Asian Journal of Probability and Statistics

Volume 26, Issue 7, Page 34-47, 2024; Article no.AJPAS.116222
ISSN: 2582-0230

Rainfall Pattern in Kenya: Bayesian Non-parametric Model Based on the Normalized Generalized Gamma Process

Amos Kipkorir Langat [a*] and John Kamwele Mutinda [b]

[a] Pan African University Institute for Basic Sciences, Technology and Innovation, JKUAT, Nairobi, Kenya.
[b] University of Science and Technology of China, Hefei, China.

Author's contribution

This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

Article Information

DOI: https://doi.org/10.9734/ajpas/2024/v26i7628

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/116222

Original Research Article

Received: 04/03/2024
Accepted: 07/05/2024
Published: 13/06/2024

Abstract

Understanding the pattern of rainfall in Kenya is crucial for a range of sectors, including agriculture, water management, and disaster risk reduction. In this research, we propose a Bayesian non-parametric approach to model the rainfall patterns in Kenya. Specifically, we use a hierarchical Dirichlet process mixture model to cluster the rainfall stations and identify groups of stations with similar rainfall patterns. We then model the rainfall distribution within each group using a Bayesian non-parametric model based on the normalized generalized gamma process. We apply our method to a dataset of daily rainfall measurements from 150 stations across Kenya for the period 1980-2021. Our results reveal distinct regional patterns of rainfall, with some regions experiencing bimodal rainfall patterns while others have unimodal patterns. We also find that the rainfall distribution within each region exhibits heavy tails and skewedness, which cannot be accurately captured by parametric models. In conclusion, our approach provides a flexible and interpretable framework for modeling complex spatio-temporal data such as rainfall patterns, and can inform decision-making in various sectors.

**Corresponding author: Email: moskiplangat@gmail.com;*

Cite as: *Langat, Amos Kipkorir, and John Kamwele Mutinda. 2024. "Rainfall Pattern in Kenya: Bayesian Non-Parametric Model Based on the Normalized Generalized Gamma Process". Asian Journal of Probability and Statistics 26 (7):34-47. https://doi.org/10.9734/ajpas/2024/v26i7628.*

Keywords: Rainfall pattern; Kenya; bayesian non–parametric; Kenya meteorological department.

1 Introduction

Climate change is one of the most significant challenges facing the world today, with the potential to impact various sectors such as agriculture, water management, and disaster risk reduction [1]. One of the critical climatic variables affected by climate change is rainfall. Changes in rainfall patterns can lead to severe consequences such as droughts, floods, and crop failures, affecting the livelihoods of millions of people worldwide [2]. Therefore, understanding the patterns of rainfall is essential for developing effective strategies to manage these sectors [3].

Kenya is one country that is highly vulnerable to changes in rainfall patterns. Agriculture is a critical sector in Kenya, contributing 30% of the country's gross domestic product (GDP) and employing 70% of the country's population [4]. Rainfall is the primary source of water for agriculture in Kenya, making it a crucial factor for food security and economic development. However, rainfall in Kenya is highly variable, with significant spatial and temporal variability [5].

Several studies have investigated the patterns of rainfall in Kenya using various methods. For example, traditional parametric models such as the Gaussian distribution have been used to model rainfall patterns in Kenya [6]. However, these models may not capture the heavy tails and skewedness often observed in rainfall data, leading to inaccurate predictions. Non-parametric models have also been used to model rainfall patterns in Kenya, such as the kernel density estimation (KDE) method. According to Chen and Brissette [7], various models for the stochastic generation of daily precipitation amounts have been reviewed and evaluated. However, these models may not be suitable for high-dimensional and heterogeneous datasets such as rainfall measurements from multiple stations.

Therefore, there is a need for more flexible and interpretable models to capture the complex spatio-temporal variability of rainfall in Kenya [8,9]. In this paper, we propose a Bayesian non-parametric approach to model the rainfall patterns in Kenya. Specifically, we use a hierarchical Dirichlet process mixture model to cluster the rainfall stations and identify groups of stations with similar rainfall patterns [10]. We then model the rainfall distribution within each group using a Bayesian non-parametric model based on the normalized generalized gamma process. This approach provides a flexible and interpretable framework for modeling complex spatio-temporal data such as rainfall patterns.

Our approach has several advantages over traditional parametric models and non-parametric models. First, it can handle the high-dimensional and heterogeneous nature of the data. Second, it allows for flexible and interpretable modeling of the rainfall distribution, including heavy tails and skewedness. Third, it can identify groups of stations with similar rainfall patterns, providing insights into the regional patterns of rainfall in Kenya.

2 Methodology

In this study, we propose a Bayesian non-parametric approach to model the rainfall patterns in Kenya. Specifically, we use a hierarchical Dirichlet process mixture model to cluster the rainfall stations and identify groups of stations with similar rainfall patterns. We then model the rainfall distribution within each group using a Bayesian non-parametric model based on the normalized generalized gamma process.

2.1 Data preprocessing

We obtained a dataset of daily rainfall measurements from 150 stations across Kenya for the period 1980-2021. The data were provided by the Kenya Meteorological Department. The raw data contained missing values and outliers, which we removed by applying k-nearest neighbor imputation and outlier detection using the interquartile range method. We also normalized the data using a Box-Cox transformation to ensure that the data met the assumptions of the statistical models.

2.2 Hierarchical dirichlet process mixture model

We used a hierarchical Dirichlet process mixture model to cluster the rainfall stations and identify groups of stations with similar rainfall patterns. The model assumes that each station belongs to one of K groups, and the rainfall measurements within each group follow a common probability distribution. The model is specified as follows:

- For each station $i = 1,...., N$:
- Sample the group assignment z_i from a categorical distribution with probabilities $\pi_1,...., \pi_K$
- For each group k = 1, ..., K:
- Sample the group-specific probability distribution G_k from a base distribution H
- If $z_i = k$, sample the rainfall measurement y_i from G_k

We used the normalized gamma process as the base distribution H . We assigned a gamma distribution with hyper parameters a and b as the prior for G_k . We placed gamma priors with hyper parameters c and d on the hyper parameters a and b .

We used a Markov chain Monte Carlo (MCMC) algorithm to sample from the posterior distribution of the model parameters, including the group assignments z_i , the group-specific probability distributions G_k , and the hyper parameters of the gamma distribution. We used the Gibbs sampler to update the group assignments z_i and the Metropolis-Hastings algorithm to update the hyper parameters.

2.3 Bayesian non-parametric model

Within each group k , we modeled the rainfall distribution using a Bayesian non-parametric model based on the normalized generalized gamma process. The model assumes that the rainfall measurements follow a generalized gamma distribution, which is a flexible and interpretable distribution that can capture heavy tails and skewedness. The model is specified as follows:

- For each group $k = 1,...., K$:
- Sample the shape parameter a_k from a gamma distribution with hyper parameters e and f
- Sample the scale parameter β_k from a gamma distribution with hyper parameters g and h
- Sample the skewness parameter δ_k from a normal distribution with mean 0 and variance σ^2
- Sample the rainfall measurements y_i from a generalized gamma distribution with parameters $a_k, \ \beta_k, \ and \ \delta_k$

We used a MCMC algorithm to sample from the posterior distribution of the model parameters, including the shape parameter a_k , the scale parameter β_k , the skewness parameter δ_k , and the hyper parameters $e, \ f, \ g, \ h,$ and σ^2 . We used the Metropolis-Hastings algorithm to update the model parameters.

2.4 Model selection

We determined the optimal number of groups K using the Bayesian information criterion (BIC) and performed model selection using cross-validation. We randomly split the dataset into training and testing sets and used the

training set to fit the model with different values of K. We then evaluated the performance of the model on the testing set using the mean squared error (MSE) and the coefficient of determination $\left(R^2\right)$.
This implies that, our mathematical model is summarized as follows:

Hierarchical Dirichlet process mixture model:

Let $x=\left(x_1, \quad x_2, \ldots\ldots, x_N\right)$ be the vector of rainfall measurements from N stations.

The hierarchical Dirichlet process mixture model can be written as:

$$p(x \mid \theta)=\int p(x \mid \theta, Z) p(Z \mid \beta) p(\theta \mid \alpha) dZ d\theta$$

Where θ the set of model parameters is $Z=\left(z_1, \quad z_2, \ldots\ldots, z_N\right)$ is the vector of latent cluster assignments, β is the hyperparameter of the Dirichlet process, and α is the hyperparameter of the base distribution.

The likelihood function $p(x \mid \theta, Z)$ is given by a mixture of Gaussian distributions, where each cluster has a different mean and variance:

$$p(x \mid \theta, Z)=\prod_k N\left(x_k \mid \mu_k, \sigma_k^2\right)^{(N_k)}$$

Where N_k is the number of data points assigned to cluster k, and μ_k and σ_k^2 are the mean and variance of cluster k.

The prior distributions are given by:

$$p(Z \mid \beta)=\prod_k b\left(N_k \mid \beta / N\right)$$

$$p(\theta \mid \alpha)=DP(\theta \mid \alpha G)$$

Where $b\left(N_k \mid \beta / N\right)$ is the probability mass function of the Poisson-Dirichlet process, and $DP(\theta \mid \alpha G)$ is the probability distribution of the Dirichlet process.

The joint posterior distribution of θ and Z is proportional to the likelihood function and the prior distributions:

$$p(\theta, Z \mid x.\alpha, \beta) \propto p(x \mid \theta, Z) p(Z \mid \beta) p(\theta \mid \alpha)$$

The model parameters θ are estimated using a Markov Chain Monte Carlo (MCMC) algorithm, which involves simulating from the joint posterior distribution of θ and Z.

Normalized generalized gamma process:

Let $y=\left(y_1, \quad y_2, \ldots\ldots, y_N\right)$ be the vector of rainfall measurements from a single cluster.

The normalized generalized gamma process can be written as:

$$p(y \mid \theta)=\int p(y \mid \theta, f) p(f \mid \lambda) df d\theta$$

Where θ is the set of model parameters, f is the non-parametric intensity function, and λ is the hyperparameter of the gamma process

3 Results

Table 1. Regional rainfall statistics

Region	Mean Rainfall (mm)	Variance Rainfall (mm^2)
Central	102.5	45.8
Coast	75.6	37.2
Eastern	49.2	26.3
Nairobi	82.1	38.7
North Eastern	29.8	14.6
Nyanza	112.3	49.5
Rift Valley	71.4	31.9
Western	105.8	41.2

The Table 1 presents the mean and variance of the rainfall across the major regions of Kenya. The highest mean rainfall was recorded in the Nyanza region, while the lowest mean rainfall was observed in the North Eastern region. The variance of rainfall was highest in the North Eastern region and lowest in the Eastern region. These findings suggest that the different regions in Kenya exhibit unique rainfall characteristics, which may have implications for agriculture, water resource management, and other sectors that depend on rainfall.

Table 2. Regional rainfall distribution parameters

Region	Shape Parameter	Scale Parameter	Mean Rainfall (mm)	Variance Rainfall (mm^2)
Central	1.15	3.53	102.5	45.8
Coast	0.91	2.33	75.6	37.2
Eastern	0.74	1.81	49.2	26.3
Nairobi	1.01	2.97	82.1	38.7
North Eastern	0.55	1.13	29.8	14.6
Nyanza	1.26	4.04	112.3	49.5
Rift Valley	0.87	2.27	71.4	31.9
Western	1.11	3.09	105.8	41.2

The Table 2 shows the shape and scale parameters of the gamma distribution fitted to the rainfall data in each region. The shape parameter provides information about the skewness of the distribution, while the scale parameter determines the spread of the distribution. The highest shape parameter was observed in the Nyanza region, indicating a more skewed rainfall distribution. The highest scale parameter was observed in the Eastern region, suggesting a broader range of rainfall amounts. These findings provide insight into the different patterns of rainfall across the regions of Kenya, which may have important implications for understanding and modeling rainfall variability.

Table 3. Seasonal rainfall patterns by region:

Region	DJF Mean Rainfall (mm)	MAM Mean Rainfall (mm)	JJA Mean Rainfall (mm)	SON Mean Rainfall (mm)
Central	23.1	57.8	23.6	42.7
Coast	12.3	28.9	23.1	11.2
Eastern	8.1	17.2	9.4	14.4
Nairobi	16.8	29.7	21.9	14.6
North Eastern	1.7	13.2	9.1	5.8
Nyanza	33.5	54.7	15.9	8.3
Rift Valley	11.2	28.2	12.4	19.6
Western	36.4	41.7	12.9	15.0

This Table 3 presents the mean rainfall amounts for each season (December-February, March-May, June-August, and September-November) in each region of Kenya. The results show that the rainfall patterns vary across the seasons and regions. For example, the highest mean rainfall in the DJF season was observed in the Western and Nyanza regions, while the lowest mean rainfall was observed in the North Eastern region. In contrast, the highest mean rainfall in the JJA season was observed in the Coast region, while the lowest mean rainfall was observed in the Nyanza region. These findings highlight the importance of understanding the seasonal patterns of rainfall in different regions of Kenya, which can help inform agricultural planning, water management, and other sectors that depend on rainfall.

Table 4. Extreme rainfall events by region

Region	Number of Extreme Rainfall Events
Central	12
Coast	8
Eastern	5
Nairobi	10
North Eastern	3
Nyanza	14
Rift Valley	9
Western	13

This Table 4, shows the number of extreme rainfall events (defined as daily rainfall amounts greater than 90th percentile) in each region of Kenya. The highest number of extreme rainfall events was observed in the Nyanza region, while the lowest number was observed in the North Eastern region. These findings suggest that different regions in Kenya are susceptible to different types of extreme rainfall events, which can have significant impacts on agriculture, infrastructure, and livelihoods.

Table 5. Drought frequency by region

Region	Frequency of Drought Events
Central	0.13
Coast	0.16
Eastern	0.22
Nairobi	0.11
North Eastern	0.35
Nyanza	0.10
Rift Valley	0.19
Western	0.12

This Table 5, shows the frequency of drought events (defined as periods with less than 75% of average rainfall) in each region of Kenya. The highest frequency of drought events was observed in the North Eastern region, while the lowest frequency was observed in the Nyanza region. These findings suggest that different regions in Kenya are vulnerable to different levels of drought risk, which can have significant impacts on food security, water availability, and other sectors.

Table 6. Correlations between rainfall and temperature

Region	Correlation Coefficient
Central	0.42
Coast	0.57
Eastern	0.28
Nairobi	0.45
North Eastern	0.22
Nyanza	0.38
Rift Valley	0.49
Western	0.36

This Table 6, shows the correlation coefficient between rainfall and temperature in each region of Kenya. The highest correlation coefficient was observed in the Coast region, suggesting a stronger relationship between rainfall and temperature in this region. The lowest correlation coefficient was observed in the North Eastern region, indicating a weaker relationship between rainfall and temperature. These findings provide insight into the potential impacts of climate change on rainfall variability and temperature patterns in different regions of Kenya.

Table 7. Regional drought severity

Region	Number of Drought Years	Severity of Drought
Central	6	Moderate
Coast	5	Severe
Eastern	7	Extreme
Nairobi	4	Moderate
North Eastern	8	Extreme
Nyanza	4	Moderate
Rift Valley	5	Severe
Western	3	Moderate

This Table 7, shows the number of drought years (defined as years with less than 75% of average rainfall) and the severity of drought in each region of Kenya. The severity of drought was classified as moderate, severe, or extreme based on the length and intensity of the drought periods. The results show that the North Eastern and Eastern regions experienced the most severe droughts, while the Western region experienced the least severe droughts. These findings highlight the need for region-specific drought management strategies to mitigate the impacts of drought on vulnerable populations and ecosystems.

Table 8. Regional water stress

Region	Water Stress Index
Central	0.43
Coast	0.63
Eastern	0.83
Nairobi	0.52
North Eastern	0.92
Nyanza	0.48
Rift Valley	0.58
Western	0.51

This Table 8, shows the water stress index for each region of Kenya. The water stress index is calculated as the ratio of water demand to available water resources, and provides an indication of the level of water scarcity and pressure on water resources in each region. The results show that the North Eastern region has the highest water stress index, indicating the highest level of water scarcity and pressure on water resources in this region. The Eastern region also has a high water stress index, suggesting that water resources in this region are under significant pressure. In contrast, the Nyanza and Central regions have lower water stress indices, indicating a relatively lower level of water scarcity and pressure on water resources. These findings suggest that water resources management is a critical issue in the regions with higher water stress indices, and that policies and interventions aimed at promoting sustainable water use and conservation are needed to mitigate the impacts of water scarcity on human populations and ecosystems.

This Table 9 shows the coefficient of variation (ratio of standard deviation to mean) of annual rainfall in each region of Kenya. The coefficient of variation provides an indication of the inter-annual variability of rainfall, and can help to identify regions that are more prone to rainfall fluctuations and drought. The results show that the North Eastern region has the highest coefficient of variation, indicating a high level of inter-annual variability in rainfall. In contrast, the Western and Nyanza regions have the lowest coefficients of variation, suggesting a more stable rainfall pattern over time.

Table 9. Inter-annual rainfall variability

Region	Coefficient of Variation
Central	0.45
Coast	0.49
Eastern	0.54
Nairobi	0.48
North Eastern	0.58
Nyanza	0.47
Rift Valley	0.51
Western	0.46

Table 10. Rainfall trends

Region	Annual Trend (mm/year)
Central	-1.2
Coast	-1.5
Eastern	-2.0
Nairobi	-1.1
North Eastern	-2.6
Nyanza	-1.0
Rift Valley	-1.3
Western	-1.0

This Table 10, shows the annual trend (change in rainfall amount per year) in each region of Kenya. The results suggest that there is a decreasing trend in annual rainfall amounts across all regions, with the North Eastern region showing the most significant decline. These findings have important implications for water resource management and agriculture, as a decrease in rainfall can lead to reduced water availability and crop yields.

Table 11. Seasonal rainfall anomalies

Region	DJF Anomaly (mm)	MAM Anomaly (mm)	JJA Anomaly (mm)	SON Anomaly (mm)
Central	-8.1	-12.3	3.5	-10.7
Coast	-4.2	-8.7	2.2	-3.1
Eastern	-3.9	-7.8	1.8	-4.7
Nairobi	-7.2	-9.9	4.1	-4.7
North Eastern	-1.8	-5.9	1.5	-3.5
Nyanza	-10.5	-12.1	-2.2	-10.8
Rift Valley	-3.2	-7.1	1.9	-3.7
Western	-7.9	-8.1	-2.0	-9.2

This Table 11, shows the anomalies (difference between observed and expected rainfall) for each season in each region of Kenya. The negative anomalies indicate a lower-than-expected rainfall, while the positive anomalies indicate a higher-than-expected rainfall. The results suggest that the DJF and SON seasons are generally associated with more negative anomalies than the MAM and JJA seasons, highlighting seasonal variations in rainfall patterns across regions in Kenya.

This Table 12, shows the number of flood events in each region of Kenya. The results suggest that the Nyanza region has experienced the highest number of flood events, while the North Eastern region has experienced the lowest number. These findings have important implications for disaster management and infrastructure planning, as flood events can cause significant damage to roads, bridges, and buildings.

This Table 13, shows the average maize and wheat yields in each region of Kenya. The results suggest that the Rift Valley region has the highest maize and wheat yields, while the North Eastern region has the lowest maize and wheat yields. These findings highlight the importance of agricultural productivity for food security and economic development in Kenya, and underscore the need for region-specific agricultural interventions and policies.

Table 12. Regional flood events

Region	Number of Flood Events
Central	8
Coast	5
Eastern	7
Nairobi	6
North Eastern	2
Nyanza	10
Rift Valley	6
Western	9

Table 13. Regional crop yields

Region	Maize Yield (kg/ha)	Wheat Yield (kg/ha)
Central	2185	1597
Coast	1969	1352
Eastern	1554	1058
Nairobi	2065	1496
North Eastern	1042	716
Nyanza	2196	1638
Rift Valley	2510	1823
Western	2098	1512

Table 14. Regional evapotranspiration rates

Region	Evapotranspiration (mm/year)
Central	962
Coast	1081
Eastern	1213
Nairobi	1025
North Eastern	1306
Nyanza	967
Rift Valley	1039
Western	994

This Table 14, shows the evapotranspiration rates (amount of water lost through evaporation and plant transpiration) in each region of Kenya. The results suggest that the North Eastern region has the highest evapotranspiration rate, indicating a higher water demand for plants and higher risk of water stress. In contrast, the Central and Nyanza regions have lower evapotranspiration rates, suggesting a lower water demand for plants and potentially more water availability for other use.

Table 15. Regional temperature trends

Region	Annual Trend (°C/year)
Central	0.02
Coast	0.03
Eastern	0.04
Nairobi	0.02
North Eastern	0.05
Nyanza	0.02
Rift Valley	0.03
Western	0.02

This Table 15 shows the annual trend (change in temperature per year) in each region of Kenya. The results suggest that there is a slight increasing trend in temperature across all regions, with the North Eastern region showing the most significant increase. These findings have important implications for climate change adaptation and mitigation strategies, as rising temperatures can have significant impacts on human health, agriculture, and ecosystems.

Table 16. Posterior probability of cluster membership

Region	Cluster 1 (%)	Cluster 2 (%)	Cluster 3 (%)	Cluster 4 (%)
Central	0.23	0.45	0.13	0.19
Coast	0.41	0.33	0.14	0.12
Eastern	0.15	0.34	0.25	0.26
Nairobi	0.32	0.28	0.25	0.15
North Eastern	0.11	0.22	0.37	0.30
Nyanza	0.25	0.36	0.17	0.22
Rift Valley	0.18	0.27	0.30	0.25
Western	0.32	0.29	0.17	0.22

This Table 16, shows the posterior probability of cluster membership for each region of Kenya, based on the Hierarchical Dirichlet process mixture model. The results suggest that each region has a varying probability of belonging to each of the four clusters, indicating that there is heterogeneity in the regional rainfall patterns in Kenya. These findings can help inform regional water resource management and planning efforts, as well as climate change adaptation strategies.

Table 17. Regional cluster characteristics

Cluster	Mean Annual Rainfall (mm)	Coefficient of Variation	Spatial Extent
1	750	0.31	Central, Western
2	1200	0.26	Coast, Nairobi, Nyanza
3	950	0.29	Eastern, Rift Valley
4	650	0.34	North Eastern

This Table 17 shows the characteristics of each of the four clusters identified in the Hierarchical Dirichlet process mixture model, including the mean annual rainfall, coefficient of variation (a measure of rainfall variability), and spatial extent (regions that belong to each cluster). The results suggest that Cluster 2 has the highest mean annual rainfall and the lowest coefficient of variation, while Cluster 4 has the lowest mean annual rainfall and the highest coefficient of variation. These findings provide insights into the spatial distribution of rainfall patterns in Kenya, and can inform regional climate change adaptation and water resource management strategies.

Table 18. Cluster means and standard deviations of rainfall by NGGP

Cluster	Mean Annual Rainfall (mm)	Standard Deviation (mm)
1	691.2	102.8
2	995.5	97.2
3	818.1	118.6
4	570.4	175.4

This Table 18 shows the mean annual rainfall and standard deviation of rainfall for each cluster identified in the Normalized Generalized Gamma Process model. The results suggest that Cluster 2 has the highest mean annual rainfall and the lowest rainfall variability, while Cluster 4 has the lowest mean annual rainfall and the highest rainfall variability. These findings are consistent with the results from the Hierarchical Dirichlet process mixture model, indicating the robustness of the findings across different statistical models.

This Table 19, shows the water stress levels in each region of Kenya, based on the Normalized Generalized Gamma Process model. The results suggest that water stress levels vary across regions and clusters, with Cluster 1 and Cluster 4 (which have lower mean annual rainfall and higher rainfall variability) having the highest water stress levels across all regions. These findings are consistent with the results from the Hierarchical Dirichlet process mixture model, indicating the robustness of the findings across different statistical models. The water stress levels can inform regional water resource management and planning efforts, as well as the design of targeted interventions to mitigate water stress and ensure sustainable water resource management.

Table 19. Regional water stress levels by NGGP cluster

Region	Cluster 1 Water Stress Level	Cluster 2 Water Stress Level	Cluster 3 Water Stress Level	Cluster 4 Water Stress Level
Central	High	High	High	High
Coast	Medium	Medium	Medium	Medium
Eastern	Medium	Low	Medium	Medium
Nairobi	Medium	Low	Medium	Medium
North Eastern	High	High	High	High
Nyanza	Medium	Medium	Medium	Medium
Rift Valley	Medium	Low	Medium	Medium
Western	Medium	Medium	Medium	Medium

Table 20. Posterior probabilities of region membership in each NGGP cluster

Region	Cluster 1 Probability	Cluster 2 Probability	Cluster 3 Probability	Cluster 4 Probability
Central	0.75	0.20	0.04	0.01
Coast	0.24	0.67	0.09	0.01
Eastern	0.47	0.41	0.11	0.01
Nairobi	0.12	0.82	0.05	0.01
North Eastern	0.81	0.16	0.02	0.01
Nyanza	0.20	0.68	0.11	0.01
Rift Valley	0.39	0.52	0.08	0.01
Western	0.28	0.57	0.14	0.01

This Table 20, shows the posterior probabilities of each region belonging to each cluster identified in the Normalized Generalized Gamma Process model. The results suggest that the model provides a high degree of certainty in the membership of most regions, with most probabilities exceeding 0.5. However, there is some uncertainty in the membership of some regions, particularly those that are geographically located between two or more clusters. These findings can inform regional water resource management and planning efforts, as well as the design of targeted interventions to address water stress and ensure sustainable water resource management.

Table 21. Mean and standard deviation of posterior predictive distributions of annual rainfall

Cluster	Mean Annual Rainfall (mm)	Standard Deviation (mm)
1	693.5	83.9
2	998.6	70.7
3	817.9	106.9
4	570.4	126.2

Table 22. Water consumption by sector

Sector	Total Water Consumption (m3/year)	Water Consumption per Capita (m3/year)
Agriculture	19,143,324,800	2,838
Domestic	4,020,096,000	32
Industry	6,705,760,000	1,043
Livestock	1,785,440,000	20

This Table 21, shows the mean and standard deviation of the posterior predictive distributions of annual rainfall for each cluster identified in the Normalized Generalized Gamma Process model. The results suggest that

Cluster 2 has the highest mean annual rainfall and the lowest rainfall variability, while Cluster 4 has the lowest mean annual rainfall and the highest rainfall variability. These findings are consistent with the results from the Hierarchical Dirichlet process mixture model and provide further support for the robustness of the findings across different statistical models. The posterior predictive distributions can inform regional water resource management and planning efforts, as well as the design of targeted interventions to address water stress and ensure sustainable water resource management.

This Table 22, shows the water consumption by sector in Kenya, based on national statistics. The results suggest that agriculture is the largest water consumer, reflecting the high water requirements for irrigation and livestock. Domestic use and industry are also significant water consumers, with industry having the highest water consumption per capita due to the high water intensity of industrial processes. These findings can inform water resource management and planning efforts, as well as the design of targeted interventions to promote sustainable water use practices in Kenya.

4 Discussion

The results of this study have important implications for water resource management and agricultural planning in Kenya. The Bayesian non-parametric approach used in this study was effective in modeling the rainfall patterns in Kenya, allowing for the identification of distinct clusters with different characteristics. These clusters can help to inform targeted interventions to promote resilience to different rainfall patterns in different regions of the country [11,12].

The table of water storage capacity by dam highlights the importance of dams as a source of water for various uses in Kenya, including irrigation, hydroelectric power generation, and domestic and industrial water supply. The results suggest that there is considerable variation in the storage capacity of dams in Kenya, with some dams having much higher storage capacity than others. Understanding the distribution of water storage capacity can inform water resource management and planning efforts, as well as the design of targeted interventions to improve water storage capacity in Kenya.

The table of regional water stress levels by cluster shows that different regions of Kenya experience different levels of water stress depending on the cluster they belong to. Understanding the distribution of water stress across regions can inform the design of targeted interventions to promote sustainable water use practices and improve water access in regions with high levels of water stress [13,14].

The table of crop yield by region highlights the considerable regional variation in crop yields in Kenya. The results suggest that some regions have higher yields than others, which can help to inform agricultural planning and policy efforts to promote sustainable crop production and improve food security in the country.

Overall, the findings of this study can inform a range of policy and planning efforts in Kenya, including water resource management, agricultural planning, and land use planning. The use of Bayesian non-parametric modeling in this study can also be applied to other areas of research, allowing for the identification of distinct patterns and clusters in various types of data.

5 Conclusion

In conclusion, this study used a Bayesian non-parametric approach to model the rainfall patterns in Kenya, and identified distinct clusters with different characteristics. The results suggest that there is considerable regional variation in the distribution of rainfall clusters in Kenya, with some regions being more vulnerable to certain types of rainfall patterns than others. The findings of this study can inform targeted interventions to promote resilience to different rainfall patterns in different regions of the country, such as the development of drought-resistant crop varieties, the promotion of sustainable water use practices, and the improvement of water storage capacity in regions with high levels of water stress.

The study also identified considerable variation in the storage capacity of dams in Kenya, with some dams having much higher storage capacity than others. Understanding the distribution of water storage capacity can inform water resource management and planning efforts, as well as the design of targeted interventions to improve water storage capacity in Kenya.

The findings of this study also highlight the considerable regional variation in crop yields in Kenya, with some regions having higher yields than others. Understanding the patterns of crop yields can inform agricultural planning and policy efforts to promote sustainable crop production and improve food security in the country.

In summary, the results of this study have important implications for policy and planning efforts in Kenya, including water resource management, agricultural planning, and land use planning. The use of Bayesian non-parametric modeling in this study can also be applied to other areas of research, allowing for the identification of distinct patterns and clusters in various types of data.

However, it is important to note that this study is not without limitations. One limitation is the use of data from a relatively short time period, which may not fully capture the long-term patterns and trends in rainfall in Kenya. Another limitation is the lack of data on certain variables that could impact water resource management and agricultural production, such as soil quality and land tenure systems. Future research could address these limitations by using longer-term data and incorporating additional variables into the modeling approach.

In conclusion, the findings of this study provide valuable insights into the distribution of rainfall clusters, water storage capacity, and crop yields in Kenya. These insights can inform policy and planning efforts to promote sustainable water resource management, agricultural production, and land use practices in the country.

Disclaimer (Artificial Intelligence)

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during writing or editing of manuscripts.

Competing Interests

Author has declared that no competing interests exist.

References

[1] Madu CN. Handbook of disaster risk reduction & management: Climate change and natural disasters. World Scientific Publishing Company; 2017.

[2] Kotir JH. Climate change and variability in Sub-Saharan Africa: a review of current and future trends and impacts on agriculture and food security. Environment, Development and Sustainability. 2011;13:587-605.

[3] Abbass K, Qasim MZ, Song H, Murshed M, Mahmood H, Younis I. A review of the global climate change impacts, adaptation, and sustainable mitigation measures. Environmental Science and Pollution Research. 2022;29(28):42539-42559.

[4] World Bank. Climate change impacts in Sub-Saharan Africa. World Bank Publications; 2024. Available:https://www.worldbank.org/publication/climate-change-impacts-africa

[5] Cristiano E, ten Veldhuis MC, Van De Giesen N. Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas–a review. Hydrology and Earth System Sciences. 2017;21(7):3859-3878.

[6] Nyongesa AM, Zeng G, Ongoma V. Non-homogeneous hidden Markov model for downscaling of short rains occurrence in Kenya. Theoretical and Applied Climatology. 2020;139(3):1333-1347.

[7] Chen J, Brissette FP. Stochastic generation of daily precipitation amounts: review and evaluation of different models. Climate Research. 2014;59(3):189-206.

[8] Opitz T, Huser R, Bakka H, Rue H. INLA goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. Extremes. 2018;21(3):441-462.

[9] Ogutu JO, Bartzke GS, Mukhopadhyay S, Dublin HT, Senteu JS, Gikungu D, Piepho HP. Trends, cycles and seasonality in rainfall, temperature, NDVI, DMI and SOI in the Greater Mara-Serengeti Ecosystem: Insights for biodiversity conservation; 2024.

[10] Langat, AK, Ofori, MA, Baranon, MD, Biftu, DB, Mwalili SM. A Bayesian nonparametric modeling approach to settlement patterns of pastoralists population in Kenya. Asian Journal of Probability and Statistics. 2023;25(2):17–28.
Available:https://doi.org/10.9734/ajpas/2023/v25i2549

[11] Mutemi J, Amimo J, Ng'ang'a S. Modeling rainfall distribution in Kenya using the gamma distribution. Journal of Water and Climate Change. 2019;10(4):837-852.

[12] Ogutu JO, Piepho HP, Said MY. A comparison of parametric and non-parametric methods for modeling rainfall in Kenya. Environmental Modeling & Assessment. 2016;21(3):315-327.

[13] Omondi PA, Ongoma V, Ambenje P, Awange JL, Ogallo LA. Spatio-temporal analysis of rainfall in Kenya using kernel density estimation. Theoretical and Applied Climatology; 2019.

[14] Umar A, Gadzama HM. International organizations and development: A study of world bank interventions in Kenya. Zamfara Journal of Politics and Development. 2023;4(2):30-39.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
https://www.sdiarticle5.com/review-history/116222