2024-09

Modeling non-stationarity in extreme rainfall data and implications for climate adaptation: A case study from southern highlands region of Tanzania

Masanja, Verdiana

Elsevier

https://doi.org/10.1016/j.sciaf.2024.e02321

Scientific African 25 (2024) e02321

Contents lists available at ScienceDirect

Scientific African

journal homepage: www.elsevier.com/locate/sciaf

Modeling non-stationarity in extreme rainfall data and implications for climate adaptation: A case study from southern highlands region of Tanzania

Erick A. Kyojo [a,c,*], Silas Mirau [a], Sarah E. Osima [b], Verdiana G. Masanja [a]

[a] Department of Applied Mathematics and Computational Sciences, Nelson Mandela African Institution of Science and Technology, P.O. Box 447, Arusha, Tanzania

[b] Tanzania Meteorological Authority (TMA), P.O. Box 3056, Dar Es Salaam, Tanzania

[c] University of Dar es Salaam, Mkwawa University College of Education (MUCE), P.O. Box 2513, Iringa, Tanzania

ARTICLE INFO

Editor name: Ahmed Kenawy

Keywords:
Modeling
Extreme rainfall
Generalized extreme value
Climate change
Intensity-duration-frequency
Stationary
Non-stationary

ABSTRACT

The Southern Highlands region of Tanzania has witnessed an increased frequency of severe flash floods. This study examines rainfall data of four stations (Iringa, Mbeya, Rukwa, and Ruvuma) spanning 30 years (1991–2020) to investigate drivers of extreme rainfall and non-stationarity behavior. The Generalized Extreme Value (GEV) model, commonly used in hydrological studies, assumes constant distribution parameters, which may not be true due to climate variability, potentially leading to bias in extreme quantile estimation. Recent studies have introduced a technique for constructing non-stationary Intensity-Duration-Frequency (IDF) rainfall curves. The method incorporates trends in the parameters of the GEV distribution, only using time as a covariate. However, uncertainty exists about whether time is the most suitable covariate, highlighting the need to explore all potential covariates for modeling non-stationarity. The aim of this study is to assess the influence of other time-varying covariates on extreme daily rainfall events, considering seasonality and climate change in the rainfall data. Specifically, five processes (i.e., local temperature changes (LTC), urbanization, annual Global Temperature Anomaly (GTA), the Indian Ocean Dipole (IOD), and the El Niño-Southern Oscillation (ENSO) cycle) were studied as drivers of extreme rainfall events. Sixty two non-stationary GEV models are developed based on these covariates and their combinations, alongside two non-stationary GEV models using the time covariate to capture the seasonality of the unimodal rainfall in the region, and one stationary GEV model (S0). With the use of corrected Akaike Information Criterion (AICc), the best model for each duration (i.e., 1-, 3-, and 5-days) of rainfall series is chosen. Results indicate that local processes (i.e., LTC and urbanization) are the optimal covariates for 1 day-duration rainfall, while global processes (i.e, IOD, ENSO cycle, and GTA) are identified as the most suitable covariates for 3, and 5 day-duration rainfall across all stations. The identified best non-stationary model (with their best covariates) are then used to develop non-stationary rainfall IDF curves for all stations. According to the analysis of non-stationary extreme values, the return periods of extreme rainfall events concluded a notable decrease in comparison to the stationary approach. The study also revealed strong correlations between global climate indices (ENSO, IOD, GTA) and long-duration extreme rainfall in Tanzania's Southern Highlands. Local factors like Urbanization and temperature changes also show significant associations with 1-day duration events. These findings emphasize

* Corresponding author at: Department of Applied Mathematics and Computational Sciences, Nelson Mandela African Institution of Science and Technology, P.O. Box 447, Arusha, Tanzania.
E-mail address: kyojoe@nm-aist.ac.tz (E.A. Kyojo).

https://doi.org/10.1016/j.sciaf.2024.e02321
Received 4 November 2023; Received in revised form 4 July 2024; Accepted 11 July 2024
Available online 16 July 2024

the need for integrated climate forecasting to inform effective adaptation strategies. Finally, the study addresses associated uncertainties in our predictions of forthcoming extreme rainfall events through rigorous analysis. The study demonstrated that return levels for extreme rainfall events exhibit a rising trend with increasing return period, indicating heightened intensity over longer time spans, whereas, a relative uncertainty analysis illustrate escalating uncertainty with increasing return periods, emphasizing challenges in long-term prediction.

Introduction

The most immediate impact of extreme rainfall in Tanzania and many other parts of the world is the potential of flooding which renders fatal consequences on economy, livelihood, and ecosystem [1–3]. According to [4–6], changes in extreme climatic events have been observed since 1950, with an increase in the frequency and magnitude of climate extremes, such as droughts, floods, and heat waves, causing severe damage and loss of life and properties worldwide. Recent IPCC reports confirm an overall decrease in cold days and nights and an increase in warm days and nights, while heavy precipitation is expected to increase in the 21st century [4,7,8]. Many regions in Africa such as Southern Highlands region in Tanzania, play a vital economic role yet they experience an increasing incidence of extreme climatic events, including floods and droughts [1,9,10]. These events have severe socio-economic and ecological implications, especially for developing countries [11]. To mitigate future risks and plan for sustainable development, accurately predicting extreme rainfall events becomes crucial.

The Generalized Extreme Value distribution (GEV), introduced by [12], stands as one of the prevailing distributions for modeling extreme occurrences tied to environmental factors [13,14]. Its widespread adoption is due to its proficiency in encompassing a broad spectrum of tail behaviors [15]. The essence of fitting this distribution involves identifying the optimal values for its three parameters that effectively represent the observed maxima distribution. Among various approaches for parameter determination, the Maximum Likelihood Estimation (MLE) method, as presented by [16], garners the most prominence.

The stationarity assumption is one of the long standing tradition that extends to imply that forthcoming extreme events will conform to the same distribution, thus preempting any disparities in their behavior over time [17,18].

However, practical conditions rarely conform to the stationarity assumption [17,19–21]. Climate variability introduces non-stationarity in hydrological records, challenging the assumption of constant parameters in extreme value analysis. The impacts of a changing climate can introduce trends in the time series of environmental variables, thereby inducing a degree of temporal dependence in the distribution's parameters. Recent investigations have revealed that extreme rainfall events are becoming more intense, attributed to global climate change [19,22,23]. The escalation in global temperatures over the past century, largely due to human activities, results in heightened atmospheric moisture content, potentially increasing the likelihood of maximum rainfall events [24]. According to [25,26], more than half of the global population resided in urban areas by mid-2009, and projections suggest that by 2030, towns and cities will accommodate nearly 5 billion people, constituting approximately 80% of the world's population [19]. Over the past decade, research has highlighted potential rainfall changes linked to urbanization [27–29]. Recent studies such as [19,30,31] have specifically explored the impact of urbanization on extreme rainfall events. [31] illustrated the influence of the Indian Ocean Dipole (IOD) on extreme rainfall events in India, with an observed increase in the frequency of extreme IOD events due to greenhouse warming [32]. In the context of India, [21] demonstrated non-stationarity in heavy precipitation patterns attributed to various factors, including the global warming, local temperature changes (LTC), and El Niño-Southern Oscillation (ENSO) cycle.

On the other hand, [19] identified urbanization and local temperature changes as prime covariates for short-duration rainfall in India. However, they did not determine time to be the most suitable covariate. [33] ascertained that the non-stationary GEV model yields better estimates by considering the influence of the North Atlantic Oscillation and the annual seasonal cycle on parameters. [34] found local mean temperature to be a significant covariate for non-stationary GEV applied to extreme rainfall in their study region.

In a study by [20] conducted over United States, the nonstationary nature of the most recent extreme precipitation events occurred over different durations (1-, 3- and 5-days) were investigated by incorporating time-varying covariates, such as time, maximum temperature, mean temperature, and the ENSO cycle. The nonstationary frequency analysis for these extreme events was conducted using nonstationary GEV by incorporating the time-varying covariates. The study also observed that most of the temporal evolution of extreme precipitation events follow the nonstationary pattern, which may be due to the increase in the magnitude of recent extreme precipitation events, especially during hurricane events. The study also concluded that different combination of covariates can potentially influence the non-stationary frequency analysis, and the type of covariate may differ when the accumulated period of extreme precipitation event increased [20].

One significant application of extreme value theory in hydrology is the development of rainfall Intensity-Duration-Frequency (IDF) relationships. IDF curves, crucial for storm water management and engineering design globally, are typically created by fitting theoretical probability distributions to historical rainfall time series data of annual maximum rainfall (AMR) series [35]. Present IDF curves rely on stationary extreme value theory (EVT). However, the various physical factors mentioned earlier are anticipated to modify the intensity, duration, and frequency of rainfall extremes over time, introducing a non-stationary component into the time series [19–21].

Leveraging recent advancements in EVT, [36] devised non-stationary rainfall Intensity-Duration-Frequency (IDF) curves by introducing a linear trend in the location parameter of the GEV distribution using time as a covariate. Similarly, [37] explored non-stationarity in Melbourne, Australia's IDF curves, incorporating linear trends in both location and shape parameters of the GEV distribution with time as a covariate. Unlike earlier studies that solely employed time as a covariate for non-stationary IDF curves, recent research highlights the potential limitation of this approach [19–21]. The covariate time may not be the most suitable without exploring all possible covariates influencing climate change. The covariates to be used in a study possess the capacity to impact the non-stationary characteristics of climate-related factors, such as extreme rainfall. Hence, it is pivotal to discern the significant covariates in the process of conducting non-stationary frequency analysis.

The Southern Highlands region of Tanzania is well-known for its unique unimodal rainfall patterns known as Msimu rains, which typically occur from November to May (NDJFMA) annually [10,38]. Extreme rainfall in the region, particularly in the first part of the season, can be attributed to anomalies influencing the local climate, such as remote forcings (i.e., Walker Circulation, and IOD). These anomalies are likely to boost the inflow of moisture from the Congo Basin and the ongoing warming of the Central Indian Ocean [1,38]. The region covers a vast area, making up about 28% of Tanzania's mainland. It plays a crucial role in the country's economy, responsible for around 30% of export earnings and employing approximately 70% of Tanzania's population, mainly through farming. Surprisingly, despite its significant impact on the nation's economy, there is a noticeable absence of comprehensive studies on rainfall distribution in the area.

Therefore, this study seeks to identify the most effective covariates that influence extreme rainfall events over Southern Highlands region of Tanzania. To achieve this goal, five processes (i.e., urbanization, LTC, Global Temperature Anomaly (GTA), ENSO cycle, and IOD) considered as drivers of extreme rainfall are utilized as covariates. A total of 62 non-stationary GEV models are created by exploring different combinations of these covariates. Additionally, 2 non-stationary GEV incorporating the time to capture the seasonality of the unimodal rainfall in the region, and 1 stationary GEV are developed. The selection of the optimal GEV for each extreme rainfall duration is based on the corrected Akaike Information Criterion (AICc). The covariate(s) associated with the best-performing GEV model is considered as the most suitable covariate(s) for establishing non-stationary rainfall IDF relationships for the respective durations. The study employs MLE for parameter estimation. The impact of seasonality of the rainfall in location and scale parameters is captured through the development of sine and cosine functions as time cyclic covariate. Moreover, the study involves estimating the updated return levels and return periods of the extreme rainfall events in the context of the non-stationarity.

Study area

Fig. 1 shows the Southern Highlands that constitutes four regions of Tanzania namely; Iringa, Mbeya, Ruvuma, and Rukwa that lie within the coordinates of approximately latitude 9°19'37.20" South and longitude 34°09'54.00" East. Rainfall varies across the area, ranging from 823 mm on the Ufipa Plateau to 2,850 mm on the slopes of Mount Rungwe, the Livingstone and Poroto mountains situated near Lake Nyasa (also known as Lake Malawi).

Temperature fluctuations range from warm in the lowlands to cool in the highlands [39]. The mean annual rainfall spans from 750 to 3,500 mm and predominantly occurs from November to May as unimodal rainfall patterns [39]. Agriculture thrives in this region, cultivating over 70 crops including fruits and vegetables, while smallholder farmers raise diverse livestock species. The prominent farming systems encompass maize-based, rice-based, agropastoral, and coffee/banana-based approaches [39].

During June and July, the Southern Highlands experience notably cold conditions, with some areas recording minimum temperatures below 10 °C. This can adversely impact the well-being of both the local population and the region's animals. Moreover, the highlands confront the dual threats of droughts and floods due to climate change, leading to potential disruptions in water availability and crop yields. These extreme weather occurrences might have adverse repercussions on the local environment and the region's economic stability.

Extreme rainfall events in Tanzania

The Tanzania Meteorological Authority (TMA) had recently issued a seasonal rainfall forecast. The forecast indicated that during the period from October to December 2023, most regions in the country including Mbeya, Iringa and Rukwa were expected to receive rainfall categorized as above-normal to normal, with the patterns being influenced by El Niño conditions. Similarly, on December 21, 2011, an extreme rainfall incident occurred in various regions of the country as in Fig. 2, leading to widespread flooding, particularly in coastal areas. This flooding event had significant adverse effects on the national economy, resulting in the loss of approximately 43 lives in Dar es Salaam alone, displacing hundreds of individuals, and causing substantial damage to infrastructure such as houses, roads, and bridges. The accumulated rainfall over the three-day period (December 20th–22nd, 2011) was recorded at 260.2 mm at Julius Nyerere International Airport (JNIA) Station. The extreme rainfall within a 24-hour period, observed on December 21, 2011, amounted to 156.4 mm, surpassing the December monthly mean of 117.8 mm [1], as depicted in Fig. 2. According to TMA, this volume of rainfall set a record in the 58 years since the establishment of the station in 1953.

Data description

Rainfall data

Daily rainfall data (in millimeters) were collected from the TMA across a span of 30 years, covering the period from 1991 to 2020. The collection encompassed data from four distinct weather stations (Iringa, Mbeya, Ruvuma, and Rukwa) situated in the Southern Highlands of Tanzania. This region typically experiences a unimodal rainfall pattern, occurring from November to May. Given the individual analysis of data per station, a homogeneity test was deemed unnecessary for this study.

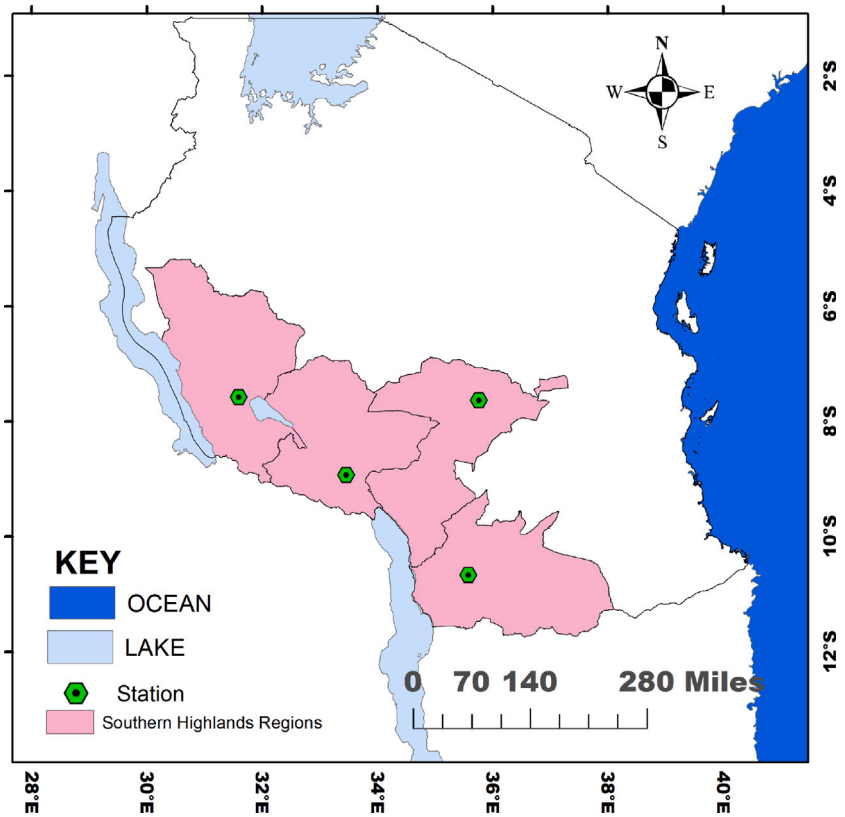


Fig. 1. Map showing four Stations in Southern Highlands Region of Tanzania.

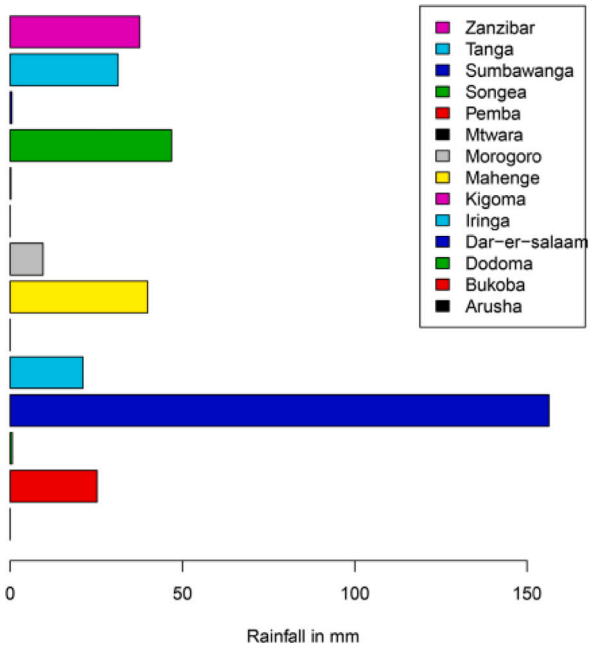


Fig. 2. The bar chart for the total observed rainfall across stations in Tanzania on December 21, 2011 [40].

Covariates

In the quest to determine the optimal covariate that influences extreme rainfall, this study explores 5 drivers of extreme rainfall (i.e., LTC, urbanization, GTA, IOD, and ENSO cycle), along with their potential combinations. In alignment with previous research such as [41,42] employing time as the only covariate, this study also includes time to model the seasonality of the unimodal rainfall in the region. Each covariate has been chosen based on its direct impact on rainfall patterns without interdependence affecting the results. This section provides the data for all drivers of extreme rainfall and the rationale justifying their inclusion.

Annual global temperature anomaly and local temperature variation

Over the last hundred years, human endeavors have led to an increase in global temperatures [8]. This rise in temperatures increases the atmosphere's capacity to hold water by approximately 7% to 8% for each 1 °C increase, directly impacting rainfall events [8]. Moreover, recent research indicates that elevated atmospheric water vapor levels can result in more intense rainfall occurrences [24]. According to some research [19,43], the way local temperatures affect rainfall events might be different from how GTA influences extreme rainfall. For instance in India, the variability in intensity and frequency of extreme rainfall in many regions is linked to anomalies in both local and global temperatures [21]. Therefore, consideration of LTC and GTA as drivers of extreme rainfall is justified. The Yearly Local Temperature Anomaly (LTA), serving as an indicator for the LTC, is derived from the maximum and minimum daily temperatures recorded by the TMA over a span of 30 years. Meanwhile, the HadCRUT4 annual GTA is employed as an index to assess the phenomenon of global warming. This anomaly is derived from the average surface air temperature measurements and can be accessed for reference at the following link http://www.metoffice.gov.uk/hadobs/hadcrut4/.

Urbanization

Within urban areas, natural land surfaces are supplanted by artificial surfaces, which possess distinct thermal characteristics such as heat capacity and thermal inertia [19]. These artificial surfaces typically excel at storing solar energy and converting it into sensible heat. Consequently, as this sensible heat permeates the air, urban air temperatures often soar 2°–10 °C higher than those in neighboring non-urban regions [19,44]. Therefore, urban areas alter boundary layer dynamics by generating an Urban Heat Island (UHI), which exerts a substantial impact on mesoscale circulations and subsequent convection [44]. In 19th century, many discussions emerged on the urbanization's influence on rainfall occurrences [45]. Early findings by [46,47] suggested a link between urbanization and increased rainfall events, particularly in summer. Recent studies have delved deeper into potential shifts in urban rainfall patterns due to urbanization [19,27,28,48]. [48] proposed urbanization's role in diurnal rainfall changes, while [49] highlighted urbanization-induced alterations in heavy rainfall phenomenon.

In this study, high-resolution remote sensing data is employed to model the urban growth, specifically the expansion of built-up land, in Southern Highlands region from 1991 to 2020. The process involved preparing a Level I Land Use Land Cover (LULC) map using Landsat satellite images available at https://earthexplorer.usgs.gov/, from which the urban settlement class was extracted. The satellite images were classified using the Maximum Likelihood Classification (MLC) algorithm. The MLC algorithm was chosen due to its high tolerance for inadequate training samples, particularly in comparison to other classification algorithms for Landsat images. This algorithm is favored by many users of remote sensing data for its effectiveness in classifying land covers globally.

Table 1 presents the satellite images utilized and the corresponding urban-built up land area extracted from the LULC map. Focused on the 2020 extent of Southern Highlands region, Landsat satellite images were cropped for the designated area and processed through the MLC algorithm for LULC mapping. Following this, the built-up land areas were computed from the LULC maps, and the urban growth models for each station revealed an exponential growth as shown in Fig. 3. The model's root mean square error (RMSE) and the coefficient of determination (R^2) are 12.04 and 0.962, 11.62 and 0.934, 13.46 and 0.971, and 12.83 and 0.986 for Iringa, Mbeya, Rukwa, and Ruvuma respectively.

El Niño-southern oscillation (ENSO) cycle

According to the TMA, El Niño refers to a climatic phenomenon characterized by the abnormal warming of surface waters in the eastern tropical Pacific Ocean [50]. This warming can lead to alterations in atmospheric circulation and weather patterns. The confirmation of an El Niño event typically occurs when sea surface temperatures in the eastern tropical Pacific exceed normal levels by at least 0.5 °C compared to long-term averages. The occurrence of El Niño is characterized by periodic warming of the Pacific Ocean waters, typically happening every three to eight years and lasting for 9-12 months [19,50,51]. Since 1900, over 30 El Niño events have been recorded worldwide, with notable occurrences in 1982–83, 1997–98, and 2014–16, which set records and caused significant damage. In Tanzania during the 2000s, El Niño events were observed in 2002–03, 2004–05, 2006–07, 2009–10, 2014–16, 2018–19, and currently in 2023, with its peak anticipated between September 2023 and February 2024 [40,50–53].

It is been found that the ENSO significantly influences the monthly and seasonal rainfall patterns in Tanzania and East Africa [50,54]. Warm ENSO events are linked to increased rainfall, while cold events are associated with reduced rainfall. The relationship between East African rainfall and ENSO is particularly robust during the OND (October–November–December) rainy season, but weaker during MAM (March–April–May) [50,54–56].

The ENSO cycle has two phases, the cold phase is known as La Niña, while the warm phase is termed El Niño [19]. These departures from typical surface temperatures can exert significant effects not only on oceanic dynamics but also on worldwide weather patterns and climate conditions [19,57–59]. The ENSO cycle stands as the most influential coupled ocean-atmosphere phenomenon, exerting a significant impact on global climate variability over interannual periods. Studies by [19–21,58,59] have underscored its impact on extreme rainfall both locally and regionally. Additionally, [60] suggests that the frequency of

Table 1
Satellite images utilized for urbanization modeling and classification outcomes for all stations.

Station	Year	Satellite	Capturing date	SR (m)	SR (b)	BUL (km^2)
Iringa	1991	4/5 TM	1991–08–25	30	7	10.62
	1995	4/5 TM	1995–10–12	30	7	11.36
	2000	4/5 TM	2000–01–01	30	7	14.04
	2006	4/5 TM	2006–07–15	30	7	19.44
	2010	7ETM+	2010–01–28	30	8	22.95
	2015	7ETM+	2015–09–30	30	8	27.05
	2020	8 OLI	2020–02–20	30	11	38.02
Mbeya	1991	4/5 TM	1991–08–25	30	7	11.74
	1995	4/5 TM	1995–10–12	30	7	13.88
	2000	4/5 TM	2000–01–01	30	7	16.80
	2006	4/5 TM	2006–07–15	30	7	22.42
	2010	7ETM+	2010–01–28	30	8	27.14
	2015	7ETM+	2015–09–30	30	8	33.54
	2020	8 OLI	2020–02–20	30	11	49.23
Rukwa	1991	4/5 TM	1991–08–25	30	7	11.10
	1995	4/5 TM	1995–10–12	30	7	13.12
	2000	4/5 TM	2000–01–01	30	7	15.89
	2006	4/5 TM	2006–07–15	30	7	20.25
	2010	7ETM+	2010–01–28	30	8	24.71
	2015	7ETM+	2015–09–30	30	8	30.76
	2020	8 OLI	2020–02–20	30	11	45.59
Ruvuma	1991	4/5 TM	1991–08–25	30	7	10.05
	1995	4/5 TM	1995–10–12	30	7	11.87
	2000	4/5 TM	2000–01–01	30	7	14.37
	2006	4/5 TM	2006–07–15	30	7	18.32
	2010	7ETM+	2010–01–28	30	8	22.36
	2015	7ETM+	2015–09–30	30	8	27.83
	2020	8 OLI	2020–02–20	30	11	41.25

Note: SR (m) is Spatial Resolution in meters, SR (b) is Spectral Resolution in bands, and BUL (km^2) is a Build-up Land in Square Kilometers.

El Niño events, which are part of the ENSO cycle, may increase with higher concentrations of greenhouse gases. Moreover, some scholars [20,21] have begun modeling the non-stationarity in intensity, duration, and frequency of daily extreme rainfall, incorporating the ENSO cycle as a covariate. There are various indices, including the Multivariate ENSO Index (MEI), Southern Oscillation Index (SOI), and Sea Surface Temperature (SST), that can be utilized to represent the ENSO cycle [19]. In this study, we adopt the methodology proposed by [19,21], which involves using the SST index as ENSO indicator available at https://psl.noaa.gov/data/timeseries/month/ as a covariate for modeling non-stationarity in extreme rainfall patterns.

Indian ocean dipole (IOD)

In the last decade, researchers uncovered a dipole mode within the tropical Indian Ocean, known as the IOD, assessed by the Dipole Mode Index (DMI) [61]. According to [61], the DMI reflects the contrast in SST between the tropical western Indian Ocean (50 ° E–70 ° E, 10 ° S–10 ° N) and the tropical southeastern Indian Ocean (90 ° E–110 ° E, 10 ° S-Equator). A positive DMI results in drought conditions in the Indonesia region and heightened rainfall and flooding in East Africa [50,62]. Conversely, a reversal in DMI sign triggers a shift in these anomalous patterns to the opposite phase [62,63]. Analyzing the IOD's impact on the connection between Indian monsoon rainfall and ENSO, [62] found that the ENSO-induced anomalous circulation during the Indian summer monsoon over the Indian region is either mitigated or amplified by the IOD-induced anomalous meridional circulation.

Additionally, some studies have shown a consistent connection between the IOD and various flooding events across the East African region [50,54–56]. [63] investigated the effects of ENSO and IOD occurrences on sub-regional Indian Summer Monsoon Rainfall (ISMR), determining that the correlations between ENSO and ISMR are stronger compared to those between IOD and ISMR. [31] examined the impact of IOD on heavy rainfall phenomenon across India, presenting the initial evidence that recent heavy rainfall events over India are significantly influenced by IOD. Moreover, according to [32] it is noteworthy that the frequency of extreme IOD events is on the rise due to greenhouse warming. This study deployed monthly DMI obtained from the HadISST dataset retrieved from https://psl.noaa.gov/data/timeseries/month/. Yearly DMI values, are computed and utilized as a covariate to represent the IOD.

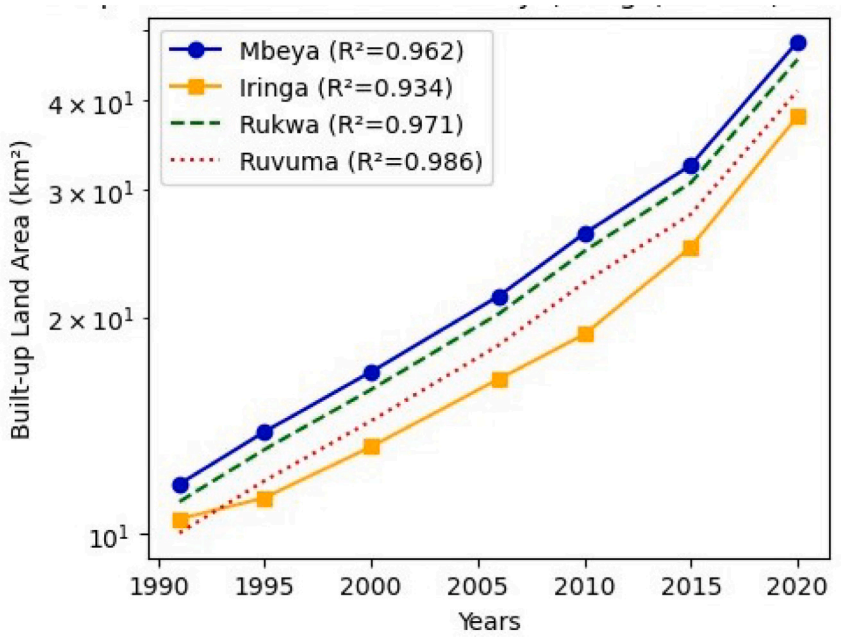


Fig. 3. The built up land models for all stations.

Time

[64] examined the variability in extreme daily rainfall in United States, by introducing a trend in the scale and location parameters of the GEV distribution using time as one of the covariates. In 2014, [36] devised non-stationary rainfall IDF curves by integrating a linear trend in the location parameter of the GEV distribution using a time covariate. [37] explored the non-stationarity in IDF curves by integrating linear trends in both location and shape parameters of the GEV distribution using time as a covariate. This study also considers time as one of the potential covariates for constructing non-stationary IDF curves. However, this study will apply non-stationary GEV model with a cyclic covariate structures in scale and location parameters to capture the reality of seasonality observed in the rainfall patterns investigation. The cyclic covariates mentioned refer to sinusoidal functions utilized to model the seasonal variations observed in extreme rainfall patterns. Specifically, we employ sine and cosine functions with a 1-year period to characterize the fluctuations in the location and scale parameters of the GEV distribution over the course of a year. The years of maxima rainfall series are utilized as a covariate representing time.

Methodology

This study involves a methodology that consists of

1. Developing various non-stationary GEV models by incorporating five drivers of extreme rainfall (i.e., urbanization, LTC, ENSO cycle, global warming (GTA), and IOD) both individually and in combination, as covariates affecting the parameters of the GEV model. In addition time will also be used independently to develop non-stationary GEV model that captures the seasonality of the unimodal rainfall patterns.
   To align the gridded GTA data with the specific locations of our study stations (Iringa, Mbeya, Rukwa, and Ruvuma), we followed a precise methodology:
   (a). Grid Mapping: We identified the grid cells in the HadCRUT4 dataset that correspond to the geographical coordinates of each station. These grid cells were selected based on their proximity to the station locations.
   (b). Data Extraction: For each station, we extracted the temperature anomaly data from the nearest grid cell. This approach ensures that the GTA data reflects the temperature anomalies representative of the station's location.

   This method of extracting GTA data ensures that it accurately represents the climate conditions at each station.
2. Determining the most suitable covariate(s) and corresponding GEV model in order to devise modern non-stationary IDF curves that consider optimal covariates.
3. Identifying the correlation between the optimal covariate(s) and the extreme rainfall durations.

Block maxima and extreme value theory

Annual maxima approach is used in this study to identify extreme values in a dataset. The dataset is divided into blocks of equal size, and the maximum value within each block is selected. Choosing the block size involves a trade-off between bias and variance: larger blocks reduce estimation variance but may introduce bias, while smaller blocks may provide a poor approximation of extreme values. To mitigate this, blocks are defined as one-year periods, ensuring that each block contains observations for an entire year, resulting in annual maxima, M_n.

Classical extreme value theory describes the behavior of

$$M_n = \max\left\{X_1, X_2, \dots, X_n\right\}$$

whereby $X_1, X_2, \dots, X_n$ having a distribution function F, is a sequence of independent and identically distributed (IID) random variables. As the value of n increases to infinity, the properties of the maximum value, M_n, become significantly important. The distribution of M_n for all n can be theoretically derived as

$$\begin{aligned} Pr\left\{M_n \le x\right\} &= Pr\left\{X_1 \le x, X_2 \le x, \dots, X_n \le x\right\} \\ &= Pr\left\{X_1 \le x\right\} \times Pr\left\{X_2 \le x\right\} \times \cdots \times Pr\left\{X_n \le x\right\} \\ &= \left\{F(x)\right\}^n, \end{aligned}$$

In practical applications, the distribution function F is usually not known. To estimate F, standard statistical methods can be used with observed data, which in turn could lead to an estimate of the distribution of M_n. This study use only extreme data to estimate approximate families of models for $\{F\}^n$. The extremal types theorem is a significant finding that provides a comprehensive range of potential limiting distributions for M_n as n approaches infinity. The theorem was initially established by [65] and subsequently confirmed by [66], with each researcher contributing different elements to the proof.

Theorem. *Extremal Types Theorem.*

If there exist sequences of constants $\left\{\sigma_n > 0\right\}$ *and* $\left\{\mu_n\right\}$ *such that*

$$Pr\left\{\frac{M_n - \mu_n}{\sigma_n} \le x\right\} \to G(x) \quad as \quad n \to \infty \tag{1}$$

where G is a non-degenerate distribution function, then G belongs to one of the following families:

$$I : G(x) = \exp\left\{-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\}, -\infty\langle x\rangle\infty$$

$$II : G(x) = \begin{cases} 0, & if \quad x \le \mu \\ \exp\left\{-\left(\frac{x-\mu}{\sigma}\right)^{-\xi}\right\}, & if \quad x > \mu, \xi > 0 \end{cases}$$

$$III : G(x) = \begin{cases} \exp\left\{-\left[-\left(\frac{x-\mu}{\sigma}\right)^{\xi}\right]\right\}, & if \quad x < \mu, \xi > 0 \\ 1, & if \quad x \ge \mu, \xi > 0 \end{cases}$$

for parameters $\sigma > 0, \mu,$ *and* $\xi > 0.$

The theorem above states that if there exist appropriate sequences $\{\sigma_n\}$ and $\{\mu_n\}$ that can stabilize M_n, then the limiting distribution belongs to one of three families of distributions (Type I, Type II and Type III) [67]. These three types of distributions are known as the Gumbel, Fréchet, and Weibull families respectively, and they are together referred to as GEV.

Stationary and non-stationary GEV distribution models

A stationary GEV model is a statistical model used to analyze extreme events, such as extreme rainfall or temperatures, over time [15]. In this model, it is assumed that the parameters of the GEV distribution (location, scale, and shape) remain constant over time. This implies that the probability distribution of extreme events does not change with time. Mathematically, the cumulative distribution function (CDF) of the stationary GEV distribution function $F(x)$ becomes

$$F(x;\xi,\mu,\sigma) = \begin{cases} \exp\left(-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right), & \sigma > 0 \\ 1+\xi\left(\frac{x-\mu}{\sigma}\right), & \sigma > 0, \quad \xi \ne 0 \\ \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right), & \sigma > 0, \quad \xi = 0 \end{cases} \tag{2}$$

where: x is the extreme event, and μ, σ, ξ are location, scale, and shape parameters, respectively. ξ determines whether the distribution is Fréchet ($\xi > 0$), Gumbel ($\xi = 0$), or Weibull ($\xi < 0$). This model describes the probability that a random variable X

is less than or equal to x. The parameters μ, σ, and ξ are estimated from the data to fit the distribution to the observed extreme values.

A non-stationary GEV model is a statistical model used to analyze extreme events over time, allowing for changes in the parameters of the GEV distribution [15]. Unlike the stationary GEV model, the parameters of the non-stationary GEV model are allowed to vary over time, capturing temporal trends or fluctuations in extreme event behavior. Non-stationary processes exhibit systematic changes over time, often due to seasonal variations or long-term trends driven by factors like climate patterns [41,42,68]. This study focuses on developing non-stationary GEV distribution models to investigate whether GEV parameters are influenced by physical factors (i.e., Urbanization, LTC, GTA, ENSO, and IOD) affecting daily rainfall in Southern Highlands, Tanzania. The non-stationary GEV cumulative distribution function is represented as follows:

$$F(x;\xi,\mu(t),\sigma(t)) = \exp\left(-\frac{1}{\xi}\left[1+\xi\left(\frac{x-\mu(t)}{\sigma(t)}\right)\right]^{-\frac{1}{\xi}}\right), \quad 1+\xi\left(\frac{x-\mu(t)}{\sigma(t)}\right) > 0 \tag{3}$$

In the analysis of non-stationarity, parameters are articulated as functions of covariates over time [69–71]. As previously mentioned, five drivers of extreme rainfall were chosen for examining non-stationarity in heavy rainfall. The non-stationary GEV models were applied by adjusting the location and scale parameters according to the selected covariates. However, the shape parameter remains constant, as assuming it to vary smoothly over time may be unrealistic [15]. It is recommended to use a constant shape parameter for a small number of stations [20,72], a presumption also adopted in recent studies [19,20,73,74]. For each station, a total of 65 models were derived, comprising one stationary and 62 non-stationary models based on five covariates and 2 time-varying parameters. A succinct overview of the model formulation is provided with comprehensive details available in Tables 2 and 3. Four model types are briefly outlined below: described by Eqs. (4), (5), (6), (10) and (11).

Type I: Stationary GEV model, with all parameters constant over time, (S0):

$$\mu(t) = \mu; \sigma(t) = \sigma; \xi(t) = \xi. \tag{4}$$

Type II: Non-stationary GEV models (NS), with a non-stationarity in location parameter only.
Table 2 gives the summary of these developed models with their corresponding covariates (i.e., NS1 to NS32):

$$\begin{aligned} \mu(t) &= \mu_0 + \mu_1\gamma_1 + \mu_2\gamma_2 + \mu_3\gamma_3 + \mu_4\gamma_4 + \mu_5\gamma_5 \\ \sigma(t) &= \sigma \\ \xi(t) &= \xi \end{aligned} \tag{5}$$

Type III: Non-stationary GEV models (NS), with non-stationarity in location and scale parameters.
Table 3 gives the summary of these models with their covariates (i.e., NS33 to NS64):

$$\begin{aligned} \mu(t) &= \mu_0 + \mu_1\gamma_1 + \mu_2\gamma_2 + \mu_3\gamma_3 + \mu_4\gamma_4 + \mu_5\gamma_5 \\ \sigma(t) &= \exp(\sigma_0 + \sigma_1\gamma_1 + \sigma_2\gamma_2 + \sigma_3\gamma_3 + \sigma_4\gamma_4 + \sigma_5\gamma_5) \\ \xi(t) &= \xi \end{aligned} \tag{6}$$

In model type II and III, γ_1, γ_2, γ_3, γ_4, and γ_5 represent five covariates for five drivers of extreme rainfall: Local temperature variation (L), Urbanization (U), Global warming (GTA) (G), IOD cycle (D), and ENSO cycle (E). In the stationary model, SO, these covariates γ_1, γ_2, γ_3, γ_4, and γ_5 are assigned zero values. In Eq. (6), an exponential function is introduced in a scale trend to ensure the positivity of the scale parameter. The influence of the covariates γ_1, γ_2, γ_3, γ_4, and γ_5 on the location parameter μ_0 is captured by the slope parameters μ_1, μ_2, μ_3, μ_4, and μ_5. Similarly, the influence of covariates γ_1, γ_2, γ_3, γ_4, and γ_5 on the scale parameter σ_0 is modeled by the slope parameters σ_1, σ_2, σ_3, σ_4, and σ_5 to illustrate the trend in the scale parameter.

Type IV: Non-stationary GEV models, with time covariate in location and scale parameters NS32 in Table 2 and NS64 in Table 3.

Here, our focus is to model μ and σ as combinations of sine and cosine functions with a 1-year period to capture the influence of seasonality in the model.

For the location parameter, we have:

$$\mu(t) = \mu_0 + \mu_1 \sin\left(\omega c_i + \phi\right), \quad i = 1, \ldots, 12 \tag{7}$$

Here, μ_1 represents the amplitude of the sinusoidal component, ϕ represents the phase angle or phase shift, it determines the starting point of the oscillation. For monthly modeling, it might represent the phase shift necessary to align the model with the observed monthly rainfall patterns. ω is the angular frequency or the rate of oscillation. In the context of seasonal modeling, ω represents the cyclic nature of the seasons within a year. It determines how quickly the parameter $\mu(t)$ or $\sigma(t)$ varies with time. c_i denotes time t, which is the independent variable. In the context of modeling monthly rainfall patterns, t would typically represent time in months, ranging from 1 to 12 (January to December). Notice that Eq. (7) is based on a trigonometric identity, which allows us to express a sine wave with a phase shift as a combination of sine and cosine terms. In view of this, we arrive at the reformulated Eqs. (8), and (9) as follows:

$$\mu(t) = \mu_0 + \mu_1 \sin\left(\omega c_i\right) + \mu_2 \cos\left(\omega c_i\right), \quad i = 1, \ldots, 12 \tag{8}$$

$$\sigma(t) = \sigma_0 + \sigma_1 \sin\left(\omega c_i\right) + \sigma_2 \cos\left(\omega c_i\right), \quad i = 1, \ldots, 12 \tag{9}$$

Table 2
A Catalog of developed Non-stationary models with their attributes (i.e., S0, and NS1 to NS32).

Model type	Model summary
S0	$X \sim \text{GEV}(\mu, \sigma, \xi)$
NS1	$X \sim \text{GEV}((\mu_0 + U\mu_1), \sigma, \xi)$
NS2	$X \sim \text{GEV}((\mu_0 + L\mu_1), \sigma, \xi)$
NS3	$X \sim \text{GEV}((\mu_0 + G\mu_1), \sigma, \xi)$
NS4	$X \sim \text{GEV}((\mu_0 + E\mu_1), \sigma, \xi)$
NS5	$X \sim \text{GEV}((\mu_0 + I\mu_1), \sigma, \xi)$
NS6	$X \sim \text{GEV}((\mu_0 + U\mu_1 + L\mu_2), \sigma, \xi)$
NS7	$X \sim \text{GEV}((\mu_0 + U\mu_1 + G\mu_2), \sigma, \xi)$
NS8	$X \sim \text{GEV}((\mu_0 + U\mu_1 + E\mu_2), \sigma, \xi)$
NS9	$X \sim \text{GEV}((\mu_0 + U\mu_1 + I\mu_2), \sigma, \xi)$
NS10	$X \sim \text{GEV}((\mu_0 + L\mu_1 + G\mu_2), \sigma, \xi)$
NS11	$X \sim \text{GEV}((\mu_0 + L\mu_1 + E\mu_2), \sigma, \xi)$
NS12	$X \sim \text{GEV}((\mu_0 + L\mu_1 + I\mu_2), \sigma, \xi)$
NS13	$X \sim \text{GEV}((\mu_0 + G\mu_1 + E\mu_2), \sigma, \xi)$
NS14	$X \sim \text{GEV}((\mu_0 + G\mu_1 + I\mu_2), \sigma, \xi)$
NS15	$X \sim \text{GEV}((\mu_0 + E\mu_1 + I\mu_2), \sigma, \xi)$
NS16	$X \sim \text{GEV}((\mu_0 + U\mu_1 + L\mu_2 + G\mu_3), \sigma, \xi)$
NS17	$X \sim \text{GEV}((\mu_0 + U\mu_1 + L\mu_2 + E\mu_3), \sigma, \xi)$
NS18	$X \sim \text{GEV}((\mu_0 + U\mu_1 + L\mu_2 + I\mu_3), \sigma, \xi)$
NS19	$X \sim \text{GEV}((\mu_0 + U\mu_1 + G\mu_2 + E\mu_3), \sigma, \xi)$
NS20	$X \sim \text{GEV}((\mu_0 + U\mu_1 + G\mu_2 + I\mu_3), \sigma, \xi)$
NS21	$X \sim \text{GEV}((\mu_0 + U\mu_1 + E\mu_2 + I\mu_3), \sigma, \xi)$
NS22	$X \sim \text{GEV}((\mu_0 + L\mu_1 + G\mu_2 + E\mu_3), \sigma, \xi)$
NS23	$X \sim \text{GEV}((\mu_0 + L\mu_1 + G\mu_2 + I\mu_3), \sigma, \xi)$
NS24	$X \sim \text{GEV}((\mu_0 + L\mu_1 + E\mu_2 + I\mu_3), \sigma, \xi)$
NS25	$X \sim \text{GEV}((\mu_0 + G\mu_1 + E\mu_2 + I\mu_3), \sigma, \xi)$
NS26	$X \sim \text{GEV}((\mu_0 + U\mu_1 + L\mu_2 + G\mu_3 + E\mu_4), \sigma, \xi)$
NS27	$X \sim \text{GEV}((\mu_0 + U\mu_1 + L\mu_2 + G\mu_3 + I\mu_4), \sigma, \xi)$
NS28	$X \sim \text{GEV}((\mu_0 + U\mu_1 + L\mu_2 + E\mu_3 + I\mu_4), \sigma, \xi)$
NS29	$X \sim \text{GEV}((\mu_0 + U\mu_1 + G\mu_2 + E\mu_3 + I\mu_4), \sigma, \xi)$
NS30	$X \sim \text{GEV}((\mu_0 + L\mu_1 + G\mu_2 + E\mu_3 + I\mu_4), \sigma, \xi)$
NS31	$X \sim \text{GEV}((\mu_0 + U\mu_1 + L\mu_2 + G\mu_3 + E\mu_4 + I\mu_5), \sigma, \xi)$
NS32	$X \sim \text{GEV}(\mu_0 + \mu_1 \sin\left(\frac{2\pi t}{12}\right) + \mu_2 \cos\left(\frac{2\pi t}{12}\right), \sigma, \xi) \quad t = 1, \ldots, 12$

This study characterize the location μ and scale σ parameters by sinusoidal functions exhibiting seasonal variations, while the shape parameter ξ is considered constant:

$$\mu(t) = \mu_0 + \mu_1 \sin\left(\frac{2\pi t}{12}\right) + \mu_2 \cos\left(\frac{2\pi t}{12}\right), \quad t = 1, \ldots, 12 \tag{10}$$

$$\sigma(t) = \sigma_0 + \sigma_1 \sin\left(\frac{2\pi t}{12}\right) + \sigma_2 \cos\left(\frac{2\pi t}{12}\right), \quad t = 1, \ldots, 12 \tag{11}$$

The model described by the equations for $\mu(t)$ and $\sigma(t)$, exhibit seasonal variations over 12 months, that is well-suited for capturing the characteristics of rainfall patterns with a single peak or unimodal rainfall patterns in a year. It allows for modeling changes in the location and scale parameters of the GEV distribution over the course of a year, which is appropriate for areas where rainfall predominantly follows a single seasonal pattern.

Parameter estimation and model selection technique

Parameter estimation for the GEV distribution is typically performed using the MLE method. The primary benefit of the MLE approach lies in its ability to analyze complex scenarios such as non-stationarity and covariate effects [75]. However, a significant drawback of the MLE method is its tendency to become highly unpredictable when dealing with as small sample sizes as 20 [75]. The log-likelihood function is maximized to obtain maximum-likelihood estimator $(\hat{\xi}, \hat{\mu}, \hat{\sigma})$ for the parameters (ξ, μ, σ). Given a dataset consisting of n years of annual maximum series, represented as $X = x_1, x_2, \ldots, x_n$, the log-likelihood when $\xi \neq 0$, is defined by [15] as:

$$L(\xi, \mu, \sigma) = -n \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^{n} \log\left[1 + \xi\left(\frac{x_i - \mu}{\sigma}\right)\right] - \sum_{i=1}^{n} \left[1 + \xi\left(\frac{x_i - \mu}{\sigma}\right)\right]^{\frac{-1}{\xi}} \tag{12}$$

Table 3
A Catalog of developed Non-stationary models (i.e., NS33 to NS64).

Model type	Model summary
NS33	$X \sim \text{GEV}((\mu_0 + U\mu_1), e^{(\sigma_0+U\sigma_1)}, \xi)$
NS34	$X \sim \text{GEV}((\mu_0 + L\mu_1), e^{(\sigma_0+L\sigma_1)}, \xi)$
NS35	$X \sim \text{GEV}((\mu_0 + G\mu_1), e^{(\sigma_0+G\sigma_1)}, \xi)$
NS36	$X \sim \text{GEV}((\mu_0 + E\mu_1), e^{(\sigma_0+E\sigma_1)}, \xi)$
NS37	$X \sim \text{GEV}((\mu_0 + I\mu_1), e^{(\sigma_0+I\sigma_1)}, \xi)$
NS38	$X \sim \text{GEV}((\mu_0 + U\mu_1 + L\mu_2), e^{(\sigma_0+U\sigma_1+L\sigma_2)}, \xi)$
NS39	$X \sim \text{GEV}((\mu_0 + U\mu_1 + G\mu_2), e^{(\sigma_0+U\sigma_1+G\sigma_2)}, \xi)$
NS40	$X \sim \text{GEV}((\mu_0 + U\mu_1 + E\mu_2), e^{(\sigma_0+U\sigma_1+E\sigma_2)}, \xi)$
NS41	$X \sim \text{GEV}((\mu_0 + U\mu_1 + I\mu_2), e^{(\sigma_0+U\sigma_1+I\sigma_2)}, \xi)$
NS42	$X \sim \text{GEV}((\mu_0 + L\mu_1 + G\mu_2), e^{(\sigma_0+L\sigma_1+G\sigma_2)}, \xi)$
NS43	$X \sim \text{GEV}((\mu_0 + L\mu_1 + E\mu_2), e^{(\sigma_0+L\sigma_1+E\sigma_2)}, \xi)$
NS44	$X \sim \text{GEV}((\mu_0 + L\mu_1 + I\mu_2), e^{(\sigma_0+L\sigma_1+I\sigma_2)}, \xi)$
NS45	$X \sim \text{GEV}((\mu_0 + G\mu_1 + E\mu_2), e^{(\sigma_0+G\sigma_1+E\sigma_2)}, \xi)$
NS46	$X \sim \text{GEV}((\mu_0 + G\mu_1 + I\mu_2), e^{(\sigma_0+G\sigma_1+I\sigma_2)}, \xi)$
NS47	$X \sim \text{GEV}((\mu_0 + E\mu_1 + I\mu_2), e^{(\sigma_0+E\sigma_1+I\sigma_2)}, \xi)$
NS48	$X \sim \text{GEV}((\mu_0 + U\mu_1 + L\mu_2 + G\mu_3), e^{(\sigma_0+U\sigma_1+L\sigma_2+G\sigma_3)}, \xi)$
NS49	$X \sim \text{GEV}((\mu_0 + U\mu_1 + L\mu_2 + E\mu_3), e^{(\sigma_0+U\sigma_1+L\sigma_2+E\sigma_3)}, \xi)$
NS50	$X \sim \text{GEV}((\mu_0 + U\mu_1 + L\mu_2 + I\mu_3), e^{(\sigma_0+U\sigma_1+L\sigma_2+I\sigma_3)}, \xi)$
NS51	$X \sim \text{GEV}((\mu_0 + U\mu_1 + G\mu_2 + E\mu_3), e^{(\sigma_0+U\sigma_1+G\sigma_2+E\sigma_3)}, \xi)$
NS52	$X \sim \text{GEV}((\mu_0 + U\mu_1 + G\mu_2 + I\mu_3), e^{(\sigma_0+U\sigma_1+G\sigma_2+I\sigma_3)}, \xi)$
NS53	$X \sim \text{GEV}((\mu_0 + U\mu_1 + E\mu_2 + I\mu_3), e^{(\sigma_0+U\sigma_1+E\sigma_2+I\sigma_3)}, \xi)$
NS54	$X \sim \text{GEV}((\mu_0 + L\mu_1 + G\mu_2 + E\mu_3), e^{(\sigma_0+L\sigma_1+G\sigma_2+E\sigma_3)}, \xi)$
NS55	$X \sim \text{GEV}((\mu_0 + L\mu_1 + G\mu_2 + I\mu_3), e^{(\sigma_0+L\sigma_1+G\sigma_2+I\sigma_3)}, \xi)$
NS56	$X \sim \text{GEV}((\mu_0 + L\mu_1 + E\mu_2 + I\mu_3), e^{(\sigma_0+L\sigma_1+E\sigma_2+I\sigma_3)}, \xi)$
NS57	$X \sim \text{GEV}((\mu_0 + G\mu_1 + E\mu_2 + I\mu_3), e^{(\sigma_0+G\sigma_1+E\sigma_2+I\sigma_3)}, \xi)$
NS58	$X \sim \text{GEV}((\mu_0 + U\mu_1 + L\mu_2 + G\mu_3 + E\mu_4), e^{(\sigma_0+U\sigma_1+L\sigma_2+G\sigma_3+E\sigma_4)}, \xi)$
NS59	$X \sim \text{GEV}((\mu_0 + U\mu_1 + L\mu_2 + G\mu_3 + I\mu_4), e^{(\sigma_0+U\sigma_1+L\sigma_2+G\sigma_3+I\sigma_4)}, \xi)$
NS60	$X \sim \text{GEV}((\mu_0 + U\mu_1 + L\mu_2 + E\mu_3 + I\mu_4), e^{(\sigma_0+U\sigma_1+L\sigma_2+E\sigma_3+I\sigma_4)}, \xi)$
NS61	$X \sim \text{GEV}((\mu_0 + U\mu_1 + G\mu_2 + E\mu_3 + I\mu_4), e^{(\sigma_0+U\sigma_1+G\sigma_2+E\sigma_3+I\sigma_4)}, \xi)$
NS62	$X \sim \text{GEV}((\mu_0 + L\mu_1 + G\mu_2 + E\mu_3 + I\mu_4), e^{(\sigma_0+L\sigma_1+G\sigma_2+E\sigma_3+I\sigma_4)}, \xi)$
NS63	$X \sim \text{GEV}((\mu_0 + U\mu_1 + L\mu_2 + G\mu_3 + E\mu_4 + I\mu_5), e^{(\sigma_0+U\sigma_1+L\sigma_2+G\sigma_3+E\sigma_4+I\sigma_5)}, \xi)$
NS64	$X \sim \text{GEV}(\mu_0 + \mu_1 \sin\left(\frac{2\pi t}{12}\right) + \mu_2 \cos\left(\frac{2\pi t}{12}\right), \sigma_0 + \sigma_1 \sin\left(\frac{2\pi t}{12}\right) + \sigma_2 \cos\left(\frac{2\pi t}{12}\right), \xi)$

given that $\left[1 + \xi\left(\frac{x_i - \mu}{\sigma}\right)\right] > 0, \quad$ for $i = 1...n$. For the case when $\xi = 0$, the log-likelihood function reduces to

$$L(0, \mu, \sigma) = -n\log(\sigma) - \sum_{i=1}^{n} \log\left(\frac{x_i - \mu}{\sigma}\right) - n\sum_{i=1}^{n} \exp\left\{-\frac{x_i - \mu}{\sigma}\right\}. \tag{13}$$

In the case of non-stationarity, we substitute the location and scale parameter functions into the Eqs. (12) and (13) according to the non-stationary conditions.

In order to assess the uncertainty associated with the estimated parameters, this study determines confidence intervals (CI) for the GEV parameters during the MLE process. CI for the GEV parameters (ξ, μ, σ) are derived from the asymptotic normality of the maximum-likelihood estimators. For instance, given $\alpha \in (0, 1)$, the $(1 - \alpha)100\%$ CI for ξ can be determined as follows:

$$\left(\hat{\xi} - Z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{\hat{v}_{2,2}}{n}}, \quad \hat{\xi} + Z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{\hat{v}_{2,2}}{n}}\right) \tag{14}$$

where $\hat{\xi}$ is the maximum-likelihood estimate of ξ, $Z_{\left(1-\frac{\alpha}{2}\right)}$ is the $\left(1 - \frac{\alpha}{2}\right)$ quantile of the normal distribution, and $\hat{v}_{2,2}$ is the second diagonal element of the inverse of the Fisher information matrix, which is given as

$$M(\theta) = -\mathbb{E}\left(\frac{\partial^2 \log g(X; \theta)}{\partial\theta\partial\theta^T}\right), \tag{15}$$

with $\theta = (\xi, \mu, \sigma)$, $\mathbb{E}$ represents the expected value operator, and g is the probability density function (PDF) of the random variable X, which is typically parameterized by θ. The PDF of X is derived as

$$g(X; \xi, \mu, \sigma) = \frac{1}{\sigma}\left(1 + \xi\frac{x - \mu}{\sigma}\right)\exp\left(-\left(1 + \xi\frac{x - \mu}{\sigma}\right)^{-1/\xi}\right) \tag{16}$$

The following are the typical elements comprised in the Fisher information matrix (M)

$$\begin{cases} M_{1,1}(\theta) &= \frac{1}{\sigma^2\xi^2}(1-2\Gamma(2+\xi)+c), \\ M_{1,2}(\theta) &= -\frac{1}{\sigma\xi^2}\left(1-\eta-k+\frac{1-\Gamma(2+\xi)}{\xi}+\frac{c}{\xi}\right), \\ M_{1,3}(\theta) &= \frac{1}{\sigma^2\xi}(c-\Gamma(2+\xi)), \\ M_{2,2}(\theta) &= \frac{1}{\xi^2}\left(\frac{\pi^2}{6}+\left(1-\eta+\frac{1}{\xi}\right)-\frac{2k}{\xi}+\frac{c}{\xi^2}\right), \\ M_{2,3}(\theta) &= -\frac{1}{\sigma\xi}\left(k-\frac{c}{\xi}\right), \\ M_{3,3}(\theta) &= \frac{c}{\sigma^2}, \end{cases} \tag{17}$$

where $\eta \approx 0.5772157$ is Euler's constant,

$$c = (1+\xi)^2\Gamma(1+2\xi),$$

$$k = \Gamma(2+\xi)\left(\psi(1+\xi)+\frac{1+\xi}{\xi}\right)$$

with $\psi(x) = \frac{d}{dx}(\log\Gamma(x))$.

The symbol Γ represents the gamma function, which is typically given as

$$\Gamma(t) = \int_0^{+\infty} e^{-x}x^{t-1}dx, \quad t>0. \tag{18}$$

However, it is crucial to recognize that inference relying solely on normal approximation may lead to inaccuracies, as the approximation to the true sampled distribution of the estimator might be inadequate [15]. According to [15,76,77], better approximations can be achieved using the profile likelihood function (PLF), as the confidence intervals derived from it are not necessarily centered around the estimated points, unlike those from normality approximation. Specifically, profile likelihood intervals might exhibit a rightward shift, whereas those from normality approximation are always symmetrically centered around the estimates due to the inherent symmetry of the normal distribution [15,76,77].

Hence, the PLF for the shape parameter ξ is obtained as

$$L_p(\xi) = \max_{(\sigma,\mu)|\xi} L(\xi,\mu,\sigma),$$

whereas, the profile likelihood ratio statistic is obtained as

$$R = \frac{L_p(\xi_0)}{L_p(\widehat{\xi})},$$

This aligns with the traditional likelihood ratio statistic used to test the null hypothesis $H_0 : \xi = \xi_0$ against alternative $H_1 : \xi \neq \xi_0$. Thus, under the assumption of H_0, as the sample size n tends towards infinity, we have

$$-2\log R \xrightarrow{d} \chi_1^2,$$

In this context, $\xrightarrow{d}$ signifies "convergence in distribution to", while χ_1^2 denotes the chi-square distribution with a single degree of freedom.

The H_0 gets rejected with a significance level of α if $-2\log R > \chi_1^2(1-\alpha)$, where $\chi_1^2(1-\alpha)$ denotes the $(1-\alpha)$ quantile of the χ_1^2 distribution. Hence, the CI for ξ, based on profile likelihood, with a confidence level of $(1-\alpha)100\%$, is obtained as

$$CI_\xi = \left\{\xi : -2\log\frac{L_p(\xi)}{L_p(\widehat{\xi})} \leq \chi_1^2(1-\alpha)\right\}$$

or alternatively we can obtain it as

$$CI_\xi = \left\{\xi : \log L_p(\xi) \geq \log L_p(\widehat{\xi}) - \frac{\chi_1^2(1-\alpha)}{2}\right\}.$$

Uncertainty analysis

To account for the uncertainty associated with the estimated parameters, 95% CI is calculated from the optimal model using asymptotic theory. For each parameter, the CI provides a range of plausible values within which the true parameter value lies with 95% probability. Furthermore, the uncertainty associated with each return levels, which represent the rainfall intensity associated with specific return periods (e.g., 10-year, 50-year events), is assessed. We calculated the relative uncertainty in percentage (%) on return levels by considering the difference between the upper and lower bounds of the confidence intervals and normalizing it by the mean return level.

Model selection technique

The Akaike Information Criterion (AIC) [78] helps in selecting the best model by considering the minimized negative log-likelihood and the number of estimated parameters, as discussed by [17]. In this study, the AICc [79], a variant designed for small sample sizes, (i.e., $n < 40k$, where n is the sample size and k is the number of parameters) [80], is employed to identify the optimal GEV model from a range of non-stationary and stationary GEV models. The AICc is preferred in practical applications due to its superior performance compared to AIC, particularly in mitigating the risk of over-fitting the data, as demonstrated by [81]. For a given model characterized by ρ parameters, θ, and a sample size of n, the AICc of the model can be calculated as follows when fitted to an AMR series.

$$AICc = -2\log L(\theta|X) + 2\rho + \frac{2\rho(\rho+1)}{n-\rho-1}$$

the term "$-logL(\theta/X)$" refers to the minimized negative log likelihood function. The initial two components of AICc are identical to those of the standard AIC, and as the sample size (n) increases, AICc converges towards AIC. Additionally, the rescaled version of AICc, denoted as Δ_i as shown in Eq. (19), is utilized for ranking GEV models:

$$\Delta_i = AICc - \min(AICc) \tag{19}$$

Here, $\min(AICc)$ represents the smallest AICc value of all given models. The model with $\Delta_i = 0$ is considered the optimal model and the models with $\Delta_i \leq 2$ are deemed considerable options [82]. In this study, the importance of the optimal non-stationary model compared to the stationary model is assessed by the Likelihood Ratio Test (LRT) at a significance level, typically 5%. If the resulting p-value is greater than the chosen significance level (0.05), it indicates that the stationary model (S0) is preferred over the non-stationary model [15]. This ratio is used to determine whether the additional parameters in the more complex model significantly improve the fit to the data compared to the simpler model. The superior non-stationary model must exhibit a trend in its parameter values. The test involves comparing the negative log likelihood of the stationary model with that of the best non-stationary model [17]. For instance, suppose $NS4 : X \sim \text{GEV}((\mu_0 + E\mu_1), \sigma, \xi)$ is identified as the optimal non-stationary model. The negative log likelihoods of the stationary model ($nllS0$) and the best non-stationary model ($nllNS4$) are expressed as:

$$nllS0 = -\log L(\mu, \sigma, \xi|X)$$
$$nllNS4 = -\log L(\mu_0, \mu_1, \sigma, \xi|X)$$

Under the assumption that $\mu_1 = 0$, the test statistic for the likelihood ratio, calculated as double the disparity between $nllS0$ and $nllNS4$, adheres to an estimated Chi-square (χ^2_k) distribution with degrees of freedom (k) equivalent to the variance in the number of parameters between the stationary model (S0) and the superior (in this case, a non-stationary model (NS4)) [17]. The Chi-square distribution for this scenario is represented by Eq. (20):

$$2(nllS0 - nllNS4) \sim \chi^2_1 \tag{20}$$

The importance of the optimal non-stationary GEV, in comparison to the stationary, can be determined by the p-value associated with the χ^2_1 distribution.

After the optimal non-stationary GEV is determined, it is vital to verify its accuracy in representing the data. Quantile–Quantile (QQ) and Probability–Probability (PP) plots serve as effective tools for assessing the quality of the fitted model, as suggested by previous studies [15,81].

Return level estimation

Return level estimates for AMR series are also provided in this study. Return levels, denoted as z_p, represent the level of extreme events (maximum rainfall) that we can expect to be exceeded, on average, within a specified number of years [83,84]. Once we determine the best-fitted non-stationary GEV model for our data, we proceed to derive the return levels for extreme daily rainfall events. We define the probability P as the likelihood of an extreme event occurring at least once, on average, within a period of T years. Consequently, $P = \frac{1}{T}$. Precisely, these return levels correspond to levels surpassed by the yearly maxima with probability p. The return levels, z_p estimation, for stationary model is achieved by inverting the cumulative function of the GEV distribution as shown by Eqs. (21) and (22).

$$z_p = \mu - \frac{\sigma}{\xi}\left[1 - (-\log(1-p))^{-\xi}\right], \quad \xi \neq 0 \tag{21}$$

$$z_p = \mu - \sigma \log(-\log(1-p)), \quad \xi = 0 \tag{22}$$

Where $G(z_p) = 1 - p$, $y_p = -\log(1-p)$, and $0 < p < 1$.

For non-stationary models denoted as NS, the return levels z_p are obtained similarly by substituting parameter functions according to the non-stationarity conditions:

$$z_p = \mu(t) - \frac{\sigma(t)}{\xi}\left[1 - (-\log(1-p))^{-\xi}\right], \quad \xi \neq 0 \tag{23}$$

$$z_p = \mu(t) - \sigma(t)\log(-\log(1-p)), \quad \xi = 0 \tag{24}$$

Where $G(z_p) = 1 - p$, $y_p = -\log(1-p)$, and $0 < p < 1$.

Correlation analysis

The study conducted a correlation analysis to explore the relationships between the five potential covariates used and various extreme rainfall durations (1-, 3-, 5-day AMR) across all stations in the Southern Highlands region of Tanzania. This analysis aimed to assess the strength and direction of association between each covariate and extreme rainfall events, providing valuable insights into the factors influencing rainfall variability in the region. Correlation coefficients were computed to quantify the degree of linear relationship between the covariates and rainfall durations, with significance levels determined to identify statistically meaningful associations. The formula below is used to calculate the correlation coefficient;

$$r_{xy} = \frac{\sum_{I=1}^{n}(x-\bar{x})(y-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x-\bar{x})^2 . \sum_{i=1}^{n}(y-\bar{y})^2}}$$

The Pearson correlation coefficient (r_{xy}) quantifies the strength and direction of the linear relationship between two variables x and y. It is computed as the covariance between x and y divided by the product of their standard deviations [85]. This normalized measure ranges from -1 to 1, where values close to 1 indicate a strong positive linear relationship, values close to -1 indicate a strong negative linear relationship, and values close to 0 indicate no linear relationship.

In our study, we investigate the correlation between climate covariates and extreme rainfall events to understand their association. Previous research in the region has employed this analysis method [86]. To assess the statistical significance of the correlation coefficient, we use the student t-test. The t-statistic is calculated using the formula:

$$t = r_{xy}\sqrt{\frac{n-2}{1-r_{xy}^2}}, \tag{25}$$

where r_{xy} is the correlation coefficient and $n - 2$ represents the degrees of freedom. This test evaluates whether the observed correlation is significantly different from zero, indicating whether there is a meaningful relationship between the variables.

Results and discussion

Analysis of non-stationarity and identification of optimal covariates for chosen models

Based on the daily observations, calculations are performed for various rainfall duration (i.e., 1-,3-, and 5-days). Subsequently, the AMR series is derived from each rainfall duration as illustrated in Fig. 4. The Fig. 4 reveals an upward trend in 1- day duration for all stations with a 5% significance level. The optimal models for the four designated stations are outlined in Table 4. The outcome of the non-stationarity analysis using 1-day duration AMR data for Iringa is presented in Table 5, while the results for the other regions with their respective parameters are supplied in Supplementary materials. The NS38 model emerges as the most suitable for 1-day duration AMR for all stations. In NS38, a linear trend is incorporated into the location and scale parameters, considering local temperature variations and urbanization as potential covariates. The stationary model S0 is ranked 5th, 3rd, 9th, and 8th in Iringa, Mbeya, Rukwa, and Ruvuma, respectively. However, it proves to be an acceptable or considerable model across all stations based on the Δ_i value, as depicted in Table 4. Fig. 5 illustrates the PP and QQ plots of S0 and NS38 for the 1-day duration AMR data at Iringa. In this study, both the empirical and model quantiles represented in the quantile plot are measured in millimeters (mm). From Fig. 6, it is evident that the non-stationary GEV NS38 has effectively captured the quantiles of heavy rainfall intensities better than S0 for all regions.

Besides NS38, the NS34 non-stationary model outperforms S0 in Mbeya. For Iringa, three non-stationary models (NS2, NS34, and NS64) exhibit superiority over S0. Similarly, in Rukwa, seven non-stationary models (NS6, NS10, NS18, NS2, NS16, NS64, and NS34) demonstrate better performance than S0. In Ruvuma, six non-stationary models (NS18, NS10, NS6, NS12, NS64, and NS16) are superior to S0. It is noteworthy that most of these significantly superior models to S0 are also based on local temperature changes and urbanization covariates. Additionally, it is noteworthy that the time-based non-stationary model NS64 outperforms S0 in Iringa, Rukwa, and Ruvuma, while in Mbeya, S0 is found to be superior to NS64. Moreover, the time-based non-stationary model NS32 is deemed insignificant in all regions.

Similarly, the analysis extends to non-stationarity in the AMR series of various duration, with a summary provided in Table 4. Among the 65 GEV models considered, NS57 is identified as the optimal model for the 3-day duration AMR series in Iringa, and Rukwa, while NS46 emerges as the most suitable model for Mbeya, and Ruvuma. NS57 incorporates a linear trend in the location

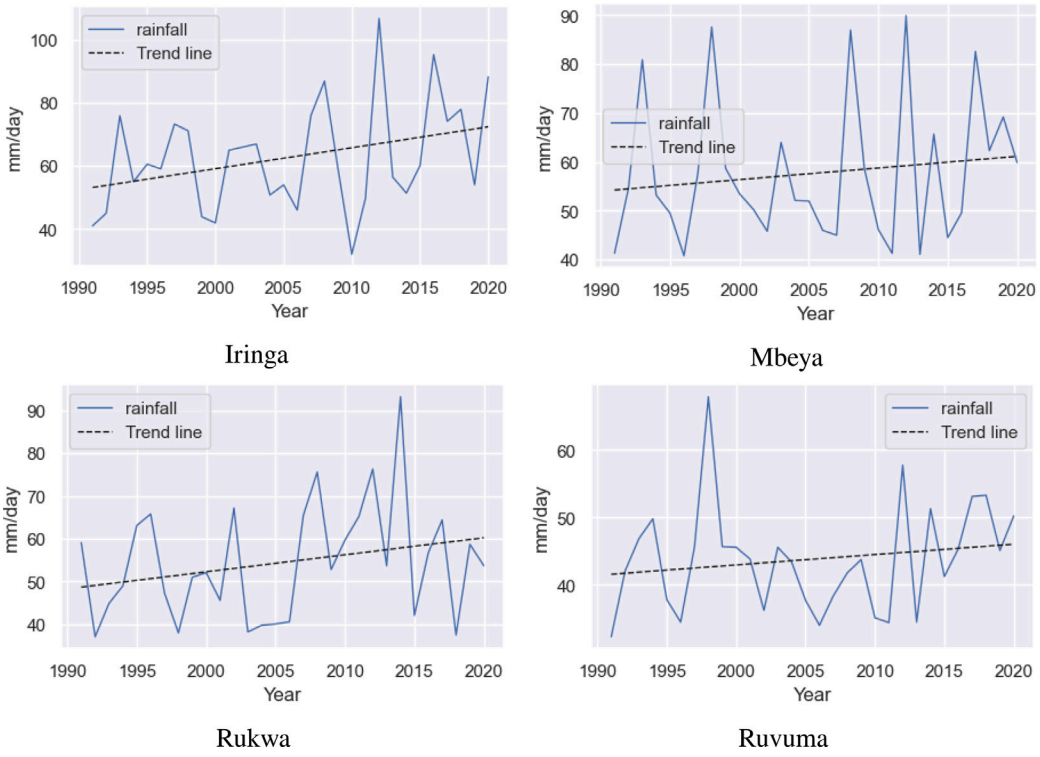


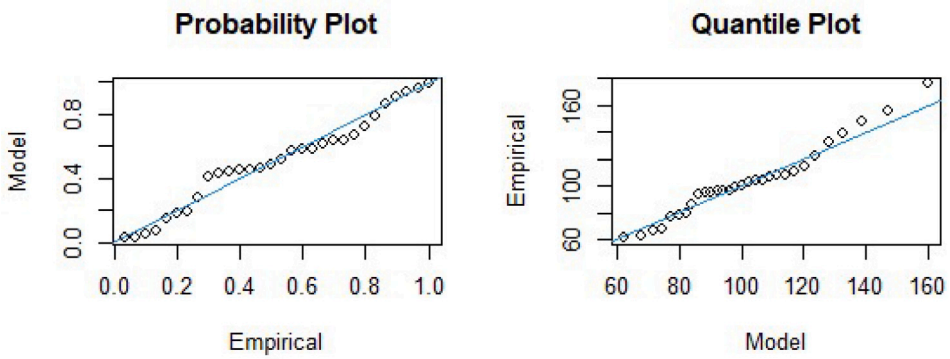
Fig. 4. Temporal trend in 1-day duration AMR series.

Table 4
A brief overview of non-stationarity assessment for all stations.

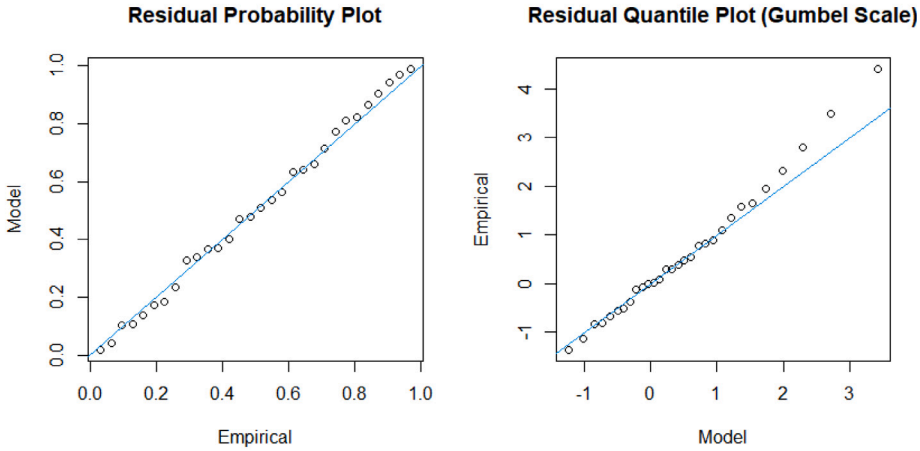
Station	Best model	1 day AMR	3 day AMR	5 day AMR
Iringa	Optimal model (covariate)	NS38 (U & LTC)	NS57 (G, E, & I)	NS57 (G, E, & I)
	Stationary model	A (S0, 5th)	NA (S0, 24th)	NA (S0, 43rd)
	Optimal time-based model	A (NS64, 4th)	NA (NS32, 31st)	NA (NS32, 46th)
	Δr	1.91	4.10	7.82
Mbeya	Optimal model (covariate)	NS38 (U & LTC)	NS46 (G, & I)	NS46 (G, & I)
	Stationary model	A (S0, 3rd)	NA (S0, 65th)	NA (S0, 10th)
	Optimal time-based model	A (NS64, 4th)	NA (NS32, 47th)	NA (NS32, 7th)
	Δr	1.58	26.64	2.65
Rukwa	Optimal model (covariate)	NS38 (U & LTC)	NS57 (G, E, & I)	NS57 (G, E, & I)
	Stationary model	A (S0, 9th)	NA (S0, 45th)	NA (SO, 26th)
	Optimal time-based model	A (NS64, 7th)	NA (NS64, 32nd)	NA (NS32, 23rd)
	Δr	1.98	11.53	5.82
Ruvuma	Optimal model (covariate)	NS38 (U & LTC)	NS46 (G, & I)	NS46 (G & I)
	Stationary model	A (S0, 8th)	NA (S0, 17th)	NA (S0, 34th)
	Optimal time-based model	A (NS64, 6th)	NA (NS32, 15th)	NA (NS32, 31st)
	Δr	1.83	3.48	7.92

Note: Δr is AICc difference between the Optimal model & S0.
A-Acceptable; NA-Not acceptable.

and scale parameters, including potential covariates such as global warming, ENSO cylcle, and the IOD. On the other hand, NS46 integrates a linear trend in the location and scale parameters, with covariates comprising global warming and the IOD. Besides NS57, there are eight considerable non-stationary models (NS55, NS60, NS52, NS45, NS42, NS39, NS26, NS37) superior to S0 in Iringa, and four considerable non-stationary models (NS30, NS50, NS46, NS41) that outperform S0 in Rukwa. In Mbeya, apart from NS46, six considerable non-stationary models (NS61, NS60, NS53, NS44, NS42, NS55) outperform S0, and ten considerable non-stationary models (NS17, NS31, NS38, NS10, NS47, NS63, NS14, NS49, NS34, NS62) are superior to S0 in Ruvuma. Additionally, for the 3-day



(a) S0



(b) NS38

Fig. 5. Model diagnostics for (a) Stationary, SO and (b) non-stationary, NS38 from 1- day AMR at Iringa station.

duration AMR, stationary models and two time-based covariate non-stationary models were found to be inconsiderable across all stations.

Similar to the 3-day duration rainfall series, NS57 emerges as the optimal model for the 5-day duration AMR series in Iringa and Rukwa, while NS46 is identified as the most suitable model for Mbeya and Ruvuma. Apart from NS46, there are five considerable non-stationary models (NS48, NS44, NS57, NS55, NS42) that outperform S0 in Mbeya, and five considerable non-stationary models (NS57, NS53, NS60, NS52, NS58) that outperform S0 in Ruvuma. Additionally, for NS57, eight considerable non-stationary models (NS56, NS45, NS60, NS24, NS51, NS35, NS62, NS28) are superior to S0 in Iringa, while eleven considerable non-stationary models (NS58, NS56, NS38, NS52, NS22, NS46, NS61, NS50, NS59, NS53, NS63) that surpass S0 in Rukwa. It is noteworthy that the optimal models and all considerable models for both the 3- and 5-day AMR series primarily incorporate global processes such as the global warming, ENSO cycle, and the IOD as potential covariates. Moreover, akin to the 3-day duration AMR, stationary models and two time-based covariate non-stationary models were deemed insignificant for the 5-day duration AMR across all stations. In addition, the QQ and PP plots of 3- and 5-day durations show a pattern similar to that of the 1-day duration rainfall series. It is evident from these plots that the optimal non-stationary model effectively represents the quantiles of extreme rainfall intensities, outperforming the stationary model S0 across all stations.

Based on the findings presented in Table 4, it can be observed that LTC and urbanization are identified as the most effective covariates for 1-day duration rainfall series, while GTA, ENSO, and IOD are deemed the most suitable covariates for 3-, 5-day duration rainfall across all regions. For a comprehensive overview of the ranking of 65 developed models, as well as overview of the QQ and PP plots for the best fit models across different duration and regions, readers are encouraged to refer to **Tables (1 to 13)** and **Figures (1 to 18)** provided in the Supplementary materials.

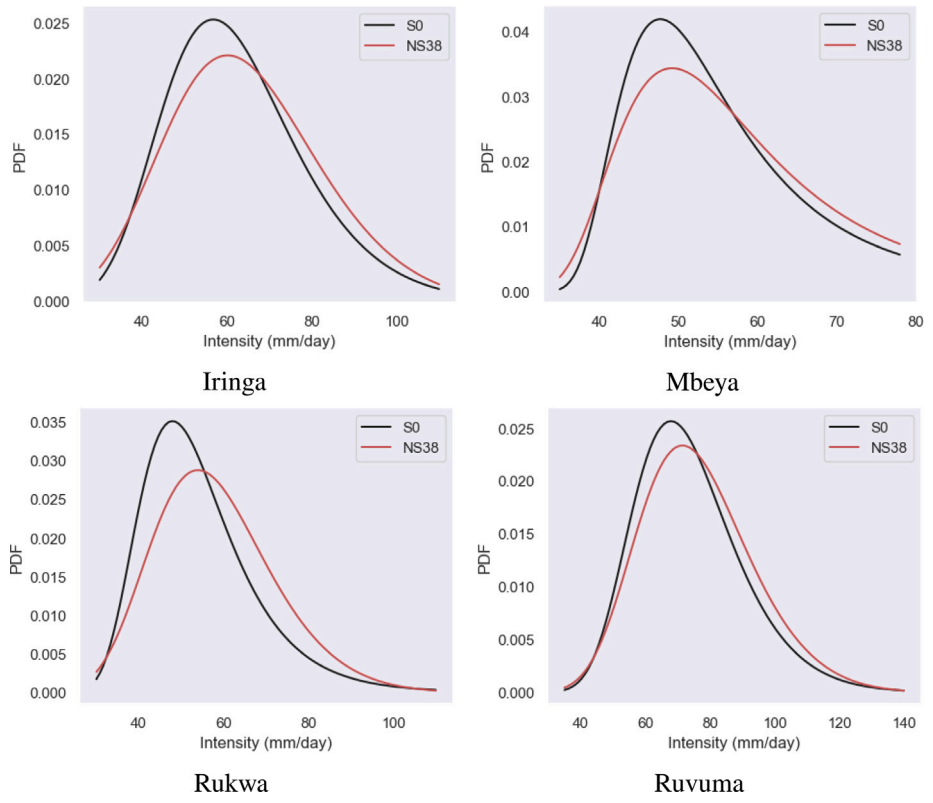


Fig. 6. Probability density function (PDF) plots of non-stationary (NS38) versus stationary (S0) for 1D AMR.

Table 5
Performance assessment of GEV models for 1-day AMR, ranked by Δ_i values for Iringa.

Model	AICc	Δ_i	Model	AICc	Δ_i	Model	AICc	Δ_i	Model	AICc	Δ_i
NS38	342.36	0.00	NS11	346.63	4.27	NS42	350.98	8.62	NS57	355.79	13.43
NS2	343.44	1.08	NS36	346.86	4.50	NS26	351.21	8.85	NS51	356.93	14.57
NS34	343.83	1.47	NS16	348.08	5.72	NS19	351.33	8.97	NS40	357.59	15.23
NS64	344.10	1.74	NS13	348.34	5.98	NS31	351.63	9.27	NS52	358.25	15.89
S0	344.27	1.91	NS28	348.46	6.10	NS20	352.23	9.87	NS50	358.90	16.54
NS43	344.85	2.49	NS30	348.56	6.20	NS29	352.57	10.21	NS53	360.03	17.67
NS10	344.90	2.54	NS9	348.64	6.28	NS44	353.02	10.66	NS58	361.21	18.85
NS24	345.17	2.81	NS14	348.78	6.42	NS47	353.1	10.74	NS60	362.17	19.81
NS4	345.24	2.88	NS8	348.87	6.51	NS56	353.25	10.89	NS59	362.78	20.42
NS5	346.05	3.69	NS12	348.94	6.58	NS45	353.42	11.06	NS41	363.11	20.75
NS23	346.19	3.83	NS25	348.99	6.63	NS54	353.83	11.47	NS62	364.45	22.09
NS18	346.21	3.85	NS27	349.06	6.70	NS46	354.14	11.78	NS15	366.23	23.87
NS1	346.21	3.85	NS33	349.13	6.77	NS49	354.31	11.95	NS61	366.45	24.09
NS39	346.26	3.90	NS21	349.17	6.81	NS32	354.39	12.03	NS63	368.67	26.31
NS3	346.39	4.03	NS7	349.33	6.97	NS48	354.8	12.44			
NS17	346.49	4.13	NS22	349.51	7.15	NS37	354.93	12.57			
NS6	346.58	4.22	NS35	350.19	7.83	NS55	354.99	12.63			

Furthermore, upon examining the non-stationary models that outperform the stationary GEV in each duration of rainfall series, it becomes evident that local factors serve as the most effective covariates for 1-day duration rainfall, while global ones prove to be more suitable for 3-, and 5-day duration rainfall across all regions. Moreover, in this study, the selection of physical factors influencing intense rainfall in the Southern Highlands region is based solely on available literature. It is important to note that these selected factors might have varying influence in various geographic areas. Therefore, it is recommended to carefully select covariates tailored to the specific area of study [19].

The fitting comparison of the best cumulative distribution function (CDF) to 1-day AMR series between NS38 and S0 in four stations is depicted in Figs. 7. From Fig. 7, it is evident that the patterns of extremely heavy rainfall can be effectively captured by non-stationary GEV models rather than stationary. Thus, the IDF curves obtained from the stationary GEV (S0) are expected to substantially underestimate the incidence of severe rainfall occurrences.

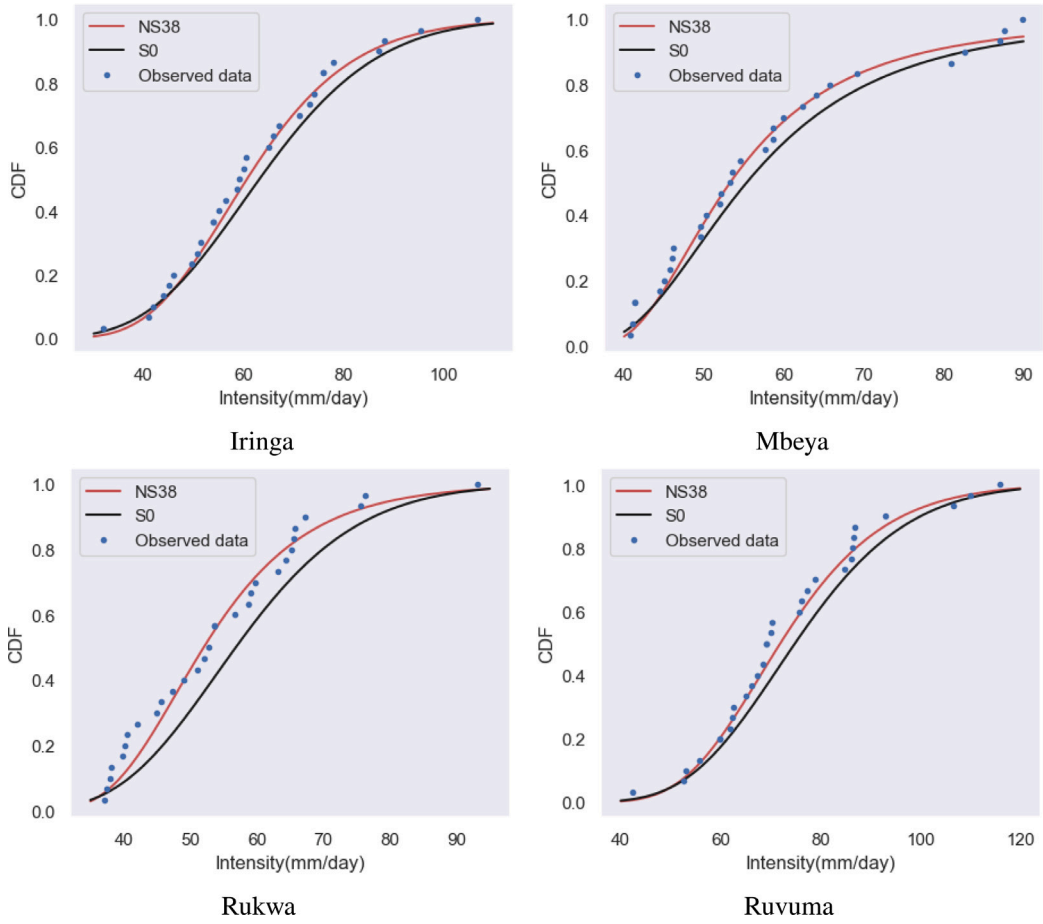


Fig. 7. CDF plots of non-stationary (NS38) versus stationary (S0) for 1D AMR.

Furthermore, the non-stationary model incorporating the time is deemed acceptable only for 1-day duration AMR series, but they never emerged as the optimal model for any rainfall duration across all stations. The LRT outcomes and parameter values of both stationary and optimal non-stationary GEV are presented in Table 12 in Supplementary materials. The LRT results distinctly indicate that the optimal non-stationary GEV outperform the stationary significantly for AMR series of all duration.

A detailed correlation study between climatic covariates and extreme rainfall

The correlation analysis reveals slight positive relationships between global climatic indices (ENSO, IOD, and GTA) and 3-, 5-day duration extreme rainfall in all stations. Conversely, local climatic factors such as Urbanization and local temperature changes exhibit slight positive relationships with 1-day duration extreme rainfall. Urbanization's impact on extreme rainfall is multifaceted, with changing land use patterns being the primary focus in this study. However, it is essential to acknowledge that urban aerosols and anthropogenic heat may also influence extreme rainfall intensity and frequency. According to [87,88], further investigations utilizing high-resolution atmospheric chemistry models are recommended to quantify these effects accurately. Additionally, incorporating additional dynamic and drivers of extreme rainfall, such as multiple-canopy parameterization, can provide deeper insights into urbanization's impact on extreme rainfall patterns [89]. Understanding the influence of urban agglomeration expansion on extreme rainfall in different directions is crucial for informed planning and development decisions.

Positive correlations between the IOD and extreme rainfall duration suggest a strong influence of anomalous warming over the western Indian Ocean on extreme rainfall in the Southern Highlands region. Comparatively, correlations with ENSO are weaker, indicating a less pronounced impact from remote influences originating in the central Pacific Ocean. Additionally, positive correlations between global warming (GTA) and extreme rainfall of all duration underscore the significant and robust changes in extreme rainfall intensities in response to global warming. The correlation results consistent with studies by [2,50,90,91], highlight the potential of climate indices as predictors of extreme rainfall in the study zone. Significant positive correlations among various climate indices suggest interdependencies that can inform predictive modeling efforts. Notably, rainfall patterns in the Southern

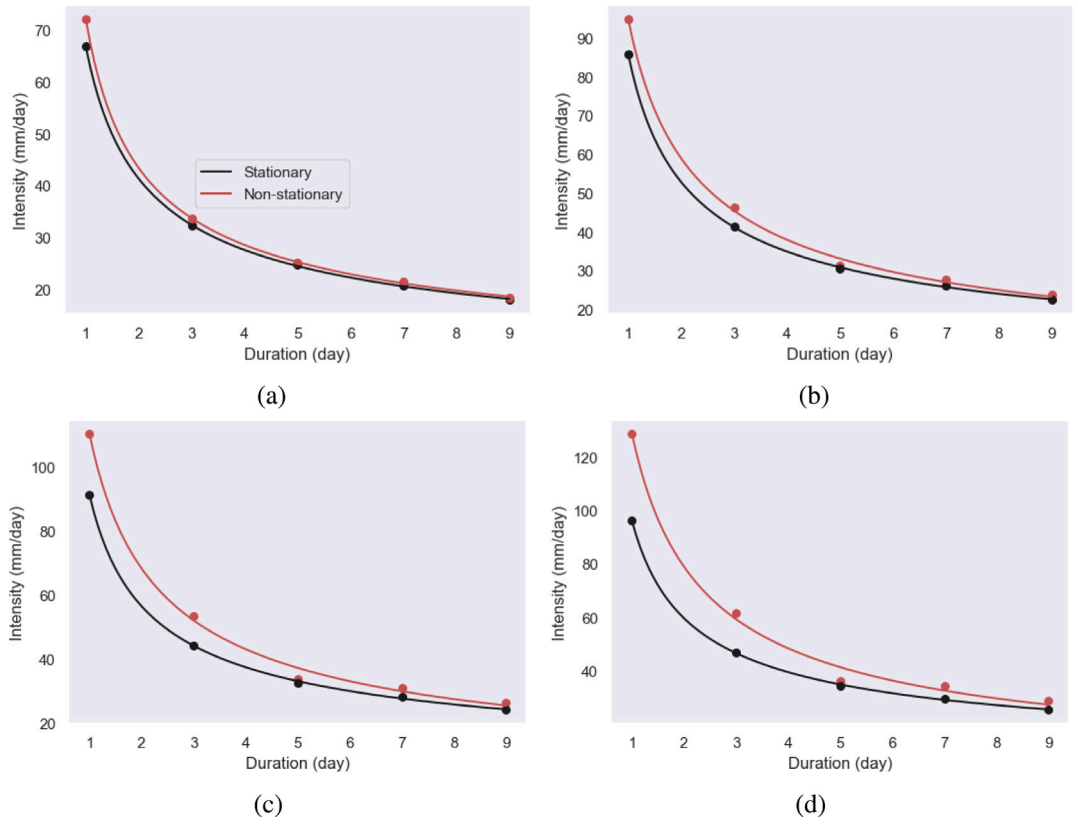


Fig. 8. Non-stationary IDF curves of Iringa station for (a) 5-, (b) 25-, (c) 50-, and (d) 100-years return period.

Highlands region exhibit sensitivity to the Indian SST, with warmer SST over the western Indian Ocean inducing heavy rainfall and cooler temperature linked to reduced rainfall [91].

Given the dependency of extreme rainfall on climate indices , it is recommended to incorporate these factors into forecasting models for the Southern Highlands region of Tanzania. By integrating climate indices into forecasting efforts, planners and policymakers can make informed decisions to mitigate the negative impacts of extreme rainfall events.

Updated return levels and the IDF curves derived from the non-stationary approach

Revised IDF curves for various return periods (2-, 5-, 10-, 25-, 50-, and 100-years) were developed based on the obtained optimal non-stationary models. The IDF curves for Iringa station corresponding to (5-, 25-, 50-, and 100-years) are depicted in Fig. 8. Similar IDF relationship curves for other stations and return periods are in the supplementary materials. The study indicates that IDF curves constructed from stationary GEV consistently downplay the intensities of severe events of all duration and return periods in all regions. Employing these IDF curves for infrastructure design could lead to projects being ill-equipped to withstand highly forthcoming extreme events. This means that, if the infrastructure designs rely on stationary rainfall IDF curves, they may become vulnerable to more frequent extreme rainfall events surpassing their capacity. For example, in Iringa region according to Fig. 8, a 25-year return period event with a 1-day duration has non-stationary extreme rainfall of 99.38 mm compared to 86 mm for stationary extreme rainfall. In other words, the non-stationary extreme rainfall for a 25-year return period and for 1- day duration is almost matching the stationary extreme rainfall intensity for a 100-year return period. This indicates a decrease in the return period of forthcoming extreme rainfall events under non-stationarity approach.

Fig. 9 illustrates the return levels against return periods obtained from the best models for 3- day AMR at each station along with their confidence intervals. This suggests that as the return period becomes longer (e.g., from 2 years to 100 years), the intensity of extreme rainfall events also tends to increase. The middle black line represent the estimated return levels, while the blue and red denote the upper and lower bound of the 95% confidence intervals around these estimates. Notably, the width of the confidence intervals tends to widen as the return period increases, indicating increasing uncertainty in the estimation of extreme rainfall intensity for longer return periods. This widening of confidence intervals reflects the greater variability and uncertainty associated with estimating extreme events over longer time scales.

Also, the relative uncertainty on return levels against the return period for 3- day AMR at each station is illustrated in Fig. 10. This analysis provides insights into how the uncertainty in extreme rainfall estimates varies with different return periods. More information on return level and relative uncertainty for other extreme rainfall duration are found in Supplementary materials.

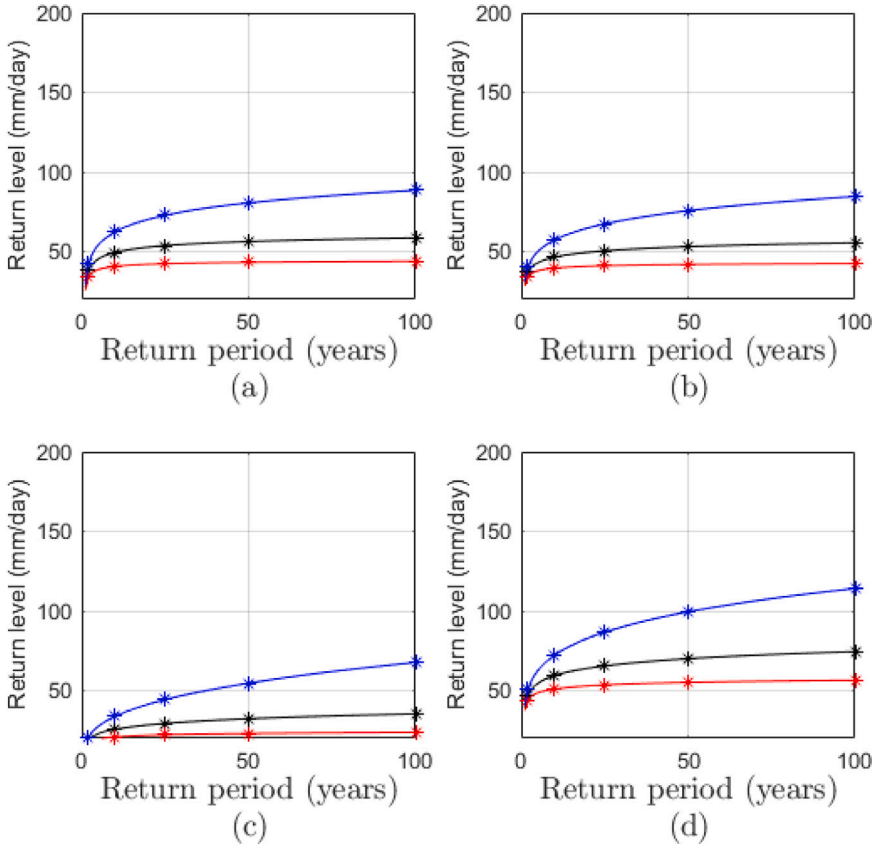


Fig. 9. Return levels with 95% CI for 3- day AMR at (a) Iringa, (b) Mbeya, (c) Rukwa, (d) Ruvuma.

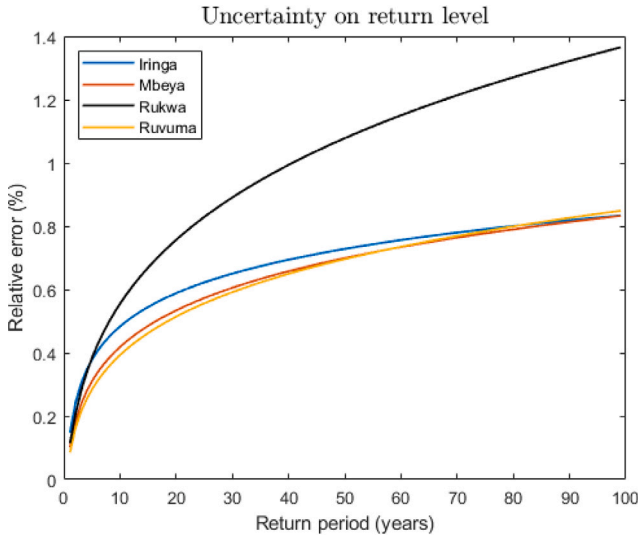


Fig. 10. Relative uncertainty on return levels against the return period for each station.

Summary and conclusions

This study delved into five key drivers of extreme rainfall, encompassing urbanization, LTC, GTA, and the large scale climatic indices (i.e., ENSO cycle and IOD). Urbanization data were derived from detailed satellite pictures, while other covariates were sourced from available data repositories. The analysis revealed notable findings: the local temperature variations and urbanization-based non-stationary model (NS38) emerged as optimal for 1-day duration AMR series across all regions, while the ENSO cycle, global warming (GTA), and IOD-based non-stationary model (NS57) proved superior for both 3- and 5-day durations AMR in Iringa and Rukwa. Also the global warming, and IOD-based non-stationary model (NS46) was found the best for 3- and 5-day durations AMR series in Mbeya and Ruvuma. Interestingly, the time-based non-stationary GEV (NS32 & NS64) never ranked as the optimal choice, except for the 1-day duration AMR series, where it was deemed considerable. Additionally, LRT results underscored the significance of the best non-stationary GEV models over stationary counterparts across all rainfall duration.

Moreover, the rainfall IDF relationship curves were established for each region based on the optimal non-stationary GEV models. A comparison between stationary and non-stationary IDF curves across various return periods highlights the fact that stationary models consistently underestimate severe events across all duration and return periods. Furthermore, it is observed that the return period of severe rainfall events in all regions is diminishing. The exploration of various covariate combinations, guided by the AICc, revealed the significance of factors such as urbanization, local temperature changes, global warming, and oceanic oscillations in modeling non-stationary IDF curves.

As observed, the relative uncertainty on return levels increases with the increase in return period. This implies that as we extrapolate our estimates to longer return periods, the uncertainty in our predictions also escalates. The increasing relative uncertainty underscores the challenges inherent in predicting extreme rainfall events over extended time horizons, necessitating cautious interpretation and decision-making in applications reliant on such forecasts.

Overall, the combined analysis of return levels with 95% confidence intervals and the relative uncertainty plot provides valuable insights into the dynamics of extreme rainfall behavior and the associated uncertainty across different return periods, aiding in robust risk assessment and mitigation strategies in diverse domains, including hydrology, infrastructure planning, and disaster management.

In conclusion, the correlation analysis highlights significant relationships between both global and local climatic indices and extreme rainfall events in the Southern Highlands region. The positive correlations observed between global climatic indices (ENSO, IOD, and GTA) and 3-, 5-day duration extreme rainfall underscore the influence of large-scale climatic phenomena on regional weather patterns. Conversely, the positive relationships between local climatic factors such as Urbanization and local temperature changes and 1-day duration extreme rainfall emphasize the importance of understanding local environmental dynamics. These findings suggest that incorporating climate indices into forecasting models can enhance the accuracy of extreme rainfall predictions, enabling planners and policymakers to make informed decisions to mitigate the negative impacts of extreme weather events.

In light of these findings, it is recommended to continue monitoring and studying the interactions between climatic indices and extreme rainfall events in the Southern Highlands region. Additionally, efforts should be made to integrate high-resolution atmospheric chemistry models and consider additional dynamic processes to better understand the influence of urbanization on extreme rainfall patterns. Overall, the findings presented in this analysis contribute valuable insights into the complex relationship between climate covariates and extreme rainfall events, providing a basis for more effective forecasting and adaptation strategies in the region. Moreover, while our study does not include sub-daily rainfall durations, we recognize the importance of this and suggest that future research could delve into shorter durations, such as hourly rainfall events, to provide a more comprehensive understanding of extreme rainfall patterns.

Finally, the conclusions drawn from this study are derived from data collected at four stations within the Southern Highlands of Tanzania. It is cautioned against generalizing these findings to broader spatial scales without additional analysis using dataset from the area of focus. Moreover, the selection of physical covariates influencing extreme rainfall in the Southern Highlands region relies solely on existing literature, which means that these same covariates may exhibit varying effects in different geographic areas. Thus, it is advisable to carefully consider physical factors specific to the area of focus.

CRediT authorship contribution statement

Erick A. Kyojo: Conceptualization, Writing manuscript, Methodology, Formal analysis. **Silas Mirau:** Visualization, Critical feedback, Review & editing, Insightful ideas and suggestions. **Sarah E. Osima:** Visualization, Critical feedback, Review & editing, Insightful ideas and suggestions. **Verdiana G. Masanja:** Supervision, Conceptualization, Manuscript - review & editing, Visualization, Critical feedback, Insightful ideas and suggestions.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors acknowledge TMA for providing data used in this study

Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.sciaf.2024.e02321.

References

[1] T.M. Authority, Statement on the Staus of Tanzania Climate in 2019, Tanzania Meteorological Authority, Dar es Salaam, Tanzania, 2019.
[2] K.B. Mafuru, I. Yonah, Assessing the variability of heavy rainfall during October to December rainfall season in Tanzania, 2021.
[3] L.P. Japheth, G. Tan, L.B. Chang'a, A.L. Kijazi, K.B. Mafuru, I.a. Yonah, Assessing the variability of heavy rainfall during October to December rainfall season in Tanzania, Atmospheric Clim. Sci. 11 (02) (2021) 267.
[4] S.I. Seneviratne, X. Zhang, M. Adnan, W. Badi, C. Dereczynski, A. Di Luca, S.M. Vicente-Serrano, M. Wehner, B. Zhou, 11 Chapter 11: weather and climate extreme events in a changing climate, 2021.
[5] L.B. Chang'a, A.L. Kijazi, P.M. Luhunga, H.K. Ng'ongolo, H.I. Mtongori, Spatial and Temporal Analysis of Rainfall and Temperature Extreme Indices in Tanzania, Scientific Research Publishing, 2017.
[6] W. Cai, C. Zhang, H.P. Suen, S. Ai, Y. Bai, J. Bao, B. Chen, L. Cheng, X. Cui, H. Dai, The 2020 China report of the lancet countdown on health and climate change, Lancet Public Health 6 (1) (2021) e64–e81.
[7] P.R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.O. Pörtner, D. Roberts, P. Zhai, R. Slade, S. Connors, R. Van Diemen, IPCC, 2019: Climate Change and Land: an IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems, Intergovernmental Panel on Climate Change (IPCC), 2019.
[8] C.C. IPCC, Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Intergovernmental Panel on Climate Change (IPCC), Cambridge University Press, New York, 2007.
[9] C.S. Omambia, Y. Gu, The cost of climate change in Tanzania: impacts and adaptations, J. Am. Sci. 6 (3) (2010) 182–196.
[10] E.A. Kyojo, S.S. Mirau, S.E. Osima, V.G. Masanja, Frequentist and Bayesian approaches in modeling and prediction of extreme rainfall series: A case study from Southern Highlands Region of tanzania, Adv. Meteorol. 2024 (2024).
[11] H.-J. Lüdecke, G. Müller-Plath, M.G. Wallace, S. Lüning, Decadal and multidecadal natural variability of African rainfall, J. Hydrol.: Reg. Stud. 34 (2021) 100795.
[12] A.F. Jenkinson, The frequency distribution of the annual maximum (or minimum) values of meteorological elements, Q. J. R. Meteorol. Soc. 81 (348) (1955) 158–171.
[13] D.F. Campos-Aranda, Joint frequency analysis of peak flow and volumes of floods with gumbel marginals, Tecnol. Cienc. Agua 14 (3) (2023) 1–50.
[14] D. Urrea Méndez, M. del Jesus, Estimating extreme monthly rainfall for Spain using non-stationary techniques, Hydrol. Sci. J. 68 (7) (2023) 903–919.
[15] S. Coles, J. Bawa, L. Trenner, P. Dorazio, An Introduction to Statistical Modeling of Extreme Values, vol. 208, Springer, 2001.
[16] R.L. Smith, Maximum likelihood estimation in a class of nonregular cases, Biometrika 72 (1) (1985) 67–90.
[17] R.W. Katz, Statistical methods for nonstationary extremes, in: Extremes in a changing climate: Detection, analysis and uncertainty, Springer, 2013, pp. 15–37.
[18] D. Jakob, Nonstationarity in extremes and engineering design, in: Extremes in a Changing Climate: Detection, Analysis and Uncertainty, Springer, 2012, pp. 363–417.
[19] V. Agilan, N. Umamahesh, What are the best covariates for developing non-stationary rainfall intensity-duration-frequency relationship? Adv. Water Resour. 101 (2017) 11–22.
[20] T.M. Vu, A.K. Mishra, Nonstationary frequency analysis of the recent extreme precipitation events in the United States, J. Hydrol. 575 (2019) 999–1010.
[21] A. Mondal, P.P. Mujumdar, Modeling non-stationarity in intensity, duration and frequency of extreme rainfall over India, J. Hydrol. 521 (2015) 217–231.
[22] L. Xu, H. Zhou, L. Du, H. Yao, H. Wang, Precipitation trends and variability from 1950 to 2000 in arid lands of Central Asia, J. Arid Land 7 (2015) 514–526.
[23] N.R. Cavanaugh, A. Gershunov, A.K. Panorska, T.J. Kozubowski, The probability distribution of intense daily precipitation, Geophys. Res. Lett. 42 (5) (2015) 1560–1567.
[24] K.E. Kunkel, T.R. Karl, D.R. Easterling, K. Redmond, J. Young, X. Yin, P. Hennon, Probable maximum precipitation and climate change, Geophys. Res. Lett. 40 (7) (2013) 1402–1408.
[25] U. Desa, World Urbanization Prospects, the 2011 Revision, Population Division, Department of Economic and Social Affairs, United Nations Secretariat, 2014.
[26] Un-Habitat, State of the World's Cities 2008/9: Harmonious Cities, Earthscan, 2008.
[27] Y. Zhang, J.A. Smith, L. Luo, Z. Wang, M.L. Baeck, Urbanization and rainfall variability in the Beijing metropolitan region, J. Hydrometeorol. 15 (6) (2014) 2219–2235.
[28] L. Yang, F. Tian, D. Niyogi, A need to revisit hydrologic responses to urbanization by incorporating the feedback on spatial rainfall patterns, Urban Clim. 12 (2015) 128–140.
[29] G. Villarini, On the seasonality of flooding across the continental United States, Adv. Water Resour. 87 (2016) 80–91.
[30] S. Miao, F. Chen, Q. Li, S. Fan, Impacts of urban processes and urbanization on summer precipitation: A case study of heavy rainfall in Beijing on 1 August 2006, J. Appl. Meteorol. Climatol. 50 (4) (2011) 806–825.
[31] R. Ajayamohan, S.A. Rao, Indian ocean dipole modulates the number of extreme rainfall events over India in a warming environment, J. Meteorol. Soc. Japan. Ser. II 86 (1) (2008) 245–252.
[32] W. Cai, A. Santoso, G. Wang, E. Weller, L. Wu, K. Ashok, Y. Masumoto, T. Yamagata, Increased frequency of extreme Indian ocean dipole events due to greenhouse warming, Nature 510 (7504) (2014) 254–258.
[33] V.R. Golroudbary, Y. Zeng, C.M. Mannaerts, Z.B. Su, Attributing seasonal variation of daily extreme precipitation events across The Netherlands, Weather Clim. Extremes 14 (2016) 56–66.
[34] S. Sharma, P. Mujumdar, On the relationship of daily rainfall extremes and local mean temperature, J. Hydrol. 572 (2019) 179–191.
[35] H. Van de Vyver, Bayesian estimation of rainfall intensity–duration–frequency relationships, J. Hydrol. 529 (2015) 1451–1463.
[36] L. Cheng, A. AghaKouchak, Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate, Sci. Rep. 4 (1) (2014) 7093.
[37] A. Yilmaz, B. Perera, Extreme rainfall nonstationarity investigation and intensity–frequency–duration relationship, J. Hydrol. Eng. 19 (6) (2014) 1160–1172.
[38] K. Näschen, B. Diekkrüger, M. Evers, B. Höllermann, S. Steinbach, F. Thonfeld, The impact of land use/land cover change (LULCC) on water resources in a tropical catchment in tanzania under different climate change scenarios, Sustainability 11 (24) (2019) 7083.
[39] S. Bisanda, W. Mwangi, H. Verkuijl, A.J. Moshi, P. Anandajayasekeram, Adoption of Maize Production Technologies in the Southern Highlands of Tanzania, CIMMYT, 1998.
[40] T.J. Ngailo, N. Shaban, J. Reuder, M.D. Mesquita, E. Rutalebwa, I. Mugume, C. Sangalungembe, Assessing weather research and forecasting (WRF) model parameterization schemes skill to simulate extreme rainfall events over Dar es Salaam on 21 December 2011, J. Geosci. Environ. Prot. 6 (01) (2018) 36.
[41] L. Cheng, A. AghaKouchak, E. Gilleland, R.W. Katz, Non-stationary extreme value analysis in a changing climate, Clim. Change 127 (2014) 353–369.

[42] B. Renard, X. Sun, M. Lang, Bayesian methods for non-stationary extreme value analysis, in: Extremes in a Changing Climate: Detection, Analysis and Uncertainty, Springer, 2012, pp. 39–95.
[43] K.E. Trenberth, Changes in precipitation with climate change, Clim. Res. 47 (1–2) (2011) 123–138.
[44] J.M. Shepherd, H. Pierce, A.J. Negri, Rainfall modification by major urban areas: Observations from spaceborne rain radar on the TRMM satellite, J. Appl. Meteorol. Climatol. 41 (7) (2002) 689–701.
[45] X. Wang, Z. Wang, Y. Qi, H. Guo, Effect of urbanization on the winter precipitation distribution in Beijing area, Sci. China Ser. D: Earth Sci. 52 (2) (2009) 250–256.
[46] S.A. Changnon, Jr., R.G. Semonin, F. Huff, A hypothesis for urban rainfall anomalies, J. Appl. Meteorol. Climatol. 15 (6) (1976) 544–560.
[47] J. Dettwiller, S. Changnon Jr., Possible urban effects on maximum daily rainfall at Paris, St. Louis and Chicago, J. Appl. Meteorol. (1962-1982) (1976) 517–519.
[48] J.M. Shepherd, S.J. Burian, Detection of urban-induced rainfall anomalies in a major coastal city, Earth Interact. 7 (4) (2003) 1–17.
[49] C.M. Kishtawal, D. Niyogi, M. Tewari, R.A. Pielke Sr, J.M. Shepherd, Urbanization signature in the observed heavy rainfall climatology over India, Int. J. Climatol. 30 (13) (2010) 1908–1916.
[50] H.K. Ame, A.L. Kijazi, L.B. Changa, K.B. Mafuru, M.K. Ngwali, M.M. Faki, A.O. Hmad, M.K. Miraji, Rainfall variability over Tanzania during October to December and its association with sea surface temperature (SST), Atmos. Clim. Sci. 11 (02) (2021) 324.
[51] B.F. Rweyemamu, R.W. Mulokozi, Understanding nexuses between precipitation changes and climate change and variability in semi-arid lowlands of Mwanga District, Tanzania, Afr. J. Environ. Sci. Technol. 13 (1) (2019) 52–65.
[52] D. Mbigi, Z. Xiao, Analysis of rainfall variability for the October to December over Tanzania on different timescales during 1951–2015, Int. J. Climatol. 41 (14) (2021) 6183–6204.
[53] E. Tumaini, Analysis of Rainfall Characteristics in Tanzania for Climate Change Signals (Ph.D. thesis), 2009.
[54] L. Ogallo, Relationships between seasonal rainfall in East Africa and the Southern oscillation, J. Climatol. 8 (1) (1988) 31–43.
[55] S. Hastenrath, A. Nicklis, L. Greischar, Atmospheric-hydrospheric mechanisms of climate anomalies in the western equatorial Indian ocean, J. Geophys. Res.: Oceans 98 (C11) (1993) 20219–20235.
[56] T. Kabanda, M. Jury, Inter-annual variability of short rains over northern Tanzania, Clim. Res. 13 (3) (1999) 231–241.
[57] H. Zelle, G. Appeldoorn, G. Burgers, G.J. van Oldenborgh, The relationship between sea surface temperature and thermocline depth in the eastern equatorial Pacific, J. Phys. Oceanogr. 34 (3) (2004) 643–655.
[58] X. Zhang, J. Wang, F.W. Zwiers, P.Y. Groisman, The influence of large-scale climate variability on winter maximum daily precipitation over North America, J. Clim. 23 (11) (2010) 2902–2915.
[59] M.Q. Villafuerte, J. Matsumoto, Significant influences of global mean temperature and ENSO on extreme rainfall in Southeast Asia, J. Clim. 28 (5) (2015) 1905–1919.
[60] A. Timmermann, J. Oberhuber, A. Bacher, M. Esch, M. Latif, E. Roeckner, Increased El Niño frequency in a climate model forced by future greenhouse warming, Nature 398 (6729) (1999) 694–697.
[61] N. Saji, B.N. Goswami, P. Vinayachandran, T. Yamagata, A dipole mode in the tropical Indian ocean, Nature 401 (6751) (1999) 360–363.
[62] K. Ashok, Z. Guan, T. Yamagata, Impact of the Indian ocean dipole on the relationship between the Indian monsoon rainfall and ENSO, Geophys. Res. Lett. 28 (23) (2001) 4499–4502.
[63] K. Ashok, N. Saji, On the impacts of ENSO and Indian ocean dipole events on sub-regional Indian summer monsoon rainfall, Nat. Hazards 42 (2007) 273–285.
[64] T.M. Vul, Nonstationary frequency analysis of the recent extreme rainfall events in the United States, J. Hydrol. (2019).
[65] R. Fisher, L. Tippett, On the estimation of the frequency distributions of the largest or smallest member of a sample, Proc. Camb. Phil. Soc. 24 (1928) 180–190.
[66] B. Gnedenko, Sur la distribution limite du terme maximum d'une série aléatoire, Ann. of Math. 44 (3) (1943) 423–453.
[67] E. Smith, Bayesian Modelling of Extreme Rainfall Data (Ph.D. thesis), University of Newcastle upon Tyne England, UK, 2005.
[68] E. Gilleland, R.W. Katz, New software to analyze how extremes change over time, EOS Trans. Am. Geophys. Union 92 (2) (2011) 13–14.
[69] F. De Paola, M. Giugni, F. Pugliese, A. Annis, F. Nardi, GEV parameter estimation and stationary vs. non-stationary analysis of extreme rainfall in African test cities, Hydrology 5 (2) (2018) 28.
[70] K. Johnson, J. Smithers, R. Schulze, A review of methods to account for impacts of non-stationary climate data on extreme rainfalls for design rainfall estimation in South Africa, J. South Afr. Inst. Civ. Eng. 63 (3) (2021) 55–61.
[71] J. Hounkpè, B. Diekkrüger, D.F. Badou, A.A. Afouda, Non-stationary flood frequency analysis in the Ouémé River Basin, Benin Republic, Hydrology 2 (4) (2015) 210–229.
[72] D. Cooley, D. Nychka, P. Naveau, Bayesian spatial modeling of extreme precipitation return levels, J. Amer. Statist. Assoc. 102 (479) (2007) 824–840.
[73] A. Sarhadi, E.D. Soulis, Time-varying extreme rainfall intensity-duration-frequency curves in a changing climate, Geophys. Res. Lett. 44 (5) (2017) 2454–2463.
[74] M.D. Risser, M.F. Wehner, Attributable human-induced changes in the likelihood and magnitude of the observed extreme precipitation during Hurricane Harvey, Geophys. Res. Lett. 44 (24) (2017) 12–457.
[75] S.G. Coles, M.J. Dixon, Likelihood-based inference for extreme value models, Extremes 2 (1999) 5–23.
[76] J. Beirlant, Y. Goegebeur, J. Segers, J.L. Teugels, Statistics of Extremes: Theory and Applications, John Wiley & Sons, 2006.
[77] M.I. Gomes, M. Fraga Alves, C. Neves, Análise de valores extremos: uma introduçao, Ediçoes SPE & INE (2013).
[78] H. Akaike, A new look at the statistical model identification, IEEE Trans. Autom. Control 19 (6) (1974) 716–723.
[79] J.E. Cavanaugh, Unifying the derivations for the Akaike and corrected Akaike information criteria, Statist. Probab. Lett. 33 (2) (1997) 201–208.
[80] C.N. Charles, Reliability and Uncertainty in Diffusion MRI Modelling (Ph.D. thesis), 2016.
[81] S. Sugahara, R.P. Da Rocha, R. Silveira, Non-stationary frequency analysis of extreme daily rainfall in Sao Paulo, Brazil, Int. J. Climatol.: J. R. Meteorol. Soc. 29 (9) (2009) 1339–1349.
[82] K.P. Burnham, D.R. Anderson, Multimodel inference: understanding AIC and BIC in model selection, Sociol. Methods Res. 33 (2) (2004) 261–304.
[83] S. Feng, S. Nadarajah, Q. Hu, Modeling annual extreme precipitation in China using the generalized extreme value distribution, J. Meteorol. Soc. Japan. Ser. II 85 (5) (2007) 599–613.
[84] J.D. Salas, J. Obeysekera, Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events, J. Hydrol. Eng. 19 (3) (2014) 554–568.
[85] N.J. Gogtay, U.M. Thatte, Principles of correlation analysis, J. Assoc. Physicians India 65 (3) (2017) 78–81.
[86] A.A. Lukali, S.E. Osima, Y. Lou, K.H. Kai, Assessing the impacts of climate change and variability on maize (zea mays) yield Over Tanzania, Atmos. Clim. Sci. 11 (3) (2021) 569–588.
[87] J.-Y. Han, J.-J. Baik, A.P. Khain, A numerical study of urban aerosol impacts on clouds and precipitation, J. Atmos. Sci. 69 (2) (2012) 504–520.
[88] N. WanShu, B. Zaitchik, N. GuangHeng, S. Ting, Impacts of anthropogenic heat on summertime rainfall in Beijing, 2017.
[89] S.C. Van Den Heever, W.R. Cotton, Urban aerosol impacts on downwind convective storms, J. Appl. Meteorol. Climatol. 46 (6) (2007) 828–850.
[90] P. Mwinuka, C. Uiso, Prediction of rainfall in the southern highlands of Tanzania, Tanzania J. Sci. 43 (1) (2017) 36–46.
[91] Y. Mbululo, F. Nyihirani, Climate Characteristics Over Southern Highlands Tanzania, Scientific Research, 2012.