International Journal of Climatology

International Journal of Climatology RMetS

RESEARCH ARTICLE OPEN ACCESS

Interannual Rainfall Variability in West Africa: Reconstruction Based on Atmospheric Circulation Patterns

Manuel Rauch[1] | Jan Bliefernicht[1] | Marlon Maranan[2] | Andreas H. Fink[2] | Harald Kunstmann[1,2]

[1]Institute of Geography, University of Augsburg, Augsburg, Germany | [2]Institute of Meteorology and Climate Research, Karlsruhe Institute of Technology, Karlsruhe, Germany

Correspondence: Manuel Rauch (manuel.rauch@geo.uni-augsburg.de)

Received: 23 July 2024 | **Revised:** 2 July 2025 | **Accepted:** 13 August 2025

Funding: This work was supported by FURIFLOOD (Current and Future Risks of Urban and Rural Flooding in West Africa) project, which is funded by the Federal Ministry of Education and Research in Germany [grant number: 01LG2086D]. We would like to thank our partners involved in the FURIFLOOD project for their support.

Keywords: circulation patterns | k-means | logistic regression | rainfall | West Africa

ABSTRACT

The West African Monsoon, known for its significant rainfall variability, led to the Sahel drought from 1968 to the 1990s, followed by a recovery in rainfall since the 1990s. In response to such variability, this study introduces a statistical approach for reconstructing interannual rainfall variability across seven rainfall regimes, each representing unique climatic zones in West Africa. Initially, a robust catalogue of daily atmospheric circulation pattern classifications over West Africa is established, based on pre-selected variables from the ERA5 reanalysis and using k-means clustering. Subsequently, the annual occurrence frequencies of these circulation pattern classifications, along with the annual rainfall conditions in the rainfall regimes, serve as inputs in a multi-class logistic regression model. This model is designed to identify dry, normal, and wet years, relative to the climatology. The rainfall regimes are determined using k-means clustering and a quality-controlled dataset from 971 rainfall stations, with daily observations ranging from 1959 to 2010. These regimes vary from the Sahelian belt, characterised by a short rainy season, to tropical regions exhibiting a bimodal rainfall regime. After comprehensive predictor screening of specific West African Monsoon patterns, such as the Tropical Easterly Jet and the African Easterly Jet, six variables at four different pressure levels under a running split-sampling cross-validation, the best models achieve an average proportion correct of 0.57 and a positive Peirce skill score for all regions over West Africa. This shows the performance in reconstructing dry, normal, and wet years for the different rainfall regimes in West Africa. Therefore, this study provides a statistical tool for the reconstruction of annual rainfall anomalies in this challenging region.

1 | Introduction

The West African Monsoon (WAM) is one of the world's most complex regional climate systems, affecting key surface fluxes including rainfall, across a broad range of spatial and temporal scales (Nicholson 2001, 2009; Lafore et al. 2011; Fink et al. 2017; Nicholson et al. 2018). Rainfall-related extremes and disasters in this region, such as the recent Nigeria floods in 2022 (Tunde 2022) or the Sahel droughts of the 1970s and 1980s (Hagos and Cook 2008), highlight the substantial vulnerabilities of both urban and rural communities (Paeth et al. 2011; Epule et al. 2013; Engel et al. 2017; Maranan et al. 2019; Atiah et al. 2023). To mitigate the impacts of natural disasters, accurate local-scale rainfall information is crucial for providing reliable warnings of seasonal rainfall anomalies, extreme events and for numerous other applications. The existing methods employed by national weather services in West Africa exhibit various limitations,

for example, in the domain of seasonal rainfall forecasting (Bliefernicht et al. 2019). Moreover, state-of-the-art general circulation models (GCMs) fail to provide accurate local-scale rainfall information for West Africa, as shown by Vogel et al. (2018) in their assessment of global numerical weather prediction models. In contrast, GCMs demonstrate good performances in capturing regional-scale atmospheric phenomena such as the West African heat low or the African Easterly Waves (AEWs) in this challenging region (Ngoungue Langue et al. 2021; Bliefernicht, Rauch, et al. 2022).

To bridge this disparity, innovative statistical downscaling methods and strategies are essential (Maraun et al. 2010; Maraun and Widmann 2018). Statistical downscaling frameworks are computationally efficient compared to dynamical downscaling (e.g., Siegmund et al. 2015), making them a practical choice for regional studies, particularly in developing countries. Moreover, they can be tailored to address specific research questions or integrated with other techniques to enhance performance (Rust et al. 2013). However, the focus on statistical relationships may not fully capture the complexity of physical climate processes. Additionally, while statistical downscaling is effective at explaining certain variability patterns, it can struggle with accounting for longer-term trends or handling non-stationarities in atmospheric circulation patterns (Maher et al. 2020). For the context of this study, statistical techniques that use circulation patterns (CPs) have emerged as a promising approach, as shown by their diverse applications across various regions of the world over recent decades (Moron et al. 2008b; Rust et al. 2013; Batté et al. 2018; Böker et al. 2023). In this study, the approach is tested for the annual areal rainfall amount of rainfall zones whose spatial extent is similar to the scales often used in climate predictions or seasonal forecasts (Bliefernicht et al. 2019). Although the current approach does not downscale the GCM output in the sense of providing higher spatiotemporal information, the statistical methods used here are based on the downscaling procedures.

For the West African region, CP-based methods are especially promising because large-scale circulation features like the African Easterly Jet, Tropical Easterly Jet or the West African heat low play a central role in modulating rainfall variability (Lafore et al. 2011; Fink et al. 2017; Lemburg et al. 2019; Bliefernicht, Rauch, et al. 2022). While GCMs may represent these features at a coarse scale, they often fail to accurately reproduce their regional structure or interactions with mesoscale convective systems, which are one of the primary rainfall drivers in the region (Mathon et al. 2002; Nkrumah et al. 2023; Semunegus et al. 2017; Atiah et al. 2023; Diedhiou et al. 2024). CP-based methods, in contrast, offer a way to statistically link these key circulation modes to rainfall anomalies without requiring full physical simulation of convection. This is particularly beneficial in West Africa, where sparse observational data limits both the initialization of dynamical models through data assimilation and the post-processing calibration (e.g., bias correction) of their output.

Despite the evident potential of CP-based statistical techniques, its application in monsoon regions has been less explored. In the WAM region, relatively few studies have explored the relationship between CP classifications and local meteorological variables. Moron et al. (2008a) employed a k-means cluster analysis to classify regional circulation in the Western Sahel, focusing on its influence on rainfall in Senegal. This classification considered three wind levels (200, 700, and 925 hPa) during the months of July to September. Subsequently, these CPs were utilised in a downscaling method that performed well in capturing the interannual variability of rainfall, including the frequency and mean duration of dry and wet spells (Moron et al. 2008b). Guèye et al. (2011) identified nine CPs to account for Senegal's daily atmospheric circulation variability, primarily focusing on AEWs and the Saharan heat low (SHL). Camberlin et al. (2020) determined six types of intense rainfall events in southern West Africa, noting westward signals at the 700 hPa wind field. Moron et al. (2018) employed cluster analysis to assess CPs' relevance for temperature fluctuations in tropical North Africa. Batté et al. (2018) used CPs to predict West African Heat Waves at both subseasonal and seasonal timescales. They proposed that these CP classifications might be integrated into a hybrid statistical-dynamical forecasting method. Instead of using direct model outputs for temperature, this method uses the ECMWF's SEAS5 forecasted CP-frequency anomalies to generate temperature anomaly forecasts. However, their findings also indicated that this technique needs further improvement. In contrast, Rust et al. (2013) demonstrated that when large-scale CPs are incorporated into their statistical regression model, there is an enhancement in the predictive ability for both the occurrence and amount of rainfall in Senegal. Bliefernicht, Rauch, et al. (2022) introduced a two-step atmospheric CP classification for West Africa, focusing on both seasonal and daily rainfall variability in the Sudan-Sahel zone (Burkina Faso). As predictor variables, they used mean sea level pressure for the seasonal classification and the stream function fields at 700 hPa to better account for daily rainfall variability. Recently, Nkrumah et al. (2023) identified nine CPs over West Africa and investigated changes associated with favourable environments for mesoscale convective systems using self-organising maps (SOMs) based on the 925 hPa geopotential height. While the aforementioned studies have primarily addressed CP classifications and their effects on daily variability, this study builds upon the work of Bliefernicht, Rauch, et al. (2022) and specifically targets CPs indicative of interannual variability.

Furthermore, the availability of in situ observations remains a persistent challenge (Salack et al. 2019; Bliefernicht, Salack, et al. 2022). Such data are crucial for establishing the relationships between regional-scale atmospheric patterns (predictors) and local weather variables (predictands). Despite significant progress in the development of improved meteorological networks through various initiatives (van de Giesen et al. 2014; Galle et al. 2018; Bliefernicht et al. 2018; Schunke et al. 2021; Sawadogo et al. 2023), the scarcity of long-term in situ observations continues to inhibit the development of reliable statistical approaches in the region. In response to these challenges, our study proposes several key contributions to enhance the understanding of interannual rainfall variability in West Africa through the development of a CP-based statistical adaptation approach. Therefore, our research is directed towards achieving three key objectives:

- Establishment of an improved rainfall dataset for the West Africa region based on long-term in situ observations.
- Development of historical catalogues of daily CP classifications using pre-selected atmospheric variables for different climate zones.
- Introduction of a straightforward method that uses daily CPs to reconstruct interannual rainfall anomalies under ideal reanalysis conditions similar to regional climate outlook forums.

The focus on interannual rainfall variability aligns with the seasonal forecasting approaches of regional climate outlook forums (Semazzi 2011) like *Prévisions Climatiques Saisonnières en Afrique Soudano-Sahélienne* (PRESASS). Seasonal rainfall amounts are categorised into terciles (below-average, near-average, and above-average) relative to the climatology (e.g., 1981–2010) and simplify complex climate data into actionable insights for stakeholders through maps (see Bliefernicht et al. (2019) and Pirret et al. (2020) for more information). However, these forecasts face several limitations (Mason and Chidzambwa 2009; Bliefernicht et al. 2019; Pirret et al. 2020), and statistical downscaling models that support this process are relatively rare in West Africa (e.g., Ndiaye et al. 2011; Manzanas 2017). Our CP-based approach also classifies rainfall into terciles using a probabilistic multi-class logistic regression, thus providing outputs that are structurally compatible with PRESASS-style products. While the current study focuses on reconstruction of the interannual variability of West Africa, a complementary study evaluates the forecasting potential of this framework in greater detail (Rauch et al. 2025), showing that CP-based models may enhance the skill and transparency of operational regional forecasts in data-scarce regions.

The article is structured as follows: Section 2 provides an overview of the study area and rainfall data with a particular focus on the ERA-5 dataset and the extensive network of rainfall stations across West Africa. Section 3 outlines the methodological workflow. This includes an explanation of the clustering method, the multi-class logistic regression, and the validation of the latter. Section 4 presents the results, including a special focus on the Sahelian belt. Section 5 offers a discussion of the results, and Section 6 provides a key summary.

2 | Study Area and Data

2.1 | Study Area

The focus of this study is on West Africa, with special emphasis on the Sahelian belt—a zone dependent on the dynamics of the WAM. The WAM is characterised by a complex interplay of atmospheric, oceanic, and terrestrial forces. A strong influence on the monsoon's northward progression during the boreal summer is the thermal gradient established between the landmass of the African continent and the relatively cooler Atlantic and Mediterranean Seas. As the land heats up due to seasonally driven insolation changes, the associated temperature differences amplify, leading to a northward shift of the low-level pressure gradient. This shift plays a central role in the seasonal rainfall distribution and the overall seasonal cycle in West Africa.

Central atmospheric features in the WAM region include the SHL, the intertropical discontinuity (ITD), the AEJ, the TEJ, and the AEWs (Lafore et al. 2010; Nicholson 2009; Fink et al. 2017). These features, each with their unique impact on the WAM, play crucial roles in determining rainfall patterns. For instance, Sheen et al. (2017) point out the significance of certain elements, such as the convergence of southern monsoon winds with northern Mediterranean winds, in enhancing rainfall. Enhanced rainfall is also associated with a more potent TEJ positioned further north and a strengthened AEJ (Grist and Nicholson 2001). The West African Westerly Jet (WAWJ), a low-level westerly jet located between 8° N and 11° N over the West African coast and the Atlantic, plays a pivotal role in moisture transport into the Sahel (Pu and Cook 2010). Thorncroft et al. (2011) emphasise the SHL and the Atlantic equatorial cold tongue's role in moisture fluxes and subsequent rainfall. Moreover, the variability in West African summer rainfall is heavily influenced by the equatorial Atlantic sea surface temperature (Fontaine and Bigot 1993; Fontaine and Janicot 1996; Vizy and Cook 2002; Reason and Rouault 2006; Losada et al. 2010).

In essence, the WAM is a complex system characterised by interactions between zonal features such as the WAWJ, AEJ, and TEJ, and meridional elements like southerly monsoon winds, northerly Mediterranean winds, and convergence zones. The strength, positioning, and interplay of these WAM features are pivotal in determining rainfall quantity and its variability across the region. Given the evident relationships between these regional-scale atmospheric processes and meteorological variables, there is a strong motivation for their use in statistical adaptation of model outputs. Harnessing these relationships allows us to infer local-scale phenomena from larger-scale atmospheric conditions. Therefore, this study leverages the fact that many of the key atmospheric components of the WAM shape both daily and seasonal rainfall patterns across the region. By identifying and classifying CPs that reflect the combined states of these atmospheric elements, we aim to capture the variability in their structure and location from year to year. This approach represents an interpretable and computationally efficient tool which enables us to statistically link large-scale circulation modes with observed rainfall anomalies, thus directly addressing our goal of reconstructing interannual rainfall variability over West Africa.

2.2 | Rainfall Data

The in situ daily rainfall records were sourced from two databases: the KASS-D (Karlsruhe Surface Station Database, indicated in red in Figure 1, 1810 stations) and a daily subset of the WAHPD (West African Historical Precipitation Database, indicated in blue in Figure 1, 410 stations). The WAHPD is described in Bliefernicht, Salack, et al. (2022), providing information on data sources and the quality algorithms applied. Furthermore, these databases have been used in various other studies, for example, Bliefernicht, Rauch, et al. (2022), Ascott et al. (2020) for WAHPD, and Schlueter et al. (2019), Vogel et al. (2018) for

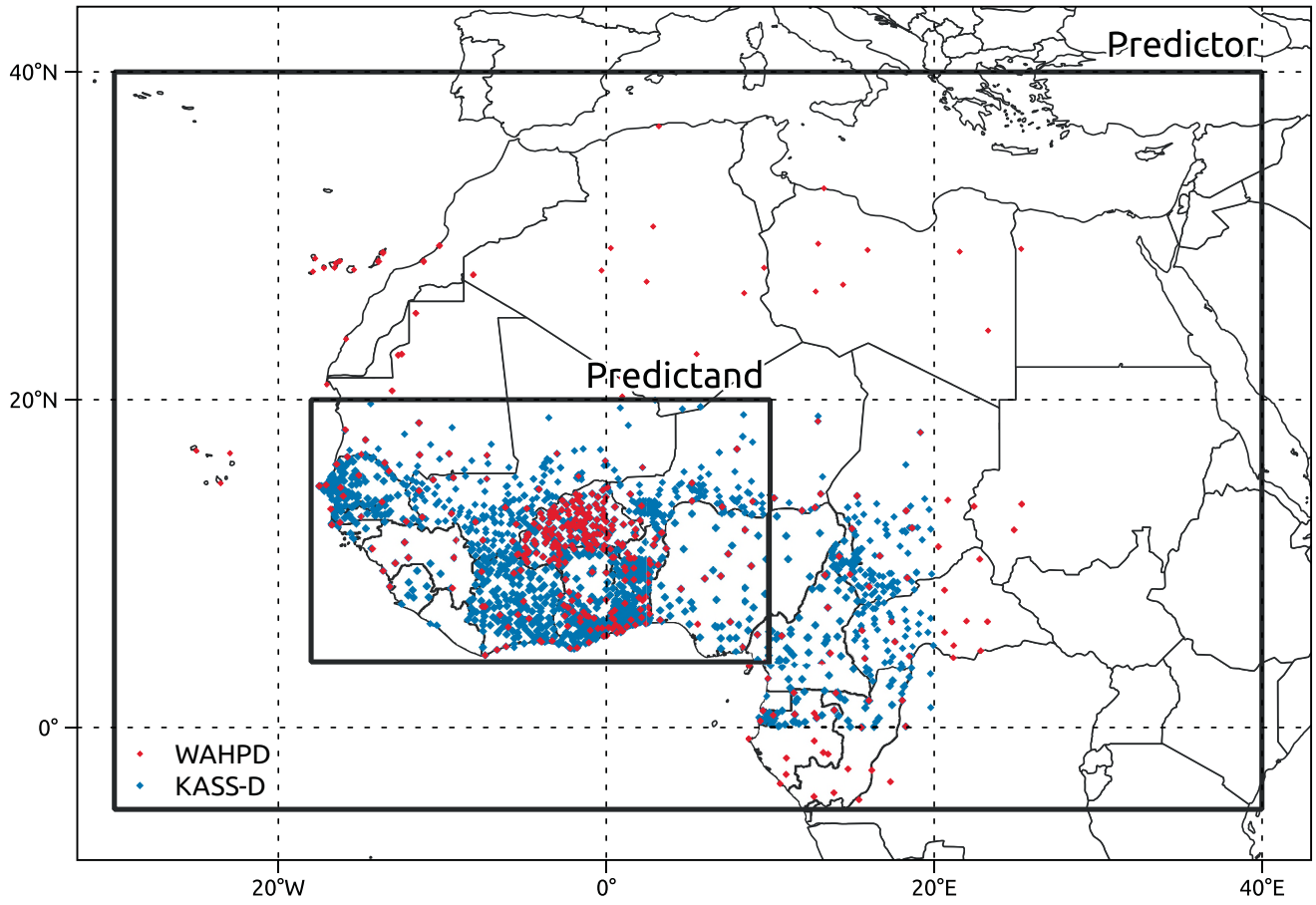

FIGURE 1 | Distribution of the rainfall stations across the study area in West Africa, with the locations of WAHPD (West African Historical Precipitation Database, red) stations and KASS-D (Karlsruhe Surface Station Database, blue) stations. The box labelled 'Predictand' denotes the target region for the chosen rainfall data, whereas the broader area labelled 'Predictor' represents the target region for the reanalysis data (ERA-5). [Colour figure can be viewed at wileyonlinelibrary.com]

KASS-D. Both databases have overlaps and include data from national hydrological and meteorological services but also draw on regional repositories (e.g., AMMA-Catch, Galle et al. 2018) and global networks (e.g., GHCNd). Although both the KASS-D and WAHPD databases include partially overlapping station data, their use is complementary and necessary. Each database is a curated collection compiled from various national, regional, and international sources, often following different metadata conventions, quality control procedures, and formatting standards. To make both usable in a unified analysis, a harmonisation step was applied to align station metadata, formats, and quality filters. This integrated approach enables us to construct a more comprehensive and spatially balanced rainfall dataset for West Africa, which is crucial for robust CP classification and interannual variability analysis.

In order to cover a significant portion of Africa north of the equator, the data selection focuses on the region between 18° W and 10° E longitude and 4° N and 20° N latitude (see Predictand in Figure 1, WAHPD: 299 stations, KASS-D: 1467 stations) and spans an overall period from 1959 to 2010. To ensure an adequate data sample for meaningful analysis, time series are required to have a minimum of 5 years of data (WAHPD: 276 stations, KASS-D: 1099 stations). A five-year threshold was chosen as a balance between ensuring sufficient temporal coverage to capture basic interannual variability and retaining enough stations for robust spatial representation.

2.3 | ERA-5 Reanalysis

The ERA-5 dataset, developed by the European Centre for Medium-Range Weather Forecasts (ECMWF), represents their most recent reanalysis dataset. It provides comprehensive insights into atmospheric, terrestrial, and oceanic climate variables (Hersbach et al. 2020). For the CP-based regression analysis, ERA-5 serves as a primary source of predictor data. For the analysis, ERA-5 hourly data from 1959 to 2010 was obtained from the Climate Data Store of the Copernicus Climate Change Service. This data was then averaged from an hourly scale to provide daily-scale predictor information. To further optimise computational efficiency, the ERA-5 reanalysis fields were pre-processed to a resolution of $2^{\circ} \times 2^{\circ}$ which is widely used to reduce computational load during clustering, particularly for variables like mean sea level pressure or high-level wind fields that exhibit low spatial variability (Bliefernicht, Rauch, et al. 2022). In total, this study uses six individual variables (stream function, SF; U-component of wind, U; V-component of wind, V; wind direction, WD; wind speed, WS; and mean sea level pressure, MSLP) on four different pressure levels (200, 700, 850, 925 hPa) from the ERA-5

TABLE 1 | Pressure-level variables: Listing of abbreviations, associated atmospheric pressures, units, and input variables for computation. The designation 'all levels' includes atmospheric pressures at 200, 700, 850, and 925 hPa. Further information on each variable can be found in Hersbach et al. (2023a, 2023b).

Name	Abbreviation	Levels	Units	Input
Pressure-level variables				
Stream function	SF	All levels	$m^2 s^{-1}$	U, V
U-component of wind	U	All levels	$m s^{-1}$	—
V-component of wind	V	All levels	$m s^{-1}$	—
Wind direction	WD	All levels	degrees	U, V
Wind speed	WS	All levels	$m s^{-1}$	U, V
Single-level variables				
Mean sea level pressure	MSLP	—	Pa	—
Total precipitation	TP	—	mm	—

dataset. Detailed information about the variables can be found in Table 1. The pre-selection of predictor variables in this study is based on existing literature, aligning with the studies by Moron et al. (2008a) (U/V200, U/V700, U/V925), Guèye et al. (2011, 2012) (U/V850), and Bliefernicht, Rauch, et al. (2022) (U/V at multiple levels, SF700, SF850). MSLP is also commonly employed (Guèye et al. 2011, 2012; Bliefernicht, Rauch, et al. 2022). These variables were selected not only based on precedent in the literature, but also due to their physical relevance for capturing the dynamics of the WAM system. For instance, the U and V wind components at 200 and 700 hPa reflect the influence and positioning of the TEJ and AEJ. The 925 and 850 hPa levels capture low-level monsoon inflow and moisture transport from the Gulf of Guinea. SF at 700 and 850 hPa provides information on rotational flow and large-scale vorticity, aiding in identifying circulation features like AEWs. MSLP is closely linked to the position and intensity of the SHL. Additionally, every possible combination of these variables up to four is considered, leading to $C(21,2) + C(21,3) + C(21,4) = 7525$ combinations. As a result, a total of 7546 initial predictor combinations are assessed.

To describe dominant regional-scale features of the WAM (e.g., SHL, TEJ or AEJ), we use a domain spanning from 30° W to 40° E and 5° S to 40° N (see predictor in Figure 1). This coverage encompasses a majority of the regions in Africa located north of the equator, aligning precisely with the domain size used by Bliefernicht, Rauch, et al. (2022). To prepare the data for modelling, a data scaling preprocessing step was executed. This step involved the normalisation of the variables' range. Specifically, standardised anomalies, represented as $z(i,t)$, were derived from the predictor data $x(i,t)$ according to:

$$z(i,t) = \frac{x(i,t) - \overline{x}(i)}{s(i)} \quad (1)$$

In this equation, $i = 1, \ldots, L$ pertains to the grid points, $t = 1, \ldots T$ denotes the time steps in days, $\overline{x}(i)$ signifies the time-inclusive long-term mean, and $s(i)$ indicates the standard deviation of the time series. This means standardised unfiltered (i.e., the annual cycle is kept) daily anomalies are used for the analysis.

In addition to the atmospheric data, the variable of total precipitation from ERA-5 at the original resolution of $0.25^\circ \times 0.25^\circ$ has been used to determine the value-added of the proposed CP-based logistic regression model. No further preprocessing steps were applied.

3 | Methods

To capture the relationships between selected predictors and desired outcomes, a multi-data preparation pipeline was designed. As shown in Figure 2, the methodology is split into two branches: one to process atmospheric predictor data and the other focusing on the different RRs over the West African continent. For the atmospheric data, the following steps were undertaken to prepare the predictors for modelling:

1. Standardisation: The initial step involved transforming the predictors into anomaly-based variables, as detailed in Equation (1).
2. CP classification: The clustering technique k-means was deployed. The aim was to generate a time series of daily CPs.
3. Frequency computation: Post-classification, the occurrence frequencies of each CP were standardised and calculated on an annual basis.
4. Setup of a regression model: The computed CP frequencies served as predictor variables in a subsequent multi-class logistic regression analysis.

For processing and integrating the rainfall data (predictand), the methodology consisted of:

1. Data integration: The different sources (WAHPD, KASS-D) of rainfall data were prepared and merged.
2. Station-based averaging: Rainfall records from multiple stations within the chosen region were aggregated to obtain an average areal rainfall measure. This average areal rainfall is then aggregated to yearly sums, denoted as R_A.

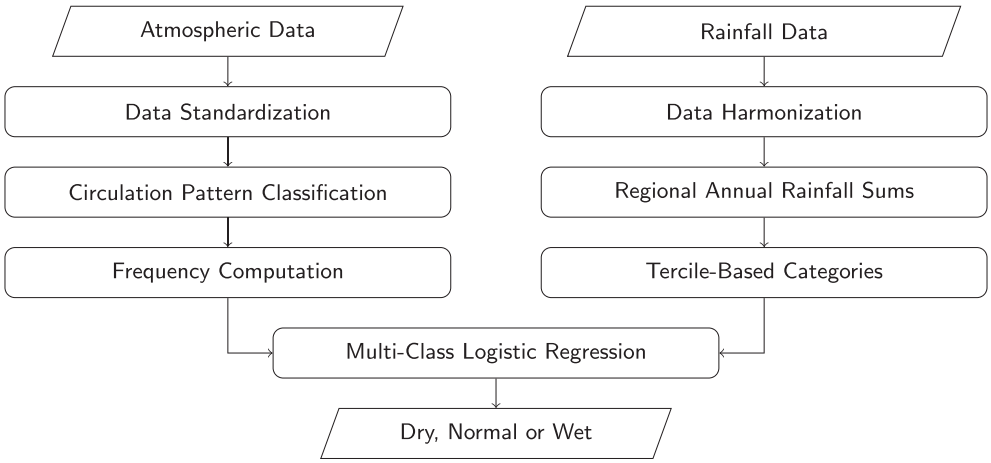

FIGURE 2 | Flowchart illustrating the multifaceted approach to reconstruct rainfall anomalies in the West African region. The flowchart is divided into two main branches: the left one focusing on atmospheric data and the right one on rainfall regions (RR). Each branch undergoes a series of transformations and classifications. They converge at a multi-class logistic regression model, which detects whether a year will be dry, normal, or wet.

3. Categorization: This R_A value was then discretized into ordinal classes using two quantile thresholds (q_1 and q_2). The classes were categorised as follows:
   - $R_A \leq q_1$: Class 1, signifying a 'dry' condition.
   - $q_1 < R_A < q_2$: Class 2, indicative of a 'normal' condition.
   - $R_A \geq q_2$: Class 3, representing a 'wet' condition.

   Here q_1 is the 0.33th quantile and q_2 is the 0.66th quantile. This choice was made to align with the seasonal forecasting approach used by regional climate outlook forums like PRESASS.
4. Preparation for regression: These categorizations then became the target variables in the multi-class logistic regression model.

Detailed discussions on important methodological steps follow in the subsequent sections.

3.1 | Rainfall Merging

To merge the KASS-D and WAHPD datasets (1375 stations in total), a multi-step approach is necessary. Initially, we address the discrepancies in time stamps between the datasets. The norm for daily rainfall measurements is a 24-h period, from 06:00 UTC to 06:00 UTC, but real-world inconsistencies often occur due to factors like local time zones, station-specific data collection practices, and occasional errors in recording or reporting. For KASS-D, stations assign date stamps for daily accumulated rainfall observations from 06:00 UTC to 06:00 UTC to the subsequent day. This practice deviates from the WAHPD database alignment. To reconcile this, we consider shifting the WAHPD time series by 1 day, aligning it with the KASS-D time stamps. We then evaluate the correlation between the adjusted WAHPD data and nearby stations within a 50 km radius. If a significant improvement in correlation (> 0.5 improvement) is observed, suggesting a closer agreement with nearby stations, the respective WAHPD station's data is shifted by 1 day for better alignment.

Prior to geostatistical analysis, the data undergoes preprocessing: years marked only by dry periods, and stations with only missing values or duplicate coordinates are excluded (1299 stations remaining). This ensures the dataset is free from anomalies like years with exclusive dry periods and stations with missing or redundant data. By considering the spatial relationships between stations, we can assess the correlation between neighbouring stations within a specific distance, which allows us to evaluate the similarity or dissimilarity of rainfall time series and identify any discrepancies between the databases. After acquiring the distance and correlation matrices, several geostatistical steps are performed to filter stations with insufficient data quality. These steps help ensure that the merged dataset includes reliable and representative stations. The following procedures are implemented:

- Highly correlated stations (less than 25 km apart, with a correlation greater than 0.9) were consolidated, retaining the one with more complete data (elimination of 280 stations).
- Poorly correlated nearby stations (less than 25 km apart, with a correlation less than 0.1) were removed (elimination of 25 stations).
- For distant but highly correlated stations (more than 25 km apart, with a correlation greater than 0.9), the station with the lowest correlation among the ten nearest was deleted (elimination of 23 stations).

By applying these geostatistical filtering steps, stations with insufficient data quality or poor spatial consistency are eliminated.

To address missing values, we employ a nearest-neighbour (NN) infilling approach. While traditional NN methods depend on Euclidean distance, our approach leans on the highest observed Pearson correlation for the specific station; this approach has outperformed others in our evaluations [not shown]. We draw upon the insights from Pegram et al. (2016), which show that the performance differences between

sophisticated statistical methods, such as dynamic copula regression (Bárdossy and Pegram 2014), and simpler NN techniques are minimal. Additional details about the NN infilling methodology can be found in the aforementioned references. Furthermore, NN methods preserve spatial variance more effectively than other interpolation techniques such as ordinary kriging or inverse distance weighting, which tend to cause smoothing (Rauch et al. 2024).

3.2 | K-Means

The k-means algorithm is a widely used clustering algorithm in atmospheric sciences (Huth et al. 2008; Govender and Sivakumar 2020), designed to group data by dividing samples into K distinct clusters, each characterised by similar variance levels. This algorithm is selected in this study for its simplicity, reproducibility and widespread use in CP classification (e.g., Solman and Menéndez 2003; Moron et al. 2008a).

The prerequisite for this algorithm is the pre-determination of the number of clusters. In k-means clustering, a set of N samples, denoted as X, is divided into K non-overlapping clusters C. These clusters are defined by their mean μ_j, known as the cluster's centroid. The primary objective of k-means is to determine centroids that minimise inertia, the within-cluster sum-of-squares criterion, represented as:

$$\sum_{i=0}^{n} \min_{\mu_j \in C} \left(\left\| x_i - \mu_j \right\|^2 \right) \tag{2}$$

For classifying the standardised anomalies of the predictors, we used the MiniBatchKMeans algorithm from scikit-learn (Pedregosa et al. 2012), an efficient adaptation of the traditional k-means. MiniBatchKMeans reduces computational time by processing mini-batches—smaller, randomly selected data subsets—in each iteration. This method significantly lowers the computational load required for converging to a local solution. Although this technique compromises the algorithm's accuracy, MiniBatchKMeans maintains high-quality outputs, typically with only marginal deviations from the standard k-means results. The k-means++ initialization method (Arthur and Vassilvitskii 2007) was used for robust centroid selection, which aids in faster convergence. To effectively uncover atmospheric CPs, we experimented with cluster counts ranging from 5 to 15. This range follows earlier studies in the region (Moron et al. 2008a; Rust et al. 2013; Bliefernicht, Rauch, et al. 2022) and reflects a balance between capturing dominant circulation regimes and avoiding overly fine clustering. It also ensures adequate sample sizes for stable logistic regression.

3.3 | Multi-Class Logistic Regression Model

Logistic regression is widely used in many atmospheric applications (Gijben et al. 2017; Moon et al. 2019; Moon and Kim 2020). It offers a method of modelling a categorical dependent variable based on one or more independent variables. Unlike linear regression, which estimates continuous outputs, logistic regression focuses on predicting binary outcomes using the logistic function. Logistic regression is chosen because it is highly interpretable, as coefficients directly reveal predictor influences. In our implementation, the computed frequencies of the CPs on an annual basis were used as predictor variables, while the classes—dry, normal, or wet—were used as the predictand variables. The use of CP frequencies as predictors, rather than direct atmospheric variables, reduces dimensionality and emphasises the role of physically interpretable CPs in modulating interannual rainfall variability.

Given a set of predictors $X_1, X_2, \ldots, X_p$, the logistic regression model expresses the probability $P(Y = 1)$ as:

$$P(Y = 1 | X) = \frac{e^{\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p}} \tag{3}$$

where $\beta_0, \beta_1, \ldots, \beta_p$ are coefficients that the model seeks to estimate. Given the multi-class nature of the dependent variable, a conventional binary logistic regression model would be insufficient. To address this, the study employed the one-vs-rest (OVR) strategy combined with logistic regression to perform multi-class classification. The OVR approach inherently breaks down a multi-class classification problem into multiple binary classification tasks. For a dataset with n classes, n distinct binary classifiers are trained. For each classifier, one class is treated as the positive class, while all other classes are treated as the negative class. In this study, the LogisticRegression function from the scikit-learn library was used (Pedregosa et al. 2012).

3.4 | Validation and Performance Measures

To evaluate the performance of our multi-class logistic regression model, we use two different metrics: Proportion correct (PC) and Peirce skill score (PSS). These metrics are implemented in accordance with the guidelines provided by Wilks (2006). For additional information on these performance measures, the cited work offers an in-depth discussion. The performance measures are derived from a 3×3 contingency table. The PC is the most straightforward metric, computed as the sum of the diagonal elements (correct hindcasts) divided by the total number of joint hindcast and observation pairs in a 3×3 contingency table. As pointed out by Grandini et al. (2020), the PC is particularly suitable for balanced datasets. In the tercile-based approach of this study, the balanced nature of the dataset ensures that the metric is not biased towards any specific class. A shortcoming of the PC is that no comparison is done versus a low skill reference. That is the reason why the PSS is chosen in addition to the PC to determine the skill of the reconstruction. The equation for the PSS is as follows:

$$\mathrm{PSS} = \frac{\sum_{i=1}^{I} p(y_i, o_i) - \sum_{i=1}^{I} p(y_i) p(o_i)}{1 - \sum_{j=1}^{J} p(o_j)^2} \tag{4}$$

Here, $p(y_i, o_i)$ represents the joint probability of hindcasts and observations, while $p(y_i)$ and $p(o_j)$ are the marginal distributions of hindcasts and observations, respectively. I and J denote the number of categories for hindcast and observations. The PSS aims to quantify the hindcast skill while accounting

for random chance. A perfect hindcast would result in PC and PSS scores of 1. Conversely, a hindcast no better than random chance would yield PSS scores of 0 and zero hits for PC. Negative PSS scores indicate a hindcast worse than random chance.

In atmospheric sciences, a variety of methods are used to assess the quality of regression models that link predictors and predictands. A key aspect of this validation process is choosing appropriate training and validation periods. This choice is crucial to avoid overfitting and ensure the reliability of the model. To achieve this, techniques like split-sampling (Bárdossy et al. 2002), cross-validation (Rust et al. 2013; Gutiérrez et al. 2019), stratified validation (Zeng and Martinez 2000), and statistical model ensembles (Hertig and Jacobeit 2008) can be applied. Particularly, the latter technique employs multiple statistical models across different training and validation periods to avoid artificial skill and address non-stationarities in predictor-predictand dynamics (Hertig and Jacobeit 2013). These non-stationarities occur on varying temporal and spatial scales, with the El Niño–Southern Oscillation being a well-documented example that significantly affects global climate (Janicot et al. 1996; Greatbatch et al. 2004; Cannon 2015).

This work is based on a 2/3 to 1/3 running split-sampling method, allocating 39 years for training and 13 years for validation. The complete approach, as illustrated in Figure 2, is applied to all variable combinations and CP number configurations leading to a total of 75,460 potential combinations for each RR in a running calibration mode. In this context, the metrics PC and PSS are calculated for each training (subscript t) and validation (subscript v) period for a selected period in a split-sampling approach. This period is then subsequently shifted forward by 1 year for the complete period (1959 to 2010). As the entire time series concludes, years from the start are progressively incorporated. In this study, we begin with a training period from 1959 to 1997, followed by a validation period from 1998 to 2010. Subsequently, we use the years 1960 to 1998 for training and 1959, 1999 to 2010 for validation, incorporating years from the beginning of the time series progressively.

The mean metrics $\widehat{\mathrm{PC}}$ and $\widehat{\mathrm{PSS}}$ are then computed over the statistical model ensembles. The final configuration is selected based on its performance during the validation period $\widehat{\mathrm{PC}}_v$, as evaluated during the ensemble step. This process ensures that the initial model solution is both robust and reliable.

3.5 | Value-Added

Based on the generally poor performance of GCMs for local-scale rainfall and their ability to simulate regional-scale meteorological phenomena, this study adopts the use of the 'raw' total precipitation variable from the ERA5 dataset to determine the added value of the proposed CP-based logistic regression model. By extracting daily rainfall sums on a point-to-pixel basis from ERA5, we directly compare our modelled outcomes with the reanalysis model output. Since rainfall exhibits high spatial variability (Rauch et al. 2024), we retain the original resolution of the fields, unlike the atmospheric variables which are coarsened to reduce the computational effort. This enables a comparative analysis that quantifies the enhancements our approach brings to rainfall reconstruction in West Africa, as opposed to the often used ERA5. We follow the same data preparation steps (averaging and categorization in RRs) and validation measures (PC, PSS) as described in Section 3.

4 | Results

The following section first discusses the creation of a rainfall dataset for West Africa and then the optimal parsimonious set of variables crucial for accurately reconstructing interannual rainfall anomalies in the WAM region. It furthermore presents a quality assessment of the CP-based multi-class logistic regression model of the best solution, with a particular focus on the Sahelian belt as a case study. This assessment encompasses an in-depth examination of the spatial composite of the selected CPs, their meteorological interpretation, CP statistics, the coefficients of the regression model, and a predictive analysis of interannual rainfall variability in the Sahel. Subsequently, the last chapter provides a brief summary of the other RRs.

4.1 | Rainfall Dataset

By applying the workflow of Section 3.1 the result is a curated dataset of 971 quality-controlled stations with an average station data availability of approximately 44% for the period of 1959 to 2010 (Figure 3), underscoring the data scarcity in West Africa. To easily apply statistical techniques like k-means, an infilled dataset is helpful. Therefore, the missing values are infilled as described in Section 3.1. Thus, we can establish an improved rainfall dataset for the West Africa region based on long-term in situ observations. This dataset serves then as the foundation for our analysis.

In literature, hard borders at specific latitudes are often employed to delineate the diverse rainfall patterns characterising the seasonal cycle across West Africa (Omotosho and Abiodun 2007; Akinsanola et al. 2017). While this conventional approach offers simplicity, it may fall short in capturing the interactions of climatic factors and the unique characterisation that defines the region's rainfall regimes (RRs). As a shift from this approach, we use a more fluid approach by employing the k-means algorithm to classify the rainfall patterns. This classification is based on standardised anomalies from the monthly sums. First, we aggregated the data into monthly sums. Afterwards, we calculated the standardised anomalies for each station by subtracting the station's long-term mean and then dividing by its standard deviation from the complete monthly time series. Thus, the k-means algorithm uses data from 624 months (52 years) across 971 stations as input. This analytical framework allows for a nuanced classification of rainfall variability.

These clusters, or RRs, offer representations of distinctive rainfall patterns in the region, thereby providing a better understanding of West Africa's rainfall dynamics. The geographical arrangement of the rainfall stations in relation to the designated RRs is illustrated in Figure 3. Additionally, their corresponding

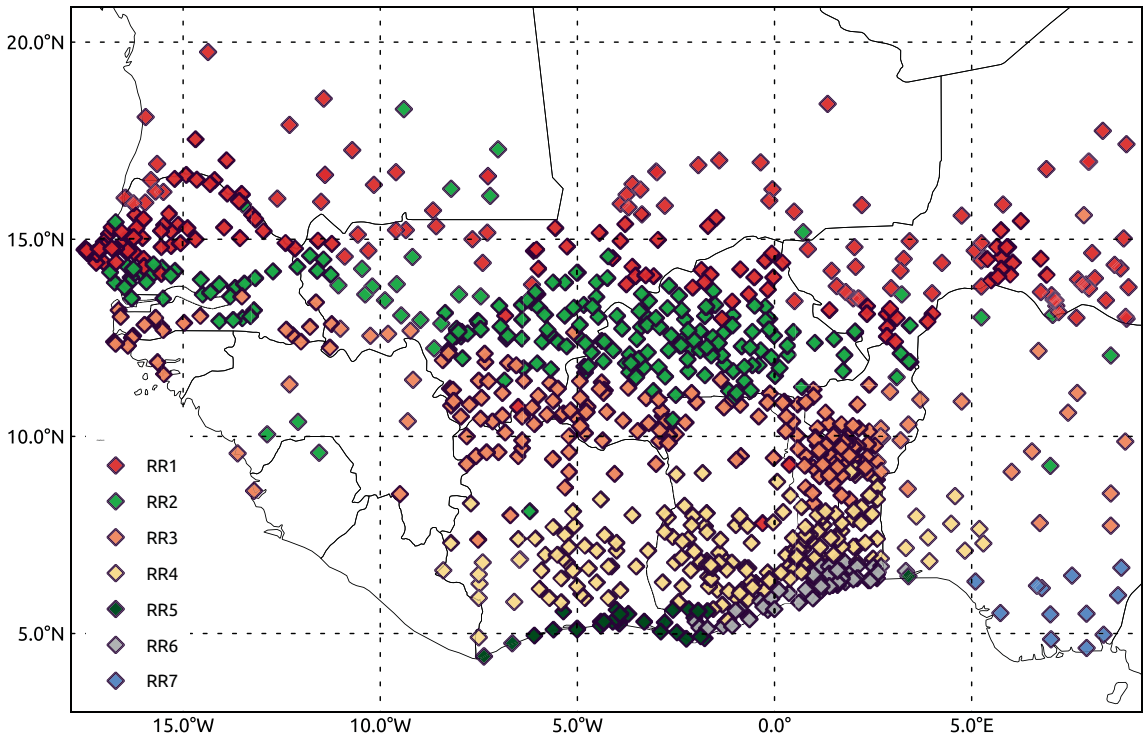

FIGURE 3 | Geographical distribution of rainfall stations across West Africa, classified into seven Rainfall Regimes (RRs) using the k-means algorithm based on monthly rainfall anomalies. Each RR represents a unique rainfall pattern, surpassing the traditional latitude-based distinctions: RR1—Sahelian Zone, RR2—Transition between Sahelian and Sudanian Zones, RR3—Sudanian Zone, RR4—Transition between Sudanian, Guinean and Coastal Zones, RR5—Guinean and Coastal Zones, Côte d'Ivoire, RR6—Guinean and Coastal Zones, Ghana, Togo, and Benin, and RR7—Guinean and Coastal Zones, Nigeria. [Colour figure can be viewed at wileyonlinelibrary.com]

monthly rainfall totals are depicted in Figure 4, while their interannual rainfall variability (target predictand) is presented in Figure 5. Within each RR, we sum the daily rainfall to obtain a regional annual sum, which is then standardised to zero mean and unit variance at the interannual scale. Here is a brief description of each RR:

- RR1: Strongly associated with the Sahelian belt, the northernmost region which typically experiences a single, short rainy season spanning from June to September. The interannual variability shows a wet phase starting in 1959, followed by the great Sahel drought from 1968 to the 1990s. The subsequent phase is a period of rainfall recovery since the 1990s (Descroix et al. 2018).
- RR2: A transition zone, bridging the Sahelian and Sudanian Zones. The rainfall pattern here reflects its intermediary position, with rainfall totals and season length increasing compared to the Sahelian Zone. This region shows a similar behaviour of the interannual rainfall variability to RR1, but the rainfall recovery seems to take longer, with more dry years extending until the early 2000s.
- RR3: Pertains predominantly to the Sudanian Zone, which is characterised by a lengthier rainy season extending from May to October. The rainfall is more abundant than in the Sahelian zone. Similar to RR1 and RR2, this region experiences comparable patterns of interannual rainfall variability, but the onset of the dry period is delayed until the early 1970s, similar to RR2.
- RR4: Serves as a climatic transition between the Sudanian Zone and the Guinean and Coastal Zones. It experiences a bi-modal rainfall pattern, although the 'August break' and rainfall peaks are less pronounced than in the zones further south. This region has fewer prolonged periods of dry or wet conditions compared to RR1 to RR3, indicating an overall more even distribution of the interannual rainfall variability categories.
- RR5: Primarily associated with the Guinean and Coastal Zones, specifically representing the coastal region of Côte d'Ivoire, experiences a unique double peak rainfall pattern due to its geographical proximity to the Atlantic Ocean and equatorial position. The maritime influence and position near the equator cause an extended rainy season with a primary maximum in June and a secondary peak in October. For this coastal region, there is a visual trend towards more normal and dry years from the 1990s to the 2010s, with fewer occurrences of wet years.
- RR6: Represents regions like Ghana, Togo, and Benin which exhibit a slightly less pronounced double peak rainfall pattern with lower peaks compared to RR5. This region shows a balance between dry and wet years and a generally more even distribution. It overlaps well with at least the coastal part of the "Dahomey Gap" (Vollmert et al. 2003).
- RR7: Represents mostly southern Nigeria specifically the wet Niger Delta region, characterised by a less pronounced

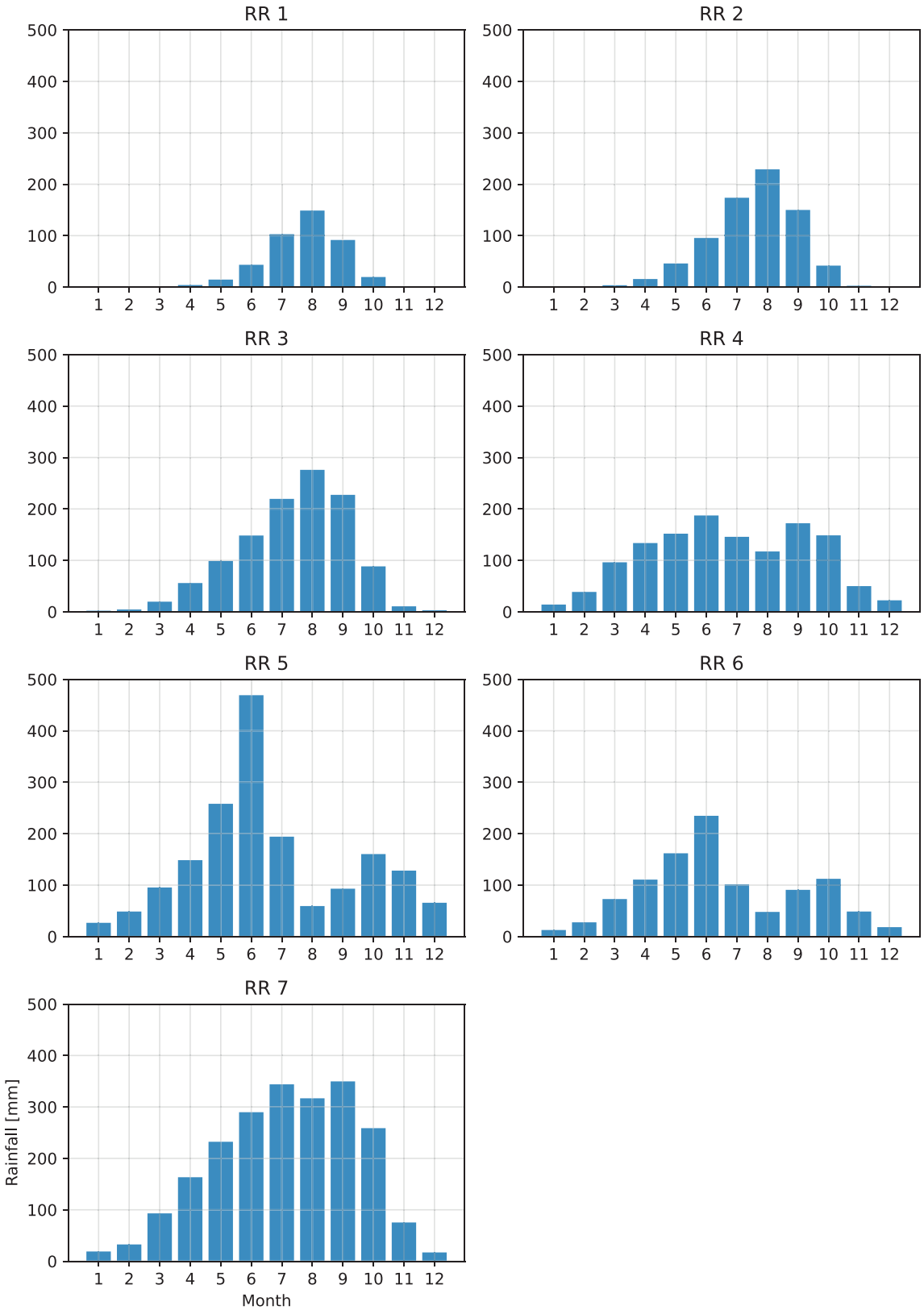

FIGURE 4 | Averaged monthly rainfall sums [mm] for the seven rainfall regimes (RRs). See Figure 3 for an overview of the different RRs. [Colour figure can be viewed at wileyonlinelibrary.com]

bimodal RR compared to RR5 or RR6. In this region, a dry phase followed by a recovery from normal to more wet years can be observed.

Because the presentation and application of the methodology for each RR are quite extensive, our main focus is on the findings of RR1, with summaries of the other RRs provided in Chapter 4.4.

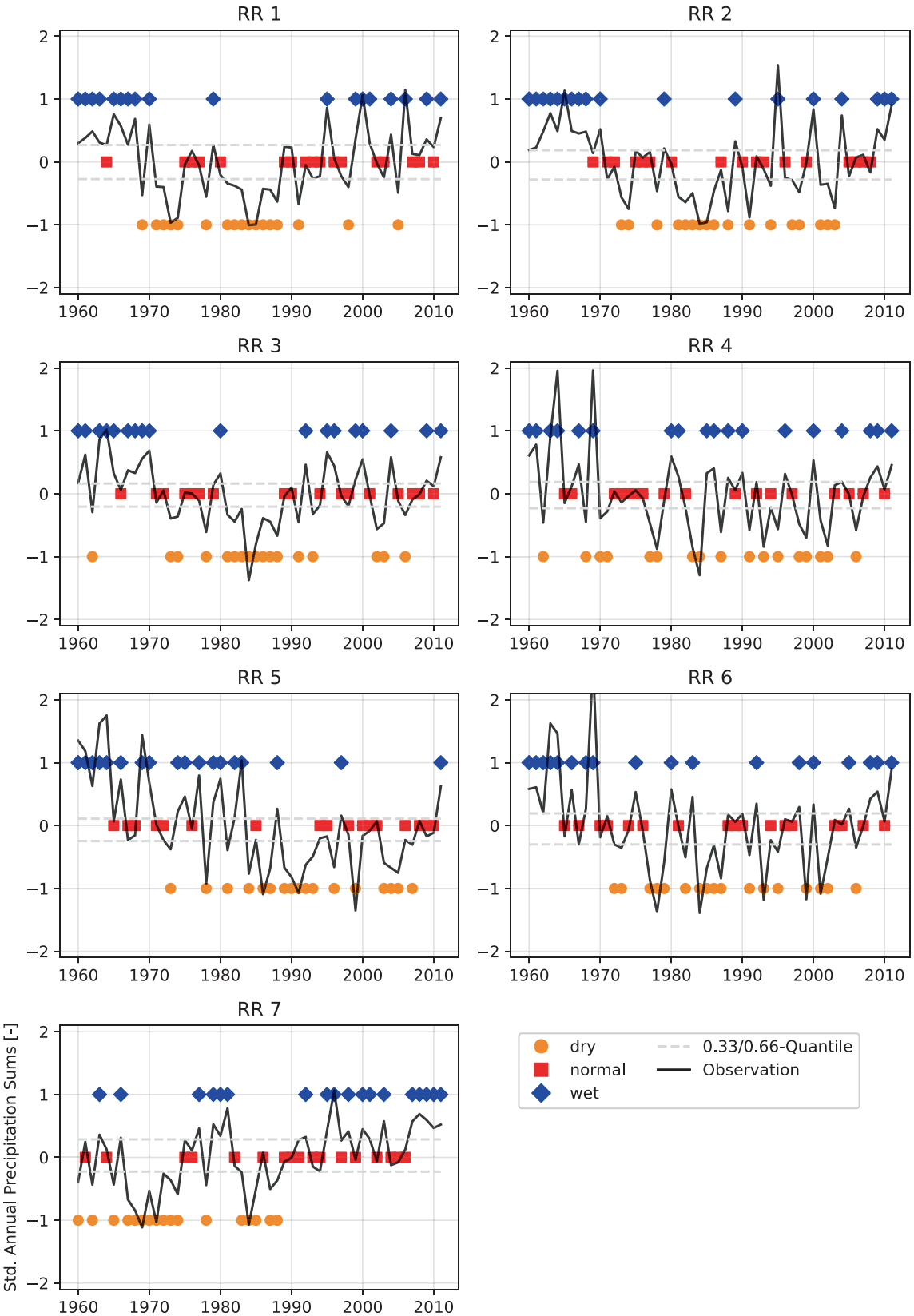

FIGURE 5 | Averaged standardised interannual rainfall variability for the seven RRs. See Figure 3 for an overview of the different RRs. The continuous black line represents the actual values of standardised annual rainfall sums, with the two dashed lines denoting the 0.33 and 0.66 quantiles, which segment the data into three categories: Dry (dark orange circle, below the first dashed line), normal (red square, between the two dashed lines), and wet (blue diamond, above the second dashed line). [Colour figure can be viewed at wileyonlinelibrary.com]

4.2 | Identifying the Optimal Parsimonious Set of Variables for Interannual Rainfall Reconstruction

This section focuses on identifying the most effective combination of atmospheric variables for reconstructing interannual rainfall in the RR1 region. We employ our methodology (Figure 2) to evaluate different variable combinations and determine which combination captures the essential dynamics influencing regional rainfall with the highest accuracy and efficiency.

Table 2 shows the best-performing models for each specific class of variable combinations, ranging from 1 to 4 variables. It reveals that the single variable WS200 delivers comparatively robust performance metrics. Specifically, WS200 as single variable achieved a $\widehat{\mathrm{PC}}_t$ of 0.627 and a $\widehat{\mathrm{PSS}}_t$ of 0.449, with the validation metrics standing at 0.602 and 0.244 respectively. When additional variables are included, the improvements in both $\widehat{\mathrm{PC}}$ and $\widehat{\mathrm{PSS}}$ are marginal.

Therefore, just a single variable and the solution of WS200 for RR1 is chosen due to its superior performance, simplicity, and lower computational demands compared to models using multiple variables. The slight improvements offered by additional variables do not outweigh the benefits of a parsimonious model, which adheres to the principle of achieving a similar level of skill scores (PC, PSS) with the least complexity.

4.3 | Sahelian Belt—RR1

Given the availability of solutions from 52 models (one for each shifted year) for the k-means clustering applied to the WS200 variable under six CPs, we chose to showcase an extraordinary case marked by an exceptionally high PC_v. This particular model underwent training during two separate intervals: from 1959 to 1977 and from 1991 to 2010. The validation period for the model covered the years from 1978 to 1990, a period notable for the great Sahel drought. During these phases, the model achieved the following performance measures: $\mathrm{PC}_t = 0.59$, $\mathrm{PSS}_t = 0.39$, $\mathrm{PC}_v = 0.85$ and $\mathrm{PSS}_v = 0.59$.

TABLE 2 | Evaluation of the optimal variable combinations for interannual rainfall variability reconstruction in RR1. This table identifies the highest-performing combinations of one to four, along with the count of CPs, and compares their performance metrics across training and validation phases.

k-means	CPs	$\widehat{\mathrm{PC}}_t$	$\widehat{\mathrm{PSS}}_t$	$\widehat{\mathrm{PC}}_v$	$\widehat{\mathrm{PSS}}_v$
WS200	6	0.627	0.449	0.602	0.244
WS200, SF200	7	0.640	0.466	0.612	0.276
U200, V200, WS200	7	0.652	0.485	0.620	0.263
U200, V700, WS200, SF200	7	0.637	0.463	0.623	0.284

Figure 6 shows the composites of the spatial CPs from WS200 anomalies which signify changes in wind speed in the upper troposphere, around an altitude of 12 km, often associated with the TEJ. The relationship between the TEJ and rainfall activity in the Sahel is well-documented (Lemburg et al. 2019; Nicholson and Klotter 2021). CP1, CP2 and CP3 mainly occur in the northern hemisphere's winter and spring months. Their occurrence decreases in April and reaches a minimum in May and October. The CPs are almost completely absent from June to September (Figure 7). The WS200 fields exhibit strong positive anomalies over the West African continent, which may be attributed to the strong westerly flow of the subtropical jet stream in winter (Krishnamurti 1961). During this period (especially December to March), the TEJ develops in the southern hemisphere, extending between the equator and 10° S (Nicholson and Klotter 2021), and is therefore mostly outside our predictor domain (Figure 1).

As the seasonal cycle progresses, key atmospheric features such as the TEJ migrate northward, in alignment with the sun's zenith position. Specifically, CP4 primarily occurs in May, October and November, while being completely absent in July and August (Figure 7). This pattern is further illustrated by the spatial reduction of positive anomalies and the emergence of higher negative anomalies in the WS200 fields (Figure 6). Moreover, the rainfall intensity is slightly increased when compared to CP1, CP2 and CP3 (Table 3).

CP5 and CP6 are primarily observed during the main monsoon period. CP5 occurs prominently in June with an average duration of 21 days (Figure 7). Notably, there is a slight positive anomaly at 0° N in the WS200 fields, which suggests a potentially weakened TEJ. CP6 exhibits a very specific temporal pattern, appearing only from June to August, but with a very high frequency, peaking at over 26 days in both July and August. This phase is characterised by pronounced positive WS200 anomalies around the Equator, indicating a strengthened TEJ, which typically signifies intense upper tropospheric winds. However, for a comprehensive analysis, it is crucial to evaluate the zonal wind component at this altitude to determine whether these anomalies correspond to a strengthened or weakened TEJ.

Table 3 summarises the behaviour of the six CPs under three tercile-based categories: dry, normal, and wet years and the climatology. Each category is evaluated using three metrics: (1) occurrence percentage, representing the relative occurrence frequency of each CP expressed as a percentage; (2) mean rainfall (R_A [mm]), indicating the average rainfall associated with each CP; and (3) occurrence anomaly, defined as the percentage increase or decrease in occurrence relative to climatology. In general, CP6 exhibits the highest amounts of rainfall, followed by CP5. Conversely, CP1 to CP4 show almost no rainfall (< 1 mm). CP1 is less common in wet years (−15.39%), while CP2 is less common in dry years (−12.39%). CP3 and CP4 are fairly stable across the different categories. In contrast, CP5 and CP6 exhibit significant variability, indicating greater sensitivity to rainfall variability. The occurrence of CP5 reduces considerably in wet years, dropping by −16.79%, which suggests it is less prevalent in wetter conditions. This decrease in both occurrence and rainfall amounts reveals CP5's sensitivity to dry years. On the other hand, CP6 occurs more often in wet years (+14.22%) with enhanced rainfall amounts.

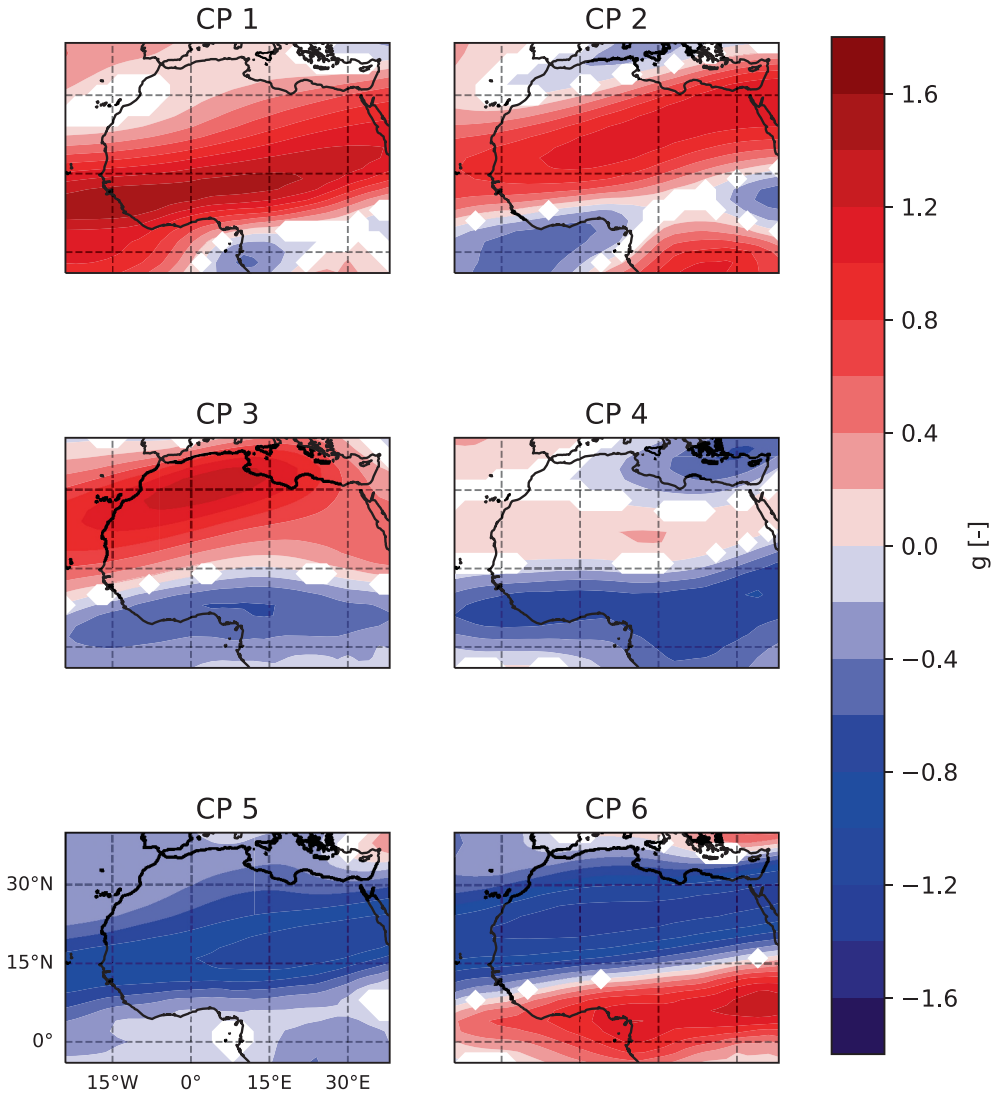

FIGURE 6 | Atmospheric CPs for the WS200 variable as the chosen solution for RR1. The colour scale signifies the mean intensity of the measured variable, where red and blue shades represent positive and negative anomalies, respectively. WS200 anomalies are represented only where the anomaly is statistically significant at the 1% level (or 99% confidence level) computed with a two-tailed *t*-test. [Colour figure can be viewed at wileyonlinelibrary.com]

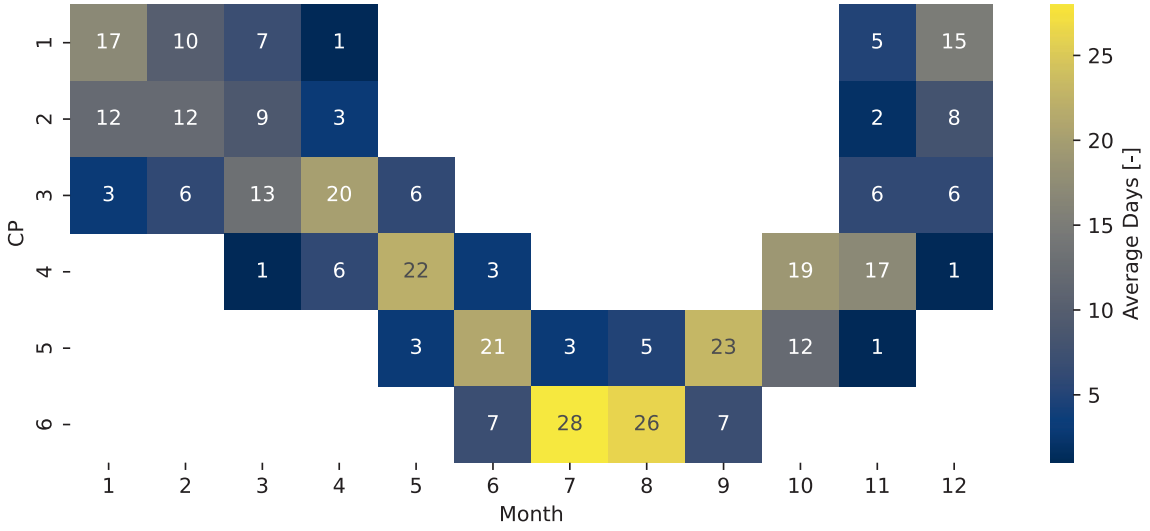

FIGURE 7 | Monthly occurrence frequencies in days for each CP of the classification using WS200. Only values exceeding more than 1 day are displayed. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 | Occurrence and rainfall statistics of the CPs in RR1 with respect to climatology, dry, normal, and wet years (1959–2010).

	Climatology		Dry years			Normal years			Wet years		
CP	Occ. [%]	R_A [mm]	Occ. [%]	R_A [mm]	Ano. [%]	Occ. [%]	R_A [mm]	Ano. [%]	Occ. [%]	R_A [mm]	Ano.[%]
1	14.99	0.03	16.66	0.03	10.00	14.22	0.02	−5.42	12.99	0.02	−15.39
2	12.80	0.03	11.34	0.03	−12.86	14.48	0.02	11.61	13.31	0.03	3.84
3	16.35	0.09	17.04	0.08	4.07	15.39	0.13	−6.24	16.30	0.09	−0.29
4	19.05	0.38	18.86	0.36	−1.02	18.42	0.31	−3.45	20.18	0.49	5.56
5	18.38	1.88	20.22	1.78	9.12	17.87	1.88	−2.82	15.73	2.13	−16.79
6	18.43	3.97	15.88	3.31	−16.08	19.62	4.21	6.06	21.48	4.57	14.22

Note: A χ^2-test assessed the statistical significance between observed (O) and expected (E) frequencies from climatology. All categories (dry, wet, and normal) showed p values below 0.05, rejecting the null hypothesis (H_0) that O equals E, indicating significant differences in all cases.
Abbreviations: Ano.: occurrence anomaly [%] with respect to the climatology; Occ.: relative occurrence of each CP [%]; R_A: spatially averaged rainfall amount [mm].

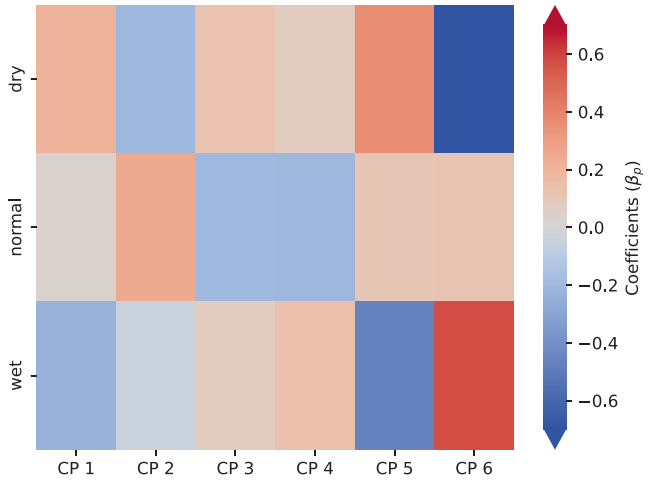

FIGURE 8 | Heat map of coefficients from the multi-class logistic regression model for RR1. The colour coding indicates the importance of each CP for dry, normal, or wet years. [Colour figure can be viewed at wileyonlinelibrary.com]

Regression coefficients from the multi-class logistic regression model further support these observations (Figure 8). The purpose of these coefficients is to gauge the likelihood of each CP occurring under specific categories, thereby providing an analytical foundation for understanding their climate sensitivity. For CP5, coefficients are positive during dry (0.36) but negative during wet years (−0.47), aligning with its occurrence data. For CP6, a negative coefficient during dry years (−0.69) and a positive one during wet years (0.57) indicate its high importance for these years. The coefficients for both CP5 and CP6 largely agree with their occurrence percentages and rainfall amounts, reinforcing their roles as key CPs for the rainy season in the Sahelian belt. Winter CPs indicating a dry year tend to persist longer. The findings from the logistic regression model support this hypothesis. CP1 exerts a positive impact (0.21) on dry conditions when it occurs more frequently, whereas it has a negative influence (−0.23) on wet conditions when it occurs less frequently. This demonstrates the significant role that winter CPs play.

Figure 9 displays the outcomes derived from the multi-class logistic regression model for RR1. This plot shows the model's reconstruction compared to the actual standardised annual rainfall amounts from 1959 to 2010. The grey continuous line shows the real values of standardised annual rainfall sums. The two dashed lines represent the terciles which cut the data into three rainfall categories: dry (below the first dashed line), normal (between the two dashed lines), and wet (above the second dashed line). Model hindcasts appear as scatter points: blue circle-markers indicate correct hindcasts, while yellow square-markers signify incorrect ones.

In the Sahel, two contrasting hydrological behaviours emerged over the past five decades. The first is the great Sahel drought from 1968 to the 1990s, while the second is the subsequent period of rainfall recovery since the 1990s (Descroix et al. 2018) (grey line in Figure 9). These phases are generally well-detected with minor inconsistencies, such as at the beginning of the time series or the years 2007 to 2009. Remarkably, the Sahel drought falls in into the validation period. In summary, built on the foundation of k-means clustering for the WS200 anomalies, this multi-class logistic regression model demonstrates an average $\widehat{PC}$ of 0.61 with a reasonable meteorological interpretation. This makes it a statistical tool for reconstructing annual rainfall amounts with respect to the Sahelian belt.

4.4 | Overall Performance and Value-Added

Table 4 provides an overview of the validation performance for each RR, the top-performing variable, their respective CP numbers, training and validation metrics. The RRs display balanced performance metrics between training and validation phases relative to RR1. The CP count varies across the regions, with values ranging from 5 to 14. Due to space constraints, a comprehensive presentation of results for each RR is beyond the scope of this paper. Therefore, we primarily focus on presenting the findings of RR1 as a representative case study.

Additionally, we evaluate the performance of our approach against the total precipitation variable from ERA5, serving as a reference(PC_{ERA5} and PSS_{ERA5} in Table 4). It is shown that our approach either outperforms or achieves comparable performance to that of ERA5 across all regions. Specifically, the proposed

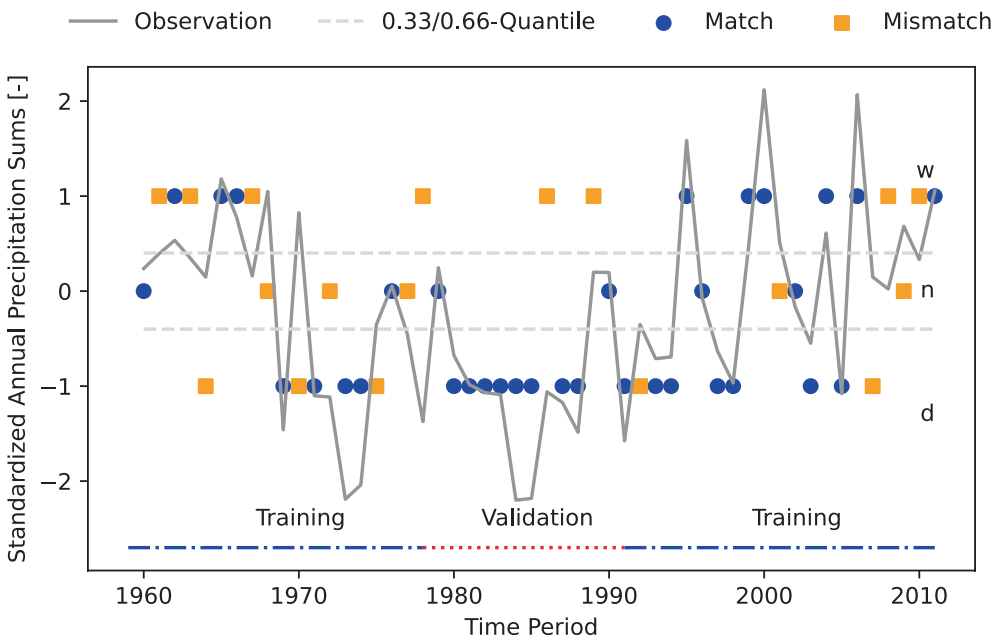

FIGURE 9 | Outcomes of the CP-based multi-class logistic regression model for the interannual rainfall variability in RR1. The continuous line represents the actual values of standardised annual rainfall sums, with the two dashed lines denoting the 0.33 and 0.66 quantiles, which segment the data into three categories: Dry (below the first dashed line), normal (between the two dashed lines) and wet (above the second dashed line). Predictions from the regression model are presented as scatter points: 'o' markers in blue represent matches (correct predictions), and 'x' markers in yellow indicate mismatches (incorrect predictions). [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 4 | Evaluation of the performance by region. The table presents the best-performing variables for each region based on the performance criterion $\widehat{PC}_v$, along with the count of CPs and the evaluation metrics for both training ($\widehat{PC}_t$, $\widehat{PSS}_t$) and validation ($\widehat{PC}_v$, $\widehat{PSS}_v$) phases. The values of PC_{ERA5} and PSS_{ERA5} show the performance of ERA5.

RR	Variable	CPs	$\widehat{PC}_t$	$\widehat{PC}_v$	$\widehat{PC}_{t,v}$	PC_{ERA5}	Diff. PC [%]	$\widehat{PSS}_t$	$\widehat{PSS}_v$	$\widehat{PSS}_{t,v}$	PSS_{ERA5}	Diff. PSS [%]
1	WS200	6	0.63	0.60	0.62	0.44	39.77	0.45	0.24	0.34	0.16	115.62
2	SF925	5	0.58	0.53	0.55	0.56	−0.89	0.37	0.20	0.29	0.34	−16.18
3	U200	6	0.62	0.55	0.58	0.56	4.46	0.43	0.22	0.32	0.34	−4.41
4	V200	14	0.82	0.48	0.65	0.46	41.30	0.73	0.19	0.46	0.19	142.11
5	SF850	6	0.66	0.55	0.60	0.42	44.05	0.49	0.19	0.34	0.13	161.54
6	U925	7	0.60	0.44	0.52	0.37	40.54	0.41	0.13	0.27	0.05	440.00
7	WD925	6	0.55	0.45	0.50	0.38	31.58	0.33	0.11	0.22	0.08	175.00

CP-based logistic model demonstrates an average performance with a PC of 0.57 and a PSS of 0.32. In contrast, ERA5 yields values of 0.46 and 0.18, respectively. Overall, a performance enhancement of more than 28% on average in terms of PC can be achieved, despite the relatively modest performance of both.

5 | Discussion

This study followed a straightforward statistical approach based on CPs to reconstruct interannual rainfall variability in the West African Sahelian belt. The further development of this study focuses on a comprehensive application covering the entire WAM region. It takes into account the variability across different RRs, relying on a long-term dataset spanning 52 years. Our approach establishes a robust catalogue of daily CP classification, which can be used for further refinement and offers greater flexibility to tackle specific research questions.

However, it is important to note that year-to-year changes in CP frequencies cannot fully account for the interannual variability of rainfall. Even if the frequency of a specific CP remains constant, it may still be associated with wetter conditions in some years than in others. This temporal non-stationarity, together with the limited sample size (i.e., one value per year), introduces the risk of overfitting in the statistical model, particularly when using many clusters or predictors. For subsequent research, it could be beneficial to delineate between high-frequency and low-frequency fluctuations—for instance, by categorising variations as either shorter or longer than a decade. This distinction is important because low-frequency (long-term) variations typically

demonstrate greater predictability compared to their high-frequency (short-term) counterparts. Such a separation could provide deeper insights into the dynamics influencing rainfall variability and enhance the accuracy of hindcasting efforts.

A study by Batté et al. (2018) also used a statistical-dynamical forecasting approach similar to ours. They employed forecasted CP-frequency anomalies from the ECMWF SEAS5 to construct temperature anomaly forecasts. Interestingly, they found no significant skill in predicting CP frequency at a seasonal timescale, a finding that should be evaluated in future for our models. This underscores the complexity of climate systems in the WAM region and highlights the challenges of seasonal forecasting for meteorological variables. Beyond their indirect utility, such as using CP frequencies for analyses, CPs can be included as additional covariates in regression models to directly model rainfall (Rust et al. 2013). They can also be used in conjunction with other techniques, such as canonical correlation analysis (Moron et al. 2008a). Furthermore, CPs can serve in an analog method (Zorita and von Storch 1999), wherein CPs simulated by climate models (e.g., CMIP6) are matched with local variables observed in conjunction with similar CPs from historical observations. For instance, this could allow us to more accurately predict the distribution of extreme values in future climate projections, which is crucial for understanding the risks associated with severe weather events. Additionally, our catalogue of classification schemes can be swiftly accessed and utilised for further regression tasks, thereby providing a valuable resource for future research.

The application of k-means clustering to delineate different RRs might be an overly broad generalisation for these diverse areas, particularly considering our initial choice of regionalisation based on the mean annual cycle, which may not fully capture the complexity of interannual variability. While the differentiation in Burkina Faso aligns well with the latitudinal subjective categorisation into the Sahelian zone, Sudano-Sahelian zone, and Sudanian zone as described by Garba et al. (2023), the longitudinal differentiation within regions like the Sahelian belt (RR1) poses challenges. For instance, projections of future rainfall patterns suggest that the central-eastern Sahel will experience increased rainfall, while the western Sahel will likely see a decrease (Monerie et al. 2012; Akinsanola and Zhou 2019). These distinctions, usually found between Eastern Senegal and Central Mali, are not visible in our analysis based on the mean annual cycle, questioning the efficacy of our chosen regionalisation approach for capturing the targeted interannual variability. This oversight underscores the need for a more nuanced approach that can account for both latitudinal and longitudinal variations in rainfall patterns, beyond the averaging of weather station data across the respective RRs.

The methodology used to define regional rainfall indices is a pivotal aspect of our analysis. It involves aggregating daily rainfall data within each cluster (RR) by mean to then obtain a regional annual sum, which is standardised to zero mean and unit variance on an interannual scale. This method was primarily selected for its potential to illuminate inter-cluster variations and trends over an extended period. However, this approach may inadvertently emphasise the contributions of the wettest stations within each region. Such stations, due to their higher rainfall totals, can disproportionately influence the regional index, potentially skewing the interpretation of regional rainfall characteristics. This issue is particularly critical as it may lead to an overestimation of wet conditions in regions that are otherwise more variably characterised by both dry and wet extremes. A better approach would be to standardise the annual rainfall at each station prior to calculating the average for the cluster. This method has the advantage of normalising the input from all stations, thereby mitigating the influence of outliers and emphasising years when rainfall anomalies are spatially consistent across the region. This could provide a more balanced view of regional climate patterns, especially in studies focused on interannual predictability where understanding spatially coherent anomalies is crucial.

The methodology presented in this study, mainly based on k-means clustering and logistic regression, offers a modular framework that can be adapted to suit different data environments, research goals, or end-user needs. In principle, each component of the workflow can be replaced by alternative techniques. The use of k-means clustering for identifying dominant CPs is widely established in atmospheric sciences (Solman and Menéndez 2003; Santos et al. 2005; Moron et al. 2008a). Nevertheless, other unsupervised or semi-supervised classification methods like hierarchical clustering (Guèye et al. 2011; Govender and Sivakumar 2020), SOMs (Hewitson and Crane 2002; Cassano et al. 2006), principal component-based methods (Galambosi et al. 1996; Esteban et al. 2005), or fuzzy rule-based methods (Bárdossy et al. 1995; Zehe et al. 2006; Wetterhall et al. 2009) could also be employed. While logistic regression was chosen for its transparency, interpretability, and compatibility with probabilistic classification tasks, more complex models such as random forests (Parmar et al. 2019), support vector machines (Noble 2006), or neural networks (Dotse 2024) could also be used. These methods may improve performance, particularly in nonlinear or high-dimensional predictor spaces.

This study highlights the promising performance of statistical adaptation approaches like the CP-based logistic regression model for rainfall reconstruction in challenging regions like West Africa, where GCM outputs alone may lack sufficient detail. By using these CP-based techniques, there is significant potential to enhance seasonal forecasts (Rauch et al. 2025) and build confidence in future climate projections for the Sahel (Biasutti 2019). These advancements underscore the value of refining statistical adaptation techniques to support robust and localised estimates of climate variability.

6 | Summary and Conclusion

This research provides an approach for a statistical tool using atmospheric circulation patterns to effectively reconstruct interannual rainfall variability in West Africa. This method integrates a reliable dataset of rainfall and atmospheric circulation patterns, offering a comprehensive view of the factors influencing rainfall.

We sourced information from 971 rainfall stations covering a broad geographical range (18°W to 10°E longitude and 4°N to 20°N latitude) spanning from 1959 to 2010. Our study categorises rainfall stations into distinct regimes representing different climatic zones, from the Sahelian Zone with a short rainy season to Nigeria's bimodal rainfall regime. The ERA-5 dataset served as our primary source of predictor data, with six pre-selected variables.

Our methodology involved several crucial steps to ensure data quality and contextual relevance. We standardised the predictors into anomaly-based measures, classified daily atmospheric circulation patterns using k-means, and established a robust historical catalogue for multiple relevant variables. After classification, we calculated the annual occurrence frequencies of each label, which then served as predictor variables in a multiclass logistic regression analysis. Our final solutions showed an average correct proportion of 0.57 and positive PSSs for all regions, mostly outperforming the TP variable from the ERA5 dataset. This makes it a valuable tool for reconstructing annual rainfall across various West African regimes. To address non-stationarities and avoid overfitting, we incorporated an extensive validation framework with statistical model ensembles using running training and validation periods.

Focusing specifically on the Sahelian belt, the study highlights two contrasting hydrological behaviours over the past five decades: the great Sahel drought (1968 to the 1990s) and the subsequent rainfall recovery since the 1990s. These periods are linked to TEJ. Wet years show high frequency with significant positive anomalies in upper air wind speed (200 hPa), indicating a strong TEJ and intense upper tropospheric winds, while dry years show a slight positive anomaly, suggesting a potentially weakened TEJ. Thus, we have directly combined atmospheric characteristics with the annual rainfall anomalies in this difficult region. In summary, our key results are:

- A quality-controlled and infilled dataset of 971 rainfall stations spanning from 1959 to 2010 for the data-scarce West African region.
- A nuanced classification of seven rainfall regimes, from the Sahelian belt to the Guinea zone.
- A robust catalogue of daily atmospheric circulation pattern classifications.
- A straightforward statistical method using daily circulation patterns to reconstruct interannual rainfall anomalies across seven rainfall regimes.

Furthermore, the presented work opens up numerous possibilities for statistical applications in climate science and other disciplines within the WAM region. For example, the target can be easily shifted from the interannual variability to other applications (e.g., the onset of the rainy season, Rauch et al. 2019), thereby enhancing the understanding of rainfall or other meteorological variables in this challenging region. Moreover, the developed methodology forms the foundation for the seasonal prediction of rainfall variability in the West African Sudan–Sahel region, serving as a subsequent study to support the West African regional climate outlook forums (Rauch et al. 2025).

Author Contributions

Manuel Rauch: conceptualization, methodology, software, data curation, investigation, validation, formal analysis, visualization, writing – original draft, writing – review and editing. **Jan Bliefernicht:** conceptualization, methodology, data curation, validation, supervision, funding acquisition, project administration, writing – original draft, writing – review and editing, formal analysis, resources. **Marlon Maranan:** writing – review and editing, writing – original draft, funding acquisition, project administration, conceptualization, data curation. **Andreas H. Fink:** data curation, funding acquisition, project administration, writing – original draft, writing – review and editing. **Harald Kunstmann:** writing – original draft, writing – review and editing, funding acquisition, project administration.

Acknowledgements

Special thanks to Andreas Fischer (University of Augsburg, Institute of Physics) for his guidance in optimising parallelisation on our computing cluster. The tools MiniBatchKMeans and Logistic Regression are from the scikit-learn library (Pedregosa et al. 2012). Data processing, handling, and manipulation were done with xarray (Hoyer and Hamman 2017), numpy (Harris et al. 2020), pandas (McKinney 2010), and ray for parallel processing (Moritz et al. 2017). Figures 1 and 3 were created using QGIS. Figures 4–9 were generated using matplotlib (Hunter 2007). Figure 2 was generated using the TikZpicture package. Manuscript drafting and formatting were done in LaTeX on the Overleaf platform. ChatGPT-4 assisted with spell-checking and grammar improvement. We also thank the editor, Dr. Annalisa Cherchi, and the reviewers for their helpful comments that improved the quality of the manuscript. Open Access funding enabled and organized by Projekt DEAL.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The in situ rainfall records were sourced from KASS-D and the WAHPD, neither of which is publicly accessible. The ERA-5 reanalysis dataset was obtained from the Climate Data Store of the Copernicus Climate Change Service (Hersbach et al. 2023a, 2023b). The CP classification data can be made available upon request by contacting the corresponding author.

References

Akinsanola, A. A., K. O. Ogunjobi, V. O. Ajayi, E. A. Adefisan, J. A. Omotosho, and S. Sanogo. 2017. "Comparison of Five Gridded Precipitation Products at Climatological Scales Over West Africa." *Meteorology and Atmospheric Physics* 129, no. 6: 669–689. https://doi.org/10.1007/s00703-016-0493-6.

Akinsanola, A. A., and W. Zhou. 2019. "Ensemble-Based CMIP5 Simulations of West African Summer Monsoon Rainfall: Current Climate and Future Changes." *Theoretical and Applied Climatology* 136, no. 3–4: 1021–1031. https://doi.org/10.1007/s00704-018-2516-3.

Arthur, D., and S. Vassilvitskii. 2007. "K-Means++: The Advantages of Careful Seeding." In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. SODA '07*, 1027–1035. Society for Industrial and Applied Mathematics.

Ascott, M. J., D. M. J. Macdonald, E. Black, et al. 2020. "In Situ Observations and Lumped Parameter Model Reconstructions Reveal Intra-Annual to Multidecadal Variability in Groundwater Levels in Sub-Saharan Africa." *Water Resources Research* 56, no. 12: e2020WR028056. https://doi.org/10.1029/2020WR028056.

Atiah, W. A., L. K. Amekudzi, and S. K. Danuor. 2023. "Mesoscale Convective Systems and Contributions to Flood Cases in Southern West Africa (SWA): A Systematic Review." *Weather and Climate Extremes* 39: 100551. https://doi.org/10.1016/j.wace.2023.100551.

Bárdossy, A., L. Duckstein, and I. Bogardi. 1995. "Fuzzy Rule-Based Classification of Atmospheric Circulation Patterns." *International Journal of Climatology* 15, no. 10: 1087–1097. https://doi.org/10.1002/JOC.3370151003.

Bárdossy, A., and G. Pegram. 2014. "Infilling Missing Precipitation Records—A Comparison of a New Copula-Based Method With Other Techniques." *Journal of Hydrology* 519, no. PA: 1162–1170. https://doi.org/10.1016/J.JHYDROL.2014.08.025.

Bárdossy, A., J. Stehlík, and H. Caspary. 2002. "Automated Objective Classification of Daily Circulation Patterns for Precipitation and Temperature Downscaling Based on Optimized Fuzzy Rules." *Climate Research* 23: 11–22. https://doi.org/10.3354/cr023011.

Batté, L., C. Ardilouze, and M. Déqué. 2018. "Forecasting West African Heat Waves at Subseasonal and Seasonal Time Scales." *Monthly Weather Review* 146, no. 3: 889–907. https://doi.org/10.1175/MWR-D-17-0211.1.

Biasutti, M. 2019. "Rainfall Trends in the African Sahel: Characteristics, Processes, and Causes." *Wiley Interdisciplinary Reviews: Climate Change* 10, no. 4: e591. https://doi.org/10.1002/wcc.591.

Bliefernicht, J., S. Berger, S. Salack, et al. 2018. "The WASCAL Hydrometeorological Observatory in The Sudan Savanna of Burkina Faso and Ghana." *Vadose Zone Journal* 17, no. 1: 1–20. https://doi.org/10.2136/vzj2018.03.0065.

Bliefernicht, J., M. Rauch, P. Laux, and H. Kunstmann. 2022. "Atmospheric Circulation Patterns That Trigger Heavy Rainfall in West Africa." *International Journal of Climatology* 42, no. 12: 6515–6536. https://doi.org/10.1002/joc.7613.

Bliefernicht, J., S. Salack, M. Waongo, T. Annor, P. Laux, and H. Kunstmann. 2022. "Towards a Historical Precipitation Database for West Africa: Overview, Quality Control and Harmonization." *International Journal of Climatology* 42, no. 7: 4001–4023. https://doi.org/10.1002/joc.7467.

Bliefernicht, J., M. Waongo, S. Salack, J. Seidel, P. Laux, and H. Kunstmann. 2019. "Quality and Value of Seasonal Precipitation Forecasts Issued by the West African Regional Climate Outlook Forum." *Journal of Applied Meteorology and Climatology* 58, no. 3: 621–642. https://doi.org/10.1175/JAMC-D-18-0066.1.

Böker, B., P. Laux, P. Olschewski, and H. Kunstmann. 2023. "Added Value of an Atmospheric Circulation Pattern-Based Statistical Downscaling Approach for Daily Precipitation Distributions in Complex Terrain." *International Journal of Climatology* 43, no. 11: 5130–5153. https://doi.org/10.1002/joc.8136.

Camberlin, P., M. Kpanou, and P. Roucou. 2020. "Classification of Intense Rainfall Days in Southern West Africa and Associated Atmospheric Circulation." *Atmosphere* 11, no. 2: 188. https://doi.org/10.3390/ATMOS11020188.

Cannon, A. J. 2015. "Revisiting the Nonlinear Relationship Between ENSO and Winter Extreme Station Precipitation in North America." *International Journal of Climatology* 35, no. 13: 4001–4014. https://doi.org/10.1002/JOC.4263.

Cassano, J. J., P. Uotila, and A. Lynch. 2006. "Changes in Synoptic Weather Patterns in the Polar Regions in the Twentieth and Twenty-First Centuries, Part 1: Arctic." *International Journal of Climatology* 26, no. 8: 1027–1049. https://doi.org/10.1002/JOC.1306.

Descroix, L., F. Guichard, M. Grippa, et al. 2018. "Evolution of Surface Hydrology in the Sahelo-Sudanian Strip: An Updated Review." *Water* 10, no. 6: 748. https://doi.org/10.3390/w10060748.

Diedhiou, S., M. Rauch, A. L. Dieng, et al. 2024. "Extreme Rainfall in Dakar (Senegal): A Case Study for September 5, 2020." *Frontiers in Water* 6: 1439404. https://doi.org/10.3389/frwa.2024.1439404.

Dotse, S. Q. 2024. "Deep Learning–Based Long Short-Term Memory Recurrent Neural Networks for Monthly Rainfall Forecasting in Ghana, West Africa." *Theoretical and Applied Climatology* 155, no. 4: 3033–3045. https://doi.org/10.1007/s00704-023-04773-x.

Engel, T., A. H. Fink, P. Knippertz, G. Pante, and J. Bliefernicht. 2017. "Extreme Precipitation in the West African Cities of Dakar and Ouagadougou: Atmospheric Dynamics and Implications for Flood Risk Assessments." *Journal of Hydrometeorology* 18, no. 11: 2937–2957. https://doi.org/10.1175/JHM-D-16-0218.1.

Epule, E. T., C. Peng, L. Lepage, and Z. Chen. 2013. "The Causes, Effects and Challenges of Sahelian Droughts: A Critical Review." *Regional Environmental Change* 14, no. 1: 145–156. https://doi.org/10.1007/S10113-013-0473-Z.

Esteban, P., P. D. Jones, J. Martín-Vide, and M. Mases. 2005. "Atmospheric Circulation Patterns Related to Heavy Snowfall Days in Andorra, Pyrenees." *International Journal of Climatology* 25, no. 3: 319–329. https://doi.org/10.1002/joc.1103.

Fink, A. H., T. Engel, V. Ermert, et al. 2017. "Mean Climate and Seasonal Cycle." In *Wiley Meteorology of Tropical West Africa*, 1–39. Wiley. https://onlinelibrary.wiley.com/doi/10.1002/9781118391297.ch110.1002/9781118391297.ch1.

Fontaine, B., and S. Bigot. 1993. "West African Rainfall Deficits and Sea Surface Temperatures." *International Journal of Climatology* 13, no. 3: 271–285. https://doi.org/10.1002/joc.3370130304.

Fontaine, B., and S. Janicot. 1996. "Sea Surface Temperature Fields Associated With West African Rainfall Anomaly Types." *Journal of Climate* 9, no. 11: 2935–2940. https://doi.org/10.1175/1520-0442(1996)009<2935:SSTFAW>2.0.CO;2.

Galambosi, A., L. Duckstein, and I. Bogardi. 1996. "Evaluation and Analysis of Daily Atmospheric Circulation Patterns of the 500 HPA Pressure Field Over the Southwestern USA." *Atmospheric Research* 40, no. 1: 49–76. https://doi.org/10.1016/0169-8095(95)00025-9.

Galle, S., M. Grippa, C. Peugeot, et al. 2018. "AMMA-CATCH, a Critical Zone Observatory in West Africa Monitoring a Region in Transition." *Vadose Zone Journal* 17, no. 1: 1–24. https://doi.org/10.2136/vzj2018.03.0062.

Garba, J. N., U. J. Diasso, M. Waongo, W. Sawadogo, and T. Daho. 2023. "Performance Evaluation of Satellite-Based Rainfall Estimation Across Climatic Zones in Burkina Faso." *Theoretical and Applied Climatology* 154: 1051–1073. https://doi.org/10.1007/s00704-023-04593-z.

Gijben, M., L. L. Dyson, and M. T. Loots. 2017. "A Statistical Scheme to Forecast the Daily Lightning Threat Over Southern Africa Using the Unified Model." *Atmospheric Research* 194: 78–88. https://doi.org/10.1016/j.atmosres.2017.04.022.

Govender, P., and V. Sivakumar. 2020. "Application of k-Means and Hierarchical Clustering Techniques for Analysis of Air Pollution: A Review (1980–2019)." *Atmospheric Pollution Research* 11, no. 1: 40–56. https://doi.org/10.1016/j.apr.2019.09.009.

Grandini, M., E. Bagli, and G. Visani. 2020. "Metrics for Multi-Class Classification: An Overview arXiv 2020 1 2008.05756." https://arxiv.org/abs/2008.05756v1.

Greatbatch, R. J., J. Lu, and K. A. Peterson. 2004. "Nonstationary Impact of ENSO on Euro-Atlantic Winter Climate." *Geophysical Research Letters* 31, no. 2: 2003GL018542. https://doi.org/10.1029/2003GL018542.

Grist, J. P., and S. E. Nicholson. 2001. "A Study of the Dynamic Factors Influencing the Rainfall Variability in the West African Sahel." *Journal of Climate* 14, no. 7: 1337–1359. https://doi.org/10.1175/1520-0442(2001)014<1337:ASOTDF>2.0.CO;2.

Guèye, A. K., S. Janicot, A. Niang, et al. 2011. "Weather Regimes Over Senegal During the Summer Monsoon Season Using Self-Organizing Maps and Hierarchical Ascendant Classification. Part I: Synoptic Time Scale." *Climate Dynamics* 36, no. 1–2: 1–18. https://doi.org/10.1007/s00382-010-0782-6.

Guèye, A. K., S. Janicot, A. Niang, et al. 2012. "Weather Regimes Over Senegal During the Summer Monsoon Season Using Self-Organizing Maps and Hierarchical Ascendant Classification. Part II: Interannual Time Scale." *Climate Dynamics* 39, no. 9–10: 2251–2272. https://doi.org/10.1007/S00382-012-1346-8.

Gutiérrez, J. M., D. Maraun, M. Widmann, et al. 2019. "An Intercomparison of a Large Ensemble of Statistical Downscaling Methods Over Europe: Results From the VALUE Perfect Predictor Cross-Validation Experiment." *International Journal of Climatology* 39, no. 9: 3750–3785. https://doi.org/10.1002/joc.5462.

Hagos, S. M., and K. H. Cook. 2008. "Ocean Warming and Late-Twentieth-Century Sahel Drought and Recovery." *Journal of Climate* 21, no. 15: 3797–3814. https://doi.org/10.1175/2008JCLI2055.1.

Harris, C. R., K. J. Millman, S. J. van der Walt, et al. 2020. "Array Programming With NumPy." *Nature 2020 585:7825* 585, no. 7825: 357–362. https://doi.org/10.1038/s41586-020-2649-2.

Hersbach, H., B. Bell, P. Berrisford, et al. 2023a. "ERA5 Hourly Data on Pressure Levels From 1940 to Present." Copernicus Climate Change Service (C3S). Climate Data Store (CDS). https://doi.org/10.24381/cds.bd0915c6.

Hersbach, H., B. Bell, P. Berrisford, et al. 2023b. "ERA5 Hourly Data on Single Levels From 1940 to Present." Copernicus Climate Change Service (C3S). Climate Data Store (CDS). https://doi.org/10.24381/cds.adbb2d47.

Hersbach, H., B. Bell, P. Berrisford, S. Hirahara, A. Horányi, and J. Muñoz-Sabater. 2020. "The ERA5 Global Reanalysis." *Quarterly Journal of the Royal Meteorological Society* 146, no. 730: 1999–2049. https://doi.org/10.1002/QJ.3803.

Hertig, E., and J. Jacobeit. 2008. "Assessments of Mediterranean Precipitation Changes for the 21st Century Using Statistical Downscaling Techniques." *International Journal of Climatology* 28, no. 8: 1025–1045. https://doi.org/10.1002/joc.1597.

Hertig, E., and J. Jacobeit. 2013. "A Novel Approach to Statistical Downscaling Considering Nonstationarities: Application to Daily Precipitation in the Mediterranean Area." *Journal of Geophysical Research: Atmospheres* 118, no. 2: 520–533. https://doi.org/10.1002/jgrd.50112.

Hewitson, B. C., and R. G. Crane. 2002. "Self-Organizing Maps: Applications to Synoptic Climatology." *Climate Research* 22, no. 1: 13–26. https://doi.org/10.3354/CR022013.

Hoyer, S., and J. J. Hamman. 2017. "Xarray: N-D Labeled Arrays and Datasets in Python." *Journal of Open Research Software* 5, no. 1: 10. https://doi.org/10.5334/JORS.148.

Hunter, J. D. 2007. "Matplotlib: A 2D Graphics Environment." *Computing in Science & Engineering* 9, no. 3: 90–95. https://doi.org/10.1109/MCSE.2007.55.

Huth, R., C. Beck, A. Philipp, et al. 2008. "Classifications of Atmospheric Circulation Patterns." *Annals of the New York Academy of Sciences* 1146, no. 1: 105–152. https://doi.org/10.1196/ANNALS.1446.019.

Janicot, S., V. Moron, and B. Fontaine. 1996. "Sahel Droughts and Enso Dynamics." *Geophysical Research Letters* 23, no. 5: 515–518. https://doi.org/10.1029/96GL00246.

Krishnamurti, T. N. 1961. "The Subtropical Jet Stream of Winter." *Journal of Meteorology* 18, no. 2: 172–191. https://doi.org/10.1175/1520-0469(1961)018<0172:TSJSOW>2.0.CO;2.

Lafore, J., C. Flamant, V. Giraud, et al. 2010. "Introduction to the AMMA Special Issue on 'Advances in Understanding Atmospheric Processes Over West Africa Through the AMMA Field Campaign'." *Quarterly Journal of the Royal Meteorological Society* 136, no. S1: 2–7. https://doi.org/10.1002/qj.583.

Lafore, J., C. Flamant, F. Guichard, et al. 2011. "Progress in Understanding of Weather Systems in West Africa." *Atmospheric Science Letters* 12, no. 1: 7–12. https://doi.org/10.1002/asl.335.

Lemburg, A., J. Bader, and M. Claussen. 2019. "Sahel Rainfall–Tropical Easterly Jet Relationship on Synoptic to Intraseasonal Time Scales." *Monthly Weather Review* 147, no. 5: 1733–1752. https://doi.org/10.1175/MWR-D-18-0254.1.

Losada, T., B. Rodríguez-Fonseca, S. Janicot, S. Gervois, F. Chauvin, and P. Ruti. 2010. "A Multi-Model Approach to the Atlantic Equatorial Mode: Impact on the West African Monsoon." *Climate Dynamics* 35, no. 1: 29–43. https://doi.org/10.1007/s00382-009-0625-5.

Maher, P., M. E. Kelleher, P. G. Sansom, and J. Methven. 2020. "Is the Subtropical Jet Shifting Poleward?" *Climate Dynamics* 54, no. 3–4: 1741–1759. https://doi.org/10.1007/s00382-019-05084-6.

Manzanas, R. 2017. "Assessing the Suitability of Statistical Downscaling Approaches for Seasonal Forecasting in Senegal." *Atmospheric Science Letters* 18, no. 9: 381–386. https://doi.org/10.1002/asl.767.

Maranan, M., A. H. Fink, P. Knippertz, et al. 2019. "Interactions Between Convection and a Moist Vortex Associated With an Extreme Rainfall Event Over Southern West Africa." *Monthly Weather Review* 147, no. 7: 2309–2328. https://doi.org/10.1175/MWR-D-18-0396.1.

Maraun, D., F. Wetterhall, A. M. Ireson, et al. 2010. "Precipitation Downscaling Under Climate Change: Recent Developments to Bridge the Gap Between Dynamical Models and the End User." *Reviews of Geophysics* 48, no. 3: RG3003. https://doi.org/10.1029/2009RG000314.

Maraun, D., and M. Widmann. 2018. *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge University Press. https://www.cambridge.org/core/books/statistical-downscaling-and-bias-correction-for-climate-research/4ED479BAA8309C7ECBE6136236E3960F.

Mason, S., and S. Chidzambwa. 2009. "Position Paper: Verification of African RCOF Forecasts." IRI Technical Report, 09-02. https://doi.org/10.7916/D85T3SB0.

Mathon, V., H. Laurent, and T. Lebel. 2002. "Mesoscale Convective System Rainfall in the Sahel." *Journal of Applied Meteorology* 41, no. 11: 1081–1092. https://doi.org/10.1175/1520-0450(2002)041<1081:MCSRIT>2.0.CO;2.

McKinney, W. 2010. "Data Structures for Statistical Computing in Python." In *Proceedings of the 9th Python in Science Conference*, edited by S. van der Walt and J. Millman, 51–56. SciPy.

Monerie, P. A., B. Fontaine, and P. Roucou. 2012. "Expected Future Changes in the African Monsoon Between 2030 and 2070 Using Some CMIP3 and CMIP5 Models Under a Medium-Low RCP Scenario." *Journal of Geophysical Research: Atmospheres* 117, no. D16: D16111. https://doi.org/10.1029/2012JD017510.

Moon, S. H., and Y. H. Kim. 2020. "An Improved Forecast of Precipitation Type Using Correlation-Based Feature Selection and Multinomial Logistic Regression." *Atmospheric Research* 240: 104928. https://doi.org/10.1016/j.atmosres.2020.104928.

Moon, S. H., Y. H. Kim, Y. H. Lee, and B. R. Moon. 2019. "Application of Machine Learning to an Early Warning System for Very Short-Term Heavy Rainfall." *Journal of Hydrology* 568: 1042–1054. https://doi.org/10.1016/j.jhydrol.2018.11.060.

Moritz, P., R. Nishihara, S. Wang, et al. 2017. "Ray: A Distributed Framework for Emerging AI Applications." In *Proceedings of the 13th USENIX Symposium on Operating Systems Design and Implementation (OSDI 18)*, 561–577. USENIX Association.

Moron, V., B. Oueslati, B. Pohl, and S. Janicot. 2018. "Daily Weather Types in February–June (1979–2016) and Temperature Variations in Tropical North Africa." *Journal of Applied Meteorology and Climatology* 57, no. 5: 1171–1195. https://doi.org/10.1175/JAMC-D-17-0105.1.

Moron, V., A. W. Robertson, M. N. Ward, and O. Ndiaye. 2008a. "Weather Types and Rainfall Over Senegal. Part I: Observational Analysis." *Journal of Climate* 21, no. 2: 266–287. https://doi.org/10.1175/2007JCLI1601.1.

Moron, V., A. W. Robertson, M. N. Ward, and O. Ndiaye. 2008b. "Weather Types and Rainfall Over Senegal. Part II: Downscaling of GCM Simulations." *Journal of Climate* 21, no. 2: 288–307. https://doi.org/10.1175/2007JCLI1624.1.

Ndiaye, O., M. N. Ward, and W. M. Thiaw. 2011. "Predictability of Seasonal Sahel Rainfall Using GCMs and Lead-Time Improvements Through the Use of a Coupled Model." *Journal of Climate* 24, no. 7: 1931–1949. https://doi.org/10.1175/2010JCLI3557.1.

Ngoungue Langue, C. G., C. Lavaysse, M. Vrac, P. Peyrillé, and C. Flamant. 2021. "Seasonal Forecasts of the Saharan Heat Low Characteristics: A Multi-Model Assessment." *Weather and Climate Dynamics* 2, no. 3: 893–912. https://doi.org/10.5194/wcd-2-893-2021.

Nicholson, S. E. 2001. "Climatic and Environmental Change in Africa During the Last Two Centuries." *Climate Research* 17, no. 2: 123–144. https://doi.org/10.3354/CR017123.

Nicholson, S. E. 2009. "A Revised Picture of the Structure of the "Monsoon" and Land ITCZ Over West Africa." *Climate Dynamics* 32, no. 7–8: 1155–1171. https://doi.org/10.1007/S00382-008-0514-3/FIGURES/18.

Nicholson, S. E., C. Funk, and A. H. Fink. 2018. "Rainfall Over the African Continent From the 19th Through the 21st Century." *Global and Planetary Change* 165: 114–127. https://doi.org/10.1016/j.gloplacha.2017.12.014.

Nicholson, S. E., and D. Klotter. 2021. "The Tropical Easterly Jet Over Africa, Its Representation in Six Reanalysis Products, and Its Association With Sahel Rainfall." *International Journal of Climatology* 41, no. 1: 328–347. https://doi.org/10.1002/JOC.6623.

Nkrumah, F., C. Klein, K. A. Quagraine, et al. 2023. "Classification of Large-Scale Environments That Drive the Formation of Mesoscale Convective Systems Over Southern West Africa." *Weather and Climate Dynamics* 4, no. 3: 773–788. https://doi.org/10.5194/WCD-4-773-2023.

Noble, W. S. 2006. "What Is a Support Vector Machine?" *Nature Biotechnology* 24, no. 12: 1565–1567. https://doi.org/10.1038/nbt1206-1565.

Omotosho, J. B., and B. J. Abiodun. 2007. "A Numerical Study of Moisture Build-Up and Rainfall Over West Africa." *Meteorological Applications* 14, no. 3: 209–225. https://doi.org/10.1002/met.11.

Paeth, H., A. H. Fink, S. Pohle, F. Keis, H. Mächel, and C. Samimi. 2011. "Meteorological Characteristics and Potential Causes of the 2007 Flood in Sub-Saharan Africa." *International Journal of Climatology* 31, no. 13: 1908–1926. https://doi.org/10.1002/joc.2199.

Parmar, A., R. Katariya, and V. Patel. 2019. "A Review on Random Forest: An Ensemble Classifier." In *International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018*, edited by J. Hemanth, X. Fernando, P. Lafata, and Z. Baig, 758–763. Springer International Publishing.

Pedregosa, F., G. Varoquaux, A. Gramfort, et al. 2012. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825–2830. http://arxiv.org/abs/1201.0490.

Pegram, G., S. Sinclair, and A. Bárdossy. 2016. "New Methods of Infilling Southern African Raingauge Records Enhanced by Annual, Monthly and Daily Precipitation Estimates Tagged With Uncertainty Report to the Water Research Commission." *Water Research Commission* 2241, no. 1: 16.

Pirret, J. S. R., J. D. Daron, P. E. Bett, N. Fournier, and A. K. Foamouhoue. 2020. "Assessing the Skill and Reliability of Seasonal Climate Forecasts in Sahelian West Africa." *Weather and Forecasting* 35, no. 3: 1035–1050. https://doi.org/10.1175/WAF-D-19-0168.1.

Pu, B., and K. H. Cook. 2010. "Dynamics of the West African Westerly Jet." *Journal of Climate* 23, no. 23: 6263–6276. https://doi.org/10.1175/2010JCLI3648.1.

Rauch, M., J. Bliefernicht, P. Laux, S. Salack, M. Waongo, and H. Kunstmann. 2019. "Seasonal Forecasting of the Onset of the Rainy Season in West Africa." *Atmosphere* 10, no. 9: 528. https://doi.org/10.3390/ATMOS10090528.

Rauch, M., J. Bliefernicht, M. Maranan, A. H. Fink, and H. Kunstmann. 2024. "Geostatistical Simulation of Daily Rainfall Fields—Performance Assessment for Extremes in West Africa." *Journal of Hydrometeorology* 25, no. 10: 1425–1442. https://doi.org/10.1175/JHM-D-23-0123.1.

Rauch, M., J. Bliefernicht, W. Sawadogo, S. Sy, M. Waongo, and H. Kunstmann. 2025. "Seasonal Prediction of Rainfall Variability for the West African Sudan-Sahel." *Frontiers in Water* 6: 1523898. https://doi.org/10.3389/frwa.2024.1523898.

Reason, C. J. C., and M. Rouault. 2006. "Sea Surface Temperature Variability in the Tropical Southeast Atlantic Ocean and West African Rainfall." *Geophysical Research Letters* 33, no. 21: 2006GL027145. https://doi.org/10.1029/2006GL027145.

Rust, H. W., M. Vrac, B. Sultan, and M. Lengaigne. 2013. "Mapping Weather-Type Influence on Senegal Precipitation Based on a Spatial–Temporal Statistical Model." *Journal of Climate* 26, no. 20: 8189–8209. https://doi.org/10.1175/JCLI-D-12-00302.1.

Salack, S., A. Bossa, J. Bliefernicht, et al. 2019. "Designing Transnational Hydroclimatological Observation Networks and Data Sharing Policies in West Africa." *Data Science Journal* 18, no. 1: 1–15. https://doi.org/10.5334/dsj-2019-033.

Santos, J. A., J. Corte-Real, and S. M. Leite. 2005. "Weather Regimes and Their Connection to the Winter Rainfall in Portugal." *International Journal of Climatology* 25, no. 1: 33–50. https://doi.org/10.1002/JOC.1101.

Sawadogo, W., J. Bliefernicht, B. Fersch, et al. 2023. "Hourly Global Horizontal Irradiance Over West Africa: A Case Study of One-Year Satellite- and Reanalysis-Derived Estimates vs. in Situ Measurements." *Renewable Energy* 216: 119066. https://doi.org/10.1016/j.renene.2023.119066.

Schlueter, A., A. H. Fink, P. Knippertz, and P. Vogel. 2019. "A Systematic Comparison of Tropical Waves Over Northern Africa. Part I: Influence on Rainfall." *Journal of Climate* 32, no. 5: 1501–1523. https://doi.org/10.1175/JCLI-D-18-0173.1.

Schunke, J., P. Laux, J. Bliefernicht, M. Waongo, W. Sawadogo, and H. Kunstmann. 2021. "Exploring the Potential of the Cost-Efficient TAHMO Observation Data for Hydro-Meteorological Applications in Sub-Saharan Africa." *Water* 13, no. 22: 3308. https://doi.org/10.3390/W13223308.

Semazzi, F. 2011. "Framework for Climate Services in Developing Countries." *Climate Research* 47, no. 1: 145–150. https://doi.org/10.3354/cr00955.

Semunegus, H., A. Mekonnen, and C. J. Schreck. 2017. "Characterization of Convective Systems and Their Association With African Easterly Waves." *International Journal of Climatology* 37, no. 12: 4486–4492. https://doi.org/10.1002/joc.5085.

Sheen, K. L., D. M. Smith, N. J. Dunstone, R. Eade, D. P. Rowell, and M. Vellinga. 2017. "Skilful Prediction of Sahel Summer Rainfall on Inter-Annual and Multi-Year Timescales." *Nature Communications* 8, no. 1: 14966. https://doi.org/10.1038/ncomms14966.

Siegmund, J., J. Bliefernicht, P. Laux, and H. Kunstmann. 2015. "Toward a Seasonal Precipitation Prediction System for West Africa: Performance of CFSv2 and High-Resolution Dynamical Downscaling." *Journal of Geophysical Research-Atmospheres* 120, no. 15: 7316–7339. https://doi.org/10.1002/2014JD022692.

Solman, S. A., and C. G. Menéndez. 2003. "Weather Regimes in the South American Sector and Neighbouring Oceans During Winter." *Climate Dynamics* 21, no. 1: 91–104. https://doi.org/10.1007/S00382-003-0320-X/FIGURES/13.

Thorncroft, C. D., H. Nguyen, C. Zhang, and P. Peyrillé. 2011. "Annual Cycle of the West African Monsoon: Regional Circulations and Associated Water Vapour Transport." *Quarterly Journal of the Royal Meteorological Society* 137, no. 654: 129–147. https://doi.org/10.1002/qj.728.

Tunde, O. 2022. "2022 Flood: 603 Dead, 1.3m Displaced Across Nigeria—Federal Govt." *Leadership Newspaper*.

van de Giesen, N., R. Hut, and J. Selker. 2014. "The Trans-African Hydro-Meteorological Observatory (TAHMO)." *WIREs Water* 1, no. 4: 341–348. https://doi.org/10.1002/wat2.1034.

Vizy, E. K., and K. H. Cook. 2002. "Development and Application of a Mesoscale Climate Model for the Tropics: Influence of Sea Surface Temperature Anomalies on the West African Monsoon." *Journal of Geophysical Research: Atmospheres* 107, no. D3: ACL2-1–ACL2-22. https://doi.org/10.1029/2001JD000686.

Vogel, P., P. Knippertz, A. H. Fink, A. Schlueter, and T. Gneiting. 2018. "Skill of Global Raw and Postprocessed Ensemble Predictions of Rainfall Over Northern Tropical Africa." *Weather and Forecasting* 33, no. 2: 369–388. https://doi.org/10.1175/WAF-D-17-0127.1.

Vollmert, P., A. H. Fink, and H. Besler. 2003. "Ghana Dry Zone und 'Dahomey Gap': Ursachen für eine Niederschlagsanomalie im tropischen Westafrika." *Die Erde* 134, no. 4: 375–393.

Wetterhall, F., A. Bárdossy, D. Chen, S. Halldin, and C. Y. Xu. 2009. "Statistical Downscaling of Daily Precipitation Over Sweden Using GCM Output." *Theoretical and Applied Climatology* 96, no. 1–2: 95–103. https://doi.org/10.1007/s00704-008-0038-0.

Wilks, D. 2006. *Statistical Methods in the Atmospheric Sciences*. second ed. Elsevier.

Zehe, E., A. K. Singh, and A. Bárdossy. 2006. "Modelling of Monsoon Rainfall for a Mesoscale Catchment in North-West India I: Assessment of Objective Circulation Patterns." *Hydrology and Earth System Sciences* 10, no. 6: 797–806. https://doi.org/10.5194/hess-10-797-2006.

Zeng, X., and T. R. Martinez. 2000. "Distribution-Balanced Stratified Cross-Validation for Accuracy Estimation." *Journal of Experimental and Theoretical Artificial Intelligence* 12, no. 1: 1–12. https://doi.org/10.1080/095281300146272.

Zorita, E., and H. von Storch. 1999. "The Analog Method as a Simple Statistical Downscaling Technique: Comparison With More Complicated Methods." *Journal of Climate* 12, no. 8: 2474–2489. https://doi.org/10.1175/1520-0442(1999)012<2474:TAMAAS>2.0.CO;2.