https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


# Physical Climate Drivers of East Africa's March-April-May (MAM) seasonal rainfall Identified through Machine Learning Analysis

Sinclair Chinyoka[1,2], Masilin Gudoshava[2], Hussen Seid Endris[2], Shingirai Shepard Nangombe[3], Jordi Vila-Guerau De Arellano[1], and Gert-Jan Steeneveld[1]

[1]Meteorology and Air Quality Section, Wageningen University, P.O.Box 47 6700 AA, Wageningen, The Netherlands
[2]Climate Diagonistic and Prediction Unit, IGAD Climate Prediction and Application Centre (ICPAC), P.O.BOX 10304-00100, Nairobi, Kenya
[3]National Center for Climate Research, Danish Meteorological Institute, Copenhagen, Denmark

**Correspondence:** Sinclair Chinyoka (sinclair.chinyoka@wur.nl)

**Abstract.** East African March–May rainfall (MAM) remains difficult to predict despite its importance for agriculture, water resources, and disaster preparedness. This study identifies pre-season physical drivers of MAM rainfall and tests their value for probabilistic seasonal prediction. Predictor basins were derived from December and January sea surface temperature (SST), 2 m air temperature (T2), and sea-level pressure (SLP) anomalies relative to 1991–2020, using correlations with the leading mode of East African MAM rainfall and subsequent SHAP-based feature selection. The selected basin-derived indices were applied in Random Forest (RF) and Extreme Gradient Boosting (XGB) models. The dominant predictors appear to be the southern Indian Ocean T2 tendency, Australian and Eurasian T2 gradients, South Pacific and Antarctic T2 signals, Atlantic Niño tendency, and the Euro–African SLP gradient. T2-related predictors dominate both the January and December initialisations, showing that near-surface thermal gradients provide useful information in addition to SST memory. Walker-circulation diagnostics show that these drivers influence rainfall through pressure-gradient changes, tropical overturning, and upper-level wave-train development. For January initialisation, RF and XGB achieve spatially averaged Brier Skill Scores of 0.48 and 0.41, respectively, while the corresponding Area Under the Receiver Operating Characteristic Curve values amounting to 0.72 and 0.65. These results demonstrate that physically constrained machine learning provides promising probabilistic skill for East African MAM rainfall prediction.

## 1 Introduction

The hydroclimate of East Africa is governed by the seasonal migration of the Intertropical Convergence Zone (ITCZ) and its interaction with Indian Ocean circulation. The region spans latitudes 11.5° S to 15.5° N and longitudes 24° E to 50° E (Figure 1), falling within the tropical zone and exhibiting pronounced spatial heterogeneity in rainfall regimes. Some countries, such as Kenya, experience a bimodal rainfall distribution, with the "short rains" occurring from October to December (OND) and the "long rains" from March to May (MAM). Other parts of the region, particularly northern East Africa, exhibit a monomodal rainfall season during June–September (JJAS) (Diro et al., 2011; Camberlin and Philippon, 2002; Nicholson, 2017; Wainwright

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


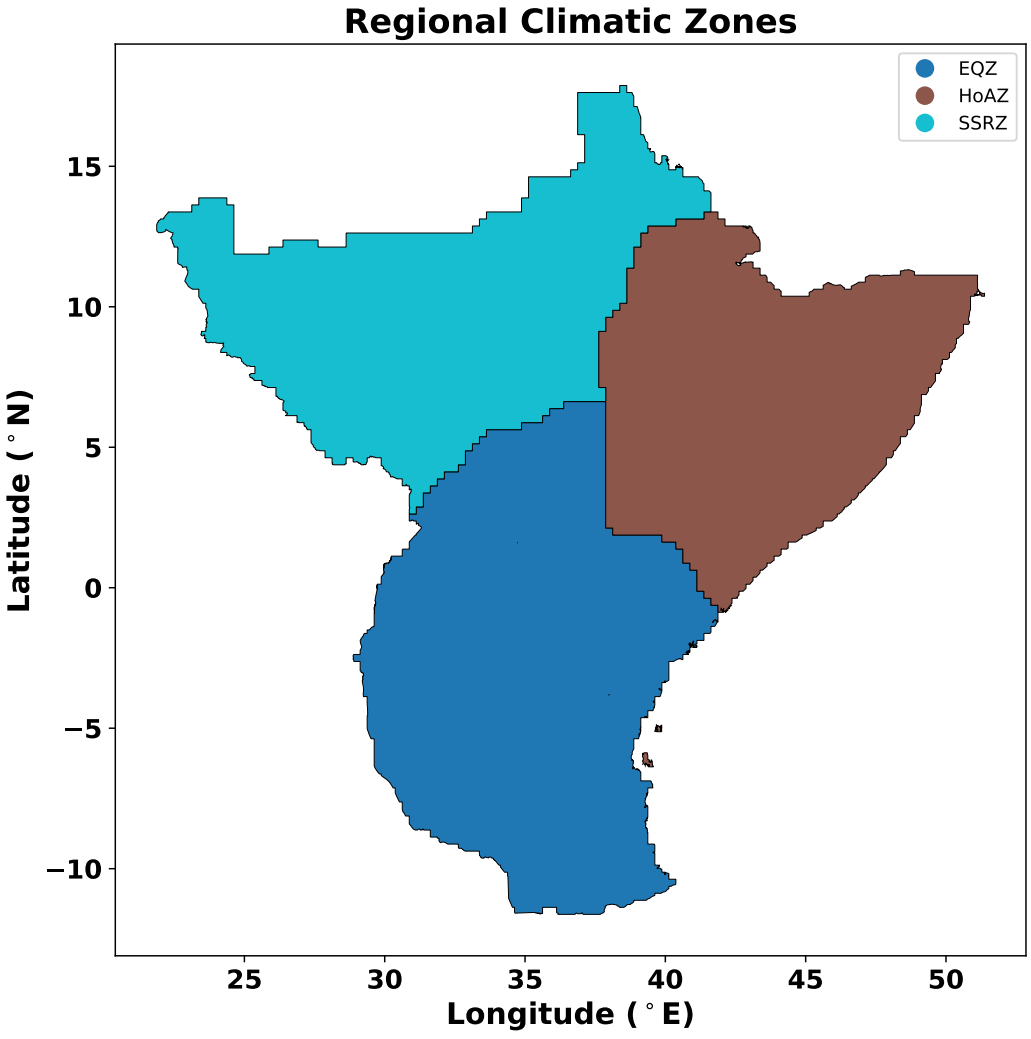


**Figure 1.** Regional climatic zones for East Africa derived from MAM rainfall. The results of the empirical orthogonal functions (EOFs) of gridded MAM rainfall and clustered grid cells in the space of the leading EOFs using k-means; clusters were then made contiguous by small-patch removal and boundary smoothing. The three objectively identified zones are: EQZ (Equatorial Zone; relatively wet MAM), HoAZ (Horn of Africa Zone, mostly covering Ethiopia and Somalia; characterised by orographic enhancement of MAM rainfall), and SSRZ (Sub-Saharan Region; low MAM totals with high interannual variability). Colours indicate zone membership.

et al., 2019). This climatic contrast, spanning arid, semi-arid, and humid tropical regimes, creates substantial challenges for generating accurate long- and short-range forecasts.

Of all rainfall seasons in East Africa, the MAM season is dynamically distinct and particularly complex. MAM is a crucial season for farming across much of the region, yet it remains one of the most difficult to predict. Variations in large-scale

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


ocean–atmosphere conditions, especially sea surface temperature (SST) gradients and associated circulation anomalies, modulate moisture transport, vertical motion, and convection over East Africa.

Large-scale climate drivers such as the Indian Ocean Dipole (IOD), Quasi-Biennial Oscillation (QBO), ENSO, and regional SST anomalies have been shown to influence East African MAM rainfall depending on their phase and interaction (Saji et al., 1999; Nicholson, 2017; Palmer et al., 2023; Tierney et al., 2015; Baldwin et al., 2001; Diro et al., 2011; Omondi et al., 2013; Wainwright et al., 2019). Recent studies have also highlighted additional drivers, including cooling within the Eurasian basin linked to MAM rainfall variability (Hagos et al., 2024), zonal winds at 700 hPa over the Congo basin (Ward et al., 2023), and warming in parts of the Indian Ocean associated with suppressed MAM rainfall (Lyon and DeWitt, 2012). Additionally, several studies have further demonstrated that tropical seasonal rainfall is closely linked to large-scale climate variability, with SST (Uvo et al., 1998; Kumar et al., 2013; Landman et al., 2012) and near-surface air temperature (Ran et al., 2025) emerging as important sources of predictability. By modulating land–atmosphere coupling, moisture transport, and large-scale circulation, these variables can provide substantial skill for seasonal rainfall prediction (Uvo et al., 1998; Kumar et al., 2013; Ran et al., 2025; Landman et al., 2012). These drivers do not operate in isolation, but through nonlinear and seasonally evolving interactions that influence vertical motion, moisture convergence, and rainfall distribution over the region.

Despite this physically grounded understanding, the MAM rainfall season remains notoriously difficult to predict because of weak and shifting large-scale climate drivers, strong spatiotemporal variability, and poorly captured seasonal transitions in global models (Dunning et al., 2016; Nicholson, 2017; Wainwright et al., 2019). Global climate models (GCMs) often misrepresent the large-scale circulation and associated SST gradients that influence East African rainfall, leading to systematic biases in rainfall simulation (James et al., 2020). Their coarse horizontal (typically 100–250 km) and vertical grid spacings limit their ability to represent localised upwelling, downwelling, and vertical velocity relatively , which are important components of the tropical circulation system (Turner and Annamalai, 2012). Because the impacts of climate drivers during MAM are often weak, transient, and nonlinear, particularly during ENSO transition and neutral IOD phases, both GCMs and multi-model ensemble (MME) forecasts continue to struggle in capturing East African rainfall variability.

Earlier studies indicate that the MAM season is becoming shorter and drier, likely in association with warming sea surface temperatures in the southern Indian Ocean and the North Arabian Sea (Wainwright et al., 2019; Palmer et al., 2023). This tendency, combined with the absence of well-defined and stationary predictors, further underscores the need for improved forecasting approaches. As Tierney et al. (2015) emphasised, there is an urgent need to develop more accurate systems for predicting MAM seasonal rainfall anomalies and for understanding the physical mechanisms that drive rainfall variability during this season.

Given the importance of rainfall predictions in East Africa, improving seasonal forecast skill is both a scientific and societal priority. In a region where rain-fed agriculture dominates, reliable seasonal rainfall forecasts are essential for food security, water-resource management, disaster preparedness, and economic stability. However, improvements in climate prediction remain closely tied to access to advanced forecasting tools, observational data, and technical capacity. Although dynamical climate models are widely used globally, their implementation and maintenance remain challenging for many African countries because of resource limitations (Milton et al., 2017).

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


Recent studies have shown that both mesoscale climate models and machine-learning (ML) techniques can enhance subseasonal to seasonal forecasting skill (Chinyoka and Steeneveld, 2023; Deman et al., 2022; Baig et al., 2024; Antonio et al., 2025; Gibson et al., 2021; Weyn et al., 2021; Pinheiro and Ouarda, 2025). ML techniques are particularly attractive because they have shown promising predictive performance while offering operational advantages in terms of adaptability and relatively low computational cost (Bauer, 2024). Building on these advances, this study investigates the potential of ML as a cost-effective framework for improving seasonal rainfall prediction over East Africa.We also build on this physical understanding by defining SST, 2-m air temperature (T2), and sea level pressure as key climate-driver variables for the ML models.

More specifically, we investigate the key climate-driver basins that influence East African MAM rainfall predictability through thermal heating and cooling, sea level pressure anomalies, and their associated atmospheric responses. We focus on how these drivers are linked to variations in the Walker circulation, the development of Rossby wave responses, and the resulting rainfall anomalies over East Africa. In doing so, the aim is not only to improve prediction skill, but also to better understand the mechanisms through which basin-scale thermodynamic and dynamic forcing affects the East African long rains.

To capture these processes, we incorporate climate drivers hypothesized to be relevant to the East African MAM season, including the Western V Gradient (WVG; Funk et al., 2018), the Quasi-Biennial Oscillation (QBO; Baldwin et al., 2001), and variability in the ITCZ (Finney et al., 2020), together with basin-scale SST, T2, and sea level pressure fields. The ML framework is therefore used not only as a predictive tool, but also as a diagnostic approach for identifying the dominant large-scale and regional drivers of MAM rainfall and for assessing both their linear and nonlinear contributions to seasonal predictability.

In this study, we therefore investigate the physical controls and predictability of East African MAM rainfall by explicitly linking mechanistic understanding of large-scale circulation to data-driven modelling. We test the hypothesis that MAM predictability can be enhanced by: (i) isolating physically consistent pre-season drivers that modulate the Walker circulation and Rossby wave pathways, and (ii) capturing their combined linear and nonlinear effects using machine-learning techniques. In this way, the ML approach is intended to help overcome some of the limitations of traditional dynamical models by focusing on physically meaningful climate drivers that are not always adequately represented in GCMs.

To address these challenges, this study is guided by the following scientific questions:

1. Which pre-MAM large-scale and regional climate drivers most strongly influence East African MAM seasonal rainfall totals?

   This question focuses on identifying the most relevant climate drivers of MAM rainfall through both physical reasoning and data-driven feature selection.

2. How do these climate drivers contribute to the skill of machine-learning forecasts of East African March–May rainfall?

   We address this question by developing and evaluating ML models tailored to use key climate drivers over East Africa as predictors. By incorporating physically relevant drivers, this research seeks to understand each driver's contribution to improving forecast accuracy and reducing uncertainty, while also providing a framework that can be operationalised by regional and national meteorological services such as the Intergovernmental Authority on Development (IGAD) Climate Prediction and Applications Centre (ICPAC).

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


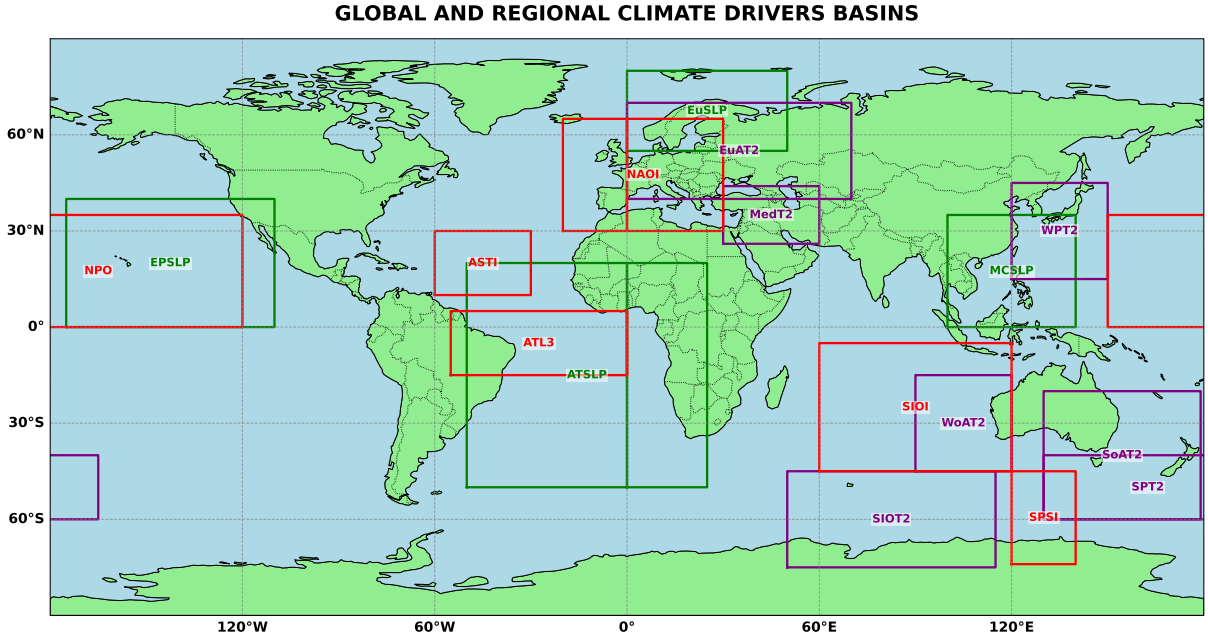


**Figure 2.** Global and regional predictor basins used to construct climate-driver indices for seasonal rainfall modelling over East Africa. Green boxes indicate SLP-based regions, purple boxes indicate 2-m air temperature regions, and red boxes indicate SST regions. The outlined domains were identified from the correlation-based screening procedure and retained for basin-mean index construction.

## 2 Methods

### 2.1 Predictor-region identification

Potential predictor regions were identified from global fields of SST, T2, SLP, and soil moisture using the leading mode of East Africa (EA) MAM rainfall variability. Empirical orthogonal function (EOF) analysis was first applied to EA MAM rainfall anomalies, and the first principal component (PC1) was retained as the target rainfall index because it explained the largest fraction of interannual variance (Scalabrini and Remoaldo, 2024; Hotelling, 1933a, b; Wilks, 2006, 2020). Grid-point correlations were then computed between EA MAM rainfall PC1 and antecedent monthly climate anomalies for December and January. Only statistically significant regions ($p < 0.1$) were retained to define candidate predictor basins.

Figure 2 summarises the final predictor basins retained from the screening analysis. The corresponding latitude–longitude boundaries and basin-mean correlation coefficients with EA MAM rainfall PC1 are given in Table 1. These domains were subsequently used for spatial averaging and index construction.

### 2.2 Climate data

Multiple observational and reanalysis datasets were used to derive predictor variables and rainfall targets.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


**Table 1.** Overview of predictor basins used for index construction and their key climate variable as well as location. Negative longitudes denote degrees west, whereas positive longitudes denote degrees east. Basins crossing the dateline are indicated explicitly in the longitude range.

| Basin | Climate variable | Latitude range | Longitude range |
|---|---|---|---|
| EPSLP (Eastern Pacific) | SLP | 0°–40°N | 175°W–110°W |
| EuSLP (Europe) | SLP | 55°–80°N | 0°–50°E |
| MCSLP (Maritime Continent) | SLP | 0°–35°N | 100°–140°E |
| ATSLP (Atlantic sector) | SLP | 50°S–20°N | 50°W–25°E |
| EuAT2 (Eurasia) | T2 | 40°–70°N | 0°–70°E |
| MedT2 (Mediterranean) | T2 | 26°–44°N | 30°–60°E |
| SIOT2 (Southern Indian Ocean) | T2 | 75°S–45°S | 50°–115°E |
| WPT2 (Western Pacific) | T2 | 15°–45°N | 120°–150°E |
| SPT2 (South Pacific) | T2 | 60°S–40°S | 130°E–165°W |
| SoAT2 (South of Australia) | T2 | 60°S–20°S | 130°E–179°E |
| WoAT2 (West of Australia) | T2 | 45°S–15°S | 90°–120°E |
| SPSI (South Pacific) | SST | 74°S–45°S | 120°–140°E |
| NPO (North Pacific Ocean) | SST | 0°–35°N | 150°E–120°W |
| ASTI (Atlantic subtropical sector) | SST | 10°–30°N | 60°W–30°W |
| SIOI (Southern Indian Ocean) | SST | 45°S–5°S | 60°–120°E |
| ATL3 (Tropical Atlantic) | SST | 15°S–5°N | 55°W–0° |
| NAOI (North Atlantic Ocean) | SST | 30°–65°N | 20°W–30°E |

*Note.* Additional gradient-based predictors were derived from the primary basins as follows: $\mathrm{EAT2GD} = \mathrm{EuAT2} - \mathrm{MedT2}$, $\mathrm{AuT2GD} = \mathrm{WoAT2} - \mathrm{SoAT2}$, and $\mathrm{SPID} = \mathrm{SIOI} - \mathrm{SPSI}$.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


#### a Sea surface temperature

Monthly SST data were obtained from COBE-SST2 (Hirahara et al., 2014), which provides global fields at $1^\circ$ spatial resolution from 1891 to the present. Basin-mean SST anomalies were computed for the identified predictor domains and used to derive oceanic indices including the IOD, Oceanic Niño Index (ONI), Niño 3.4, WVG, southern Indian Ocean SST anomalies (SIO), north Arabian Sea SST anomalies (NAS), Atlantic sector variability (ATL3 and ASTI), North Atlantic Ocean variability (NAOI), and other identified basin-scale SST predictors.

#### b Zonal wind at 30 hPa: QBO

Monthly mean zonal wind at 30 hPa was obtained from ERA5 pressure-level reanalysis at $0.25^\circ$ resolution (Copernicus Climate Change Service (C3S), 2023a). These data were used to derive the Quasi-Biennial Oscillation (QBO) index, which was included as a large-scale dynamical predictor.

#### c Soil moisture

Monthly soil moisture was obtained from ERA5 single-level data (Copernicus Climate Change Service (C3S), 2023b). Soil-moisture anomalies were computed over selected continental basins with potential influence on EA rainfall variability, including the Democratic Republic of Congo Basin, Southern Africa, East Africa, Ethiopia, and the southern Sahara. Basin-mean anomalies from these regions were used as predictor variables.

#### d Monthly rainfall totals

Seasonal rainfall targets were derived from Global Precipitation Climatology Centre (GPCC) products. We combined GPCC First Guess (Schneider et al., 2018) and GPCC Precipitation $0.25^\circ$ V2020 Full Reanalysis (Schneider et al., 2020) to obtain a continuous rainfall record from 1891 to the present. Because ERA5 soil-moisture data are available since 1940, the common period used for model development was restricted to 1940 onward.

### 2.3 Feature engineering and selection

In addition to the primary predictor indices, tendency (TND) features were computed as month-to-month differences to represent changes in the strength and direction of the climate drivers. These features were designed to capture whether a given driver was strengthening or weakening prior to the MAM season. Predictor construction was based on antecedent December and January conditions for forecasting MAM rainfall totals.

Because rainfall controls vary across East Africa, separate models were developed for each rainfall zone and the outputs were later merged into a regional forecast. Predictor selection for each zone was based on SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017), which provides a robust measure of feature importance in nonlinear models. SHAP-based feature selection was implemented iteratively until the ranked contributions stabilised, retaining 75% of predictors at each step subject to a minimum of four predictors.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


### 2.4 Machine-learning models

Two ensemble-learning algorithms were used: Extreme Gradient Boosting (XGBoost) and Random Forest (RF). Both methods are suitable for nonlinear hydroclimatic prediction problems and have shown good skill in rainfall forecasting (Zhang et al., 2021; Meenachi et al., 2023; Baig et al., 2024).

#### a XGBoost

This algorithm is considered an optimization method in function space rather than parameter space (Friedman, 2001). Its core concept revolves around stagewise additive expansion and its connection to steepest-descent minimization. We developed a model using a regression booster aimed at reproducing deterministic seasonal rainfall amounts, which were later categorized into three classes: below normal, normal, and above normal seasonal rainfall, with associated probabilities to capture the level of uncertainty.

The model was implemented using the XGBoost (Chen and Guestrin, 2016b) package in Python, which provides flexibility and ease of use. A key advantage of the gradient boosting algorithm as implemented by XGBoost is its ability to run on GPU processors, significantly reducing computation time, particularly given the size and resolution of our data. XGBoost produces competitive results compared to other machine learning algorithms.

The choice of the model structure and hyperparameters depends on the size and complexity of the dataset, as well as the specific requirements of the model being developed (Chen and Guestrin, 2016a). Thus, hyperparameter tuning with actual training data is essential to achieve optimal performance, within predetermined acceptable thresholds. After conducting comprehensive hyperparameter tuning, we selected the final model structure and key hyperparameter settings as on Table 2. Additionally, a custom evaluation metric based on the coefficient of determination ($R^2$) was used, together with a custom objective function that penalised underforecasting and overforecasting differently.

#### b Random Forest

Another ensemble learning method that combines predictions from multiple tree predictors is Random Forests (Breiman, 2001). The model's accuracy relies on the performance of each individual tree predictor. Its strength lies in its robustness against overfitting, especially as the number of trees grows based on the dataset size. The choice of model configurations and features was made using random search and cross-validation for each country. The best-performing model was retained for predictions (Table 3).

### 2.5 Training and probabilistic prediction

Both models were trained using the same predictor set and targeted EA MAM seasonal rainfall over 1940–2025. Since our ultimate goal is produce probability tercile rainfall forecasts, tercile thresholds for below-normal, normal, and above-normal conditions were computed from the 1991–2020 climatological reference period. Seasons were classified according to the dominant tercile category over the domain, and stratified sampling was used to reduce class imbalance.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


**Table 2.** XGBoost model hyperparameter configuration used for seasonal rainfall prediction.

| Parameter | Value | Description |
|---|---|---|
| max_depth | 6 | Maximum depth of each regression tree. |
| learning_rate | 0.05 | Step size shrinkage applied after each boosting iteration. |
| subsample | 0.7 | Fraction of training samples used to fit each tree. |
| colsample_bytree | 0.6 | Fraction of predictor variables sampled for each tree. |
| reg_alpha | 2.0 | L1 regularisation term used to reduce overfitting. |
| reg_lambda | 8.0 | L2 regularisation term used to improve model stability. |
| min_child_weight | 10 | Minimum instance weight required in a child node. |
| gamma | 1.5 | Minimum loss reduction required for a further split. |
| eval_metric | rmse | Metric used to evaluate model error during training. |
| random_state | 42 | Random seed used for reproducibility. |
| tree_method | hist | Histogram-based algorithm used for efficient tree construction. |
| device | cuda | GPU-based model training. |
| objective | reg:squarederror | Regression objective that minimises squared prediction error. |

**Table 3.** Optimised Random Forest hyperparameter settings for each climatic zone.

| Zone | n_estimators | max_depth | min_samples_split | min_samples_leaf | max_features | ccp_alpha | criterion |
|---|---|---|---|---|---|---|---|
| HoAZ | 120 | 30 | 2 | 2 | 0.7 | 0.01 | squared_error |
| SSRZ | 120 | 10 | 2 | 2 | 0.5 | 0.01 | squared_error |
| EQZ | 120 | 30 | 2 | 2 | 0.7 | 0.01 | squared_error |

The full dataset was partitioned into training, validation, hindcast, and reforecast components. Hindcasts were used to estimate forecast uncertainty and to derive probabilistic tercile forecasts following (Min et al., 2009). For each grid point, the hindcast mean ($\mu_h$), hindcast standard deviation ($\sigma_h$), forecast mean ($\mu_f$), and reforecast standard deviation ($\sigma_{rf}$) were used to compute lower and upper tercile bounds and the corresponding category probabilities.

The lower and upper tercile bounds were defined as:

$$\text{LB} = \mu_h - 0.43\sigma_h \tag{1}$$

$$\text{UB} = \mu_h + 0.43\sigma_h \tag{2}$$

The threshold and probability calculations were applied at each grid point while masking areas that are off-season during MAM.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


$$P(\text{Below Normal}) = P(\mu_f < \text{LB}) = \int_{-\infty}^{\text{LB}} \frac{1}{\sqrt{2\pi\sigma_{rf}^2}} \exp\left(-\frac{(\mu_f - \mu_h)^2}{2\sigma_{rf}^2}\right) d\mu_f \tag{3}$$

$$P(\text{Above Normal}) = P(\mu_f > \text{UB}) = \int_{\text{UB}}^{\infty} \frac{1}{\sqrt{2\pi\sigma_{rf}^2}} \exp\left(-\frac{(\mu_f - \mu_h)^2}{2\sigma_{rf}^2}\right) d\mu_f \tag{4}$$

$$P(\text{Normal}) = P(\text{LB} < \mu_f < \text{UB}) = \int_{\text{LB}}^{\text{UB}} \frac{1}{\sqrt{2\pi\sigma_{rf}^2}} \exp\left(-\frac{(\mu_f - \mu_h)^2}{2\sigma_{rf}^2}\right) d\mu_f \tag{5}$$

## 3 Results

In this section, we present results that address the questions raised in the introduction by examining whether distinct and physically consistent climate drivers can serve as reliable precursors of the forthcoming MAM rainfall season over East Africa (EA). We first analyse the dominant climatic conditions in January, a few months before MAM, to establish the large-scale circulation and oceanic anomalies that may foreshadow seasonal rainfall outcomes. This is achieved through EOF/PCA and SHAP analyses to systematically identify the most influential climate drivers, which are then incorporated into machine-learning (ML) frameworks for predictive evaluation. From the 22 available validation seasons, four representative cases are selected to demonstrate the ability of the ML models to reproduce contrasting wet and dry seasons. The discussion then assesses the relative strengths and limitations of the ML approach, with particular emphasis on the predictive relevance of January drivers. Finally, we synthesise these results by evaluating model skill and reliability, thereby assessing their suitability as operational tools for seasonal climate prediction across EA's highly complex MAM rainfall regime.

### 3.1 Identification of dominant climate drivers of EA MAM rainfall

As outlined in Section 2.1, EOF analysis of EA MAM rainfall was first performed, and the significant predictor basins were subsequently identified from Figure 2. First we examine the relationship between global SST anomalies and EA MAM rainfall identified on Figure 3. The eastern Pacific warming region (NPO) shows a positive relationship with EA rainfall, indicating that warming in this region is associated with enhanced EA rainfall in both December and January . Additional SST-based basins that appear significant include the southern Indian Ocean near Australia (SIOI), particularly in January, which exhibits a negative relationship with EA MAM rainfall. This suggests that warming in this region is associated with drier conditions over EA, and vice versa. Similarly, the North Atlantic Ocean basin (NAOI) also shows a negative association with EA rainfall.

Similarly, T2 predictor basins were identified, and the main regions are shown in Figure 4. Broadly consistent thermal patterns are evident in both SST- and T2-based fields. Key regions include the Eurasian basin, the southern Pacific Ocean, and the southern Indian Ocean. Negative associations with EA rainfall are evident for WoAT2 and Eurasian T2 anomalies, whereas

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


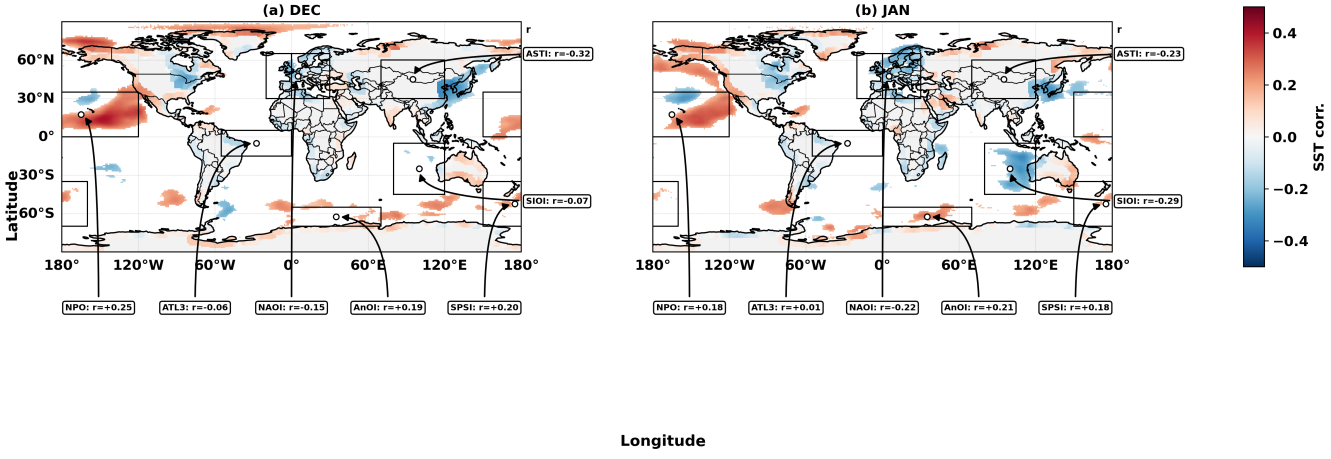


**Figure 3.** Correlation between antecedent sea surface temperature (SST) anomalies and East Africa (EA) March–May (MAM) rainfall PC1 for (a) December and (b) January. Shading shows statistically significant grid-point correlations at $p < 0.1$, and black boxes mark the retained SST predictor domains used to derive basin-mean oceanic indices. These domains represent candidate oceanic forcing regions linked to interannual EA MAM rainfall variability.

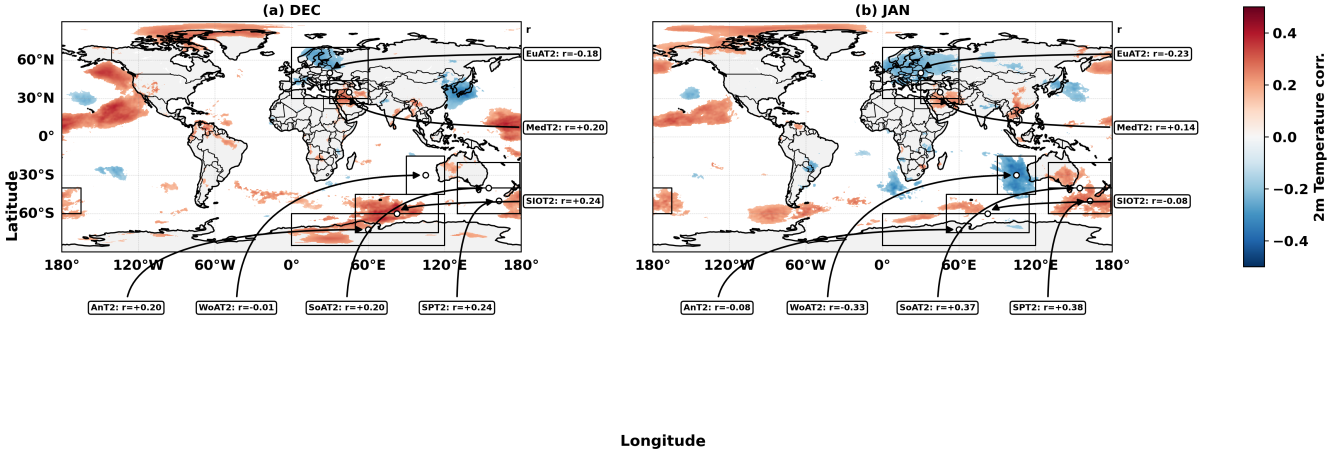


**Figure 4.** Correlation between antecedent T2 anomalies and East Africa (EA) March–May (MAM) rainfall PC1 for (a) December and (b) January. Shading shows statistically significant grid-point correlations at $p < 0.1$, while black boxes indicate the retained T2 predictor domains used for basin-mean index construction. The identified thermal regions were used to represent large-scale land–atmosphere temperature contrasts relevant to EA MAM rainfall variability.

SIOT2 (southern Indian Ocean basin) and southern Pacific T2 (SPT2) show positive relationships with EA rainfall. These contrasting signals indicate that antecedent thermal forcing over different continental and oceanic sectors contributes differently to subsequent EA MAM rainfall variability.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


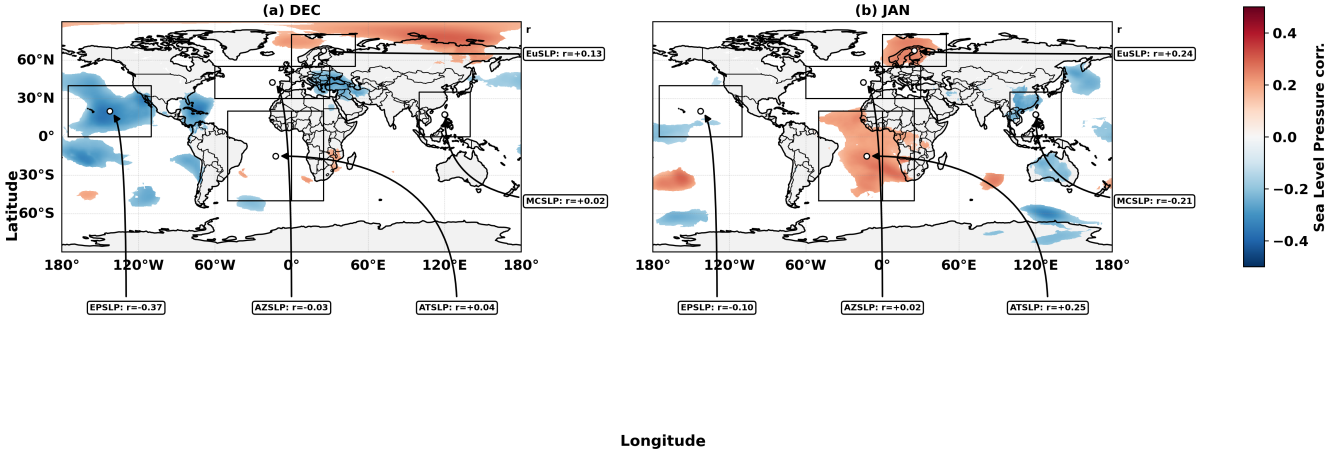


**Figure 5.** Correlation between antecedent sea level pressure (SLP) anomalies and East Africa (EA) March–May (MAM) rainfall PC1 for (a) December and (b) January. Shading shows statistically significant grid-point correlations at $p < 0.1$, while black boxes indicate the retained SLP predictor domains used for basin-mean index construction. Positive correlations denote regions where above-normal SLP is associated with a positive phase of EA MAM rainfall PC1, whereas negative correlations indicate the opposite relationship.

Furthermore, a similar analysis was performed for SLP (Figure 5), and nearly the same broad regions were identified. This indicates a close link between thermal heating or cooling, SLP anomalies, and the associated atmospheric circulation response. The eastern Pacific basin (EPSLP), Atlantic ocean basin (ATSLP), Eurasian ocean basin (EuSLP) and Azore basin covering north of Africa region (AZSLP) all show a considerable relationship with EA rainfall in both December and January. In December, EPSLP exhibits a negative relationship with EA rainfall. It is also notable that these systems appear to shift from December to January. This may reflect a seasonal reorganisation or eastward displacement of the dominant circulation features, consistent with the weakening or relocation of the relationships identified in the SST, T2, and SLP fields.

Following this screening step, basin-mean indices were computed from the identified regions and used as input features in the Random Forest (RF) and XGBoost (XGB) models. Iterative SHAP analysis, together with physical understanding of the identified systems, was then used to refine the final set of predictors. In addition to the primary indices, derived predictors such as SPID, AuT2GD, and EAT2GD were also computed; these represent gradients between selected basins. Driver tendencies were further calculated and included as ML features. Based on this procedure, Figure 6 shows the dominant January predictors for each rainfall zone within EA.

Each zone is influenced by a distinct set of drivers, and their contributions to rainfall predictability can also be inferred. The EQZ zone appears to be primarily driven by thermal conditions, in which higher values of most predictors are associated with enhanced rainfall, and vice versa (Figure 6a). The exception is southern Pacific 2-m temperature, which shows a negative contribution to EQZ rainfall. In contrast, the HoAZ zone shows both positive and negative contributions depending on the driver. Higher and lower values of these predictors may both contribute to rainfall variability, while near-neutral values appear to produce mixed effects for most drivers (Figure 6b). The SSRZ zone is influenced more strongly by SLP-related drivers.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


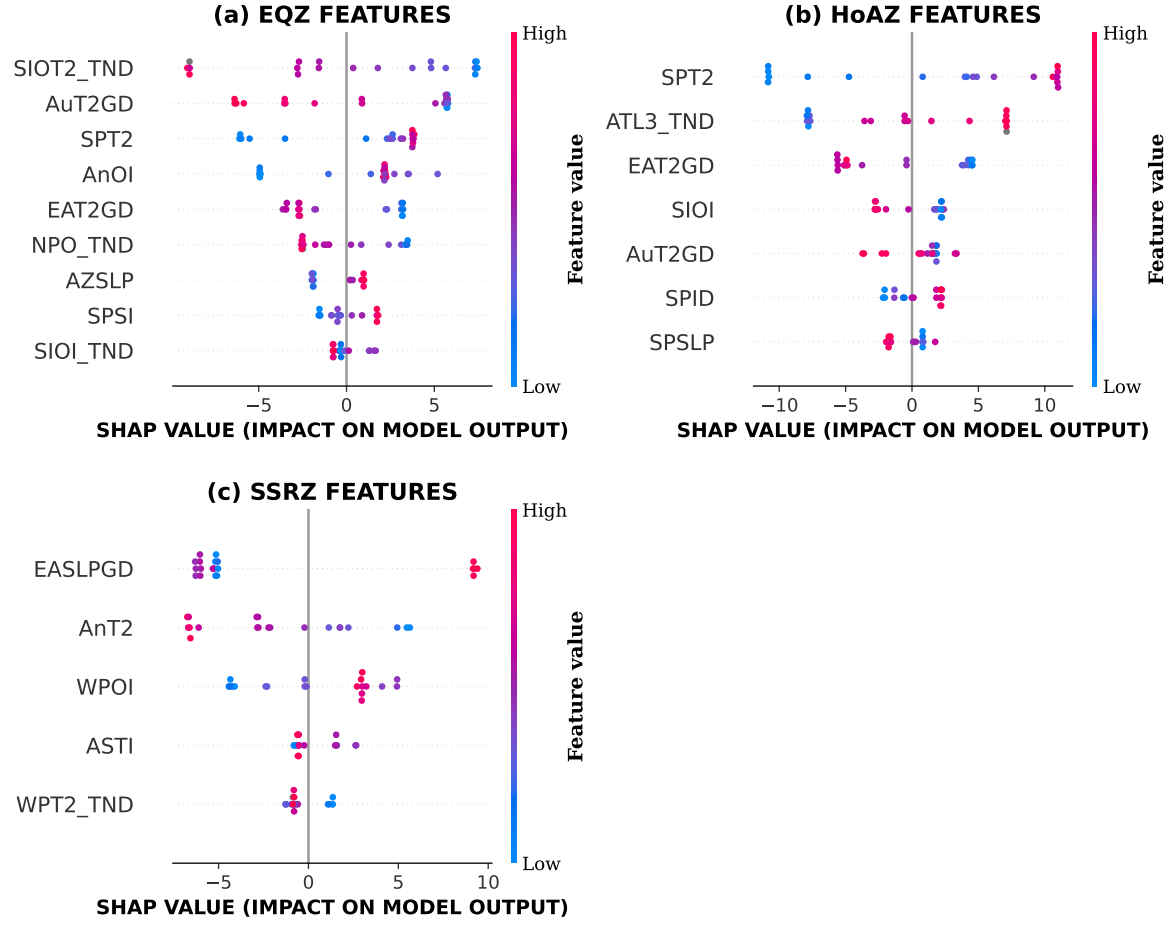


**Figure 6.** SHAP summary plots for the most influential January predictors retained in the XGBoost models for the three East Africa rainfall zones: (a) EQZ, (b) HoAZ, and (c) SSRZ. Each point represents one sample, positioned according to its SHAP value, which measures the contribution of a given predictor to the model output. Point colour indicates the magnitude of the predictor value from low (blue) to high (pink). Predictors with larger absolute SHAP values exert stronger influence on the forecast, while the sign of the SHAP value indicates whether the predictor increases or decreases the predicted seasonal rainfall anomaly.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


However, it is also the least predictable zone, likely because of its marked heterogeneity. Much of this zone is off-season during MAM, leaving only limited areas that are climatically in season, which makes prediction more challenging.

### 3.2 Dynamical Pathways Linking the Drivers to EA MAM Rainfall Variability

The leading Maximum Covariance Analysis (MCA) mode between SST anomalies and Walker circulation shows that the strongest coupled SST influence is located over the eastern Pacific, centred around 120–170°W in December (Figure 7a) and becoming more defined in January (Figure 7b). Most SST basins exhibit negative loadings, indicating that cooler basin conditions are associated with the Walker phase characterised by enhanced subsidence, whereas warmer conditions favour the opposite phase. The subtropical Atlantic is an exception, where cooler conditions are linked to enhanced convection.

The corresponding Walker structure displays a clear zonal overturning pattern, with alternating ascent and descent across the Pacific, Maritime Continent, Atlantic, Indian Ocean, and Africa. Over East Africa, January shows more pronounced ascent from the mid- to upper troposphere compared to December, suggesting a more convection-favourable Walker configuration during January.

The coupled MCA analysis between Walker circulation and T2 drivers reveals a spatial structure that is broadly consistent with the SST-forced Walker mode (Figure 7c,d). Enhanced subsidence is evident over the eastern Pacific, with multiple narrow centres in December (Figure 7c), and the strongest descending branch located near 160–175°W in December. In January, subsidence over the Pacific is weaker, with a broader but less intense descending region extending from about 120–175°W (Figure 7d).

Over the Atlantic, the ascending branch is stronger in January than in December, and also stronger than that associated with SST forcing alone. The zonal overturning signal extends across the Indian Ocean and Atlantic sectors, confirming a coherent Walker-like circulation with alternating regions of ascent and descent across the tropical belt.

The strength of the T2–Walker coupling exhibits clear seasonal variability. Several basins that strongly influence the Walker mode in December contribute less in January, leading to a reduction in the explained covariance fraction from about 75% in December to 60% in January. Nearly all T2 basins show negative loadings with the leading Walker mode in December, whereas only southern Pacific 2-m temperature (SPT2) shows positive contributions. Although this spatial pattern persists into January, the magnitude of the loadings generally weakens.

The MCA analysis between SLP and Walker circulation indicates that the strongest arm of the Walker mode is centred over the western Pacific, around 160°E–180° in December. Although this location differs slightly from the centres identified in the SST- and T2-forced Walker patterns, the overall zonal overturning structure remains similar.

In January, the dominant Walker centre shifts eastward to approximately 150–160°W, indicating a seasonal displacement of the main overturning branch. This shift may reflect changes in the large-scale circulation background, including possible modulation by Rossby-wave propagation. The SLP–Walker coupling also strengthens in January, with the explained covariance fraction increasing from about 64% in December to 86%.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


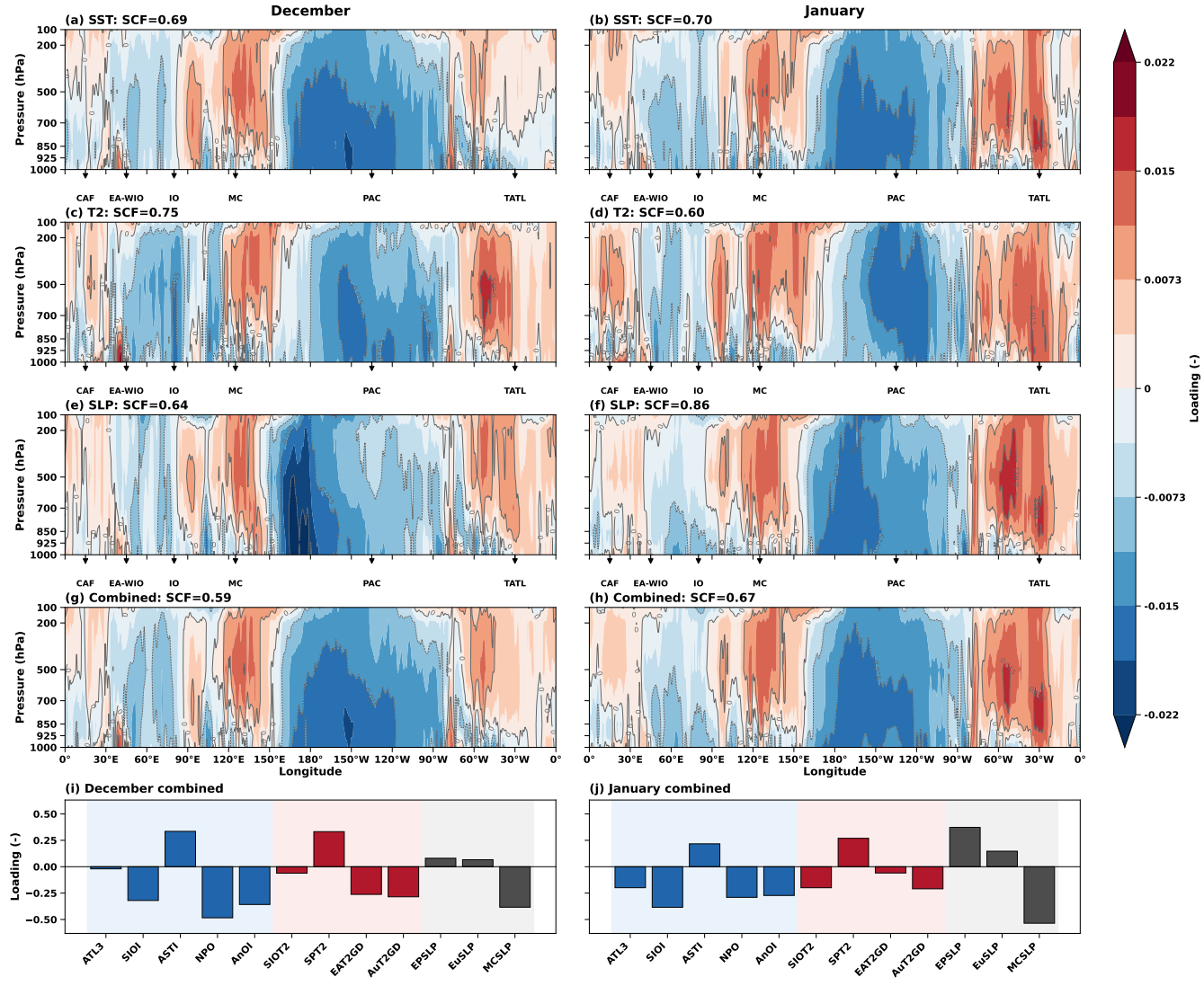


**Figure 7.** Walker circulation patterns associated with the leading maximum covariance analysis (MCA) mode between East Africa (EA) March–May (MAM) rainfall and antecedent climate drivers. Panels (a) and (b) show December and January SST-related Walker vertical-structure loadings, respectively; (c) and (d) show the corresponding T2-related loadings; (e) and (f) show the SLP-related loadings; and (g) and (h) show the combined-driver Walker loadings. Shading (blue denotes subsidence and red denotes ascent/convection) indicates MCA loadings in pressure–longitude space averaged over the equatorial belt (latitude $\pm 5^{\circ}$), with pressure on the vertical axis and longitude on the horizontal axis. Panels (i) and (j) summarise the combined December and January basin loadings for the retained predictors across the main Walker-circulation branches: TATL (tropical Atlantic/Gulf of Guinea sector), CAF (Central Africa/Congo Basin sector), EA–WIO (East Africa and western Indian Ocean sector, including the Horn of Africa and adjacent ocean), IO (Indian Ocean sector), MC (Maritime Continent sector, including Indonesia and surrounding seas), and PAC (Pacific sector, representing the ENSO-related central–eastern Pacific). The squared covariance fraction (SCF) shown in each panel indicates the fraction of shared covariance explained by the leading mode.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


Over the Atlantic, the ascending branch is better defined in January, with a broader convection centre extending from roughly 30°W to 60°W. Basin loading patterns show that eastern Pacific and European SLP regions have positive loadings with the leading Walker mode, whereas Maritime Continent and northern African SLP regions exhibit negative loadings.

The relationship between SLP anomalies and vertical motion reflects the expected dynamical behaviour: lower surface pressure is generally associated with enhanced convergence and upward motion, while higher surface pressure corresponds to divergence and subsidence. This pressure–vertical motion coupling is therefore broadly inverse to the thermal forcing patterns identified from SST and T2 anomalies.

After examining the individual climate drivers (panels a–f), panels g and h illustrate the combined contribution of SST, T2, and SLP forcings to the Walker circulation. The dominant large-scale structure remains anchored over the eastern Pacific, with the main centre extending from about 120°W to 180°W in December. In January, this centre becomes weaker and more spatially confined, with the strongest influence located near 150°W.

This combined response highlights the joint role of thermal forcing and surface pressure gradients shaping the zonal overturning circulation. Although the overall Walker structure remains similar between the two months, important differences arise from the longitudinal displacement of the Walker arms and from variations in the intensity of the ascending and descending branches.

The persistence of an eastern Pacific-centred Walker response in both months suggests that the integrated influence of oceanic thermal anomalies, continental heating, and pressure variability acts coherently to organise tropical overturning. The narrower and weaker January structure indicates seasonal modulation of forcing strength, likely reflecting changes in the relative contributions of thermal gradients and dynamical circulation processes. These results show that, while the overall Walker configuration remains stable, its intensity and zonal position evolve in response to the combined effects of multiple climate drivers.

### 3.3 Selected Case Studies: Probabilistic Forecasts and Associated Large-Scale Circulation Patterns

To further assess the physical consistency of the identified driver–rainfall relationships, we examined four contrasting case-study years: 2007 and 2012 as dry MAM seasons, and 2018 and 2020 as wet MAM seasons over East Africa (Figure 8). The analysis focuses on the January pre-season state, prior to the onset of the MAM rains, in order to evaluate whether the antecedent large-scale thermal and circulation anomalies are consistent with the subsequent rainfall outcomes. Both the Random Forest (RF) and Extreme Gradient Boosting (XGB) models reproduce the broad dry and wet rainfall anomalies reasonably well, suggesting that the selected predictors capture a physically meaningful component of East African MAM rainfall variability and therefore have potential utility for seasonal prediction.

The January SST anomaly fields reveal a clear contrast between the dry and wet case-study years (Figure 9). During the dry MAM seasons of 2007 and 2012, the pre-season oceanic state is characterised by cooling over key predictor regions, particularly in the Pacific ocean and Antarctica ocean , as indicated by negative anomalies within the NPO-related domain , South Pacific region (SPSI) and Antarctica region (AnOI) (Figure 9a,b). However, the two dry years are not identical. The 2012 case exhibits a broader and more coherent cooling pattern, with negative anomalies also extending into the Atlantic sector, including the ATL3

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


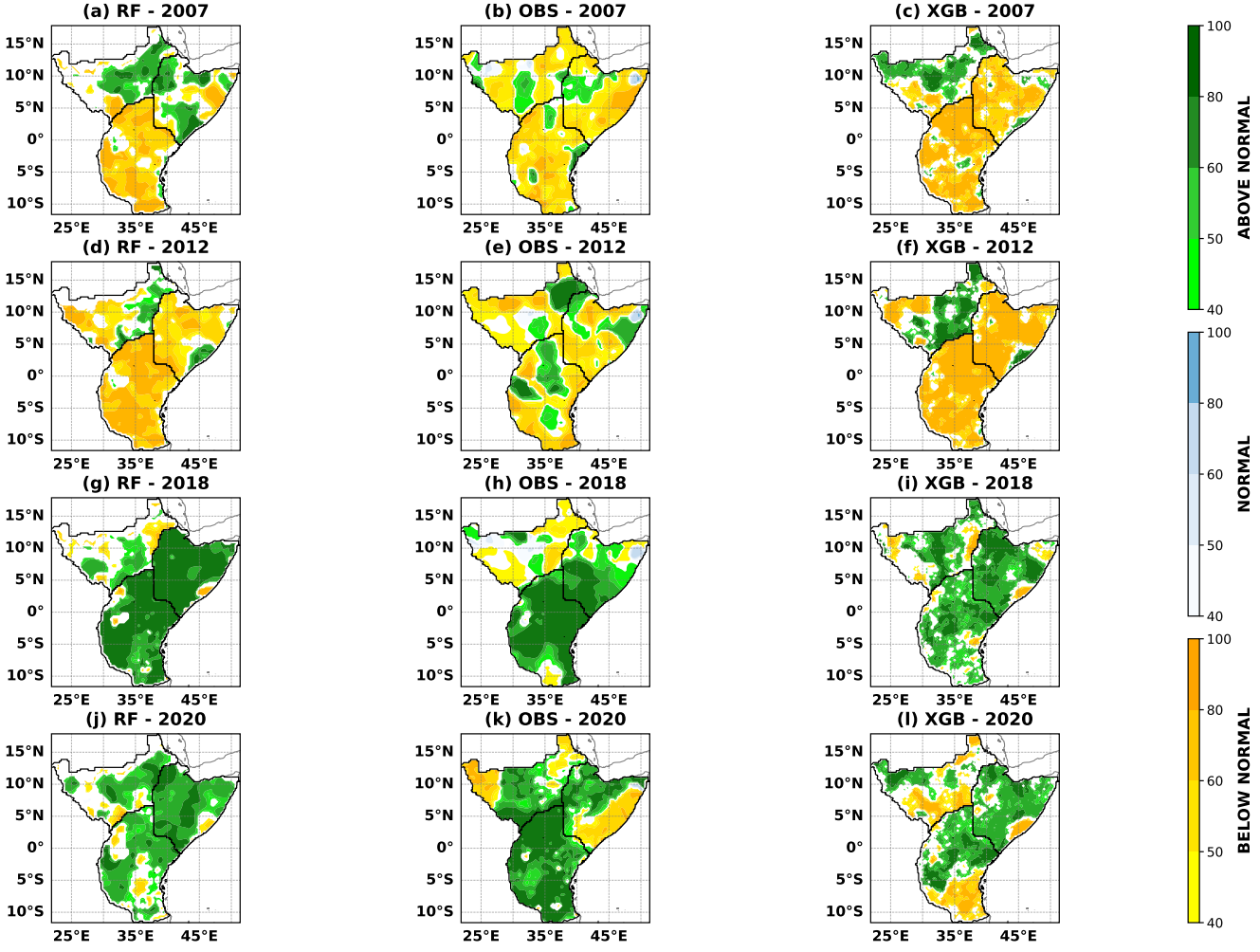


**Figure 8.** Probabilistic tercile rainfall forecasts from the Random Forest (RF) and XGBoost (XGB) models, compared with observed rainfall categories (OBS), for four representative East Africa MAM seasons: 2007, 2012, 2018, and 2020. For each year, the thre/e panels show RF, OBS, and XGB, respectively. Shading indicates the dominant tercile category and its associated probability: above normal, normal, and below normal rainfall. The figure is used to assess how well the two machine-learning models reproduce the observed spatial distribution of contrasting wet and dry MAM rainfall seasons across East Africa.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


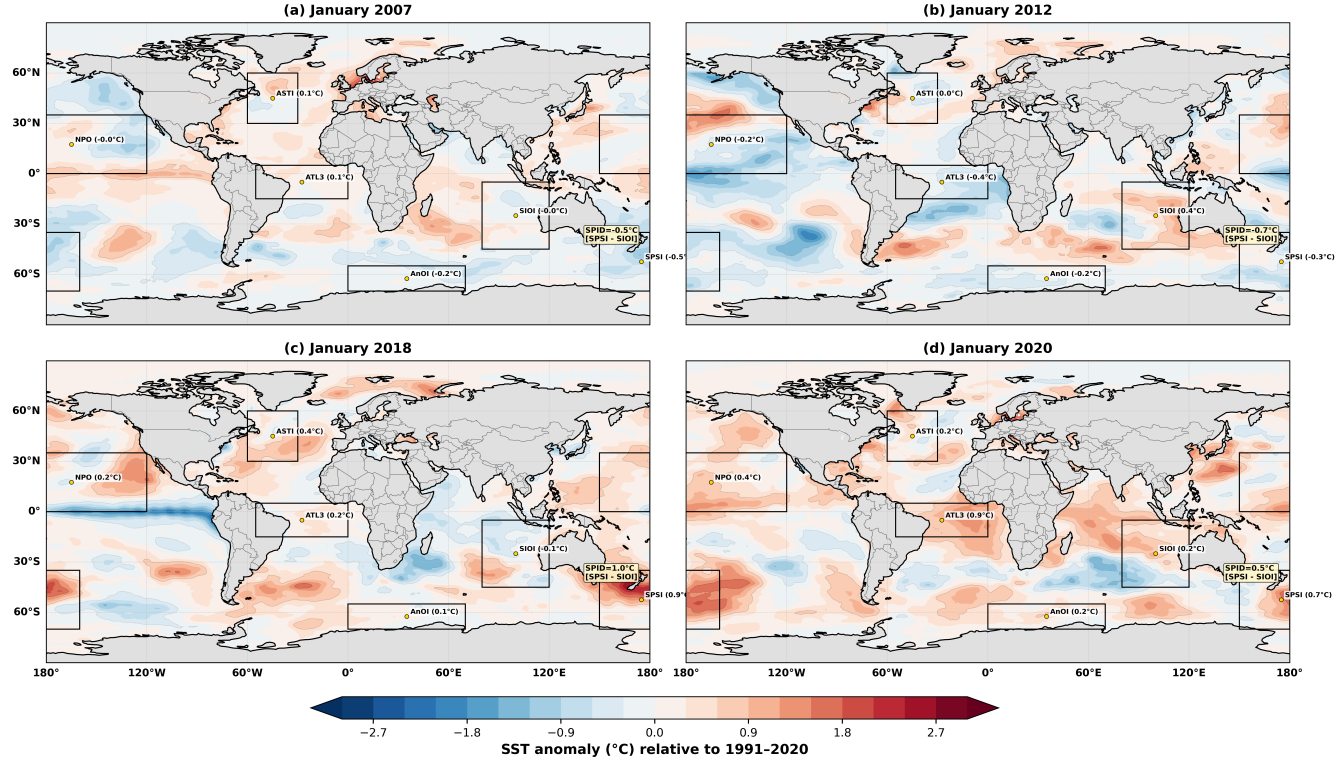


**Figure 9.** January sea surface temperature (SST) anomalies for four representative years: (a) 2007, (b) 2012, (c) 2018, and (d) 2020, relative to the 1991–2020 climatology. Shading denotes SST anomalies (°C), with warm colours indicating positive anomalies and cool colours indicating negative anomalies. Black boxes indicate the retained SST predictor domains, and labels show the corresponding basin-mean January anomalies and derived gradients where applicable. These patterns illustrate the antecedent oceanic thermal conditions associated with contrasting East Africa MAM rainfall outcomes.

and ASTI regions (Figure 9b). In contrast, 2007 shows weaker cooling and a relatively warmer background state, suggesting that it was closer to neutral conditions during January (Figure 9a). Despite these differences, both years exhibit pre-season thermal conditions that are consistent with suppressed convective support and deficient MAM rainfall over East Africa. By contrast, the wet MAM seasons of 2018 and 2020 are characterised by warmer SST anomalies across several key oceanic basins, including the Pacific, Atlantic, and Indian Ocean sectors (Figure 9c,d). This opposite thermal configuration suggests that basin-scale lower-boundary forcing plays an important role in modulating East African MAM rainfall. Overall, the case-study comparison indicates that cooling within key SST predictor regions is associated with dry MAM conditions, whereas warming in these regions provides a more favourable background for enhanced convection and wetter MAM rainfall. These results support the broader statistical relationship between the retained SST predictors and East African MAM rainfall variability.

The January near-surface air temperature ($T_2$) anomalies provide additional evidence that large-scale thermal gradients influence East African MAM rainfall (Figure A1). In the dry years, 2007 and 2012, cooling is evident over several predictor

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


regions linked to the Indo-Pacific and south-west Pacific–Australian sector (Figure A1a,b). These anomalies are physically important because they influence zonal and meridional thermal gradients that can modify large-scale circulation, moisture transport, and vertical motion over the western Indian Ocean and East Africa. The Eurasian temperature gradient, represented by EAT2GD, shows a less consistent relationship with the MAM rainfall outcome. For example, EAT2GD is stronger in 2007, reaching approximately $3°C$, but is much weaker in 2012, at about $0.2°C$ (Figure A1a,b). Since both years were dry despite the different magnitudes of EAT2GD, this gradient appears to act as a complementary rather than a dominant driver of East African MAM rainfall. Its influence may therefore depend on the concurrent state of other oceanic and atmospheric drivers.

In contrast, the Indo-Pacific and south-west Pacific–Australian temperature drivers, including SIOT2, SPT2, SoAT2, and AuT2GD, show a more coherent relationship with the rainfall anomalies. Cooling over the SoAT2 region, which extends from the south-west Pacific towards the Australian sector, is associated with dry MAM conditions over East Africa. This cooling modifies the Austral temperature gradient, AuT2GD, defined by the thermal contrast between the western Australian–eastern Indian Ocean sector and the south-west Pacific sector. When the south-west Pacific sector cools while the western Australian–eastern Indian Ocean sector remains relatively warm, the AuT2GD gradient strengthens. This configuration is dynamically consistent with circulation anomalies that reduce convective support over East Africa and favour dry MAM rainfall. The wet years, 2018 and 2020, display the opposite behaviour. Warming over the SoAT2 region weakens or reverses the Austral temperature gradient, producing negative AuT2GD values (Figure A1c,d). This configuration is associated with wetter-than-normal MAM conditions, particularly over the Horn of Africa Zone (HoAZ) and the Equatorial Zone (EQZ). The contrast between the dry and wet case-study years therefore suggests an inverse relationship between thermal anomalies over the south-west Pacific–Australian sector and East African MAM rainfall: cooling in this region tends to favour deficient rainfall, whereas warming is associated with enhanced rainfall.

Similarly, the sea-level-pressure (SLP) predictors show an inverse relationship with the thermal anomalies. During years with warmer sea surface temperature (SST) and 2 m air temperature (T2) anomalies, pressure anomalies generally weaken or become negative, whereas cooler thermal anomalies are associated with stronger or positive pressure anomalies (Figure A4). After machine-learning screening, the Euro–Asian sea-level-pressure gradient (EASLPGD) and the Azores sea-level-pressure index (AZSLP) were the main SLP-related predictors retained for MAM rainfall prediction, particularly for the Sub-Saharan Region (SSRZ).

However, the case-study years do not show a simple linear relationship between these SLP predictors and MAM rainfall. The dry seasons of 2007 and 2012 (Figure A4a,b) and the wet seasons of 2018 and 2020 (Figure A4c,d) exhibit mixed pressure-anomaly patterns. This suggests that EASLPGD and AZSLP do not act as isolated predictors of rainfall anomalies. Instead, their predictive value likely depends on how they interact with concurrent SST and T2 anomalies to modify pressure gradients, regional circulation, moisture transport, and vertical motion over East Africa.

Taken together, the SST , SLP and T2 case-study analyses indicate that East African MAM rainfall variability is strongly linked to the pre-season thermal state of the Indo-Pacific, Atlantic, and adjacent continental sectors. The dry years are associated with cooler antecedent conditions in key predictor basins and reduced thermal support for convection, whereas the wet years are associated with broader warming and a more favourable background for moisture convergence and uplift. These physically

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


consistent contrasts provide additional confidence that the selected machine-learning predictors are not only statistically informative but also dynamically interpretable.

The Walker circulation structure provides further dynamical support for these results (Figure 10). The key sectors of interest include the eastern Pacific, the Maritime Continent, the western Indian Ocean adjacent to the East African coast, and the Atlantic basin.In 2007 and 2012, the western Indian Ocean–East African tropical coastal sector, approximately 40°E–80°E, is characterised by coherent and persistent subsidence. This region forms part of the Indian Ocean Walker circulation, where vertical motion strongly modulates moisture convergence and convective activity along the East African coast. The anomalous downward motion over this sector is therefore consistent with suppressed convection over the western Indian Ocean and reduced moisture supply into East Africa, thereby favouring dry MAM rainfall conditions.

The circulation structure in both 2007 and 2012 further suggests a weakened, displaced, or partially reversed Walker-cell configuration. Under this orientation, ascent is shifted away from the western Indian Ocean–East African coastal sector, while moisture transport is favoured towards the eastern Indian Ocean and the Maritime Continent. This displacement of convective support limits uplift over East Africa and provides a plausible dynamical explanation for the rainfall deficits observed during the 2007 and 2012 MAM seasons.

In contrast, the wet years 2018 and 2020 show enhanced convection over the western Indian Ocean and East African coastal sector, supporting increased moisture convergence and enhanced rainfall during MAM. These wet cases are therefore dynamically consistent with the thermal anomalies described above, whereby warmer conditions in the Pacific and Atlantic basins favour a Walker circulation configuration with ascending motion over the western Indian Ocean. The eastern Pacific signal is also consistent with the MCA results, which indicate that cooling in this basin is linked to enhanced subsidence, whereas warming is associated with greater convection. In addition, the Atlantic basin appears to play an important role in regulating the location and intensity of the Indian Ocean Walker circulation branch over Africa. Taken together, these case studies provide event-scale support for the statistical relationships identified by the machine learning models and MCA, and demonstrate that the selected drivers are physically relevant for explaining both dry and wet MAM rainfall anomalies over East Africa.

### 3.4 Forecast Skill Evaluation Using BSS, AUC, and ROC

The evaluation focused on the Below Normal (BN) and Above Normal (AN) categories only, because the Normal category represented less than 2% of the available data and had limited spatial coverage in the observations. Model skill was assessed using 22 randomly selected validation seasons. The metrics retained for evaluation are the Brier Skill Score (BSS; Figure 11), the Area Under the Curve (AUC; Figure 12), and the Receiver Operating Characteristic and reliability analysis (ROC; Figure 13).

The BSS maps (Figure 11) indicate that both the RF and XGB models provide meaningful probabilistic skill for MAM rainfall prediction across EA. Spatially averaged BSS values are positive for both models, with RF outperforming XGB: RF achieves a spatial mean BSS of 0.48, compared with 0.40 for XGB. These values correspond to average reductions in probabilistic forecast error of approximately 48% and 40%, respectively, relative to climatology. However, the spatial distribution shows that skill is not uniform across the domain. Many regions attain BSS values above 0.70, indicating strong local probabilistic

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


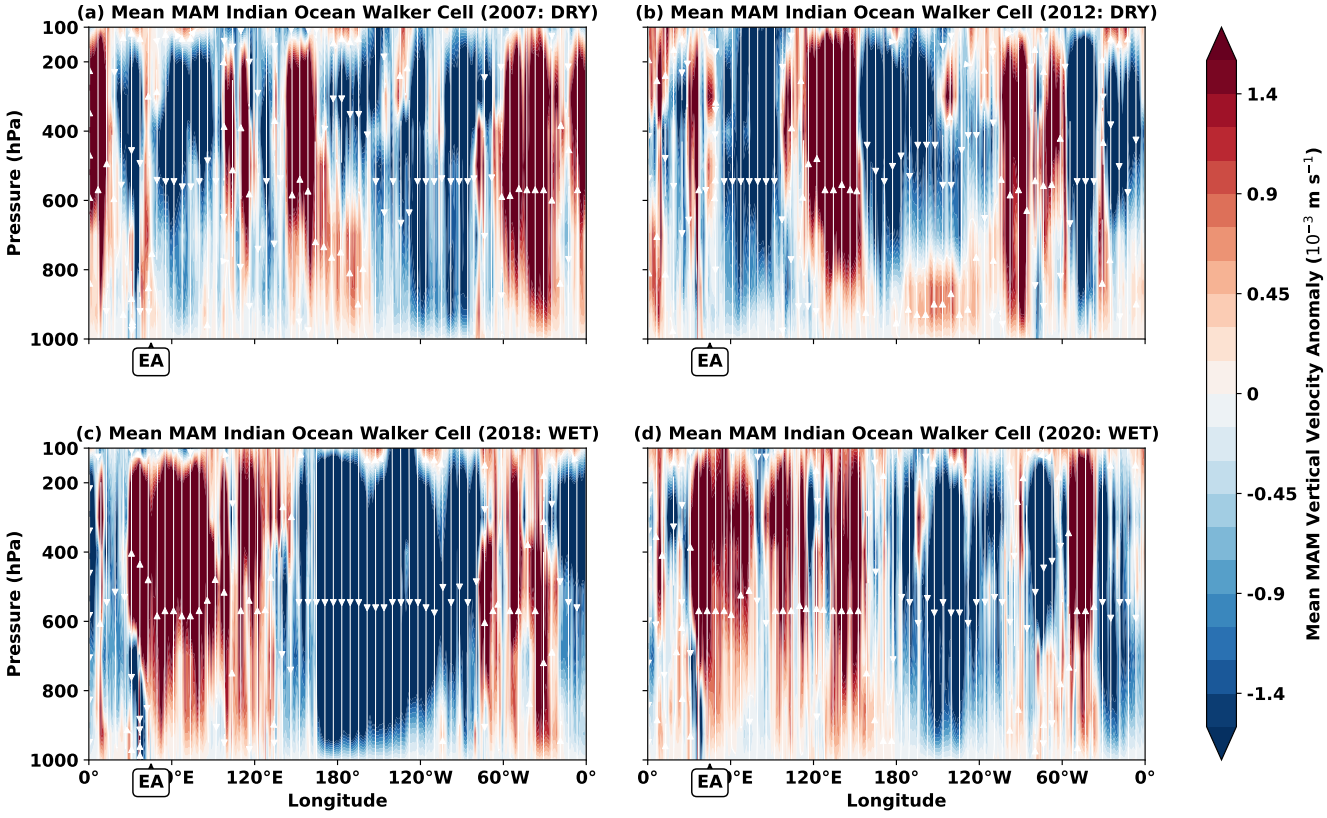


**Figure 10.** Mean March–May (MAM) Indian Ocean Walker circulation (IOWC) vertical-velocity anomalies for four representative years: (a) 2007, (b) 2012, (c) 2018, and (d) 2020. Shading denotes anomalous vertical velocity ($10^{-3}$ m s$^{-1}$) in pressure–longitude coordinates, relative to the 1991–2020 climatology, while white streamlines indicate the associated zonal–vertical circulation anomalies. The selected years represent contrasting dry and wet East Africa MAM rainfall seasons and illustrate differences in the structure and intensity of the Walker circulation during those cases.

skill, particularly within parts of the EQZ and HoAZ regions. The lower domain-mean values therefore reflect the influence of weaker-skill areas, highlighting the spatial heterogeneity of MAM rainfall predictability while confirming the added probabilistic value of the ML-based forecasts.

The AUC results further demonstrate that both ML algorithms discriminate between below-normal (BN) and above-normal (AN) MAM rainfall conditions at levels clearly above random chance (Figure 12). The reported AUC values represent spatially averaged skill over the evaluation domain. Averaged over the 22 seasons, RF achieves domain-mean AUC values of approximately 0.72 for both the BN and AN categories, while XGB achieves corresponding values of about 0.65. These values indicate moderate but meaningful discriminatory skill in identifying anomalously dry and wet seasonal rainfall conditions over East Africa.

The spatial distribution of AUC reveals substantially stronger local performance than suggested by the domain averages alone. In particular, most regions for the RF model attain AUC values above 0.80 for the BN, AN, and overall skill categories,

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


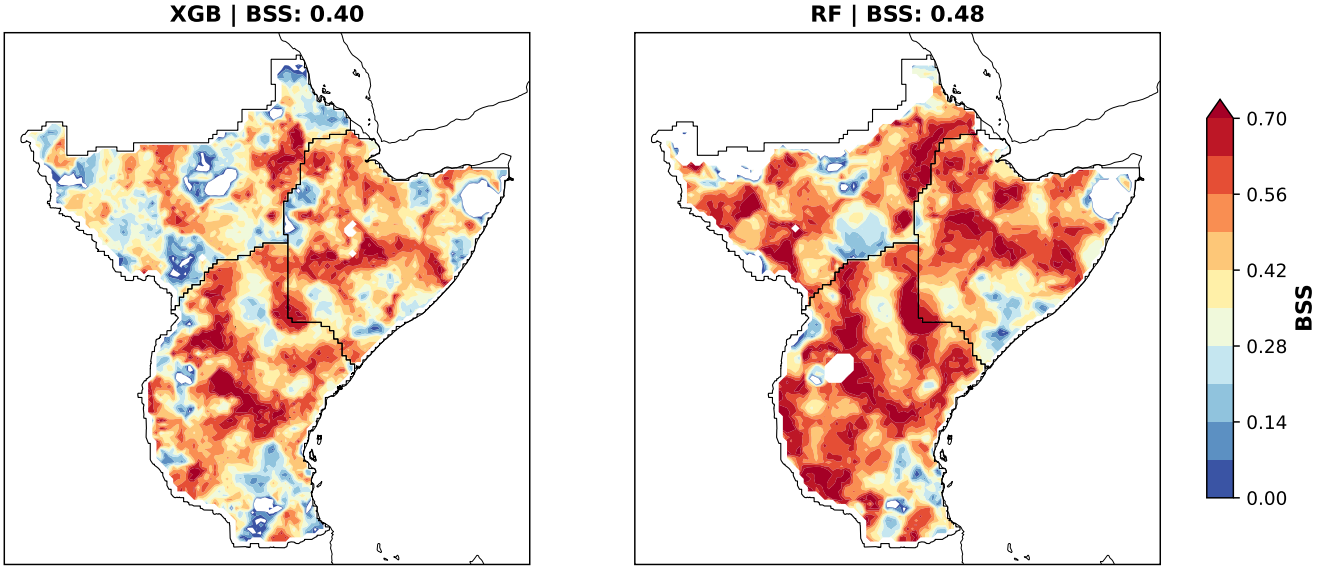


**Figure 11.** Spatial distribution of Brier Skill Score (BSS) for the RF and XGB models computed from the 22 validation seasons. A perfect score is 1, while values below 0 indicate performance worse than climatology; these negative values are masked and appear as white areas.

indicating strong discrimination skill in key sub-regions. The lower spatially averaged values therefore reflect the influence of areas with weaker predictability, highlighting the spatial heterogeneity of MAM rainfall forecast skill across the domain.

Relative to the current ICPAC operational multi-model ensemble, which has produced AUC values below 0.40 over recent MAM seasons, the ML models provide a substantial improvement in discriminatory performance. The RF and XGB models also outperform ECMWF SEAS5 over Africa, for which mean AUC values are approximately 0.545 and 0.54 for the BN and AN categories, respectively. Overall, these results show that the machine-learning models, particularly RF, provide more reliable discrimination of dry and wet seasonal rainfall outcomes than both the operational multi-model ensemble and the SEAS5 benchmark.

The ROC and reliability analysis confirms these findings. Both models show under-confidence at lower forecast probabilities (below about 50%) and over-confidence at higher probabilities. This behaviour is consistent with the probability distributions of the selected seasons, where markedly dry or wet years often produce event probabilities exceeding 60%. Despite this calibration issue, the ROC curves show that both RF and XGB retain useful discriminative skill, with RF consistently outperforming XGB. Overall, the retained metrics demonstrate that the machine-learning models provide useful probabilistic guidance for extreme MAM rainfall categories over East Africa, with RF showing the most robust performance.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


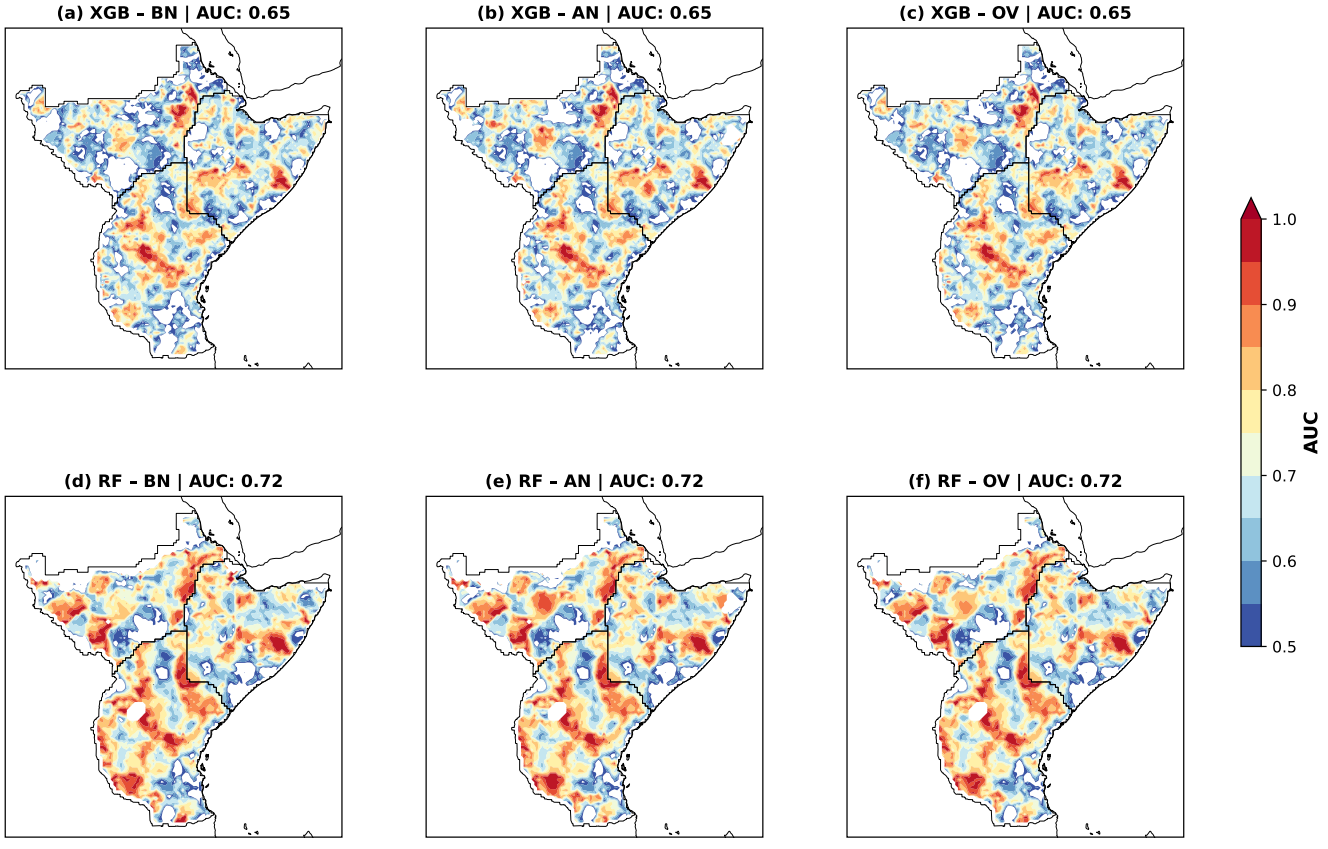


**Figure 12.** Spatial distribution of Area Under the Curve (AUC) for the RF and XGB models computed from the 22 validation seasons. A perfect score is 1, whereas values below 0.5 indicate no useful discrimination skill.

## 4 Discussion

The spatial configuration of the dominant drivers changes markedly between the December- and January-initialised analyses, indicating that the physical sources of MAM predictability evolve during the pre-season period. In December, 2-m air temperature (T2) anomalies over the western Pacific basin exert a clear influence on the SSRZ, whereas this relationship is not evident in January (Figure A6). This enhanced role of T2-related predictors is physically plausible given recent evidence that remote land–atmosphere thermal anomalies can modulate East African long-rains variability. In particular, Hagos et al. (2024) showed that northwestern Eurasian warming is negatively associated with East African MAM rainfall through a weakening of the regional overturning circulation, enhanced subsidence over East Africa, and reduced moisture convergence. Although the T2 centres identified here differ spatially from the Eurasian region emphasized by Hagos et al. (2024), the result supports the broader interpretation that remote near-surface thermal anomalies can provide dynamically relevant information for MAM rainfall prediction.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


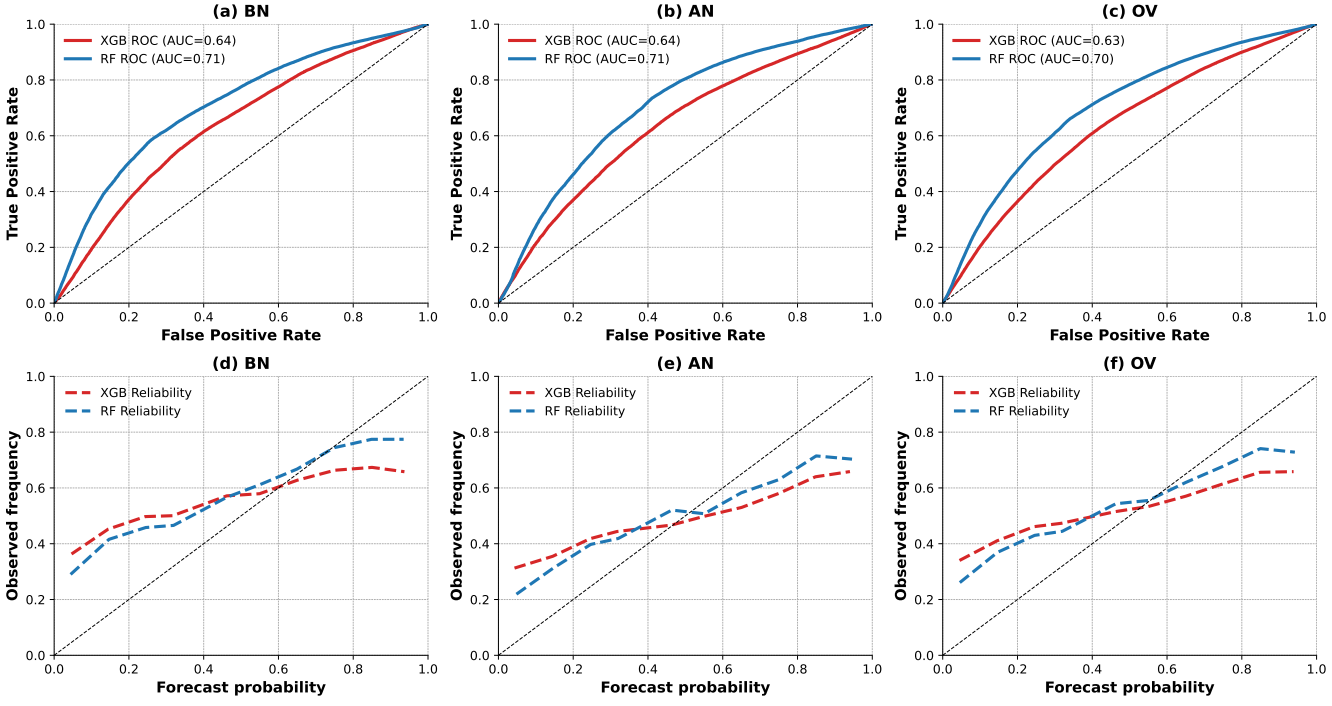


**Figure 13.** Receiver operating characteristic (ROC) curves and reliability diagrams for the RF and XGB models, constructed from the 22 validation seasons across East Africa. ROC curves summarise the ability of the models to discriminate between event and non-event outcomes, while reliability diagrams assess the consistency between forecast probabilities and observed frequencies.

In contrast to the more variable T2 influence, the dominant SST signal remains more closely linked to the Atlantic Ocean, particularly through ATL3. This is consistent with previous studies showing that East African rainfall is influenced by remote SST teleconnections across the Atlantic, Indian and Pacific Ocean basins (Ogallo et al., 1988; Nicholson, 2017; Palmer et al., 2023). The importance of Atlantic and Indo-Pacific variability for the potential predictability of the East African long rains has also been emphasized by Ward et al. (2023), who linked MAM rainfall predictability to large-scale SST forcing and Congo zonal wind variability. Similarly, Williams and Funk (2011) showed that Indo-Pacific warming can alter the Walker circulation and contribute to drying over eastern Africa, while Wainwright et al. (2019) demonstrated that SST and pressure-gradient changes over the Arabian Sea and adjacent regions affect the timing and duration of the long rains. Therefore, the SST predictors selected by the models should be interpreted not only as local oceanic drivers, but also as indicators of basin-scale circulation adjustments that influence moisture transport and convergence over East Africa.

Figures A2–A3 further show that both the T2 and SST driver regions are generally warmer in December than in January. This cooling from December toward January suggests that both oceanic and near-surface atmospheric precursor conditions evolve substantially as the system approaches the MAM season. Such evolution may partly explain why MAM is often regarded as a transition season and why canonical teleconnections such as ENSO and the IOD tend to provide weaker and less stationary

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


predictive information during the long rains than during seasons closer to their peak influence. The shift in the relevant predictor basins between December and January also highlights a potential limitation of relying only on traditional teleconnection regions, such as those used to define ENSO and the IOD, for MAM prediction. Instead, the results suggest that broader and seasonally evolving SST, T2 and SLP driver basins may be needed to capture the physical pathways through which predictability arises.

A further notable result is the strong influence of T2-related drivers in both the December and January initialisations (Figures 2 and A6). Although their relative importance varies between the two lead times, the repeated dominance of T2 predictors suggests that near-surface atmospheric thermal gradients contain important information on the evolving large-scale circulation prior to the MAM season. This is particularly noteworthy because SST is generally expected to provide stronger seasonal memory, owing to the slower thermal adjustment of the ocean compared with the atmosphere. However, the emergence of similar key basins in both the SST- and T2-based predictors suggests that these two sources of predictability are not independent. Instead, they likely reflect closely linked ocean–atmosphere processes, whereby SST anomalies contribute to heating or cooling of the overlying near-surface air, thereby influencing 2-m temperature gradients, surface pressure contrasts, and the associated circulation response.

The prominence of T2 drivers therefore does not imply that SST forcing is unimportant; rather, it suggests that MAM predictability emerges from coupled ocean–land–atmosphere interactions. In this framework, SST anomalies provide slowly varying boundary forcing, T2 anomalies represent the near-surface atmospheric expression of remote and regional thermal contrasts, and SLP gradients translate these contrasts into circulation and moisture-transport responses. The similarity between some SST and T2 driver basins further supports a physically consistent link between oceanic thermal anomalies and lower-atmospheric circulation adjustments. This interpretation is consistent with the established role of global SST basins in East African rainfall variability (Ogallo et al., 1988), the non-stationary nature of long-rains teleconnections (Nicholson, 2017; Palmer et al., 2023), and recent evidence for the combined importance of Atlantic, Indo-Pacific and land-surface thermal forcing (Ward et al., 2023; Hagos et al., 2024). This interpretation is further supported by Rowell (2019), who showed that warming of the southern Indian Ocean may substantially increase East African long-rains rainfall by the end of the twenty-first century, with some projections suggesting increases approaching a doubling relative to present-day conditions. This finding highlights the additional complexity introduced by climate change when interpreting SST–MAM rainfall relationships. While historical SST anomalies may be linked to suppressed rainfall through Walker-circulation changes, moisture-transport anomalies and regional pressure-gradient adjustments, future warming patterns may alter these relationships and modify the strength, location and even sign of SST teleconnections. Therefore, the role of SST drivers in MAM rainfall prediction should be interpreted within a non-stationary climate context, where both historical variability and forced long-term trends can influence the apparent relationship between ocean basins and East African rainfall.

In addition, OND_ANOM_PY, representing October–December rainfall anomalies in the preceding season, shows a consistent positive relationship with MAM rainfall in both the December-initialised analyses. This indicates that antecedent wet or dry conditions during OND contribute to the likelihood of wetter or drier conditions in the subsequent MAM season. The persistence of this predictor suggests that, in addition to remote SST, T2 and SLP forcing, regional hydroclimatic memory may provide an

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


additional source of predictability. Such memory may arise through soil-moisture persistence, land–atmosphere feedbacks, or the tendency for large-scale circulation anomalies to persist across adjacent seasons.

Despite the differences in dominant physical drivers, the December- and January-initialised models exhibit broadly similar forecast skill, with AUC and BSS values differing by only about seven percentage points and January showing a marginal improvement over December (Figures A8 and A7). This small difference suggests that both initialisation periods are comparably effective for forecasting MAM rainfall over East Africa, although they draw predictive information from partly different precursor states. The similarity in skill, despite the changing combinations of SST, T2 and SLP drivers, highlights the complexity of the MAM rainfall system. It also suggests that comparable forecast performance can emerge from different but physically meaningful pathways: December skill appears to be more strongly linked to remote thermal forcing and early-season circulation adjustment, whereas January skill reflects a later-stage configuration of SST, pressure-gradient and hydroclimatic persistence signals. Overall, these results support the view that MAM rainfall predictability over East Africa is not controlled by a single dominant teleconnection, but by seasonally evolving interactions among oceanic forcing, land–atmosphere thermal gradients, pressure anomalies and regional moisture transport.

## 5 Conclusions

This study examined the physical controls and seasonal predictability of East African MAM seasonal total rainfall using basin-derived sea surface temperature (SST), 2-m air temperature (T2), sea level pressure (SLP), and land-surface predictors together with RF and XGB models. The predictor basins were first identified from December and January anomalies relative to the 1991–2020 climatology and their correlation with the leading mode of EA MAM rainfall variability. Basin-mean indices, spatial gradients, and tendency features were then selected using SHAP and applied in probabilistic ML forecasts.

The main finding is that East Africa (EA) March-May (MAM) rainfall predictability is controlled by a combination of remote and regional thermal and pressure-gradient drivers. The dominant predictors are South Indian Ocean T2 Tendency (SIOT2_TND), Australia T2 temperature gradient (AuT2GD) , South Pacific T2 (SPT2) , Antarctic Ocean Index (AnOI) , Atlantic Nino Tendency (ATL3_TND), EuroAsia T2 gradient (EAT2GD), Euro-Africa SLP gradient (EASLPGD), and Antarctic T2 (AnT2). These drivers make the largest contributions to prediction in both models, while additional lower-ranked predictors add complementary climate information. The repeated selection of these basin-derived predictors shows that MAM predictability is not controlled by one teleconnection, but by interacting SST, T2, and SLP signals across the Indian Ocean, Pacific, Atlantic, Eurasian, Australian, Antarctic, and Euro–African sectors.

A key result is that T2-related predictors dominate both December and January initialisations. This demonstrates that near-surface thermal gradients provide useful pre-season information in addition to SST memory. Similar SST and T2 predictor basins indicate a coupled ocean–atmosphere pathway in which SST anomalies heat or cool the overlying air, modify T2 gradients, alter SLP contrasts, and influence circulation and moisture transport. Several T2 predictors, including SIOT2_TND, AuT2GD, EAT2GD, and AnT2, are negatively related to EA MAM rainfall anomalies, while other selected drivers contribute positively. This confirms that rainfall anomalies depend on the combined configuration of thermal and pressure-gradient forcing.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


The physical diagnostics show that these drivers affect EA MAM rainfall through SLP gradients, Walker-circulation changes. Dry seasons are associated with subsidence over the western Indian Ocean and East African coastal sector, while wet seasons are associated with enhanced ascent and moisture convergence. The dominant drivers also differ between December and January, confirming that the sources of predictability evolve with lead time.

The ML models provide useful probabilistic skill. For January initialisation, RF reaches a spatially averaged BSS of about 0.48, while XGB reaches about 0.41, corresponding to reductions in probabilistic forecast error of about 48% and 41% relative to climatology. The corresponding AUC values are about 0.72 for RF and 0.65 for XGB, showing useful discrimination of below-normal and above-normal rainfall categories. Comparable skill from December initialisation demonstrates that forecast information is already present before January.

Overall, the study shows that physically constrained ML can improve East African MAM rainfall prediction while preserving a clear physical interpretation. Basin-derived SST, T2, and SLP predictors link forecast skill to identifiable ocean–land–atmosphere processes, providing a stronger basis for seasonal early-warning information across East Africa.

*Code availability.* The code used in this study is available from the corresponding author upon reasonable request.

*Data availability.* Open-source datasets were used in this study and are publicly available. The specific data sources are described in Sect. 2.

*Author contributions.* The first author conceived the study, designed and implemented the modelling framework, conducted the analyses, and prepared the manuscript with supervision from the last two authors. All co-authors contributed to scientific discussions, interpretation of the results, and revision of the manuscript to ensure scientific rigour and clarity.

*Competing interests.* The authors declare that they have no conflict of interest.

*Acknowledgements.* The author gratefully acknowledges NORCAP for supporting the pursuit of doctoral studies while fulfilling professional duties across Africa, including East Africa and Malawi. This support has been invaluable. The author also acknowledges Wageningen University for providing access to its high-performance computing facilities, which were essential for conducting the numerical simulations and machine learning analyses presented in this study. ChatGPT (OpenAI) was used to assist with Python code development, debugging, and language editing. All scientific analyses, methodological choices, interpretation of results, validation of outputs, and final manuscript decisions were conducted by the author. The author takes full responsibility for the manuscript content.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


## References

Antonio, B., McRae, A. T., MacLeod, D., Cooper, F. C., Marsham, J., Aitchison, L., Palmer, T. N., and Watson, P. A.: Postprocessing East African Rainfall Forecasts Using a Generative Machine Learning Model, Journal of Advances in Modeling Earth Systems, 17, https://doi.org/10.1029/2024MS004796, 2025.

Baig, F., Alghamdi, A. S., Al-Quraishi, A. M. F., and Ahmad, F.: How accurate are the machine learning models in improving monthly rainfall prediction in hyper-arid environment?, Journal of Hydrology, 629, 129 040, https://doi.org/10.1016/j.jhydrol.2024.129040, 2024.

Baldwin, M. P., Gray, L. J., Dunkerton, T. J., Hamilton, K., Haynes, P. H., Randel, W. J., Holton, J. R., Alexander, M. J., Hirota, I., Horinouchi, T., Jones, D. B. A., Kinnersley, J. S., Marquardt, C., Sato, K., and Takahashi, M.: The Quasi-Biennial Oscillation, Reviews of Geophysics, 39, 179–229, https://doi.org/10.1029/1999RG000073, 2001.

Bauer, P.: What if? Numerical weather prediction at the crossroads, Journal of the European Meteorological Society, 1, 100 002, https://doi.org/https://doi.org/10.1016/j.jemets.2024.100002, 2024.

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Camberlin, P. and Philippon, N.: The East African March–May rainy season: Associated atmospheric dynamics and predictability over the 1968–97 period, Journal of Climate, 15, 1002–1019, https://doi.org/10.1175/1520-0442(2002)015<1002:TEAMMR>2.0.CO;2, 2002.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, pp. 785–794, ACM, New York, NY, USA, https://doi.org/10.1145/2939672.2939785, 2016a.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, ACM, https://doi.org/10.1145/2939672.2939785, 2016b.

Chinyoka, S. and Steeneveld, G.: Evaluation of downscaling seasonal climate forecasts for crop yield forecasting in Zimbabwe, Climate Services, 30, 100 380, https://doi.org/https://doi.org/10.1016/j.cliser.2023.100380, 2023.

Copernicus Climate Change Service (C3S): ERA5 monthly data on pressure levels from 1940 to present, https://doi.org/10.24381/cds.bd0915c6, 2023a.

Copernicus Climate Change Service (C3S): ERA5 monthly data on single levels from 1940 to present, https://doi.org/10.24381/cds.adbb2d47, 2023b.

Deman, V. M. H., Koppa, A., Waegeman, W., MacLeod, D. A., Bliss Singer, M., and Miralles, D. G.: Seasonal prediction of Horn of Africa long rains using machine learning: The pitfalls of preselecting correlated predictors, Frontiers in Water, 4, https://doi.org/10.3389/frwa.2022.1053020, 2022.

Diro, G. T., Grimes, D. I. F., and Black, E.: Teleconnections between the Indian Ocean Dipole and East African short rains, International Journal of Climatology, 31, 557–578, https://doi.org/10.1002/joc.2086, 2011.

Dunning, C. M., Black, E., and Allan, R. P.: The Onset and Cessation of Seasonal Rainfall over Africa, Journal of Geophysical Research: Atmospheres, 121, 11 405–11 424, https://doi.org/10.1002/2016JD025428, 2016.

Finney, D. L., Marsham, J. H., Walker, D. P., Birch, C. E., Woodhams, B. J., Jackson, L. S., and Hardy, S.: The Effect of Westerlies on East African Rainfall and the Associated Role of Tropical Cyclones and the Madden–Julian Oscillation, Quarterly Journal of the Royal Meteorological Society, 146, 647–664, https://doi.org/10.1002/qj.3705, 2020.

Friedman, J. H.: Greedy function approximation: a gradient boosting machine, Annals of Statistics, 29, 1189–1232, https://doi.org/10.1214/aos/1013203451, 2001.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


Funk, C., Harrison, L., Shukla, S., Pomposi, C., Galu, G., Korecha, D., Husak, G., Magadzire, T., Davenport, F., Hillbruner, C., Eilerts, G., Zaitchik, B., and Verdin, J.: Examining the role of unusually warm Indo-Pacific sea-surface temperatures in recent African droughts, Quarterly Journal of the Royal Meteorological Society, 144, 360–383, https://doi.org/https://doi.org/10.1002/qj.3266, 2018.

Gibson, P. B., Chapman, W. E., Altinok, A., Delle Monache, L., DeFlorio, M. J., and Waliser, D. E.: Training machine learning models on climate model output yields skillful interpretable seasonal precipitation forecasts, Communications Earth & Environment, 2, 159, https://doi.org/10.1038/s43247-021-00225-4, 2021.

Hagos, S. M. et al.: Synchronization of the Recent Decline of East African Long Rains and Northwestern Eurasian Cooling, JGR Atmospheres, 129, https://doi.org/10.1029/2023JD039097, 2024.

Hirahara, S., Ishii, M., and Fukuda, Y.: Centennial-Scale Sea Surface Temperature Analysis and Its Uncertainty, Journal of Climate, 27, 57 – 75, https://doi.org/10.1175/JCLI-D-12-00837.1, 2014.

Hotelling, H.: Analysis of a Complex of Statistical Variables into Principal Components, Journal of Educational Psychology, 24, 417–441, https://doi.org/10.1037/h0071325, 1933a.

Hotelling, H.: Analysis of a Complex of Statistical Variables into Principal Components, Journal of Educational Psychology, 24, 498–520, https://doi.org/10.1037/h0070888, 1933b.

James, A. K., Washington, R., and Engelstaedter, S.: Representation of the Indian Ocean Walker Circulation in Climate Models and Links to Kenyan Rainfall, International Journal of Climatology, 40, 1085–1101, https://doi.org/10.1002/joc.6254, 2020.

Kumar, A., Chen, M., and Wang, W.: Understanding Prediction Skill of Seasonal Mean Precipitation over the Tropics, Journal of Climate, 26, 5674–5681, https://doi.org/10.1175/JCLI-D-12-00731.1, 2013.

Landman, W. A., DeWitt, D., Lee, D.-E., Beraki, A., and Lotter, D.: Seasonal Rainfall Prediction Skill over South Africa: One- versus Two-Tiered Forecasting Systems, Weather and Forecasting, 27, 489–501, https://doi.org/10.1175/WAF-D-11-00078.1, 2012.

Lundberg, S. M. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, in: Advances in Neural Information Processing Systems, edited by Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., vol. 30, Curran Associates, Inc., https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf, 2017.

Lyon, B. and DeWitt, D. G.: A recent and abrupt decline in the East African long rains, Geophysical Research Letters, 39, https://doi.org/10.1029/2011GL050337, 2012.

Meenachi, P., Aruna, P., Masilla, S., Thouraya, B., Dhia, B., Monira, K., and Tabrej, K.: Machine Learning-Based Rainfall Prediction for Various Rainfall Densities, Advances in Engineering Research, 9, 145–150, https://doi.org/10.31098/aer.v9i4.1591, 2023.

Milton, S., Diongue-Niang, A., Lamptey, B., Bain, C., Birch, C., and Bougeault, P.: Numerical Weather Prediction over Africa, chap. 10, pp. 380–422, John Wiley & Sons, Ltd, ISBN 9781118391297, https://doi.org/https://doi.org/10.1002/9781118391297.ch10, 2017.

Min, Y.-M., Kryjov, V. N., and Park, C.-K.: A Probabilistic Multimodel Ensemble Approach to Seasonal Prediction, Weather and Forecasting, 24, 812–828, https://doi.org/10.1175/2008WAF2222140.1, 2009.

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

Ogallo, L. J., Janowiak, J. E., and Halpert, M. S.: Teleconnection between Seasonal Rainfall over East Africa and Global Sea Surface Temperature Anomalies, Journal of the Meteorological Society of Japan. Ser. II, 66, 807–822, https://doi.org/10.2151/jmsj1965.66.6_807, 1988.

Omondi, P. et al.: Influence of low frequency teleconnections on East African rainfall, International Journal of Climatology, 33, 1250–1265, 2013.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


Palmer, P. I., Wainwright, C. M., Dong, B., et al.: Drivers and impacts of Eastern African rainfall variability, Nature Reviews Earth and Environment, 4, 254–270, https://doi.org/10.1038/s43017-023-00397-x, 2023.

Pinheiro, E. and Ouarda, T. B. M. J.: An interpretable machine learning model for seasonal precipitation forecasting, Communications Earth & Environment, 6, 222, https://doi.org/10.1038/s43247-025-02207-2, 2025.

Ran, G., Meng, J., and Fan, J.: Tropical monsoon rainfall can be predicted with lead times up to 10 months, Communications Earth & Environment, 6, 417, https://doi.org/10.1038/s43247-025-02391-1, 2025.

Rowell, D. P.: An Observational Constraint on CMIP5 Projections of the East African Long Rains and Southern Indian Ocean Warming, Geophysical Research Letters, 46, 6050–6058, https://doi.org/10.1029/2019GL082847, 2019.

Saji, N. H., Goswami, B. N., Vinayachandran, P. N., and Yamagata, T.: A dipole mode in the tropical Indian Ocean, Nature, 401, 360–363, 1999.

Scalabrini, E. C. B. and Remoaldo, P. C.: Principal Component Analysis, pp. 823–824, Springer Nature Switzerland, Cham, ISBN 978-3-030-74923-1, https://doi.org/10.1007/978-3-030-74923-1_580, 2024.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., and Rudolf, B.: Global Precipitation Climatology Centre (GPCC), First Guess Monthly Product Version 2018 at 1.0°: Near Real-Time Monthly Land-Surface Precipitation from Rain Gauges Built on GTS-Based Data, https://doi.org/10.5676/DWD_GPCC/FG_M_100, accessed from GPCC First Guess Monthly dataset, 2018.

Schneider, U., Fuchs, T., Meyer-Christoffer, A., and Rudolf, B.: Global Precipitation Climatology Centre (GPCC), Full Data Monthly Product Version 2020 at 0.25°: Monthly Land-Surface Precipitation from Rain Gauges Built on GTS-Based and Historical Data, https://doi.org/10.5676/DWD_GPCC/FD_M_V2020_025, accessed from GPCC dataset from 1981 to 2019 at 0.25° resolution, 2020.

Tierney, J. E., Ummenhofer, C. C., and deMenocal, P. B.: Past and future rainfall in the Horn of Africa, Science Advances, 1, e1500 682, https://doi.org/10.1126/sciadv.1500682, 2015.

Turner, A. G. and Annamalai, H.: Climate change and the South Asian summer monsoon, https://doi.org/10.1038/nclimate1495, 2012.

Uvo, C. B., Repelli, C. A., Zebiak, S. E., and Kushnir, Y.: The Relationships between Tropical Pacific and Atlantic SST and Northeast Brazil Monthly Precipitation, Journal of Climate, 11, 551–562, https://doi.org/10.1175/1520-0442(1998)011<0551:TRBTPA>2.0.CO;2, 1998.

Wainwright, C. M., Marsham, J. H., Keane, R. J., Rowell, D. P., Finney, D. L., Black, E., and Allan, R. P.: 'Eastern African Paradox' rainfall decline due to shorter not less intense Long Rains, npj Climate and Atmospheric Science, 2, 34, https://doi.org/10.1038/s41612-019-0091-7, 2019.

Ward, P. J. et al.: Predictability of the East Africa long rains through Congo zonal winds, Atmospheric Science Letters, 24, https://doi.org/10.1002/asl.1124, 2023.

Weyn, J. A., Durran, D. R., Caruana, R., and Cresswell-Clay, N.: Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models, Journal of Advances in Modeling Earth Systems, 13, e2021MS002 502, https://doi.org/10.1029/2021MS002502, 2021.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, vol. 91 of *International Geophysics Series*, Academic Press, Amsterdam, 2 edn., ISBN 978-0127519654, https://doi.org/10.1002/met.16, 2006.

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, Elsevier, Amsterdam, 4 edn., 2020.

Williams, A. P. and Funk, C. C.: A westward extension of the warm pool leads to a westward extension of the Walker circulation, drying eastern Africa, Climate Dynamics, 37, 2417–2435, https://doi.org/10.1007/s00382-010-0984-y, 2011.

Zhang, M., Ye, H., Xie, T., and Zhang, Y.: Machine Learning for Precipitation Forecasts Postprocessing, ArXiv e-prints, https://arxiv.org/abs/2103.13361, 2021.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


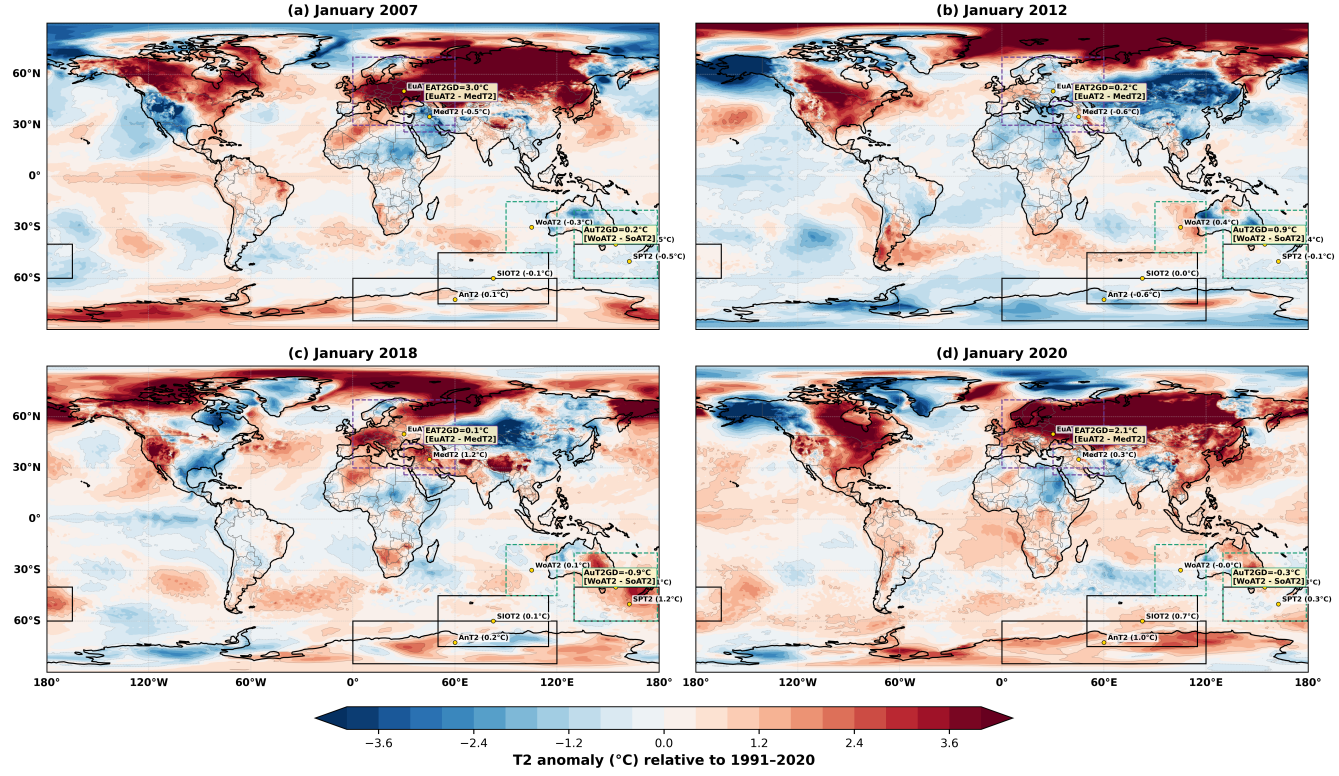


**Figure A1.** January 2-m air temperature (T2) anomalies for four representative years: (a) 2007, (b) 2012, (c) 2018, and (d) 2020, relative to the 1991–2020 climatology. Shading denotes T2 anomalies (°C), with warm colours indicating positive anomalies and cool colours indicating negative anomalies. Black and dashed boxes mark the retained T2 predictor domains, while labels indicate the corresponding basin-mean January anomalies and the derived temperature-gradient indices (EAT2GD and AuT2GD). The maps highlight the antecedent thermal conditions linked to contrasting East Africa MAM rainfall seasons.

## Appendix A: Additional Model Diagnostics

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


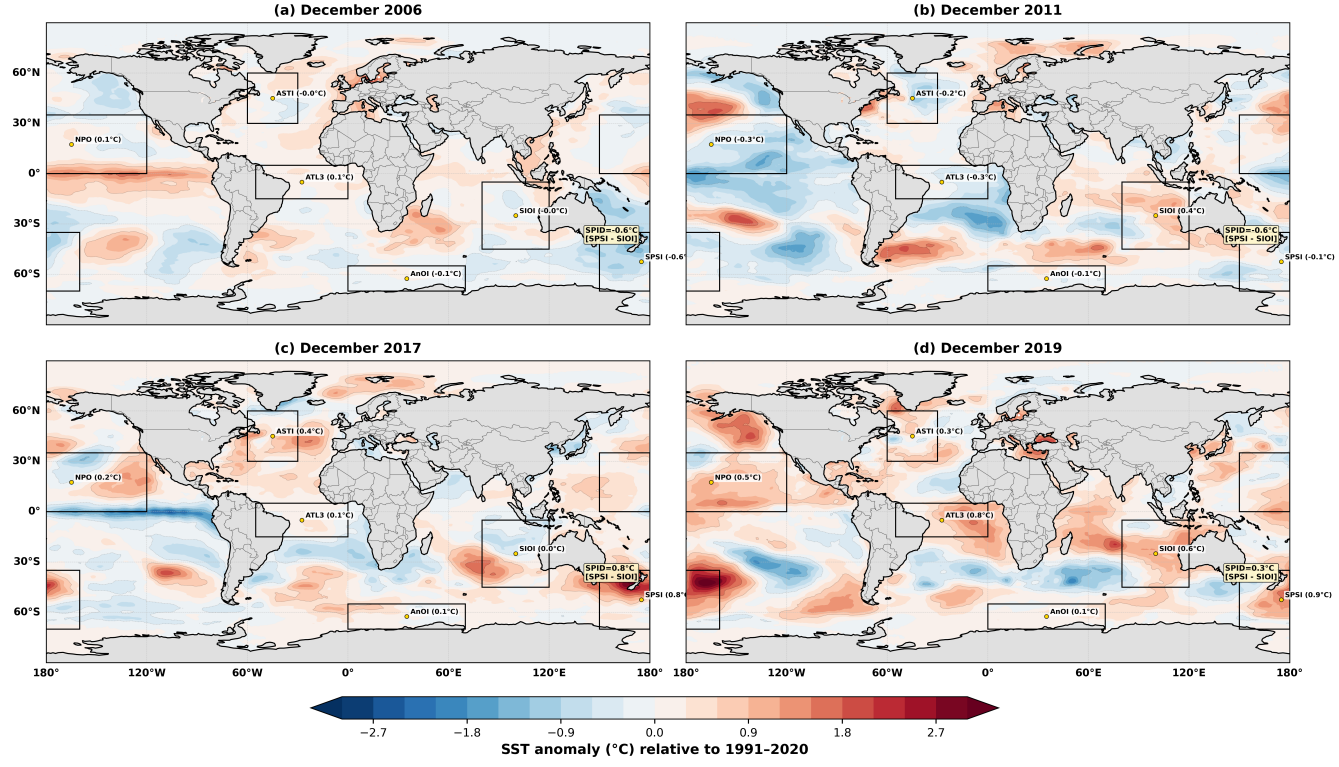


**Figure A2.** As in Fig. 9, but using December initialization.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


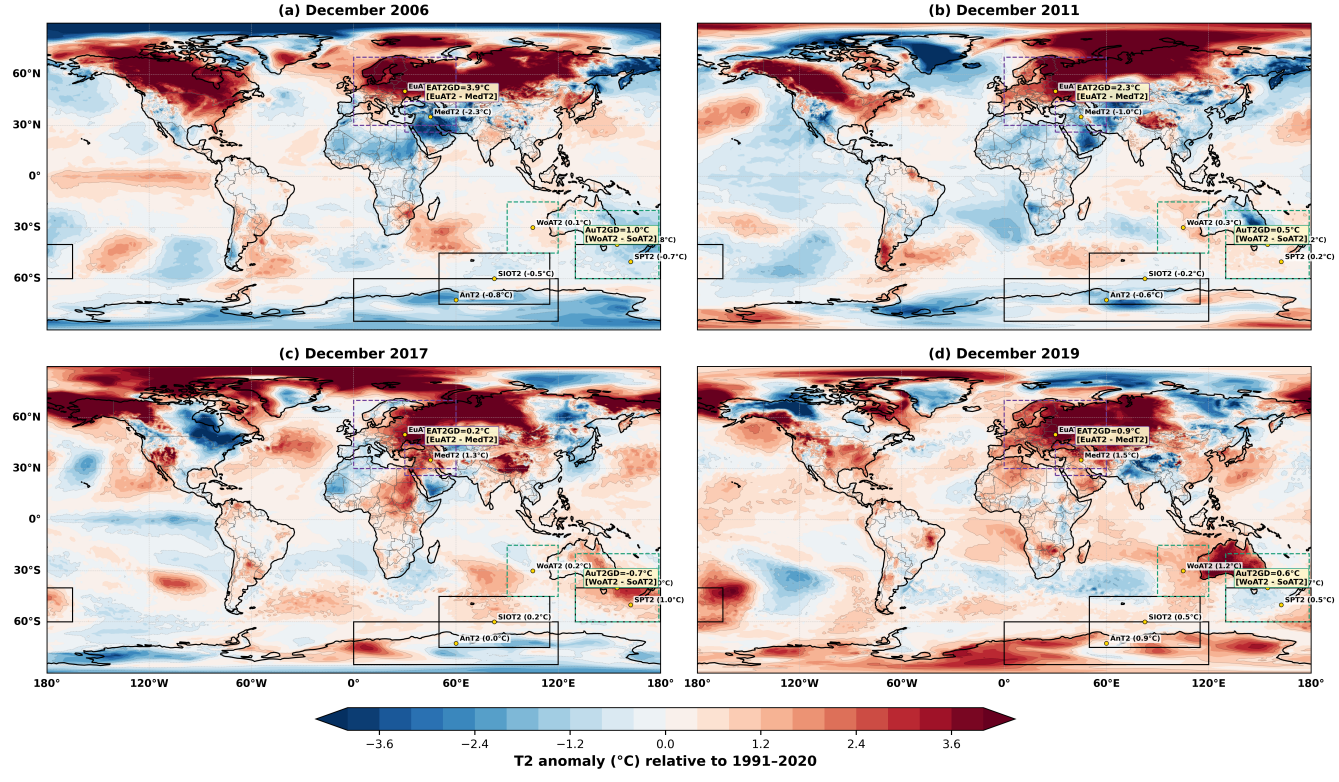


**Figure A3.** As in Fig. A1, but using December initialization.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


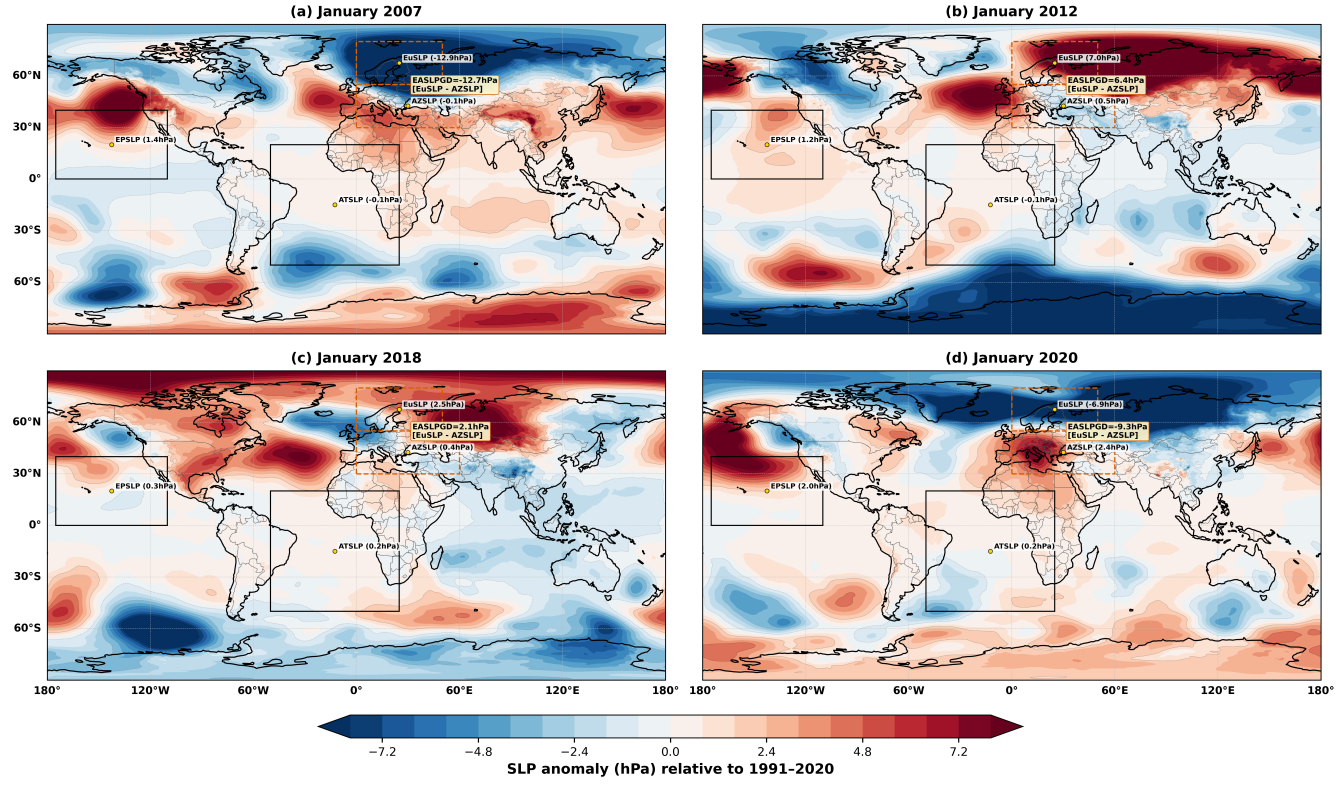


**Figure A4.** January sea level pressure (SLP) anomalies for four representative years: (a) 2007, (b) 2012, (c) 2018, and (d) 2020, relative to the 1991–2020 climatology. Shading denotes SLP anomalies (°C), with warm colours indicating positive anomalies and cool colours indicating negative anomalies. Black and dashed boxes mark the retained SLP predictor domains, while labels indicate the corresponding basin-mean January anomalies and the derived pressure-gradient index EASLPGD. The maps highlight the antecedent thermal conditions linked to contrasting East Africa MAM rainfall seasons.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


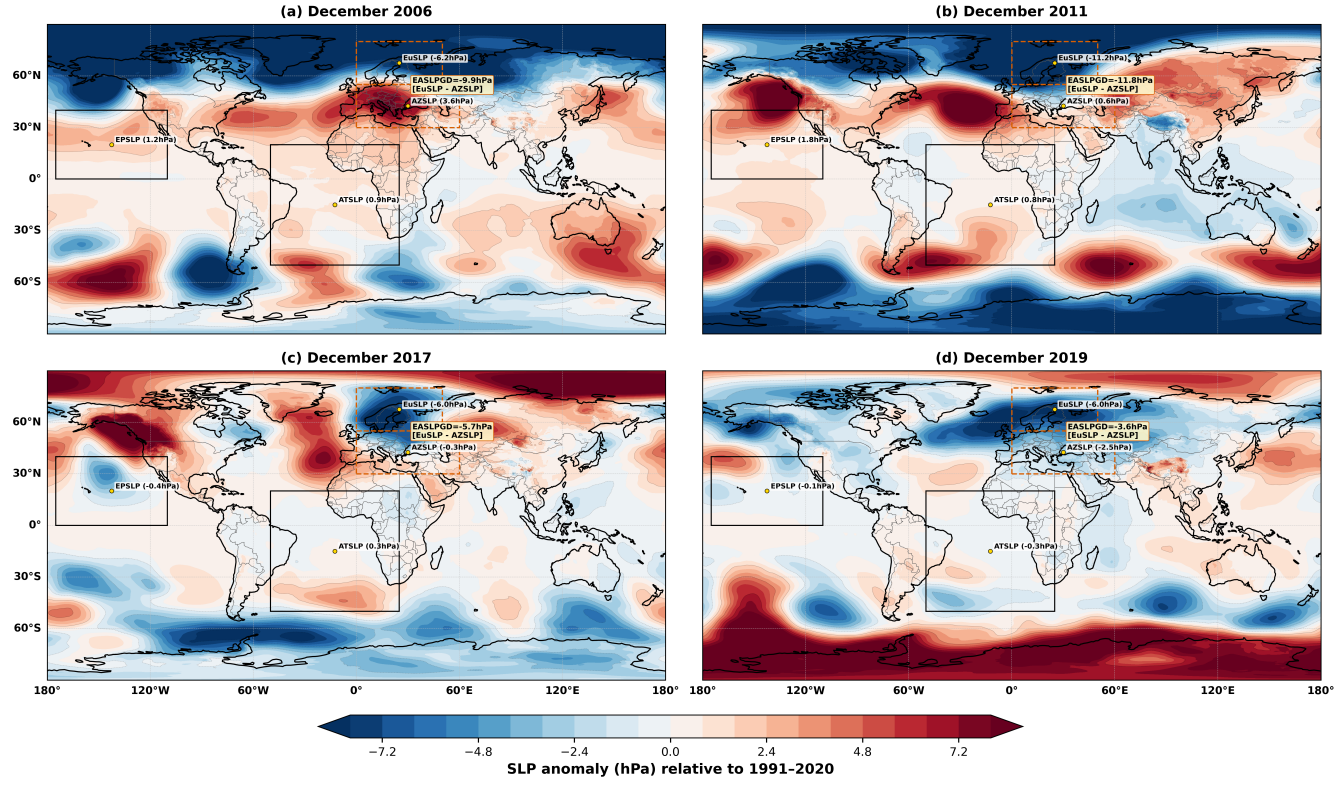


**Figure A5.** January sea level pressure (SLP) anomalies for four representative years: (a) 2006, (b) 2011, (c) 2017, and (d) 2019, relative to the 1991–2020 climatology. Shading denotes SLP anomalies (°C), with warm colours indicating positive anomalies and cool colours indicating negative anomalies. Black and dashed boxes mark the retained SLP predictor domains, while labels indicate the corresponding basin-mean January anomalies and the derived pressure-gradient index EASLPGD. The maps highlight the antecedent thermal conditions linked to contrasting East Africa MAM rainfall seasons.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


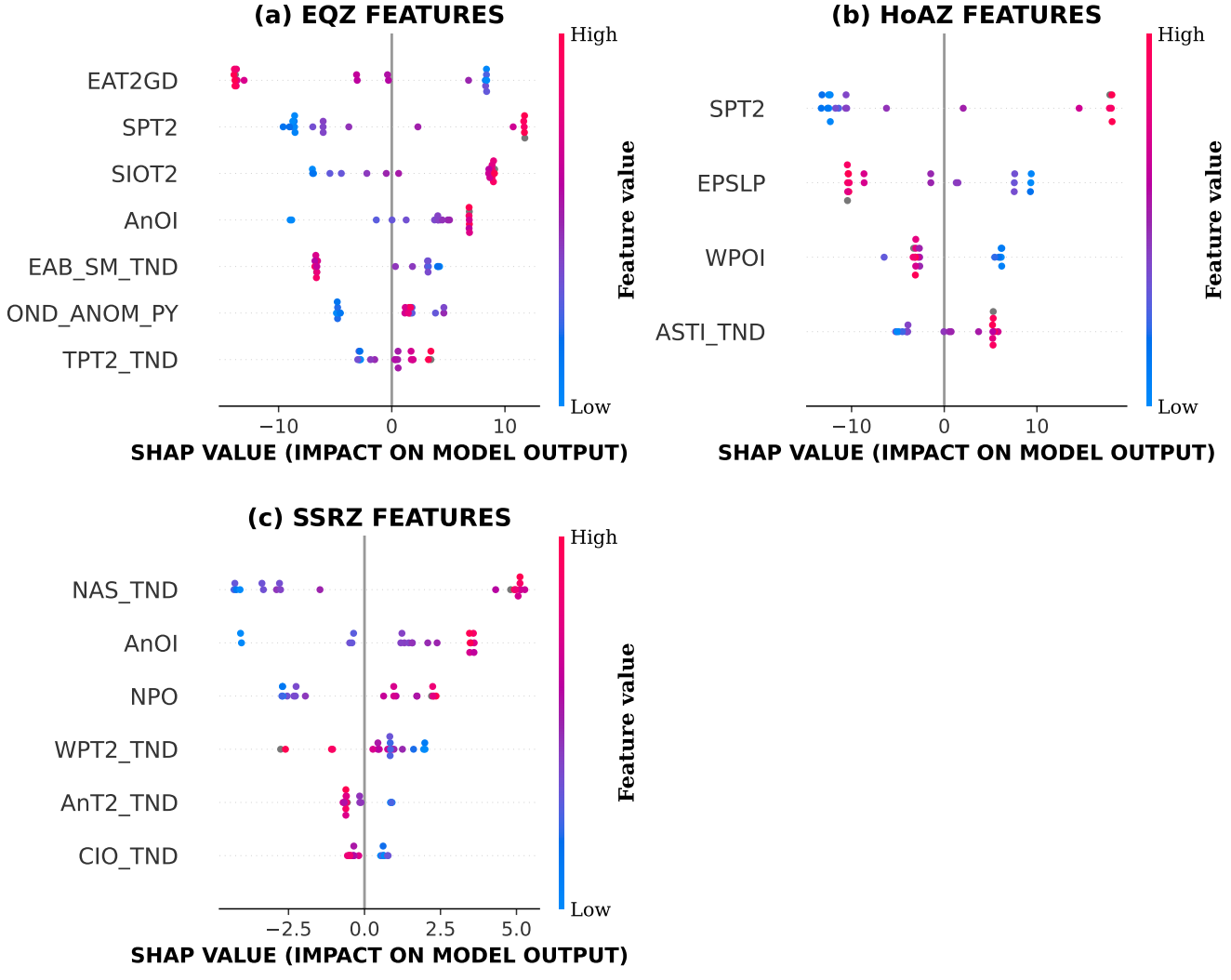


**Figure A6.** As in Fig. 6, but using December initialization.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


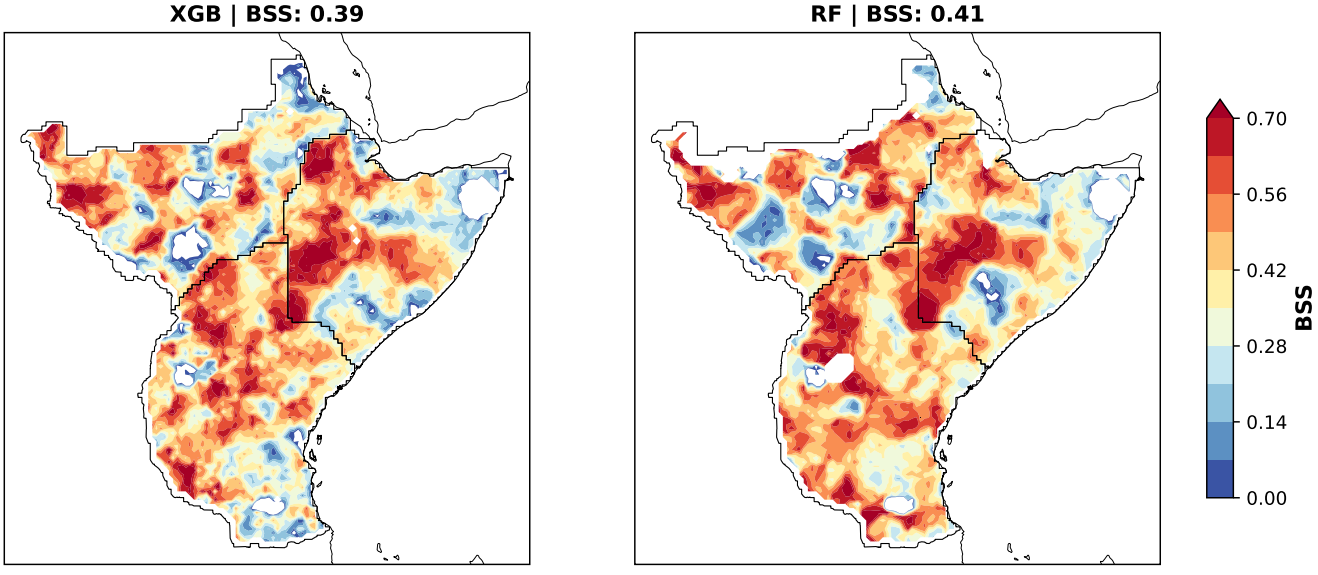


**Figure A7.** As in Fig. 11, but using December initialization.

https://doi.org/10.5194/egusphere-2026-2832
Preprint. Discussion started: 24 June 2026


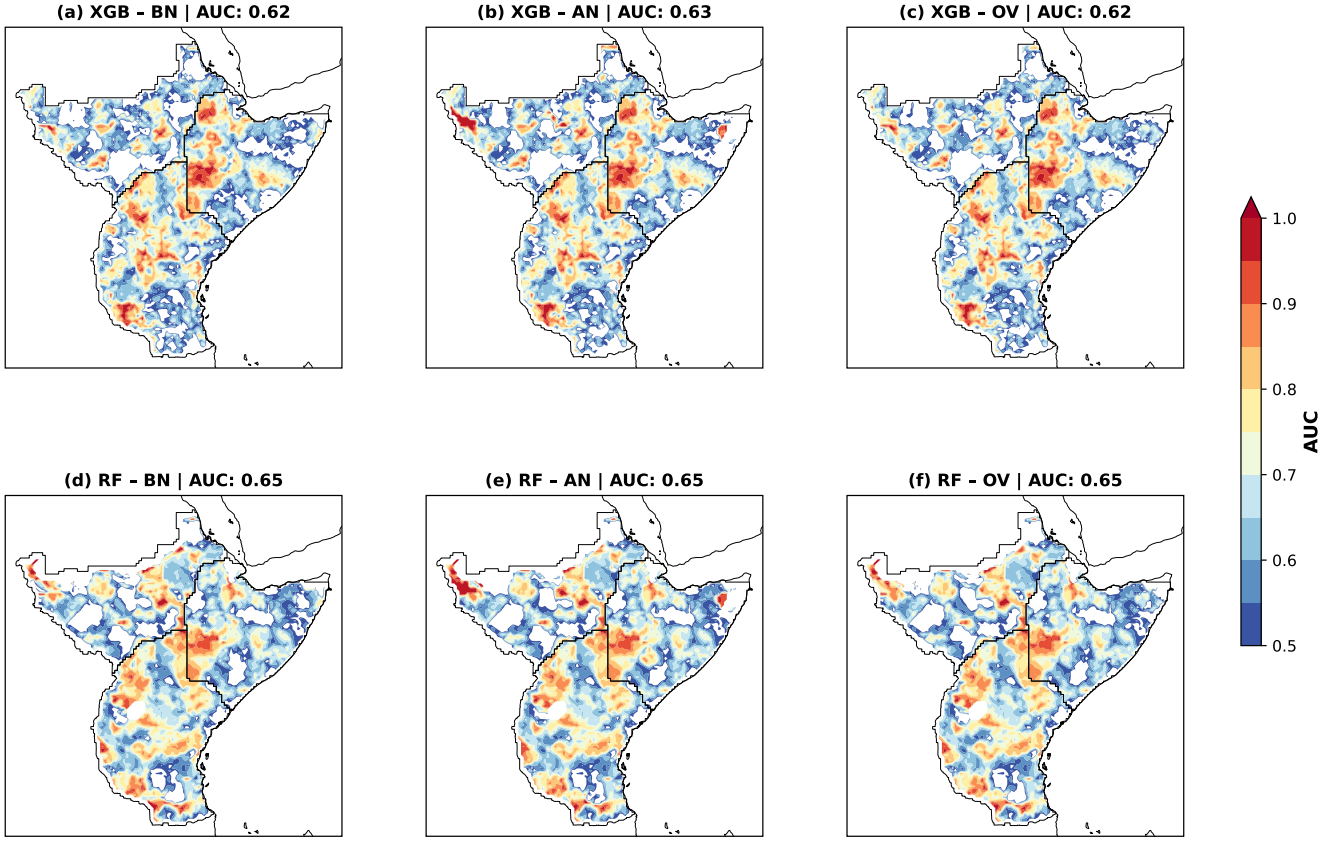


**Figure A8.** As in Fig. 12, but using December initialization.