

## RESEARCH ARTICLE OPEN ACCESS

# Mid-Latitude Versus Tropical Scales of Predictability and Their Implications for Forecasting

Richard J. Keane<sup>1,2</sup>  | Douglas J. Parker<sup>3,4,5,6</sup> | Etienne Dunn-Sigouin<sup>5</sup> | Erik W. Kolstad<sup>5,7</sup>  | John H. Marsham<sup>3</sup>

<sup>1</sup>Centre for Environmental Modelling and Computation, School of Earth and Environment, University of Leeds, Leeds, UK | <sup>2</sup>Met Office, Exeter, UK | <sup>3</sup>Institute for Climate and Atmospheric Science, School of Earth and Environment, University of Leeds, Leeds, UK | <sup>4</sup>School of Mathematics, University of Leeds, Leeds, UK | <sup>5</sup>NORCE Norwegian Research Centre, Bjerknes Centre for Climate Research, Bergen, Norway | <sup>6</sup>National Centre for Atmospheric Science (NCAS), UK | <sup>7</sup>Chr. Michelsen Institute, Bergen, Norway

**Correspondence:** Richard J. Keane ([r.j.keane@leeds.ac.uk](mailto:r.j.keane@leeds.ac.uk))

**Received:** 22 November 2024 | **Revised:** 17 April 2025 | **Accepted:** 30 April 2025

**Funding:** This work was supported by UK Research and Innovation as part of the Global Challenges Research Fund African SWIFT program, Grant NE/P021077/1; the Research Council of Norway through Climate Futures (grant 309562); the European Union through CONFER (Horizon 2020 grant 869730) and European Union through ACACIA (Horizon Europe grant 101137847); and the Norwegian Agency for Development Cooperation (Norad) through ARCS (grant RAF:23/006).

**Keywords:** Africa | machine learning | predictability | tropical weather forecasting

## ABSTRACT

Weather predictability varies between tropical and middle latitudes: rotational effects enable forecasts on moderate spatial scales up to 10 days in middle latitudes, while longer term predictions are less reliable; in contrast, tropical weather is challenging to predict at short lead times, but seasonal forecasts are more accurate due to the influence of larger-scale oscillations, such as slowly varying oceanic surface conditions. This behaviour has been demonstrated in previous studies, but has yet to be focused on in detail, despite its importance to the development of forecasting systems in Tropical regions. This study systematically evaluates precipitation in weather prediction models across both regions using the fractions skill score, evaluating performance at progressively longer lead times and averaging scales, and compares the results with an evaluation based on upper air error kinetic energy. The results confirm that the prediction systems perform better on smaller scales and shorter lead times at middle latitudes and on larger scales and longer lead times at tropical latitudes. A “crossover” in performance is seen at forecast lead times of 5–7 days, a result that appears to be consistent across a range of model resolutions, and occurs both when specifically comparing European and African domains and when comparing whole latitude bands. This differential pattern of model skill even occurs for machine learning-based forecast models, suggesting that it is a fundamental property of the atmosphere rather than an effect of the construction of currently used operational forecasting systems. These findings highlight the need for different forecasting methodologies in tropical regions to address the lack of short-term predictability and leverage long-term statistical predictability.

## 1 | Introduction

Numerical weather prediction (NWP) is the cornerstone of weather forecasting worldwide. In the mid-latitudes, it is central to an enterprise worth billions of pounds per year (Lazo et al. 2009). Major efforts have been made to translate these

benefits to the developing countries of the tropics, but progress has been slow (e.g., Cullmann et al. 2020; Lamptey et al. 2024). Weather prediction is economically important throughout the world, and in tropical regions, daily weather events can be severe enough to be a significant hazard to lives and economic activity at the personal level.

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

© 2025 Crown copyright and The Author(s). *Meteorological Applications* published by John Wiley & Sons Ltd on behalf of Royal Meteorological Society. This article is published with the permission of the Controller of HMSO and the King's Printer for Scotland.

It has been known since the early days of dynamical meteorology that the physics of tropical meteorology is fundamentally different from that of the mid-latitudes (e.g., Riehl 1954). This is not too hard to explain to a non-specialist: close to the equator the effects of the Earth's rotation are not as strong as they are near the poles, but the power of the Sun's heating is much greater. While extra-tropical weather systems tend to be dominated by large-scale rotating weather systems (cyclones and anticyclones) on scales of thousands of kilometres, tropical weather is dominated by convective storms (many of which are thunderstorms), with horizontal scales of tens of kilometres. The effects of rotation mean that the mid-latitude cyclones and anticyclones have some kind of stability over time periods of a few days, and our computer models can predict them with useful accuracy for a week or two. In contrast, the convective weather events dominating the tropics develop rapidly, over timescales of a few hours, and are mostly quite unpredictable, at least in their details.

The heuristic explanation of differing forecast skill between the tropics and mid-latitudes is backed up by more rigorous theory. Lorenz (1969) used scaling arguments based on simple fluid dynamics to argue that the predictability timescale of a flow depends on its length scale—he argued that for a large-scale cyclone on the scale of hundreds to thousands of km, this timescale is on the order of days, while for a convective storm it is a matter of hours. Lorenz (1969) then argued that the timescale for propagation of small-scale errors onto the larger scale circulation depends on the shape of the background energy spectrum. In particular, it is useful to compare spectra with the form  $E(k) \sim k^{-\alpha}$ , where  $E$  is kinetic energy and  $k$  is the spectral wavenumber. Here  $\alpha = 5/3$ , a  $-5/3$  gradient, characterizes three-dimensional isotropic turbulence and  $\alpha = 3$  is characteristic of geostrophic turbulence (Charney 1971). Lorenz (1969) indicated that for the  $-3$  gradient, the timescale of propagation of errors to large scales diverges with increasing spatial scale and that we can consequently increase the time over which our forecast can be skilful if we improve the accuracy of the initial conditions of the forecast. In contrast, for a flow with a  $-5/3$  energy spectrum, that timescale of error growth converges with length scale, implying that there is a finite time over which errors propagate onto the large scales and therefore over this timescale no improvement in initial conditions will improve the accuracy of the forecast. Effectively, flows with  $-5/3$  spectra have finite absolute limits to predictability, whereas flows with the  $-3$  spectrum, while still chaotic, can be predicted better if we improve the initial, observationally based analysis (Palmer et al. 2014).

Observations and analysis of models have been consistent in showing that mid-latitude energy spectra have behaviour on the  $-5/3$  gradient for mesoscale dynamics (up to about 400 km) and a  $-3$  gradient for larger scales (e.g., Nastrom and Gage 1985), consistent with the mesoscales being dominated by convective dynamics and the larger scales controlled by rotational, geostrophic dynamics. This would imply that mesoscale convective dynamics are unpredictable beyond a finite timescale, probably a matter of hours, but longer length scales can be predicted for a matter of days (in practice, a couple of weeks, Selz et al. 2022).

Recently, it has become computationally feasible to investigate this behaviour in more detail using “identical twin” experiments

involving global simulations that differ only by a small initial condition perturbation. Error growth is modelled by calculating the evolution of the difference between two simulations for a representative field (e.g., kinetic energy or potential vorticity), which can be compared with a saturation error calculated from the climatological variance of the same field. The simulations are carried out either at kilometre-scale grid spacings, to allow convective processes to be simulated explicitly (Judt 2018; Zhang et al. 2019), or include a stochastic convection scheme to simulate sub-grid scale convective error growth (Baumgart et al. 2019; Selz 2019; Selz et al. 2022). They generally confirm fast error growth on smaller scales, associated with convection and a  $-5/3$  gradient in the energy spectrum, and slower error growth on larger scales, associated with rotational effects and a  $-3$  gradient in the energy spectrum.

The identical-twin studies have generally focused on mid-latitude regions or on the global domain as a whole. However, Judt (2020) used output from such simulations to analyse error growth for tropical, mid-latitude and polar regions separately, and showed that the spectra for the tropics exhibit the  $-5/3$  slope through the mesoscale and into longer length scales, quite understandably since tropical regions do not have the same influence of rotation which provide the geostrophic,  $-3$  gradient regime seen for longer length scales in mid-latitudes. It was found that errors saturated more quickly on smaller scales in the tropics than in mid-latitudes, but more slowly in the tropics on larger scales. Judt (2020) attributed this to error growth being dominated by convective processes (which are relatively more important in the tropics) on smaller scales and baroclinic processes (which are relatively more important in mid-latitudes) on larger scales; see also Charney et al. (1981). It was also found that there was a further increase in the error growth after 10 days in the tropics, indicative of the influence of equatorial waves.

While convective regimes in the tropics typically have short predictability timescales, the tropical atmosphere benefits from the slower-changing conditions of the underlying surface (Charney et al. 1981; Shukla 1998; Bach et al. 2019). This relationship has been understood and utilized for decades in statistical prediction models, which in some cases can accurately forecast large-scale rainfall patterns based on antecedent ocean surface conditions. For example, this approach has been effective in East Africa, where the October–December rainy season is strongly influenced by El Niño–Southern Oscillation (ENSO) conditions several months in advance (Kolstad and MacLeod 2022). Moron and Robertson (2020) showed that forecast skill on weekly time scales in the tropics is stronger over oceans and where the influence of intra-seasonal and interannual modes of variability is strong. The fluid dynamical ideas of Lorenz (1969), Palmer et al. (2014) and related studies relate to the dynamics of one fluid—the atmosphere—rather than a coupled system (e.g., coupling the atmosphere with the ocean or the land surface). Through the coupling, there is evidence of impressive skill on subseasonal-to-seasonal (S2S) timescales (e.g., Hoskins 2013, de Andrade et al. 2021), if we treat the rainfall statistically, averaged in time and space. We may be able to predict generally wet and dry periods over large areas, but topography and other geographical features may give rise to large differences between regions within these large areas (Kolstad et al. 2024).

Two additional factors contribute to the relative skill of mid-latitude and tropical forecasts, namely the relative sparsity of routine weather observations in the tropics, and our lack of scientific understanding of some of the key physical processes which control tropical weather. Both of these factors were illustrated by an analysis of the European Centre for Medium-Range Weather Forecasts (ECMWF) forecasting system for the period of the AMMA (Redelsperger et al. 2006) field campaign in West Africa (Lebel et al. 2010). By re-running the forecasts with and without the full, enhanced AMMA observations, Agustí-Panareda et al. (2010) were able to show that the observations corrected the model analyses, particularly in the boundary layer, but that “the impacts of the extra AMMA data on the forecast disappear after 24 h” into the forecast. Analysis of the model errors indicated problems with the model’s representation of land-atmosphere interactions and aerosol-radiation interactions in particular. Linsenmeier and Shrader (2023) showed that forecast skill increases with *per capita* GDP, with lower national GDP often related to the relative lack of investment in observations in poorer countries, in turn degrading the skill of temperature forecasts. This is almost certainly a contributing factor, but evidence that observational density alone cannot explain forecast skill comes from the analysis of Haiden et al. (2012), who showed significantly lower forecast skill in Northern Hemisphere extratropical summer than winter, consistent with the greater significance of convective dynamics in the summer, over regions where GDP and observational density are reasonably constant over time.

On the basis of this body of theoretical work, observations and modelling studies, it appears that errors initially grow more quickly in the tropics than in mid-latitudes. This derives in part from the weak effects of planetary rotation in the tropics and the dominant effects of convection, meaning that the energy spectrum approximates the  $-5/3$  behaviour. For any latitude, there is shorter predictability on smaller scales, which Lorenz (1969) estimated to be a matter of hours for convective storms on scales of 10 km. In convective environments, which characterize the climate of many tropical regions, those shorter scales represent the critical scales of high-impact weather in the form of thunderstorms, heavy rain and squall winds. Conversely, on longer time scales, errors in the tropics grow more slowly, making it more feasible to provide generalized predictions weeks in advance. However, the extent to which this slower error growth is influenced by weaker rotational effects, equatorial waves, and stronger coupling to the ocean remains unclear.

We believe that these general concepts about the comparative skill of tropical and mid-latitude NWP forecasts are well known qualitatively within the meteorological community.

However, in our view, they are less recognized in allied fields and among agencies responsible for funding capacity building in the developing world. The aim of this paper is to make a synthesis of forecast model predictive skill across time and space scales, quantifying differences in NWP forecast skill between a European and an African domain as a function both of spatial and temporal scales, and explore its implications for African/tropical weather forecasting services. Given that the principles of forecast predictability arising from the work of Lorenz and later researchers pertain to the physics of the underlying fluid dynamics, not the method of predicting that fluid, we also compare very different forecasting systems, including models based on machine learning (ML).

Theoretical work on predictability has generally focused on other variables such as atmospheric kinetic energy and upper-level geopotential, which are chosen for their suitability for physically representing the fluid flow, rather than for their direct impacts on society. Another aim of this study is to connect this theoretical work to issues of practical forecast skill. This study therefore primarily focuses on precipitation, as it is the most societally relevant variable, particularly for agriculture, hydrological management, and disaster risk reduction, especially in Africa (e.g., Rockström and Falkenmark 2003; Beven 2012; Norris et al. 2021). We also conduct a similar evaluation for upper-level error kinetic energy in order to provide a closer link to the theoretical work and to investigate to what extent the results vary when using different methods and variables.

## 2 | Method

### 2.1 | Data

Forecasts are evaluated for five forecasting systems: GloSea6, the current operational version of the UK Met Office seasonal forecasting system, which is based on GC3 (Williams et al. 2018) and described by MacLachlan et al. (2015); the Unified Model, used for operational weather forecasts at the UK Met Office and forming the atmosphere and land component of GC3 (Walters et al. 2019); the Integrated Forecasting System (IFS), the operational weather forecasting system at ECMWF; and two ML-based systems, FuXi (Chen et al. 2023) and GraphCast (Lam et al. 2023). The model setups are summarized in Table 1. The aim is to cover multiple years for one model (GloSea6) and to then compare with other models covering a single year. Precipitation forecasts from each system are compared with IMERG precipitation observations (Huffman

**TABLE 1** | Summary of model setups evaluated in this study.

Model	Evaluation period	Forecast length evaluated	Grid spacing	Forecast frequency
GloSea6	2011–2016	30 days	$0.556^{\circ} \times 0.833^{\circ}$	approximately 8 days
GC3	2020	7 days	$0.141^{\circ} \times 0.0938^{\circ}$	12 h
IFS	2022	10 days	$0.25^{\circ} \times 0.25^{\circ}$	12 h
FuXi	2022	10 days	$0.25^{\circ} \times 0.25^{\circ}$	12 h
GraphCast	2022	10 days	$0.25^{\circ} \times 0.25^{\circ}$	12 h

et al. 2019), which are available globally (with restrictions at higher latitudes) from June 2000. GloSea forecasts of 500-hPa kinetic energy are also compared with ERA5 reanalysis data (Hersbach et al. 2023).

GloSea forecasts, in common with most seasonal prediction systems, are calibrated using hindcasts for a fixed period in the past. For each year the operational forecast is run, a set of hindcasts is produced for the period 1993–2016. Here we verify the hindcast set generated operationally from March 2021 to February 2022, restricting to the years 2001–2016 to coincide with the available observational data. The primary aim of this study is to compare performance in the tropics with that in the mid-latitudes, rather than to evaluate the differences between models. Therefore, we use the larger available hindcast dataset instead of the forecast dataset, which only covers approximately 1 year for each model version, despite the fact that forecasts are the data actually used to predict precipitation. The hindcasts are run on an N216 grid (0.556° spacing in latitude and 0.833° spacing in longitude), initialized on the 1st, 9th, 17th, and 25th of each month with seven ensemble members for each day. The simulations run for 216 days (7 months) and daily output is produced; in this study we evaluate the first 30 days.

Operational weather forecasts produced by the UK Met Office are verified here for the year 2020. These were run on an N1280 grid (0.141° spacing in latitude and 0.0938° spacing in longitude), initialized every 6 h. The simulations run for 48 h (06 UTC and 18 UTC initializations) or 168 h (00 UTC and 12 UTC initializations); in this study, we assess the longer forecasts starting at 00 UTC and 12 UTC.

Data from the other three systems were produced by ECMWF, as operational forecasts for IFS and as hindcasts for the ML systems. All three are verified here for the year 2022 and are provided on a 0.25° grid on lead times of up to 240 h. In this study, we assess the forecasts/hindcasts produced starting at 00 UTC and 12 UTC.

## 2.2 | Precipitation Verification

The forecasts are verified using the Fractions Skill Score (FSS, Roberts and Lean 2008; Roberts 2008), which is widely used for assessing precipitation forecasts (Mittermaier 2021; Keane et al. 2016; Zhao and Zhang 2018; Schwartz 2019; Cafaro et al. 2021). It is chosen here because it is specifically designed to account for the spatial scale at which the forecast is assessed, making it particularly well suited for comparing forecast performance across different scales. It is also a threshold-based measure, which makes it appropriate for assessing variables with heavily skewed distributions such as precipitation.

The FSS is defined as:

$$f(s, \theta, t) = 1 - \frac{\langle \sum_i (F_i(s, \theta, t) - O_i(s, \theta, t))^2 \rangle}{\langle \sum_i (F_i(s, \theta, t)^2 + O_i(s, \theta, t)^2) \rangle} \quad (1)$$

$$= \frac{2 \langle \sum_i (F_i(s, \theta, t) O_i(s, \theta, t)) \rangle}{\langle \sum_i (F_i(s, \theta, t)^2 + O_i(s, \theta, t)^2) \rangle}$$

where  $t$  is the forecast lead time, the sums are over all grid points in the domain and the angle brackets are averages over all available forecast start times. The quantities  $F_i$  and  $O_i$  are the fraction of grid points in a region of size  $s \times s$  points, centred on point  $i$ , where the verification quantity is above a threshold  $\Theta_i$ . For the GloSea ensemble  $F_i$  is the fraction of grid points in the region and over the 7 ensemble members (i.e., out of  $7s^2$  “points”) with values above the threshold; this follows Schwartz et al. (2010) and is the method recommended by Necker et al. (2024) for applying the FSS to ensemble forecasts (their “pFSS” method). The threshold  $\Theta_i$  is based on the climatology at the grid point  $i$ , calculated separately for the observation data set and for the model data set at each lead time, and is equal to the value corresponding to the percentile  $\theta$ . Both fields are padded with zeros outside the domain to allow for grid points within  $s/2$  points of the edge of the domain. This calculation of the FSS was executed using a method based on pysteps (Pulkkinen et al. 2019).

Forecasts are verified based on daily accumulations of precipitation, at lead times up to 30 days for GloSea, 7 days for the Met Office weather forecasts, and 10 days for the simulations run at ECMWF. This was based simply on the availability of data, with an upper limit set to 30 days as initial investigations found that the scores for GloSea were relatively constant with lead time by this stage.

The verification is carried out over two regions, both with longitude ranges from 30°W to 60°E, and with latitude ranges from 15°S to 15°N, representing tropical latitudes, and 30°N to 60°N, representing mid-latitudes. The tropical region corresponds to most of tropical Africa, where the performance of short-range weather forecasts has been found to be particularly poor (Vogel et al. 2020). The northern region corresponds to most of Europe and the very north of Africa, and enables a comparison at the same longitudes between two regions clearly demarcated by the Sahara Desert between them. To enable a direct comparison between the forecasts and observations, the observations are re-gridded onto the forecast grid using area weighting. An alternative method of interpolating both fields onto a 1° grid was applied to the Met Office seasonal hindcasts and found not to substantially affect the results. For some of the models a global evaluation is also carried out. The method here is the same except that the zero-padding outside the domain is only required near the poles, in the latitude direction. For GloSea the global evaluation is applied only to a single ensemble member. Results for GloSea are also compared with a method applying a seasonally varying threshold. In particular, this is defined, for each day of year  $d$ , based on all forecast or observed values, at the relevant grid point, with a day of year in the range  $d - 45$  to  $d + 45$  (where  $d \pm 365$  is the same day of year as  $d$ ).

The number of grid points  $s$  constituting the averaging region is converted to a spatial scale in km by considering the area of each region and the grid spacing of each model. The average area represented by each grid point is given by the area of the region  $A$  divided by the number of grid points  $n$ . These are defined as

$$A = r_E^2 (\sin \lambda_{\max} - \sin \lambda_{\min}) (\phi_{\max} - \phi_{\min}) \quad (2)$$

and

$$n = \frac{\lambda_{\max} - \lambda_{\min}}{\delta_\lambda} \frac{\phi_{\max} - \phi_{\min}}{\delta_\phi} \quad (3)$$

where  $r_E$  is the radius of the Earth (taken as 6371 km),  $\lambda_{\min}$  and  $\lambda_{\max}$  are the minimum latitude (here  $-\pi/12$  radians ( $-15^\circ$ ) for the tropical region and  $\pi/6$  ( $30^\circ$ ) for the mid-latitude region) and maximum latitude (here  $\pi/12$  ( $15^\circ$ ) for the tropical region and  $\pi/3$  ( $60^\circ$ ) for the mid-latitude region), respectively,  $\phi_{\min}$  and  $\phi_{\max}$  are the minimum longitude (here  $-\pi/6$  or  $-30^\circ$ ) and maximum longitude (here  $\pi/3$  or  $60^\circ$ ), respectively, and  $\delta_\lambda$  and  $\delta_\phi$  are the model grid spacings in, respectively, the latitude and longitude directions. The length scale  $l$  is then given by

$$l = s \sqrt{\frac{A}{n}} \quad (4)$$

In practice, the physical length scale varies with latitude within each region, but this method provides a means of comparing overall results from the two regions directly.

### 2.3 | Error Kinetic Energy

In order to provide a closer link with idealized studies on predictability, we also investigate errors in kinetic energy. This evaluation is carried out over entire latitude bands (i.e., the full range of longitudes, and latitudes varying between  $-\pi/12$  and  $\pi/12$  (tropics) and between  $\pi/6$  and  $\pi/3$  (mid-latitudes)). The error kinetic energy is defined, analogously to Judt (2018, 2020), as

$$E = \frac{1}{2} \left\{ (u_f - u_a)^2 + (v_f - v_a)^2 \right\} \quad (5)$$

where  $u$  and  $v$  are the horizontal velocity components at a height of 500 hPa and the subscripts refer to the forecast (f, here from the GloSea hindcasts for a single member) and analysis (a, here from ERA5, re-gridded onto the GloSea grid).  $E$  is a function of forecast lead time, and is normalized by a saturation value, representing the error that would result from simply selecting a random forecast from the whole distribution. This saturation value is defined here by averaging  $E$  over the longest lead times, corresponding to 199–216 days for the GloSea hindcasts.

Spectra for the error kinetic energy are created using Fourier transforms in longitude:

$$F(t, \lambda, \kappa) = \sum_{n=1}^{2\pi/\delta_\phi} e^{-i\kappa n \delta_\phi} E(t, \lambda, \phi = n \delta_\phi) \quad (6)$$

The longitude wavenumbers are given by  $\kappa_j = 2\pi/j$ . These can be converted to length scales  $l_j = 2\pi(r_E \cos \lambda)/j$ . Then the spectral value  $P$  for a length scale bin varying from  $l_{\min}$  to  $l_{\max}$  is given by:

$$P(l_{\min} \rightarrow l_{\max}) = \left\langle |F(t, \lambda_k, l_j)| \right\rangle \quad (7)$$

where the angle brackets now represent an average over all available forecast start times and all  $j$  and  $k$  such that  $l_{\min} \leq 2\pi(r_E \cos \lambda_k)/j < l_{\max}$ .

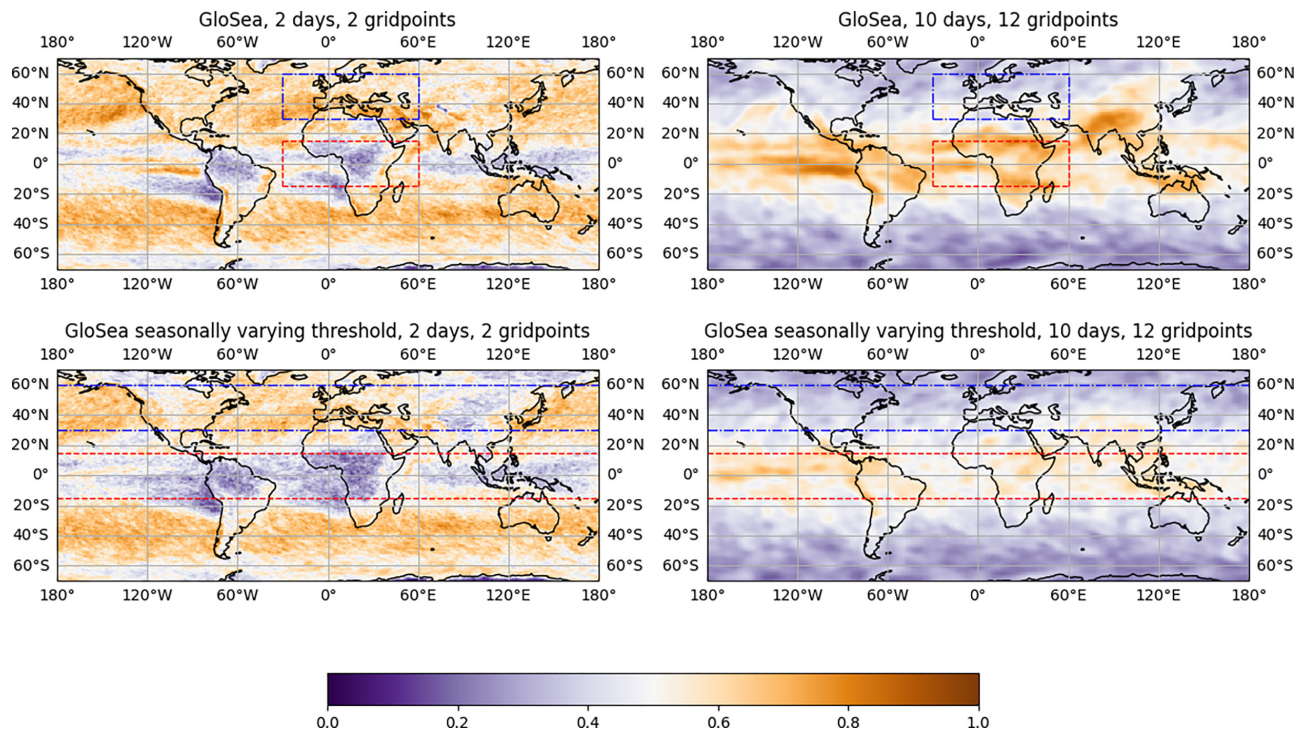
## 3 | Results

### 3.1 | Fraction Skill Score Maps

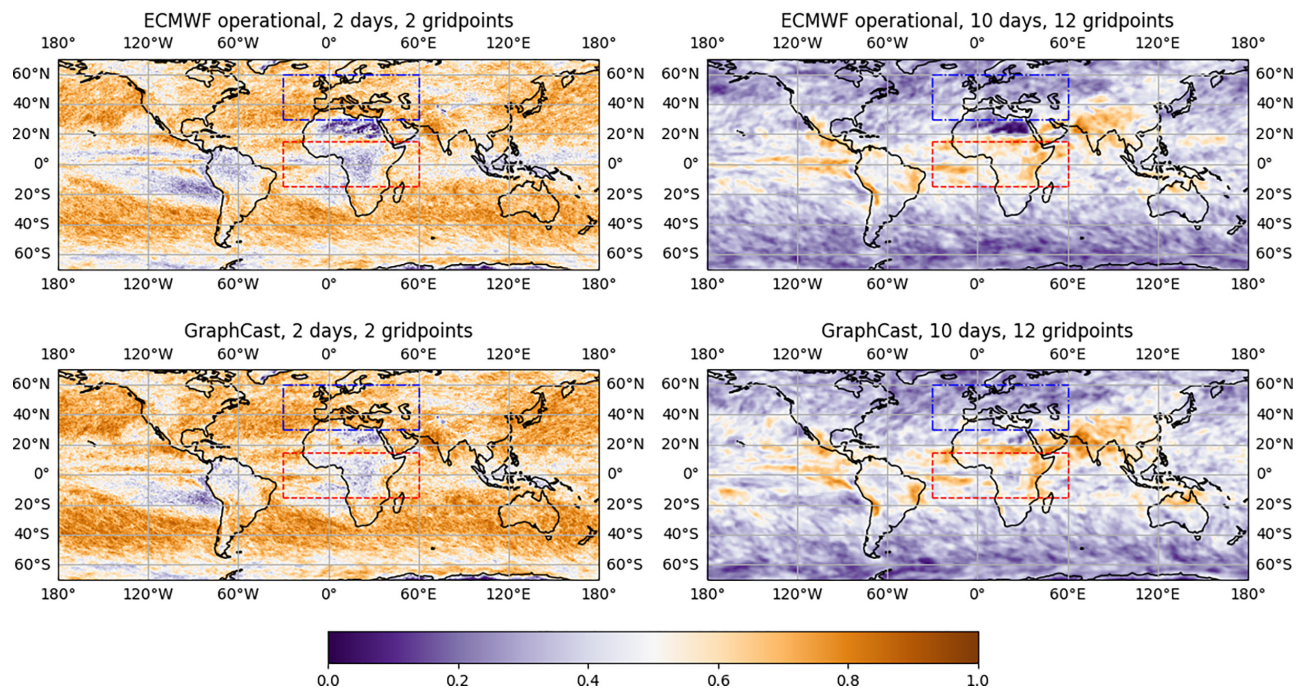
Figure 1 (focusing first on the upper panels) shows how the FSS varies in space for a pair of shorter (in terms of lead time) and smaller (in space) scales and a pair of longer/larger scales for GloSea6. Here, the FSS was calculated by not applying the sums over grid points in Equation (1). Scores would generally be expected to be higher on larger spatial scales, where the atmosphere is better resolved by the model and errors generally grow more slowly, and at shorter lead times, where the model has had less time to depart from its initial state and should be closer to the observed atmospheric state. The panels on the left show that the model performs better on the shorter/smaller scales (higher FSS values) at mid-latitudes than in the tropics, indicating that it is easier to predict local precipitation 2 days ahead in Europe than in tropical Africa. Conversely, the right-hand panels demonstrate that the performance is better on longer/larger scales in the tropics than in the middle latitudes, indicating that predicting spatially aggregated regional weather a week or more ahead is easier in tropical Africa than in Europe.

As well as this broad variation in latitude, there is also some longitudinal variation, particularly at tropical latitudes. For example, on shorter/smaller scales, the forecasts generally perform better over ocean than land, with the exception of the ocean to the west of South America and southern Africa. This is in broad agreement with Vogel et al. (2020) (comparing with their Figure 5, which includes verification of heavy precipitation), who mention that the anomalously relatively poor forecast over the “oceanic deserts”, where there is very little precipitation, could partly be due to issues with the observational data. We note here that very low FSS values are also seen over Antarctica and Greenland, where there is also very little precipitation and IMERG coverage is relatively sparse (for which reason these regions are not included in these plots). On longer/larger scales, there is less longitudinal variation, although FSS values are particularly high over the equatorial Pacific, where high skill was associated by Moron and Robertson (2020) with strong interannual variability.

A particularly striking aspect of the longitudinal variation is how high the FSS values are for India on both sets of scales, despite the fact that the Indian monsoon is one of the most challenging meteorological phenomena to simulate. One possible explanation is that precipitation has particularly strong seasonal variation over India, so by using a constant threshold, the model can attain a large FSS simply by predicting values above the threshold quite often during the monsoon season and very rarely at other times, even if the actual values predicted during the monsoon are not particularly accurate. To test this hypothesis, FSSs using a seasonally varying threshold are plotted in the lower row of Figure 1. The FSS values are indeed much lower over India for both sets of scales. The overall variation is smoother, with lower values in the Tropics at longer/larger scales, but the general behaviour in terms of latitude variation is very similar. Interestingly, values are much lower over northern Asia and northern North America at shorter/smaller scales. This should be investigated further in future work; it may be due to the increased importance of initial conditions at mid-latitudes, so that forecasts are poorer over more sparsely populated regions with fewer observations.



**FIGURE 1** | Spatial maps of FSS for the Met Office seasonal forecasting system (hindcasts) for the period 2001–2016, for lead time and averaging area shown in the panel title, for a threshold of 90%, defined separately for hindcast and observations, and for each grid point. In the top row the threshold is constant in time, based on values over the whole verification period; in the bottom row the threshold varies with season, based on values in a 90-day day-of-year window centred on the relevant time (for all years). The boxes in the upper panel show the regions investigated in this study and the lines in the bottom panel show the latitude bands investigated in this study (red dashed: Tropics; blue dash-dotted: Mid-latitudes).



**FIGURE 2** | Spatial maps of FSS for the ECMWF operational forecasting system (upper panels) and GraphCast (lower panels) for the year 2022, for lead time and averaging area shown in the panel title, for a threshold of 90%, defined separately for forecast and observations, and for each grid point. The boxes show the regions investigated in this study (red dashed: Tropics; blue dash-dotted: Mid-latitudes).

Figure 2 shows similar FSS maps for the ECMWF operational forecast and GraphCast, one of the machine-learning-based models. These both look similar to the maps from GloSea, but with more small-scale spatial variation as would be expected

from higher-resolution models. In particular, the sets of maps shown for the two models in Figure 2 are very similar, suggesting that spatial patterns of predictability are due to fundamental properties of the atmosphere and its surface forcing, and are not

strongly affected by the structure of the model used to simulate them.

### 3.2 | Comparison of FSSs Between Europe and Tropical Africa

This subsection investigates the behaviour illustrated in Figures 1 and 2 more systematically by looking at domain-averaged FSS values for different ranges of spatial scales and forecast lead times in various models. The two domains are over Tropical Africa and Europe, marked by the red and blue lines in Figure 2, respectively.

To investigate the differing forecast skill systematically, FSS values were calculated for averaging regions of sizes  $s = \{1, 2, 3, 5, 8, 12, 18, 27, 42\}$  (increasing by a factor of roughly 1.5 each time) and percentile thresholds  $\theta = \{75, 90, 95\} \%$ , where the maximum corresponds to 100%. These results are presented in Figure 3 for GloSea6, showing each region separately and the differences between the two regions. Note that the size of the grid boxes varies with latitude, and the nominal length scale for each region is given by Equation (4). This is the value used in the left two columns. For the right column, representing the difference between the two regions, the FSS values for the tropical region have been interpolated to match the equivalent spatial scales in the mid-latitude region, so that a direct comparison can be made. In the difference plots, blue indicates better performance at mid-latitudes, while red indicates better performance in the tropics. This description also applies to Figures 4, 5 and 7.

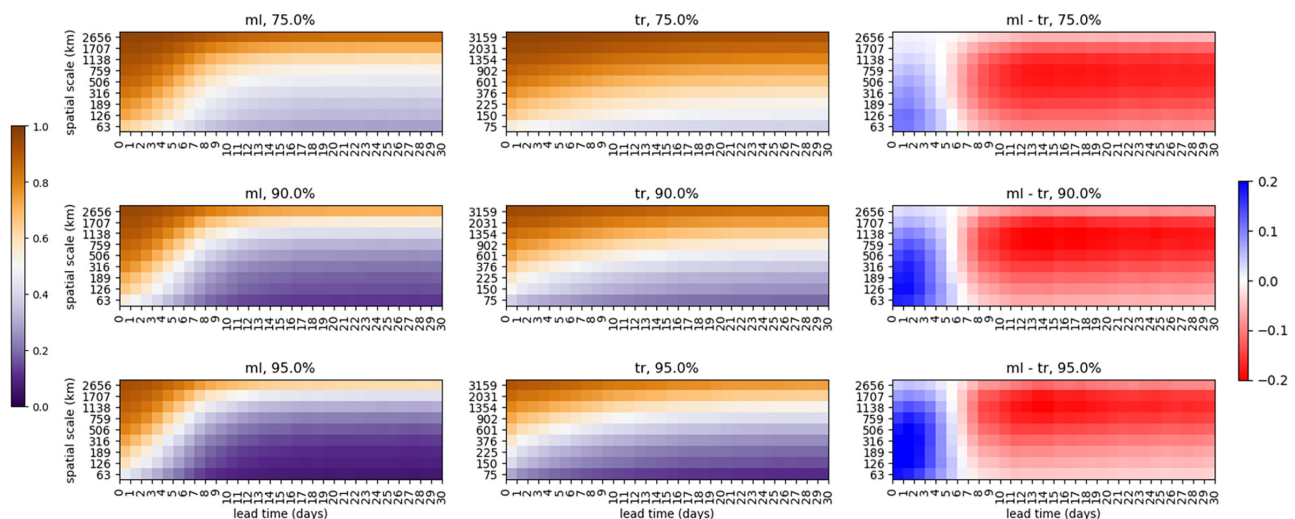
Both at mid-latitudes and in the tropics, the FSS values for GloSea6 (Figure 3, left and middle columns) show scores increasing with spatial scale and decreasing with lead time, as would be expected. The scores for short lead times are generally better at mid-latitudes, but drop off more quickly with lead time than in the tropics. The scores appear to be approximately independent of lead time after 10–15 days. The variation of FSS with spatial and temporal scale is mostly similar for the three thresholds shown.

The difference plots in the right column of Figure 3 show a clear separation in relative model performance, with the mid-latitudes more predictable at shorter lead times and smaller scales and the tropics at longer lead times and larger scales. The separation is stronger on the temporal scale, with the difference changing sign at lead times of about 5–7 days on all spatial scales, but this “crossover” lead time is slightly shorter for larger spatial scales compared to smaller scales, and the improvement in the tropics is generally stronger (or the degradation weaker) on larger scales. The crossover lead time is slightly earlier for the lowest threshold, in agreement with the findings of Gehne et al. (2022), who found the crossover time with FSS to increase with threshold when comparing similar tropical and mid-latitude bands for a different model. At the largest spatial scales (above 2000 km), there is little difference between the regions, with both showing very high FSS values. This suggests that, on this scale, the proportion of a region experiencing precipitation above a given percentile threshold is relatively stable and easier to predict; it may also be a consequence of the total size of the domain investigated here.

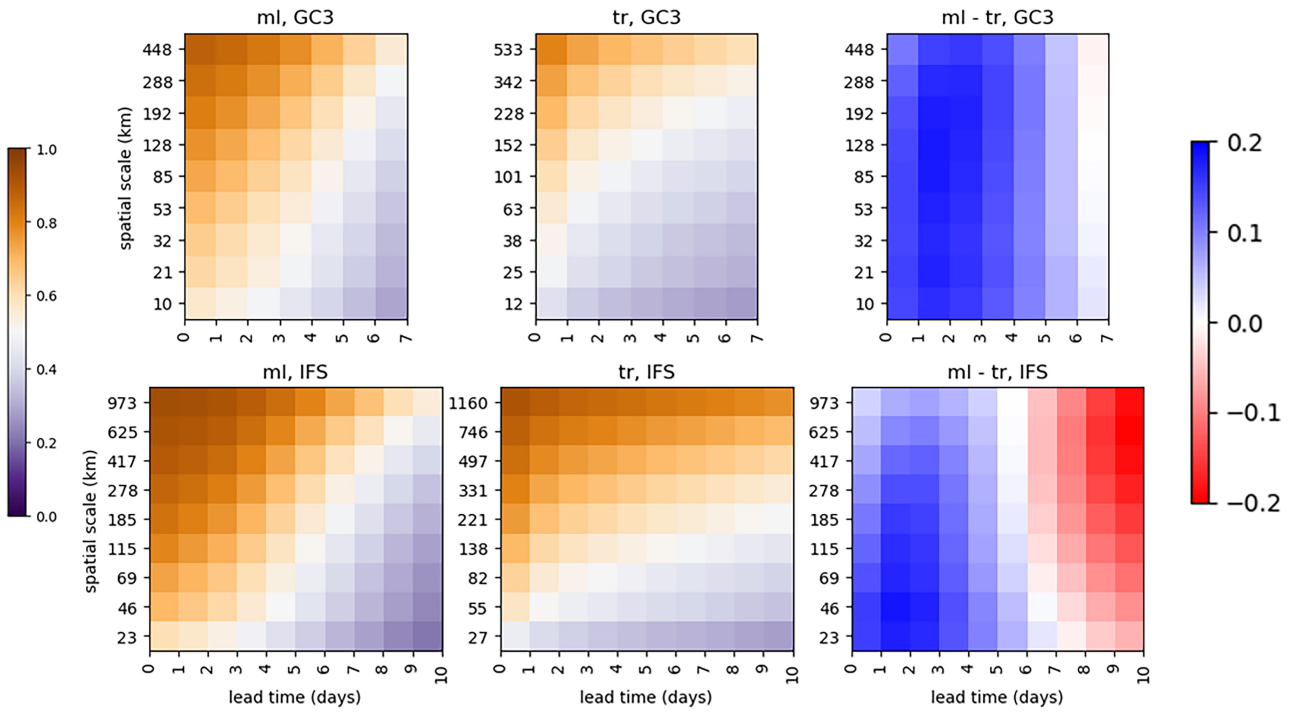
FSS values are plotted in the same way in Figure 4, for operational NWP forecasts from the Met Office (GC3) and ECMWF (IFS). The behaviour is broadly similar, with a reduction in performance with increasing lead time that is slower in the tropics than at mid-latitudes. There is a similar crossover in performance, with the difference in scores depending strongly on lead time and weakly on spatial scale.

This demonstrates that the general result—models perform better at shorter and smaller scales in mid-latitudes and at longer and larger scales in the tropics—holds both for a single model operating at different spatial and temporal scales and for models with different underlying structures developed by various operational centres. The crossover occurs at lead times of 5–7 days in both weather models, which is also similar to that for GloSea6.

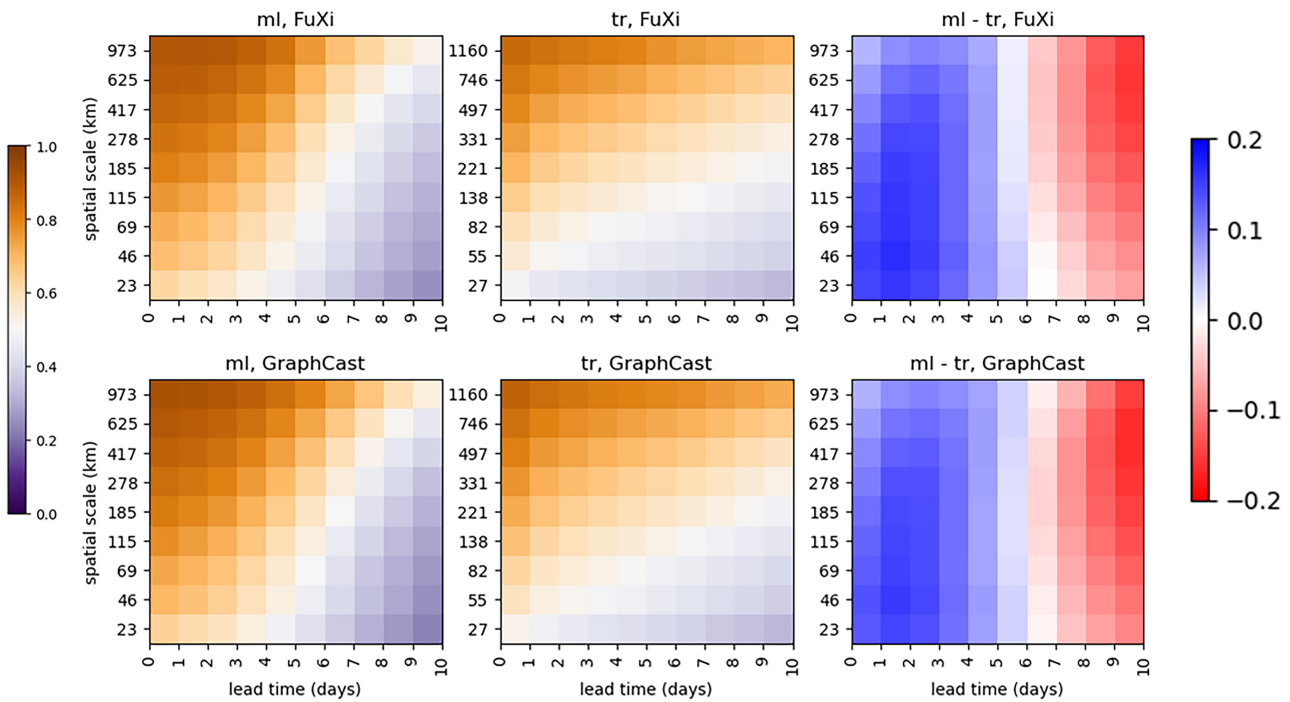
The same verification for two machine-learning-based models is shown in Figure 5, for FuXi and GraphCast forecasts. These



**FIGURE 3** | Fractions Skill Scores for Met Office seasonal hindcasts (GloSea6), for the period 2001–2016, shown as a function of spatial scale and lead time. The left column displays results for mid-latitudes, the middle column for the tropics, and the right column shows the difference (mid-latitudes minus tropics). Each row represents a percentile threshold, indicated in the title above the panels. For each panel, the vertical axis represents spatial scale in kilometres, and the horizontal axis represents lead time in days.



**FIGURE 4** | Fractions skill scores for Met Office operational weather forecasts (GC3, top row) from 2020 and ECWMF operational forecasts (IFS, bottom row) from 2022 as a function of spatial scale and lead time at mid-latitudes (left column) and tropics (middle column), and the difference between them (mid-latitudes minus tropics, right column), for a percentile threshold of 90%.



**FIGURE 5** | Fractions skill scores for ECMWF ML-based hindcasts (FuXi, top row, and GraphCast, bottom row) for 2022 as a function of spatial scale and lead time at middle (left column) and tropical (middle column) latitudes, and the difference between them (middle latitudes minus tropics, right column), for a percentile threshold of 90%.

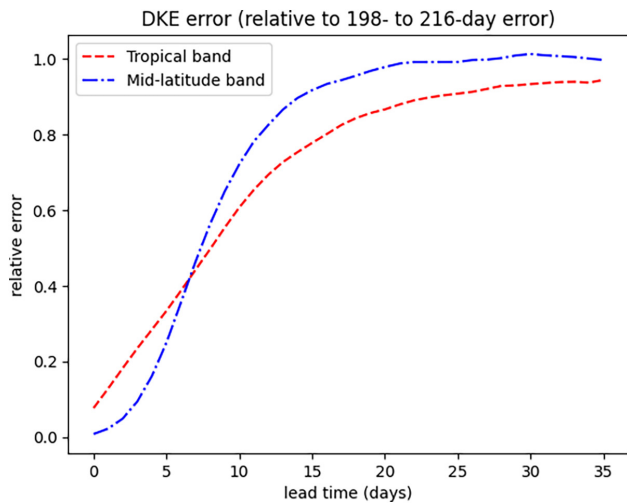
show very similar dependence on lead time and spatial scale to the NWP models, with a similar crossover time of 5–7 days. This is a very important result, as it provides strong evidence that the difference in performance is due to fundamental differences in the scales of predictability of the atmosphere at different latitudes, rather than an aspect of the particular

construction of dynamical models. New machine-learning-based weather forecasting models are rapidly being developed, and it will be interesting in future work to investigate whether any new models can produce forecasts that are not subject to this constraint, or whether it will remain impossible to produce models that perform as well in tropical regions as

in mid-latitude regions on short time scales, and vice-versa on long time scales.

### 3.3 | Comparison Between Precipitation FSS Evaluation and Error Kinetic Energy Evaluation

In this subsection, the results regarding precipitation forecasts are compared with those obtained using a different metric, namely 500-hPa error kinetic energy, and a different method for comparing different horizontal scales, namely taking Fourier spectra of the error fields. Error kinetic energy is calculated



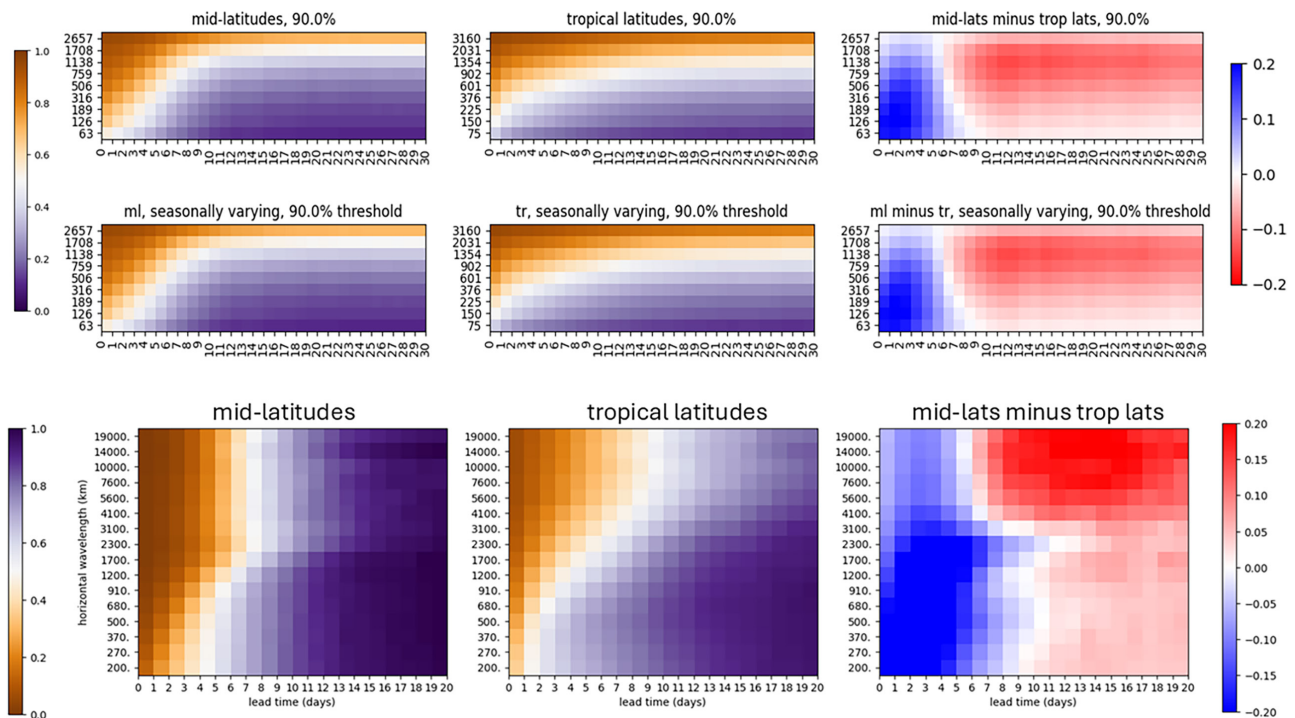
**FIGURE 6** | 500-hPa error kinetic energy for Met Office seasonal hindcasts averaged over the period 2001–2016, relative to the value at the end of the forecast. The tropical band represents latitudes from 15°S to 15°N and the mid-latitude band represents values from 30°N to 60°N.

following Judt (2020) and compared for two full latitude bands (a tropical band with latitudes from 15°S to 15°N and a mid-latitude band with values from 30°N to 60°N). This quantity is plotted as a function of forecast lead time for the GloSea hindcasts in Figure 6.

As with precipitation FSS, the errors are initially larger in the tropical band, but then grow more quickly in the mid-latitude band so that there is a crossover time when the mid-latitude error overtakes the tropical error. This crossover occurs at about 7 days, similar to that seen for precipitation FSS, and also similar to Judt (2020) (comparing panels a and b of their Figure 3). It is interesting to note that an earlier “identical twin” study by Straus and Paolino (2009) finds a later crossover time, at about 15 days (their Figure 2a,b). They evaluated zonal wind error variance, corresponding to only the first term on the right-hand side of Equation (5), but perhaps a more important difference is that they used a model with a much coarser grid spacing than Judt (2020), so it may not have correctly captured the smaller-scale error growth.

Figure 7 shows verification of the GloSea hindcasts over full latitude bands, as a function of spatial scale, using two different methods. The top two rows show FSS values for a constant (in time) threshold and for a seasonally varying threshold. These are very similar to each other and to Figure 3, with differences only in the precise crossover lead time at which the values in the tropics overtake those at mid-latitudes. This indicates that, while using a seasonally varying threshold can affect the spatial patterns of FSS, it does not strongly affect the general behaviour across length and time scales.

Also shown in Figure 7 are error kinetic energy values, separated into spatial scales using Fourier spectra. The behaviour



**FIGURE 7** | Comparison between mid-latitude (30°N–60°N) and tropical (15°S–15°N) bands for Met Office seasonal hindcasts, measured in terms of FSS (top two rows) and error kinetic energy (bottom row) for the period 2001–2016.

is overall very similar; some of the smaller differences could be due to the fact that the Fourier transform method only accounts for scale in the meridional direction, whereas the FSS aggregation method accounts for scale in both horizontal directions, and that the Fourier transform method accounts only for information at the relevant scale, whereas the FSS aggregation method at a given scale accumulates information on all scales up to and including that scale. In terms of the differences, there is a similar crossover at about 7 days for all spatial scales, although the behaviour is somewhat noisier for the error kinetic energy than for the FSS.

A clear difference between the two methods is that, with increasing lead time, scores tend to unity (i.e., poor) on all scales for the error kinetic energy method, while they tend to values increasing from near zero (i.e., poor) to unity (i.e., good), with increasing spatial scale, for the FSS method. Generally, forecasts regress to being uncorrelated with increasing lead time (i.e., essentially random), and the discrepancy between the two methods at longer lead times can be understood by considering the scores they produce for random forecasts. For the error kinetic energy method, the score is simply unity at all scales, as it is here defined as the ratio of the error to that of the long-lead time forecast. For the FSS, it can be shown that a random forecast will produce a score increasing from  $100\% - \theta$  for no averaging (i.e., calculated at the grid scale) toward 100% for increasing spatial averaging; for example, an averaging region of 625 points and a threshold  $\theta = 90\%$  yields an FSS very close to unity (Skok and Roberts 2016, see in particular their Figure 1a and [equation 13]).

## 4 | Conclusions and Discussion

In this article we have quantified patterns of forecast skill according to temporal lead time and spatial aggregation scale, comparing middle latitudes with tropical latitudes. In simple terms, the middle latitudes have better predictability on small scales over a few days while the tropical latitudes start to have better predictability, if spatially averaged, beyond about a week. This qualitative pattern of differences was probably already understood by most meteorologists, and does appear in the literature (e.g., Zhu et al. 2014; Buizza and Leutbecher 2015; Gehne et al. 2022). In this study, we have demonstrated a timescale beyond which the tropical domain starts to be more statistically predictable than the mid-latitude domain at around 5–7 days, in broad agreement with these previous studies, and shown that the patterns of differences are remarkably similar across different forecast model systems.

We believe, with a basis in theory, that these patterns of differing predictive skill between latitudes are a fundamental property of our world, related to the rotation of the Earth and the inherent chaos of the atmosphere as a convecting fluid. That understanding has been backed up by the observation that relative forecast skill between the tropics and mid-latitudes has a very similar pattern across the range of models we tested, including traditional NWP models and new ML-based codes. We have also shown that very similar results are obtained using two very different evaluation methods (fraction skill score for precipitation and spectral analysis of upper atmosphere

error kinetic energy). It therefore seems likely that the overall behaviour is rather independent of which methods and variables are used, but it will be interesting to investigate in more detail how the precise crossover time depends on these factors, and also investigate the spatial scale dependence shown in Figure 7 in more detail. We note that Gehne et al. (2022) obtain similar, but not identical, crossover times for different verification metrics of precipitation, surface and upper air wind, surface humidity, surface temperature, and upper air geopotential height.

Although the qualitative behaviour is probably already well-known to meteorologists, there are still many open questions around the possible predictive skill which could be achieved in the future for different parts of the world. Are there absolute, finite-time limits of predictability (the “real” butterfly effect of Palmer et al. (2014)), or can we continue to push the limits of prediction through improved observations and modelling? Our theoretical understanding of the limits of predictability (e.g., from Lorenz 1969, and following work) is based on simple models of the chaotic atmosphere in isolation, while our general understanding of the longer-term predictability of tropical weather invokes the slower variations of the underlying land and ocean surfaces. Bach et al. (2019) have started to quantify the improvements in predictability conferred by the ocean on the tropical atmosphere.

We should also not forget that the global atmosphere is connected: our tropical and mid-latitude domains interact on timescales from days upwards. The arguments of Lorenz (1969) show that planetary-scale circulations have longer error-growth timescales than smaller systems, and Hoskins (2013) gives a number of examples where global connections confer predictability on timescales of a week or two. More specifically for the African domain, a number of authors have shown how North and Central African rainfall variability can be driven by mid-latitude wave trains (e.g., Knippertz 2004; Bekele-Biratu et al. 2018; Vizi and Cook 2014; Ward et al. 2023). In other words, it is thought that the good predictability of weather systems over the North Atlantic and Europe can, in the right circumstances, drive weather in Africa with a similar predictability timescale.

The paper also shows geographic patterns of forecast skill on different spatial and temporal scales (Figures 1 and 2). While these patterns quite dramatically show the contrasts between continents, they also offer provocative suggestions as to smaller-scale controls on predictability. Topography and land-sea differences seem in some places to improve predictive skill. Exploring these differences, alongside the seasonality of skill (with an expectation of reduced European skill in more convective summer conditions) would be an interesting next step.

Answering such questions has become increasingly important with the recent revolution in ML-based forecasting systems. A key question is whether the predictability limits seen in current NWP models are fundamental constraints that ML-based systems cannot surpass, or whether substantial improvements in lead time and spatio-temporal localization of accurate forecasts are still achievable.

However, it is important to recognize that current global ML-based forecasts are trained on NWP model analyses, meaning their accuracy—when evaluated against observations—is unlikely to exceed that of the NWP analyses that trained them. Recent work by Vonich and Hakim (2024) suggests that initial condition uncertainties are a more significant constraint on predictability than model uncertainties and that applying ML techniques to optimize initial conditions from observational data can yield substantial improvements over traditional data assimilation. This suggests that ML methods could enhance both initial condition estimation and forecast generation.

Selz and Craig (2023) showed that an ML-based model fails to capture the initial rapid error growth of very small perturbations, unlike a physically based system, but for errors comparable to current initial condition uncertainties, ML models exhibit similar error growth to physics-based systems. This aligns with the behaviour observed in this study and suggests that as improved observation and assimilation technologies reduce initial condition uncertainties, the nature of error growth in ML-based models may diverge further from that in NWP models.

Since ML-based models are computationally cheaper to run than full NWP systems, their continued development and refinement could lead to their gradual adoption. Once trained, the reduced cost of running these models, and their versatility for locally-focused post processing, also offers the chance that competitive models will be run within African centres within the coming years (Vaughan et al. 2024).

While the qualitative forecast differences between Europe and Africa are well known to meteorologists, they are certainly not well understood beyond the meteorological community, and we would argue that better communication of the science with policy-makers is needed. It is widely accepted that African weather prediction services are not as effective as they need to be to meet the needs of forecast users (e.g., Cullmann et al. 2020; World Bank 2021; Tzachor et al. 2023) and it has also been argued that progress may be achievable within a few years (Parker et al. 2022). Rectifying the situation is the focus of some major international development programs (e.g., Early Warnings for All, Egerton et al. 2022), in which there is an acknowledged need for “co-production” of services with their users (e.g., Carter et al. 2019). The results of this paper are strong evidence of the requirement, in these global efforts, for different approaches to the weather prediction enterprise in Africa from those approaches that are established for mid-latitudes. Initiatives to exploit rainfall forecasts in Africa are likely to be disappointed if new services are based on unrealistic expectations of daily forecast skill at the local scale. African and tropical prediction systems need to deal with the relative unpredictability of convective rainfall on the small scale and to exploit the relatively high statistical predictability of tropical weather on the larger scales.

What does this disparity in methods between mid-latitudes and tropical Africa mean in practice? The mid-latitude approaches still tend to focus on “deterministic” prediction of the weather, despite the wealth of ensemble data now available, and this is somewhat justified by models having acceptably good skill at 1 day’s lead time, on the smaller spatial scales of interest to most

users (as seen in Figures 3 and 4). In contrast, African forecasting needs to recognize that beyond a few hours, all convective forecasts are necessarily probabilistic, giving users an increased or decreased chance of rain. Probabilistic forecasts are more difficult to interpret and use, and it is unfortunate that for populations most vulnerable to the effects of climate change (in the tropics) the successful “deterministic” approach is not so viable. However, the forecasts do still provide very useful information and the challenge for the tropics is in refining the understanding of this available skill, using ensembles, statistics, and ML to produce more accurate (probabilistic) predictions, and in working with users so that they can exploit probabilistic skill effectively in their decision-making. This is also the case for S2S timescales; it is ironic that S2S models are being used more in the mid-latitudes, where they do not work so well, while we are failing to exploit the greater opportunities in tropical Africa for skilful forecasts at these longer ranges of a few weeks and a season. Some progress in using probabilistic and data-driven post-processing to improve S2S forecast skill has, for example, been made by de Andrade et al. (de Andrade et al. 2021) and Bach et al. (2024).

Ultimately, we should recognize that global prediction models—whether traditional NWP or ML-based—are necessary but insufficient for tropical forecasting. With skill emerging at broader spatial and temporal scales, effective prediction requires statistical post-processing to refine and downscale raw model output. Given the low confidence in convective forecasts beyond a day, nowcasting is of the highest priority in Africa, providing critical warnings of storms as they move and evolve. Yet, few African countries currently have operational nowcasting capability, despite its potential for improving short-range forecasting (Roberts et al. 2022). Tackling these gaps will require co-production of services in a way that reaches into the forecast office, the computational products, the technical infrastructure, communications, and the standard operating procedures of forecasters.

## Author Contributions

**Richard J. Keane:** writing – original draft, writing – review and editing, visualization, methodology, formal analysis, software, validation. **Douglas J. Parker:** conceptualization, writing – original draft, writing – review and editing, methodology. **Etienne Dunn-Sigouin:** methodology, writing – review and editing. **Erik W. Kolstad:** methodology, writing – review and editing. **John H. Marsham:** methodology, writing – review and editing.

## Acknowledgments

This work was supported by: UK Research and Innovation as part of the Global Challenges Research Fund African SWIFT program, Grant NE/P021077/1; the Research Council of Norway through Climate Futures (grant 309562); the European Union through CONFER (Horizon 2020 grant 869730) and ACACIA (Horizon Europe grant 101137847); and the Norwegian Agency for Development Cooperation (Norad) through ARCS (grant RAF:23/006). The IMERG data were provided by the NASA Goddard Space Flight Center and PPS, which develop and compute IMERG as a contribution to GPM, and are archived at the NASA GES DISC. ERA5 data were obtained using the Copernicus Climate Change Service information [2025]. Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains. ECMWF operational forecast data Copyright statement: Copyright “© 2025 European Centre for Medium-Range Weather Forecasts (ECMWF)”. Source:

www.ecmwf.int License Statement: ECMWF open data is published under a Creative Commons Attribution 4.0 International (CC BY 4.0). <https://creativecommons.org/licenses/by/4.0/>. Disclaimer: ECMWF does not accept any liability whatsoever for any error or omission in the data, their availability, or for any loss or damage arising from their use. The FuXi and DeepMind GraphCast forecast models are licensed under the Apache License 2.0, see <https://github.com/tpys/ai-models-fuxi/blob/main/LICENSE> and <https://github.com/google-deepmind/graphcast/blob/main/LICENSE>, respectively.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Data Availability Statement

Due to intellectual property right restrictions, we cannot provide either the source code or the documentation papers for the Met Office Unified Model (MetUM). For information on accessing MetUM data, see <https://www.metoffice.gov.uk/research/approach/modelling-systems/unified-model>. JULES is available under licence free of charge. For further information on how to gain permission to use JULES for research purposes, see <https://jules.jchmr.org/>. The model code for NEMO v3.4 is available from the NEMO Consortium and can be downloaded from their repository (<https://forge.ipsl.jussieu.fr/nemo/chrome/site/doc/NEMO/guide/html/install.html>; <https://doi.org/10.5281/zenodo.1464816>). The model code for CICE is freely available from the CICE Consortium, a group of stakeholders and primary developers of the Los Alamos sea ice model, and can be downloaded from the CICE repository (<https://github.com/CICE-Consortium/CICE/wiki>). ERA5 data were obtained using Copernicus Climate Change Service information [2025]. Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains. For information about accessing ECMWF operational forecast data, see <https://www.ecmwf.int/en/forecasts/dataset/operational-archive>. For information about running the FuXi forecast model, see <https://github.com/tpys/ai-models-fuxi>. For information about running the GraphCast model, see <https://github.com/google-deepmind/graphcast>.

## References

- Agustí-Panareda, A., A. Beljaars, C. Cardinali, I. Genkova, and C. Thorncroft. 2010. "Impacts of Assimilating Amma Soundings on Ecmwf Analyses and Forecasts." *Weather and Forecasting* 25, no. 4: 1142–1160. <https://doi.org/10.1175/2010WAF222370.1>.
- Bach, E., S. Motescharrei, E. Kalnay, and A. Ruiz-Barradas. 2019. "Local Atmosphere–Ocean Predictability: Dynamical Origins, Lead Times, and Seasonality." *Journal of Climate* 32, no. 21: 7507–7519. <https://doi.org/10.1175/JCLI-D-18-0817.1>.
- Bach, E., V. Krishnamurthy, S. Mote, et al. 2024. "Improved Subseasonal Prediction of South Asian Monsoon Rainfall Using Data-Driven Forecasts of Oscillatory Modes." *Proceedings of the National Academy of Sciences* 121, no. 15: e2312573121. <https://doi.org/10.1073/pnas.2312573121>.
- Baumgart, M., P. Ghinassi, V. Wirth, T. Selz, G. C. Craig, and M. Riemer. 2019. "Quantitative View on the Processes Governing the Upscale Error Growth up to the Planetary Scale Using a Stochastic Convection Scheme." *Monthly Weather Review* 147, no. 5: 1713–1731. <https://doi.org/10.1175/MWR-D-18-0292.1>.
- Bekele-Biratu, E., W. M. Thiaw, and D. Korecha. 2018. "Sub-Seasonal Variability of the Belg Rains in Ethiopia." *International Journal of Climatology* 38, no. 7: 2940–2953. <https://doi.org/10.1002/joc.5474>.
- Beven, K. 2012. *Rainfall-Runoff Modelling: The Primer*. 2nd ed. John Wiley & Sons.
- Buizza, R., and M. Leutbecher. 2015. "The Forecast Skill Horizon." *Quarterly Journal of the Royal Meteorological Society* 141, no. 693: 3366–3382. <https://doi.org/10.1002/qj.2619>.
- Cafaro, C., B. J. Woodhams, T. H. M. Stein, et al. 2021. "Do Convection-Permitting Ensembles Lead to More Skillful Short-Range Probabilistic Rainfall Forecasts Over Tropical East Africa?" *Weather and Forecasting* 36, no. 2: 697–716. <https://doi.org/10.1175/WAF-D-20-0172.1>.
- Carter, S., A. Steynor, K. Vincent, E. Visman, and K. Waagsaether. 2019. "Co-Production of African Weather and Climate Services." Manual, Cape Town: Future Climate for Africa and Weather and Climate Information Services for Africa. <https://futureclimateafrica.org/coproduction-manual>.
- Charney, J. G. 1971. "Geostrophic Turbulence." *Journal of the Atmospheric Sciences* 28, no. 6: 1087–1095.
- Charney, J. G., J. Shukla, J. Lighthill, and R. P. Pearce. 1981. *Predictability of Monsoons*, 99–110. Cambridge University Press.
- Chen, L., X. Zhong, F. Zhang, et al. 2023. "Fuxi: A Cascade Machine Learning Forecasting System for 15-Day Global Weather Forecast." *Npj Climate and Atmospheric Science* 6, no. 1: 190. <https://doi.org/10.1038/s41612-023-00512-1>.
- Cullmann, J., M. Dilley, P. Egerton, et al. 2020. "2020 State of Climate Services: Risk Information and Early Warning Systems." WMO Report No. 1252. <https://library.wmo.int/idurl/4/57191>.
- de Andra, F. M., M. P. Young, D. MacLeod, L. C. Hirons, S. J. Woolnough, and E. Black. 2021. "Subseasonal Precipitation Prediction for Africa: Forecast Evaluation and Sources of Predictability." *Weather and Forecasting* 36, no. 1: 265–284. <https://doi.org/10.1175/WAF-D-20-0054.1>.
- Egerton, P., J. Stander, L. Paterson, R. Devillier, and A. Champagne-Fall. 2022. "Early Warnings for All: The UN Global Early Warning Initiative for the Implementation of Climate Adaptation." WMO. <https://library.wmo.int/idurl/4/58209>.
- Gehne, M., B. Wolding, J. Dias, and G. N. Kiladis. 2022. "Diagnostics of Tropical Variability for Numerical Weather Forecasts." *Weather and Forecasting* 37, no. 9: 1661–1680. <https://doi.org/10.1175/WAF-D-21-0204.1>.
- Haiden, T., M. J. Rodwell, D. S. Richardson, A. Okagaki, T. Robinson, and T. Hewson. 2012. "Intercomparison of Global Model Precipitation Forecast Skill in 2010/11 Using the Seeps Score." *Monthly Weather Review* 140, no. 8: 2720–2733. <https://doi.org/10.1175/MWR-D-11-00301.1>.
- Hersbach, H., B. Bell, P. Berrisford, et al. 2023. "ERA5 Hourly Data on Pressure Levels From 1940 to Present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS)." <https://doi.org/10.24381/cds.bd0915c6>.
- Hoskins, B. 2013. "The Potential for Skill Across the Range of the Seamless Weather-Climate Prediction Problem: A Stimulus for Our Science." *Quarterly Journal of the Royal Meteorological Society* 139, no. 672: 573–584. <https://doi.org/10.1002/qj.1991>.
- Huffman, G., E. Stocker, D. Bolvin, E. Nelkin, and J. Tan. 2019. "GPM IMERG Final Precipitation L3 Half Hourly 0.1 Degree x 0.1 Degree V06." Greenbelt, MD, Goddard Earth Sciences Data and Information Services Center (GES DISC). <https://doi.org/10.5067/GPM/IMERG/3B-HH/06>.
- Judt, F. 2018. "Insights Into Atmospheric Predictability Through Global Convection-Permitting Model Simulations." *Journal of the Atmospheric Sciences* 75, no. 5: 1477–1497. <https://doi.org/10.1175/JAS-D-17-0343.1>.
- Judt, F. 2020. "Atmospheric Predictability of the Tropics, Middle Latitudes, and Polar Regions Explored Through Global Storm-Resolving Simulations." *Journal of the Atmospheric Sciences* 77, no. 1: 257–276. <https://doi.org/10.1175/JAS-D-19-0116.1>.

- Keane, R. J., R. S. Plant, and W. J. Tennant. 2016. "Evaluation of the Plant–Craig Stochastic Convection Scheme (v2.0) in the Ensemble Forecasting System Mogreps-r (24 Km) Based on the Unified Model (v7.3)." *Geoscientific Model Development* 9: 1921–1935. <https://doi.org/10.5194/gmd-9-1921-2016>.
- Knippertz, P. 2004. "A Simple Identification Scheme for Upper-Level Troughs and Its Application to Winter Precipitation Variability in Northwest Africa." *Journal of Climate* 17, no. 6: 1411–1418. [https://doi.org/10.1175/1520-0442\(2004\)017<1411:ASISFU>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<1411:ASISFU>2.0.CO;2).
- Kolstad, E. W., and D. MacLeod. 2022. "Lagged Oceanic Effects on the East African Short Rains." *Climate Dynamics* 59, no. 3: 1043–1056. <https://doi.org/10.1007/s00382-022-06176-6>.
- Kolstad, E. W., D. J. Parker, D. A. MacLeod, C. M. Wainwright, and L. C. Hirons. 2024. "Beyond the Regional Average: Drivers of Geographical Rainfall Variability During East Africa's Short Rains." *Quarterly Journal of the Royal Meteorological Society* 150, no. 764: 4550–4566. <https://doi.org/10.1002/qj.4829>.
- Lam, R., A. Sanchez-Gonzalez, M. Willson, et al. 2023. "Learning Skillful Medium-Range Global Weather Forecasting." *Science* 382, no. 6677: 1416–1421. <https://doi.org/10.1126/science.adi2336>.
- Lamprey, B., S. Sahabi Abed, and M. Gudoshava. 2024. "Challenges and Ways Forward for Sustainable Weather and Climate Services in Africa." *Nature Communications* 15: 2664. <https://doi.org/10.1038/s41467-024-46742-6>.
- Lazo, J. K., R. E. Morss, and J. L. Demuth. 2009. "300 Billion Served: Sources, Perceptions, Uses, and Values of Weather Forecasts." *Bulletin of the American Meteorological Society* 90, no. 6: 785–798. <https://doi.org/10.1175/2008BAMS2604.1>.
- Lebel, T., D. J. Parker, C. Flamant, et al. 2010. "The Amma Field Campaigns: Multiscale and Multidisciplinary Observations in the West African Region." *Quarterly Journal of the Royal Meteorological Society* 136, no. S1: 8–33. <https://doi.org/10.1002/qj.486>.
- Linsenmeier, M., and J. G. Shrader. 2023. "Global Inequalities in Weather Forecasts." 2023. Soc ArXiv 7e2jf, Center for Open Science. <https://ideas.repec.org/p/osf/socarx/7e2jf.html>.
- Lorenz, E. N. 1969. "The Predictability of a Flow Which Possesses Many Scales of Motion." *Tellus* 21, no. 3: 289–307. <https://doi.org/10.1111/j.2153-3490.1969.tb00444.x>.
- MacLachlan, C., A. Arribas, K. A. Peterson, et al. 2015. "Global Seasonal Forecast System Version 5 (glosea5): A High-Resolution Seasonal Forecast System." *Quarterly Journal of the Royal Meteorological Society* 141, no. 689: 1072–1084. <https://doi.org/10.1002/qj.2396>.
- Mittermaier, M. P. 2021. "A 'Meta' Analysis of the Fractions Skill Score: The Limiting Case and Implications for Aggregation." *Monthly Weather Review* 149, no. 10: 3491–3504. <https://doi.org/10.1175/MWR-D-18-0106.1>.
- Moron, V., and A. Robertson. 2020. "Tropical Rainfall Subseasonal-To-Seasonal Predictability Types." *Npj Climate and Atmospheric Science* 3, no. 1: 4. <https://doi.org/10.1038/s41612-020-0107-3>.
- Nastrom, G. D., and K. S. Gage. 1985. "A Climatology of Atmospheric Wavenumber Spectra of Wind and Temperature Observed by Commercial Aircraft." *Journal of Atmospheric Sciences* 42, no. 9: 950–960. [https://doi.org/10.1175/1520-0469\(1985\)042<0950:ACOAWS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1985)042<0950:ACOAWS>2.0.CO;2).
- Necker, T., L. Wolfgruber, L. Kugler, M. Weissmann, M. Dorninger, and S. Serafin. 2024. "The Fractions Skill Score for Ensemble Forecast Verification." *Quarterly Journal of the Royal Meteorological Society* 150, no. 764: 4457–4477. <https://doi.org/10.1002/qj.4824>.
- Norris, J., A. Hohner, and L. McGuire. 2021. "Precipitation as a Driver of Post-Wildfire Hydrologic Hazards." *WIREs Water* 8, no. 5: e1728. <https://doi.org/10.1002/wat2.1728>.
- Palmer, T. N., A. Döring, and G. Seregin. 2014. "The Real Butterfly Effect." *Nonlinearity* 27, no. 9: R123. <https://doi.org/10.1088/0951-7715/27/9/R123>.
- Parker, D. J., A. M. Blyth, S. J. Woolnough, et al. 2022. "The African Swift Project: Growing Science Capability to Bring About a Revolution in Weather Prediction." *Bulletin of the American Meteorological Society* 103, no. 2: E349–E369. <https://doi.org/10.1175/BAMS-D-20-0047.1>.
- Pulkkinen, S., D. Nerini, A. A. Pérez Hortal, et al. 2019. "Pysteps: An Open-Source Python Library for Probabilistic Precipitation Nowcasting (v1.0)." *Geoscientific Model Development* 12, no. 10: 4185–4219. <https://doi.org/10.5194/gmd-12-4185-2019>.
- Redelsperger, J.-L., C. D. Thorncroft, A. Diedhiou, T. Lebel, D. J. Parker, and J. Polcher. 2006. "African Monsoon Multidisciplinary Analysis: An International Research Project and Field Campaign." *Bulletin of the American Meteorological Society* 87, no. 12: 1739–1746. <https://doi.org/10.1175/BAMS-87-12-1739>.
- Riehl, H. 1954. *Tropical Meteorology*. McGraw-Hill.
- Roberts, A. J., J. K. Fletcher, J. Groves, et al. 2022. "Nowcasting for Africa: Advances, Potential and Value." *Weather* 77, no. 7: 250–256. <https://doi.org/10.1002/wea.3936>.
- Roberts, N. 2008. "Assessing the Spatial and Temporal Variation in the Skill of Precipitation Forecasts From an Nwp Model." *Meteorological Applications* 15, no. 1: 163–169. <https://doi.org/10.1002/met.57>.
- Roberts, N. M., and H. W. Lean. 2008. "Scale-Selective Verification of Rainfall Accumulations From High-Resolution Forecasts of Convective Events." *Monthly Weather Review* 136, no. 1: 78–97.
- Rockström, J., and M. Falkenmark. 2003. "Rainfall Variability and Its Impact on Crop Production in a Water Scarce Region: Upscaling From Field to Watershed Scale." *Agricultural Water Management* 61, no. 2: 109–128. [https://doi.org/10.1016/S0378-3774\(02\)00155-8](https://doi.org/10.1016/S0378-3774(02)00155-8).
- Schwartz, C. S. 2019. "Medium-Range Convection-Allowing Ensemble Forecasts With a Variable-Resolution Global Model." *Monthly Weather Review* 147, no. 8: 2997–3023. <https://doi.org/10.1175/MWR-D-18-0452.1>.
- Schwartz, C. S., J. S. Kain, S. J. Weiss, et al. 2010. "Toward Improved Convection-Allowing Ensembles: Model Physics Sensitivities and Optimizing Probabilistic Guidance With Small Ensemble Membership." *Weather and Forecasting* 25, no. 1: 263–280. <https://doi.org/10.1175/2009WAF2222267.1>.
- Selz, T. 2019. "Estimating the Intrinsic Limit of Predictability Using a Stochastic Convection Scheme." *Journal of the Atmospheric Sciences* 76, no. 3: 757–765. <https://doi.org/10.1175/JAS-D-17-0373.1>.
- Selz, T., and G. C. Craig. 2023. "Can Artificial Intelligence-Based Weather Prediction Models Simulate the Butterfly Effect?" *Geophysical Research Letters* 50, no. 20: e2023GL105747. <https://doi.org/10.1029/2023GL105747>.
- Selz, T., M. Riemer, and G. C. Craig. 2022. "The Transition From Practical to Intrinsic Predictability of Midlatitude Weather." *Journal of the Atmospheric Sciences* 79, no. 8: 2013–2030. <https://doi.org/10.1175/JAS-D-21-0271.1>.
- Shukla, J. 1998. "Predictability in the Midst of Chaos: A Scientific Basis for Climate Forecasting." *Science* 282, no. 5389: 728–731. <https://doi.org/10.1126/science.282.5389.728>.
- Skok, G., and N. Roberts. 2016. "Analysis of Fractions Skill Score Properties for Random Precipitation Fields and Ecmwf Forecasts." *Quarterly Journal of the Royal Meteorological Society* 142, no. 700: 2599–2610. <https://doi.org/10.1002/qj.2849>.
- Straus, D. M., and D. Paolino. 2009. "Intermediate Time Error Growth and Predictability: Tropics Versus Mid-Latitudes." *Tellus A: Dynamic Meteorology and Oceanography* 61, no. 5: 579–586. <https://doi.org/10.1111/j.1600-0870.2009.00411.x>.

- Tzachor, A., C. Richards, M. Gudoshava, et al. 2023. "How to Reduce Africa's Undue Exposure to Climate Risks." *Nature* 620, no. 7974: 488–491. <https://doi.org/10.1038/d41586-023-02557-x>.
- Vaughan, A., S. Markou, W. Tebbutt, et al. 2024. "Aardvark Weather: End-To-End Data-Driven Weather Forecasting." *arXiv:2404.00411 [physics.ao-ph]*. <https://doi.org/10.48550/arXiv.2404.00411>.
- Vizy, E., and K. Cook. 2014. "Impact of Cold Air Surges on Rainfall Variability in the Sahel and Wet African Tropics: A Multi-Scale Analysis." *Climate Dynamics* 43: 1057–1081. <https://doi.org/10.1007/s00382-013-1953-z>.
- Vogel, P., P. Knippertz, A. H. Fink, A. Schlueter, and T. Gneiting. 2020. "Skill of Global Raw and Postprocessed Ensemble Predictions of Rainfall in the Tropics." *Weather and Forecasting* 35, no. 6: 2367–2385. <https://doi.org/10.1175/WAF-D-20-0082.1>.
- Vonich, P. T., and G. J. Hakim. 2024. "Predictability Limit of the 2021 Pacific Northwest Heatwave From Deep-Learning Sensitivity Analysis." *Geophysical Research Letters* 51, no. 19: e2024GL110651. <https://doi.org/10.1029/2024GL110651>.
- Walters, D., A. J. Baran, I. Boutle, et al. 2019. "The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 Configurations." *Geoscientific Model Development* 12, no. 5: 1909–1963. <https://doi.org/10.5194/gmd-12-1909-2019>.
- Ward, N., A. H. Fink, R. J. Keane, and D. J. Parker. 2023. "Upper-Level Midlatitude Troughs in Boreal Winter Have an Amplified Low-Latitude Linkage Over Africa." *Atmospheric Science Letters* 24, no. 1: e1129. <https://doi.org/10.1002/asl.1129>.
- Williams, K. D., D. Copsey, E. W. Blockley, et al. 2018. "The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations." *Journal of Advances in Modeling Earth Systems* 10, no. 2: 357–380. <https://doi.org/10.1002/2017MS001115>.
- World Bank. 2021. *A Regional Analysis of Weather, Climate, Water and Early Warning Services in Southern Africa: Status Quo and Proposed Actions*. World Bank.
- Zhang, F., Y. Q. Sun, L. Magnusson, et al. 2019. "What Is the Predictability Limit of Midlatitude Weather?" *Journal of the Atmospheric Sciences* 76, no. 4: 1077–1091. <https://doi.org/10.1175/JAS-D-18-0269.1>.
- Zhao, B., and B. Zhang. 2018. "Assessing Hourly Precipitation Forecast Skill With the Fractions Skill Score." *Journal of Meteorological Research* 32: 135–145. <https://doi.org/10.1007/s13351-018-7058-1>.
- Zhu, H., M. C. Wheeler, A. H. Sobel, and D. Hudson. 2014. "Seamless Precipitation Prediction Skill in the Tropics and Extratropics From a Global Model." *Monthly Weather Review* 142, no. 4: 1556–1569. <https://doi.org/10.1175/MWR-D-13-00222.1>.