American Journal of Mathematical and Computer Modelling
2026; 11(1): 1-12
http://www.sciencepublishinggroup.com/j/ajmcm
doi: 10.11648/j.ajmcm.20261101.11

Modelling the Dependence Between Drought and Heatwave Extremes in Kenya Using an Integrated EVT-Copula-XGBoost Framework

Charity Mueni Mulwa*, **Herbert Imboga** , **Susan Mwelu**

Department of Statistics and Acturial Sciences, Jomo Kenyatta University of Agriculture and Technoogy (JKUAT), Nairobi , Kenya

Email address:
charitymueni028@gmail.com (Charity Mueni Mulwa), imbogaherbert@jkuat.ac.ke (Herbert Imboga), smwelu@jkuat.ac.ke (Susan Mwelu)
*Corresponding author

To cite this article:
Charity Mueni Mulwa, Herbert Imboga, Susan Mwelu. (2026). Modelling the Dependence Between Drought and Heatwave Extremes in Kenya Using an Integrated EVT-Copula-XGBoost Framework. *American Journal of Mathematical and Computer Modelling, 11*(1), 1-12. https://doi.org/10.11648/j.ajmcm.20261101.11

Received: 26 October 2025; **Accepted:** 8 November 2025; **Published:** 15 January 2026

Abstract: Compound drought-heatwave events pose serious threats to agriculture, ecosystems, and livelihoods in Kenya, where increasing climate variability amplifies their frequency and intensity. This study developed a hybrid Extreme Value Theory (EVT)-Copula-XGBoost framework to characterize and predict concurrent drought and heatwave extremes using ERA5 reanalysis data (2005-2024).The EVT component modeled the marginal tails of temperature and precipitation, revealing that temperature extremes follow a bounded Weibull-type tail, while rainfall deficits exhibit heavy tails, indicating a high potential for severe droughts. Copula modeling captured the dependence structure between drought and heatwave indices, showing weak but significant negative dependence (Kendall's $\tau = -0.189$ to 0.034), strongest during the short rains season (SON), implying that hot and dry conditions often co-occur. Joint risk analysis estimated return periods of 2.5-4.7 years, with five-year joint thresholds of 2.3-2.7 mm rainfall and 25.1-25.3°C temperature, suggesting that compound drought-heatwave events recur roughly every three years. The XGBoost model achieved high predictive skill (AUC = 0.989), with EVT and Copula derived features contributing most to performance. This hybrid framework provides a robust, data driven foundation for early detection, risk mapping, and climate adaptation planning, supporting proactive management of compound climate extremes in Kenya.

Keywords: Drought, Heatwave, Extreme Value Theory (EVT), Copula, XGBoost, Joint Risk Analysis, Kenya

1. Introduction

Climate change remains one of the most pressing challenges of the twenty-first century, with widespread consequences for ecosystems, human health, and socio-economic development [10]. The accelerated rise in global temperatures, increasing variability in precipitation patterns, and more frequent extreme weather events underscore the urgent need for improved understanding of climate-related risks [6]. These climatic changes exert significant stress on natural resources and human systems, especially in regions where livelihoods depend directly on climate-sensitive sectors [12].

Sub-Saharan Africa is particularly vulnerable to the impacts of climate change due to its limited adaptive capacity, rapid population growth, and reliance on rain-fed agriculture [19]. Kenya exemplifies these vulnerabilities, as over 80% of its land area is classified as arid or semi-arid and highly sensitive to rainfall fluctuations [15]. Agriculture contributes substantially to national income and employment, meaning that even slight variations in rainfall or temperature can have devastating economic and social impacts [18]. Reports by the World Meteorological Organization indicate that East Africa's temperatures have been rising faster than the global average, with significant shifts in rainfall distribution patterns [20]. Similarly, recent assessments highlight that droughts and heat extremes have intensified in both frequency and magnitude across Kenya and the broader Horn of Africa [18].

Kenya's rainfall regime is bimodal, characterized by long rains from March to May and short rains from

October to December. However, these seasons have become increasingly unpredictable, with notable declines in rainfall totals and changes in onset dates [15]. Droughts such as those recorded in 1984, 2000, 2011, and 2022 have caused widespread agricultural losses, water shortages, and ecosystem degradation [10]. When droughts and heatwaves occur simultaneously, their combined impacts often exceed the sum of their individual effects, leading to intensified evapotranspiration, soil moisture depletion, and crop failure [21]. The resulting consequences include diminished agricultural productivity, reduced hydropower generation, and heightened food insecurity [8]. These cascading impacts disproportionately affect vulnerable communities, contributing to rural poverty, resource conflicts, and large-scale displacement [6].

Globally, the scientific community increasingly recognizes that many high-impact climate events arise from the interaction of multiple drivers rather than single variables [1, 9]. Compound events, such as concurrent droughts and heatwaves, represent a key example of such interactions [2]. However, in most parts of Africa, climate research still treats these hazards independently, neglecting their co-dependence and joint effects [21]. Traditional indices like the Standardized Precipitation Index (SPI) or the Excess Heat Factor (EHF) are useful for detecting individual hazards but fail to capture concurrent processes or inter-variable dependencies [19]. Consequently, the actual risk of co-occurrence is often underestimated, leading to inadequate preparedness and misinformed decision-making [20]. To address this gap, researchers increasingly advocate for integrated statistical and machine learning approaches capable of modeling multivariate dependencies and extremes [17].

Extreme Value Theory (EVT) offers a fundamental statistical framework for analyzing rare and extreme phenomena by focusing on the tails of probability distributions [4]. It provides essential tools for quantifying exceedance probabilities, estimating return levels, and assessing the frequency of extreme occurrences [5]. EVT has been successfully applied to characterize hydrological, temperature, and wind extremes across diverse climatic regions [11]. However, traditional univariate EVT models do not adequately capture dependencies among multiple variables that may jointly influence extreme outcomes [7]. For compound events such as drought-heatwave interactions, a multivariate approach is necessary to represent joint tail behavior and co-occurrence probabilities [4].

Copula theory provides a flexible means of capturing and quantifying dependence structures between variables while retaining their marginal distributions [16]. Copulas enable joint probability modeling without assuming linear relationships, making them ideal for analyzing complex climate interactions [2]. They have been widely used to assess joint rainfall-runoff behavior, drought-heatwave occurrences, and flood-storm dependencies across different regions [17]. Through the copula framework, it is possible to describe the degree and type of dependence (e.g., upper or lower tail) between climate drivers, providing a more realistic understanding of multi-hazard interactions [7]. Nevertheless, copula-based models, while powerful, are often constrained when dealing with high-dimensional or dynamic dependence structures common in climate systems [11].

Machine learning (ML) techniques, particularly ensemble and boosting algorithms, have emerged as valuable tools for modeling nonlinear and high-dimensional relationships in large datasets [3]. Among these, Extreme Gradient Boosting (XGBoost) has gained prominence for its ability to handle complex interactions, improve predictive accuracy, and manage large amounts of heterogeneous data [13]. Recent advances have demonstrated the potential of combining ML and statistical frameworks to enhance the detection, modeling, and prediction of compound extremes [22]. Such hybridization leverages the interpretability and probabilistic rigor of traditional methods alongside the flexibility and predictive power of ML [23]. For instance, hybrid ML-copula frameworks have been applied in hydrological forecasting and climate risk analysis to better capture nonlinear dependencies and improve generalization performance [17].

The integration of EVT, copula theory, and XGBoost offers a comprehensive modeling approach that unites theoretical and data-driven perspectives [13]. EVT provides the statistical foundation for modeling extreme tails, copulas describe dependence between hazards, and XGBoost captures nonlinear relationships and complex feature interactions [16]. Together, these methods enable a holistic understanding of the frequency, magnitude, and joint occurrence of compound drought-heatwave events [4]. Previous research has shown that such hybrid models can significantly improve predictive skill in compound hazard analysis by effectively combining statistical structure with machine learning adaptability [24]. Moreover, entropy-based and ensemble ML-copula methods have proven capable of uncovering hidden dependencies and providing probabilistic risk estimates with greater accuracy [25].

Although hybrid EVT-Copula-ML frameworks have been successfully applied in Europe, Asia, and North America, their application in Sub-Saharan Africa remains limited [19]. Data scarcity, inadequate computational resources, and limited research capacity have constrained the advancement of such methods in many African contexts [20]. Nevertheless, the growing availability of reanalysis datasets such as ERA5, along with open-source statistical and machine learning tools, has opened new possibilities for regional-scale compound event studies [15]. Kenya, with its diverse climatic gradients ranging from humid highlands to arid lowlands, provides a valuable context for testing and applying such advanced frameworks [18].

This study therefore develops an integrated EVT-Copula-XGBoost modeling framework to characterize and predict compound drought-heatwave events in Kenya [3, 5, 14]. The proposed framework combines the tail modeling strength of EVT, the dependence modeling flexibility of copulas, and the predictive capability of XGBoost [16, 17]. Through this integration, the model aims to capture the joint probability, intensity, and spatial-temporal dynamics of compound extremes across Kenyan counties [13, 18, 20].

Furthermore, this framework advances the understanding of compound event risk by enabling multi-dimensional modeling of drought and heatwave interactions under varying climatic conditions [22]. By integrating machine learning into an EVT-Copula structure, the framework is capable of learning from complex climatic signals and extracting relationships that traditional parametric models often overlook [23]. The model also facilitates probabilistic forecasting, which is essential for early warning systems and adaptive management in regions with limited observation networks [25].

In addition, the hybrid design contributes to policy relevance by providing a scientific foundation for evidence-based climate adaptation planning [6]. The model's capacity to quantify co-occurrence probabilities can guide decision-makers in prioritizing interventions such as irrigation infrastructure, crop diversification, and drought preparedness programs [19]. This research thus bridges methodological, computational, and applied dimensions of compound climate risk modeling, offering a replicable framework for other data-scarce regions facing similar multi-hazard threats [1, 2]. By enhancing analytical capability and predictive precision, the study contributes toward strengthening climate resilience and long-term sustainability in Kenya and beyond [7, 10].

2. Methodology

2.1. Data Source and Description

This study used the monthly rainfall and 2-m air temperature of the ERA5 reanalysis dataset (2005–2024) at a spatial resolution of $0.25^\circ \times 0.25^\circ$. The data were averaged into the four climatological seasons namely, December to February (DJF), March to May (MAM), June to August (JJA), and September to November (SON) to establish seasonal patterns of extremes. Seasonal maxima of temperature and minima of precipitation were utilized in modeling heatwave and drought events, respectively, under the Extreme Value Theory (EVT) framework. The seasonal approach provides a statistically significant and policy relevant scale in which compound climate extremes can be assessed in Kenya with sufficient data to allow for rare event estimation reliably.

2.2. Extreme Value Theory (EVT)

Extreme Value Theory (EVT) provides a robust statistical basis for modelling the tails of climate distributions, where Gaussian assumptions fail to capture the behavior of extremes. In this study, EVT was used to estimate the probability and return levels of record breaking heatwaves and droughts, forming the univariate foundation for subsequent dependence modeling.

Let $\{X_1, X_2, \ldots, X_n\}$ denote independent and identically distributed (i.i.d.) seasonal climate observations with distribution function $F(x)$. The block maximum is defined as:

$$M_n = \max(X_1, \ldots, X_n), \tag{1}$$

and under suitable normalization $a_n > 0$ and $b_n \in \mathbb{R}$, the Fisher–Tippett–Gnedenko theorem states that

$$P\left(\frac{M_n - b_n}{a_n} \le x\right) \to G(x), \quad n \to \infty, \tag{2}$$

where $G(x)$ follows the Generalized Extreme Value (GEV) distribution:

$$\begin{gathered} G(x; \mu, \sigma, \xi) = \exp\left\{-\left[1 + \xi \frac{(x-\mu)}{\sigma}\right]^{-1/\xi}\right\}, \\ 1 + \xi \frac{(x-\mu)}{\sigma} > 0. \end{gathered} \tag{3}$$

Here, μ, $\sigma > 0$, and ξ are the location, scale, and shape parameters, respectively, with special cases corresponding to Gumbel ($\xi = 0$), Fréchet ($\xi > 0$), and Weibull ($\xi < 0$) families.

Seasonal maximum temperature $T_s^{\max}$ and minimum precipitation $P_s^{\min}$ were modeled as:

$$F_T(x; \mu_T, \sigma_T, \xi_T) = \exp\left\{-\left[1 + \xi_T \frac{(x-\mu_T)}{\sigma_T}\right]^{-1/\xi_T}\right\}, \tag{4}$$

$$\begin{gathered} F_P(y; \mu_P, \sigma_P, \xi_P) = \exp\left\{-\left[1 + \xi_P \frac{(y-\mu_P)}{\sigma_P}\right]^{-1/\xi_P}\right\}, \\ y = -P_s^{\min}. \end{gathered} \tag{5}$$

Parameters $\theta = (\mu, \sigma, \xi)$ were estimated via Maximum Likelihood Estimation (MLE):

$$\ell(\mu, \sigma, \xi) = \sum_{i=1}^{n} \log f(x_i; \mu, \sigma, \xi), \tag{6}$$

where

$$\begin{aligned} f(x; \mu, \sigma, \xi) = \frac{1}{\sigma} &\left(1 + \xi \frac{x-\mu}{\sigma}\right)^{-\left(\frac{1}{\xi}+1\right)} \\ &\times \exp\left\{-\left(1 + \xi \frac{x-\mu}{\sigma}\right)^{-1/\xi}\right\}. \end{aligned} \tag{7}$$

Return levels z_T for T-season events were computed as:

$$\begin{aligned} z_T &= \mu + \frac{\sigma}{\xi}\Big[(-\log(1-1/T))^{-\xi} - 1\Big], \quad \xi \neq 0, \\ z_T &= \mu - \sigma \log\big(-\log(1-1/T)\big), \quad \xi \to 0. \end{aligned} \tag{8}$$

Goodness-of-fit was evaluated using Q–Q and P–P plots, the Kolmogorov–Smirnov statistic:

$$D_n = \sup_x |F_n(x) - F(x)|, \tag{9}$$

The univariate EVT analysis provided tail characterizations of individual hazards before their interdependence was modeled via copulas.

2.3. Copula Modelling of Dependence

To quantify the dependence between drought and heatwave extremes, copula functions were applied. According to Sklar's theorem, any joint cumulative distribution function (CDF) $H(x, y)$ can be represented as:

$$H(x, y) = C\big(F_X(x), F_Y(y); \theta\big), \tag{10}$$

where $C(\cdot, \cdot; \theta)$ is a copula parameterized by θ. The corresponding joint probability density function (PDF) is:

$$h(x, y) = c\big(F_X(x), F_Y(y); \theta\big) f_X(x) f_Y(y), \tag{11}$$

where $c(u, v; \theta) = \frac{\partial^2 C(u,v;\theta)}{\partial u \partial v}$ is the copula density.

Multiple copula families were tested to capture diverse dependence structures, including the Independence, Clayton, Gumbel, Frank, and Joe copulas. Parameters were estimated via maximum likelihood:

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} \log c\big(F_X(x_i), F_Y(y_i); \theta\big), \tag{12}$$

and the optimal copula was selected using the Akaike Information Criterion (AIC) and empirical dependence diagnostics. The 270° rotated Clayton and Joe copulas effectively captured asymmetric upper–lower tail dependencies characteristic of compound drought–heatwave behavior.

2.4. Joint Risk Analysis

After modeling marginal and joint dependencies, a joint risk framework was constructed to quantify the probability and recurrence of compound extremes. The joint survival function is given by:

$$\begin{aligned} \bar{H}(x, y) &= P(X > x, Y > y) \\ &= 1 - F_X(x) - F_Y(y) + C(F_X(x), F_Y(y); \theta), \end{aligned} \tag{13}$$

which represents the probability that precipitation is below a drought threshold and temperature is above a heatwave threshold. The joint return period is obtained as:

$$RP(x, y) = \frac{1}{\bar{H}(x, y)}. \tag{14}$$

Bootstrap resampling (B = 10,000) was also used to estimate uncertainty in joint return levels, with 95% confidence intervals which were derived from empirical percentiles. Seasonal risk measures for DJF, MAM, JJA, and SON were calculated to account for Kenya's climatic variability across the seasons.

2.5. XGBoost-Based Prediction of Compound Extremes

In the final stage of the proposed EVT–Copula–XGBoost framework, the probabilistic features obtained from the marginal and dependence analyses are used to predict compound drought–heatwave occurrences. For each observation i in season s, the predictor vector is defined as

$$\mathbf{x}_i = \big[p_i^{(D)},\ p_i^{(H)},\ c_i,\ \chi_i^{(s)},\ d_{i-1},\ h_{i-1},\ \bar{d}_{i,2},\ \bar{h}_{i,2},\ S_i\big], \tag{15}$$

where the predictor components are defined as:

$$\begin{aligned} p_i^{(D)} &= F_{\text{GEV}}^{(D)}(x_i), \\ p_i^{(H)} &= 1 - F_{\text{GEV}}^{(H)}(x_i), \\ c_i &= f_C\left(p_i^{(D)}, p_i^{(H)}; \boldsymbol{\theta}_s\right), \\ \chi_i^{(s)} &\ \text{is the tail dependence coefficient,} \\ d_{i-1}, h_{i-1} &\ \text{are lagged drought and heatwave values,} \\ \bar{d}_{i,2}, \bar{h}_{i,2} &\ \text{are two-period moving averages,} \\ S_i &\ \text{denotes the seasonal indicator variable.} \end{aligned}$$

Specifically, $p_i^{(D)}$ denotes the drought non-exceedance probability from the GEV fit, $p_i^{(H)}$ the heatwave exceedance probability, c_i the Copula-based joint density capturing drought-heatwave dependence, $\chi_i^{(s)}$ the seasonal tail-dependence coefficient, d_{i-1} and h_{i-1} the one-period lagged drought and heatwave indices, $\bar{d}_{i,2}$ and $\bar{h}_{i,2}$ the two-period moving averages, and S_i the seasonal indicator variable.

Let $y_i \in \{0, 1\}$ represent the occurrence of a compound drought–heatwave event, where $y_i = 1$ indicates occurrence and $y_i = 0$ otherwise. The XGBoost model predicts the probability of occurrence as

$$\hat{y}_i = P(y_i = 1 \mid \mathbf{x}_i) = \sigma\left(\sum_{k=1}^{K} f_k(\mathbf{x}_i)\right), \tag{16}$$

where $\sigma(z) = \frac{1}{1+e^{-z}}$ is the logistic activation function, and each $f_k \in \mathcal{F}$ represents a regression tree in the ensemble.

The model parameters are obtained by minimizing the regularized objective function

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} \ell(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k), \tag{17}$$

where

$$\begin{aligned} \ell(y_i, \hat{y}_i) &= -\big[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)\big], \\ \Omega(f_k) &= \gamma T_k + \tfrac{1}{2}\lambda \|w_k\|^2, \end{aligned} \tag{18}$$

with T_k denoting the number of leaves and w_k the leaf weights of tree f_k. To handle class imbalance, a weighting factor is introduced as

$$\text{scale\_pos\_weight} = \frac{N_0}{N_1}, \tag{19}$$

where N_0 and N_1 represent the number of non-extreme and extreme events, respectively.

The trained XGBoost model learns a nonlinear mapping between the EVT–Copula-derived predictors and the probability of compound extreme occurrence:

$$f_{\text{XGB}} : \left(F_{\text{GEV}}^{(D)},\, F_{\text{GEV}}^{(H)},\, f_C,\, \chi,\, \mathbf{Z}\right) \longrightarrow P(y=1), \tag{20}$$

where $\mathbf{Z} = [d_{t-1}, h_{t-1}, \bar{d}_{t,2}, \bar{h}_{t,2}, S_t]$ encapsulates the temporal and seasonal covariates.

Finally, the complete hybrid framework can be summarized as

$$P(y_t=1) = f_{\text{XGB}}\Big(F_{\text{GEV}}^{(D)}(x_t),\ 1-F_{\text{GEV}}^{(H)}(x_t),\ f_C(p_t^{(D)}, p_t^{(H)}; \boldsymbol{\theta}),\ \chi_t,\ \mathbf{Z}_t\Big), \tag{21}$$

where the inputs are derived from the EVT–Copula analysis and the output represents the probability of a compound drought–heatwave occurrence at time t.

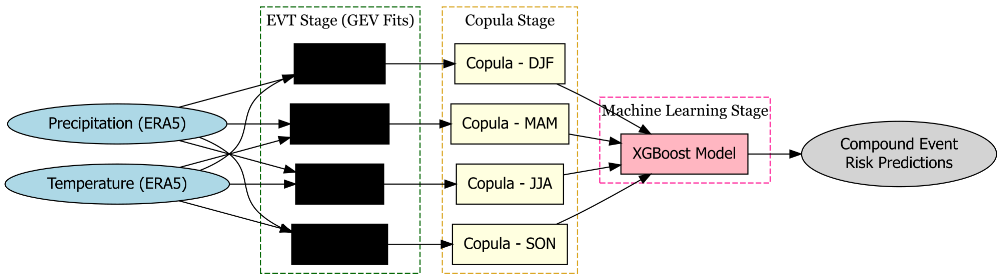

Figure 1. *A scheme representing the integrated EVT–Copula–XGBoost framework for modeling and predicting compound drought–heatwave extremes in Kenya.*

3. Data Analysis and Results

The descriptive analysis of Kenya's climate from 2005 to 2024, highlighted patterns in precipitation and temperature that influence compound drought–heatwave risk. Long-term trends and seasonal cycles were examined in order to identify high-risk periods that provide context for the subsequent extreme value and dependence modeling.

3.1. Descriptives

3.1.1. Long-Term Trends

Figure 2 below shows the trend of monthly precipitation and temperature for the period 2005 to 2024. Rainfall exhibits substantial interannual variability with a slight downward trend, while temperatures steadily increase.

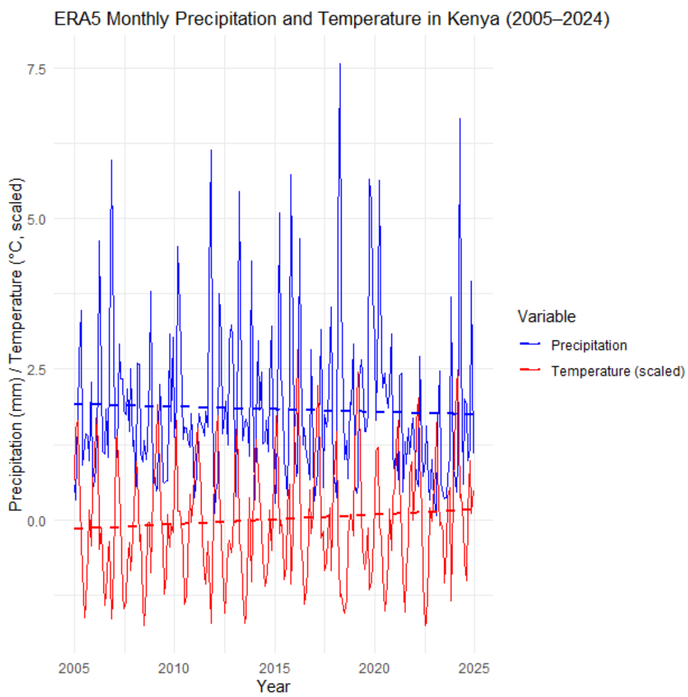

Figure 2. *Trends of monthly precipitation and temperature in Kenya (2005–2024).*

Over the 20 years, the diminishing rainfall and rising temperature brings on greater potential evapotranspiration, which amplifies water deficit and soil moisture loss. As such, even fairly moderate rainfall deficits can quickly form into drought. Occurrence of high temperatures coupled with dry spells brings on an elevated incidence of compound drought–heatwave events, which pose enormous dilemmas for agriculture, water resources, and overall climate resilience especially in the ASALs.

3.1.2. Seasonal Cycle

Kenya has two rainy seasons of long rains between March and May (MAM) and short rains between October and December (OND). The driest period is June to August (JJA), while February to March is the hottest and July the coldest. These seasonal trends produce periods of high risk for compound extremes like hot–dry from February to March and dry–cool from JJA. Recognition of these patterns is essential for targeted risk monitoring and early warning. These qualitative patterns emphasize the suitability of applying extreme value theory, copulas, and machine learning techniques, like XGBoost, to properly quantify and predict compound drought–heatwave events.

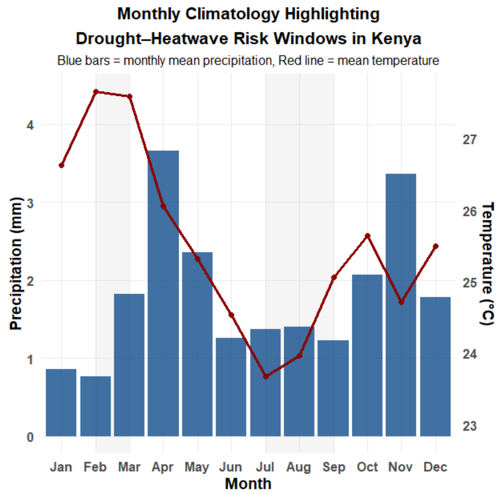

Figure 3. *Monthly precipitation (bars) and temperature (line) in Kenya (2005–2024).*

3.2. Extreme Value Modeling

The seasonal temperature maxima and precipitation minima were modeled using the Generalized Extreme Value (GEV) distribution to determine the behavior of the univariate climate extremes. Trend analysis showed no statistically significant change across any season. Therefore, the extremes were presumed stationary.

Table 1. *GEV parameters for temperature maxima and precipitation minima by season.*

Season	mu	$sigma$	xi	KS	p	mu	$sigma$	xi	KS	p
DJF	24.15	0.96	-0.26	0.063	0.955	1.507	0.983	0.233	0.034	0.988
MAM	24.08	0.93	-0.19	0.070	0.910	1.502	0.790	0.159	0.064	0.883
JJA	24.09	1.03	-0.13	0.079	0.831	1.339	0.761	0.157	0.070	0.868
SON	24.18	1.00	-0.14	0.067	0.941	1.461	0.852	0.321	0.087	0.817

The fitted GEV parameters provide useful insights into the variability of extreme rainfall and temperature events across Kenya's climatic seasons. The shape parameter (ξ) for maximum temperature is negative across all seasons. This indicates that temperature extremes are *bounded* which means that there is a physical threshold above which very high temperatures are not likely to occur (Weibull-type tail). The scale parameter (σ) is approximately 1.0 across all seasons. This implies a moderate variability in seasonal extremes. The mean parameter (μ) indicates average seasonal maxima of around 24°C. Amongst all seasons, SON and JJA experience the highest extremes, with 10-year return levels estimated between 25.6°C and 26.0°C. These results show that the most extreme hot conditions are more probable to occur in the dry and short-rain transition seasons.

For minimum precipitation, positive shape parameters ($\xi > 0$) during all seasons imply a *heavy-tailed* distribution– that is , extremely low rainfall instances (extreme droughts) are conceivable, although improbable.The SON season holds the greatest ξ value (0.321), indicating increased probabilities of extreme droughts during this short-rain season.The relatively lower μ and moderate σ values indicate that rainfall deficits during this period can be serious but spatially heterogeneous.

The Kolmogorov–Smirnov (KS) statistics and the p-values are all above 0.80, indicating that the GEV model provides a good fit to temperature and precipitation extremes.Diagnostic plots (Figure 4) also ensure that model residuals follow the theoretical distributions as required.

These results highlight that temperature extremes are bounded in Kenya, while rainfall deficits have heavier tails indicating that, drought conditions can become excessively severe, especially for SON. These findings thus serve as the foundation for follow-up steps of compound drought–heatwave event dependence and risk modeling.

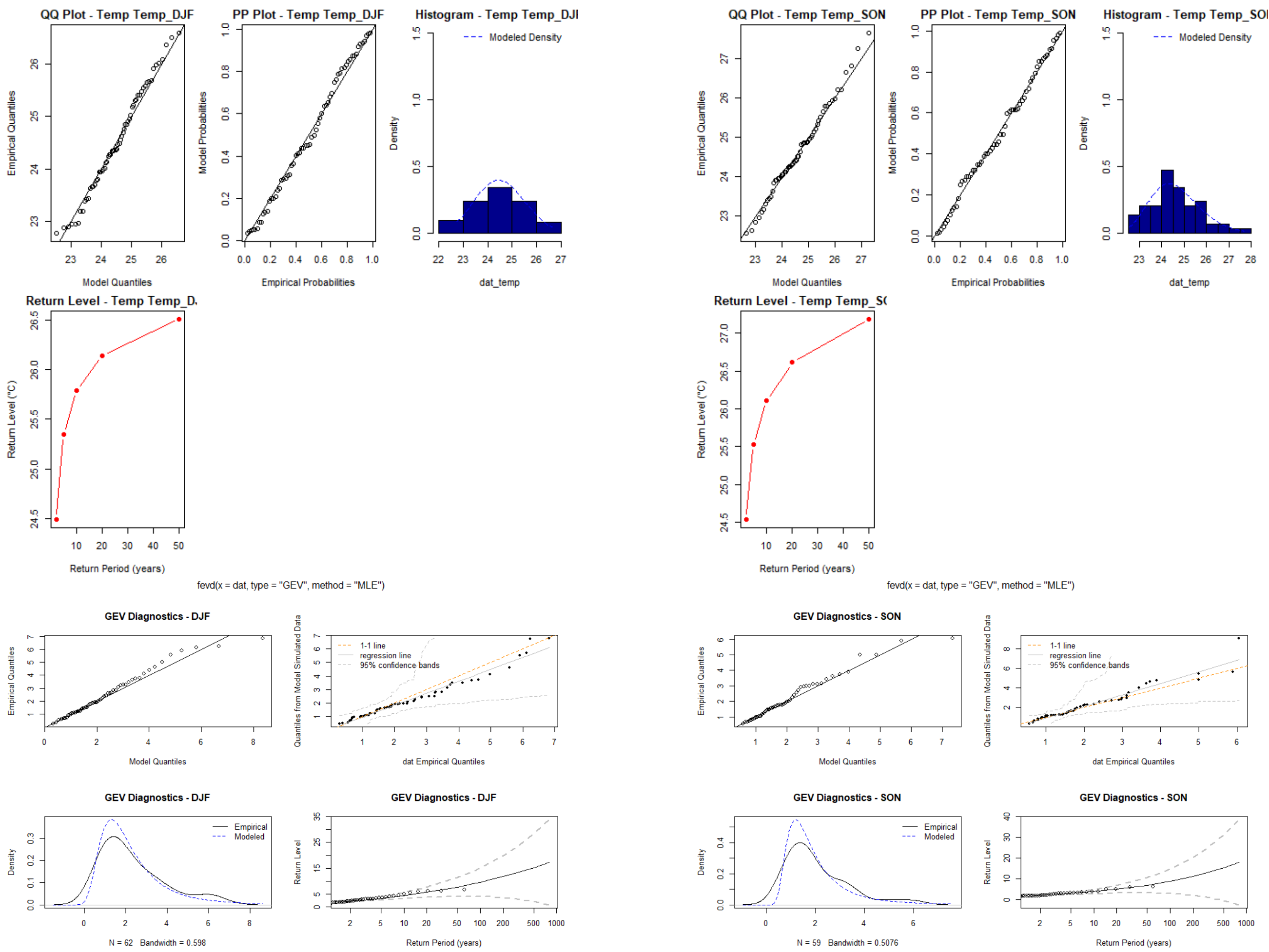

Figure 4. *A representation of GEV diagnostic plots for temperature (top) and precipitation (bottom) extremes in DJF and SON.*

3.3. Dependence Analysis with Copulas

To examine the interaction of extremely hot temperatures with minimal rainfall, we employed *copulas*, which enable flexible dependence modeling between variables with different marginal distributions. This approach reveals co-occurring patterns of drought (minimal rainfall) and heatwave (hot temperature) extremes that are not identifiable when employing conventional correlation analysis.

Table 2. *Seasonal copula summary: dependence measures and significance.*

Season	**Copula**	**Kendall's tau**	**Tail Dependence**	**Independence p**
DJF	Rotated Joe (90°)	-0.157	none	0.071
MAM	Independence	0.034	none	0.702
JJA	Rotated Clayton (270°)	-0.121	none	0.176
SON	Rotated Clayton (270°)	-0.189	none	0.035

All fitted copulas are the optimal fitting models for the respective period, chosen with the minimum AIC and goodness of fit statistics. Rotated Joe (90°) copula provided the best fit during December–February and indicated weak negative dependence ($\tau = -0.157$). This indicates hotter dry season conditions are likely to be linked with sub-average rainfall. Although the Joe family is generally fitted for upper tail dependence, its rotated version can provide lower tail dependence, which means dryness can reinforce the extremes in temperature during this period.

For March–May, the independence copula was selected, revealing no meaningful relationship ($\tau = 0.034$) between temperature and rainfall. This agrees with the dominance of large scale ocean atmosphere systems that control the long rains, with little influence from local temperature.

The June–August quarter was best described by the rotated Clayton (270°) copula, which captured weak but persistent negative dependence ($\tau = -0.121$). The Clayton family emphasizes lower tail dependence, which suggests that during the dry winter months, even fairly modest rainfall deficits have a tendency to be associated with temperatures somewhat higher than average.

A similar pattern appeared in September–November, where the rotated Clayton (270°) copula again provided the best fit, showing the strongest negative dependence ($\tau = -0.189$) and a statistically significant result ($p = 0.035$). This finding highlights that during the short rains, dry spells and high temperatures are more likely to co-occur, making SON the season most prone to compound drought–heatwave episodes in Kenya.

Although tail dependence was not detected for any season, this outcome remains insightful. It implies that the most extreme droughts and the most severe heatwaves rarely peak simultaneously. However, the consistent weak negative dependencies captured by Kendall's τ reveal that even moderate co-occurrences of heat and dryness can emerge seasonally, influencing overall climate stress.

These findings are essential for the next step of predicting compound extremes using the XGBoost model. While copulas quantify dependence patterns, machine learning enables us to leverage these relationships, along with other environmental predictors, to forecast the likelihood and intensity of concurrent drought–heatwave conditions across Kenya.

3.4. Joint Risk Diagnostics

Understanding how often compound drought–heatwave events occur together is crucial for climate risk assessment and preparedness. This subsection translates the dependence results into practical measures of how frequently joint extremes happen and what rainfall and temperature levels characterize such high-risk conditions.

3.4.1. Joint Return Periods

Joint return periods provide a probabilistic measure of how frequently drought and heatwave extremes co-occur within a given season. They indicate the expected time interval between successive compound events that exceed pre-defined severity thresholds. A shorter return period therefore reflects higher co-occurrence likelihood, implying more persistent climate stress.

Results in Table 3 reveal that compound drought–heatwave events are recurrent features of Kenya's seasonal climate. Across all seasons, joint return periods range between 2.5 and 4.7 years based on empirical and copula-based estimates, highlighting that such extremes are not rare anomalies but moderately frequent occurrences. The bootstrap-derived median return periods (10–15 years) are longer, reflecting the adjustment for sampling variability and uncertainty in dependence estimation. The wide 95% confidence intervals further emphasize that true recurrence intervals may vary substantially across years depending on prevailing large-scale drivers such as the El Niño–Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD).

Seasonal differences are also evident. The *MAM* season shows the shortest return period (approximately 2.5 years), implying that compound extremes during the long rains are relatively frequent, likely due to high interannual rainfall variability and rapid temperature rises preceding the onset of the wet season. In contrast, *DJF* exhibits the longest return period (around 4.7 years), indicating lower but still notable co-occurrence risk during this transition season. The *JJA* and *SON* seasons fall in between, suggesting that drought–heatwave linkages persist even outside the main rainy periods.

Overall, these diagnostics indicate that Kenya experiences a compound drought–heatwave event roughly every 3 years on average. Such frequency underscores the importance of integrated drought and heat early-warning systems, as recurrent compound stressors can erode ecosystem resilience, intensify water scarcity, and undermine agricultural productivity even in years without record-breaking extremes.

Table 3. *Joint return period diagnostics for compound drought–heatwave events.*

Season	Empirical RP	Copula RP	Bootstrap Median	95% CI
DJF	3.9	4.7	15.5	[7.8, 62]
MAM	2.5	2.5	10.0	[5.5, 30]
JJA	3.0	4.1	11.8	[6.6, 59]
SON	3.0	3.2	11.8	[5.9, 59]

3.4.2. Five-Year Thresholds

A five-year joint return period represents the intensity of rainfall and temperature expected to occur together roughly once every five years. These joint conditions correspond to very low rainfall (about 2.3–2.7 mm) and high temperatures (around 25.1–25.3°C), as summarized in Table 4. Such thresholds offer practical benchmarks for developing early-warning systems and climate adaptation strategies.

Table 4. *Five-year joint return period thresholds.*

Season	Precip. (mm)	Temp. (°C)
DJF	2.74	25.08
MAM	2.71	25.22
JJA	2.35	25.25
SON	2.34	25.23

Several important insights emerge from these diagnostics. Compound drought–heatwave events occur moderately often (every 2–5 years) but have disproportionately high impacts. The SON season (October–December) shows the highest drought-related risk, while JJA (June–August) remains persistently dry. The absence of tail dependence implies that the most severe droughts and heatwaves rarely peak together, though moderate co-occurrences still drive significant climate stress.

These joint risk insights provide a solid foundation for the next stage, *predicting compound extremes using the*

XGBoost model, which integrates these dependencies with environmental predictors to forecast future drought–heatwave likelihood and severity.

3.5. Prediction of Rare Events Using XGBoost

This section evaluates how effectively the integrated EVT–Copula–XGBoost framework predicts rare compound drought–heatwave events. The model integrates EVT-derived drought and heat probabilities, copula-based dependence metrics, and lagged climatic predictors. Despite the low occurrence of such joint extremes (about 2%), the model exhibits strong predictive ability.

3.5.1. Model Performance

The XGBoost model achieved high predictive skill with a testing AUC of 0.989 and balanced accuracy of 0.989 (Table 5). High sensitivity ensures that the few existing compound events are captured, while strong specificity minimizes false alarms.

Table 5. *Performance metrics for the XGBoost model predicting compound drought–heatwave events.*

Dataset	Accuracy	Sensitivity	Specificity	Kappa
Training	0.969	0.714	0.978	0.533
Testing	0.979	1.000	0.979	0.657

Table 6. *Confusion matrices for training and testing datasets.*

	Training		Testing	
	Predicted No	Predicted Yes	Predicted No	Predicted Yes
Actual No	TN = 181	FP = 4	TN = 45	FP = 1
Actual Yes	FN = 2	TP = 5	FN = 0	TP = 2

The confusion matrices make it clear that during testing, all compound drought–heatwave events were correctly identified (no false negatives), confirming high sensitivity and reliability for early warning. A few false positives indicate slight overprediction-a favorable bias in risk management contexts where missing an event is costlier than a false alarm.

3.5.2. Feature Importance

Feature importance analysis (Figure 5) reveals that EVT-based drought and heat probabilities contributed most to predictive power, followed by copula joint probabilities and short-term climatic lags. This indicates that both the magnitude of extremes and their interdependence drive compound risk patterns.

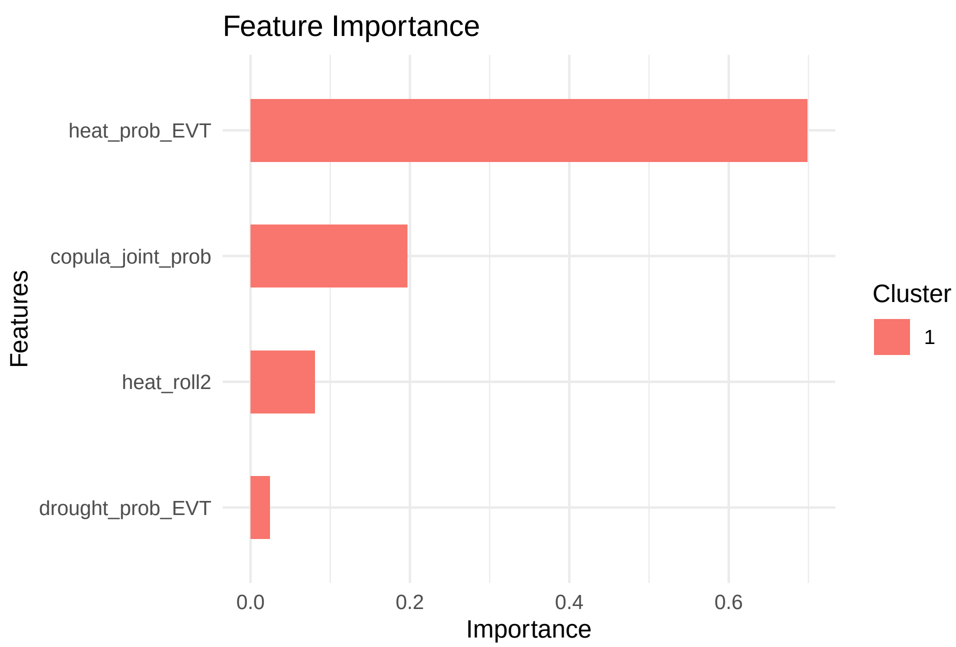

Figure 5. *Feature importance from XGBoost model highlighting the dominant role of EVT and copula features in predicting compound events.*

XGBoost accurately detects rare compound drought-heatwave events. EVT and copula-based predictors provide interpretability and robustness. Sensitivity and AUC values confirm strong early-warning potential. These findings provide a reliable basis for constructing the Compound Climate Risk Index (CCRI).

4. Conclusion and Recommendations

This study employed an integrated Extreme Value Theory (EVT)-Copula-XGBoost framework to analyze and predict compound drought-heatwave events across Kenya from 2005 to 2024. The univariate EVT analysis showed that temperature extremes are bounded, while precipitation extremes display heavy lower tails, particularly during the SON season, indicating increased drought susceptibility. Copula dependence modeling revealed weak but meaningful co-movements between drought and heatwave extremes, with Kendall's τ ranging from -0.189 to 0.034, suggesting asymptotic independence but notable co-occurrence at finite return levels. Joint risk diagnostics confirmed that compound extremes, though rare, are climatically significant, with precipitation and temperature thresholds of 2.34-2.74 mm and 25.08-25.25°C for a 5-year return period. The XGBoost model achieved outstanding performance (AUC = 0.989, balanced accuracy = 0.989), effectively identifying rare compound events with minimal false alarms. EVT-based and copula-derived probabilities were the most influential predictors, demonstrating the value of combining statistical extremal analysis with machine learning for robust climate-risk prediction. Overall, the proposed EVT-Copula-XGBoost framework provides both scientific and operational value by improving the understanding and forecasting of compound extremes. Meteorological and disaster-management agencies are encouraged to integrate EVT- and copula-informed thresholds into early-warning systems, especially during SON and JJA when compound risks are elevated. Future work should incorporate non-stationary models and higher-resolution data to capture evolving climate trends and enhance local-scale prediction accuracy. The framework is adaptable and transferable, offering a replicable foundation for compound climate-risk assessment and early-warning development in Kenya and beyond.

Data Availability

The datasets used in this study originate from the *ERA5 reanalysis* archive, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) under the Copernicus Climate Change Service (C3S). ERA5 provides hourly estimates of a wide range of atmospheric, land, and oceanic variables at a horizontal resolution of 0.25°. For this study, monthly mean temperature and total precipitation data spanning the period 2005–2024 were extracted and spatially averaged over Kenya to represent national-scale climate variability. These data were obtained from the *Copernicus Climate Data Store (CDS)*, which offers free and open access to users worldwide.The ERA5 dataset can be accessed directly from the CDS portal at https://cds.climate.copernicus.eu/ All data processing and statistical analyses were performed in the `R` environment. The processed datasets supporting the results of this study are available from the corresponding author upon reasonable request.

ORCID

0009-0008-9238-2996 (Charity Mueni Mulwa)

Abbreviations

AIC	Akaike Information Criterion
ASALs	Arid and Semi-Arid Lands
AUC	Area Under the Receiver Operating Characteristic Curve
BIC	Bayesian Information Criterion
CDS API	Copernicus Data Store Application Programming Interface
DEM	Digital Elevation Model
DJF	December–January–February (season)
EHF	Excess Heat Factor
ENSO	El Niño–Southern Oscillation
ERA5	ECMWF Reanalysis v5
ECMWF	European Centre for Medium-Range Weather Forecasts
EVT	Extreme Value Theory
GEV	Generalized Extreme Value (distribution)
GPD	Generalized Pareto Distribution
IOD	Indian Ocean Dipole
IPCC	Intergovernmental Panel on Climate Change
JJA	June–July–August (season)
KS	Kolmogorov–Smirnov (test)
MLE	Maximum Likelihood Estimation
MAM	March–April–May (season)
MJO	Madden–Julian Oscillation
NDMA	National Drought Management Authority
NDVI	Normalized Difference Vegetation Index
OND/SON	October–November–December / September–October–November (seasons)
PDF	Probability Density Function
Q–Q	Quantile–Quantile (plot)
RP	Return Period
RMSE	Root Mean Square Error
SHAP	Shapley Additive Explanation
SMOTE	Synthetic Minority Oversampling Technique
SPI	Standardized Precipitation Index
SPEI	Standardized Precipitation Evapotranspiration Index
TN, TP, FP, FN	True Negative, True Positive, False Positive, False Negative
XGBoost	Extreme Gradient Boosting (Machine Learning Algorithm)

Acknowledgments

I am deeply grateful to the Almighty God for His protection and guidance throughout the study and in the completion of this project. I extend my sincere thanks to my supervisors, Dr. Herbert Imboga and Dr. Susan Mwelu, for their invaluable guidance, encouragement, and constructive feedback throughout this work. Special appreciation goes to my family for their unwavering love, prayers, and moral support during this academic journey. I also acknowledge my friends and colleagues for their encouragement, insightful discussions, and support, which greatly contributed to the progress of this study. Finally, I wish to acknowledge the researchers and institutions whose data and prior research were instrumental to this work; without their contributions, this study would not have been possible.

Author Contributions

Charity Mueni Mulwa: Conceptualization, Data curation, Formal Analysis , Funding acquisition, Investigation, Methodology, Project administration, Resources, Software , Validation, Visualization, Writing - original draft, Writing - review & editing

Ethics Clearance

This study utilized publicly available climate data from the ERA5 reanalysis archive, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) under the Copernicus Climate Change Service (C3S). The data are provided under the C3S licensing terms, permitting free use for research purposes, ensuring that the study fully complies with ethical and legal standards for data use.

Conflicts of Interest

The authors declare that there are no conflicts of interest related to this study.

References

[1] Bevacqua, E., Lall, U., Maraun, D., Friederichs, P., & Lehner, F. (2019). Higher probability of compound flooding from precipitation and storm surge in Europe under anthropogenic climate change. *Science Advances, 5*(9), eaaw5531. https://doi.org/10.1126/sciadv.aaw5531

[2] Bevacqua, E., Maraun, D., Hobæk Haff, I., Widmann, M., & Vrac, M. (2017). Multivariate statistical modelling of compound events via pair-copula constructions: Analysis of floods in Ravenna (Italy). *Hydrology and Earth System Sciences, 21*(5), 2701–2723. https://doi.org/10.5194/hess-21-2701-2017

[3] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

[4] Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer. https://doi.org/10.1007/978-1-4471-3675-0

[5] Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling Extremal Events: For Insurance and Finance*. Springer. https://doi.org/10.1007/978-3-642-33483-2

[6] Food and Agriculture Organization of the United Nations. (2018). *The Impact of Disasters and Crises on Agriculture and Food Security*. FAO. https://www.fao.org/3/I8656EN/i8656en.pdf

[7] Genest, C., Rémillard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and Economics, 44*(2), 199–213. https://doi.org/10.1016/j.insmatheco.2007.07.002

[8] Hao, Z., Singh, V. P., Xia, Y., & Sun, F. (2020). Impact of dependence changes on the likelihood of hot extremes under drought conditions in the United States. *Journal of Hydrology, 581*, 124410. https://doi.org/10.1016/j.jhydrol.2019.124410

[9] Imboga, H. O., Orwa, G. O., & Otieno, R. O. (2017). Optimal nonparametric regression estimation of finite population total using Nadaraya-Watson incorporating jackknifing. *International Journal of Theoretical and Applied Mathematics, 3*(3), 102–110. https://doi.org/10.11648/j.ijtam.20170303.14

[10] Intergovernmental Panel on Climate Change. (2021). *Sixth Assessment Report*. Intergovernmental Panel on Climate Change. https://www.ipcc.ch/assessment-report/ar6/

[11] Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources, 25*(8–12), 1287–1304. https://doi.org/10.1016/S0309-1708(02)00056-8

[12] Kenya National Bureau of Statistics. (2020). *Economic Survey 2020*. Government Printer. https://www.knbs.or.ke/wp-content/uploads/2023/09/2020-Economic-Survey.pdf

[13] Koh, J. (2023). Gradient boosting with extreme-value theory for wildfire prediction. *Extremes, 26*(2), 273–299. https://doi.org/10.1007/s10687-022-00454-6

[14] Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Stationarity is dead: Whither water management? *Science, 319*(5863), 573–574. https://doi.org/10.1126/science.1151915

[15] National Drought Management Authority. (2019). *Kenya Drought Early Warning Bulletin: April 2019*. National Drought Management Authority. https://reliefweb.int/report/kenya/national-drought-early-warning-bulletin-april-2019

[16] Nelsen, R. B. (2006). *An Introduction to Copulas*. Springer. https://doi.org/10.1007/0-387-28678-0

[17] Song, M., Zhang, J., Liu, Y., Liu, C., Bao, Z., Jin, J., He, R., Bian, G., & Wang, G. (2024). Time-varying copula-based compound flood risk assessment of extreme rainfall and high water level under a non-stationary environment. *Journal of Flood Risk Management, 17*(2), e13032. https://doi.org/10.1111/jfr3.13032

[18] United Nations Environment Programme. (2021). *Kenya Country Profile: Climate Change Overview*. United Nations Environment Programme. https://dicf.unepgrid.ch/kenya/climate-change

[19] United Nations Office for the Coordination of Humanitarian Affairs. (2017). *Horn of Africa: Humanitarian Impacts of Drought – Issue 2 (13 April 2017)*. ReliefWeb. https://reliefweb.int/report/somalia/horn-africa-humanitarian-impacts-drought-issue-2-13-april-2017

[20] World Meteorological Organization. (2022). *State of the Climate in Africa 2022*. World Meteorological Organization. https://wmo.int/publication-series/state-of-climate-africa

[21] Zscheischler, J., & Seneviratne, S. I. (2018). Dependence of drivers affects risks associated with compound events. *Science Advances, 3*(6), e1700263. https://doi.org/10.1126/sciadv.1700263

[22] Jiang, T., Su, X., Zhang, G., Zhang, T., & Wu, H. (2023). Estimating propagation probability from meteorological to ecological droughts using a hybrid machine learning-copula method. *Hydrology and Earth System Sciences, 27*(2), 559–577. https://doi.org/10.5194/hess-27-559-2023

[23] Liu, R., Yin, J., Slater, L., Kang, S., Yang, Y., Liu, P., Guo, J., Gu, X., Zhang, X., & Volchak, A. (2024). Machine-learning-constrained projection of bivariate hydrological drought magnitudes and socioeconomic risks over China. *Hydrology and Earth System Sciences, 28*, 3305–3326. https://doi.org/10.5194/hess-28-3305-2024

[24] Alencar, P. H. L., & Paton, E. N. (2024). Which droughts are becoming more frequent? A copula-entropy analysis on the return period of droughts in Europe. *Natural Hazards, 121*, 543–565. https://doi.org/10.1007/s11069-024-06848-y

[25] M., et al. (2024). Forecasting of meteorological drought using ensemble and machine learning models. *Environmental Sciences Europe, 36*, 160. https://doi.org/10.1186/s12302-024-00975-w