

Performance evaluation of real-time sub-to-seasonal (S2S) rainfall forecasts over West Africa of 2020 and 2021 monsoon seasons for operational use

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Olaniyan, E. A., Woolnough, S. J. ORCID: <https://orcid.org/0000-0003-0500-8514>, Andrade, F. M. D. ORCID: <https://orcid.org/0000-0001-6653-3916>, Hirons, L. C. ORCID: <https://orcid.org/0000-0002-1189-7576>, Thompson, E. ORCID: <https://orcid.org/0000-0003-4250-8075> and Lawal, K. A. ORCID: <https://orcid.org/0000-0002-8198-8844> (2025) Performance evaluation of real-time sub-to-seasonal (S2S) rainfall forecasts over West Africa of 2020 and 2021 monsoon seasons for operational use. *Atmosphere*, 16 (9). 1072. ISSN 2073-4433 doi: 10.3390/atmos16091072 Available at <https://centaur.reading.ac.uk/124480/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.3390/atmos16091072>

Publisher: MDPI

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur





CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Article

Performance Evaluation of Real-Time Sub-to-Seasonal (S2S) Rainfall Forecasts over West Africa of 2020 and 2021 Monsoon Seasons for Operational Use

Eniola A. Olaniyan ^{1,*}, Steven J. Woolnough ² , Felipe M. De Andrade ³ , Linda C. Hirons ², Elisabeth Thompson ⁴  and Kamoru A. Lawal ^{5,6} 

¹ Department of Meteorology, African Aviation and Aerospace University, Abuja 900102, Nigeria

² National Centre for Atmospheric Science, University of Reading, Reading RG6 6BB, UK; s.j.woolnough@reading.ac.uk (S.J.W.); l.c.hirons@reading.ac.uk (L.C.H.)

³ Empresa de Pesquisa Agropecuária e Extensão Rural de Santa Catarina (Epagri), Florianópolis 88034-901, SC, Brazil; felipestratus@gmail.com

⁴ Met Office, Exeter EX1 3PB, UK; elisabeth.thompson@metoffice.gov.uk

⁵ African Centre of Meteorological Applications for Development (ACMAD), Niamey PL6 BP 13184, Plateau, Niger; lawal@acmad.org

⁶ Climate and Development Initiative (ACDI), Department of Environmental and Geographical Science, University of Cape Town, Cape Town 7701, South Africa

* Correspondence: eniola.ajani@aaaau.edu.ng

Abstract

Accurate sub-seasonal-to-seasonal (S2S) forecasts are critical for mitigating extreme weather impacts and supporting development in West Africa. This study evaluates real-time ECMWF S2S rainfall forecasts during the 2020–2021 West African monsoon (March–October) and uses corresponding hindcasts for comparison. We verify forecasts at 1–4 dekads lead against two satellite-based rainfall datasets (TAMSAT and GPM-IMERG) to cover observational uncertainty. The analysis focuses on spatio-temporal monsoon patterns over the Gulf of Guinea (GoG) and Sahel (SAH). The results show that ECMWF-S2S captures key monsoon features. The forecast skill is generally higher over the Sahel than the GoG, and peaks during the main monsoon period (July–August). Notably, forecasts achieve approximately 80% synchronization with observed rainfall-anomaly timing, indicating that roughly 4 out of 5 dekads have correctly predicted wet/dry phases. Probabilistic evaluation shows strong reliability. The debiased ranked probability skill score (RPSS) is high across thresholds, whereas the average ROC AUC (~0.68) indicates moderate discrimination. However, forecasts tend to under-predict very low rains in the GoG and very high rains in the Sahel. Using multiple datasets and robust metrics helps mitigate observational uncertainty. These results, for the first real-time S2S pilot over West Africa, demonstrate that ECMWF rainfall forecasts are skillful and actionable (especially up to 2–3 dekads ahead), providing confidence for early-warning and planning systems in the region.

Keywords: sub-seasonal-to-seasonal (S2S) forecasting; West Africa monsoon (WAM); rainfall variability; ECMWF-S2S model; satellite observations; evaluation



Academic Editor: Tin Lukić

Received: 30 June 2025

Revised: 29 August 2025

Accepted: 5 September 2025

Published: 11 September 2025

Citation: Olaniyan, E.A.; Woolnough, S.J.; Andrade, F.M.D.; Hirons, L.C.; Thompson, E.; Lawal, K.A.

Performance Evaluation of Real-Time Sub-to-Seasonal (S2S) Rainfall Forecasts over West Africa of 2020 and 2021 Monsoon Seasons for Operational Use. *Atmosphere* **2025**, *16*, 1072.

<https://doi.org/10.3390/atmos16091072>

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the Creative Commons Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

West Africa has a growing need for accurate and actionable weather forecasts to build resilient livelihoods and infrastructure in support of sustainable socio-economic development [1,2]. Understanding the predictability and enhancing the provision of

extended-range sub-seasonal-to-seasonal (S2S; 2 to 2 months ahead) forecasts has the potential to significantly inform anticipatory action before high-impact weather events [3,4]. S2S forecasts bridge the gap between short-term (1 to 15 days) and long-term (3 to 7 months) forecasts. Predictive skill in the former deriving from the initial conditions and in the latter from slowly varying modes of variability (e.g., oceanic circulations).

Huge potential exists for S2S forecasts to facilitate effective decision-making through early warning and proactive disaster mitigation [3]. This potential can only be realized if we have a thorough understanding of the large-scale drivers of predictability on these timescales and a thorough evaluation of their representation in models [5–7]. In addition, Ref. [8] emphasized the need for a comprehensive evaluation of numerical weather prediction products and services. Failure to do so may result in the disclosure of life-critical information whose value is unknown and which may be harmful.

The dynamics of boreal summer monsoon rainfall play a fundamental role in numerous socioeconomic activities (agriculture, hydropower, livestock and pastoralism, water resources, human health, transportation and construction, and climate change adaptation) in West Africa [9]. Therefore, it is crucial to have an accurate representation of the evolution and sustainability of these dynamics in models, especially for forecasting on S2S timescales. To achieve this objective, the purpose of this study is to evaluate the capability of operational S2S prediction systems over West Africa in 2020 and 2021 (the ECMWF-GCRF-AFRICAN SWIFT S2S Real Time Pilot (RTP) years), to gain a better understanding of the model's representation of monsoon dynamics.

The consensus from several studies [10–16] that have examined the skill of the S2S precipitation forecast over West Africa in recent years is that there is predictive skill in S2S precipitation forecasts, albeit with inherent biases. Despite such studies demonstrating the potential value of S2S forecasts, access to real-time S2S data from the Global Modelling centers for operational forecasting remains a challenge for National Meteorological and Hydrological Services (NMHS) across Africa. Global S2S forecast data is available, but on a three-week delay rather than in real-time, which, although useful for research, cannot inform the co-development of more appropriate services to support in-country anticipatory action.

However, as part of the World Meteorological Organization (WMO) S2S Prediction Project Real-Time Pilot (RTP), a joint initiative between the World Weather Research Programme (WWRP) and the World Climate Research Programme (WCRP), 16 international projects were granted access to S2S data in real-time for a pilot period of two years (extended to three due to COVID). One of these international project partnerships was the Global Challenges Research Fund African Science for Weather Information and Forecasting Techniques (GCRF African-SWIFT) project. By supporting partnerships between academic institutions, operational agencies, and forecast users in four African partner countries (Nigeria, Ghana, Senegal, and Kenya), the African-SWIFT aimed to improve the continent's long-term weather forecasting capacity [8]. As one of the operational institutions participating in the project, the Nigerian Meteorological Agency (NiMet) had real-time access to the European Center for Medium Range Weather Forecast (ECMWF) S2S forecast data. The primary objective of the African-SWIFT project was to use a coproduction approach to create tailored forecast products to better support decision-making across a range of weather-sensitive sectors [3]. The meteorological verification of such forecasts is essential for ensuring the sustainability and reliability of actionable coproduced forecasts. Therefore, this study aims to provide robust scientific evaluation of S2S rainfall prediction in West Africa and in doing so improve our understanding of windows of opportunity for operational real-time S2S forecasting in the region.

While previous studies have demonstrated S2S skill using hindcast data [5,11,17], two critical gaps limit operational implementation in West Africa. First, verification studies predominantly evaluate retrospective hindcasts rather than real-time operational forecasts that decision-makers actually receive, potentially overestimating practical utility [5]. Second, reliance on single observational datasets masks substantial uncertainty inherent in satellite-based precipitation estimates over data-sparse tropical regions. This study addresses these limitations through comprehensive real-time evaluation of ECMWF S2S forecasts during the 2020–2021 West African monsoon seasons. Our approach uniquely combines the following: (1) systematic assessment of operational forecasts ECMWF sub-seasonal as received by meteorological services, (2) quantifying observational uncertainty using two independent satellite datasets, and (3) development of skill-informed operational guidance for regional climate services.

This framework addresses three fundamental questions essential for the operational implementation of ECMWF S2S forecasts over West Africa in real-time.

This will establish the dependability of using the ECMWF S2S forecasts to answer the following real-time questions in West Africa. First, how accurately do real-time S2S forecasts capture West African monsoon evolution? Secondly, to what extent does observational uncertainty influence forecasting skill assessment? Thirdly, what is the comparative performance of real-time forecasts versus hindcast climatology? These insights provide essential foundations for confident operational implementation of S2S forecasts across West Africa.

While Section 1 introduces the paper’s concept, Section 2 will describe the ECMWF-S2S rainfall data, the satellite dataset used, and the method of analysis utilized in this study. Section 3 examines the results of the analyses, while Section 4 summarizes and concludes the work’s findings.

2. Data and Methods

2.1. S2S-Rainfall Data

In this study, both hindcast and real-time ensemble S2S rainfall forecasts are utilized. They are obtained from the ECMWF-S2S database [18], supported by the WMO through the WWRP and WCRP, as well as the GCRF African SWIFT project. As shown in Tables 1 and 2, the real-time rainfall forecasts for 2020 and 2021 are retrieved every 14 days between the last initialized forecast in March and the first initialized forecast in October. This indicates that there are two to three start dates (hereafter referred to as SDs) in a month, excluding March and October. April, May, June, July, and September all have two SDs, while only August has three SDs.

Table 1. Summary of data sources and characteristics.

Data	Variable	Source	Months	Periods	Resolution	Ensemble Size	Time-Step
ECMWF Realtime	Rainfall	ECMWF	March–October	2020–2021	150 km	51	Daily
ECMWF Hindcast	Rainfall	ECMWF	March–October	2001–2019	150 km	11	Daily
Observed	Rainfall	GPM_IMERG	March–November	2001–2021	10 km	NA	Daily
Observed	Rainfall	TAMSAT	March–November	2001–2021	5 km	NA	Daily

Table 2. Summary of initialization dates and forecast validity periods for real-time ECMWF S2S evaluation, aligned with West African monsoon phases.

Fct. Days	Start-Date (SD; 2020)	Valid Until	Start-Date (SD; 2021)	Valid Until	Period	Lead Time (LT)
1	30 March	8 May	29 March	7 May	40 days	10 days
2	13 April	22 May	12 April	21 May	40 days	10 days
3	27 April	5 June	26 April	4 June	40 days	10 days
4	11 May	19 June	12 May	18 June	40 days	10 days
5	25 May	3 July	24 May	2 July	40 days	10 days
6	8 June	17 July	7 June	16 July	40 days	10 days
7	22 June	31 July	21 June	30 July	40 days	10 days
8	6 July	14 August	5 July	13 August	40 days	10 days
9	20 July	28 August	19 July	27 August	40 days	10 days
10	3 August	11 September	2 August	10 September	40 days	10 days
11	17 August	25 September	16 August	24 September	40 days	10 days
12	31 August	09 October	30 August	8 October	40 days	10 days
13	14 September	23 October	13 September	22 October	40 days	10 days
14	28 September	7 November	27 September	6 November	40 days	10 days
15	12 October	21 November	13 October	20 November	40 days	10 days

Although the ECMWF S2S rainfall forecast is initialized twice a week, we utilized the 2 weekly (14 days) forecasts for strategic data management, operational efficiency, trend detection, and noise minimization. For instance, the strategic data management helps in maintaining a manageable dataset, avoiding potential information overload and computational strain. Secondly, a 2 weekly data interval is long enough to detect meaningful trends and shifts in weather patterns, but short enough to remain responsive to significant changes.

The SDs are also divided into pre-monsoon (PREMON; March and April), first southern monsoon peak (FSMP; May and June), northern monsoon peak (NMP; July and August), and second southern monsoon peak (SSMP; September and October) seasons. This may differ slightly from standard classification, but this approach (dividing SDs into specific monsoon phases) enhances the meteorological relevance, operational applicability, predictive accuracy, and historical consistency of the study.

Following the 15 SDs, the hindcast rainfall data for the same months from 2001 to 2019 were retrieved “on the fly” (hindcast produced concurrently with the near real-time forecasts) using only the 2020 forecasts.

2.2. Satellite-Rainfall Data

Two independent satellite-derived observed daily rainfall datasets (Table 1) are used to validate the S2S rainfall real-time forecasts and hindcasts to ensure a comprehensive evaluation of skill and also to assess the sensitivity of skill under different reference datasets. These are the (Tropical Applications of Meteorology using SATellite (TAMSAT) version 3 [19] and the version 6 of the Integrated Multi-satellitE Retrievals for the Global Precipitation Measurement mission (GPM_IMERG) [20]. GPM_IMERG rainfall data is available with a grid spacing of 10 km and global coverage. Several studies (e.g., [21–23]) have utilized GPM_IMERG for model verification over Africa despite its inherent biases [22,24]. The GPM_IMERG system is run twice in near-real-time, with the Early run (IMERG-E; about six hours after normal observation time) and the Late run (IMERG-L after about 18

h) [20,25,26]. For this study, however, the dataset from the final run (IMERG-F; about 3 months after the observation month) was used. With both forward and backward morphing schemes [22,24], the IMERG final run better described the intensity in the rainfall features.

The TAMSAT-V3 daily rainfall estimates are derived from the TAMSAT rainfall estimation algorithm [19,27,28]. It employs thermal infrared (TIR) imagery to monitor the cold cloud tops of rain-producing convective cumulonimbus systems, which are a useful indicator of rainfall in the tropics [29]. The pentad rainfall estimates are disaggregated into daily values based on the amount of cold cloud duration (CCD) observed on each day. According to [19], each daily CCD map is created by analyzing all TIR images from 0600 to 0600 UTC the following day, to correspond with the typical daily observation period of West Africa rain gauges. It is available with a grid spacing of 5 km.

The two satellite-derived datasets are utilized for two primary purposes. First, the rainfall network in Africa is non-stationary and, in some regions, sparse [23]. So, utilizing satellite-derived observations guarantees continent-wide consistency. The second reason is that the verification against two satellite-derived rainfall datasets permits the estimation of rainfall uncertainty based on the application of different algorithms [30]. Using GPMIMERG with both oceanic and land estimates, for instance, will permit the evaluation of the evolution of the monsoon from the ocean to the continent.

2.3. Methods

The quality of the S2S rainfall forecasts and hindcasts is evaluated using verification metrics following [5]. Here, the forecast quality assessment is conducted by using dekadal (10-day) temporal aggregation with four distinct lead times (LTs) ahead of each of the 15 SDs: LT1: LT1 forecasts days 1–10 (first dekad); LT2 forecasts days 11–20 (second dekad); LT3 forecasts days 21–30 (third dekad); and LT4 forecasts days 31–40 (fourth dekad), (see Table 2). Utilized are both deterministic and probabilistic forecast evaluation methods. This is to provide a comprehensive and nuanced understanding of rainfall forecast performance in order to have clear insights into forecast accuracy, temporal alignment, and reliability and uncertainty of forecasts.

The deterministic forecast evaluation is based on bias and Pearson correlation (r), while the two probabilistic forecast evaluations used are the ranked probability skill score (RPSS) and relative operating characteristic (ROC). To quantify the inherent systematic error, bias in this study is calculated as the mean error (ME) between the S2S forecast and the observed rainfall from both GPMIMERG and TAMSAT datasets, as expressed in Equation (1).

$$BIAS = \frac{1}{N} \sum_{i=1}^N F_i - O_i \quad (1)$$

where N denotes the sample size, F_i , the forecast totals, and O_i , the observation totals. It is computed from the ensemble mean (forecast and hindcast) for all LTs from each SD.

The Pearson correlation coefficient (r) is calculated to quantify the strength and direction of the linear relationship between the S2S forecasts and the observed rainfall data. This information is vital for decision-makers in agriculture, water resource management, and other sectors that depend on accurate rainfall predictions especially within the S2S time-scale. So, the correlation is determined between the ensemble mean of S2S forecasts and observations for all LTs from each SD, and from all SDs for each LT. Then, seasonal mean correlation and correlation for all LTs from all SDs were determined. This approach is to ensure that the evaluation captures the variability and consistency of forecasts across different temporal horizons and initial conditions, enhances the reliability of seasonal forecasts, and supports informed decision-making across various sectors.

Following established practices in S2S verification for West Africa [14,31], we employ the synchronization metric [32–34] to quantify temporal alignment between forecast and observed rainfall anomaly signals. Synchronization (Equation (2)) measures the percentage of forecast periods in which the sign of the predicted anomaly matches the observed anomaly. In other words, it quantifies the fraction of dekads when the model correctly predicts above- vs. below-normal rainfall, crucial information for operational decision-making.

$$\text{Synchronization} = \left(\frac{1}{N} \right) \times \sum [\text{sign}(F_i) = \text{sign}(O_i)] \times 100\% \quad (2)$$

where F_i , and O_i are standardized forecast and observed anomalies at time step i , N is the total number of time steps, $\text{sign}()$ function returns +1 for positive values, and −1 for negative values. The equality operator returns 1 when signs match, and 0 otherwise.

Here, the anomalies are computed using a leave-one-out cross-validation method [20]. RPSS is computed from each SD for each LT and from all SD for each LT with subsequent computation of the seasonal RPSS following [35].

The RPSS (Equation (3)) is calculated using rainfall thresholds as described in [36,37]. Rainfall amounts less than 10 mm, greater than 25 mm but less than 50 mm ($50 \text{ mm} > r > 25 \text{ mm}$), greater than 50 mm but less than 100 mm ($100 \text{ mm} > r > 50 \text{ mm}$), and greater than 100 mm are used as thresholds. Each of these thresholds corresponds to various rainfall levels, representing a spectrum of meteorological conditions and impacts.

For instance, rainfall of <10 mm is sufficient for dust suppression and could result in surface cooling. Also, rainfall in the range of 25–50 mm is effective in soil moisture recharge and promoting plant growth. The threshold of 50–100 mm could be crucial for monitoring increases in river flow and flood preparedness. This range represents heavy rainfall that normally led to localized flooding, especially in urban areas with impervious surfaces. Additionally, the threshold > 100 mm is critical for assessing major flood risks and could potentially trigger landslides in susceptible areas.

$$\text{RPSS} = 1 - \frac{\text{RPS}}{\text{RPS}_r} \quad (3)$$

For the hindcast, the RPSS is expressed as

$$\text{RPSS}_D = 1 - \frac{\text{RPS}}{\text{RPS} + D} \quad (4)$$

The D in Equation (4) is a debiased (discrete) RPSS to overcome the sensitivity which can be introduced due to the small ensemble size in the hindcast [35]. It is expressed as

$$D = \frac{1}{M} \left(\frac{K^2 - 1}{6K} \right) \quad (5)$$

where M is the ensemble size for K equiprobable categories (0.25 for four thresholds in this case).

To evaluate the ability of the forecast to discriminate between events and non-events across different probability thresholds, we computed the ROC [38]. Here, the ROC diagram is created by plotting the hit rate against the false alarm rate for all LTs from all SDs in the hindcast at different probability thresholds. The area under the ROC curve (AUC) is computed as a score with rainfall thresholds of less than 10 mm, greater than 100 mm, and the anomaly.

In addition, West Africa is divided into four zones based on the common climatology of the onset of rainfall. These include the Gulf of Guinea (GoG; latitudes 4 to 8 degrees north; region 1), the Guinea Forest (GFR; latitudes 8 to 10 degrees north; region 2), the

Savannah (SAV; latitudes 10 to 12 degrees north; region 3), and the Sahel (SAH; latitudes 12 to 16 degrees north; region 4), (Figure 1a). All datasets utilized in this study were re-gridded to $1^\circ \times 1^\circ$ for uniformity and because it is the lowest common resolution.

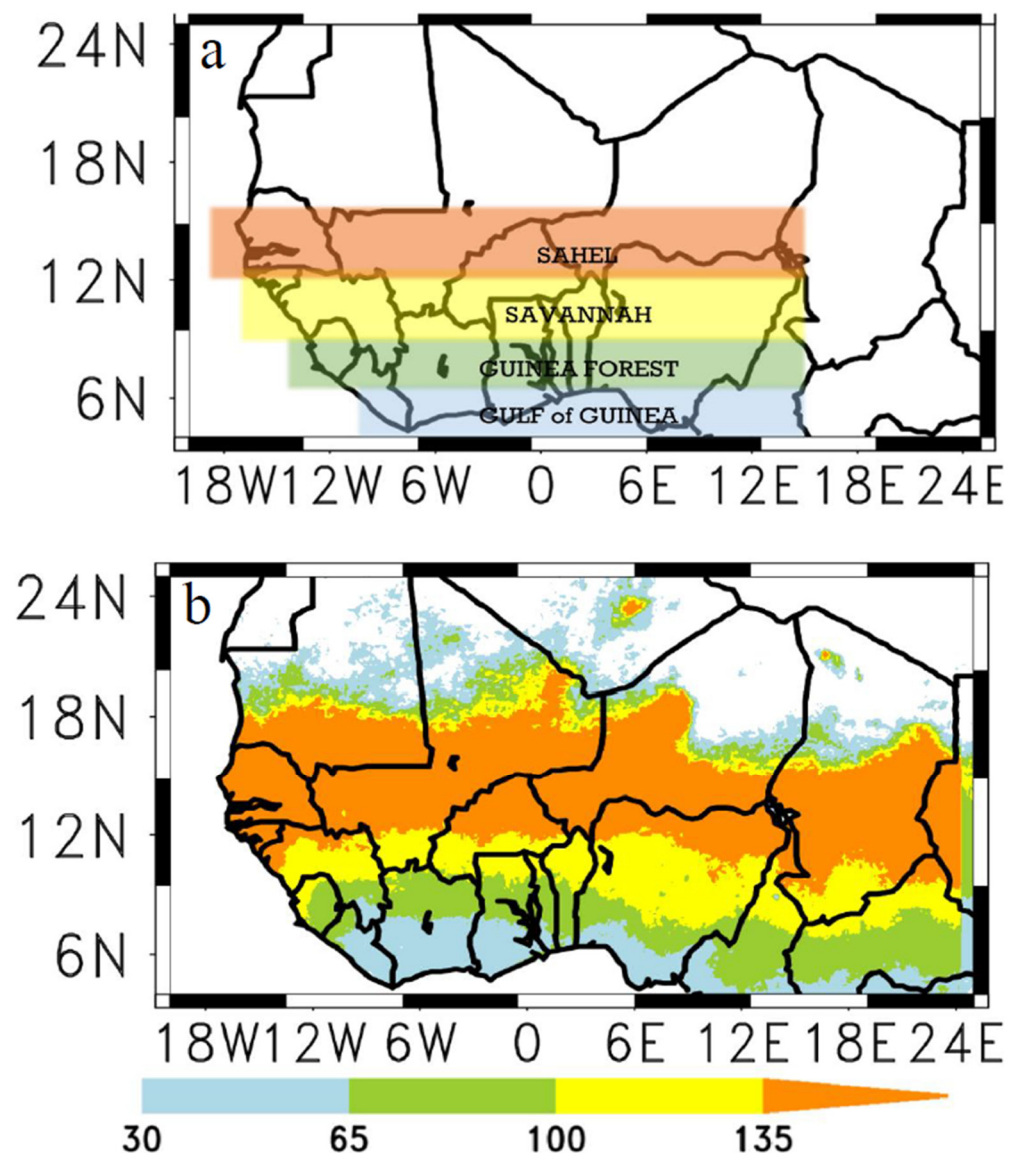


Figure 1. (a) Classification of ecological zones in West Africa: Gulf of Guinea (GoG region1; blue; 4° N– 8° N, 10° W– 15° NE), Guinea Forest (GFR region2; green; 8° N– 10° N, 12° W– 15° E), Savannah, (SAV region3; yellow; 10° N– 12° N, 14° W– 15° E), Sahel (SAH region4; brown; 12° N– 16° N, 18° W– 15° E), and (b) the rainfall onset climatology derived from the GPM IMERG satellite dataset.

3. Results

3.1. Rainfall Characteristics

On S2S timescales, the ECMWF model captures the evolution of monsoon rainfall throughout the forecast periods. This includes the southern bi-modal and the uni-modal attribute of the North, as well as the monsoon jump [39] from the Gulf of Guinea (GoG) to the Sahel (SAH) (Figure 2). The model effectively captures both the onset and retreat of the Little Dry Season (LDS), particularly over the GoG.

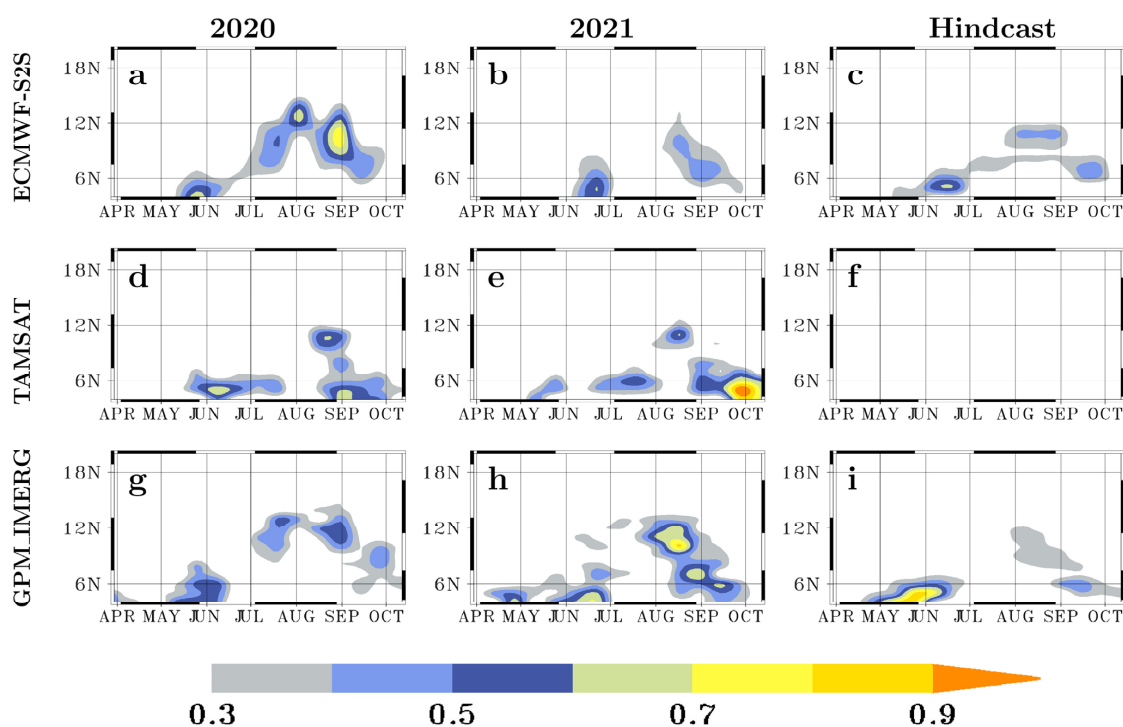


Figure 2. Hovmöller diagram (averaged along longitude 18° W to 15° E) illustrating the ECMWF-S2S probability forecast (a–c) and observed dekadal rainfall accumulation (binary format) exceeding 100mm, from, TAMSAT (d–f), and GPM IMERG (g–i) for lead time 1 (LT1) from start date 1 (SD1) to start date 15 (SD15) of 2020, 2021, and the hindcast.

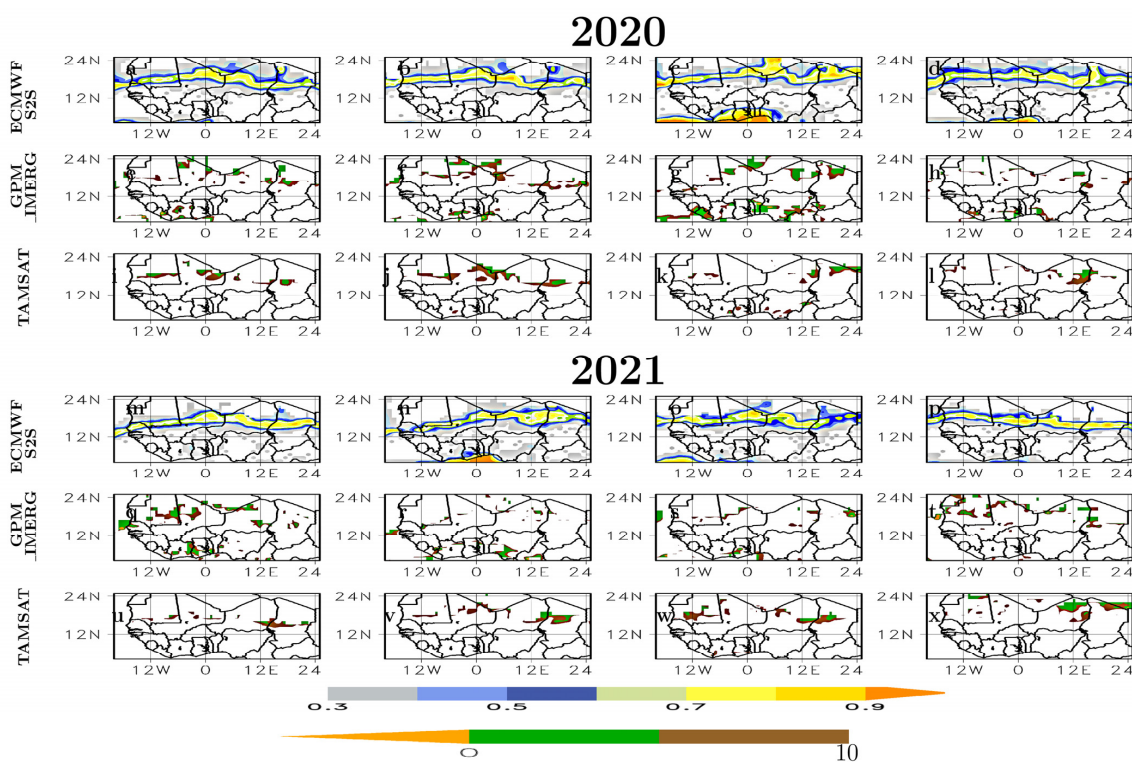


Figure 3. Probability forecasts of dekadal rainfall accumulation exceeding 100 mm from ECMWF-S2S for LT1 of SD8–SD11 (panels (a–d,m–p)), compared with observed dek-adal rainfall accumulation ≥ 100 mm from GPM IMERG (e–h,q–t) and TAMSAT (i–l,u–x) during 2020 and 2021.

As displayed in Figure 2a–i, rainfall events greater than 100 mm observed in both TAMSAT and GPM_IMERG may have similar latitudinal structure, except for the temporal shift of the coastal rainfall high. However, significant spatial inconsistencies exist between the datasets and discrepancies are evident when compared to hindcast simulations. Notably, the extreme rainfall events exceeding 100 mm as observed by GPM_IMERG are not captured by TAMSAT in the hindcast. These disparities may be attributed to the unique characteristics of each satellite dataset, including differences in sensors, algorithms, and spatial resolution [40, 41]. For instance, GPM_IMERG provides global coverage including oceanic regions, whereas TAMSAT focuses exclusively on the African continent, excluding ocean areas.

However, there are notable discrepancies in rainfall representation between TAMSAT and GPM_IMERG datasets. Although varying between years, the model shows high probability for rainfall less than 10 mm during the peak of the LDS, as illustrated in Figure 3c,o. The intensity of the LDS appears less pronounced in TAMSAT compared to GPM_IMERG for both years.

Although the performance varies over different years, the model adequately predicts areas of heavy rainfall, albeit with some discrepancies in intensity and spatial errors during the peak of the monsoon (Figure 4a–x). For instance, areas such as Guinea, Liberia, and Sierra Leone, where the model indicated high probability, are regions where heavy rainfall (>100 mm) was observed in both GPM_IMERG and TAMSAT. Other areas of consistent high probability and observed heavy rainfall in both GPM_IMERG and TAMSAT include the Mambila plateau regions of Nigeria and Cameroon. However, the intensity of heavy rainfall observed in GPM_IMERG, particularly in August (Figure 4c,g,k,p,t,x), is not captured in TAMSAT. The differences in detecting extreme rainfall between the two satellite datasets may be attributed to disparities in measurement techniques, sensor sensitivities, and temporal and spatial resolutions. For instance, the multi-sensor approach employed by GPM_IMERG, which integrates data from various satellite instruments, may enhance its ability to detect and quantify extreme rainfall events.

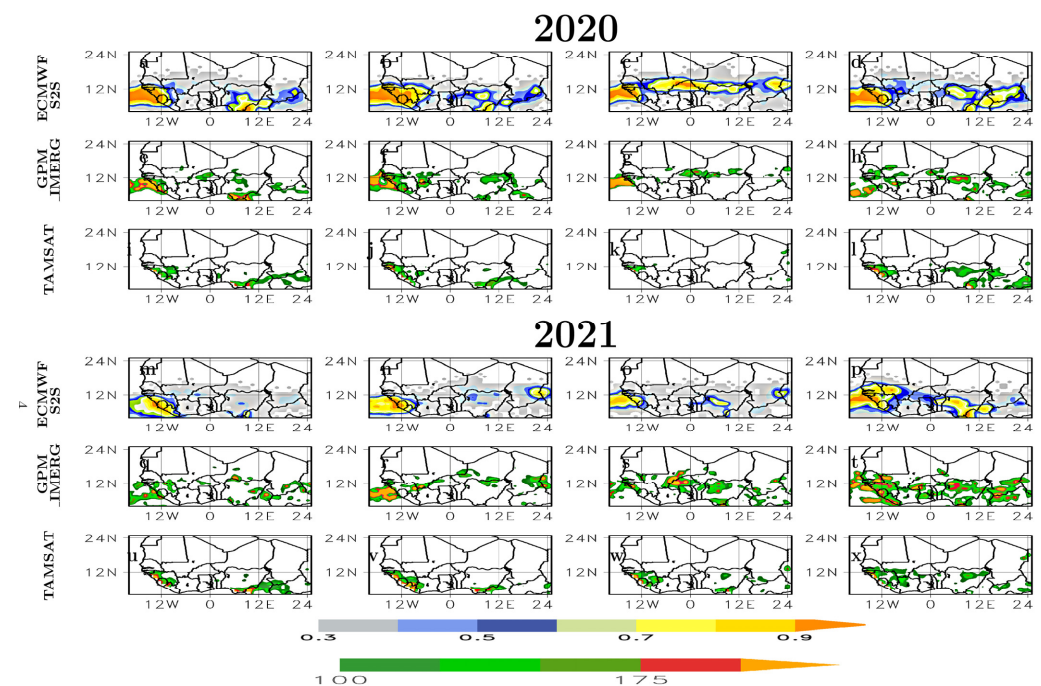


Figure 4. Probability forecasts of dekadal rainfall accumulation exceeding 100 mm from ECMWF-S2S for LT1 of SD8–SD11 (panels (a–d,m–p)), compared with observed dekadal rainfall accumulation ≥ 100 mm from GPM IMERG (e–h,q–t) and TAMSAT (i–l,u–x) during 2020 and 2021.

3.2. Rainfall Bias

The rainfall biases between the ECMWF-S2S model, GPM-IMERG, and TAMSAT during the RTP years (2020, 2021) and the hindcast period show notable differences across various west Africa regions (Figure 5). For instance, the model bias ranges from -14.3 mm to 25.6 mm in 2020, -21 mm to 28 mm in 2021, and -24.1 mm to 36.7 mm in the hindcast. The model's bias relative to GPM-IMERG between June and August is predominantly wet in both 2020 and 2021, with the strongest positive bias observed in July 2020 and between July and August 2021 (Figure 5a). Over the Sahel, the model exhibits a negative bias in both years, more pronounced in 2021 (Figure 5c), indicating drier conditions especially before and after the peak of the monsoon. Despite the wet conditions observed between August and September in the GFR to SAH regions and in October from the GoG to GFR, the model consistently shows drier conditions in the hindcast period (Figure 5e).

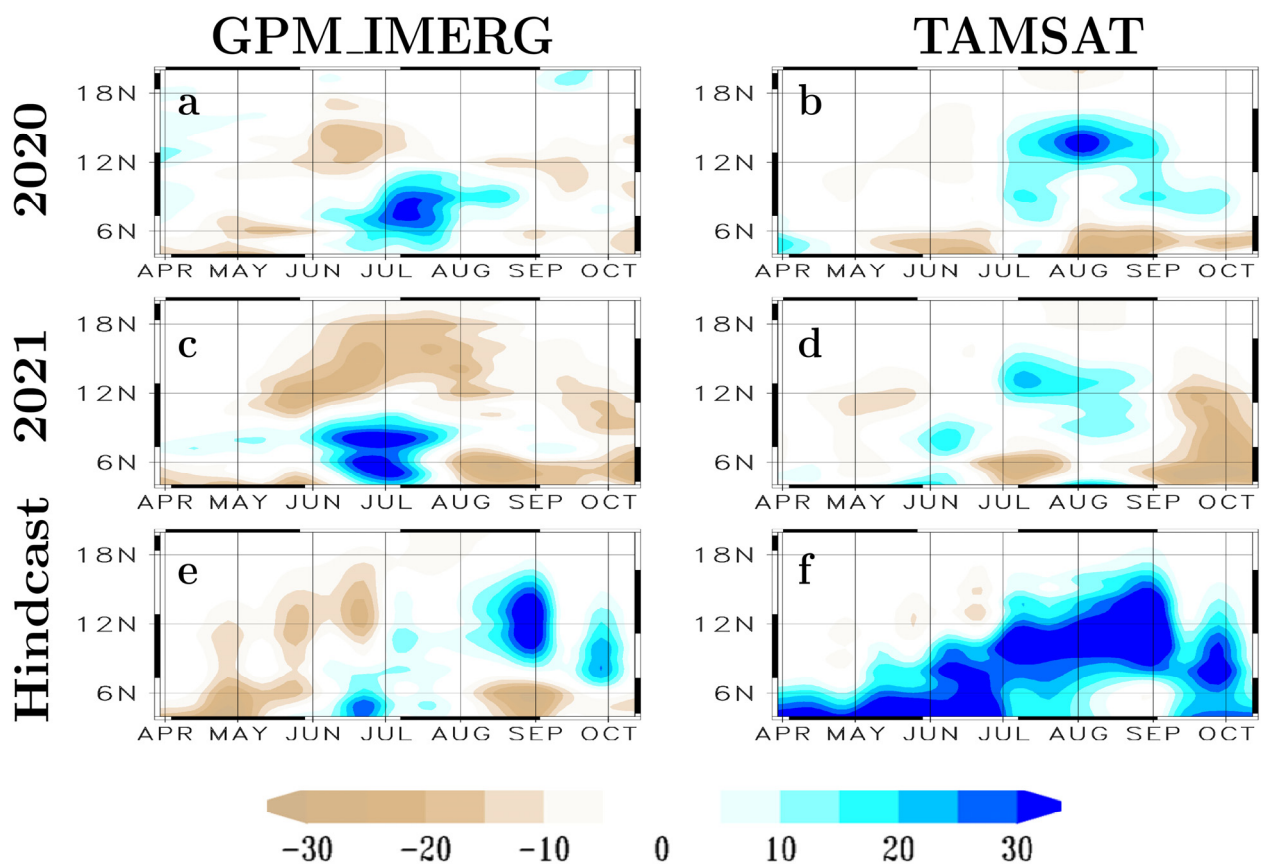


Figure 5. Hovmöller diagram (average along longitude 18° W to 15° E) showing the dekadal rainfall bias (mm) between ECMWF-S2S ensemble mean and observed (GPM IMERG (a,c,e), and TAMSAT (b,d,f)) in real-time of 2020, 2021, and hind-cast.

When comparing the ECMWF bias to TAMSAT in real-time forecasts, a consistent pattern of dryness in the northern (above latitude 12° N) regions and wetness in the southern (below latitude 12° N) regions is observed, particularly from July to August (Figure 5b,d). The model tends to be wetter in the hindcast compared to the real-time forecast, with a noticeable decrease in wetness towards the Sahel from April to July and an increase from July to October (Figure 5f).

The discrepancies in rainfall biases between the ECMWF-S2S model, GPM-IMERG, and TAMSAT demonstrate numerous significant aspects of the model's performance and regional climate variability [42]. For instance, the range of biases observed across different years and subregions highlights the inherent difficulties in reliably estimating rainfall in a

climatic system as complex and highly changeable as that of West Africa. The wetter conditions predicted by the model relative to GPM-IMERG during the monsoon season could be due to the model's sensitivity to convective processes and parameterization approaches.

Also, the pronounced negative bias over the Sahel, particularly evident in 2021, can be attributed to multiple factors within the model's framework. Primarily, it may stem from inadequacies in the model's representation of dry season dynamics and the transition periods surrounding the monsoon peak. This bias could be a consequence of imprecise soil moisture initialization, which plays a crucial role in land–atmosphere coupling processes that significantly influence rainfall distribution in this region.

Additionally, the consistent dryness observed in the hindcast period, especially before and after the monsoon peak, might reflect the model's tendency to underestimate rainfall persistence and the occurrence of smaller-scale convective events that are not well-resolved at coarser resolutions. Furthermore, the observed north–south gradient in biases relative to TAMSAT, with wetness in the south and dryness in the north, could be influenced by regional differences in satellite observation capabilities.

3.3. Rainfall Anomaly

The ECMWF-S2S model effectively captures the inter-annual variability of decadal rainfall anomalies across different lead times (LTs) from all start dates (SDs) and regions, albeit with some discrepancies (Figure 6). When compared to GPM-IMERG observations, the model successfully reproduced the interannual variability of anomalies across all regions, except in 2018 and 2019 for various LTs over the Gulf of Guinea (GoG) (Figure 6a–d). Notably, LT2 showed no years of deviation, whereas deviations were observed in 2006 and 2012 for LT1 over the Guinea Forest Region (GFR), and in 2006 and 2010 for LT3 over the same area (Figure 6e,g,h). Additionally, consistent deviations from observed anomalies were noted in 2003 and 2012 across all LTs over the Savannah (SAV) region, and in 2019 for LT2 and LT3. Other deviations over the SAV region included 2006 for LT1 and 2008 for LT2 (Figure 6i–l). Over the Sahel (SAH) region, the model consistently deviated in 2003 and 2012 across all LTs, except in 2010 for LT2 and LT4.

Unlike GPM-IMERG, TAMSAT observations show consistent negative rainfall anomalies in 2001, 2002, 2017, 2018, and 2019 across all regions and LTs (Figure 6). Despite some synchronization skill, TAMSAT anomalies often fall outside the spread of the model ensemble members. This indicates a need for further refinement in the model to better capture the variability observed in TAMSAT data.

While minor deviations exist, the model's ensemble-mean synchronization is generally strong but varies depending on the region, LT, and observation used (Figure 6a,b). For instance, synchronization with TAMSAT improves from the GoG to the SAH, but remains consistently high across all LTs when using GPM-IMERG. With a synchronization value of approximately 80 percent across the regions, it implies that the model can predict on average 8 out of 10 correct decadal rainfall anomalies, regardless of LT. However, using TAMSAT data, the model can reliably predict 6 out of 10 observed anomalies over the SAV and SAH, but it may fail to predict 3 out of 10 observed anomalies over the GoG. Additionally, the model's synchronization skill across the LTs is not sequential but varies for each region and observation used. The model's strong synchronization skill, particularly with GPM-IMERG data, demonstrates its ability to capture the overall pattern of rainfall anomalies.

The model's ability to capture inter-annual variability of decadal rainfall anomalies across different regions and LTs, despite some deviations, is indicative of its potential and limitations. However, the inconsistencies between the ECMWF-S2S model and both GPM-IMERG and TAMSAT observations highlight a few key aspects of model performance and observational challenges.

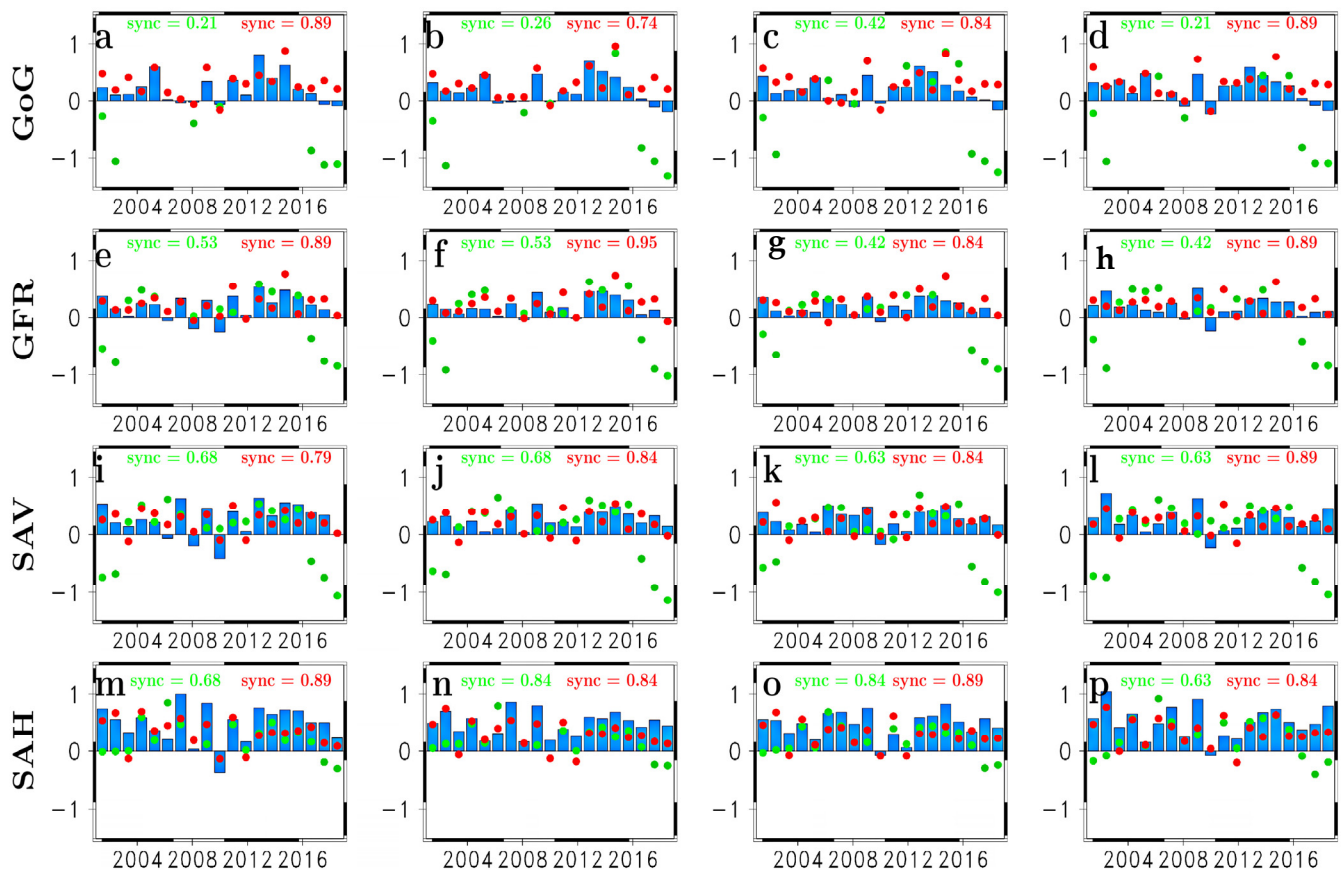


Figure 6. Inter-annual variability of the standardized dekadal rainfall anomaly over the GoG (a–d), Guinea Forest (e–h), Savanah (i–l), and Sahel (m–p) from the ensemble mean (blue bar), with the GPM IMERG (red circle) and TAMSAT (green circle). Also, the synchronization (sync) skill of the model anomalies with GPM IMERG (red) and TAMSAT (green).

The observed deviations, particularly in specific years and regions, may be suggesting areas where the model's representation of climate processes could be improved. For example, the deviations in 2018 and 2019 over the GoG when compared to GPM-IMERG data might be due to regional climatic anomalies or deficiencies in the model's parameterization of convective processes. Similarly, the consistent deviations observed over SAV and SAH in 2003 and 2012 across multiple regions and LTs indicate potential issues in the model's representation of large-scale atmospheric circulation patterns influencing these areas.

Also, the fact that TAMSAT anomalies often fall outside the model's ensemble spread may be connected to the inability of the model to fully capture the variability seen in TAMSAT data. This could also be due to differences in spatial resolution, observation techniques, or inherent biases in the satellite data, particularly over the GoG.

3.4. Correlation Skill

Across regions, start dates (SDs), lead times (LTs), seasons, and anomalies, the relationship between observations and the ECMWF-S2S model exhibits considerable variability (Figure 7b,d, 8 and 9). The correlation between the GPM-IMERG and the model in real-time, hindcast, and anomaly analysis varies significantly. For instance, the correlation between the model and the rainfall anomaly from all SDs for all LTs across every region and season is generally weak (Figure 7a,b), whereas the hindcast exhibits a stronger correlation. Notably, the strongest correlation is observed over the Sahel (SAH), but there is no significant, steady progression in correlation strength from the Gulf of Guinea (GoG) to the SAH (Figure 7a).

A similar correlation pattern, with some deviations, is observed between the model and TAMSAT (Figure 7b). Like the correlation with GPM-IMERG, the model's correlation with TAMSAT is strongest over the SAH, except for 2020. However, contrary to the GPM-IMERG-model relationship, the hindcast shows the strongest correlation only over the Savannah (SAV) and the SAH. Interestingly, the correlation strength generally increases towards the SAH, but the strongest correlation in 2020 is over the SAV (Figure 7b). Seasonal correlations with GPM-IMERG reveal that hindcast correlations are the strongest, while anomaly correlations are the weakest (Figure 7c). Consistently low correlations are observed during the First Southern Monsoon Peak (FSMP) (May–June) and the Northern Monsoon Peak (NMP) (July–August), while the pre-monsoon (PREMON) (March–April) and Second Southern Monsoon Peak (SSMP) (September–October) exhibit stronger correlations. The model's correlation with TAMSAT mirrors this pattern, except in 2020, and shows inconsistent strength in the hindcast correlation (Figure 7d).

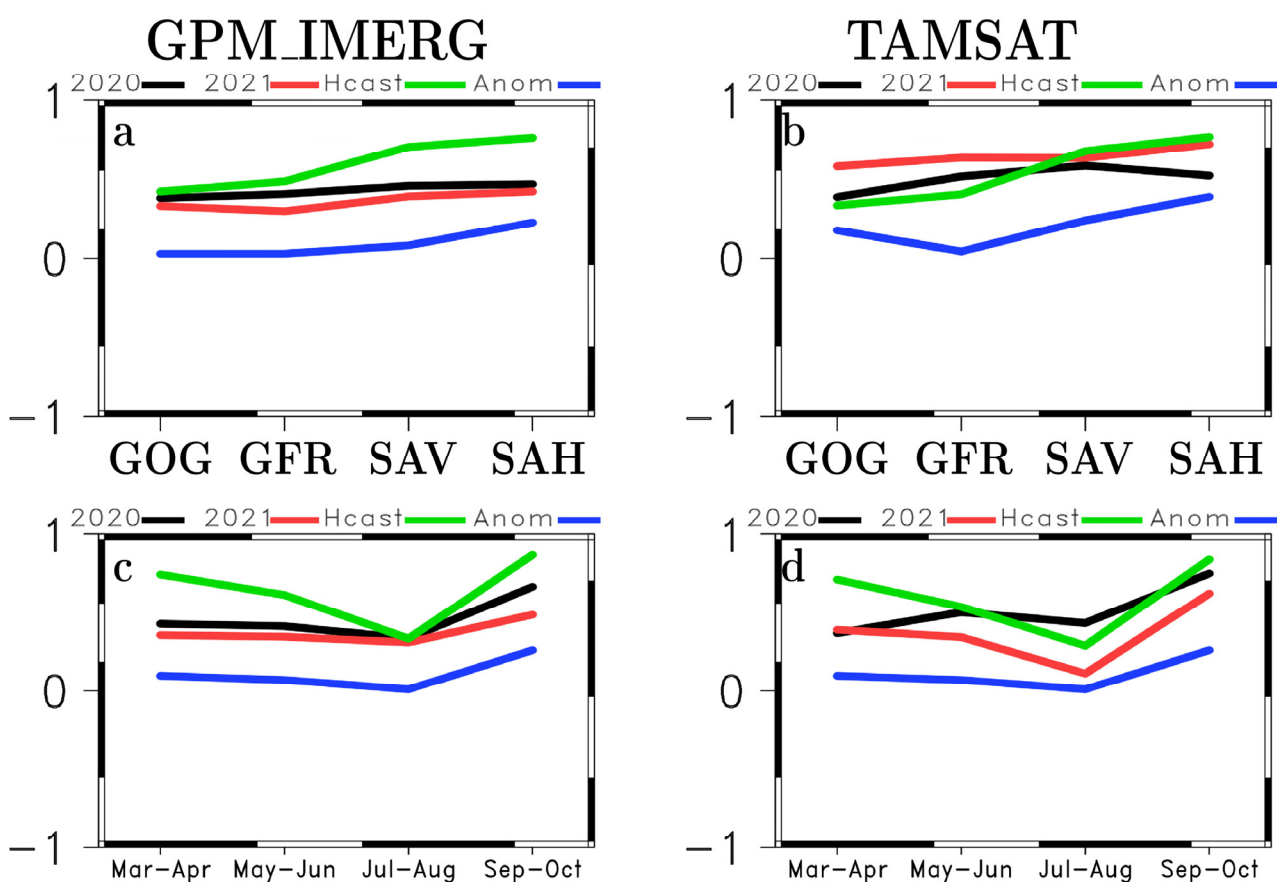


Figure 7. ECMWF-S2S ensemble mean correlation from all start dates (SDs) for all lead times (LTs) between the ECMWF-S2S ensemble mean and GPM IMERG (a) and TAMSAT (b). Seasonal correlation between the ECMWF-S2S ensemble mean and GPM IMERG (c) and TAMSAT (d) across all regions in real-time (black line for 2020 and red line for 2021), hindcast (red line), and anomaly (blue line).

The variability in correlations between the ECMWF-S2S model and both GPM-IMERG and TAMSAT observations highlights several important aspects of model performance and the challenges of accurately predicting rainfall anomalies across different regions and seasons. The observation that the strongest correlations are found over the SAH, while not showing a clear progression from GoG to SAH, indicates that the model performs better in certain climatic zones than others. This could be due to better representation of the climatic and meteorological processes specific to the SAH region within the model. Seasonal variations in correlation strength reveal that the model performs better during

the pre-monsoon and second monsoon peak periods compared to the FSMP and NMP seasons. This could be attributed to the model's handling of seasonal transitions and its ability to capture smaller-scale, less intense rainfall events more effectively than the major monsoon peaks.

Additionally, across each region, the model's correlation with GPM-IMERG and TAMSAT displays varying spatio-temporal attributes in the anomaly, hindcast, and real-time forecasts for each lead time (LT) from all start dates (SDs) (Figure 8). As depicted in Figure 8, there is no discernible pattern of correlation strength increasing or decreasing across LTs for all regions. For instance, in the anomaly, the model's correlation with GPM-IMERG in LT4 is generally the strongest across all regions except the Gulf of Guinea (GoG), where LT3 shows the highest correlation (Figure 8g).

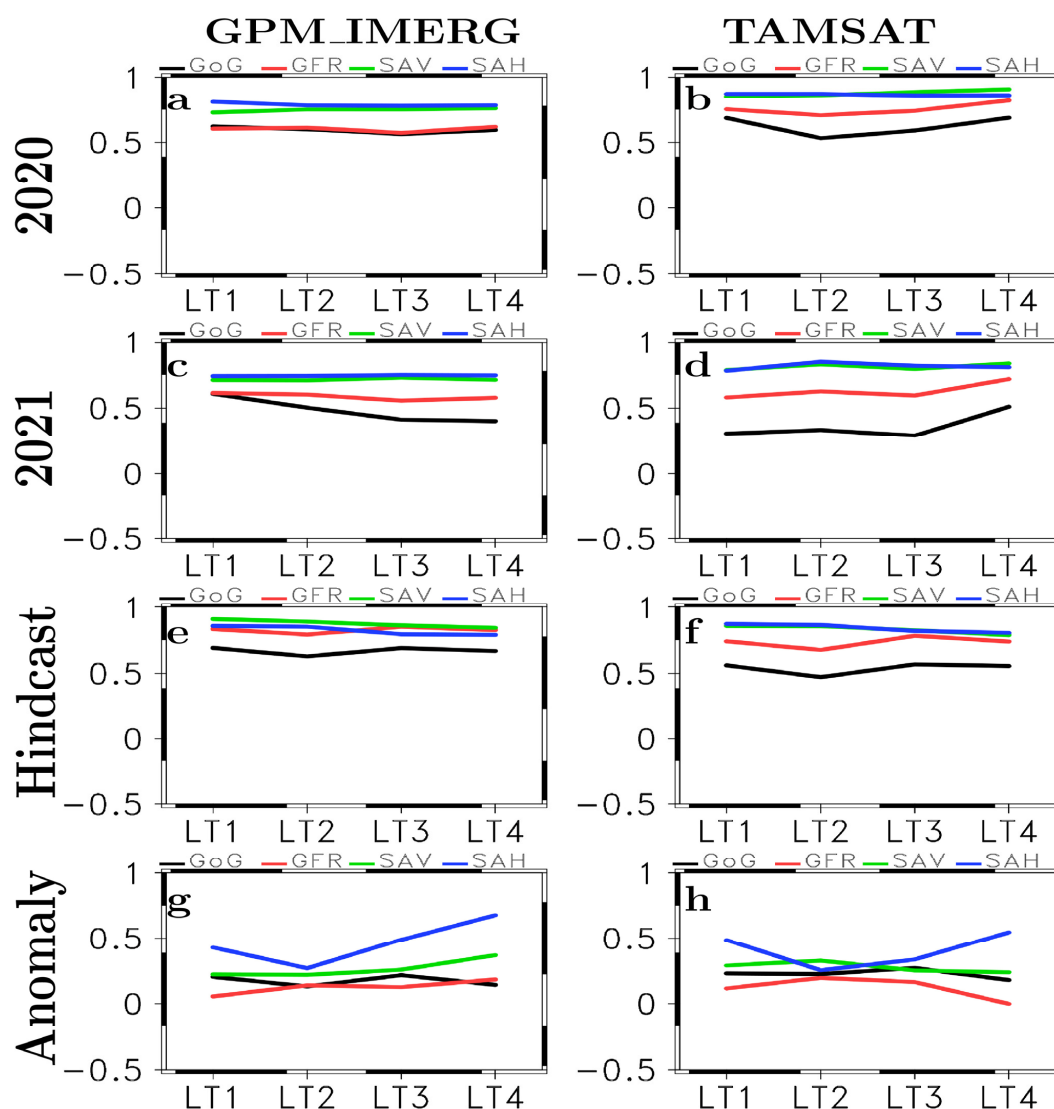


Figure 8. The correlation between ECMWF-S2S ensemble mean and observed (GPM IMERG (a,c,e,g) and TAMSAT (b,d,f,h)) for each LT from all SDs, over Gulf of Guinea (Reg1; black line), Guinea Forest (Reg2; red line), SAV (Reg3; green line), and SAH (Reg4; blue line).

The LT with the strongest and weakest correlations relative to GPM-IMERG in real-time and hindcast also varies across regions. In the hindcast, for example, the Savannah (SAV) and Sahel (SAH) regions showed LT1 and LT4 to have the strongest and weakest correlations, respectively, while LT2 had the weakest correlation over both the GoG and Guinea Forest Region (GFR) (Figure 8e). In 2020, LT1 showed the strongest correlation

over the GoG and the SAH, while LT4 had the strongest correlation over the GFR and SAV (Figure 8a). However, except for the SAV, LT3 exhibited the weakest correlation across all regions. In 2021, LT1 had the strongest correlation over the GoG, LT4 over the GFR, and LT3 over the SAV and the SAH (Figure 8c).

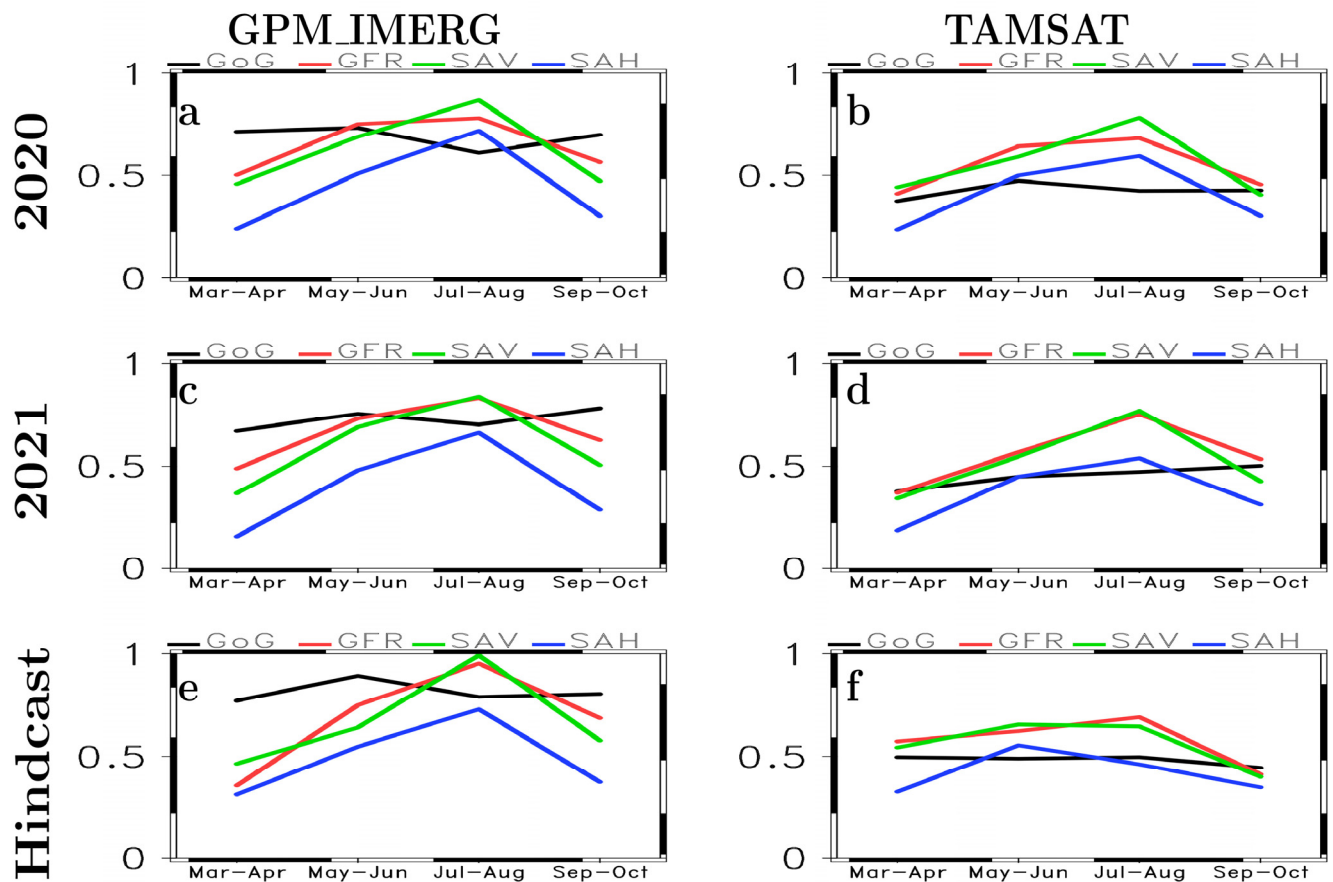


Figure 9. Rank and Discrete Rank Probability Skill Score (RPSS and RPSSD) in real-time and hindcast between S2S and GPM IMERG (a,c,e) and between S2S and TAMSAT (b,d,f) for each season, of 2020, 2021, and hindcast over the GOG (black line), GF (red line), SAV (green line), and the SAH (blue line).

Similarly, the model's correlation with TAMSAT for each LT from all SDs is comparable to that with GPM-IMERG but with notable differences. In the anomaly (Figure 8h), LT4, LT3, and LT2 exhibit the strongest correlations over the SAH, GoG, GFR, and SAV, respectively. In the hindcast, LT1 and LT4 show the strongest correlations over SAV and SAH, while LT2 has the weakest correlation over both the GoG and GFR (Figure 8f). As with GPM-IMERG, the LT with the highest correlation skill in real-time is inconsistent across regions. Figure 8b,d demonstrate that, except for the SAH where LT2 has the strongest correlation, LT4 has the strongest correlation in both 2020 and 2021. The inconsistencies in the strongest and weakest correlations across LTs indicate that the model's ability to predict rainfall anomalies is not uniform. This variability could be due to several factors, including differences in the model's representation of atmospheric processes, the quality and resolution of the observational data, and the inherent unpredictability of weather patterns in different regions.

3.5. Ranked Probability Skill Score

As the Rank Probability Skill Score (RPSS) decreases toward the Sahel (SAH), the model's reliability skill across all rainfall thresholds exhibits varying degrees of disparity over each region from each start date (SD) at each lead time (LT) (Figures 9 and 10). Despite

the consistently low correlation during the Northern Monsoon Peak (NMP), the RPSS during this season is the strongest except over the Gulf of Guinea (GoG), regardless of the type of observation used (Figure 9). Although weak during the NMP (July–August), the RPSS relative to GPM-IMERG during the First Southern Monsoon Peak (FSMP) is strong over the GoG and Guinea Forest Region (GFR) in both real-time and hindcast.

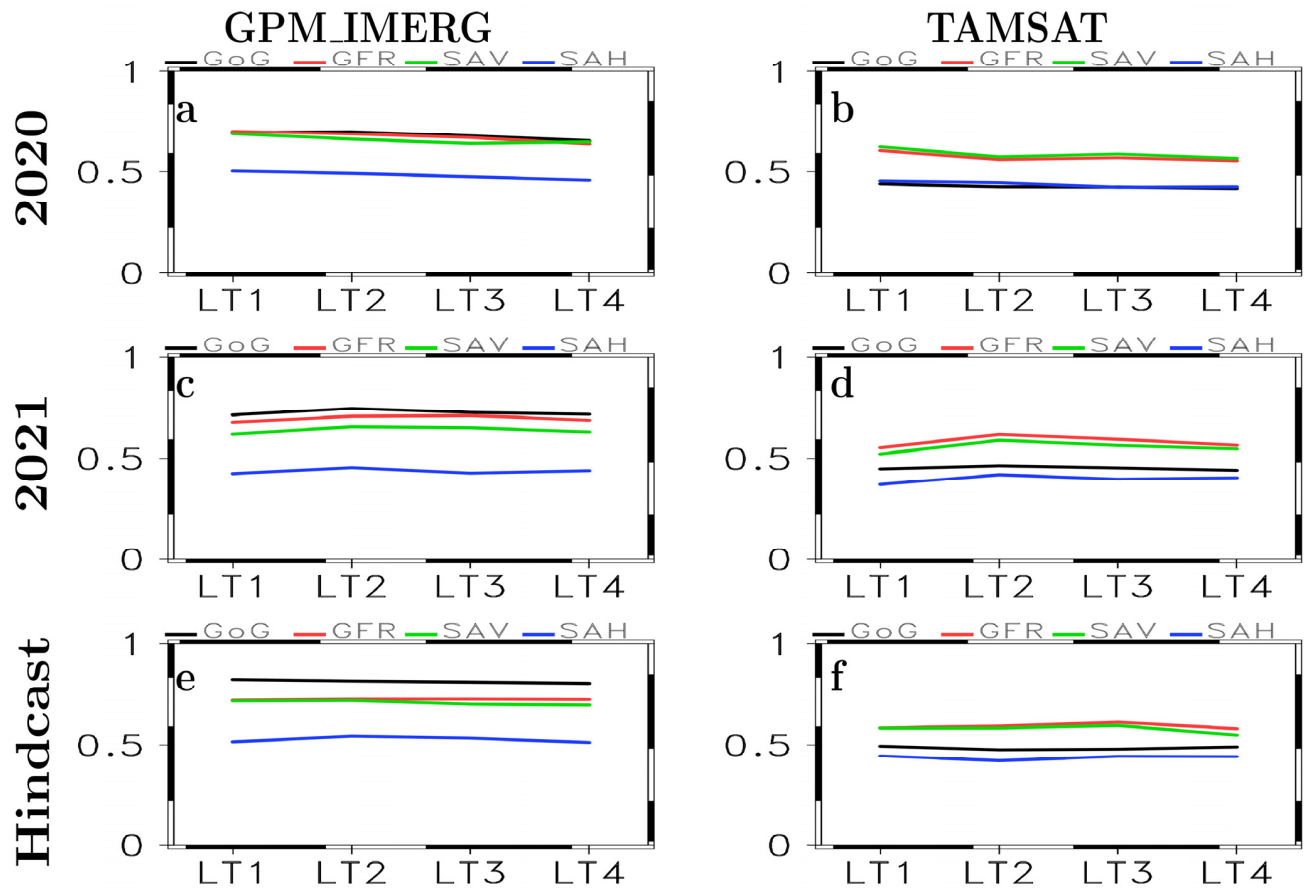


Figure 10. Rank and Discrete Rank Probability Skill Score (RPSS and RPSSD) in real-time and hindcast between S2S and GPM IMERG (a,c,e) and between S2S and TAMSAT (b,d,f) for each LT, from all SDs of 2020, 2021, and hindcast over the GOG (black line), GF (red line), SAV (green line), and the SAH (blue line).

A similar pattern is displayed by the model's RPSS relative to TAMSAT, as depicted in Figure 9b,d,f). The RPSS of the model is also strongest during the NMP except over the GoG and the SAH in the hindcast (Figure 9f). Apart from a slight improvement in 2020 (Figure 9b), the RPSS over the GoG is weak during the FSMP season. However, the model shows strong RPSS during the same season over the GFR in 2020 (Figure 9b) and over both the Savannah (SAV) and the SAH in the hindcast (Figure 9f).

In addition, there are significant disparities in RPSS across LTs from all SDs, whether in real-time or in hindcast (Figure 10). Except over the SAV, where LT3 is the weakest, the model's RPSS across other regions in 2020 decreases with increasing LT (Figure 10a). Despite the progressive fall in RPSS with increasing LT in 2020, Figure 9c depicts LT1 as having the weakest RPSS across all regions in 2021. Additionally, LT3 displays the strongest RPSS over the GoG and the GFR, whereas LT2 displays the strongest RPSS over the SAV and the SAH in 2021. Similar to 2020, the LT4 forecast in the hindcast exhibits the weakest RPSS in most regions, except over the GFR, where the LT1 forecast has the weakest skill. In contrast to real-time, however, the second LT displays the strongest RPSS across other regions, except for the GoG, where LT1 displays the strongest skill (Figure 10e).

Similar to GPM-IMERG, the RPSS relative to TAMSAT in LT1 of 2020 across all regions is the strongest (Figure 10b). Despite the strong skill in LT1 of 2020, the RPSS of LT1 is the weakest across all regions in 2021 except for the GoG (Figure 10d). In the hindcast (Figure 10f), over the GoG and SAH, LT1, and LT2 forecasts exhibit the strongest and weakest RPSS, respectively, while over the GFR and SAV, LT3, and LT4 forecasts display the strongest and weakest skills. This implies that, apart from 2020 relative to GPM-IMERG, the order of RPSS increase or decrease is not sequential regardless of observation.

The variability in RPSS across different regions reflects the complexity and challenges in accurately predicting rainfall with the ECMWF-S2S model. The decreasing RPSS toward the SAH, for instance, may be indicating potential limitations of the model in capturing the climatic and meteorological dynamics of this region. Also, the observation that RPSS is strongest during the NMP, except over the GoG, indicates that the model performs better in certain seasonal conditions. Furthermore, the significant disparities in RPSS across LTs from all SDs, whether in real-time or hindcast, underscore the influence of lead time on forecast skill. The generally stronger RPSS in hindcast scenarios suggests that while the model can effectively simulate past events, real-time forecasting is challenging.

3.6. Relative Operating Characteristics (ROC)

As depicted in Figure 11, the model's ability to distinguish between very low (10 mm), very heavy (>100 mm), and rainfall anomaly events is robust but varies across regions and for each lead time (LT), regardless of the observation dataset used. For instance, the model struggles to differentiate between extremely low rainfall over the Gulf of Guinea (GoG) and extremely high rainfall over the Sahel (SAH). Additionally, during light rainfall events, the model's discrimination strength varies across the region, whereas it increases from the GoG to the Sudanian Savanna (SAV) during heavy rainfall events (Supplementary Material).

The model's separability of rainfall anomalies relative to GPM-IMERG decreases with increasing LT, particularly over the GoG and SAH, except over the Guinea Forest Region (GFR) and SAV (Figure 11a,c,e,g). While LT1 exhibits strong area under the curve (AUC) over the GFR and SAV, LT2, and LT3 show the weakest AUC over these regions, respectively. The model's separability relative to TAMSAT (Figure 11b,d,f,h) shows similar and unique attributes compared to GPM-IMERG. For instance, the skill decreases with increasing LT across most regions except the SAH, where LT4 displays the weakest AUC, but LT2 shows the strongest AUC, contrary to the GPM-IMERG results.

The strong separability of the model, especially over the SAH and particularly relative to GPM-IMERG, validates the utility of the synchronization technique that compensates for the low correlation of the ensemble mean of the model, particularly in the anomaly.

The model's robust but regionally variable ability to differentiate between rainfall extremes underscores the complexity of accurately forecasting rainfall across diverse climatic regions. Also, the model's variation in discrimination strength during light and heavy rainfall events may be connected to the influenced of rainfall intensity on the model's performance, necessitating further refinement to enhance accuracy across all rainfall thresholds.

Additionally, the decreasing separability of rainfall anomalies with increasing LT relative to GPM-IMERG, particularly over the GoG and SAH, highlights the challenges in maintaining forecast skill over longer lead times. Furthermore, the strong AUC in certain LTs and regions, especially over the SAH, demonstrates the effectiveness of the synchronization technique in compensating for low ensemble mean correlation, particularly in anomaly.

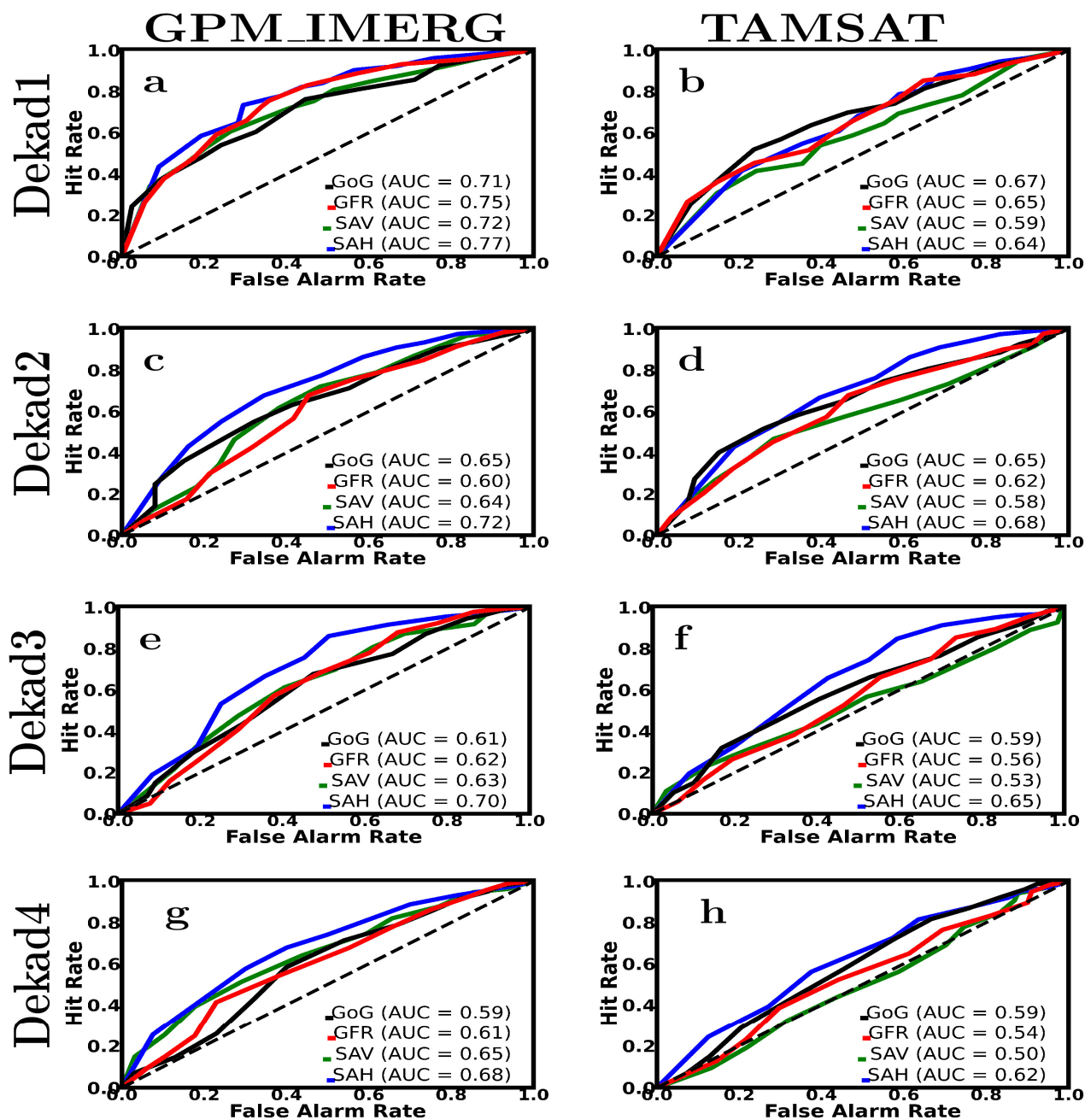


Figure 11. Showing the relative operating characteristic (ROC) diagram from the S2S hindcast probabilities and binary observation (GPM_IMERG (a,c,e,g)) and TAMSAT (b,d,f,h) from the precipitation standardized anomaly over the GoG (black line), GFR (red line), SAV (green line), and SAH (blue line), for all SDs and each LT.

4. Discussion

This work demonstrated that the evaluation of real-time ECMWF S2S forecasts for the 2020–2021 during the RTP over West Africa reliably captures the broad spatial and temporal structure of the West African Monsoon (WAM). In agreement with previous studies, we find significant forecast skill during the main rainy season. For example, ref. [5] also reported strong correlations between ECMWF S2S forecasts and observed rainfall over West Africa in boreal summer, with significant skill extending out to 3–4 weeks lead.

Similarly, our results show high synchronization ($\approx 80\%$) of rainfall anomalies with observations during the northern monsoon peak (July–August) and the second southern monsoon peak (September–October). In agreement with [14], the model tends to perform

better over the Sahel than over the Gulf of Guinea, consistent with known difficulties in predicting equatorial precipitation. At the same time systematic biases persist. For instance, a dry bias relative to TAMSAT in most regions, and a mixed wet/dry bias in different latitudes with GPM-IMERG. These model biases may be attributed in part to known physics limitations. Deep convection parameterizations (as utilized in the ECMWF model) might result in an excessively moist boundary layer and an abnormally dry mid-troposphere. According to [43], in places with boundary-layer moistening and mid-level drying, the convective scheme is most likely responsible. This can lead to an underestimation of heavy rainfall events. Another element is land–atmosphere coupling, with observations and modeling studies indicating substantial soil–moisture–precipitation feedback over West Africa. For instance, ref. [44] noted that wetter soils can boost Sahel rainfall by up to 40% in hypothetical testing, while dry soils suppress convection. If the model underestimates soil moisture, it might accentuate wet/dry biases. Coastal dynamics could also play a role. This is because sea-breeze and Atlantic cold-tongue processes influence Gulf of Guinea rainfall in ways that coarse models may fail to capture. These combined causes most likely explain the geographical bias patterns in the model.

Despite such biases, the deterministic skill of the forecasts is promising. Linear correlations between ensemble-mean forecasts and observations improves generally in the hindcasts (reflecting larger sample size) and comparable across lead times. Consistent with earlier results, we find that correlations increase northward from the Gulf of Guinea into the Sahel. The authors of [5] showed especially strong summer (JJA) correlations along the Gulf coast, which accords with our finding that the model well captures the late-summer onset over coastal West Africa. Notably, we observe the strongest correlations during the September–October period (SSMP), suggesting the model also skillfully represents the retreat phase of the monsoon in those years, whereas [5] emphasized the early (JJA) season skill. This difference here could be as a result of seasonal approach classification approach by [5]. In any case, the relatively improved correlation in both early and late monsoon phases indicates that the model effectively tracks the north–south migration of the ITCZ and associated rains on S2S time scales, in line with prior findings (e.g., [6]).

The probabilistic skill of the forecasts is also encouraging but not without limitations. For instance, the RPSS results confirm substantial probabilistic skill in most thresholds, with peak values in mid-summer (July–August) and some weakening toward the Sahel, again reflecting better predictability during the northern monsoon. These seasonal patterns mirror those in ECMWF hindcast studies. On the other hand, the averaged ROC AUC (~0.68) indicates moderate discrimination (which is well above the no-skill 0.5 level) but still with room for improvement. This is consistent with [5] of ECMWF AUC scores around 0.7 in week 1 for tropical Africa, decreasing with increasing lead time.

The real-time forecasts from our results show a similar trend of strong discrimination in week 1 (when we see 80% synchronization) that decreases by the third and fourth dekads. The implication of these attributes from our results is that actionable S2S information is most robust in the first 2–3 dekads, though useful signals can persist through the fourth dekads for certain cases. This may align with broader analyses of S2S predictability over Africa particularly [7], which emphasize the role of large-scale drivers modulating forecast quality, implying that improving their representation tends to improve skill in rainfall. For instance, it is clear from our results that skill peaks coincide with coherent climate patterns (e.g., a strong wet Sahel in 2020), suggesting potential links to driver phases.

Forecasting of rainfall extremes stands out as a critical challenge from our findings. For instance, we observed that the model struggles to predict extreme low rainfall in the Gulf of Guinea and extreme high rainfall in the Sahel [45]. This is consistent with the general difficulty of mesoscale convective extremes in models [46]. Similar works in tropical

Africa have shown that convection-permitting or higher-resolution ensembles prediction models can better capture localized heavy events (e.g., [21,23]), suggesting a path for future improvement. However, our verification framework, which included multi-threshold and ROC analysis, demonstrates that even when extremes are missed, the model retains reasonable ability to distinguish wet vs. dry spells in most regions.

The use of two observational datasets (TAMSAT, GPM-IMERG) is another notable aspect of our study to validate forecasts. The differences we found between TAMSAT and GPM biases underscore the observational uncertainty in African rainfall products. By using both datasets, this study mitigates reliance on a single observational truth, which is especially important in data-sparse regions. We also applied debiased RPSS approach to account for ensemble size effects, in line with best practices [35].

Our results have clear positive implications in operational application. For instance, the demonstration that ECMWF S2S forecasts are skillful and reliable (particularly up to 2–3 dekads) supports their use in early warning and planning systems across West Africa. Indeed, the integration of S2S products into decision making is already ongoing. For example, S2S products are being used by centers such as NiMet, ACMAD, and national agencies for applications ranging from crop advisories, malaria-transmission to meningitis-risk mapping. This study underpins these efforts by quantifying forecast confidence. It emphasizes that NMHSs and humanitarian planners can trust ECMWF S2S rainfall forecasts to signal above-/below-normal rain with 70–80% skill up to a few dekads ahead, enabling proactive action. Simultaneously, our findings caution that calibration (to correct any over-confidence) and multi-model ensemble thinking may be needed to address residual biases and extreme cases [47].

Four critical future research directions may emerge from this study looking ahead. First, further work could explore forecast calibration and bias-correction tailored to West Africa, building on the observed forecast–climate relationships. For example, the suggestion in our results that dekadal averaging may improve skill at longer leads is intriguing; future studies should test whether using longer aggregated lead periods (e.g., 10-day vs. weekly) consistently enhances predictability, possibly by filtering out high-frequency noise. Second, investigating the mechanistic causes of sub-seasonal predictability in this region (e.g., via MJO/ENSO filtering as in [5,48,49] could identify windows of enhanced skill or particular drought/wet spell precursors. Third, expanding the analysis to additional monsoon years and other S2S systems (e.g., UK Met Office) would strengthen confidence in these results. Finally, bridging to the last-mile user community, co-production studies should assess how to best communicate the probabilistic information to decision makers. Our results suggest that up to 3 dekads lead, forecasts carry real actionability, but that is if end users understand confidence levels (for instance, a 60% chance of above-normal rain should be weighed differently than a 90% chance).

Finally, this work thus not only evaluates a forecast product, it identifies when and where the ECMWF S2S rainfall forecasts are most trustworthy, and suggests how they can be improved (e.g., through longer lead aggregations or higher resolution) for even greater societal benefit.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/atmos16091072/s1>: Figures S1 and S2: title Relative Operating Characteristic (ROC) diagrams derived from S2S hindcast probabilities and binary observations (GPM_IMERG and TAMSAT).

Author Contributions: All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by E.A.O., L.C.H., S.J.W., K.A.L., F.M.D.A., and E.T. The first draft of the manuscript was written by E.A.O. and all authors commented on

previous versions of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by U.K. Research and Innovation through the Global Challenges Research Fund, Grant NE/P021077/1 (GCRF African SWIFT; 2017–2022).

Data Availability Statement: The following datasets used in this study are openly accessible: S2S data through the S2S Prediction Project (<http://s2sprediction.net/> (accessed on the fly from 30 March to 13 October 2020 and 29 March to 12 October 2021). TAMSAT rainfall data via the TAMSAT Data Portal (<https://data.tamsat.org.uk/data/download/rainfall/> (accessed on 10 March 2022)). GPM_IMERG precipitation data from NASA's GSFC DAAC (<https://daac.gsfc.nasa.gov/datasets?page=1&measurement=Rain&project=GPM> (accessed on 12 March 2022)). All datasets can be provided in NetCDF format upon request.

Acknowledgments: We are grateful to Steven Woolnough for his invaluable contribution to this work. The first author acknowledges the support provided by the Nigerian Meteorological Agency and the Abdus Salam International Centre for Theoretical Physics (ICTP).

Conflicts of Interest: The authors have no relevant financial or non-financial interests to disclose.

References

1. Vitart, F.; Robertson, A.W. The sub-seasonal to seasonal prediction project (S2S) and the prediction of extreme events. *Npj Clim. Atmos. Sci.* **2018**, *1*, 3. [CrossRef]
2. White, C.J.; Carlsen, H.; Robertson, A.W.; Klein, R.J.; Lazo, J.K.; Kumar, A.; Vitart, F.; de Perez, E.C.; Ray, A.J.; Murray, V.; et al. Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorol. Appl.* **2017**, *24*, 315–325. [CrossRef]
3. Hirons, L.; Thompson, E.; Dione, C.; Indasi, V.S.; Kilavi, M.; Nkiaka, E.; Talib, J.; Visman, E.; Adefisan, E.A.; de Andrade, F.; et al. Using co-production to improve the appropriate use of sub-seasonal forecasts in Africa. *Clim. Serv.* **2021**, *23*, 100246. [CrossRef]
4. White, C.J.; Domeisen, D.I.V.; Acharya, N.; Adefisan, E.A.; Anderson, M.L.; Aura, S.; Balogun, A.A.; Bertram, D.; Bluhm, S.; Brayshaw, D.J.; et al. Advances in the application and utility of subseasonal-to-seasonal predictions. *Bull. Am. Meteorol. Soc.* **2021**, *103*, E1448–E1472. [CrossRef]
5. de Andrade, F.M.; Young, M.P.; MacLeod, D.; Hirons, L.C.; Woolnough, S.J.; Black, E. Subseasonal rainfall prediction for africa: Forecast evaluation and sources of predictability. *Weather Forecast.* **2021**, *36*, 265–284. [CrossRef]
6. Vitart, F.; Balsamo, G.; Buizza, R.; Ferranti, L.; Keeley, S.; Magnusson, L.; Molteni, F.; Weisheimer, A. Sub-seasonal predictions. *ECMWF Tech. Memo.* **2014**, *45*, 738.
7. Vitart, F.; Robertson, A.W.; Anderson, D.T. Sub-seasonal to Seasonal Prediction Project: Bridging the gap between weather and climate. *WMO Bull.* **2012**, *61*, 23–28.
8. Parker, D.J.; Blyth, A.M.; Woolnough, S.J.; Dougill, A.J.; Bain, C.L.; de Coning, E.; Diop-Kane, M.; Foamouhoue, A.K.; Lampitey, B.; Ndiaye, O.; et al. The African SWIFT project: Growing science capability to bring about a revolution in weather prediction. *Bull. Am. Meteorol. Soc.* **2021**, *103*, E349–E369. [CrossRef]
9. Fink, A.H.; Vincent, D.G.; Ermert, V. Rainfall types in the West African Sudanian zone during the summer monsoon 2002. *Mon. Weather Rev.* **2006**, *134*, 2143–2164. [CrossRef]
10. Bombardi, R.J.; Pegion, K.V.; Kinter, J.L.; Cash, B.A.; Adams, J.M. Sub-seasonal Predictability of the Onset and Demise of the Rainy Season over Monsoonal Regions. *Front. Earth Sci.* **2017**, *5*, 14. [CrossRef]
11. Kumi, N.; Abiodun, B.J.; Adefisan, E.A. Performance Evaluation of a Subseasonal to Seasonal Model in Predicting Rainfall Onset Over West Africa. *Earth Space Sci.* **2020**, *7*, 1–13. [CrossRef]
12. Leutbecher, M. Ensemble size: How suboptimal is less than infinity? *Q. J. R. Meteorol. Soc.* **2019**, *145*, 107–128. [CrossRef]
13. McCrary, R.R.; Randall, D.A.; Stan, C. Simulations of the West African monsoon with a superparameterized climate model. Part II: African easterly waves. *J. Clim.* **2014**, *27*, 8323–8341. [CrossRef]
14. Olaniyan, E.; Adefisan, E.A.; Balogun, A.A.; Lawal, K.A. The influence of global climate drivers on monsoon onset variability in Nigeria using S2S models. *Model. Earth Syst. Environ.* **2019**, *5*, 1405–1428. [CrossRef]
15. Vellinga, M.; Arribas, A.; Graham, R. Seasonal forecasts for regional onset of the West African monsoon. *Clim. Dyn.* **2013**, *40*, 3047–3070. [CrossRef]
16. Kolstad, E.W.; Lee, S.H.; Butler, A.H.; Domeisen, D.I.V.; Wulff, C.O. Drivers of subseasonal forecast errors of the East African short rains. *Geophys. Res. Lett.* **2021**, *48*, e2021GL093292. [CrossRef]
17. Coelho, C.A.S.; Firpo, M.A.F.; de Andrade, F.M. A verification framework for South American sub-seasonal rainfall predictions. *Meteorol. Z.* **2018**, *27*, 503–520. [CrossRef]

18. Vitart, F. Madden—Julian Oscillation prediction and teleconnections in the S2S database. *Q. J. R. Meteorol. Soc.* **2017**, *143*, 2210–2220. [\[CrossRef\]](#)
19. I Maidment, R.; Grimes, D.; Black, E.; Tarnavsky, E.; Young, M.; Greatrex, H.; Allan, R.P.; Stein, T.; Nkonde, E.; Senkunda, S.; et al. A new, long-term daily satellite-based rainfall dataset for operational monitoring in Africa. *Sci. Data* **2017**, *4*, 170063. [\[CrossRef\]](#)
20. Skofronick-Jackson, G.; Petersen, W.A.; Berg, W.; Kidd, C.; Stocker, E.F.; Kirschbaum, D.B.; Kakar, R.; Braun, S.A.; Huffman, G.J.; Iguchi, T.; et al. The Global Precipitation Measurement (GPM) mission for science and society. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 1679–1695. [\[CrossRef\]](#)
21. Cafaro, C.; Woodhams, B.J.; Stein, T.H.M.; Birch, C.E.; Webster, S.; Bain, C.L.; Hartley, A.; Clarke, S.; Ferrett, S.; Hill, P. Do convection-permitting ensembles lead to more skillful short-range probabilistic rainfall forecasts over tropical east africa? *Weather Forecast.* **2021**, *36*, 697–716. [\[CrossRef\]](#)
22. Stein, T.H.M.; Keat, W.; Maidment, R.I.; Landman, S.; Becker, E.; Boyd, D.F.A.; Bodas-Salcedo, A.; Pankiewicz, G.; Webster, S. An evaluation of clouds and rainfall in convection-permitting forecasts for South Africa. *Weather Forecast.* **2019**, *34*, 233–254. [\[CrossRef\]](#)
23. Woodhams, B.J.; Birch, C.E.; Marsham, J.H.; Bain, C.L.; Roberts, N.M.; Boyd, D.F.A. What is the added value of a convection-permitting model for forecasting extreme rainfall over tropical East Africa? *Mon. Weather Rev.* **2018**, *146*, 2757–2780. [\[CrossRef\]](#)
24. Maranan, M.; Fink, A.H.; Knippertz, P.; Amekudzi, L.K.; Atiah, W.A.; Stengel, M. A process-based validation of gpm imerg and its sources using a mesoscale rain gauge network in the west african forest zone. *J. Hydrometeorol.* **2020**, *21*, 729–749. [\[CrossRef\]](#)
25. Keikhosravi-Kiany, M.S.; Balling, R.C. Evaluation of GPM IMERG Early, Late, and Final Run in representing extreme rainfall indices in Southwestern Iran. *Remote Sens.* **2024**, *16*, 2779. [\[CrossRef\]](#)
26. Watters, D.C.; Gatlin, P.N.; Bolvin, D.T.; Huffman, G.J.; Joyce, R.; Kirstetter, P.; Nelkin, E.J.; Ringerud, S.; Tan, J.; Wang, J.; et al. Oceanic validation of IMERG-GMI Version 6 precipitation using the GPM Validation Network. *J. Hydrometeorol.* **2024**, *25*, 125–142. [\[CrossRef\]](#)
27. Grimes, D.I.F.; Pardo-Igúzquiza, E.; Bonifacio, R. Optimal areal rainfall estimation using raingauges and satellite data. *J. Hydrol.* **1999**, *222*, 93–108. [\[CrossRef\]](#)
28. Tarnavsky, E.; Grimes, D.; Maidment, R.; Black, E.; Allan, R.P.; Stringer, M.; Chadwick, R.; Kayitakire, F. Extension of the TAMSAT satellite-based rainfall monitoring over Africa and from 1983 to present. *J. Appl. Meteorol. Climatol.* **2014**, *53*, 2805–2822. [\[CrossRef\]](#)
29. Boulton, V.L.; Asfaw, D.T.; Young, M.; Maidment, R.; Mwangi, E.; Ambani, M.; Waruru, S.; Otieno, G.; Todd, M.C.; Black, E. Evaluation and validation of TAMSAT-ALERT soil moisture and WRSI for use in drought anticipatory action. *Meteorol. Appl.* **2020**, *27*, e1959. [\[CrossRef\]](#)
30. Macharia, D.; Maidment, R.I.; Grimes, D.I.F.; Black, E. Validation and intercomparison of satellite-based rainfall products in Africa. *J. Hydrometeorol.* **2022**, *23*, 1087–1104.
31. Lawal, K.A.; Akintomide, O.M.; Olaniyan, E.; Bowery, A.; Sparrow, S.N.; Wehner, M.F.; Stone, D.A. Performance Evaluation of Weather@home2 Simulations over West African Region. *Atmosphere* **2025**, *16*, 392. [\[CrossRef\]](#)
32. Misra, J. Phase synchronization. *Inf. Process. Lett.* **1991**, *38*, 101–105. [\[CrossRef\]](#)
33. Gouda, K.C.; Rajeevan, M.; Rajeevan, K.; Raju, P.V.S.; Aditi, A. Performance of a variable resolution global model in simulating Indian summer monsoon rainfall at multiple scales. *Pure Appl. Geophys.* **2020**, *177*, 5903–5919.
34. Joshi, J.; Tripathi, S.; Dey, S.; Pandey, A.C. Seasonal prediction of Indian summer monsoon rainfall over Uttarakhand using an atmospheric general circulation model. *Theor. Appl. Climatol.* **2020**, *139*, 1037–1047. [\[CrossRef\]](#)
35. Weigel, A.P.; Liniger, M.A.; Appenzeller, C. The discrete Brier and ranked probability skill scores. *Mon. Weather Rev.* **2007**, *135*, 118–124. [\[CrossRef\]](#)
36. Wu, X.; Liu, Y.; Liu, S.; Jin, Y.; Xu, H. Assessment of Satellite Products in Estimating Tropical Cyclone Remote Precipitation over the Yangtze River Delta Region. *Atmosphere* **2024**, *15*, 667. [\[CrossRef\]](#)
37. Wilks, D.S. *Statistical Methods in the Atmospheric Sciences*, 4th ed.; Academic Press: Cambridge, MA, USA, 2019; ISBN 978-0-12-815823-4.
38. Ageet, S.; Fink, A.H.; Maranan, M.; Diem, J.E.; Hartter, J.; Ssali, A.L.; Ayabagabo, P. Validation of Satellite Rainfall Estimates over Equatorial East Africa. *J. Hydrometeorol.* **2022**, *23*, 129–151. [\[CrossRef\]](#)
39. Hagos, S.M.; Cook, K.H. Dynamics of the West African monsoon jump. *J. Clim.* **2007**, *20*, 5264–5284. [\[CrossRef\]](#)
40. Yu, L.; Leng, G.; Python, A. A comprehensive validation for GPM IMERG precipitation products to detect extremes and drought over mainland China. *Weather Clim. Extrem.* **2022**, *36*, 100458. [\[CrossRef\]](#)
41. Wang, Y.; Miao, C.; Zhao, X.; Zhang, Q.; Su, J. Evaluation of the GPM IMERG product at the hourly timescale over China. *Atmos. Res.* **2023**, *285*, 106656. [\[CrossRef\]](#)
42. Mekonnen, K.; Velpuri, N.M.; Leh, M.; Akpoti, K.; Owusu, A.; Tinonetsana, P.; Hamouda, T.; Ghansah, B.; Paranamana, T.P.; Munzimi, Y. Accuracy of satellite and reanalysis rainfall estimates over Africa: A multi-scale assessment of eight products for continental applications. *J. Hydrol. Reg. Stud.* **2023**, *49*, 101514. [\[CrossRef\]](#)

43. Beljaars, A.C.M. The parametrization of moist convection in the ECMWF model: Recent developments and future challenges. *ECMWF Newsl.* **2006**, *24*, 24–29.
44. Koné, B.; Diedhiou, A.; Diawara, A.; Anquetin, S.; Touré, N.E.; Bamba, A.; Koba, A.T. Influence of initial soil moisture in a regional climate model study over West Africa—Part 2: Impact on the climate extremes. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 731–754. [[CrossRef](#)]
45. Sanogo, S.; Peyrillé, P.; Roehrig, R.; Guichard, F.; Ouedraogo, O. Extreme Precipitating Events in Satellite and Rain Gauge Products over the Sahel. *J. Clim.* **2022**, *35*, 1915–1938. [[CrossRef](#)]
46. Vigaud, N.; Robertson, A.W.; Tippett, M.K.; Acharya, N. Sub-seasonal Predictability of Boreal Summer Monsoon Rainfall from Ensemble Forecasts. *Front. Environ. Sci.* **2017**, *5*, 67. [[CrossRef](#)]
47. Zhang, L.; Yang, T.; Gao, S.; Hong, Y.; Zhang, Q.; Wen, X.; Cheng, C. Improving Subseasonal-to-Seasonal forecasts in predicting the occurrence of extreme precipitation events over the contiguous U.S. using machine learning models. *Atmos. Res.* **2022**, *281*, 106502. [[CrossRef](#)]
48. Kim, H.; Son, S.-W.; Kim, H.; Seo, K.H.; Kang, M.J. MJO influence on subseasonal-to-seasonal prediction in the Northern Hemisphere extratropics. *J. Clim.* **2023**, *36*, 7943–7956. [[CrossRef](#)]
49. Vashisht, A.; Zaitchik, B. Modulation of East African boreal fall rainfall: Combined effects of the Madden–Julian Oscillation (MJO) and El Niño–Southern Oscillation (ENSO). *J. Clim.* **2022**, *35*, 2019–2034. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.