Received: 1 February 2024 | Revised: 7 June 2024 | Accepted: 17 July 2024

DOI: 10.1002/qj.4829

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Beyond the regional average: Drivers of geographical rainfall variability during East Africa's short rains

Erik W. Kolstad[1] | **Douglas J. Parker**[1,2] | **David A. MacLeod**[3] | **Caroline M. Wainwright**[3] | **Linda C. Hirons**[4]

[1]NORCE Norwegian Research Centre, Bjerknes Centre for Climate Research, Bergen, Norway

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

[3]School of Earth and Environmental Sciences, Cardiff University, Cardiff, UK

[4]National Centre for Atmospheric Science (NCAS), University of Reading, Reading, UK

Correspondence
Erik W. Kolstad, Bjerknes Centre for Climate Research, Jahnebakken 5, 5007 Bergen, Norway.
Email: ekol@norceresearch.no

Funding information
Natural Environment Research Council, Grant/Award Number: NE/X006263/1; Horizon 2020 Framework Programme, Grant/Award Numbers: 869550, 869730

Abstract

The East African "short rains" from October–December (OND) are crucial for the region's cultural and agricultural landscape. Traditional climate studies have often treated these rains as a single mode, representing the average rainfall across the region. This approach, however, fails to capture the complex geographical variations in seasonal rainfall. In our study, we analyse 4200 reforecasts from a seasonal prediction system spanning 1981–2022, identifying distinct clusters that represent different geographical patterns of the short rains. We explore the influence of tropical sea-surface temperature patterns, upper-level tropospheric flow, and low-level moisture fluxes on these clusters. A key revelation of our research is the limited predictability of certain geographical rainfall structures based on large-scale climatic drivers. This finding highlights a gap in current forecasting methodologies, emphasising the necessity for further research to understand and predict these intricate patterns. Our study illuminates the complexities of regional rainfall variability in East Africa, underlining the importance of continued investigation to improve climate resilience strategies in the region.

KEYWORDS

atmosphere, ensembles, rainfall, seasonal, Tropics

1 | INTRODUCTION

The rainy season that occurs at the end of the year in East Africa, around October, November, and December (OND), is known as the "short rains" (Nicholson, 2015). In the northern part of the region, such as in Ethiopia and Somalia, the season usually starts earlier than in the southern part, including Kenya and northern Tanzania (Dunning *et al.*, 2016; Gudoshava *et al.*, 2022). Evidence of the cultural importance of this season is found in its local names; the short rains are known as "deyr" in Somalia, "bega" in Ethiopia, and "vuli" in Kenya. In light of this cultural backdrop, which reflects the importance of rain-dependent agriculture and pastoralism in the region, the climatic understanding of these rains becomes crucial.

East Africa is characterised by significant annual weather fluctuations, and the short rains stand out for their extreme interannual variability (Nicholson, 2017), often oscillating between contrasting conditions. For instance, the year 1996 experienced notably dry conditions

(Doi *et al.*, 2022), whereas 1997 saw exceptionally wet weather (Black *et al.*, 2003). This pattern of sharp contrast in rainfall is exemplified by similar occurrences in 2005/2006, 2010/2011, and 2018/2019, illustrating the region's highly variable climate.

In studies of the short rains, they are often represented as the geographical average of the precipitation in a large area, typically from Tanzania in the south to Ethiopia in the north, and between Uganda in the west and the tip of the Horn of Africa in the east. One reason for this simplification is that the regional average rainfall is tightly linked to predictable oceanic states not only in the nearby Indian Ocean, but much farther away in the tropical Pacific Ocean. There is a strong positive correlation between the geographical average of East African rainfall during OND and both the El Niño–Southern Oscillation, or ENSO (Indeje *et al.*, 2000) and the Indian Ocean Dipole, or IOD (Bahaga *et al.*, 2015; Bahaga *et al.*, 2019; Hirons & Turner, 2018; Kebacho, 2021; Wenhaji Ndomeni *et al.*, 2018). These correlations are significant even when ENSO and the IOD lead the rainfall by several months. However, Kolstad and MacLeod (2022) showed using mediation analysis (e.g., MacKinnon *et al.*, 2007) that the main effect of ENSO prior to the season (e.g., in August) is that it exerts a lagged influence on the IOD, which then influences the East African rainfall directly during the OND season (this mechanism has also been suggested by Black, 2005, and others).

The explanation for the strong relationship between the IOD and East African rainfall is that anomalies in the IOD are associated with anomalies in the Walker circulation over the Indian Ocean, which is characterised by subsidence over the western part and low-level westerlies over the central part of the basin (e.g., Pohl & Camberlin, 2011). There is a strong anticorrelation between the strength of this Walker circulation and East African rainfall during OND (Zhao & Cook, 2021). In other words, high seasonal rainfall tends to coincide with a weakening of this Walker circulation, which implies weakened subsidence or enhanced convection over the western Indian Ocean and East Africa.

Linear models can skilfully predict some aspects of the short rains based on the ENSO and IOD states before the start of the season. For instance, Kolstad and MacLeod (2022) demonstrated that such a model has skill in predicting the dominant mode of interannual variability of East African rainfall during OND, specifically the leading empirical orthogonal function (EOF), which is equivalent to the area-averaged rainfall anomalies across the region (Kolstad & MacLeod, 2022). However, their study, along with the subsequent study by MacLeod *et al.* (2024), found evidence of a strong asymmetry between large positive and negative area-averaged rainfall anomalies. In short, negative rainfall anomalies are bounded because the rainfall cannot be lower than zero, while positive anomalies increase nonlinearly with the strength of the positive IOD. This implies that East African rainfall should be investigated using methods that account for nonlinear effects.

An additional concern is that linear models describing the leading mode of variability (i.e., the regional average) do not provide information about geographical differences within the region. Recognising these regional contrasts is vital, as they reflect more than just deviations from the mean, but are key to understanding the true nature of East African climate dynamics. Moreover, the lives of East African citizens will be influenced by local rainfall anomalies more than regional ones, so the regional mean rainfall will not account for the experiences of every person or every country.

In this study, we explore three core ideas relating to the geographical rainfall distribution in East Africa and the mechanisms that dictate this distribution. The first of these ideas is that a high degree of geographical variability within the region becomes apparent when moving beyond the analysis of the regional average. We believe that it is possible to understand this rainfall variability as the projection of large-scale climate drivers, including circulation and moisture transport, on to the regional, continental dynamics. These projections are modulated by local topography, land cover (e.g., lakes), and proximity to the ocean, which lead to geographical differences in the response of the short rains. In particular, the Indian Ocean is the dominant source of moisture for rainfall over large parts of East Africa. For instance, Koppa *et al.* (2023) estimated that the Arabian Sea and the southern Indian Ocean contribute around 70%–80% of the rainfall over the Horn of Africa drylands during the short rains.

However, the Congo Basin can also be a significant moisture source for the western parts of East Africa (Diem *et al.*, 2019; Dyer *et al.*, 2017). A recent article by Finney *et al.* (2019) studied the effect of low-level westerly wind anomalies on equatorial East Africa. A key finding was that, for all months, days with stronger westerly anomalies experienced more rainfall in the region surrounding Lake Victoria (described by Finney et al. as an "expanded Lake Victoria Basin region") and depressed rainfall along the Kenya/Tanzania coastline, which is in a rain shadow in such cases. Conversely, days with strong easterlies experienced lower rainfall rates across the Lake Victoria Basin than days with weak easterlies (Finney *et al.*, 2019). The main driver of enhanced rainfall in westerly regimes is an increase in moisture advection from the Congo Basin. Figure 7a in Finney *et al.* (2019) provides a useful visualisation of these conditions.

The second core idea explored herein is that the patterns of moisture import into East Africa from the Indian Ocean and the Congo Basin are influenced, at least in part, by sea-surface temperature (SST) anomalies in the western Indian Ocean. In equatorial regions, strong oceanic heat anomalies lead to deep convective warming in the atmosphere and induce a Matsuno–Gill response (Gill, 1980; Matsuno, 1966), with anticyclonic gyres at upper levels both north and south of the Equator to the west of the area of enhanced convection and above-normal divergence aloft. A reversed pattern is then seen in the flow at lower levels; however, these lower levels are also more prone to influences from coastlines, mountains, and other surface features. In East Africa, the result is that positive IOD conditions lead to a Gill-type response that causes low-level westerly wind anomalies over equatorial Africa, moving the moisture-flux convergence zone from central equatorial Africa towards and over the East African region (Bahaga *et al.*, 2015; Liu *et al.*, 2020; Maybee *et al.*, 2023; Ummenhofer *et al.*, 2009; Wenhaji Ndomeni *et al.*, 2018). This Matsuno–Gill response will be examined for its influence on regional circulation and rainfall patterns to deepen our understanding of how it impacts rainfall in the area.

The third core idea motivating this study is that the location of SST anomalies in the western Indian Ocean is as crucial as their existence in influencing rainfall in East Africa. For instance, warm SST anomalies along the southern coast of East Africa may induce inland westerly anomalies south of the Equator, directed from the Congo Basin into Tanzania, whilst cold SST anomalies in the same location are likely to contribute to dry anomalies over land. When the SST anomalies are more widespread, such as during strongly negative or positive phases of the IOD, they are likely to be linked to regional scale rainfall anomalies inland.

To contribute to a more thorough understanding of the mechanisms that drive spatial rainfall differences within the East African region during the short rains, we present an analysis of various climatic variables and relate them to different rainfall distributions. Whilst most studies in this vein rely on observational or reanalysis data, the limited sample size of about 40 seasons since the start of the satellite era poses a challenge. One problem with data with a limited sample size is that the results may be unduly influenced by single data points or events. For the short rains, the exceptionally wet season in 2019 (Wainwright *et al.*, 2021) is an example of an event with such potential. Even with the extension of the ERA5 reanalysis (Hersbach *et al.*, 2020) back to 1940, the small sample size remains problematic (and compounded by the scarcity of reliable observational and satellite data before 1980).

To circumvent the limitations of small sample sizes in observational data, a string of recent studies have turned to analysing extensive collections of retroactive forecasts, known as *reforecasts* or *hindcasts*. These reforecasts, which include an ensemble of multiple simulations for each season, expand the sample size dramatically compared with traditional observational datasets. Similar approaches have been used effectively in other rainfall studies (Kelder *et al.*, 2020; Thompson *et al.*, 2017), including those focusing on East Africa (MacLeod *et al.*, 2024). We utilised reforecasts from the European Centre for Medium-Range Weather Forecasts' (ECMWF) seasonal prediction system SEAS5 (Johnson *et al.*, 2019), comprising 4200 simulations: 100 times the number of seasons in the study period.

To address the core ideas underpinning this study, we pursued three key objectives that aim to provide new insights into short-rains variability. Firstly, we evaluated whether SEAS5 captures the spatial variability of rainfall in East Africa accurately. This was achieved by comparing SEAS5 outputs with a dataset derived from observations and satellite imagery. Secondly, we explored the possibility of categorising SEAS5 samples into distinct, representative rainfall patterns. This involved contrasting a traditional EOF analysis with a clustering method, which accounts better for asymmetries between large positive and negative rainfall anomalies, as well as spatial gradients within East Africa. Finally, we examined the extent to which these patterns correlate with large-scale climatic drivers and local circulation anomalies. To this end, we analysed composite average anomalies in various diagnostic variables: SSTs, wind patterns, divergence, and moisture fluxes.

The article is organised as follows. The subsequent section details the datasets, clustering technique, and other methodologies employed in this study. Section 3 presents the results of our analyses, and a comprehensive discussion of these results follows in Section 4. Finally, the article is wrapped up with some concluding remarks in Section 5.

2 | DATA AND METHODS

2.1 | Data

Our study is based on reforecasts and forecasts from SEAS5, covering the period from 1981–2022. With an initial date at the beginning of each month in this period, a set of ensemble members was integrated for seven months. For the reforecast period (1981–2016), 25 ensemble members were produced each month, and for the forecast period (2017–2022) 51 members were produced. To ensure an equal contribution from each period, we only used the first 25 members from the forecast period. The strategy behind the ensemble generation is explained in Johnson *et al.* (2019).

We used monthly mean data and extracted the subset of data that covered the period between October and December. Following the convention that a lead time of one month corresponds to the month of model initialisation, we could use data from the June simulations for lead times of 5, 6, and 7 months. From the July simulations, we could use lead times from 4–6 months, and so on up to the October initialisations, from which we could use lead times of 1–3 months. However, we opted not to include the October-initialised runs. The background for this decision was the close relationship between the ocean surface and the mechanisms that drive rainfall in East Africa. Therefore, because the spread between the slowly varying oceanic fields is narrow at the start of each model run, the atmospheric fields in the model runs initialised in October are likely also to have a narrow spread. We desired a wide ensemble spread to gain as complete a picture as possible of plausible atmospheric behaviour.

Another potential issue relates to the model runs initialised several months before the OND season. Model drift over extended lead times has been observed in some seasonal forecast systems (Manzanas, 2020). Specifically, a report on SEAS5 highlighted notable biases during the 1981–2016 reforecast period (Stockdale *et al.*, 2018). To account for this drift issue, we calculated lead-time-dependent climatologies for each variable. When we refer to "anomalies" henceforth, we mean deviations relative to these climatologies.

To check whether the model drift influenced our results despite this precaution, we performed parts of the analysis using only June and July initialisations and compared these with a separate analysis based on August and September initialisations. The results were so similar that we decided it was acceptable to perform the analysis using the pooled set of forecasts and reforecasts initialised from June to September. Hence, the total number of samples available from the model runs initialised in June–September was 4200 OND seasons from 1981–2022 (4 start dates × 25 members × 42 years). These seasons were pooled and treated equally in the analysis.

From each of the model simulations, we used monthly means of total precipitation, divergence and wind components at 200 and 700 hPa, specific humidity at 700 hPa, and SST. The precipitation data were downloaded with a grid spacing of 0.5°, and the other data with a spacing of 1°.

Using the wind vector ($\boldsymbol{v}$) and the specific humidity (q) at 700 hPa, we computed the moisture-flux convergence (MFC) as $-\nabla \cdot (q\boldsymbol{v})$. We also split the MFC into an advection term ($-\boldsymbol{v} \cdot \nabla q$) and a convergence term ($-q\nabla \cdot \boldsymbol{v}$).

To validate the SEAS5 precipitation data, we used Climate Hazards group Infrared Precipitation with Stations (CHIRPS) data (Funk *et al.*, 2015) from 1981 to 2022. The grid spacing of the CHIRPS data is 0.05° (5–6 km at the Equator), and the dataset only provides data over land.

As a further validation, we compared the climatologies of the seasonal mean atmospheric circulation over East Africa and the Indian Ocean with data from the ERA5 reanalysis, downloaded with the same horizontal grid spacing as the SEAS5 data.

Although several of the variables, such as SST and rainfall, have significant trends over the course of the study period, some of these trends differ widely from one location to the next. This makes it a complicated exercise to detrend the data, as one has to make a choice between detrending for each grid point individually or subtracting or adding a common trend, such as the global averaged SSTs. To avoid these complications, we present results obtained with non-detrended data.

2.2 | Methods

Two SST-based indices are used, the Niño 3.4 index (hereafter N34), which was calculated as the area-averaged SST anomalies between 5°S and 5°N and from 170°W to 120°W (Trenberth, 1997), and the Dipole Mode Index (DMI hereafter; Saji *et al.*, 1999), which is the standard metric for the IOD. The DMI was calculated as the difference between the area-averaged SST anomalies in the western (10°S–10°N, 50°E–70°E) and eastern (10°S–0°N, 90°E–110°E) Indian Ocean. Both SST indices were standardised.

The area between 11°S and 12°N and from 31°E to 51°E is our East African reference region (outlined in Figure 1a). This region was selected to encompass areas where the rainfall during the OND season exceeds the average rainfall during the rest of the year, and to avoid the influence of rainfall over the Congo Basin (see Figure 1b,c). We used the seasonal means of rainfall inside this region to compute the rainfall EOFs, as well as to carry out a cluster analysis across all pooled reforecast members.

The principal components for seasonal rainfall anomalies were computed using the *eofs* Python package (Dawson, 2016). This package enables the application of specific weights to each grid point. Oceanic grid points were assigned a zero weight, whilst land and lake points were weighted based on their area. We present the first four EOFs, each characterised by its principal component (PC). The PCs are series matching the length of the original dataset: 42 for the CHIRPS data and 4200 for the SEAS5 data.

One of the main objectives of our study is to illustrate the nonlinearity and diversity of geographical rainfall patterns. Whilst it would have been possible to represent the geographical distribution of each rainfall season as a

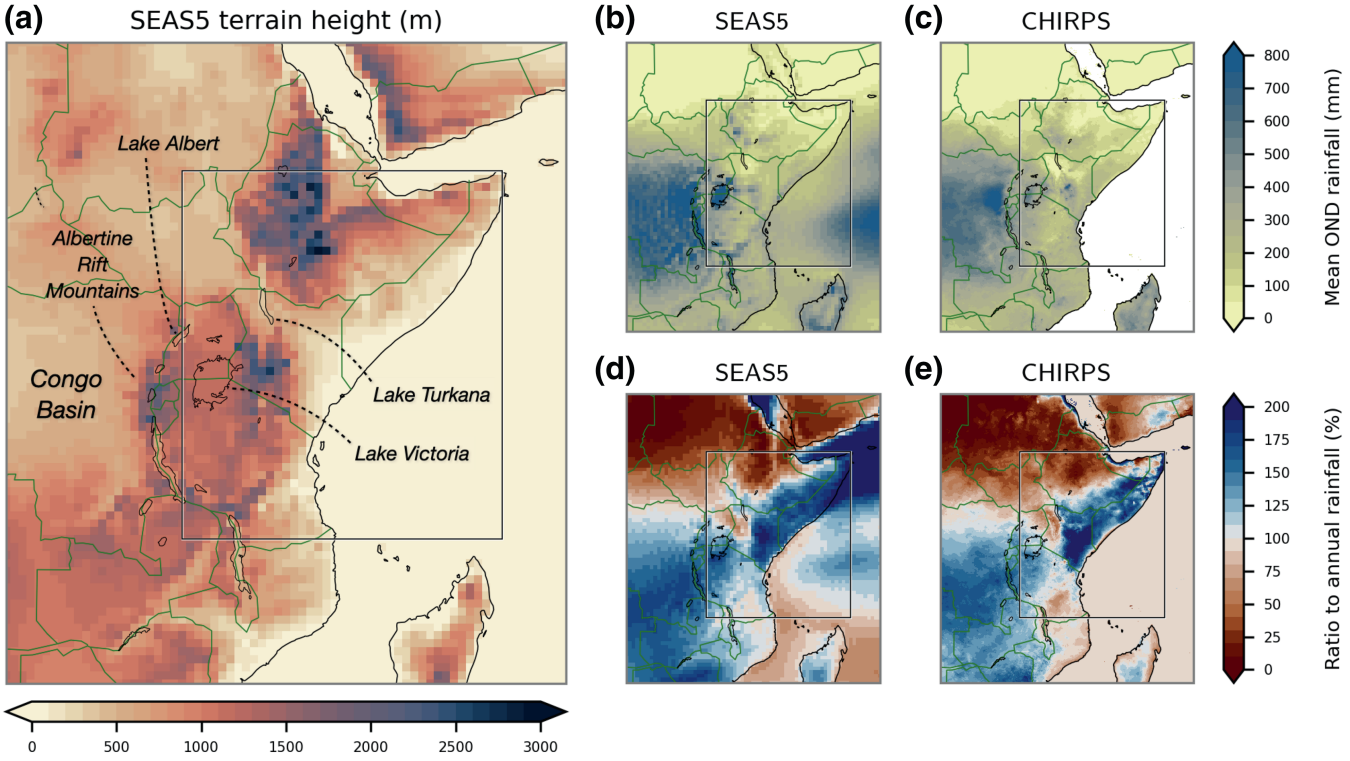

FIGURE 1 (a) Terrain height in SEAS5 on the same 0.5° grid spacing as the precipitation data, shown in metres. Key geographical features are indicated. The remaining panels show the mean total precipitation during OND in millimetres for (b) SEAS5 and (c) CHIRPS, and the ratio of average monthly rainfall during OND to the average monthly rainfall throughout the year for (d) SEAS5 and (e) CHIRPS. In each panel the East African reference region is indicated with a rectangle.

linear combination of the EOFs, we wished to divide the actual data into a specified number of clusters that are representative of recurring geographical patterns. A unique best approach for clustering three-dimensional data does not exist. Techniques such as self-organising maps (SOM) have been used successfully on climate data (Hewitson & Crane, 2002), but we chose to use a *k*-means algorithm on seasonal rainfall anomalies, a machine-learning approach designed to create groups that are as different as possible from each other (e.g., Steinley, 2006).

Whilst the *k*-means algorithm traditionally operates on Euclidean distances between two-dimensional points, it also proves effective for partitioning three-dimensional climate data into distinct clusters. Before applying the algorithm, we organised the data in a two-dimensional structure where the first dimension was time (4200 samples) and the samples along the second dimension consisted of spatial information collapsed into a single dimension (which was then reconstituted following the clustering). As a weighting strategy, we multiplied each sample for each grid point by the cosine of the latitude of land grid points and by zero for oceanic grid points. Because the *k*-means algorithm requires large data samples, we only applied it to the SEAS5 data and not to CHIRPS.

Method selection typically involves subjective decisions. In our usage of the *k*-means implementation in the SciKit-Learn Python library (Pedregosa *et al.*, 2011), we predetermined only the random state initialisation (the *random_state* parameter) for consistent results, and the number of clusters (*n_clusters*). Various *n_init* values, determining the iteration count, were also tested. Whilst this parameter significantly impacts small datasets, its influence was minimal on the large SEAS5 dataset. Identifying the optimal cluster number often involves multiple strategies, but in this case we relied on trial and error, seeking clusters that in our subjective judgement captured the variability among the model runs aptly. The results of this testing are discussed in Section 3.3.

Two methods of estimating statistical significance were used. For Pearson correlation calculations for time series, we used standard software, which estimates *p*-values based on two-sided *t*-tests. Choosing a significance level of 5% throughout, correlations were deemed significant if $p < 0.05$. We also used bootstrapping to estimate significance when working with the ensemble members. This was done by creating *N* synthetic time series by drawing random ensemble members from the pool of 4200 members (with replacement) and then calculating the metric under investigation for all the *N* elements of the synthetic

set. When the actual metric value fell outside the interval spanned by the 2.5th and 97.5th percentiles of the synthetic set, we defined it to be significant at the 5% level.

2.3 | Colour maps

Standard colour maps tend to accentuate extreme values disproportionately and are sometimes difficult to interpret for people with colour-vision deficiencies (Crameri *et al.*, 2020). To mitigate these issues, we use scientifically derived colour maps by Crameri (2021).

3 | RESULTS

3.1 | Mean rainfall

To provide context for interpreting the mean OND rainfall in East Africa, the terrain height in the SEAS5 model is shown in Figure 1a. The remaining maps in Figure 1 compare the mean OND rainfall in SEAS5 and CHIRPS, where the latter dataset is based on observations and satellite imagery.

The mean SEAS5 rainfall map in Figure 1b indicates that the rainfall over the Congo Basin is higher than in most of the East African reference region, except over Lake Victoria. A number of well-known features are also seen on the map, such as the low rainfall in the Rift Valley depression that crosses Lake Turkana (also known as the Turkana Channel; Munday *et al.*, 2023), the low rainfall in northern Somalia and northern Ethiopia, the high rainfall in the highlands of western Ethiopia and western Kenya, and the relatively high rainfall along the coastal regions of the Indian Ocean. The observed shift from high seasonal-mean rainfall in the Congo Basin to lower rainfall east of and around the Lake Victoria Basin is marked and abrupt, aligning with the mountainous terrain of the Albertine Rift. This pattern is consistently reflected in both SEAS5 and CHIRPS, underscoring the significant impact of topography on regional rainfall distribution.

When comparing SEAS5 rainfall with CHIRPS (Figure 1c), a few discrepancies are evident. One notable difference is that SEAS5 simulates more rainfall over the Congo Basin compared with CHIRPS. It is difficult to assess whether this reflects an error in SEAS5. Funk *et al.* (2015) acknowledged that some areas, "like the Democratic Republic of Congo (DRC), report virtually no rain gauge information." In an evaluation of several satellite-based precipitation products, Camberlin *et al.* (2019) found low correlations between daily CHIRPS rainfall and rain-gauge observations in both the central and eastern parts of the DRC. We mention this not to cast doubt on the CHIRPS data, but to emphasise that we do not know for certain which dataset is more correct.

A discrepancy is also observed over Lake Victoria, with higher rainfall over the lake in SEAS5 compared with CHIRPS. As with the Congo Basin, it is not immediately clear if this constitutes a model bias. Nicholson *et al.* (2021) compared rainfall over the lake across several datasets, including CHIRPS, and concluded that, whilst CHIRPS represents the spatial rainfall pattern well, it "clearly underestimate[s] overlake rainfall." If the high rainfall in SEAS5 is indeed a bias, it is probably linked to too strong convection over Lake Victoria in the model (see discussion in Section 3.3).

Aside from the discrepancies over the Congo Basin and Lake Victoria, there appears to be a good correspondence between the two datasets.

For each dataset, Figure 1d,e shows the ratio of the average monthly rainfall during the OND period to the average monthly rainfall throughout the year. Any instance where this ratio exceeds 100% indicates that the rainfall during OND is higher than the average for the year. Whilst there are some differences between the datasets, such as the lower CHIRPS ratios in southern Tanzania and higher ratios in western Kenya, both compared with SEAS5, the geographical structures are quite similar in the two datasets. This indicates that SEAS5 produces a realistic seasonal footprint for the OND rainfall compared with the rainfall during the rest of the year.

3.2 | EOF analysis

Our analysis of the patterns of interannual rainfall variability starts with a comparison of the leading rainfall EOFs in SEAS5 and CHIRPS. The EOF analysis serves a dual purpose. Firstly, it demonstrates that, despite the dominance of the leading EOF, important geographical patterns not encapsulated by EOF1 exist. Secondly, it diagnoses the consistency in the representation of rainfall variability between the datasets.

The patterns associated with the first four EOFs are shown for both SEAS5 and CHIRPS in Figure 2. In visual terms, it is clear that the first EOF is representative of the geographically averaged rainfall in the study region. This is in agreement with previous studies (e.g., Kolstad & MacLeod, 2022). EOF1 explains 47% of the variance in SEAS5 (Figure 2a) and 63% in CHIRPS (Figure 2e). The dominance of the leading EOF is consistent with other studies (e.g., Schreck & Semazzi, 2004; Wenhaji Ndomeni *et al.*, 2018). Although the large-scale spatial pattern of EOF1 is broadly similar across the datasets, some differences stand out. For instance, there is a much sharper drop-off in power west of the Albertine Rift in CHIRPS than in SEAS5. This gradient even leads to a change in sign over the Congo basin in CHIRPS. The extension of positive EOF1 correlations into the Congo Basin in SEAS5 is

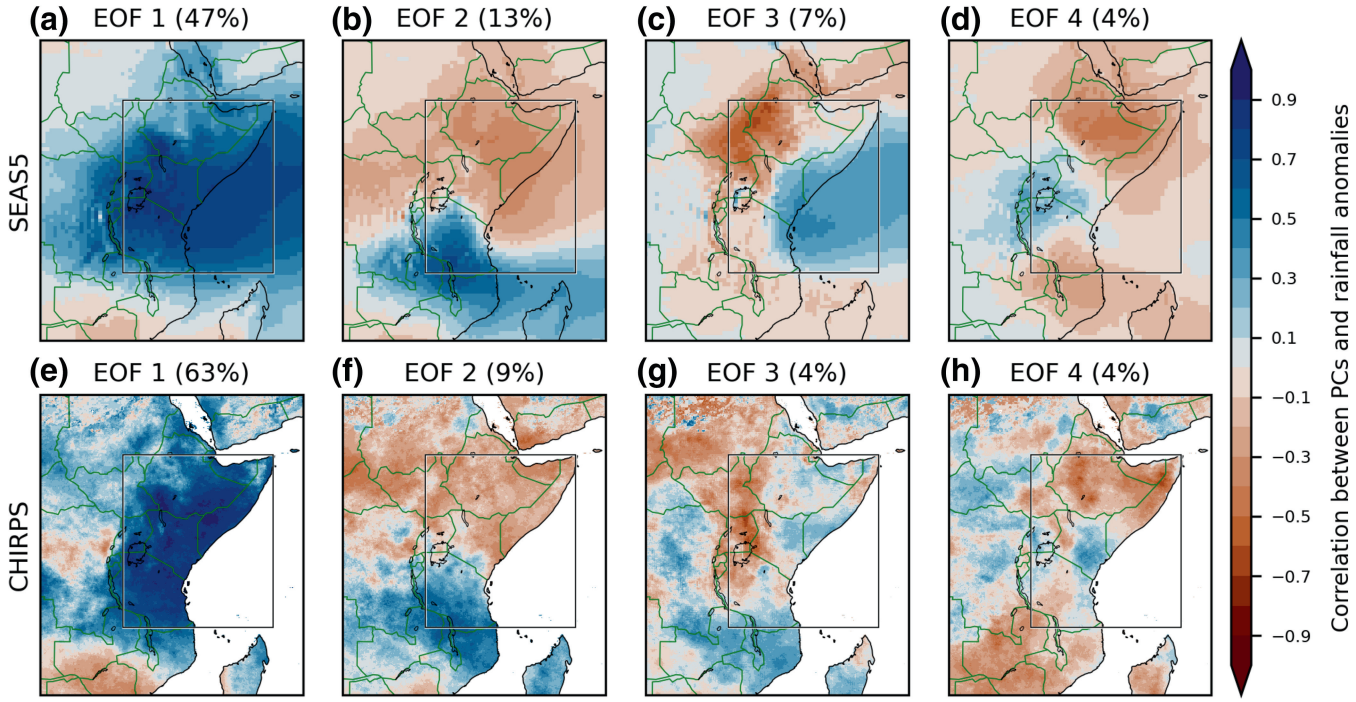

FIGURE 2 Correlations between OND precipitation and the first four PCs of East African rainfall, for (a–d) SEAS5 and (e–h) CHIRPS. The rectangle in each panel shows the region used to compute the EOFs, and the percentages in the captions indicate the fraction of total variance explained by each EOF. Note that, even though the oceanic grid points were masked when computing the EOFs for SEAS5, we show the rainfall over the ocean in panels (a)–(d).

probably linked to the higher average rainfall in this region compared with CHIRPS (Figure 1b vs. Figure 1c).

The dominance of EOF1 is influenced by a factor not seen in Figure 2, namely that the magnitude of the East African area-averaged rainfall is asymmetric for wet and dry anomalies. This imbalance arises because the rainfall has a hard lower threshold (as it cannot be lower than zero), whilst there is no such clearly defined upper threshold. This has important consequences for the flood risk due to extreme rainfall, which lacks a hard upper boundary (MacLeod *et al.*, 2024). In Section 3.3 on rainfall clusters below, it will be shown that this asymmetry gives rise to an inflated difference in the variance explained by EOF1 compared with the lower-order EOFs.

Arguably, the EOF2 structure is quite similar across the two datasets, as shown in Figure 2b,f. It mainly describes a gradient between Tanzania in the south and the areas to the north and northeast of Lake Victoria. In SEAS5, EOF2, which explains 13% of the variance, is distinctly separated from EOF3 (Figure 2c), which only explains 7%. A similar separation is observed in CHIRPS, where EOFs 2 and 3 (Figure 2f,g) explain 9% and 4% of the variance, respectively. In SEAS5, EOF3 describes a distinctive east-/west gradient, and its spatial pattern is well separated from EOF4 (Figure 2d). This is not the case in CHIRPS, although EOFs 3 and 4 (Figure 2g,h) both share similarities with their SEAS5 counterparts. Hence it is only in SEAS5 that a distinctive east/west gradient is seen.

Despite some minor discrepancies between SEAS5 and CHIRPS, especially for higher order EOFs, the EOF analysis gives us confidence that SEAS5 represents the geographical characteristics of OND rainfall in East Africa adequately. In the remainder of the article, we present results using the SEAS5 data to understand relationships with climatic drivers, leveraging the larger sample size compared with CHIRPS.

3.3 | Cluster analysis

We now present the results of a k-means algorithm to divide the 4200 simulated OND seasons in SEAS5 into n clusters based on the geographical distribution of seasonal rainfall anomalies inside the reference region. Our assessment is that $n = 5$ maintains a suitable balance between preserving spatial detail and keeping the number of clusters low for ease of analysis. The results for this choice are shown in Figure 3. For comparison, Figure A1 in the Appendix illustrates the rainfall clusters for $n = 6$ and $n = 7$.

A consistent feature across these varying n values is that the clustering algorithm collects cases with extremely

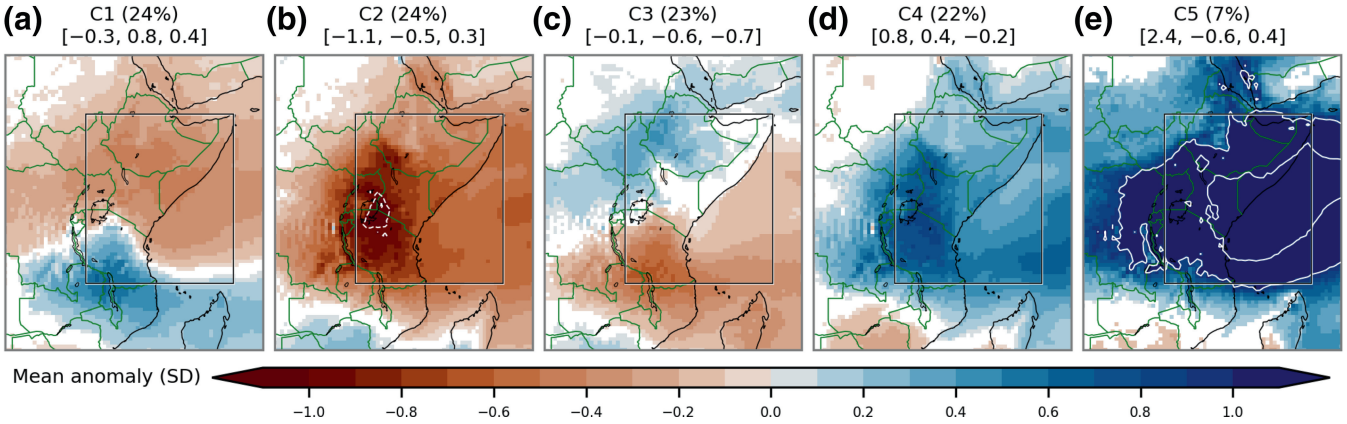

FIGURE 3 Mean standardised precipitation anomalies in each of the rainfall clusters ($n = 5$), sorted by cluster size. The captions include the name of each cluster, the frequency as a percentage in parentheses, as well as the mean PC value in brackets, for PCs 1, 2, and 3. Oceanic grid points are shown, although they were ignored when computing the clusters. Non-significant anomalies according to a bootstrapping test are masked, and the unit is standard deviation (SD). Because the colours saturate, white contours show isolines for anomalies of ± 1, 2, and 3 SD (negative contours dashed, positive contours solid).

high seasonal rainfall in the smallest cluster. When n is 6 or 7, this cluster contains 5%–7% of the model runs, and the corresponding C5 cluster contains 7% when $n = 5$ (Figure 3e). For $n > 5$, there is an additional relatively small cluster that also has high rainfall amounts, but these are less extreme than the smallest cluster. In our chosen configuration with $n = 5$, the cases in these wet clusters appear to be merged into C4 (Figure 3d).

The presence of one extremely wet cluster aligns well with the previously discussed asymmetry between extremely wet and dry cases. None of the cluster groups has an extremely dry cluster (in the sense that its mean negative PC1 value is close in magnitude to the positive PC1 value of C5). To assess the consequences of this asymmetry, we computed new EOFs for SEAS5 when leaving out the cases in cluster C5. This yielded almost identical spatial signatures for the EOFs (not shown). The recalculated EOFs excluding C5 cases only explained 37% of the variance for EOF1 (compared with 47% when all model runs were included). The second and third recalculated EOFs excluding C5 cases were found to explain more variance, 15% and 8%, respectively (compared with 13% and 7%, respectively, when all model runs were included).

The dichotomy between wet and dry extremes highlights one of the key reasons for adopting a cluster-based approach over EOFs. By isolating the most extreme rainfall cases in a distinct cluster, it becomes possible to investigate the mechanisms driving such cases separately. In contrast, an EOF-based analysis might have led to the erroneous inference that opposite mechanisms could lead to a similar group of extremely dry cases.

However, this does not mean that dry conditions are not captured by the clustering method. The C2 cluster shown in Figure 3b contains 24% of the model runs, has a mean PC1 value of −1.1, and is drier than normal within the entire reference region. Its anomalies are roughly mirrored by the wet cluster C4, which has a mean PC1 value of 0.8.

A consistent result for n between 5 and 7 is that the k-means algorithm selects a cluster that is dominated by EOF2 as the largest among the clusters. The mean PC2 value of C1 (Figure 3a), which contains 24% of the cases, is 0.8. This cluster has a counterpart in C3 (Figure 3c), for which the mean PC2 value is −0.6.

The cluster analysis confirms that, whilst EOF1 is the primary EOF and represents the average rainfall in East Africa effectively, it fails to capture the region's diverse rainfall geography on its own. EOF2 plays a crucial role in identifying the north–south distribution of rainfall anomalies, whereas EOF3 delineates the differences between the east and west. The mean PC values for the first three leading EOFs are significant across all clusters, highlighting their importance in representing the full variability of rainfall patterns within East Africa.

In essence, the five clusters shown in Figure 3 provide a useful representation of the diversity in the geographical variance across OND rainy seasons. To understand the potential drivers of the rainfall distribution in these five clusters better, we examine the large-scale climate anomalies associated with each. This starts with a composite analysis of seasonal mean tropical SST anomalies, followed by composite patterns of atmospheric humidity, winds at low levels and aloft, as well as divergence and convergence.

The composite mean SST anomalies for each cluster are illustrated in Figure 4. It is anticipated that clusters closely aligned with EOF1 would be associated with the

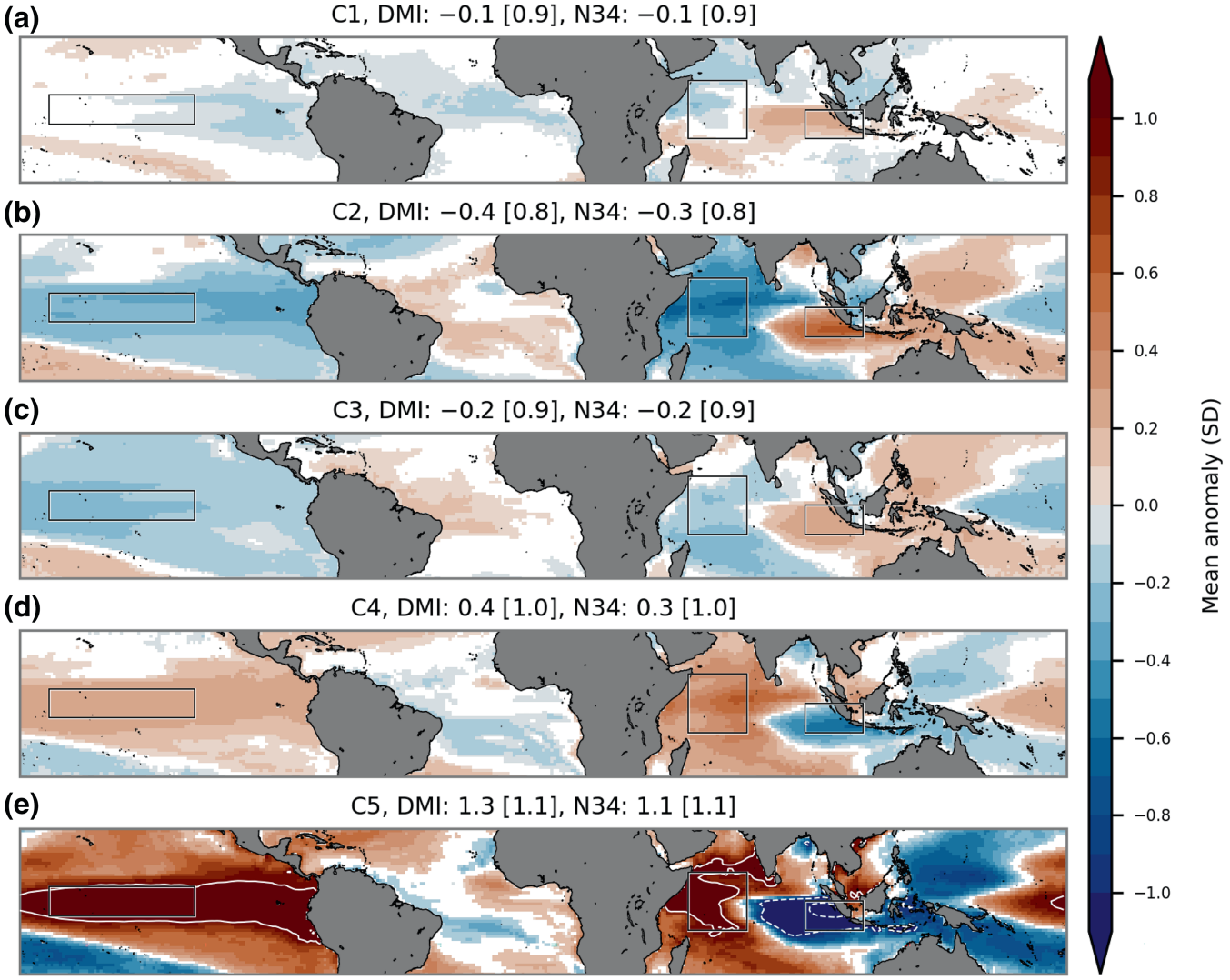

FIGURE 4 Mean standardised SST anomalies for each of the rainfall clusters, sorted by cluster population size as in Figure 3. The captions include the mean DMI and N34 for the members in each cluster, with the standard deviation of these indices in brackets. Non-significant (at the 5% level) anomalies according to a bootstrapping test are masked, and the unit is standard deviation (SD). Because the colours saturate, solid white contours show isolines for anomalies of 1 SD. Dashed white contours show isolines for anomalies of 1 and 2 SD. The rectangles show the reference region for ENSO and the western and eastern IOD regions.

phase of the IOD, and indeed, the mean DMI and N34 values in the extremely wet C5 cluster are higher than +1 SD and higher than or equal to the standard deviation across the cluster members (Figure 4e). However, C5 is the only cluster where any of the oceanic indices exceeds this standard deviation. In both the dry C2 cluster (Figure 4b) and the wet C4 cluster (Figure 4d), the mean DMI and N34 values are well below the standard deviation.

There is clearly a need to look beyond the IOD and ENSO to map the relationship between SSTs and the geographical diversity of the short rains. A simple and consistent pattern in Figure 4 is that the location of the strongest SST anomalies off the East African coast appears to correspond geographically with localised inland rainfall anomalies. For example, in C1, cold SST anomalies off the northern coast align with dry anomalies inland, whilst warm SST anomalies further south correspond to wet anomalies inland (Figure 4a). The C3 cluster, characterised mainly by dry conditions in the southern part of the region, has below-normal SST anomalies off the coast in this area (Figure 4c).

In the upcoming analysis, we explore these relationships by examining the Matsuno–Gill response to anomalous heating over the ocean. Our exploration of the atmospheric driving mechanisms in each cluster involves studying composite averages in various diagnostic variables. Specifically, we focus on the specific humidity and zonal and meridional wind components at 700 hPa, the divergence and wind components at 200 hPa, and the MFC and moisture transport ($q\boldsymbol{v}$) at 700 hPa.

Examining the ERA5 climatology and SEAS5 bias for diagnostic variables during the OND season (Figure 5), we note specific patterns. At 200 hPa, there is anticyclonic circulation in both hemispheres over land, shifted slightly eastward in the north, leading to an "S-like" shape in the mean cross-equatorial flow (Figure 5a). Notably, the easterly wind direction aloft over the Congo Basin is consistent with the wind direction at 700 hPa (Figure 5b), whereas the upper-level subtropical jets to the north and south of the East African reference region are distinctly westerly, in contrast to the low-level winds. Over the areas with weak vertical wind shear in the western part of the map, there is strong average 200-hPa divergence, indicating frequent deep convection. The Congo Basin divergence is especially pronounced over the boundary between the Basin and the Albertine Rift mountains. Highly divergent mean conditions also occur over Lake Victoria, but, apart from this area, the average conditions over the reference region are characterised by convergence aloft. This indicates subsidence in the air column below.

At 700 hPa (Figure 5b), mean flows from the northeast, east, and southeast are observed north, around, and south of the Equator, respectively. Upstream of northeastern East Africa, drier air masses prevail offshore to the northeast and over the Arabian peninsula, indicating dry advection in the region's climatology. This is illustrated by the mean ERA5 moisture advection, which is shown in Figure A2a in the Appendix. Similar dry advection is seen in the southern part of East Africa, with upstream air masses offshore being drier than over land. In Figure 5b, we note especially the relatively dry air masses off the coast of Tanzania and Mozambique downstream from Madagascar. These patterns of dry advection are consistent with a mainly negative MFC climatology over the East African reference region (Figure 5c). Figure A2b in the Appendix shows that the climatology of the conver-

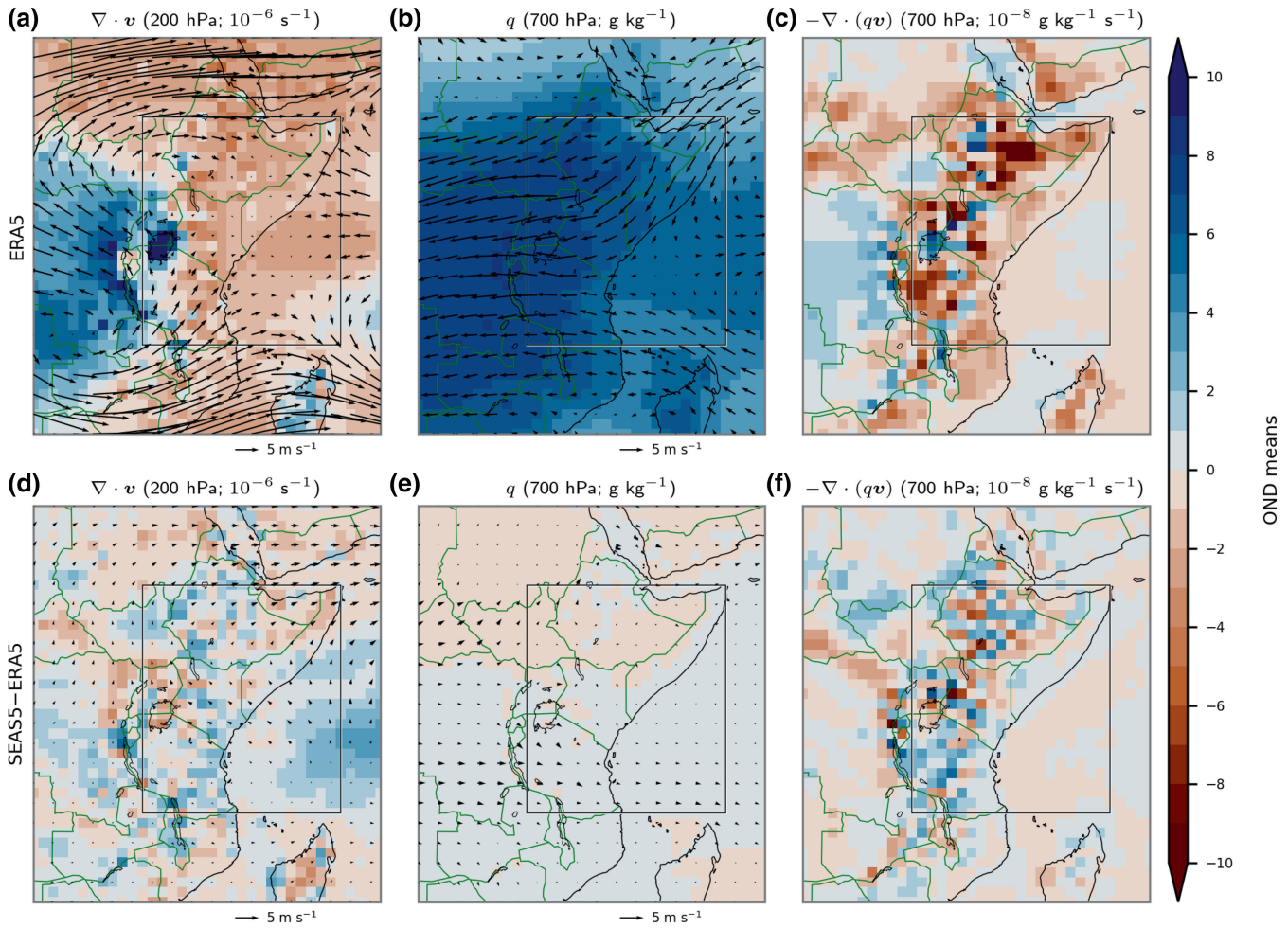

FIGURE 5 Top row: ERA5 climatology during OND of (a) 200-hPa divergence, (b) specific humidity at 700 hPa, and (c) moisture-flux convergence at 700 hPa. The units are specified in the captions. In (a,b), the vectors show the climatological wind components, with the scale and unit indicated below each panel. Bottom row: SEAS5−ERA5 climatologal differences for the same variables.

gence term of the MFC is almost identical to the total MFC, signifying that the convergence term dominates the MFC over the advection term.

Returning to Figure 5b, it is evident that the most moisture-laden air masses are concentrated over inland areas, particularly the western part of the reference region, which is part of a larger moist region extending into the Congo Basin. However, it is noteworthy that only over the Congo Basin does the MFC exhibit a consistently positive value over a large, contiguous area (Figure 5c). In contrast, over the highlands surrounding Lake Victoria, the MFC is predominantly negative. This spatial pattern, characterised by positive MFC over the Basin and negative MFC over western East Africa, aligns well with the east–west gradients in rainfall seen earlier in Figure 1a,b.

We now investigate the biases in SEAS5 relative to ERA5 in the bottom row of Figure 5. Figure 5d shows a mean positive bias in the divergence. While there is still convergence on average over the reference region in SEAS5, its magnitude is reduced by about a third. Figure 5e demonstrates that SEAS5 generally reproduces the wind and moisture patterns depicted in Figure 5b with reasonable accuracy, although there is a slight dry bias in the north and a comparable wet bias in the south. Moreover, a systematic bias is observed in the winds, where the average easterlies over the southern part of the Congo Basin and the East African region are slightly weaker in SEAS5 compared with ERA5. This discrepancy contributes to a positive MFC bias to the south and east of Lake Victoria (Figure 5f). This bias may stem from several factors, including the possibility of too frequent episodes of moisture import from the Congo Basin, or insufficient moisture transport into the Basin from East Africa. Similar positive MFC biases are seen in some areas north of the Equator, such as over parts of Ethiopia and along the southern Sudanese borders. Over the reference region, the average MFC bias is positive. The average negative MFC found in ERA5 remains negative in SEAS5, but is about 10% smaller in the mean.

Having explored the climatology and biases of the diagnostic variables, we now turn to Figure 6, which shows the mean standardised anomalies in these variables across the members of each cluster. It is illustrative to start by analysing C5, which is extremely wet in East Africa (Figure 3e). Its upper-level pattern is dominated by intense divergence anomalies over the western Indian Ocean, indicating deep convection (Figure 6e). Over land there is also widespread anomalous divergence and deep convection, with a notable exception over the eastern branch of the Rift Valley. This branch extends from east of Lake Victoria towards the northeast into Ethiopia via Lake Turkana. Anomalous convergence is also observed over the southern part of the East African region.

At 700 hPa, the air over the entire reference region is considerably more humid than usual, attributed to onshore winds transporting moist air masses from the ocean (Figure 6j). The distinct geographical differences in rainfall in C5, as observed in Figure 3e, are closely tied to local MFC anomalies depicted in Figure 6o. These anomalies seem to be influenced by interactions with topography and land–sea contrasts. For instance, the notable convergence in the western part of East Africa (Figure 6o) results from the convergence between the easterly wind anomalies over the region and the abrupt shift to southerly wind anomalies along the Albertine Rift Mountains (the boundary between the highlands and the eastern Congo Basin; Figure 6j). Moreover, the areas previously identified with convergence aloft display weaker or neutral MFC anomalies at 700 hPa. These interactions between air masses and underlying surfaces create a mosaic of predominantly positive mean MFC anomalies. The overarching pattern suggests a robust moisture inflow from the Indian Ocean, serving as the primary moisture source in C5.

The C4 cluster, which like C5 is anomalously wet in most of East Africa but particularly in the south (Figure 3d), has a flow pattern that shares many similarities with C5. Aloft, the same pattern of onshore divergent winds is seen, with the same contrasting convergence over the eastern branch of the Rift Valley (Figure 6d) as in C5. At 700 hPa, however, distinct differences between the two clusters are seen in the wind anomalies. Where C5 was found to exhibit very strong onshore wind anomalies from over the Indian Ocean, there is no such clear pattern of anomalous onshore flow in C4 (Figure 6i). Another difference is that C4 exhibits a distinct signature of westerly wind anomalies from the Congo Basin along the western boundary of the East African reference region not seen in C5 (Figure 6j). These westerly wind anomalies along the Equator align better with the expected Matsuno–Gill response to the warm SST anomalies outside the southern part of the region (Figure 4d) compared with the wind anomalies in C5. This suggests that moisture from the Congo Basin, relative to moisture originating over the Indian Ocean, contributes more to the East African rainfall in C4 than in C5. However, in terms of MFC, the spatial structures in C4 and C5 are quite similar (Figure 6n,o).

The behaviour of the dry cluster C2 can mainly be understood as the reverse of its wet counterpart C4. C2, consistent with an inverse of C4, has upper-level cyclonic flow on each side of the Equator (Figure 6b) and negative anomalous low-level MFC over East Africa (Figure 6l), which is due to anomalous outflow mainly towards the Congo Basin in the west (Figure 6g). This structure corresponds well with a Matsuno–Gill response to the cool SST anomalies off the southern coast (Figure 4b).

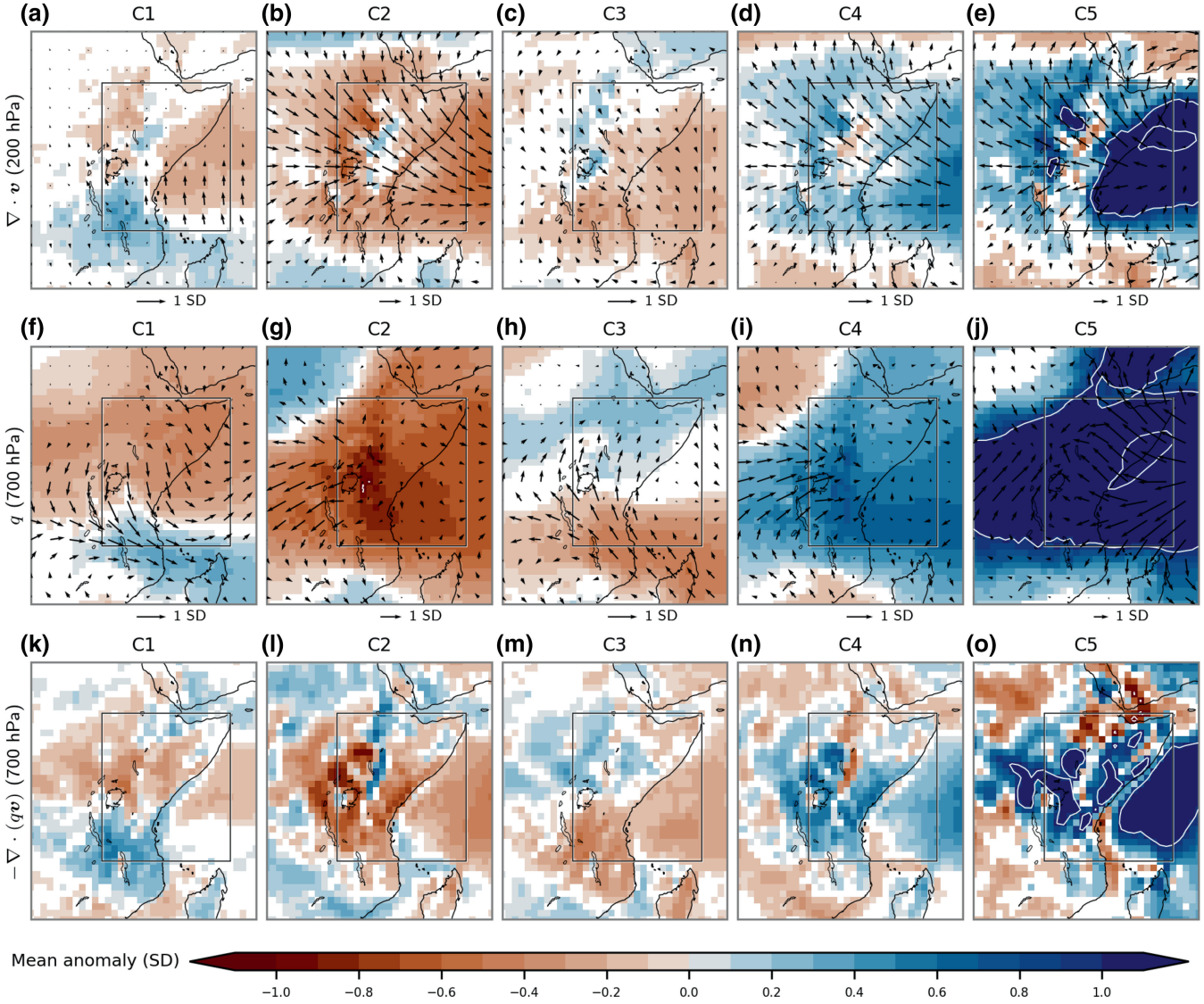

FIGURE 6 Colours: mean standardised anomalies of specific humidity at 700 hPa (top row), divergence at 200 hPa (middle row), and moisture-flux convergence at 700 hPa (bottom row) in the rainfall clusters. Non-significant anomalies (at the 5% level) according to a bootstrapping test are masked. The vectors in (a–j) show the mean standardised anomalies in the wind components at the same levels, with the scale indicated below each panel (note the varying scale). The unit for all variables is standard deviation (SD).

As seen in Figure 3, the remaining two clusters, C1 and C3, involve rainfall contrasts between the northern and southern parts of East Africa. Considering the wet southern region in C1 first, warm SST anomalies offshore (Figure 4a) are consistent with 200-hPa divergence, indicating deep convection (Figure 6a). This is coupled with anomalous flow from the west and northwest towards the warm SST anomalies north of Madagascar at 700 hPa (Figure 6f), in almost direct opposition to the mean flow (Figure 5b). The westerly direction of the anomalous flow is consistent with a Matsuno–Gill response, shifted off the Equator, and the convergent anomalous wind field yields positive low-level MFC anomalies over the anomalously wet areas (Figure 6k). Off the coast of the anomalously dry northern part of the region in C1, there is anomalous convergence at 200 hPa (Figure 6a), which is indicative of a strengthening of the climatological subsidence over this region seen in Figure 5a. It is noteworthy that the anomalous flow at 700 hPa over the northern part of the region (Figure 6f) is not directed away from this region of anomalous subsidence. Instead, the direction of the wind anomalies is northwesterly, leading towards the sinking air masses. Considering that the typical flow over the northern part of East Africa, as shown in Figure 5a, originates from the northeast, the resultant actual flow in cluster C1 is approximately from the north. This northerly flow brings air from an area of lower moisture, as seen in Figure 6f. The divergent nature of this anomalous flow at

700 hPa leads to negative MFC anomalies over northern East Africa, illustrated in Figure 6k. This anomaly aligns with the observed below-normal rainfall in the north in the C1 cases, as shown in Figure 3a.

Over the southern part of the region, the behaviour of the diagnostic variables in C3 is the opposite of their behaviour in C1. The dryness over the southern region in C3 is likely influenced by the cold SST anomalies off the southern coast, as indicated in Figure 4c. These SST anomalies are associated with convergence aloft, as shown in Figure 6c. Additionally, there is an anomalously dry southeasterly flow over the southern region at 700 hPa (Figure 6h), which strengthens the climatological flow from the same direction (Figure 5b). The resulting negative MFC anomalies (Figure 6m) over this region are consistent with the dryness on the ground. Recalling from Figure 3c that the northern part of the region is weakly wetter than normal in C3, we see that this appears to be linked to above-normal MFC at 700 hPa (Figure 6m). This enhanced MFC seems to arise from the convergent nature of the anomalous flow (Figure 6h).

4 | SUMMARY AND DISCUSSION

In our investigation of East Africa's short rains, our research was steered by three core ideas. First, we investigated whether looking beyond the dominant regional average would reveal a high degree of geographical rainfall variability. Second, we explored the idea that SST anomalies in the western Indian Ocean influence the rainfall variability significantly, potentially triggering a Matsuno–Gill response in the overland circulation, with associated feedbacks through the moisture budget. We indicated how this response differs for positive and negative rainfall anomalies, through the changing geographic pattern of thermodynamic heating when the positive anomalies extend over land. Third, we examined the notion that the precise location of the SST anomalies plays a critical role in shaping the rainfall patterns across East Africa, and how these detailed patterns can again be interpreted in terms of the Matsuno–Gill response.

To examine these ideas, we initially evaluated the SEAS5 model to assess its accuracy in depicting the region's rainfall diversity. We then applied a clustering method to represent the complex regional rainfall patterns better. Finally, we investigated the relationship between these patterns and various large-scale climatic and local factors, including SSTs, wind patterns, and moisture fluxes. This section aims to summarise our approach to these issues and the key insights we have gained.

A comparative analysis of SEAS5 and CHIRPS rainfall data revealed that average OND rainfall quantities are broadly similar across both datasets, though certain biases emerged. Notably, the model tends to produce more rainfall in the Congo Basin and around Lake Victoria than what is found in CHIRPS. The ratio of average OND rainfall to annual rainfall exhibited reassuringly similar patterns in both datasets, with higher ratios in Somalia and south of Lake Victoria and lower ratios in the southern and northern extremities of the region. A notable success of the model is its ability to replicate the low rainfall ratio around Lake Turkana accurately, underlining its effectiveness in local-scale rainfall representation.

By comparing rainfall EOF patterns across SEAS5 and CHIRPS, we identified a strong dominance by the primary EOF in explaining the rainfall variability, with a slightly greater dominance in CHIRPS compared with SEAS5. This first EOF was found to be characterised by large rainfall anomalies spanning the entire study area. The subsequent EOFs describe gradients between different parts of the region, offering nuance to the picture of nearly uniform anomalies across the region represented by EOF1. A critical, implicit assumption in EOF analysis (because it is a linear method) is a symmetry between the positive and negative phases, which our clustering analysis of SEAS5 data challenged. For all the specified number of clusters, a small cluster with extremely high rainfall emerged, highlighting an important asymmetry in the data. In light of these findings, we shifted our focus to examining the oceanic and atmospheric conditions associated with the clusters.

A key feature shared by the clusters is the dominant influence of a Matsuno–Gill response on rainfall anomalies, which effectively extends these anomalies from the ocean westward across the land (Bahaga *et al.*, 2015). This phenomenon is largely driven by westerly wind anomalies that favour increased inland rainfall (Camberlin & Wairoto, 1997; Finney *et al.*, 2019), a response triggered by offshore heating (also strongly correlated with strengthened easterlies in the central equatorial Indian Ocean; Hastenrath *et al.*, 2011). Conversely, dry offshore anomalies often lead to easterly anomalies over land, resulting in reduced rainfall. However, in scenarios with a pronounced positive rainfall anomaly over land, the atmospheric latent heating transcends the usual SST patterns and extends inland. Under these circumstances, the enhanced rainfall over land is sustained by deep easterly moisture flux from the Indian Ocean.

The SST patterns related to the rainfall clusters indicate a connection between the north–south SST gradient over the western Indian Ocean and the spatial distribution of anomalously wet and dry areas on land. This association has previously been established for the March–May (MAM) season (Vellinga & Milton, 2018), but to the best of our knowledge not for the OND period. In the MAM

season, warm SST anomalies in the northern part of the western Indian Ocean correlate with increased rainfall over the Horn of Africa. Conversely, SST anomalies towards the southern end of the basin are associated with rainfall anomalies concentrated in northern Tanzania and southern Kenya (Seregina *et al.*, 2021). Our findings for the OND season reveal a similar pattern, where the location of cold SST anomalies aligns with dry areas over land. A recommended future study could delve into the relationship between the geographical variations of SST anomalies in the Indian Ocean and East African rainfall's spatial variability, extending beyond the east/west gradient linked with the IOD.

Our findings reinforced the core ideas underpinning this study. The first key finding was that the largest of the rainfall clusters was consistently dominated by the second EOF pattern, rather than the first. This observation affirmed our initial idea that it is important to look at more nuanced spatial patterns to understand the region's rainfall variability. Secondly, we largely confirmed that rainfall clusters can be understood within a framework linked to anomalous flows consistent with a Matsuno–Gill response to SST anomalies in the western Indian Ocean. However, in extremely wet seasons this pattern is confounded, with rainfall over land being supported by strong easterly moisture fluxes from the Indian Ocean. Thirdly, our analysis suggested that the specific locations of SST anomalies critically influence the anomalous flow and, by extension, the distribution of rainfall over land. It is important to note that these interpretations stem from condensing 4200 individual seasons into only five rainfall clusters. Inevitably, there will be numerous exceptions to these generalised patterns.

An attempt was made to link the rainfall clusters to drivers that may be predictable. The first driver is the ocean surface, represented by SSTs. Only the most extremely wet cluster was convincingly linked to the IOD and ENSO, an association that is well-known from previous studies (e.g., Black, 2005; Hirons & Turner, 2018; MacLeod *et al.*, 2024). However, clusters exhibiting moderate anomalies, whether wet or dry, or those presenting geographical gradients across East Africa, did not demonstrate a consistent association with either the IOD or ENSO.

Although we did not mention this in the previous section, we also attempted to link the clusters to the MJO, using standard real-time multivariate (RMM; Wheeler & Hendon, 2004) MJO data calculated from daily SEAS5 reforecasts (kindly provided by Frédéric Vitart). The MJO's spatial influence in East Africa is closely aligned with the pattern associated with EOF1, likely due to its characteristic of eastward-propagating waves. As a result, the RMM index seems to offer limited insight into the north–south rainfall gradients within East Africa. Consequently, the clusters most closely related to EOF1 also exhibit the strongest associations with MJO phases. In summary, the MJO's predictable influence seems to be primarily limited to indicating whether the OND season will be wetter or drier than average. On the seasonal time-scale, variations in geographical rainfall within East Africa do not appear to be influenced significantly by the MJO.

5 | CONCLUDING REMARKS

In this study, we challenge the conventional approach that simplifies East Africa's short rains to a regional mean, thereby uncovering significant geographical variations. Our findings demonstrate that the regional average fails to represent the complex rainfall patterns experienced across different areas. Whilst large-scale factors like SST patterns and the MJO are effective in predicting the regional mean rainfall, they are inadequate for forecasting the nuanced geographical rainfall variations within the region. These variations appear to be influenced by regional climatic factors, including north–south SST gradients in the western Indian Ocean, air-mass interactions with topography, and the spectrum of easterly to westerly wind flows, which influence moisture dynamics from the Congo Basin. It remains uncertain whether these features are currently predictable. Research into their predictability is crucial, as more accurate localised forecasts have the potential to bolster the resilience to rainfall anomalies in a region deeply dependent on rain-fed agriculture and vulnerable to droughts and floods.

ACKNOWLEDGEMENTS

We gratefully acknowledge Frédéric Vitart at ECMWF for the provision of the processed MJO data from the SEAS5 reforecast. We also thank Rondrotiana Barimalala and Teferi Dejene Demissie, who took part in our discussions. CW thanks the Bjerknes Centre for Climate Research for a grant to visit Bergen to work on this article. The research was supported by the European Union through the Horizon 2020 grants 869730 (CONFER) and 869550 (DOWN2EARTH), and by the National Centre for Atmospheric Science through the NERC National Capability International Programmes Award NE/X006263/1. Finally, we acknowledge the assistance of OpenAI's ChatGPT in providing preliminary guidance and language editing for this article.

DATA AVAILABILITY STATEMENT

The SEAS5 data are available in the ECMWF's MARS archive. An exception is the MJO data derived from the SEAS5 data. These were computed and provided to us by Frédéric Vitart at ECMWF. The ERA5 reanalysis is available from the EU via the Copernicus Climate Data

Store at https://climate.copernicus.eu/climate-reanalysis. The CHIRPS data are available from the University of Santa Barbara's repository at https://data.chc.ucsb.edu/products/CHIRPS-2.0/.

ORCID

Erik W. Kolstad https://orcid.org/0000-0001-5394-9541
Douglas J. Parker https://orcid.org/0000-0003-2335-8198

REFERENCES

Bahaga, T.K., Fink, A.H. & Knippertz, P. (2019) Revisiting interannual to decadal teleconnections influencing seasonal rainfall in the Greater Horn of Africa during the 20th century. *International Journal of Climatology*, 39(5), 2765–2785. Available from: https://doi.org/10.1002/joc.5986

Bahaga, T.K., Mengistu Tsidu, G., Kucharski, F. & Diro, G.T. (2015) Potential predictability of the sea-surface temperature forced equatorial East African short rains interannual variability in the 20th century. *Quarterly Journal of the Royal Meteorological Society*, 141(686), 16–26. Available from: https://doi.org/10.1002/qj.2338

Black, E. (2005) The relationship between Indian Ocean sea-surface temperature and East African rainfall. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 363(1826), 43–47. Available from: https://doi.org/10.1098/rsta.2004.1474

Black, E., Slingo, J. & Sperber, K.R. (2003) An observational study of the relationship between excessively strong short rains in Coastal East Africa and Indian Ocean SST. *Monthly Weather Review*, 131(1), 74–94. Available from: https://doi.org/10.1175/1520-0493(2003)131<0074:Aosotr>2.0.Co;2

Camberlin, P., Barraud, G., Bigot, S., Dewitte, O., Makanzu Imwangana, F., Maki Mateso, J.-C. et al. (2019) Evaluation of remotely sensed rainfall products over Central Africa. *Quarterly Journal of the Royal Meteorological Society*, 145(722), 2115–2138. Available from: https://doi.org/10.1002/qj.3547

Camberlin, P. & Wairoto, J. (1997) Intraseasonal wind anomalies related to wet and dry spells during the "long" and "short" rainy seasons in kenya. *Theoretical and Applied Climatology*, 58, 57–69.

Crameri, F. (2021) *Scientific Colour Maps*. Zenodo. Available from: https://doi.org/10.5281/zenodo.4491293

Crameri, F., Shephard, G.E. & Heron, P.J. (2020) The misuse of colour in science communication. *Nature Communications*, 11(1), 5444. Available from: https://doi.org/10.1038/s41467-020-19160-7

Dawson, A. (2016) eofs: A library for eof analysis of meteorological, oceanographic, and climate data. *Journal of Open Research Software*, 4(1), e14. Available from: https://doi.org/10.5334/jors.122

Diem, J.E., Sung, H.S., Konecky, B.L., Palace, M.W., Salerno, J. & Hartter, J. (2019) Rainfall characteristics and trends–and the role of Congo Westerlies–in the Western Uganda Transition Zone of Equatorial Africa From 1983 to 2017. *Journal of Geophysical Research: Atmospheres*, 124(20), 10712–10729. Available from: https://doi.org/10.1029/2019JD031243

Doi, T., Behera, S.K. & Yamagata, T. (2022) On the predictability of the extreme drought in East Africa during the short rains season. *Geophysical Research Letters*, 49(22), e2022GL100905. Available from: https://doi.org/10.1029/2022GL100905

Dunning, C.M., Black, E.C.L. & Allan, R.P. (2016) The onset and cessation of seasonal rainfall over Africa. *Journal of Geophysical Research: Atmospheres*, 121(19), 11,405–11,424. Available from: https://doi.org/10.1002/2016JD025428

Dyer, E.L.E., Jones, D.B.A., Nusbaumer, J., Li, H., Collins, O., Vettoretti, G. et al. (2017) Congo Basin precipitation: Assessing seasonality, regional interactions, and sources of moisture. *Journal of Geophysical Research: Atmospheres*, 122(13), 6882–6898. Available from: https://doi.org/10.1002/2016JD026240

Finney, D.L., Marsham, J.H., Walker, D.P., Birch, C.E., Woodhams, B.J., Jackson, L.S. et al. (2019) The effect of westerlies on East African rainfall and the associated role of tropical cyclones and the Madden-Julian Oscillation. *Quarterly Journal of the Royal Meteorological Society*, 146, 647–664. Available from: https://doi.org/10.1002/qj.3698

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S. et al. (2015) The climate hazards infrared precipitation with stations–a new environmental record for monitoring extremes. *Scientific Data*, 2(1), 150066. Available from: https://doi.org/10.1038/sdata.2015.66

Gill, A.E. (1980) Some simple solutions for heat-induced tropical circulation. *Quarterly Journal of the Royal Meteorological Society*, 106(449), 447–462.

Gudoshava, M., Wainwright, C., Hirons, L., Endris, H.S., Segele, Z.T., Woolnough, S. et al. (2022) Atmospheric and oceanic conditions associated with early and late onset for Eastern Africa short rains. *International Journal of Climatology*, 42(12), 6562–6578.

Hastenrath, S., Polzin, D. & Mutai, C. (2011) Circulation Mechanisms of Kenya Rainfall Anomalies. *Journal of Climate*, 24(2), 404–412.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz Sabater, J. et al. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. Available from: https://doi.org/10.1002/qj.3803

Hewitson, B.C. & Crane, R.G. (2002) Self-organizing maps: Applications to synoptic climatology. *Climate Research*, 22(1), 13–26.

Hirons, L. & Turner, A. (2018) The impact of Indian Ocean mean-state biases in climate models on the representation of the East African short rains. *Journal of Climate*, 31(16), 6611–6631. Available from: https://doi.org/10.1175/jcli-d-17-0804.1

Indeje, M., Semazzi, F.H. & Ogallo, L.J. (2000) ENSO signals in East African rainfall seasons. *International Journal of Climatology*, 20(1), 19–46. Available from: https://doi.org/10.1002/(SICI)1097-0088(200001)20:1<19::AID-JOC449>3.0.CO;2-0

Johnson, S.J., Stockdale, T.N., Ferranti, L., Balmaseda, M.A., Molteni, F., Magnusson, L. et al. (2019) SEAS5: the new ECMWF seasonal forecast system. *Geoscientific Model Development*, 12(3), 1087–1117.

Kebacho, L.L. (2021) Large-scale circulations associated with recent interannual variability of the short rains over East Africa. *Meteorology and Atmospheric Physics*, 134(1), 10. Available from: https://doi.org/10.1007/s00703-021-00846-6

Kelder, T., Müller, M., Slater, L.J., Marjoribanks, T.I., Wilby, R.L., Prudhomme, C. et al. (2020) Using UNSEEN trends to detect decadal changes in 100-year precipitation extremes. *npj Climate and Atmospheric Science*, 3(1), 47. Available from: https://doi.org/10.1038/s41612-020-00149-4

Kolstad, E.W. & MacLeod, D. (2022) Lagged oceanic effects on the East African short rains. *Climate Dynamics*, 59(3), 1043–1056. Available from: https://doi.org/10.1007/s00382-022-06176-6

Koppa, A., Keune, J., MacLeod, D.A., Singer, M., Nieto, R., Gimeno, L. et al. (2023) A Lagrangian analysis of the sources of rainfall over the horn of Africa Drylands. *Journal of Geophysical Research: Atmospheres*, 128(12), e2022JD038408. Available from: https://doi.org/10.1029/2022JD038408

Liu, W., Cook, K.H. & Vizy, E.K. (2020) Influence of Indian Ocean SST regionality on the East African short rains. *Climate Dynamics*, 54(11-12), 4991–5011. Available from: https://doi.org/10.1007/s00382-020-05265-8

MacKinnon, D.P., Fairchild, A.J. & Fritz, M.S. (2007) Mediation analysis. *Annual Review of Psychology*, 58, 593–614. Available from: https://doi.org/10.1146/annurev.psych.58.110405.085542

MacLeod, D., Kolstad, E.W., Michaelides, K. & Singer, M.B. (2024) Sensitivity of rainfall extremes to unprecedented Indian Ocean dipole events. *Geophysical Research Letters*, 51(5), e2023GL105258. Available from: https://doi.org/10.1029/2023GL105258

Manzanas, R. (2020) Assessment of Model Drifts in Seasonal Forecasting: Sensitivity to Ensemble Size and Implications for Bias Correction. *Journal of Advances in Modeling Earth Systems*, 12(3), e2019MS001751. Available from: https://doi.org/10.1029/2019MS001751

Matsuno, T. (1966) Quasi-geostrophic motions in the equatorial area. *Journal of the Meteorological Society of Japan. Ser. II*, 44(1), 25–43.

Maybee, B., Ward, N., Hirons, L.C. & Marsham, J.H. (2023) Importance of Madden-Julian oscillation phase to the interannual variability of East African rainfall. *Atmospheric Science Letters*, 24(5), e1148. Available from: https://doi.org/10.1002/asl.1148

Munday, C., Engelstaedter, S., Washington, R., Ogutu, G., Olago, D., Ouma, G. et al. (2023) The Turkana Jet diurnal cycle in observations and reanalysis. *Journal of Climate*. Available from: https://doi.org/10.1175/JCLI-D-23-0325.1

Nicholson, S.E. (2015) Long-term variability of the East African 'short rains' and its links to large-scale factors. *International Journal of Climatology*, 35(13), 3979–3990. Available from: https://doi.org/10.1002/joc.4259

Nicholson, S.E. (2017) Climate and climatic variability of rainfall over eastern Africa. *Reviews of Geophysics*, 55, 590–635. Available from: https://doi.org/10.1002/2016RG000544

Nicholson, S.E., Klotter, D. & Hartman, A.T. (2021) Lake-effect rains over lake victoria and their association with mesoscale convective systems. *Journal of Hydrometeorology*, 22(6), 1353–1368. Available from: https://doi.org/10.1175/JHM-D-20-0244.1

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. et al. (2011) Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Pohl, B. & Camberlin, P. (2011) Intraseasonal and interannual zonal circulations over the equatorial Indian Ocean. *Theoretical and Applied Climatology*, 104, 175–191. Available from: https://doi.org/10.1007/s00704-010-0336-1

Saji, N.H., Goswami, B.N., Vinayachandran, P.N. & Yamagata, T. (1999) A dipole mode in the tropical Indian Ocean. *Nature*, 401(6751), 360–363. Available from: https://doi.org/10.1038/43854

Schreck, C.J.I. & Semazzi, F.H.M. (2004) Variability of the recent climate of eastern Africa. *International Journal of Climatology*, 24(6), 681–701. Available from: https://doi.org/10.1002/joc.1019

Seregina, L.S., Fink, A.H., van der Linden, R., Funk, C. & Pinto, J.G. (2021) Using seasonal rainfall clusters to explain the interannual variability of the rain belt over the Greater Horn of Africa. *International Journal of Climatology*, 41(S1), E1717–E1737. Available from: https://doi.org/10.1002/joc.6802

Steinley, D. (2006) K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, 59(1), 1–34. Available from: https://doi.org/10.1348/000711005X48266

Stockdale, T., Balmaseda, M., Johnson, S., Ferranti, L., Molteni, F., Magnusson, L. et al. (2018) *SEAS5 and the future evolution of the long-range forecast system*. Reading, UK: ECMWF.

Thompson, V., Dunstone, N.J., Scaife, A.A., Smith, D.M., Slingo, J.M., Brown, S. et al. (2017) High risk of unprecedented UK rainfall in the current climate. *Nature Communications*, 8(1), 107. Available from: https://doi.org/10.1038/s41467-017-00275-3

Trenberth, K.E. (1997) The definition of El Nino. *Bulletin of the American Meteorological Society*, 78(12), 2771–2778.

Ummenhofer, C.C., Sen Gupta, A., England, M.H. & Reason, C.J. (2009) Contributions of Indian Ocean sea surface temperatures to enhanced East African rainfall. *Journal of Climate*, 22(4), 993–1013. Available from: https://doi.org/10.1175/2008JCLI2493.1

Vellinga, M. & Milton, S.F. (2018) Drivers of interannual variability of the East African "Long Rains". *Quarterly Journal of the Royal Meteorological Society*, 144(712), 861–876. Available from: https://doi.org/10.1002/qj.3263

Wainwright, C.M., Finney, D.L., Kilavi, M., Black, E. & Marsham, J.H. (2021) Extreme rainfall in East Africa, October 2019-January 2020 and context under future climate change. *Weather*, 76(1), 26–31. Available from: https://doi.org/10.1002/wea.3824

Wenhaji Ndomeni, C., Cattani, E., Merino, A. & Levizzani, V. (2018) An observational study of the variability of East African rainfall with respect to sea surface temperature and soil moisture. *Quarterly Journal of the Royal Meteorological Society*, 144(S1), 384–404. Available from: https://doi.org/10.1002/qj.3255

Wheeler, M.C. & Hendon, H.H. (2004) An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Monthly Weather Review*, 132(8), 1917–1932.

Zhao, S. & Cook, K.H. (2021) Influence of Walker circulations on East African rainfall. *Climate Dynamics*, 56, 2127–2147. Available from: https://doi.org/10.1007/s00382-020-05579-7

How to cite this article: Kolstad, E.W., Parker, D.J., MacLeod, D.A., Wainwright, C.M. & Hirons, L.C. (2024) Beyond the regional average: Drivers of geographical rainfall variability during East Africa's short rains. *Quarterly Journal of the Royal Meteorological Society*, 150(764), 4550–4566. Available from: https://doi.org/10.1002/qj.4829

APPENDIX. SUPPORTING FIGURES

Figure A1 shows the results of the clustering algorithm for $n = 6$ and $n = 7$. Figure A2 shows the ERA5 climatology of the advection and convergence terms of the moisture-flux convergence.

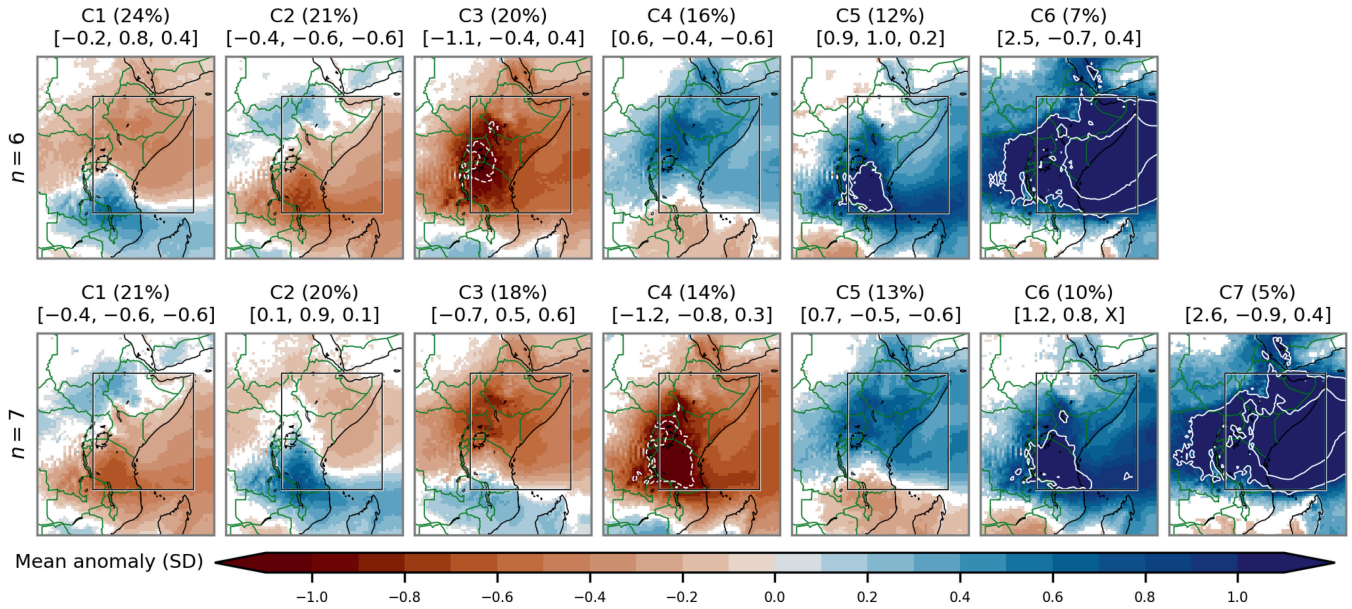

FIGURE A1 Mean standardised precipitation anomalies in each of the rainfall clusters, for $n = 6$ (top row) and $n = 7$ (bottom row), sorted by cluster size. The captions include the name of each cluster, the frequency as a percentage in parentheses, and the mean PC value in brackets, for PCs 1, 2, 3, and 4 (the letter "X" indicates that the mean anomalies are non-significant). Oceanic grid points are shown, although they were ignored when computing the clusters. Non-significant anomalies according to a bootstrapping test are masked, and the unit is standard deviation (SD). Because the colours saturate, white contours show isolines for anomalies of ± 1, 2, and 3 SD (negative contours dashed, positive contours solid).

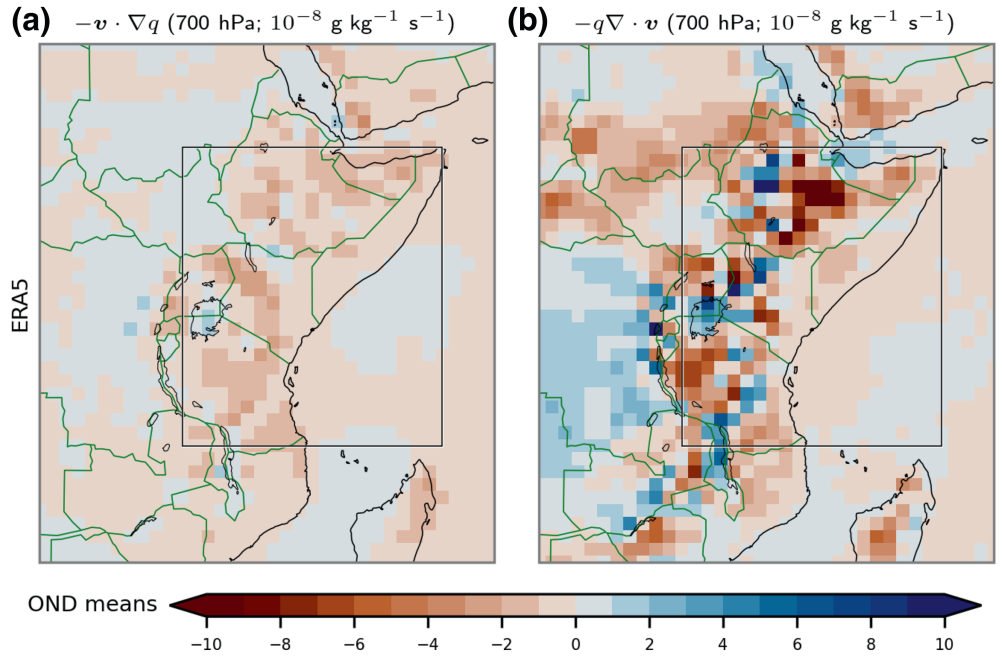

FIGURE A2 ERA5 climatology during OND of the two terms of the MFC at 700 hPa: (a) the advection term $-\boldsymbol{v} \cdot \nabla q$ and (b) the convergence term $-q\nabla \cdot \boldsymbol{v}$. The unit is specified in the captions.