Received: 27 April 2023 | Revised: 17 August 2023 | Accepted: 15 September 2023

DOI: 10.1002/qj.4581

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

Sources of predictability of synoptic-scale rainfall during the West African summer monsoon

Athul Rasheeda Satheesh[1] | **Peter Knippertz**[1] | **Andreas H. Fink**[1] | **Eva-Maria Walz**[2,3] | **Tilmann Gneiting**[2,3]

[1]Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[2]Institute for Stochastics, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[3]Computational Statistics Group, Heidelberg Institute for Theoretical Studies, Heidelberg, Germany

Correspondence
Athul Rasheeda Satheesh, Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany.
Email: athul.satheesh@kit.edu

Funding information
Deutsche Forschungsgemeinschaft

Abstract

Numerical-model-based forecasts of precipitation exhibit poor skill over northern tropical Africa when compared with climatology-based forecasts and with other tropical regions. However, as recently demonstrated, purely data-driven forecasts based on spatio-temporal dependences inferred from gridded satellite rainfall estimates show promise for the prediction of the 24-hr precipitation occurrence rate in this region. The present work explores this potential further by advancing the statistical model and providing meteorological interpretations of the performance results. Advances include (a) the use of a recently developed correlation metric, the Coefficient of Predictive Ability (CPA), to identify predictors, (b) forecast evaluation with robust reliability diagrams and score decompositions, (c) a study domain over tropical Africa nested in a considerably enlarged spatio-temporal domain to identify coherent propagating features, and (d) the introduction of a novel coherent-linear-propagation factor to quantify the coherence of propagating signals. The statistical forecast is compared with a climatology-based benchmark, the European Centre for Medium-Range Weather Forecasts (ECMWF) operational ensemble forecast, and a statistically postprocessed ensemble forecast. All methods show poor skill within the main rainbelt over northern tropical Africa, where differences in Brier scores between the different approaches are hardly statistically significant. However, the data-driven forecast outperforms the other methods along the fringes of the rainbelt, where meridional rainfall gradients are large. The coherent-linear-propagation factor, in concert with metrics of convective available potential energy and convective instability, reveals that high stochasticity in the rainbelt limits predictability. At the fringes of the rainbelt, the data-driven approach leverages coherent precipitation features associated with propagating tropical weather systems such as African Easterly Waves.

KEYWORDS

African Easterly Waves, coefficient of predictive ability, coherence, data-driven forecast, West African rainfall

1 | INTRODUCTION

Rain-fed agriculture is a significant source of food and income for livelihoods in tropical Africa, a region that would benefit greatly from accurate forecasts of rainfall on timescales from hours to seasons. Early warning systems for floods and droughts and reservoir management are some other areas that could take advantage of this. However, while economies in the midlatitudes have greatly profited from the "quiet revolution" in numerical weather prediction (Bauer *et al.*, 2015), such a development has not taken place in tropical Africa. In addition to the poor availability of in situ observations (Parker *et al.*, 2008; Fink *et al.*, 2011), this is also due to the lower intrinsic predictability in the Tropics at the meso- and synoptic scales (Judt, 2020), that is, forecast errors at these scales grow faster than in the midlatitudes. This faster error growth relates to the convective nature of rainfall, in contrast to frontal-dominated precipitation in the midlatitudes. Haiden *et al.* (2012) show that, in terms of skill, a six-day lead-time precipitation deterministic forecast in the extratropics is comparable to a one-day lead-time in the Tropics. Specifically for tropical Africa, Vogel *et al.* (2018; 2020) and Walz *et al.* (2021) demonstrate that—particularly in regions where long-lived Mesoscale Convective Systems (MCSs) provide the bulk of the annual rainfall (Mathon *et al.*, 2002; Feng *et al.*, 2021)—an Extended Probabilistic Climatology (EPC) based on records of past rainfall can outperform or reach skill comparable to state-of-the-art global Numerical Weather Prediction (NWP) models for 24-hr and five-day ensemble rainfall forecasts. While statistical postprocessing of NWP forecasts can correct for biases and dispersion errors, it still does not enhance the skill beyond that of EPC-based forecasts (Vogel *et al.*, 2018). Compared with other parts of the Tropics, the vast lowlands of West and Central Equatorial Africa stand out as the continental region with the lowest predictive skill of rainfall.

In addition to the low intrinsic predictability, deficiencies in operational systems may contribute to the low skill of rainfall predictions over tropical Africa. Current global NWP models at spatial resolutions of a few tens of kilometres parameterize moist convection. This leads to fundamental misrepresentations of MCSs and thus rainfall intensity and frequency over West Africa (Marsham *et al.*, 2011; Pante and Knippertz, 2019), though recent improvements in parameterization may have alleviated this problem (Becker *et al.*, 2021). Moreover, it is challenging to forecast the initiation of convection, as triggers tend to stem from unresolved small-scale processes. Taylor *et al.* (2012) investigate the potential of soil moisture gradients as a storm-triggering mechanism, which gives rise to better short-term forecasts of convective initiation and tracks (Taylor *et al.*, 2022). Lafore *et al.* (2017) discuss the enhanced skill of precipitation forecasts related to orographic forcing over tropical Africa, as highlighted in Vogel *et al.* (2020) for the East African Highlands. However, northern tropical Africa and the Congo Basin have few orographic features. Thus, the overall inability of the forecast models to predict triggers and simulate the organization of convection are major reasons that NWP models struggle to produce skilful precipitation forecasts over West Africa (Fink *et al.*, 2011). The poor model representation of convective processes also contributes to the fact that additional upper-air observations and thus an improved definition of the initial state are basically lost after 24 hr of model integration, as data denial experiments for the African Monsoon Multidisciplinary Analysis (AMMA) and Dynamics–Aerosol–Chemistry–Cloud Interactions in West Africa (DACCIWA) campaigns show (Agustí-Panareda *et al.*, 2010; van der Linden *et al.*, 2020).

The new generation of convection-permitting (CP) models allows a better representation of MCSs, the diurnal cycle of precipitation, and monsoon circulation over West Africa (Marsham *et al.*, 2011; Pante and Knippertz, 2019). The latter authors note that the explicit simulation of Sahelian thunderstorms improves not only tropical but also midlatitude weather forecasts at lead times of 5–8 days. Operational short-range forecasts with pan-Africa CP models were recently started at the UK Met Office, showing improved skill but also revealing problems with spin-up in the first 24 hr of integration (Hanley *et al.*, 2021; Warner *et al.*, 2023). Cafaro *et al.* (2021) demonstrate that CP ensemble forecasts for tropical East Africa show improved skill over deterministic forecasts at the 24-hr lead time but are underdispersed in terms of the locations of heavy rainfall and domain-average rainfall. Thus, while CP models improve the climatological representation of rainfall systems in Africa, the gain in predictive skill appears limited and the computational costs are enormous (Senior *et al.*, 2021).

Given the general challenge of forecasting nontopographically triggered mesoscale convection, it is conceivable that windows of enhanced predictability may exist when large-scale, more predictable trigger and forcing mechanisms act. For boreal summer, Kniffka *et al.* (2020) note enhanced model skill in rainfall predictions over southern West Africa during the DACCIWA campaign when long-lived synoptic vortices were present. It is also well-known that tropical waves, in particular African Easterly Waves (AEWs), modulate rainfall in West Africa during the boreal summer wet season (Reed *et al.*, 1977; Fink and Reiner, 2003; Lavaysse *et al.*, 2006; Schlueter *et al.*, 2019a; Schlueter *et al.*, 2019b). These and other studies suggest that the waves do not trigger convection directly but change the larger scale

(thermo)dynamical environment and, thereby, enhance the odds of triggering and organization of MCSs.

The links between waves and rainfall in West Africa suggest two alternative avenues of forecasting precipitation. Firstly, one can leverage the more predictable large-scale waves in mixed statistical–dynamical forecast approaches, similar to methods developed for tropical cyclone occurrence in the North Atlantic Ocean by Maier-Gerber *et al.* (2021), or in wave-pattern-based calibration approaches (de Andrade *et al.*, 2021). A second approach is to use purely data-driven forecast models. Such models have a long tradition going back to the times prior to NWP, as they are computationally cheap and tend to be well calibrated by construction. Early approaches often relied on Markov processes, as in Gabriel and Neumann (1962) and Gates and Tong (1976). More recently, advanced statistical and machine-learning tools have been employed (Diez-Sierra and del Jesus, 2020; Ravuri *et al.*, 2021). While studies of this type may not seek physical explanation, their successes rely on the spatio-temporal coherence of rainfall and systematic relationships with other meteorological variables. Following this avenue, Vogel *et al.* (2021) show that a relatively simple, purely data-driven logistic regression model for 24-hr precipitation occurrence, which draws on spatio-temporal patterns of daily rainfall, outperforms forecasts based on climatology or NWP models during the summer monsoon season in West Africa. The model relies on the identification of grid points with the strongest correlation to a target grid point at lags of 1 and 2 days. The success of this method suggests that statistical models can reliably represent the systematic relationships between rainfall and circulation features that dynamical models appear to struggle with (Elless and Torn, 2018). Vogel *et al.* (2021) argue that their approach can, in principle, be employed whenever there is a dominant coherent forcing that triggers and modulates rainfall.

The study presented here builds on the foundations laid by Vogel *et al.* (2021) and develops them in several ways: We strengthen the purely data-driven statistical model for 24-hr precipitation occurrence by (a) using the recently developed Coefficient of Predictive Ability (CPA) for predictor selection and (b) expanding the logistic regression model to consider temporal lags of up to three days.[1] The CPA bridges Spearman's rank correlation coefficient, as used by Vogel *et al.* (2021), and the Area Under the Receiver Operating Characteristic (ROC) curve (AUC), in a way that makes CPA a non-negative asymmetric quantity bound between 0 and 1 (Gneiting and Walz, 2022). Moreover, (c) we introduce the coherent-linear-propagation factor as a metric that quantifies the coherence of a propagating rainfall signal identified over lags of one, two, and three days, (d) we provide—at least qualitatively—physical interpretations of the most significant patterns using auxiliary meteorological fields, and (e) we replace the standard analysis of Brier scores and reliability diagrams with more robust methods, namely, the recently developed Consistency Optimality Reproducibility PAV Algorithm-Based (CORP) approach (Dimitriadis *et al.*, 2021). Finally, as minor updates, (f) we now use the newer Global Precipitation Measurement Integrated Multi-satellitE Retrievals (hereafter, GPM IMERG) product instead of the discontinued *Tropical Rainfall Measuring Mission* (*TRMM*) data, and (g) the study domain has been extended significantly beyond West Africa, now including various environments (e.g., ocean, flatlands, highlands) in order to facilitate the identification of coherent propagating features. The ultimate goal of our work is to understand better the potential sources of predictability related to the time–space behaviour of rainfall in tropical Africa and to test the skill of the purely data-driven statistical model against the latest benchmarks with respect to predicting precipitation occurrence in the next 24 hr.

The article is structured as follows. Section 2 provides an overview of the data and methods used. In Section 3, spatio-temporal patterns, the direction and speed of propagating rainfall, and the degree of coherence are discussed for the summer monsoon season, and we compare the predictive performance of the purely data-driven logistic regression model with a climatological forecast (termed EPC15), the European Centre for Medium-Range Weather Forecasts (ECMWF) raw ensemble (ENS), and a statistically postprocessed ensemble forecast (ENS-IDR). The article closes with Section 4, where we summarize and discuss the major findings of this study and provide a brief outlook.

2 | DATA AND METHODS

2.1 | Data

2.1.1 | Observational data and ECMWF ensemble forecasts

To represent precipitation, we use the satellite-based, gridded GPM IMERG final version product (Hou *et al.*, 2014; Huffman *et al.*, 2020). GPM IMERG uses both radar-calibrated microwave radiance from polar-orbiting satellites and infrared radiance from geostationary satellites to create grids with half-hourly temporal and 0.1°

[1]Three days appears to be a good compromise between including all relevant information and keeping the domain size and computations at a reasonable level.

spatial resolution. We remap conservatively to $1° \times 1°$ spatial resolution in order to suppress small-scale noise in the precipitation field and ease the computational burden, and we consider daily accumulations (0600 UTC to 0600 UTC) that reflect the diurnal cycle of convection. To define precipitation occurrence, we use a binary cutoff at 0.2 mm of daily accumulation, as Vogel (2019) found only minimal changes with other thresholds between 0.0 and 1.0 mm.

The ERA5 dataset (Hersbach *et al.*, 2020) is used for horizontal wind at 600 and 925 hPa, surface-based Convective Available Potential Energy (CAPE), and Convective INhibition (CIN). To match the spatial resolution of GPM-IMERG in this study, the ERA5 data have been remapped using bilinear interpolation from the native $0.25° \times 0.25°$ resolution to the $1° \times 1°$ grid.

We compare forecasts from our data-driven model with the operational global 24-hr ensemble forecast from ECMWF, which consists of 50 perturbed members and a control run at a spatial resolution of $0.25° \times 0.25°$ (ECMWF Directorate, 2012). We remap conservatively to $1° \times 1°$ resolution and daily accumulations from 0600 UTC to 0600 UTC, and we refer to this forecast as ENS. As Wang *et al.* (2023) demonstrate, ECMWF forecasts outperform National Centers for Environmental Prediction–Global Forecast System (NCEP–GFS) forecasts of precipitation over Africa.

2.1.2 | Study domain and period

We concentrate on a study domain that covers most of tropical Africa and over which we issue forecasts (25°W–35°E, 0–18°N, red dashed box in Figure 1). The analyses over the study domain consider signals propagating from outside, comprising the adjacent oceans, as well as parts of South America and Southwestern Asia (12°S–30°N, 68°W–75°E, the whole area shown in Figure 1). The use of a significantly larger zonal extent compared with the meridional extent is motivated by our interest in identifying zonally propagating features. Temporally, the observations cover the summer monsoon season (July, August, and September) in the 19-year period from 2001–2019, and the forecasts cover the 13 years from 2007–2019, reflecting the ENS data availability in the THORPEX Interactive Grand Global Ensemble (TIGGE: Bougeault *et al.*, 2010; Swinbank *et al.*, 2016) archive.

2.2 | Methods

As noted, our study builds on the purely data-driven statistical approach to probability forecasts of 24-hr precipitation occurrence developed by Vogel *et al.* (2021), but extends it in several ways. We now describe the methodological innovations—some original to our article, others that have only very recently been developed—that underline these extensions.

2.2.1 | Selection of predictor variables via CPA

To identify predictor variables for our purely data-driven statistical model, we use 0600 UTC to 0600 UTC accumulated GPM IMERG precipitation within the nested core area (Figure 1) as target data, with the same field over the full area at temporal lags of 1, 2, and 3 days serving to provide candidate variables for the statistical model. For each target grid point, we compute the CPA measure between the local precipitation accumulation and all of the candidate variables.

The recently developed CPA measure is tailored to the selection of predictor variables for forecasting mixed discrete–continuous variables such as accumulated precipitation (Gneiting and Walz, 2022). CPA is asymmetric and thus takes the distinct roles of the target

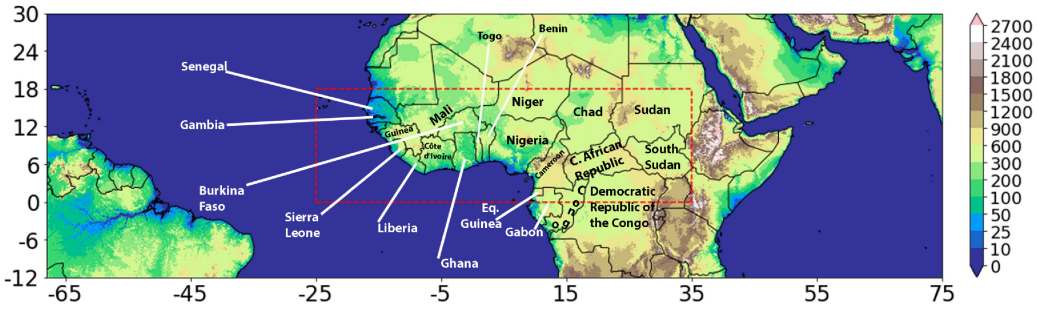

FIGURE 1 Overview of the study domain. We restrict the development of the purely data-driven statistical model and its evaluation to the nested core area that comprises 25°W–35°E in longitude and 0°–18°N in latitude. The shading indicates altitude in metres. [Colour figure can be viewed at wileyonlinelibrary.com]

variable—here, 24-hr precipitation accumulation at a target grid point (B)—and the potential predictor variables into account (A), that is, in general,

$$\mathrm{CPA}(A, B) \neq \mathrm{CPA}(B, A). \tag{1}$$

For data without ties, CPA becomes symmetric and can be expressed in terms of Spearman's rank correlation coefficient ρ_S, which was used by Vogel *et al.* (2021), namely, as

$$\mathrm{CPA} = \frac{\rho_S + 1}{2}. \tag{2}$$

While precipitation accumulations tend to show ties, this relationship only holds to a good degree of approximation over most of our analysis region. If the outcome is binary, CPA reduces to the classical AUC measure. Like the AUC measure, CPA is a non-negative quantity that is bound between 0 and 1, with a value of 0.50 indicating the absence of predictive skill and a value of 1 representing a deterministic, strictly increasing relationship. Values of CPA below 0.50 correspond to decreasing relationships and negative correlations.

As we operate in cross validation mode—with data from a single year left out and the rest of the data used to fit the statistical model—data from the season in consideration for the forecast are not included. For example, when developing the forecast model for the year 2016, data from 2016 are left out but data from all other years between 2001 and 2019 are used, and this is repeated for all years from 2001–2019. As an illustrative example for the CPA analysis, Figure 2 considers the target grid point closest to Niamey (Niger), with data from 2016 left out. At a lag of one day, the maximum arises 765 km to the east of Niamey, reaching a CPA value of 0.68. At a lag of two days, CPA generally decreases in magnitude and grid points with high CPA are shifted away from Niamey. At a lag of three days, the maximum CPA of 0.61 is reached at the border of Chad and Sudan nearly 2400 km east of Niamey in roughly the same direction as the other lags. The results are indicative of westward propagation with an approximate phase speed of 8.5 $\mathrm{m \cdot s^{-1}}$, which is typical for rainfall affected by AEWs. These patterns are exploitable for statistical forecasting and resemble those in fig. 1 of Vogel *et al.* (2021), suggesting good consistency between TRMM and IMERG data.

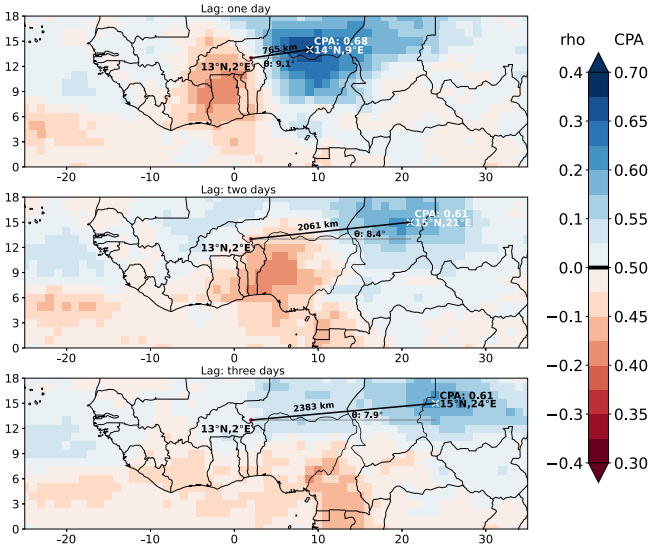

FIGURE 2 CPA with respect to the grid point (marked by a red dot) closest to Niamey (13°N, 2°E) at temporal lags of (a) one, (b) two, and (c) three days, based on data from July–September 2001–2019, except for 2016. The white cross (×) marks the grid point of maximum CPA at the given lag, with distance and direction (θ) from Niamey. The shading can also be interpreted in terms of the approximate value of Spearman's rank correlation ρ_S according to Equation 2. [Colour figure can be viewed at wileyonlinelibrary.com]

2.2.2 | Coherent-linear-propagation factor

To quantify the coherence of the patterns identified by the above analysis, we define a coherent-linear-propagation factor (coh) at any given target grid point. Figure 3 illustrates the concept on the same grid point as in Figure 2. Specifically, let X_0 be the target grid point under consideration, and let X_1, X_2, and X_3 (blue dots in Figure 3) be the grid points of maximum CPA at lags of one, two, and three days, respectively. We aim to quantify the extent to which the positions of these points agree with propagation at constant speed and in a constant direction. To this end, we introduce auxiliary, equispaced points P_1, P_2, and P_3 on a great-circle segment that starts from X_0 with distance D between successive points (Figure 3). This way, the great-circle distance (gcd) between X_0 and P_3 equals $3D$. The deviation from a linear propagation can now be quantified by the error term (orange lines in Figure 3),

$$E = \left(\mathrm{gcd}\,(X_1, P_1)^2 + \mathrm{gcd}\,(X_2, P_2)^2 + \mathrm{gcd}\,(X_3, P_3)^2 \right)^{1/2}. \tag{3}$$

Analogous to a regression problem, we can now find the optimal choice of the grid point P_3 at which E achieves a minimum. We denote this minimum by E_0, and we let D_0 be one third of the great-circle distance between X_0 and the optimal P_3.

The magnitude of E_0 will tend to be larger for grid points showing fast propagation, that is, for large D_0, such that these quantities should be considered in relation to

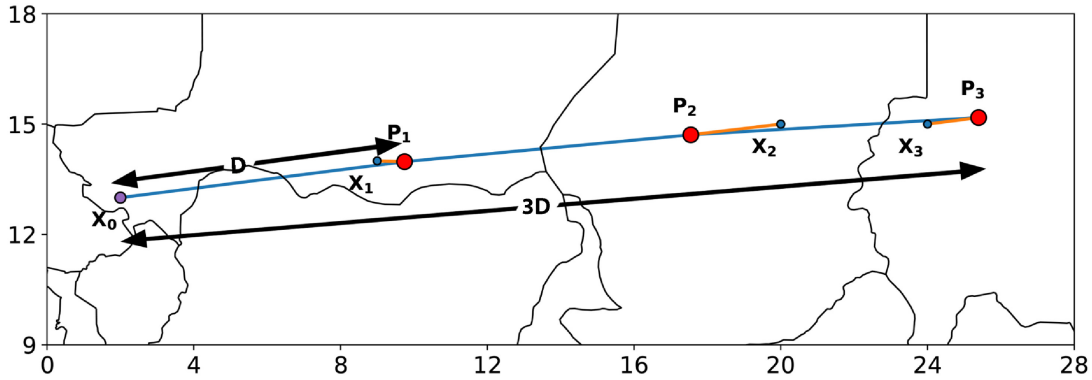

FIGURE 3 Computation of the coherent-linear-propagation factor (coh) at the grid point X_0 closest to Niamey (13°N, 2°E). The grid points X_1, X_2, and X_3 (blue dots) are the locations of maximum CPA with respect to X_0 at lags of one, two, and three days, respectively, based on data from July–September 2001–2019 except for 2016. The points P_1 and P_2 lie on the great-circle segment from X_0 to P_3 and mark the corresponding ideal locations of maximum CPA in the case of a perfectly linear propagation with constant phase speed. The grid point P_3 depicted here is chosen such that the error term in Equation 3 (orange lines) becomes minimal. [Colour figure can be viewed at wileyonlinelibrary.com]

each other. We therefore define[2] the coh at the target grid point X_0 as

$$\text{coh} = \frac{3D_0}{3D_0 + E_0}. \tag{4}$$

In Figure 3, the grid point at P_3 is such that the error term in Equation 3 is minimized. The coh measure then equals the ratio of the length of the blue great-circle segment to the sum of the length of the blue segment and the error term; specifically, $D_0 = 844.4$ km, $E_0 = 315.8$ km, and coh = 0.889. Generally, under near-perfect linear propagation, E_0 will be small and thus coh will be close to its ideal value of 1. If the points of maximum CPA scatter at random, E_0 will be large and coh will be small (but remain non-negative). In a nutshell, high values of the coh measure characterize regions dominated by stable propagation of long-lived atmospheric features that modulate conditions for rainfall, implying comparably high predictability that could potentially be exploited with statistical models. Conversely, low coh characterizes regions that do not experience stable, coherent linear propagation. For these regions, rainfall might be largely stochastic or might be subject to distinct forcings. It is important to note that these characterizations are in terms of linear propagation; more complex methods might be able to detect acceleration and deceleration as well as nonlinear tracks. Also, the definition is only sensitive to geographic distances and not the values of CPA itself, which of course, do matter for the skill of statistical models built on these relationships.

[2]If $X_0 = X_1 = X_2 = X_3$, we note a lack of propagating features at the target grid point at X_0, and thus we set coh = 0. In all other cases we proceed as stated.

2.2.3 | Purely data-driven logistic model forecast

Following Vogel *et al.* (2021), we now construct a purely data-driven, statistical logistic regression (hereafter Logistic) model. To this end, let o_1, o_2, and o_3 denote the observed precipitation accumulations at the grid points of highest CPA at lags of 1, 2, and 3 days, respectively. Then the logistic model for the probability p of 24-hr precipitation occurrence at the target grid point depends on o_1, o_2, o_3 and the day of the year d, positing that

$$\log \frac{p}{1-p} = a_0 + a_1 f(o_1) + a_2 f(o_2) + a_3 f(o_3) + b_1 \sin \frac{2\pi d}{365} + b_2 \cos \frac{2\pi d}{365}, \tag{5}$$

where the final two terms depend periodically on the day of the year. The function $f(x) = \log(x + 0.001)$ is used to transform a non-negative rainfall amount to a real value. We estimate the statistical parameters a_0, a_1, a_2, a_3, b_1, and b_2 in Equation 5 using the LogisticRegression subroutine from the scikit-learn package (Pedregosa *et al.*, 2011). The parameters are estimated for each target grid point individually, and so they vary spatially. Following Vogel *et al.* (2021), we operate in cross-validation mode, that is, the logistic model is trained on data from all years available except the year for which the prediction is issued. This is repeated until all the years from 2007–2019 are used. Cross-validation makes the best use of the data available and limits artefacts due to a low number of samples, thereby producing more robust results. As a consequence of this approach we developed a separate model for every year from 2007–2019. It is to be noted that we only show

the results from 2016 as an example in Section 3.1, but the results for other years are similar.

2.2.4 | Benchmarking: EPC15, ENS, and ENS-IDR forecasts

In order to benchmark the performance of the purely data-driven Logistic probability of precipitation forecasts, we compare it with an Extended Probabilistic Climatology with a 15-day window (EPC15; Walz *et al.*, 2021), the raw ECMWF ensemble (ENS), and a statistically post-processed ensemble (ENS-IDR). For a fair comparison in cross-validation mode, we treat the EPC15 forecast in the same way as the Logistic forecast, that is, the probability of precipitation is determined from a ±15-day window around the target date, based on observations from 2001–2019 except for the year for which the prediction is issued. This is repeated for each of the years from 2007–2019, for which we assess and compare the four types of forecast.

The operational ECMWF ensemble consists of 50 perturbed members and a control run, so the ENS probability of precipitation forecast equals the number of ensemble members that predict a precipitation accumulation of at least 0.2 mm divided by 51. Statistical postprocessing can correct systematic deficiencies in NWP ensemble output (Vannitsem *et al.*, 2018) and various methods, ranging from simple linear regression to sophisticated neural-network-based techniques, have been proposed for this purpose. In this study, we use Isotonic Distributional Regression (IDR; Henzi *et al.*, 2021) since, unlike every single competitor, IDR does not depend on user decisions and has been shown by Henzi *et al.* (2021) to be competitive with Bayesian Model Averaging (BMA; Sloughter *et al.*, 2007) and Ensemble Model Output Statistics (EMOS; Scheuerer, 2014) for precipitation (Henzi *et al.*, 2021). Furthermore, in the case of probability forecasts for a binary outcome, such as the occurrence of precipitation, IDR coincides with isotonic calibration (Zadrozny and Elkan, 2002), which is a widely used, classical method in machine learning. We retain the cross-validation setting used to generate the Logistic and EPC15 forecasts, that is, we train IDR[3] on observations from the monsoon seasons (July–September, hereafter JAS) in 2007–2019 except the year for which the prediction is issued, and repeat for each of the evaluation seasons from 2007–2019.

To assess whether or not an observed difference in the mean Brier score of the Logistic forecast and a benchmark forecast could be explained by chance alone, even though the theoretical skill is equal, we employ the Diebold–Mariano test (Diebold and Mariano, 2002) in concert with the Benjamini–Hochberg correction (Benjamini and Hochberg, 1995), as recommended by Wilks (2016) for the spatio-temporal setting at hand. For details on EPC15, IDR, and the test procedure we refer to the aforementioned references.

[3]To implement IDR we use the `isodisreg` package for Python; see https://github.com/evwalz/isodisreg.

3 | RESULTS

We first discuss spatio-temporal patterns of rainfall in the summer monsoon season by analysing the characteristics of maximum CPA relative to a target grid point and by studying the associated coherent-linear-propagation factor along with auxiliary meteorological fields and results from previous studies. We assess the predictive performance of our purely data-driven Logistic probability of precipitation forecast (hereafter, Logistic forecast) in comparison with the three benchmark forecasts, both for the example grid point near Niamey and over the entire evaluation domain (Figure 1).

3.1 | Maximum CPA and coherent-linear-propagation factor

Using the methodology described in Section 2.2.1, Figure 4 provides a detailed analysis of the points of maximum CPA relative to the target grid points in the evaluation domain.

The maximum CPA (Figure 4a–c) peaks at a temporal lag of one day and for target grid points over the northern tropical Atlantic Ocean, where it reaches values of about 0.75. The African rainbelt (Nicholson, 2008; Nicholson, 2009) is broadly co-located with the lowest values of maximum CPA, albeit with some east–west gradient (see dark red isohyets in Figure 4a–c). The latitudinal position of its axis varies from 6°N in central Africa to 10°N over West Africa, and monthly rainfall amounts range from well above 400 mm per month over the warm waters of the Atlantic to below 200 mm per month over South Sudan. Also note the slight orographic enhancement on the western sides of the Guinea Highlands and the Cameroon Line Mountains. High values of maximum CPA are generally found in areas of large climatological rainfall gradients straddling the rainbelt. On the southern side, the zone of high CPA at a lag of one day stretches from the eastern Atlantic at about 5°N past the Guinea Coast towards southern Nigeria, Cameroon, and Gabon. Interestingly, the distortion of the rainfall gradient over the relatively drier Ivory Coast and Ghana is also reflected in a northward excursion of higher CPA values. To the north of the rainbelt, over the Sahel, CPA also is enhanced.

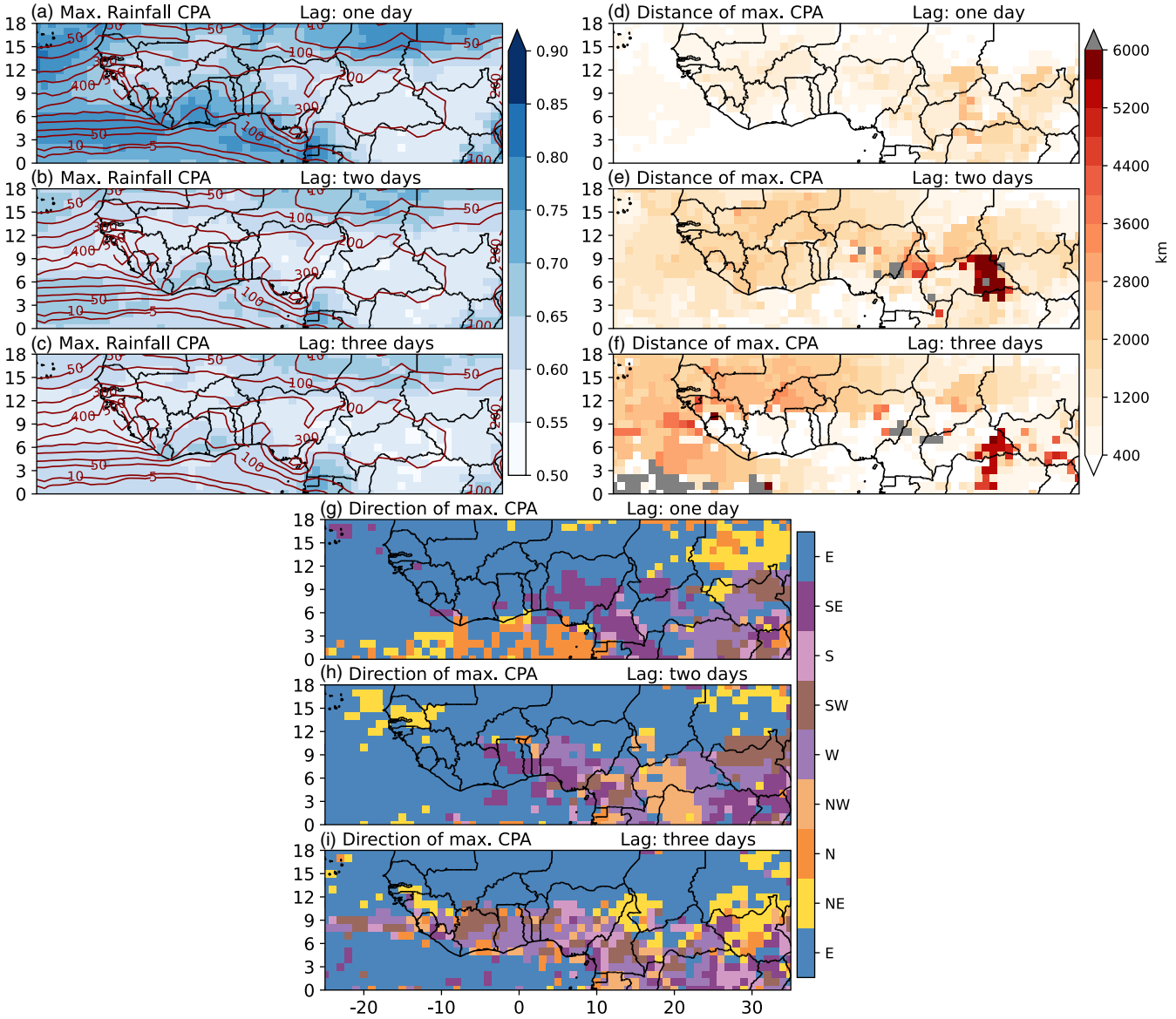

FIGURE 4 (a–c) Maximum CPA along with (d–f) distance and (g–i) direction between the target grid point and the grid point with maximum CPA, at temporal lags of (a,d,g) one, (b,e,h) two, and (c,f,i) three days, respectively, for JAS 2001–2019 except 2016. The dark red contours in (a–c) show monthly average rainfall in the period considered in mm. [Colour figure can be viewed at wileyonlinelibrary.com]

A possible interpretation of these patterns is that the rainbelt is characterized by favourable conditions for convection, which creates a high degree of stochasticity, whereas the fringes are more sensitive to coherent synoptic-scale features that trigger or organize convection, such as slow, westward-propagating cyclonic vortices that persist for more than a day, or nighttime systems that last across the 0600 UTC dividing time (Knippertz *et al.*, 2017; Kniffka *et al.*, 2020). Comparing results for a lag of one day with those for lags of two and three days, we find good agreement in the geographic patterns of maximum CPA, despite a considerable reduction in the values (as low as 0.50).

Considering the distance (Figure 4d–f) and direction (Figure 4g–i) of the point with maximum CPA relative to the target grid point, we leave grid points with distances below 400 km blank to signal very slow propagation. The largest distances are generally found for points in the continent's interior along the African rainbelt. Two types of behaviour can be distinguished. In a broad region stretching from the West Coast to 20°E, propagation is mostly westward (occasionally northwestward) with distances at a lag of one day typically at 600–1000 km, which corresponds to about 6–10° longitude per day or a phase speed of about 6.9–11.5 $m \cdot s^{-1}$. This type of behaviour is also evident in the example from Figure 2. It extends out to the downstream Atlantic, with propagation distances in the lower end of the range. The second type occurs over Central Africa and is characterized by mostly eastward propagation over larger distances.

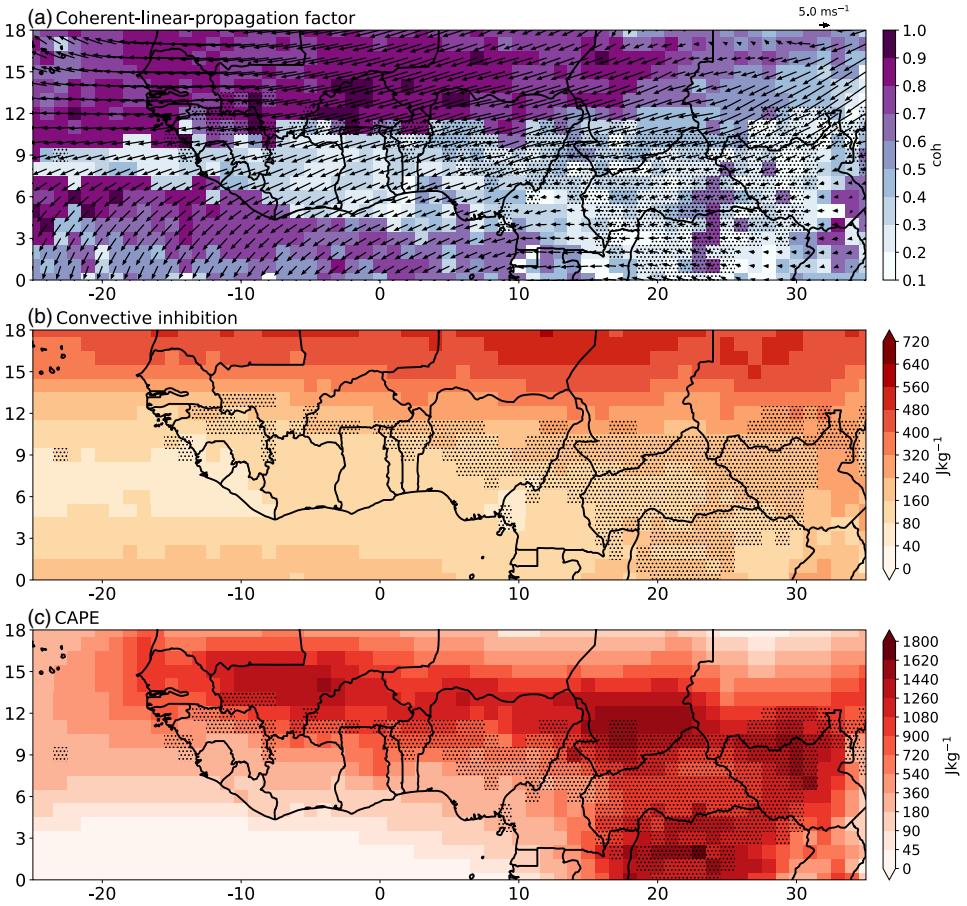

FIGURE 5 Diagnostic features: (a) coherent-linear-propagation factor (coh) and vertical shear between 600 and 925 hPa (arrows), (b) Convective INhibition (CIN), and (c) Convective Available Potential Energy (CAPE) for JAS 2001–2019 except 2016. Stippled grid points have maximum CPA below 0.575 at all three lags. [Colour figure can be viewed at wileyonlinelibrary.com]

We proceed to an interpretation of these results via the coh and auxiliary meteorological fields, such as vertical shear, CIN, and CAPE, in Figure 5. The highest values of coh are found in the western Sahel north of 10°N, where coh reaches an impressive 0.90 over Burkina Faso, western Niger, and northwestern Nigeria (Figure 5a). In view of the relatively low maximum CPA of around 0.65 at a lag of one day, this insinuates relatively fast, westward-propagating rainfall features with phase speeds in the range 7.5–12 $m \cdot s^{-1}$, in concert with smaller-scale variation, possibly weakening the spatio-temporal relationships. This aspect is also evident in Figure 2, where high values of CPA are spread out over a relatively large area. These results are consistent with large easterly vertical wind shear between 600 and 925 hPa (arrows), which is key to the formation of MCSs (Maranan *et al.*, 2018). Given mean propagation speeds of 9 $m \cdot s^{-1}$ for AEWs and 15 $m \cdot s^{-1}$ for MCSs (Fink and Reiner, 2003), the high coherence might stem from a dominating influence of AEW propagation, which is known to modulate rainfall strongly in this area (Schlueter *et al.*, 2019b). At the same time, shorter-lived MCSs form within the AEW envelope but propagate much faster due to the high shear, leading to relatively low CPA, as noted.

Another relevant factor in explaining the high coherence over the western Sahel is CIN (Figure 5b), which shows high values throughout, with a marked south-to-north increase, with values ranging from 80 $J \cdot kg^{-1}$ over southern Mali to over 400 $J \cdot kg^{-1}$ over northern Mali and Niger. High CIN implies that relatively strong triggers are needed to generate rainfall: for example, associated with AEWs and MCSs. Combined with high CAPE (Figure 5c), long-lived rainfall events tend to be relatively infrequent but, if they occur, rather intense, resulting in relatively low CPA and moderate overall rainfall in the western Sahel.

In the eastern Sahel, values of coh drop below 0.50, despite CPA values being similar to the west Sahel region and even higher in northern parts (Figure 4a–c). This is likely due to weaker AEWs, which grow in amplitude while propagating westward across the Sahel along the African Easterly Jet (AEJ), which is climatologically weaker in the east (Fink and Reiner, 2003; Lafore *et al.*, 2017). Shear also is weaker than in the western Sahel and less homogeneous in direction and magnitude. Further reasons for lower coh may lie in higher CIN and smaller CAPE, which make it harder to trigger convection, as also reflected in lower rainfall (dark red isohyets in Figure 4a–c).

A sharp drop in coh is observed as we move southward from the Sahel towards the centre of the rainbelt, where values of coh are generally very low. An exception is Central Africa (eastern Central African Republic and northern Democratic Republic of the Congo), where coh is locally much higher, though in combination with CPA below 0.575, as indicated by stippling in Figure 5a–c, which suggests possible artefacts. Indeed, there is no dominant propagation direction in this area (Figure 4g–i). Generally, westward propagation dominates in the rainbelt (Figure 4g–i) combined with low CIN and high CAPE (Figure 5b–c) such that convection can be triggered with ease almost anywhere, inducing high area-averaged rainfall and a high degree of stochasticity.

To the southwest of the rainbelt there is a mostly oceanic region (tropical eastern Atlantic) with moderate to large values of coh, where values of CPA at one day lag peak around 0.75 near the coast of Liberia and Sierra Leone (Figure 4a). Since convective maxima over the ocean tend to occur in the early morning (Albright *et al.*, 1985; Hendon and Woodberry, 1993), it is conceivable that in some cases the spatio-temporal relationship is caused by a convective storm surviving until the next morning and therefore affecting consecutive 24-hr periods, as corroborated by distances mostly below 400 km (Figure 4d) and variable directions from northerly to easterly. Interestingly, for lags of 2 and 3 days, larger distances and more consistent westward propagation are found (Figure 4e–f,h–i), suggesting that at these timescales propagating disturbances dominate over local effects. Vertical shear (Figure 5a) in this region is mostly northeasterly and thus does not follow the largely easterly direction to the north of the rainbelt. Despite this, propagation is predominantly westward, indicating that levels other than 600 and 925 hPa may play a role. Another characteristic of this region is low CIN (Figure 5b), which slightly increases southward towards the equatorial cold tongue, and low CAPE (Figure 5c). Given the absence of triggers common over land (diurnal heating, orography, coastal effects, etc.), this suggests that triggers from meso- or synoptic-scale circulation features may be necessary to release the weak convective instability or even to cause stratiform rain over the oceanic region. While Knippertz *et al.* (2017) describe slowly westward-moving, oceanic, synoptic-scale cyclonic systems during the DACCIWA campaign in 2016 and their effects on rainfall, there are no extant in-depth studies of this phenomenon.

3.2 | Forecast performance

As noted, our purely data-driven Logistic probability of precipitation forecast uses the CPA analysis to identify predictor variables from rainfall records. We now assess its predictive performance relative to the EPC15, ENS, and ENS-IDR forecasts.

For an initial illustration at the previously considered grid point closest to Niamey, Figure 6 shows reliability diagrams in the recently proposed CORP form (Dimitriadis *et al.*, 2021) for the Logistic forecast and the three competitors. The CORP approach assumes a monotonic, nondecreasing shape of the conditional event probability (CEP) or reliability curve and yields an associated decomposition of the mean Brier score (BS) into miscalibration (MCB), discrimination (DSC), and uncertainty (UNC) components, namely,

$$\mathrm{BS} = \mathrm{MCB} - \mathrm{DSC} + \mathrm{UNC}, \tag{6}$$

while avoiding the shortcomings of the traditional binning and counting approach to reliability diagrams and score decomposition. For a comparison with the classical approach, see Figure S1 and the detailed discussion in Dimitriadis *et al.* (2021). Supplementing the CORP reliability curves and confidence bands, which reflect the natural variability of the reliability curves under the assumption of perfect calibration, the inset histograms illustrate the marginal distribution of the forecast values. In the case of discrete probabilities (ENS, ENS-IDR), bar plots are used, and in the case of (nearly) continuously distributed forecast probabilities (Logistic, EPC15) density estimates are shown.

For the Niamey grid point, the Logistic forecast shows a mean Brier score of 0.209.[4] It is well calibrated (MCB = 0.005), showing moderate discrimination ability (DSC = 0.024) and forecast probability range $\sim 0.30 - 0.87$. In terms of the Brier score, it outperforms both the EPC15 and the ENS forecast but not the ENS-IDR forecast. The EPC15 forecast is even better calibrated (MCB = 0.003) but has poor discrimination ability (DSC = 0.012), with forecast probabilities ranging between 0.40 and 0.75 only. The low values in the marginal distribution are likely indicative of the monsoon onset, which is highly variable from year to year, while the high values indicate the peak of the monsoon season. In contrast, the ENS forecast is poorly calibrated (MCB = 0.027), but shows superior discrimination ability (DSC = 0.040), with forecast probabilities attaining the full range from 0 to 1. However, when the rainfall probability of the ENS forecast is 1.00 (i.e., all 51 members indicating rain), which is surprisingly

[4]We repeated the analysis considering only lags up to two days, that is, by setting $a_3 = 0$ in Equation 5, which resulted in a slightly worse Brier Score (BS = 0.211), mainly due to a decrease in discrimination (DSC = 0.020) despite a slight improvement in calibration (MSC = 0.004).

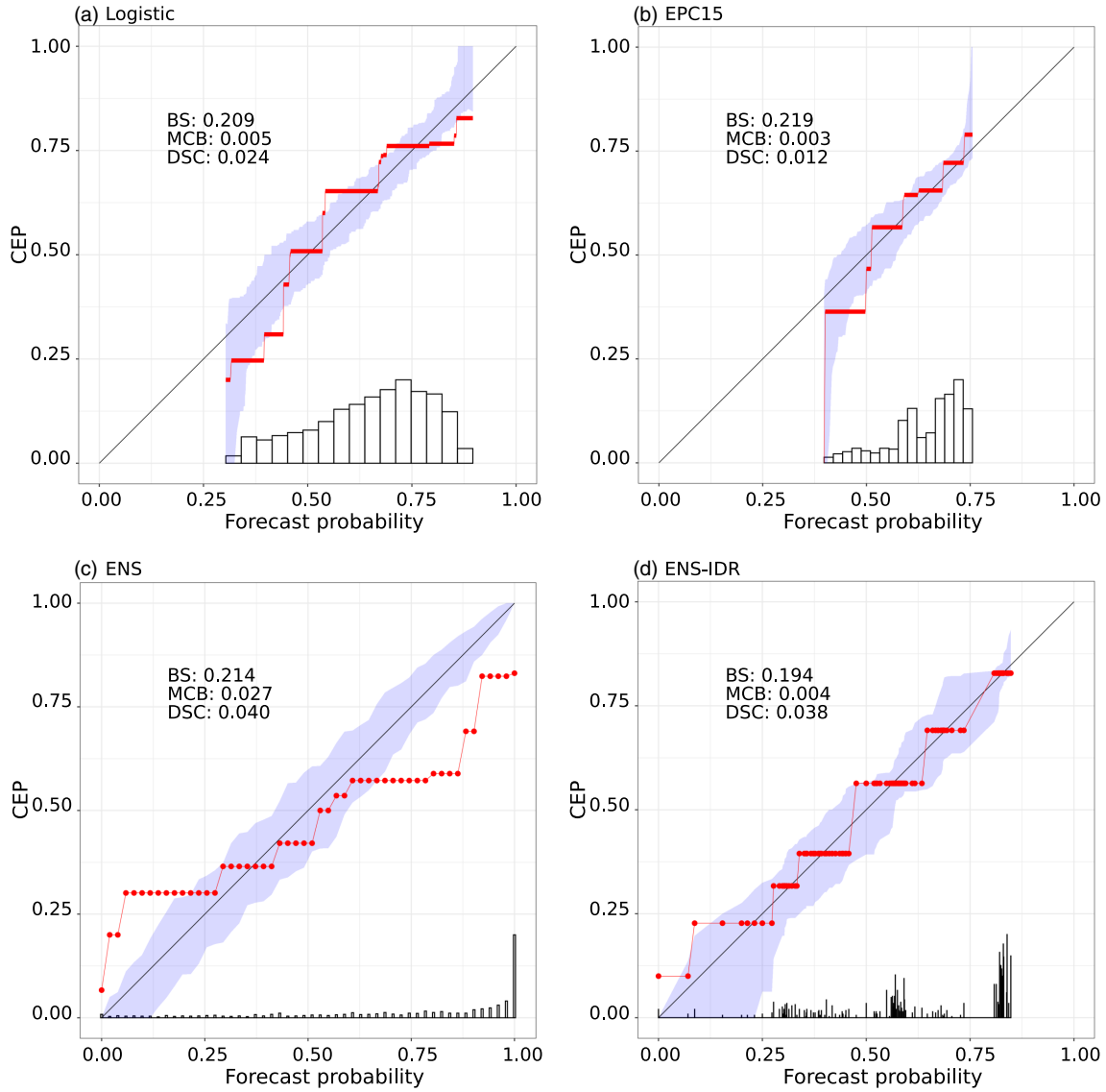

FIGURE 6 CORP reliability diagrams for (a) Logistic, (b) EPC15, (c) ENS, and (d) ENS-IDR probability of precipitation forecasts at the grid point closest to Niamey (13°N, 2°E) for JAS 2007–2019. The CORP reliability curves plot the conditional event probability (CEP) against the forecast value, along with 90% consistency bands under the assumption of perfect reliability. We also show the mean Brier score (BS) and its CORP miscalibration (MCB) and discrimination (DSC) components. The UNC component is 0.228 and only depends on climatology and not the forecast. [Colour figure can be viewed at wileyonlinelibrary.com]

often (50.9%) the case, rainfall occurs with a conditional event probability (CEP) of 0.85 only. Finally, the postprocessed ENS-IDR forecast is well calibrated (MCB = 0.004), while retaining most of the discrimination ability from the ensemble (DSC = 0.038), leading to the best mean Brier score of 0.194, 0.015 better than Logistic. However it is to be noted that the postprocessing reduces the range of forecast probabilities to ∼ 0.00–0.85.

We repeat this comparison at every target grid point within the nested core area, to obtain the mean Brier score (BS, Figure S2) and the associated MCB and DSC components (Figure 7) from the CORP decomposition. The most striking result is that the ENS forecast suffers from poor calibration over much of the analysis region, and very poor calibration over the Gulf of Guinea and the equatorial Atlantic Ocean in particular (Figure 7e). However, it needs to be mentioned that, similarly to Vogel *et al.* (2020) who pointed out that the uncertainties associated with *TRMM* estimates over the oceanic *deserts* may be responsible for the high miscalibration in ENS, uncertainties associated with the IMERG precipitation estimates (e.g., Bolvin *et al.*, 2021) over the oceans may be partly responsible for the high MCB values of ENS in this study. The MCB values of the postprocessed ENS-IDR forecasts resemble those of the Logistic and EPC15 forecasts, with all three being well calibrated (Figure 7a,c,g). Not surprisingly, the

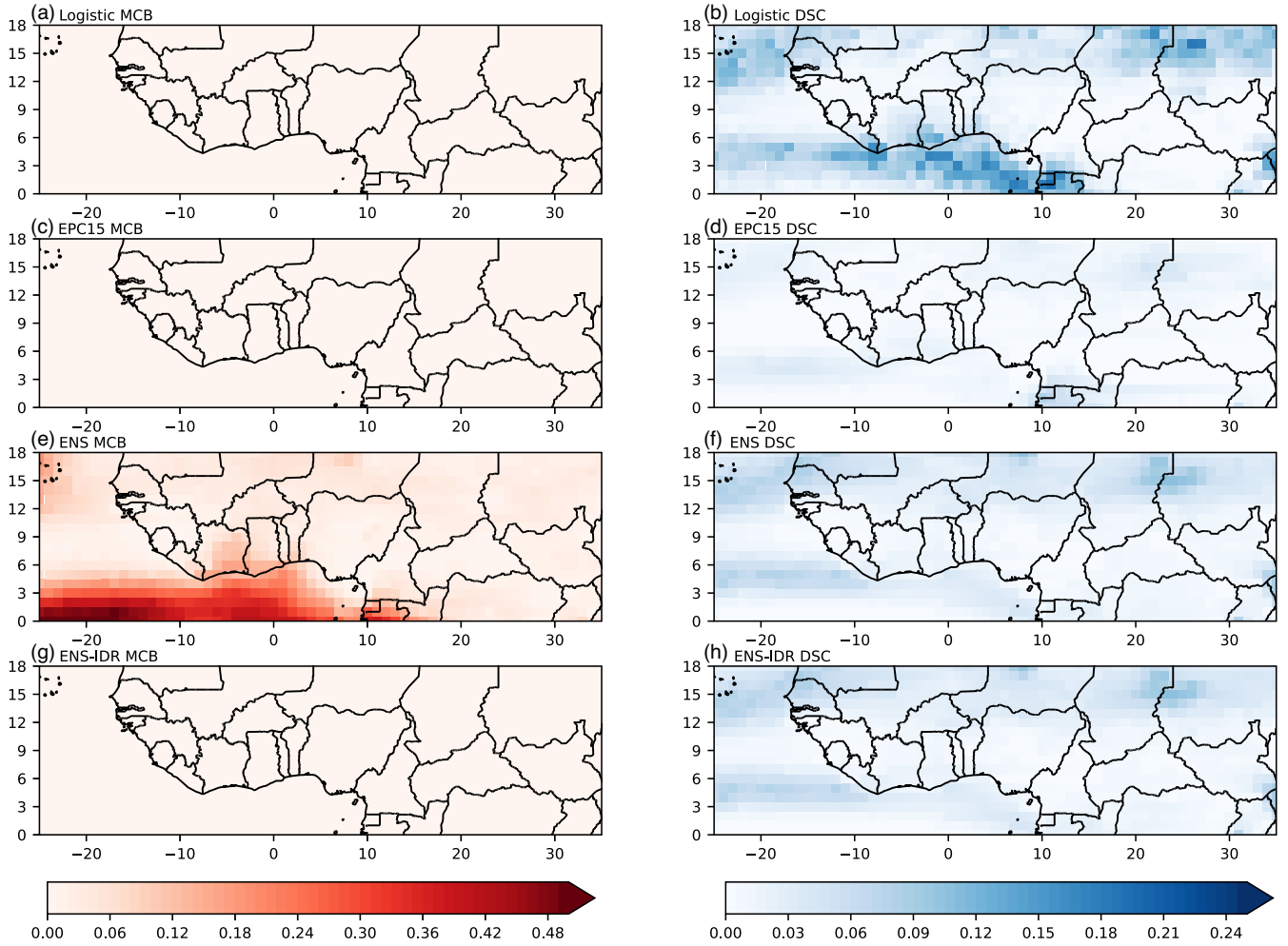

FIGURE 7 (a,c,e,g) CORP miscalibration (MCB) and (b,d,f,h) discrimination (DSC) components of the mean Brier score (BS) for (a–b) logistic, (c–d) EPC15, (e–f) ENS, and (g–h) ENS-IDR probability of precipitation forecasts for JAS 2007–2019. [Colour figure can be viewed at wileyonlinelibrary.com]

climatological EPC15 forecast shows the least discrimination ability, as reflected by low DSC values (Figure 7d). The spatial patterns of the DSC values for the Logistic forecast (Figure 7b) resemble those for CPA at a lag of one day (Figure 4a) and peak in the off-equatorial Atlantic Ocean near the Cape Verde Islands, the eastern Sahel, the Gulf of Guinea, and the coasts of southern West Africa. DSC values for ENS and ENS-IDR (Figure 7f,h) show some spatial agreement with Logistic but are generally much smoother. The good agreement between the two indicates that the postprocessing retains the discrimination ability of the dynamical model. Grid points within the African rainbelt generally show very low DSC values for all four forecasts considered, which is likely related to the high level of stochasticity discussed above.

To facilitate direct comparison between the Logistic forecast and the benchmark techniques, Figure 8 shows the Brier skill score (BSS) of the Logistic forecast relative to the EPC15, ENS, and ENS-IDR forecasts. Relative to EPC15 (Figure 8a), the Logistic forecast shows better skill over most of our evaluation domain, and particularly in regions of high rainfall gradients, where coherent (Figure 5a) propagating features create predictability through convective triggers or ambient conditions. Improvements over EPC15 in the rainbelt, however, are mostly small and statistically insignificant (hatched areas). In comparison with the ENS forecast (Figure 8b), the Logistic forecast is superior practically everywhere, with some very high values over the ocean. Nonetheless, according to the Diebold–Mariano test for equal predictive ability, the differences between the Logistic and EPC15 and ENS forecasts fail to be statistically significant over some parts of the climatological rainbelt.

The BSS of the Logistic forecast relative to the ENS-IDR forecast (Figure 8c) shows an overall mixed result. The Logistic forecast performs significantly better over regions with high CPA, namely, the northeastern Sahel, the Gulf of Guinea, and the majority of the highly populated coastal

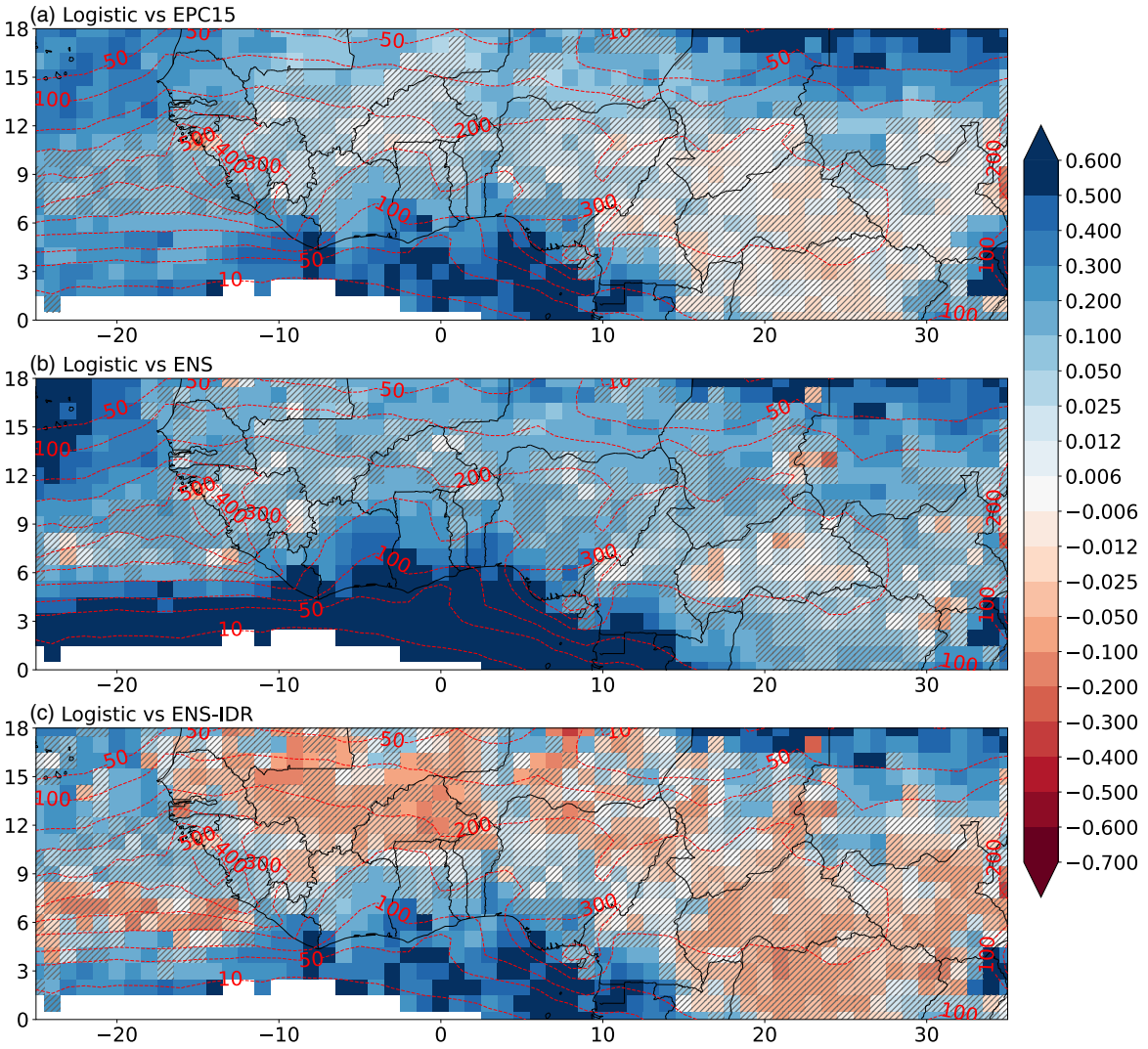

FIGURE 8 Brier skill score (BSS) for the Logistic forecast relative to the (a) EPC15, (b) ENS, and (c) ENS-IDR forecasts for JAS 2007–2019. Hatching indicates grid points where the difference in the mean Brier score (BS) fails to be statistically significant at a level of 0.05, according to the Benjamini–Hochberg corrected Diebold–Mariano test for equal predictive ability. The contours show average monthly rainfall accumulation in mm. Grid points with less than than 5 mm of average monthly rainfall are left blank. [Colour figure can be viewed at wileyonlinelibrary.com]

regions along it (comprising Côte d'Ivoire, Ghana, Togo, Benin, Nigeria, and Cameroon), the oceanic regions surrounding the Cape Verde islands near the coast of Senegal, and nearby coastal regions, due to superior discrimination ability (Figure 7b). The ENS-IDR forecast performs better over a narrow strip of grid points over the Atlantic ocean off the coast of Sierra Leone along the southern boundary of the rainbelt, in the western Sahel, in the Central African Republic, and in the Congo basin. However, the associated differences in the mean Brier score fail to be statistically significant. This confirms that, even with sophisticated postprocessing, a forecast with a dynamical model can hardly outperform a relatively simple data-driven one.

Interestingly, the findings in this section remain largely unchanged if the Logistic forecast is based on predictor variables at a temporal lag of one day only, barring some deterioration of the predictive performance over the western Sahel (Figure S3). This somewhat surprising result indicates that the bulk of the useful information is contained in yesterday's rainfall, be it a local storm that persists from one day to the next (as is often the case over the ocean) or an MCS or AEW that propagates towards the point of interest (as is often the case over the Sahel).

4 | CONCLUSIONS

Operational forecasts of rainfall in tropical Africa and the adjacent Atlantic Ocean have low skill (Vogel *et al.*, 2018), but purely data-driven statistical models that

exploit spatio-temporal correlations show promise (Vogel *et al.*, 2021). To explore the potential of the data-driven approach more systematically, we have used the recently developed CPA in a spatio-temporal analysis of 24-hr accumulated GPM-IMERG rainfall from 2001–2019 for the West African summer monsoon season, spanning up to lags of three days temporally and covering up to several thousand km spatially. The CPA patterns found can be attributed predominantly to synoptic-scale drivers causing characteristic time–space behaviour of precipitation anomalies. To substantiate this, we have introduced a novel coherent-linear-propagation factor (coh) that quantifies the extent to which the locations of lagged maximum CPA reflect a propagation with constant phase speed and direction. Large values of coh combined with substantial CPA over the three days indicate physically interpretable and statistically stable relationships and, thus, the potential for high statistical predictability. Such patterns are usually associated with high values of CAPE, CIN, and shear, which necessitates the presence of relatively strong triggers like AEWs and MCSs to generate rainfall. Land regions located within the climatological rainbelt show very low values of CPA and coh, indicating high stochasticity, while oceanic regions located within the rainbelt show sizeable CPA and moderate coh. High values of coh along the northern and southern fringes of the rainbelt over land reflect the presence of a coherent synoptic forcing that triggers and sustains precipitation over the three days.

To leverage these findings, we have carried out a systematic comparison of the predictive performance of a purely data-driven Logistic probability of precipitation forecast that employs predictors from the identified locations of maximum CPAs with three benchmark forecasts, namely, the climatological EPC15 forecast, the ENS forecast, which is heavily miscalibrated, and ENS-IDR. All forecasts struggle over regions where the precipitation is dominated by highly stochastic processes—in particular, in the African rainbelt—due to the lack of coherent synoptic-scale signals. In these regions, the differences between the mean Brier scores of the Logistic, EPC15, and ENS-IDR forecasts are small and fail to be statistically significant. However, along the fringes of the rainbelt—in particular, in the northeastern Sahel, the Gulf of Guinea, and nearby coastal regions—where stochasticity is much less pronounced, statistical approaches benefit from dominant forcing that triggers and sustains precipitation events such as AEWs, well in line with earlier results by Vogel *et al.* (2021).

With a view toward operational forecasting, we summarize that raw ensemble forecasts issued by operational services tend to be poorly calibrated in our study region and may offer little to no skill over northern tropical Africa. However, it needs to be mentioned that the operational forecasts from ECMWF are based on convection-parameterized simulations. While CP models have been shown to perform better in this regard over some regions (e.g., Woodhams *et al.*, 2018), the computational costs of a large-domain CP ensemble forecast will remain too high for years to come. Therefore, we recommend that operational services consider the use of statistical post-processing techniques that improve the skill of the raw forecasts by proper calibration (Henzi *et al.*, 2021; Hewson and Pillosu, 2021). Climatological forecasts such as the EPC forecast offer very cheap alternatives if postprocessing is not available. However, our key recommendation is that purely data-driven techniques, such as the Logistic forecast studied here, be used to obtain the probability of precipitation forecasts over northern tropical Africa, with the added benefit of being computationally much cheaper than NWP-based forecasts. We anticipate and encourage extensions of the data-driven, statistical approach, which currently is restricted to probability forecasts for the occurrence of precipitation, to full probabilistic forecasts of precipitation accumulations.

Finally, one might employ advanced techniques of machine learning, such as convolutional neural networks or graph neural networks, to develop more sophisticated data-driven approaches, as demonstrated impressively by Lam *et al.* (2022), Bi *et al.* (2023), Nguyen *et al.* (2023), and Chen *et al.* (2023). Whilst artificial intelligence (AI) based data-driven methods such as these show immense potential for many aspects of forecasting, they still face hurdles in forecasting precipitation. The approach demonstrated in this article may help advance AI-based methods in this direction, as our rather simple data-driven Logistic forecast combines expertise from data science and atmospheric science to leverage insights about spatio-temporal patterns of precipitation over northern tropical Africa and adjacent regions. While modern neural networks already have the ability to learn such patterns from data alone, without the need for human expertise for variable selection, it remains to be seen whether they can outperform much simpler data-driven or hybrid (statistical-dynamical) methods in regions such as northern tropical Africa.

AUTHOR CONTRIBUTIONS

Athul Rasheeda Satheesh: conceptualization; data curation; formal analysis; investigation; methodology; visualization; writing – original draft; writing – review and editing. **Peter Knippertz:** conceptualization; formal analysis; investigation; methodology; supervision; writing – review and editing. **Andreas H. Fink:** conceptualization; formal analysis; investigation; methodology; supervision;

writing – review and editing. **Eva-Maria Walz:** data curation; resources; software. **Tilmann Gneiting:** formal analysis; investigation; methodology; supervision; writing – review and editing.

ACKNOWLEDGEMENTS

The research leading to these results has been accomplished within phase 2 of project C2, "Statistical-dynamical forecasts of tropical rainfall" of the Transregional Collaborative Research Center SFB/TRR 165 "Waves to Weather" funded by the German Science Foundation (DFG). This work is based on TIGGE data. TIGGE (The Interactive Grand Global Ensemble) is an initiative of the World Weather Research Programme (WWRP). The authors thank their colleagues Alexander Jordan and Benedikt Schulz for code review. Tilmann Gneiting and Eva-Maria Walz are grateful for support by the Klaus Tschira Foundation. Open Access funding enabled and organized by Projekt DEAL.

ORCID

Athul Rasheeda Satheesh https://orcid.org/0000-0002-1915-3539
Peter Knippertz https://orcid.org/0000-0001-9856-619X
Andreas H. Fink https://orcid.org/0000-0002-5840-2120

REFERENCES

Agustí-Panareda, A., Beljaars, A., Cardinali, C., Genkova, I. and Thorncroft, C. (2010) Impacts of assimilating AMMA soundings on ECMWF analyses and forecasts. *Weather and Forecasting*, 25, 1142–1160.

Albright, M.D., Recker, E.E., Reed, R.J. and Dang, R. (1985) The diurnal variation of deep convection and inferred precipitation in the central tropical Pacific during January–February 1979. *Monthly Weather Review*, 113, 1663–1680.

Bauer, P., Thorpe, A. and Brunet, G. (2015) The quiet revolution of numerical weather prediction. *Nature*, 525, 47–55.

Becker, T., Bechtold, P. and Sandu, I. (2021) Characteristics of convective precipitation over tropical Africa in storm-resolving global simulations. *Quarterly Journal of the Royal Meteorological Society*, 147, 4388–4407.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, 289–300.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X. and Tian, Q. (2023) Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619, 533–538.

Bolvin, D.T., Huffman, G.J., Nelkin, E.J. and Tan, J. (2021) Comparison of monthly IMERG precipitation estimates with PACRAIN atoll observations. *Journal of Hydrometeorology*, 22, 1745–1753.

Bougeault, P., Toth, Z., Bishop, C., Brown, B., Burridge, D., Chen, D.H., Ebert, B., Fuentes, M., Hamill, T.M. and Mylne, K. (2010) The THORPEX interactive grand global ensemble. *Bulletin of the American Meteorological Society*, 91, 1059–1072.

Cafaro, C., Woodhams, B.J., Stein, T.H., Birch, C.E., Webster, S., Bain, C.L., Hartley, A., Clarke, S., Ferrett, S. and Hill, P. (2021) Do convection-permitting ensembles lead to more skillful short-range probabilistic rainfall forecasts over tropical East Africa? *Weather and Forecasting*, 36, 697–716.

Chen, K., Han, T., Gong, J., Bai, L., Ling, F., Luo, J.-J., Chen, X., Ma, L., Zhang, T. and Su, R. (2023) FengWu: Pushing the skillful global medium-range weather forecast beyond 10 days lead. Preprint, https://arxiv.org/abs/2304.02948.

de Andrade, F.M., Young, M.P., MacLeod, D., Hirons, L.C., Woolnough, S.J. and Black, E. (2021) Subseasonal precipitation prediction for Africa: forecast evaluation and sources of predictability. *Weather and Forecasting*, 36, 265–284.

Diebold, F.X. and Mariano, R.S. (2002) Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 20, 134–144.

Diez-Sierra, J. and del Jesus, M. (2020) Long-term rainfall prediction using atmospheric synoptic patterns in semi-arid climates with statistical and machine learning methods. *Journal of Hydrology*, 586, 124789.

Dimitriadis, T., Gneiting, T. and Jordan, A.I. (2021) Stable reliability diagrams for probabilistic classifiers. *Proceedings of the National Academy of Sciences*, 118, e2016191118.

ECMWF Directorate. (2012) Describing ECMWF's forecasts and forecasting system. ECMWF Newsletter No. 133 – Autumn 2012. http://www.ecmwf.int/sites/default/files/elibrary/2012/14576-newsletter-no133-autumn-2012.pdf.

Elless, T.J. and Torn, R.D. (2018) African easterly wave forecast verification and its relation to convective errors within the ECMWF ensemble prediction system. *Weather and Forecasting*, 33, 461–477.

Feng, Z., Leung, L.R., Liu, N., Wang, J., Houze, R.A., Jr., Li, J., Hardin, J.C., Chen, D. and Guo, J. (2021) A global high-resolution mesoscale convective system database using satellite-derived cloud tops, surface precipitation, and tracking. *Journal of Geophysical Research: Atmospheres*, 126, e2020JD034202.

Fink, A.H., Agustí-Panareda, A., Parker, D.J., Lafore, J.-P., Ngamini, J.-B., Afiesimama, E., Beljaars, A., Bock, O., Christoph, M. and Didé, F. (2011) Operational meteorology in West Africa: observational networks, weather analysis and forecasting. *Atmospheric Science Letters*, 12, 135–141.

Fink, A.H. and Reiner, A. (2003) Spatiotemporal variability of the relation between African easterly waves and west African squall lines in 1998 and 1999. *Journal of Geophysical Research: Atmospheres*, 108, 1–17.

Gabriel, K. and Neumann, J. (1962) A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quarterly Journal of the Royal Meteorological Society*, 88, 90–95.

Gates, P. and Tong, H. (1976) On Markov chain modeling to some weather data. *Journal of Applied Meteorology and Climatology*, 15, 1145–1151.

Gneiting, T. and Walz, E.-M. (2022) Receiver operating characteristic (ROC) movies, universal ROC (UROC) curves, and coefficient of predictive ability (CPA). *Machine Learning*, 111, 2769–2797.

Haiden, T., Rodwell, M.J., Richardson, D.S., Okagaki, A., Robinson, T. and Hewson, T. (2012) Intercomparison of global model precipitation forecast skill in 2010/11 using the SEEPS score. *Monthly Weather Review*, 140, 2720–2733.

Hanley, K.E., Pirret, J.S., Bain, C.L., Hartley, A.J., Lean, H.W., Webster, S. and Woodhams, B.J. (2021) Assessment of convection-permitting versions of the unified model over the

Lake Victoria basin region. *Quarterly Journal of the Royal Meteorological Society*, 147, 1642–1660.

Hendon, H.H. and Woodberry, K. (1993) The diurnal cycle of tropical convection. *Journal of Geophysical Research: Atmospheres*, 98, 16623–16637.

Henzi, A., Ziegel, J.F. and Gneiting, T. (2021) Isotonic distributional regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 83, 963–993.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R. and Schepers, D. (2020) The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049.

Hewson, T.D. and Pillosu, F.M. (2021) A low-cost post-processing technique improves weather forecasts around the world. *Communications Earth & Environment*, 2, 132.

Hou, A.Y., Kakar, R.K., Neeck, S., Azarbarzin, A.A., Kummerow, C.D., Kojima, M., Oki, R., Nakamura, K. and Iguchi, T. (2014) The global precipitation measurement mission. *Bulletin of the American Meteorological Society*, 97, 701–722.

Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E.J., Sorooshian, S., Tan, J. and Xie, P. (2020) NASA global precipitation measurement (GPM) integrated Mult-satellitE retrievals for GPM (IMERG), Algorithm Theoretical Basis Document (ATBD) Version 06. https://gpm.nasa.gov/sites/default/files/2020-05/IMERG_ATBD_V06.3.pdf.

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, 257–276.

Kniffka, A., Knippertz, P., Fink, A.H., Benedetti, A., Brooks, M.E., Hill, P.G., Maranan, M., Pante, G. and Vogel, B. (2020) An evaluation of operational and research weather forecasts for southern West Africa using observations from the DACCIWA field campaign in June–July 2016. *Quarterly Journal of the Royal Meteorological Society*, 146, 1121–1148.

Knippertz, P., Fink, A.H., Deroubaix, A., Morris, E., Tocquer, F., Evans, M.J., Flamant, C., Gaetani, M., Lavaysse, C. and Mari, C. (2017) A meteorological and chemical overview of the DACCIWA field campaign in West Africa in June–July 2016. *Atmospheric Chemistry and Physics*, 17, 10893–10918.

Lafore, J.P., Chapelon, N., Diop, M., Gueye, B., Largeron, Y., Lepape, S., Ndiaye, O., Parker, D.J., Poan, E. and Roca, R. (2017) Deep convection. In: Parker, D.J. and Diop-Kane, M. (Eds.) *Meteorology of Tropical West Africa: The Forecasters' Handbook*. Chichester: Wiley, pp. 90–129.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Pritzel, A., Ravuri, S., Ewalds, T., Alet, F., Eaton-Rosen, Z., Hu, W., Merose, A., Hoyer, S., Holland, G., Stott, J., Vinyals, O., Mohamed, S. and Battaglia, P. (2022) GraphCast: Learning skillful medium-range global weather forecasting. Preprint, https://arxiv.org/abs/2212.12794.

Lavaysse, C., Diedhiou, A., Laurent, H. and Lebel, T. (2006) African easterly waves and convective activity in wet and dry sequences of the west African monsoon. *Climate Dynamics*, 27, 319–332.

Maier-Gerber, M., Fink, A.H., Riemer, M., Schoemer, E., Fischer, C. and Schulz, B. (2021) Statistical–dynamical forecasting of subseasonal North Atlantic tropical cyclone occurrence. *Weather and Forecasting*, 36, 2127–2142.

Maranan, M., Fink, A.H. and Knippertz, P. (2018) Rainfall types over southern West Africa: objective identification, climatology and synoptic environment. *Quarterly Journal of the Royal Meteorological Society*, 144, 1628–1648.

Marsham, J.H., Knippertz, P., Dixon, N.S., Parker, D.J. and Lister, G.M. (2011) The importance of the representation of deep convection for modeled dust-generating winds over West Africa during summer. *Geophysical Research Letters*, 38, L16803.

Mathon, V., Laurent, H. and Lebel, T. (2002) Mesoscale convective system rainfall in the Sahel. *Journal of Applied Meteorology and Climatology*, 41, 1081–1092.

Nguyen, T., Brandstetter, J., Kapoor, A., Gupta, J.K. and Grover, A. (2023) ClimaX: A foundation model for weather and climate. Preprint, https://arxiv.org/abs/2301.10343.

Nicholson, S.E. (2008) The intensity, location and structure of the tropical rainbelt over West Africa as factors in interannual variability. *International Journal of Climatology*, 28, 1775–1785.

Nicholson, S.E. (2009) On the factors modulating the intensity of the tropical rainbelt over West Africa. *International Journal of Climatology*, 29, 673–689.

Pante, G. and Knippertz, P. (2019) Resolving Sahelian thunderstorms improves mid-latitude weather forecasts. *Nature Communications*, 10, 3487.

Parker, D.J., Fink, A., Janicot, S., Ngamini, J.-B., Douglas, M., Afiesimama, E., Agusti-Panareda, A., Beljaars, A., Didé, F. and Diedhiou, A. (2008) The amma radiosonde program and its implications for the future of atmospheric monitoring over Africa. *Bulletin of the American Meteorological Society*, 89, 1015–1028.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R. and Dubourg, V. (2011) Scikit-learn: machine learning in python. *Journal of Machine Learning Research*, 12, 2825–2830.

Ravuri, S., Lenc, K., Willson, M., Kangin, D., Lam, R., Mirowski, P., Fitzsimons, M., Athanassiadou, M., Kashem, S. and Madge, S. (2021) Skilful precipitation nowcasting using deep generative models of radar. *Nature*, 597, 672–677.

Reed, R.J., Norquist, D.C. and Recker, E.E. (1977) The structure and properties of African wave disturbances as observed during phase III of GATE. *Monthly Weather Review*, 105, 317–333.

Scheuerer, M. (2014) Probabilistic quantitative precipitation forecasting using ensemble model output statistics. *Quarterly Journal of the Royal Meteorological Society*, 140, 1086–1096.

Schlueter, A., Fink, A.H. and Knippertz, P. (2019a) A systematic comparison of tropical waves over northern Africa. Part II: dynamics and thermodynamics. *Journal of Climate*, 32, 2605–2625.

Schlueter, A., Fink, A.H., Knippertz, P. and Vogel, P. (2019b) A systematic comparison of tropical waves over northern Africa. Part I: influence on rainfall. *Journal of Climate*, 32, 1501–1523.

Senior, C.A., Marsham, J.H., Berthou, S., Burgin, L.E., Folwell, S.S., Kendon, E.J., Klein, C.M., Jones, R.G., Mittal, N. and Rowell, D.P. (2021) Convection-permitting regional climate change simulations for understanding future climate and informing decision-making in Africa. *Bulletin of the American Meteorological Society*, 102, E1206–E1223.

Sloughter, J.M.L., Raftery, A.E., Gneiting, T. and Fraley, C. (2007) Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Monthly Weather Review*, 135, 3209–3220.

Swinbank, R., Kyouda, M., Buchanan, P., Froude, L., Hamill, T.M., Hewson, T.D., Keller, J.H., Matsueda, M., Methven, J. and Pappenberger, F. (2016) The TIGGE project and its achievements. *Bulletin of the American Meteorological Society*, 97, 49–67.

Taylor, C.M., de Jeu, R.A., Guichard, F., Harris, P.P. and Dorigo, W.A. (2012) Afternoon rain more likely over drier soils. *Nature*, 489, 423–426.

Taylor, C.M., Klein, C., Dione, C., Parker, D.J., Marsham, J.H., Diop, C.A., Fletcher, J., Chaibou, A.A.S., Nafissa, D.B. and Semeena, V.S. (2022) Nowcasting tracks of severe convective storms in West Africa from observations of land surface state. *Environmental Research Letters*, 17(3), 034016.

van der Linden, R., Knippertz, P., Fink, A.H., Ingleby, B., Maranan, M. and Benedetti, A. (2020) The influence of DACCIWA radiosonde data on the quality of ECMWF analyses and forecasts over southern West Africa. *Quarterly Journal of the Royal Meteorological Society*, 146, 1719–1739.

Vannitsem, S., Wilks, D.S. and Messner, J. (Eds.). (2018) *Statistical Postprocessing of Ensemble Forecasts*. Amsterdam: Elsevier.

Vogel, P. (2019) *Assessing Predictive Performance: From Precipitation Forecasts Over the Tropics to Receiver Operating Characteristic Curves and Back*. Ph.D. thesis. Karlsruhe: Karlsruher Institut für Technologie (KIT). https://doi.org/10.5445/IR/1000091649

Vogel, P., Knippertz, P., Fink, A.H., Schlueter, A. and Gneiting, T. (2018) Skill of global raw and postprocessed ensemble predictions of rainfall over northern tropical Africa. *Weather and Forecasting*, 33, 369–388.

Vogel, P., Knippertz, P., Fink, A.H., Schlueter, A. and Gneiting, T. (2020) Skill of global raw and postprocessed ensemble predictions of rainfall in the tropics. *Weather and Forecasting*, 35, 2367–2385.

Vogel, P., Knippertz, P., Gneiting, T., Fink, A.H., Klar, M. and Schlueter, A. (2021) Statistical forecasts for the occurrence of precipitation outperform global models over northern tropical Africa. *Geophysical Research Letters*, 48, e2020GL091022.

Walz, E.-M., Maranan, M., van der Linden, R., Fink, A.H. and Knippertz, P. (2021) An IMERG-based optimal extended probabilistic climatology (EPC) as a benchmark ensemble forecast for precipitation in the tropics and subtropics. *Weather and Forecasting*, 36, 1561–1573.

Wang, Y., Gueye, M., Greybush, S.J., Greatrex, H., Whalen, A.J., Ssentongo, P., Zhang, F., Jenkins, G.S. and Schiff, S.J. (2023) Verification of operational numerical weather prediction model forecasts of precipitation using satellite rainfall estimates over Africa. *Meteorological Applications*, 30, e2112.

Warner, J.L., Petch, J., Short, C.J. and Bain, C. (2023) Assessing the impact of a NWP warm-start system on model spin-up over tropical Africa. *Quarterly Journal of the Royal Meteorological Society*, 149, 621–636.

Wilks, D.S. (2016) "The stippling shows statistically significant grid points": how research results are routinely overstated and overinterpreted, and what to do about it. *Bulletin of the American Meteorological Society*, 97, 2263–2273.

Woodhams, B.J., Birch, C.E., Marsham, J.H., Bain, C.L., Roberts, N.M. and Boyd, D.F. (2018) What is the added value of a convection-permitting model for forecasting extreme rainfall over tropical East Africa? *Monthly Weather Review*, 146, 2757–2780.

Zadrozny, B. and Elkan, C. (2002) Transforming classifier scores into accurate multiclass probability estimates. KDD'02: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 694–699.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Rasheeda Satheesh, A., Knippertz, P., Fink, A.H., Walz, E.-M. & Gneiting, T. (2023) Sources of predictability of synoptic-scale rainfall during the West African summer monsoon. *Quarterly Journal of the Royal Meteorological Society*, 1–17. Available from: https://doi.org/10.1002/qj.4581