African reference region are comparable to those observed over Malawi. This suggests that a properly bias-corrected and calibrated prediction model could be developed for this region.

4 Discussion

Our study addresses three key questions: (1) How do dry spells vary within a season and from year to year in Malawi? (2) Can dry spells around the start of the rainy season be skilfully predicted? (3) Are our findings applicable beyond Malawi to other parts of Sub-Saharan Africa?

In response to the first question, we showed that dry spells in Malawi exhibit pronounced seasonality, with the highest frequency in early October, followed by a gradual decline through November and December. While early-season dry

spells are more common, their agricultural impact intensifies later in the season when farmers start committing resources to planting.

A critical challenge in agricultural planning is distinguishing between genuine rainy season onset and false onset, i.e. brief rainfall followed by detrimental dry spells. Traditional onset definitions typically combine rainfall thresholds with dry spell absence requirements (Fitzpatrick et al. 2015). Our findings suggest improved outcomes when separating rainfall onset occurrence analysis from dry spell risk assessment. The limited forecast horizon of a maximum of 3–4 weeks makes it difficult to predict both the initial onset (based on rainfall thresholds) and the subsequent risk of dry spells with sufficient lead time for decision-making. Our approach of considering seasonally varying dry spell prediction separately from the rainfall onset criterion gets away from this mismatch of timescales.

Second, the IFS-based model demonstrates significant forecast skill at a three-week lead time from mid-October onwards, whether validated against reanalysis or against IMERG data. By dynamically incorporating real-time atmospheric conditions, the IFS-based model consistently outperforms a simple climatology-based model across multiple verification metrics, including Brier Score, forecast sharpness, and discriminatory skill. This superior performance supports potential operational deployment of the model for mitigating risks associated with false onsets.

It is important to emphasise that most of the skill appears to stem from the first two weeks or so after the initial time. Although the skill was found to be significant for weeks 2–3, the model showed no significant skill during weeks 3–4. This degradation highlights a familiar forecaster's dilemma. There is likely a strong desire for forecasts that go beyond two or three weeks, but should the forecaster oblige and provide these forecasts? A mitigating factor is that the IFS-based model still performs as well as the climatological model at longer lead times, and climatology is not a poor predictor in this context. In our opinion there is no right or wrong solution to this dilemma. What matters is that the limitations and decline in forecast skill with lead time are communicated clearly and transparently to users.

The fact that the IFS-based model retains positive skill even when calibrated against and validated using IMERG is a promising result. IMERG is independent of the forecast model, and yet a basic quantile-based bias correction yields a forecast model that outperforms climatology. This is despite the fact that the IFS forecasts typically require large adjustments (e.g., the 1 mm threshold in IMERG corresponds to about 3 mm in IFS). It is likely that more sophisticated bias correction methods could further enhance performance.

The proven skill of subseasonal dry spell forecasts positions them as a crucial bridge between seasonal outlooks, which provide broad probabilistic guidance, and short-term weather forecasts, which lack lead time for agricultural planning. While seasonal forecasts offer insights into rainfall anomalies (e.g. ENSO-driven patterns; Nicholson 2017), they lack the accuracy for onset characterisation. This underscores a need for an integrated approach that blends seasonal outlooks with subseasonal updates, real-time monitoring and improved communication strategies. Farmers should not only receive probabilistic onset forecasts at the start of the season; they should also get updates on the evolving progress of the rains. The need for such updates was stressed by Streefkerk (2020) and might ensure that early planting decisions remain flexible in the face of forecast uncertainty. Integration of subseasonal forecasts into the portfolio of DCCMS could enhance trust in forecasts and support the design of confident decision-making.

Third, our findings indicate that dry spells are predictable during the main East African rainy seasons as well as in Malawi. However, since forecast skill is likely modulated by local land-atmosphere interactions and regional circulation patterns that vary significantly across the continent, region-specific verification studies are needed before generalising our findings to other agricultural regions.

A caveat of our work is that it did not use actual observations, relying instead on reanalysis data, which are known to be biased towards producing too much light rain. Our

study should be viewed as a proof of concept, demonstrating the feasibility of IFS-based subseasonal dry spell forecasting using a physically consistent reanalysis dataset as a benchmark. Future work should focus on integrating observational data to improve site-specific reliability and enhance the practical applicability of subseasonal dry spell forecasts.

Another suggestion for future studies is to address key gaps to enhance the practical value of subseasonal forecasts. Investigating the differential impact of predictable climate modes, such as the Madden-Julian Oscillation (MJO), which has been linked to enhanced forecast skill for rainfall in many regions (e.g. Vitart 2014; Lim et al. 2018; Nsubuga et al. 2021), could facilitate the development of hybrid forecast approaches, combining dynamical models with statistical post-processing to address systematic biases. Furthermore, quantifying the economic value of subseasonal forecasts using robust methods such as randomised controlled trials would be valuable. Such experiments are currently underway in Malawi and Ethiopia as part of our ARCS project. Complementing this quantitative approach, qualitative research should explore the forecast value across different agricultural decision contexts (planting timing, variety selection, input application).

By demonstrating the feasibility of subseasonal dry spell forecasting in Malawi and East Africa, this study provides a foundation for integrating probabilistic forecasts into agricultural decision-making. For operational use, subseasonal forecasts should be tailored to local decision-making needs, requiring investment in institutional capacity, communication strategies, and stakeholder engagement. A crucial prerequisite is to strengthen institutional capacity within national meteorological services to interpret and customise subseasonal model outputs. Effective implementation depends on co-production with local stakeholders, ensuring forecasts complement traditional knowledge systems and existing advisory services. This collaborative approach would enhance usability, accessibility, and trust among end users (Streefkerk et al. 2022).

Author Contributions All authors contributed to the study conception and design. The data analysis was performed by Erik W. Kolstad. The first draft of the manuscript was written by Erik W. Kolstad and all authors provided comments and text. All authors read and approved the final manuscript.

Funding Open access funding provided by NORCE Research AS. The authors acknowledge funding from NORAD through the ARCS project (grant RAF:23/0006) and the European Union through its Horizon Europe programme (project ACACIA; grant 101137847).

Data Availability The ERA5 reanalysis and the IFS forecast and reforecast data are available from ECMWF. IMERG data are available from NASA.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adam JC, Lettenmaier DP (2003) Adjustment of global gridded precipitation for systematic bias. *Journal of Geophysical Research Atmospheres* 108(D9):4257. <https://doi.org/10.1029/2002JD002499>
- Ahn M-S, Ullrich PA, Lee J, Gleckler PJ, Ma H-Y, Terai CR, Bogen-schutz PA, Ordóñez AC (2024) Bimodality in simulated precipitation frequency distributions and its relationship with convective parameterizations. *npj Climate and Atmospheric Science* 7(1):132
- Bandhauer M, Isotta F, Lakatos M, Lussana C, Båserud L, Izsák B, Szentes O, Tveito OE, Frei C (2022) Evaluation of daily precipitation analyses in E-OBS (v19. 0e) and ERA5 by comparison to regional high-resolution datasets in European regions. *Int J Climatol* 42(2):727–747
- Barron J, Rockström J, Gichuki F, Hatibu N (2003) Dry spell analysis and maize yields for two semi-arid locations in east Africa. *Agric For Meteorol* 117(1–2):23–37
- Brier GW (1950) Verification of forecasts expressed in terms of probability. *Mon Weather Rev* 78(1):1–3
- Bröcker J, Smith LA (2007) Scoring probabilistic forecasts: On the importance of being proper. *Weather Forecast* 22(2):382–391. <https://doi.org/10.1175/WAF966.1>
- Coulibaly YJ, Kundhlande G, Tall A, Kaur H, Hansen J (2015) Which climate services do farmers and pastoralists need in malawi? baseline study for the gfcs adaptation program in africa. CCAFS Working Paper 112, CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS), Copenhagen, Denmark
- Chimimba EG, Ngongondo C, Li C, Minoungou B, Monjerezi M, Eneya L (2023) Characterisation of dry spells for agricultural applications in Malawi. *SN Applied Sciences* 5(7):199. <https://doi.org/10.1007/s42452-023-05413-9>
- Camberlin P, Philippon N (2002) The east african march-may rainy season: Associated atmospheric dynamics and predictability over the 1968–97 period. *J Clim* 15(9):1002–1019.
- Cramer F, Shephard GE, Heron PJ (2020) The misuse of colour in science communication. *Nat Commun* 11(1):5444. <https://doi.org/10.1038/s41467-020-19160-7>
- Department of Climate Change and Meteorological Services (2023) The 2023–2024 seasonal climate outlook. Technical report, Department of Climate Change and Meteorological Services, Malawi (September)
- Department of Climate Change and Meteorological Services (2024) The 2024–2025 seasonal climate outlook. Technical report, Department of Climate Change and Meteorological Services, Malawi (September)
- Demissie T, Gebrechorkos SH (2024) Spatio-temporal trends in precipitation, temperature, and extremes: A study of Malawi and Zambia (1981–2021). *Sustainability*. <https://doi.org/10.3390/su16103885>
- de Andrade FM, Young MP, MacLeod D, Hiron LC, Woolnough SJ, Black E (2021) Subseasonal precipitation prediction for Africa: Forecast evaluation and sources of predictability. *Weather Forecast* 36(1):265–284. <https://doi.org/10.1175/WAF-D-20-0054.1>
- Fitzpatrick RG, Bain CL, Knippertz P, Marsham JH, Parker DJ (2015) The West African monsoon onset: A concise comparison of definitions. *J Clim* 28(22):8673–8694
- Fitzpatrick RG, Bain CL, Knippertz P, Marsham JH, Parker DJ (2016) On what scale can we predict the agronomic onset of the West African Monsoon? *Mon Weather Rev* 144(4):1571–1589
- Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, Husak G, Rowland J, Harrison L, Hoell A et al (2015) The climate hazards infrared precipitation with stations-a new environmental record for monitoring extremes. *Scientific data* 2(1):1–21
- Goddard L, Baethgen WE, Bhojwani H, Robertson AW (2014) The International Research Institute for climate & society: why, what and how. *Earth Perspectives* 1:1–14. <https://doi.org/10.1186/2194-6434-1-10>
- Gomis-Cebolla J, Rattayova V, Salazar-Galán S, Francés F (2023) Evaluation of ERA5 and ERA5-Land reanalysis precipitation datasets over Spain (1951–2020). *Atmos Res* 284:106606
- Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D et al (2020) The era5 global reanalysis. *Q J R Meteorol Soc* 146(730):1999–2049. <https://doi.org/10.1002/qj.3803>
- Haghtalab N, Moore N, Ngongondo C (2019) Spatio-temporal analysis of rainfall variability and seasonality in Malawi. *Reg Environ Change* 19:2041–2054. <https://doi.org/10.1007/s10113-019-01535-2>
- Hansen JW, Mason SJ, Sun L, Tall A (2011) Review of seasonal climate forecasting for agriculture in sub-Saharan Africa. *Exp Agric* 47(2):205–240. <https://doi.org/10.1017/S0014479710000876>
- Hermanson L, Ren H-L, Vellinga M, Dunstone ND, Hyder P, Ineson S, Scaife AA, Smith DM, Thompson V, Tian B, Williams KD (2021) Different types of drifts in two seasonal forecast systems and their dependence on ENSO. *Clim Dyn*. <https://doi.org/10.1007/s00382-017-3962-9>
- Huffman GJ, Stocker EF, Bolvin DT, Nelkin EJ, Tan J (2023) GPM IMERG Final Precipitation L3 1 day 0.1 degree x 0.1 degree V07. NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC). <https://doi.org/10.5067/GPM/IMERGDF/DAY/07>
- Hirons L, Wainwright CM, Nying'uro P, Quayle D, Ashong J, Kiptum C, Opoku NK, Thompson EM, Lamptey B (2023) Experiences of co-producing sub-seasonal forecast products for agricultural application in Kenya and Ghana. *Weather* 78(5):148–153. <https://doi.org/10.1002/wea.4381>
- Kolstad EW, MacLeod D (2022) Lagged oceanic effects on the East African short rains. *Clim Dyn* 59(3):1043–1056. <https://doi.org/10.1007/s00382-022-06176-6>
- Kolstad EW, MacLeod D, Demissie TD (2021) Drivers of subseasonal forecast errors of the East African short rains. *Geophys Res Lett* 48(14):2021–093292. <https://doi.org/10.1029/2021GL093292>
- Lavers DA, Harrigan S, Prudhomme C (2021) Precipitation biases in the ECMWF integrated forecasting system. *J Hydrometeorol* 22(5):1187–1198
- Li X, O S, Wang N, Liu I, Huang Y, (2021) Evaluation of the GPM IMERG v06 products for light rain over mainland China.

- Atmospheric Research 253. <https://doi.org/10.1016/j.atmosres.2021.105510>
- Lawson JR, Potvin CK, Nelson K (2024) Decoding the atmosphere: Optimising probabilistic forecasts with information gain. *Meteorology* 3(2):212–231. <https://doi.org/10.3390/meteorology3020010>
- Lim Y, Son S-W, Kim D (2018) MJO prediction skill of the subseasonal-to-seasonal prediction models. *J Clim* 31(10):4075–4094. <https://doi.org/10.1175/JCLI-D-17-0545.1>
- Lavers DA, Simmons A, Vamborg F, Rodwell MJ (2022) An evaluation of ERA5 precipitation for climate monitoring. *Q J R Meteorol Soc* 148(748):3152–3165
- Müller WA, Appenzeller C, Doblas-Reyes FJ, Liniger MA (2005) A debiased ranked probability skill score to evaluate probabilistic ensemble forecasts with small ensemble sizes. *J Clim* 18(10):1513–1523. <https://doi.org/10.1175/JCLI3361.1>
- MacLeod DA, Dankers R, Graham R, Guigma K, Jenkins L, Todd MC, Kiptum A, Kilavi M, Njogu A, Mwangi E (2021) Drivers and subseasonal predictability of heavy rainfall in equatorial east africa and relationship with flood risk. *J Hydrometeorol* 22(4):887–903
- Maidment RI, Grimes D, Black E, Tamavsky E, Young M, Greatrex H, Allan RP, Stein T, Nkonde E, Senkunda S et al (2017) A new, long-term daily satellite-based rainfall dataset for operational monitoring in africa. *Scientific data* 4(1):1–19
- Mkwambisi DD, Jew EKK, Dougill AJ (2020) Farmer preparedness for building resilient agri-food systems: Lessons from the 2015/2016 El Niño drought in Malawi. *Frontiers in Climate* 2. <https://doi.org/10.3389/fclim.2020.584245>
- Mkwambisi DD, Jew EK, Dougill AJ (2021) Farmer preparedness for building resilient agri-food systems: Lessons from the 2015/2016 El Niño drought in Malawi. *Frontiers in Climate* 2:584245. <https://doi.org/10.3389/fclim.2020.584245>
- Marongwe LS, Kwazira K, Jenrich M, Thierfelder C, Kassam A, Friedrich T (2011) An African success: the case of conservation agriculture in Zimbabwe. *Int J Agric Sustain* 9(1):153–161. <https://doi.org/10.3763/ijas.2010.0556>
- Mittal N, Pope E, Whitfield S, Bacon J, Bruno Soares M, Dougill AJ, Homberg M.v.d, Walker D.P., Vanya C.L., Tibu A, et al (2021) Co-designing indices for tailored seasonal climate forecasts in Malawi. *Frontiers in Climate* 2:578553
- Nsubuga FW, Berhane F, Beraki A (2021) The impact of the madden-julian oscillation on southern african rainfall predictability. *Clim Dyn* 57:2071–2089. <https://doi.org/10.1007/s00382-021-05809-5>
- Nicholson SE (2017) Climate and climatic variability of rainfall over eastern Africa. *Rev Geophys* 55(3):590–635. <https://doi.org/10.1002/2016RG000544>
- Ndlovu J, Nyathi D, Ndlovu CN (2024) Climate change vulnerability, impact, and adaptive capacity of rural populace in Malawi: A systematic literature review. *African Renaissance* 2024:217–242
- Palmer PI, Wainwright CM, Dong B, Maidment RI, Wheeler KG, Gedney N, Hickman JE, Madani N, Folwell SS, Abdo G et al (2023) Drivers and impacts of eastern african rainfall variability. *Nature Reviews Earth & Environment* 4(4):254–270. <https://doi.org/10.1038/s43017-023-00397-x>
- Ratnam J, Behera S, Masumoto Y, Yamagata T (2014) Remote effects of El Niño and Modoki events on the austral summer precipitation of southern Africa. *J Clim* 27(10):3802–3815. <https://doi.org/10.1175/JCLI-D-13-00431.1>
- Rockström J, Karlberg L, Wani SP, Barron J, Hatibu N, Oweis T, Bruggeman A, Farahani J, Qiang Z (2010) Managing water in rainfed agriculture-the need for a paradigm shift. *Agric Water Manag* 97(4):543–550. <https://doi.org/10.1016/j.agwat.2009.09.009>
- Roudier P, Muller B, D'Aquino P, Roncoli C, Soumaré MA, Batté L, Sultan B (2014) The role of climate forecasts in smallholder agriculture: Lessons from participatory research in two communities in senegal. *Clim Risk Manag* 2:42–55. <https://doi.org/10.1016/j.crm.2014.02.001>
- ROSEA O, (2024) Malawi drought flash appeal: July 2024 - April 2025. Technical report, United Nations Office for the Coordination of Humanitarian Affairs (OCHA) (July)
- Sharma T (1996) Simulation of the Kenyan longest dry and wet spells and the largest rain-sums using a Markov model. *J Hydrol* 178(1–4):55–67
- Sivakumar M (1992) Empirical analysis of dry spells for agricultural applications in West Africa. *J Clim* 5(5):532–539
- Streefkerk I (2020) Linking drought forecast information to smallholder farmer's agricultural strategies and local knowledge in southern Malawi. Master's thesis, Delft University of Technology, Netherlands
- Streefkerk IN, Homberg MJ, Whitfield S, Mittal N, Pope E, Werner M, Winsemius HC, Comes T, Ertsen MW (2022) Contextualising seasonal climate forecasts by integrating local knowledge on drought in malawi. *Climate Services* 25:100268
- Thoithi W, Blamey RC, Reason CJC (2021) Dry spells, wet days, and their trends across southern Africa during the summer rainy season. *Geophys Res Lett* 48(5):2020–091041. <https://doi.org/10.1029/2020GL091041>
- Vitart F (2014) Monthly forecasting at ecmwf. *Mon Weather Rev* 142(2):516–527. <https://doi.org/10.1175/MWR-D-13-00076.1>
- Vitart F, Robertson AW (2019) Chapter 1 - introduction: Why subseasonal to seasonal prediction (S2S)? In: Robertson AW, Vitart F (eds) *Sub-Seasonal to Seasonal Prediction*. Elsevier, Amsterdam, pp 3–15. <https://doi.org/10.1016/B978-0-12-811714-9.00001-2>
- White CJ, Carlsen H, Robertson AW, Klein RJT (2017) Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorol Appl* 24(3):315–325. <https://doi.org/10.1002/met.1654>
- White CJ, Domeisen DIV, Acharya N.e.a, (2022) Advances in the application and utility of subseasonal-to-seasonal predictions. *Bulletin of the American Meteorological Society* <https://doi.org/10.1175/BAMS-D-20-0224.1>
- Wei D, Di W, Tian W, Cheng S, Xie H, Xie L, Jing Z (2025) Assessment of integrated multi-satellite retrievals for global precipitation measurement (IMERG) precipitation products in northwest China. *Remote Sensing*. <https://doi.org/10.3390/rs17081364>
- Weigel AP (2011) Ensemble forecasts. In: Jolliffe IT, Stephenson DB (eds) *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Wiley, Chichester, UK, pp 141–166
- Wilks DS (2019) *Statistical Methods in the Atmospheric Sciences*, 4th edn. Elsevier, Amsterdam
- Young M, Heinrich V, Emily B, Asfaw D (2020) Optimal spatial scales for seasonal forecasts over africa. *Env Res Lett* 15:094023
- Yang M, Liu G, Chen T, Chen Y, Xia C (2020) Evaluation of GPM IMERG precipitation products with the point rain gauge records over Sichuan, China. *Atmospheric Research*, p 246. <https://doi.org/10.1016/j.atmosres.2020.105101>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.