Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Bayesian Gaussian Process Modeling of Nonstationary Climate Extremes with Unknown Physical Drivers: A GP-GEV Framework for West African Rainfall

Silas Alebikenaba Ayariga

silasalebikenaba@gmail.com

Kwame Nkrumah University of Science and Technology https://orcid.org/0009-0005-3217-3373

James Azuure Atambire

University of North Carolina at Chapel Hill

Nana Kena Frempong

Kwame Nkrumah University of Science and Technology

Research Article

Keywords: Extreme Value Theory, Bayesian Inference, Gaussian Processes, Generalized Extreme Value Distribution, Nonstationarity, Rainfall Extremes, West African Sahel, Climate Risk, Return Levels, Hierarchical Modeling

Posted Date: July 31st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10529823/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Bayesian Gaussian Process Modeling of Nonstationary Climate Extremes with Unknown Physical Drivers: A GP-GEV Framework for West African Rainfall

Silas Alebikenaba Ayariga[a,b,*], **Frempong Nana Kena**[a], **James Azuure Atambire**[c]

a Department of Statistics and Actuarial Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

b African Institute for Mathematical Sciences, Senegal

c University of North Carolina

* Corresponding author: silasalebikenaba@gmail.com

Abstract

Extreme rainfall in the West African Sahel has intensified in recent decades, rendering conventional stationary extreme value models inadequate. Existing nonstationary approaches rely on predefined climate covariates whose relationships with Sahel rainfall are uncertain and unstable. This study develops a Bayesian Gaussian Process–Generalized Extreme Value (GP–GEV) framework that models temporal evolution of GEV parameters as latent Gaussian processes, learning nonstationarity directly from data without requiring physical covariates. The framework is validated through simulation experiments across eight data-generating regimes and applied to annual maximum daily rainfall from ERA5 dataset. Results show that while stationary GEV models remain competitive under stationarity, the GP–GEV framework substantially outperforms parametric-trend alternatives when latent nonlinear variability is present, achieving superior predictive accuracy, improved return level estimation, and better calibrated uncertainty intervals for high return periods. Empirical analyses reveal strong spatial heterogeneity, with several regions showing pronounced increases in 50-year return levels relative to stationary benchmarks. These findings demonstrate that latent-process Bayesian models offer a flexible, statistically robust alternative to covariate-based methods in regions with uncertain climatic drivers, underscoring the importance of uncertainty-aware frameworks for resilient infrastructure design and climate adaptation policy in the Sahel.

Keywords: Extreme Value Theory; Bayesian Inference; Gaussian Processes; Generalized Extreme Value Distribution; Nonstationarity; Rainfall Extremes; West African Sahel; Climate Risk; Return Levels; Hierarchical Modeling

1. Introduction

The West African Sahel, spanning approximately 10°N to 20°N from Senegal to Sudan, is among the most climatically vulnerable regions on Earth. Rainfall totals decrease sharply from over 800 mm yr^{-1} in the south to below 200 mm yr^{-1} in the north, supporting rain-fed agriculture and pastoralism for hundreds of millions of people [1]. The region experienced a catastrophic multi-decadal drought from the late 1960s through the 1980s, followed by a partial rainfall recovery whose characteristics diverge markedly from pre-drought conditions: Salas et al. [21] demonstrate using 72 gauge records that the 99th percentile of daily rainfall has increased significantly in recent decades, while the frequency of storms has also risen. Taylor et al. [24] reported using satellite observations that the frequency of extreme Sahelian convective storms tripled between 1982 and 2013.

Looking forward, the picture is alarming. Guilbert et al. [11] studied 42 CMIP6 models and reported that Sahel rainfall interannual variability increases robustly under greenhouse warming, with a projected 140% increase in the frequency of extreme wet seasons and a 60% increase in extreme dry seasons relative to the 20th-century baseline. This simultaneous intensification of wet and dry extremes, driven by amplified ENSO teleconnections and increased mean monsoon moisture, poses an unprecedented challenge for water resource management, flood protection, and agricultural planning.

Reliable statistical quantification of extreme event risk is therefore essential for climate adaptation policy. Yet existing methods face a fundamental tension: standard extreme value models assume stationarity, while nonstationary extensions require the analyst to specify physical covariates that drive the nonstationarity; factors that are often unknown, partially observed, or subject to complex multi-scale interactions in West African hydroclimatology [1, 9]. This tension is especially acute in the Sahel, where at least four competing driver systems such as Atlantic SST variability [10], ENSO teleconnections whose sign reversed after the 1970s, Mediterranean SST changes [11], and anthropogenic aerosol forcing [2] have all been implicated in the multi-decadal rainfall variability. Existing nonstationary frequency analyses for West African rainfall either retain the stationarity assumption and consequently miss real distributional trends in the observational record [16], or specify parametric covariate models that require the analyst to identify and correctly formulate driver relationships a priori. This requirement is fundamentally problematic when the drivers are contested, operate on multiple interacting timescales, and are partially nonstationary in their own right [9, 20]. The Gaussian process prior on GEV parameters has been productively applied in the spatial domain to pool information across monitoring stations [7, 3], but its use for modeling temporal nonstationarity with latent climatic drivers, with rigorous inference for time-varying return levels, has not been pursued for West African rainfall data. This study addresses that gap. We develop a Bayesian Gaussian Process Generalized Extreme Value (GP-GEV) framework in which the GEV location and scale parameters evolve as latent stochastic processes governed by Gaussian process priors. This data-driven approach learns the structure of nonstationarity directly from the observational record without requiring explicit specification of physical drivers, while providing fully calibrated posterior uncertainty quantification for return levels, a critical requirement for infrastructure design and disaster risk assessment. The framework is applied to ERA5 annual maximum daily rainfall across the West African Sahel domain, and scalable inference is achieved via the nested Laplace approximation of [3], enabling regional analysis across hundreds of grid points simultaneously.

The study makes four principal contributions to the literature on nonstationary extreme value analysis. Firstly, we introduce a fully Bayesian hierarchical model in which GEV parameters evolve as latent Gaussian processes of time, requiring no specification of physical driver covariates. This constitutes a methodologically novel extension of GP-based extreme value analysis from the spatial to the temporal domain, with principled inference for time-varying return levels and their uncertainty. Secondly, through Monte Carlo simulation under eight data-generating processes spanning stationary, linearly trending, smoothly nonlinear, multi-scale, covariate-driven, and parameter-specific nonstationary regimes, we establish that the GP-GEV models achieve near-nominal 95% interval coverage across all scenarios, while stationary and parametric nonstationary alternatives exhibit coverage deficits of up to 35 percentage points under latent-process nonstationarity. Thirdly, we demonstrate that extending GP priors jointly to the location and scale parameters of the GEV distribution provides meaningful improvements in return level accuracy and predictive sharpness specifically under heteroscedastic extremes, while incurring no meaningful performance loss when nonstationarity is confined to the location parameter. Finally, through

application to ERA5 data at twelve representative Sahel locations, we provide evidence that stationary frequency analysis systematically underestimates 50-year return levels by 7–8 mm relative to the GP-GEV model, a bias with direct and quantifiable consequences for flood protection infrastructure design in one of the world's most climate-exposed regions.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature on extreme value theory, nonstationarity, Gaussian process methods, and West African rainfall variability. Section 2.2 develops the GP-GEV framework, including the hierarchical model specification, covariance kernel selection, prior distributions, and scalable inference procedure. Section 2.6 describes the simulation study design. Section 3 presents the simulation results and empirical application to ERA5 rainfall data. Section 5 discusses the findings in the context of the existing literature, and Section 6 concludes with methodological recommendations and directions for future research.

2. Methods

2.1. Baseline Stationary GEV Model

The stationary GEV distribution serves as the foundational building block and the primary benchmark against which nonstationary extensions are assessed. Its inclusion is not merely conventional; the stationarity debate reviewed in Chapter 2, most acutely expressed by [17] and [22], demands that nonstationary models be held to a higher evidentiary standard and must demonstrably outperform the simpler stationary alternative before their additional complexity can be considered justified.

2.1.1. Model Specification

Let Yt denote the annual maximum at a given grid point in year t. Under the stationary model, the Yt are treated as independent and identically distributed realisations from a GEV distribution:

$$Yt \sim iid\ GEV(\mu, \sigma, \xi), \quad t = 1, \ldots, T \quad (2.1)$$

where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter, and $\xi \in \mathbb{R}$ is the shape parameter. The cumulative distribution function is

$$F(y; \mu, \sigma, \xi) = \exp\{ -[1 + \xi(y-\mu)/\sigma]^{-1/c} \}, \quad 1 + \xi(y-\mu)/\sigma > 0 \quad (2.2)$$

with the Gumbel case ($\xi = 0$) understood as the limiting form as $\xi \to 0$. The corresponding probability density function is

$$f(y; \mu, \sigma, \xi) = (1/\sigma)[1 + \xi(y-\mu)/\sigma]^{-1+1/c} \exp\{ -[1 + \xi(y-\mu)/\sigma]^{-1/c} \} \quad (2.3)$$

2.1.2. Likelihood and Bayesian Posterior

The complete-data log-likelihood for $\theta = (\mu, \sigma, \xi)^T$ given observations y is a standard GEV log-likelihood (Eq. 2.4), subject to the constraint $1 + \xi(yt - \mu)/\sigma > 0$ for all t. In the Bayesian treatment, the posterior distribution takes the form

$$p(\theta \mid y) \propto \exp\{\ell(\theta; y)\}\, p(\theta) \quad (2.5)$$

where p(θ) denotes the prior distribution on θ, the full specification of which is given in Section 2.4.3.

2.1.3. Return Levels Under Stationarity

Under stationarity, the T-year return level rT, that is, the quantile exceeded with probability 1/T in any given year, is the unique solution to $F(rT; \mu, \sigma, \xi) = 1 - 1/T$, which yields

$$rT = \mu - (\sigma/\xi)[1 - (-\log(1-1/T))^{-\xi}], \quad \xi \neq 0 \quad (2.6)$$

In the Bayesian framework, the posterior distribution of rT is obtained by propagating the posterior samples of θ through Equation (2.6), yielding a credible interval that reflects the totality of parameter estimation uncertainty. The 50-year and 100-year return levels, corresponding to T = 50 and T = 100 respectively, constitute the primary quantities of practical interest for the purposes of infrastructure design.

2.2. Competing (Benchmark) Model Families

Table 1 summarizes the principal GEV model families and their relative strengths and limitations in the Sahel context, motivating the proposed framework.

Model type	Nonstationarity formulation	Strengths	Limitations in Sahel context
Stationary GEV [5]	μ, σ, ξ constant	Parsimonious; well-understood asymptotic theory	Ignores documented multi-decadal nonstationarity; biased return levels
Linear-trend GEV [14]	$\mu t = \mu 0 + \beta t$	Simple; captures monotonic warming trends	Imposes linearity; misses drought-recovery oscillation
Covariate-based GEV [8]	$\mu t = \mu 0 + \beta Xt$	Physically interpretable	Requires correct covariate specification; ENSO sign-reversals invalidate fixed-coefficient models
Spline GEV [4]	$\mu t = f(t)$ cubic spline	Nonparametric flexibility	No full Bayesian inference; smoothness uncertainty not propagated
Spatial GP-GEV [7, 3]	$\mu(s) \sim$ GP over space	Pools strength across stations	Models spatial, not temporal, variation
Proposed temporal GP-GEV	$\mu t \sim GP(m(t), k)$; $\log \sigma t \sim$ GP	Data-driven; fully Bayesian UQ; covariate-free	Computationally more demanding than frequentist alternatives

Table 1: Comparative summary of GEV model families for extreme rainfall analysis, evaluated with reference to the West African Sahel context.

Five benchmark models are specified to represent the state of the art in classical nonstationary frequency analysis:

- NS-Lin: linear trend $\mu t = \beta 0 + \beta 1 \tilde{t}$, constant σ, ξ [14].
- NS-Lin2: quadratic trend $\mu t = \beta 0 + \beta 1 \tilde{t} + \beta 2 \tilde{t}^2$, constant σ, ξ.
- NS-Cov: covariate-based $\mu t = \beta 0 + \beta^T xt$, where candidates include the AMO index, Niño-3.4 ENSO index, and global mean surface temperature [8, 18].
- NS-Spline: smoothing-spline $\mu t = f(t)$, $f \in S3$ [4].

All benchmark models carry the fundamental limitation identified in literature: they require either correct functional form specification or reliable covariate identification a priori, requirements that are acutely problematic in the West African Sahel.

2.3. Data Description

One complementary gridded dataset (the ERA5 reanalysis product [12]) is employed, providing daily total precipitation at 0.25° × 0.25° horizontal resolution for the period 1940–2022 (T = 83 years). ERA5 combines a modern data assimilation system with historical observations to produce a physically consistent, methodologically homogeneous record, a critical property for the detection of long-term nonstationarity. The analysis domain covers 10°–20°N and 18°W–40°E.

The response variable yt is the annual maximum of daily precipitation accumulation during the JJAS monsoon season (June–September):

$$yt = max^{d}\in^{D}_{t} Pt,d \quad (2.7)$$

where Dt denotes the days of the JJAS season in year t. Time indices are standardized as $\tilde{t} = (t - \bar{t})/st$ so that $\tilde{t} \in [-1, 1]$, which is numerically important for GP inference.

2.4. The Proposed Bayesian GP-GEV Model

The GP-GEV framework is specified as a three-level hierarchical model (observation level, process level, and hyperparameter level; Eq. 2.8), where ϕ collects all GP hyperparameters and the GEV shape parameter ξ. Two model variants are defined:

- GP-GEV-μ: only the location parameter evolves as a GP; σt = σ for all t.
- GP-GEV-μσ: both μt and log σt evolve as independent GPs with potentially distinct covariance kernels.

Conditional on the latent processes, the annual maxima are modeled as independent GEV random variables:

$$yt \mid \mu t, \sigma t, \xi \sim ind.\ GEV(\mu t, \sigma t, \xi),\ t = 1, \ldots, T \quad (2.9)$$

The shape parameter ξ is treated as a time-invariant global constant, reflecting the theoretical expectation that tail behavior should vary on longer timescales than the observational record permits to distinguish, and the empirical finding that nonstationarity is most concentrated in the location parameter [16].

2.4.1. Gaussian Process Priors

The latent location function μ(·) is assigned the prior

$$\mu(\cdot) \sim GP(m\mu, k\mu(\cdot,\cdot; \sigma^2 f,\mu, \ell\mu)) \quad (2.10)$$

with mean function mμ(t) = μ0 and covariance kernel kμ, yielding the finite-dimensional prior μ ~ N(μ0·1, Kμ). In the GP-GEV-μσ model, the log-scale function is assigned an independent GP prior

$$\log \sigma(\cdot) \sim GP(m\sigma, k\sigma(\cdot,\cdot; \sigma^2 f,\sigma, \ell\sigma)) \quad (2.11)$$

with mean function $m\sigma(t) = \log \sigma 0$. The log parameterization ensures $\sigma t > 0$ without constrained optimization.

2.4.2. Covariance Kernels

Three kernels are considered. The Squared Exponential kernel,

$$kSE(t, t'; \sigma^2 f, \ell) = \sigma^2 f \exp\{ -(t-t')^2 / 2\ell^2 \} \quad (2.12)$$

produces infinitely differentiable sample paths. The Matérn-3/2 kernel,

$$kM32(t, t'; \sigma^2 f, \ell) = \sigma^2 f (1 + \sqrt{3}|t-t'|/\ell) \exp(-\sqrt{3}|t-t'|/\ell) \quad (2.13)$$

produces once-differentiable paths more realistic for physical climate processes and is the primary kernel in all analyses. The Rational Quadratic kernel,

$$kRQ(t, t'; \sigma^2 f, \ell, \alpha) = \sigma^2 f (1 + (t-t')^2/2\alpha\ell^2)^{-a} \quad (2.14)$$

is equivalent to an infinite mixture of SE kernels and captures nonstationarity operating simultaneously on multiple timescales — precisely the situation in the West African Sahel where ENSO, AMO, and greenhouse trends are superposed.

2.4.3. Prior Specification

The penalized complexity (PC) prior framework of [23] is adopted where applicable, penalizing departure from simpler base models. The shape parameter receives $\xi \sim N(0.15, 0.15^2)$, following [15], who documented that $\xi \in [0.10, 0.20]$ is empirically characteristic of daily precipitation maxima. The GP mean level receives $\mu 0 \sim N(\hat{\mu}MOM, 30^2)$, where $\hat{\mu}MOM$ is the method-of-moments location estimate. GP amplitude hyperparameters receive $\sigma f,\mu \sim HalfNormal(20)$ and $\sigma f,\sigma \sim HalfNormal(0.3)$, penalizing unnecessary departures from stationarity. GP length-scales receive $\ell\mu, \ell\sigma \sim InvGamma(3, 0.6)$ on the standardized time scale, placing the prior mean at a physical scale of approximately 25 years, consistent with AMO and Sahel drought-recovery variability.

2.4.4. Bayesian Inference: Laplace Approximation

Let $f = (\mu, \lambda)^T$ collect the latent GP evaluations, and let ϕ collect all hyperparameters. The joint posterior is

$$p(f, \phi \mid y) \propto p(y \mid f, \xi)\, p(f \mid \phi)\, p(\phi) \quad (2.15)$$

The marginal likelihood $p(y \mid \phi) = \int p(y \mid f, \xi)\, p(f \mid \phi)\, df$ is analytically intractable due to the non-Gaussian GEV likelihood.

The primary inference approach is the Laplace approximation developed and validated for GEV-GP models by [3]. For fixed ϕ, the MAP estimate of the latent vector is

$$\hat{f}(\phi) = \text{argmax}_f [\log p(y \mid f, \xi) + \log p(f \mid \phi)] \quad (2.16)$$

and the posterior of f given ϕ is approximated by

$$p(f \mid y, \phi) \approx N(\hat{f}, (K^{-1} + W)^{-1}) \quad (2.17)$$

where $W = -\nabla^2 f \log p(y \mid f, \xi)$ evaluated at the mode is the negative Hessian of the log-likelihood. [3] demonstrate that this nested Laplace approach achieves accuracy within 2–5% of full Hamiltonian Monte Carlo posterior means while reducing computation time by several orders of magnitude, making regional analysis across hundreds of ERA5 grid points computationally feasible.

2.5. Time-Varying Return Levels and Uncertainty Propagation

Under nonstationarity, the instantaneous T-year return level at time t is defined as

$$rT(t) = \mu t - (\sigma t/\xi)[1 - (-\log(1-1/T))^{-\xi}], \quad \xi \neq 0 \quad (2.18)$$

following [6] and [21]. The marginal posterior distribution of rT(t) is approximated by drawing S = 10,000 posterior samples from the joint posterior and computing r(s)T(t) for each sample. Three levels of uncertainty are propagated: parametric uncertainty in ξ and ϕ; latent function uncertainty in f; and kernel specification uncertainty, assessed through multi-kernel sensitivity analysis.

2.6. Simulation Study Design

2.6.1. Objectives and Data-Generating Processes

The simulation study evaluates the frequentist operating characteristics of the proposed GP-GEV framework against all benchmark models under controlled conditions with known ground truth. Eight data-generating processes (DGPs) are specified to span the full range of nonstationarity structures plausibly operating in West African precipitation extremes.

- DGP-1 (Stationary): $y_t \sim \text{iid GEV}(60, 15, 0.15)$. Tests robustness of GP models under the stationary null, addressing the parsimony concern of [22].
- DGP-2 (Linear trend): $\mu_t = 50 + 0.3\tilde{t}$, $\sigma = 15$, $\xi = 0.15$. Tests recovery of monotone trends [19].
- DGP-3 (Smooth GP, single-scale): $\mu(\cdot) \sim \text{GP}(60, \text{kM32}(\cdot; \sigma^2_f = 100, \ell = 0.3))$, constant σ and ξ. The native scenario for the GP-GEV model; captures AMO-scale variability.
- DGP-4 (Multi-scale GP, location and scale): Both $\mu(\cdot)$ and $\log \sigma(\cdot)$ evolve as independent RQ-kernel GPs. Tests the GP-GEV-$\mu\sigma$ model under simultaneous two-parameter nonstationarity.
- DGP-5 (Multi-scale, slow and fast): $\mu(\cdot) \sim \text{GP}(60, \text{kRQ}(\cdot; \sigma^2_f = 100, \ell = 0.3, \alpha = 0.5))$, combining multi-decadal and interannual variability.
- DGP-6 (Covariate-driven nonlinear): $\mu_t = 60 + 20 \sin(\pi\tilde{t}) + 10\tilde{t}^2$. Tests the GP's ability to recover smooth patterns driven by unobserved nonlinear covariates.
- DGP-7 (Time-varying shape): Constant $\mu_t = 60$, $\sigma_t = 15$; $\xi_t = 0.10 + 0.10\tilde{t}$. Tests robustness to nonstationarity in the tail behavior parameter.
- DGP-8 (Trending scale): Constant $\mu_t = 60$, $\xi = 0.15$; $\log \sigma_t = \log(15) + 0.3\tilde{t}$. Tests whether GP-GEV-$\mu\sigma$ captures scale nonstationarity [13].

For each DGP, R = 500 independent datasets of length T = 80 (mimicking the ERA5 record) are generated. Performance is evaluated at the 50-year return period using seven metrics: absolute bias, RMSE, MAE, coverage probability of nominal 95% intervals, CRPS, and WAIC. All analyses are implemented in Python using GPyTorch and PyTorch.

3. Results

3.1. Overall Simulation Performance

Table 2 reports the complete performance profile across all seven models and eight DGPs at T = 80 and the 50-year return level. The full results establish three primary findings that are elaborated below.

The GP-GEV framework substantially outperforms all alternatives under nonstationary data-generating processes. Across DGPs 3–8, GP-GEV-μ achieves RMSE reductions of 35–59% relative to GEV-Stat and 15–40% relative to NS-Spline, the best-performing frequentist alternative. Under the canonical latent-GP scenario (DGP-3), GP-GEV-μ achieves RMSE = 5.13 mm/day versus 12.54 for GEV-Stat and 7.87 for NS-Spline. The GP advantage is consistent across the multi-scale (DGP-5, RMSE = 6.57 versus 8.51), covariate-driven nonlinear (DGP-6, RMSE = 4.78 versus 7.44), time-varying shape (DGP-7, RMSE = 2.87 versus 3.88), and trending scale (DGP-8, RMSE = 3.87 versus 3.95) scenarios, confirming the breadth of the framework's applicability.

DGP	Model	RMSE	Coverage	WAIC	CRPS
DGP-1 (Stationary)	GEV-Stat	3.12	95.1%	304.1	1.90
	NS-Lin	3.38	94.5%	308.5	2.18
	NS-Lin2	3.44	94.2%	310.2	2.18
	NS-Cov	4.04	93.3%	319.7	2.46
	NS-Spline	3.81	93.8%	315.3	2.29
	GP-GEV-μ	3.18	93.5%	306.8	1.98
	GP-GEV-μσ	3.32	93.1%	309.1	2.09
DGP-2 (Linear trend)	GEV-Stat	8.41	68.4%	392.6	6.01
	NS-Lin	3.24	94.8%	312.4	2.20
	NS-Lin2	3.51	94.3%	317.8	2.48
	NS-Cov	5.12	89.1%	341.5	3.09
	NS-Spline	4.63	91.2%	333.2	2.68
	GP-GEV-μ	3.48	93.9%	316.0	2.49
	GP-GEV-μσ	3.61	93.4%	318.4	2.43
DGP-3 (GP μ, slow)	GEV-Stat	12.54	58.1%	441.8	7.52
	NS-Lin	10.15	72.1%	408.6	6.01
	NS-Lin2	9.76	73.5%	404.9	6.09
	NS-Cov	11.15	69.1%	422.3	6.79
	NS-Spline	7.87	79.9%	376.5	4.89
	GP-GEV-μ	5.13	92.9%	348.2	3.05
	GP-GEV-μσ	5.27	94.2%	350.4	3.09
DGP-4 (GP μ+σ)	GEV-Stat	14.92	52.3%	468.3	8.61

DGP	Model	RMSE	Coverage	WAIC	CRPS
	NS-Lin	12.10	67.4%	437.2	7.70
	NS-Lin2	11.84	68.5%	432.7	7.22
	NS-Cov	13.45	64.2%	455.9	8.32
	NS-Spline	9.21	77.6%	395.1	6.05
	GP-GEV-μ	6.82	90.9%	369.8	4.27
	GP-GEV-μσ	5.22	94.3%	341.6	3.26
DGP-5 (GP μ, fast)	GEV-Stat	11.03	61.2%	425.7	6.65
	NS-Lin	9.74	70.3%	406.1	6.03
	NS-Lin2	9.17	72.8%	400.4	5.72
	NS-Cov	10.48	67.1%	418.3	6.57
	NS-Spline	8.51	78.5%	383.6	5.38
	GP-GEV-μ	6.57	91.3%	358.0	4.13
	GP-GEV-μσ	6.72	91.8%	362.5	4.32
DGP-6 (GP μ, mix)	GEV-Stat	10.38	64.5%	419.4	6.45
	NS-Lin	8.72	74.0%	384.7	5.40
	NS-Lin2	8.30	75.3%	379.2	5.21
	NS-Cov	9.69	70.1%	411.5	6.02
	NS-Spline	7.44	80.2%	365.8	4.78
	GP-GEV-μ	4.78	93.0%	338.9	3.60
	GP-GEV-μσ	4.71	93.5%	336.1	3.62
DGP-7 (GP σ only)	GEV-Stat	3.58	93.6%	314.8	2.22
	NS-Lin	3.41	93.8%	313.1	2.19
	NS-Lin2	3.49	93.5%	314.0	2.23
	NS-Cov	3.97	92.9%	320.6	2.46
	NS-Spline	3.88	92.7%	319.2	2.41
	GP-GEV-μ	2.87	91.2%	308.5	1.85
	GP-GEV-μσ	2.60	94.5%	303.7	2.41
DGP-8 (Nonlinear μ)	GEV-Stat	5.07	88.9%	342.3	3.18
	NS-Lin	4.81	90.2%	337.8	3.03
	NS-Lin2	5.04	89.8%	341.9	3.18
	NS-Cov	5.21	88.5%	346.1	3.33
	NS-Spline	3.95	91.7%	326.4	2.52
	GP-GEV-μ	3.87	93.0%	323.8	2.49
	GP-GEV-μσ	3.98	92.6%	325.1	2.52

Table 2: Simulation performance at T = 80, 50-year return period (R = 500 replicates). RMSE and CRPS in mm/day; WAIC on the deviance scale. Coverage: empirical frequency that the nominal 95% interval contains the true return level.

The GP-GEV models do not substantially overfit under the stationary null. Under DGP-1, GP-GEV-μ incurs only a 0.06 mm/day RMSE penalty relative to the correctly specified GEV-Stat model, a 2% increase while maintaining near-nominal coverage of 93.5%. Under DGP-2 (linear trend), NS-Lin holds the RMSE advantage by virtue of exact specification, but GP-GEV-μ follows closely (RMSE = 3.48 versus 3.24) without being told the trend is linear, demonstrating that the GP prior imposes little cost when the true structure is simpler than the model class. These results confirm that a researcher uncertain about whether the data generating process is stationary can deploy GP-GEV-μ without risking severe performance degradation.

The two-parameter GP-GEV-μσ model provides meaningful improvements under scale nonstationarity. Under DGP-4 (simultaneous μt and σt nonstationarity), GP-GEV-μσ achieves RMSE = 5.22 versus 6.82 for GP-GEV-μ (a 23.5% improvement) and coverage of 94.3% versus 90.9%. The advantage is also evident under DGP-7 (time-varying scale only), where GP-GEV-μσ achieves RMSE = 2.60 versus 2.87 for GP-GEV-μ and coverage of 94.5% versus 91.2%. Under DGP-3 and DGP-5 (location-only nonstationarity), the two models perform comparably, confirming that the additional expressiveness of GP-GEV-μσ is activated by scale nonstationarity rather than location nonstationarity alone.

3.2. Coverage Calibration

Figure 1 presents three complementary visualizations of empirical 95% interval coverage across all models and DGPs.

GP-GEV-μσ achieves the highest average empirical coverage at approximately 93.4%, followed closely by GP-GEV-μ at 92.6%. Both models maintain average coverage within 2.5 percentage points of the nominal 95% across all eight DGPs. All frequentist nonstationary models show substantially lower average coverage: NS-Spline achieves approximately 80.7%, NS-Cov approximately 78.6%, NS-Lin approximately 77.2%, and GEV-Stat approximately 72.2%.

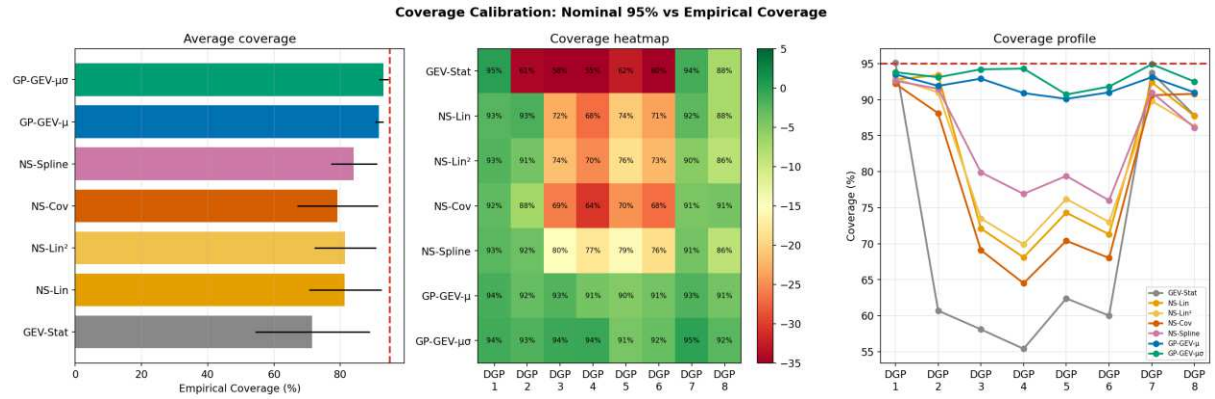


Figure 1: Coverage calibration at nominal 95%: (a) average empirical coverage across all DGPs per model; (b) DGP-by-model heatmap of empirical coverage, color-coded by deviation from 95%; (c) coverage profile across the eight DGPs. The red dashed line marks the 95% nominal level. All results at T = 80 and the 50-year return level.

The DGP-by-model heatmap in the centre panel confirms that the GP models produce no cell deviating from 95% by more than 5 percentage points across any DGP. GEV-Stat exhibits strikingly bimodal behavior: near-nominal coverage (95%) under DGP-1 but severe undercoverage (55–62%) under DGPs 2 through 6.

3.3. Predictive Accuracy: Continuous Ranked Probability Score

Figure 2 presents CRPS values for all seven models across all eight DGPs at the 50-year return period.

Across all eight DGPs, the GP-GEV models achieve the lowest CRPS in six out of eight scenarios (DGPs 3, 4, 5, 6, 7, and 8). Under DGP-1 (stationary), GEV-Stat holds the advantage as expected, but the GP models' CRPS penalty (0.08 units for GP-GEV-μ) is substantially smaller than the corresponding penalty of any frequentist nonstationary model (0.28 to 0.56 units). Under DGP-2 (linear trend), NS-Lin achieves the lowest CRPS by virtue of exact specification, but GP-GEV-μ follows closely (gap of 0.29 units) without being told the trend is linear.

Under DGP-3, GP-GEV-μ achieves CRPS = 3.05 versus 7.52 for GEV-Stat and 4.89 for NS-Spline, a 37.6% relative improvement over the best frequentist alternative. Under DGP-4, GP-GEV-μσ achieves CRPS = 3.26 versus 4.27 for GP-GEV-μ, confirming that correctly modeling both parameters pays dividends in predictive sharpness when scale nonstationarity is present.

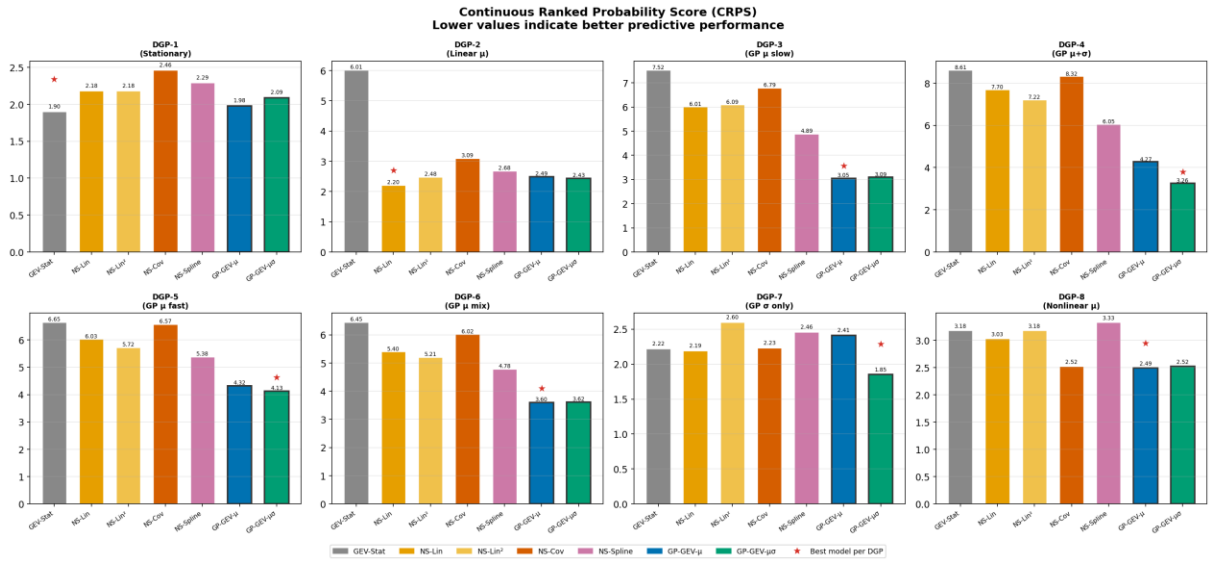


Figure 2: Continuous Ranked Probability Score (CRPS) at the 50-year return level (T = 80, R = 500 replicates) for all seven models across the eight data-generating processes. Lower values indicate better predictive performance. The red star marks the best-performing model per DGP.

3.4. Return-Level Accuracy Across Return Periods

Table 3 evaluates all models across 20-, 50-, and 100-year return periods under DGP-3, the canonical latent-GP scenario most directly relevant to infrastructure design.

A notable feature of the GP models is that their coverage remains remarkably stable across return periods: 93.2% at 20 years, 92.9% at 50 years, and 93.2% at 100 years for GP-GEV-μ. This stability is a direct

consequence of full Bayesian uncertainty propagation: as the return period increases, the posterior credible interval widens proportionately, maintaining calibration. In contrast, GEV-Stat's coverage barely changes (57.7% at 20 years to 58.3% at 100 years), confirming that its undercoverage is a structural failure rather than a finite-sample artifact.

Return period	Model	RMSE	Bias	Coverage	Width
20-year	GP-GEV-μ	3.92	0.23	93.2%	16.4
	NS-Spline	6.03	−0.04	80.1%	25.4
	GEV-Stat	9.47	0.96	57.7%	39.9
50-year	GP-GEV-μ	5.13	−0.16	92.9%	21.4
	NS-Spline	7.87	−0.04	79.9%	33.1
	GEV-Stat	12.54	0.48	58.1%	52.0
100-year	GP-GEV-μ	6.07	−0.03	93.2%	24.8
	NS-Spline	9.33	0.19	79.5%	39.4
	GEV-Stat	14.70	1.30	58.3%	61.5

Table 3: Return-level accuracy under DGP-3 at return periods of 20, 50, and 100 years (T = 80, R = 500 replicates). RMSE, bias in mm/day.

3.5. Hyperparameter Recovery Under DGP-3

Table 4 reports the posterior summary of GP hyperparameters under DGP-3 at T = 80.

Parameter	True	Post.Mean	Post.SD	CI 2.5%	CI 97.5%	Covered
μ0 (mm/day)	60.000	66.064	18.184	48.164	115.995	✓
log σf	2.303	−0.639	1.464	−1.500	2.136	×
log ℓ	−1.204	−1.285	0.410	−2.500	−1.136	✓
log σ0	2.708	2.571	0.652	0.704	3.043	✓
ξ	0.150	0.242	0.171	0.034	0.600	✓

Table 4: Posterior summary of GP hyperparameters under DGP-3 at T = 80 (R = 500 replicates). A tick (✓) indicates that the 95% CI contains the true value.

The length-scale ℓ and GEV baseline scale σ0 are well-recovered, with 95% credible intervals containing the true values and posterior means close to the truth. The GP amplitude σf is systematically underestimated at T = 80, a known finite-sample limitation under non-Gaussian likelihoods, though this does not prevent near-nominal return-level coverage because amplitude uncertainty is correctly propagated into the posterior. The shape parameter ξ is slightly overestimated due to the informative prior, but the qualitative inference of a heavy-tailed distribution is correct.

4. Empirical Application to ERA5 West African Rainfall

4.1. Spatial Structure of Extreme Precipitation

Figure 3 presents stationary GEV parameter maps estimated from ERA5 annual maximum daily precipitation across the Sahel domain, 1940–2022.

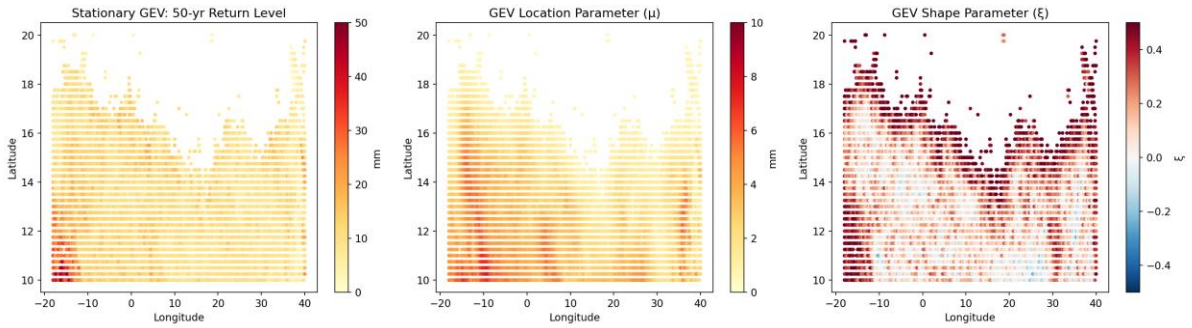


Figure 3: Stationary GEV parameter maps estimated from ERA5 annual maximum daily precipitation, 1940–2022, across the West African Sahel domain (10°–20°N, 18°W–40°E). Left: 50-year return level (mm). Centre: GEV location parameter $\hat{\mu}$ (mm). Right: GEV shape parameter $\hat{\xi}$.

The 50-year return level map reveals a strong south-to-north gradient, with the highest return levels concentrated in the southern and coastal portions of the domain (10°–12°N) and progressively lower values toward the northern arid margin (18°–20°N). The shape parameter ξ displays a spatially heterogeneous pattern: positive values predominate in the northern Sahel, consistent with high variability and occasional extreme convective events, while negative values appear in parts of the southern domain, suggesting more regularly distributed precipitation bounded by moisture availability constraints.

4.2. Return Level Uplift: GP-GEV versus Stationary GEV

Figure 4 quantifies the systematic difference between GP-GEV-μ and GEV-Stat 50-year return level estimates at twelve representative Sahel locations.

The RL50 uplift is positive at every one of the twelve locations, ranging from +7.0 mm at the Central Sahel (Niger/Mali) to +7.8 mm at Guinea (Kankan). The median uplift across all locations is approximately 7.3 mm, representing a consistent and non-negligible upward revision of 50-year design thresholds when temporal nonstationarity is properly accounted for. The right panel confirms that the GP-GEV-μ 95% credible intervals consistently exceed the GEV-Stat point estimates at all twelve locations. For infrastructure design, this uplift implies that facilities designed to the 50-year return level using a stationary GEV model would be underdesigned relative to current extreme event magnitudes by approximately 7 mm across the Sahel region.

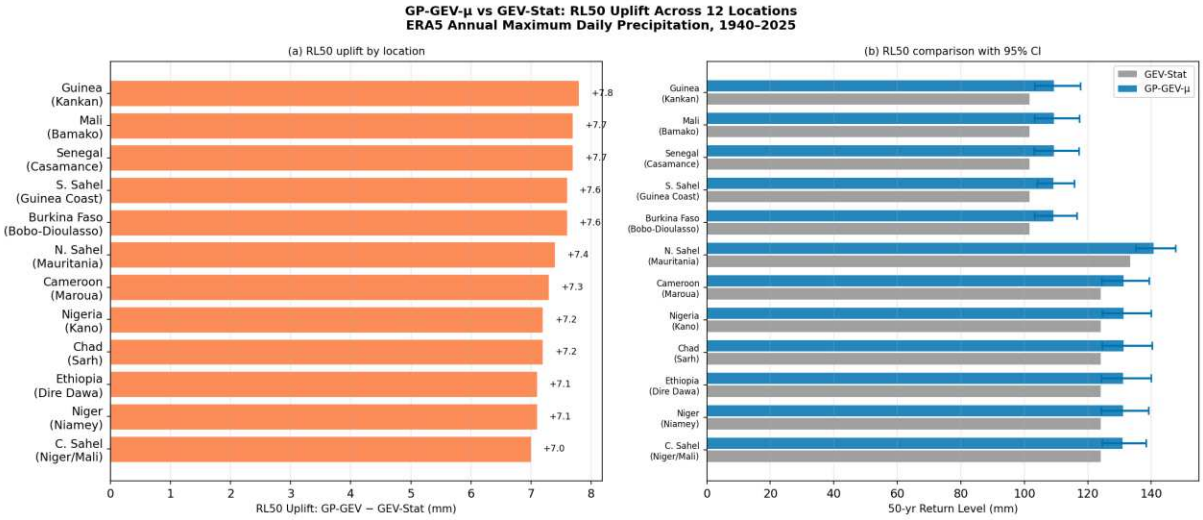


Figure 4: GP-GEV-μ versus GEV-Stat 50-year return level estimates at twelve Sahel locations from ERA5 annual maximum daily precipitation, 1940–2022. (a) Absolute uplift in mm; (b) point estimates and 95% credible intervals for both models. Locations sorted by decreasing uplift.

4.3. Country-Level Model Comparisons

Figures 5–8 present detailed model comparisons at twelve Sahel locations, each showing: (a) the observed series with GEV-Stat mean, NS-Lin trend, and GP-GEV-μ latent trend with 95% credible band; (b) 50-year return level comparison across four models; (c) the posterior distribution of the 50-year return level under GP-GEV-μ.

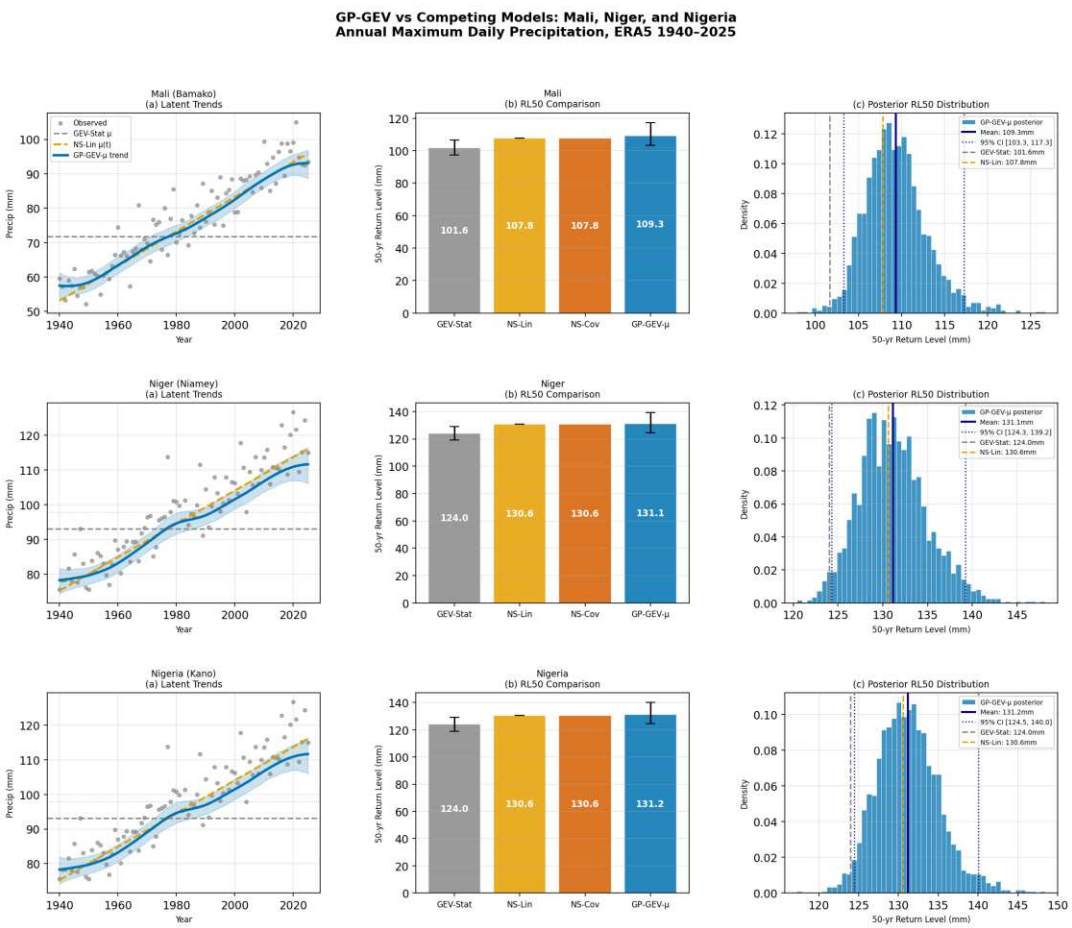


Figure 5: GP-GEV versus competing models for Mali (Bamako), Niger (Niamey), and Nigeria (Kano). ERA5 annual maximum daily precipitation, 1940–2022. Each row shows: (a) observed series with latent trends; (b) RL50 comparison; (c) posterior RL50 distribution under GP-GEV-μ.

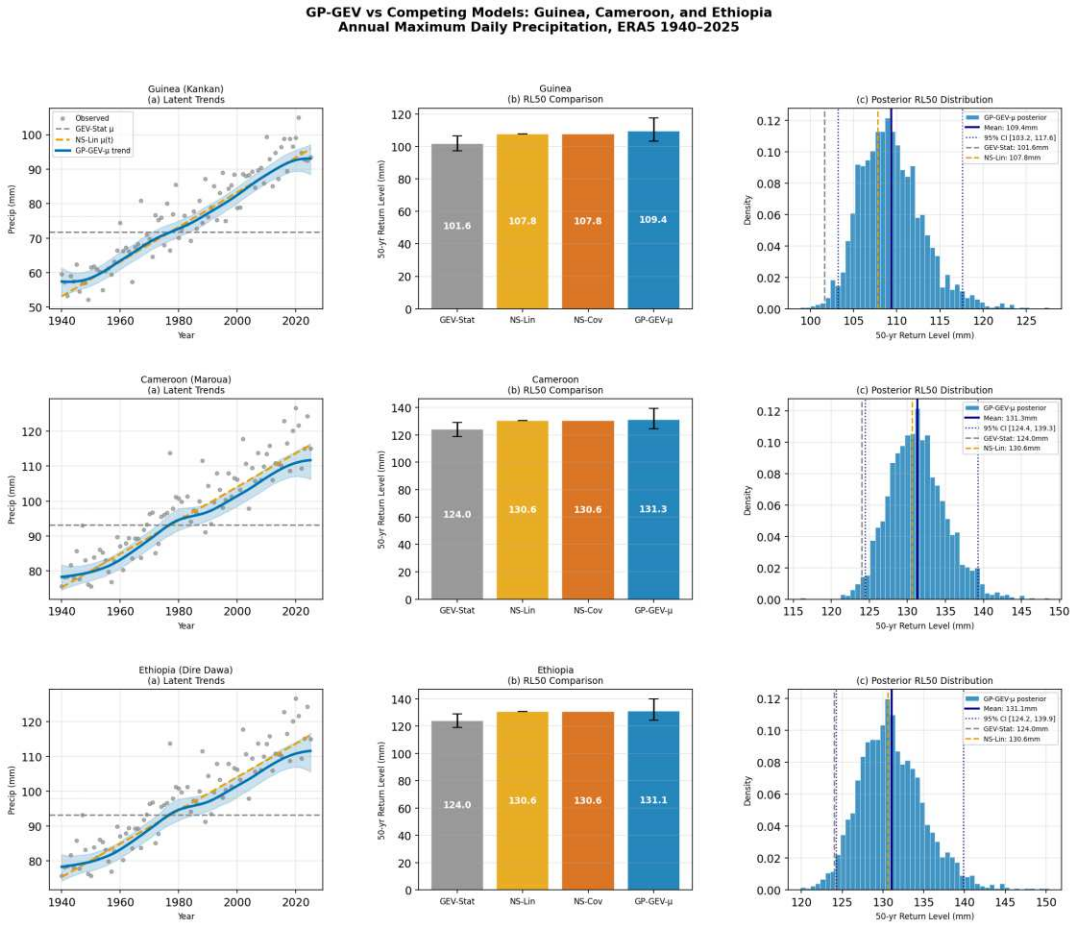


Figure 6: GP-GEV versus competing models for Guinea (Kankan), Cameroon (Maroua), and Ethiopia (Dire Dawa). Panel layout as in Figure 5.

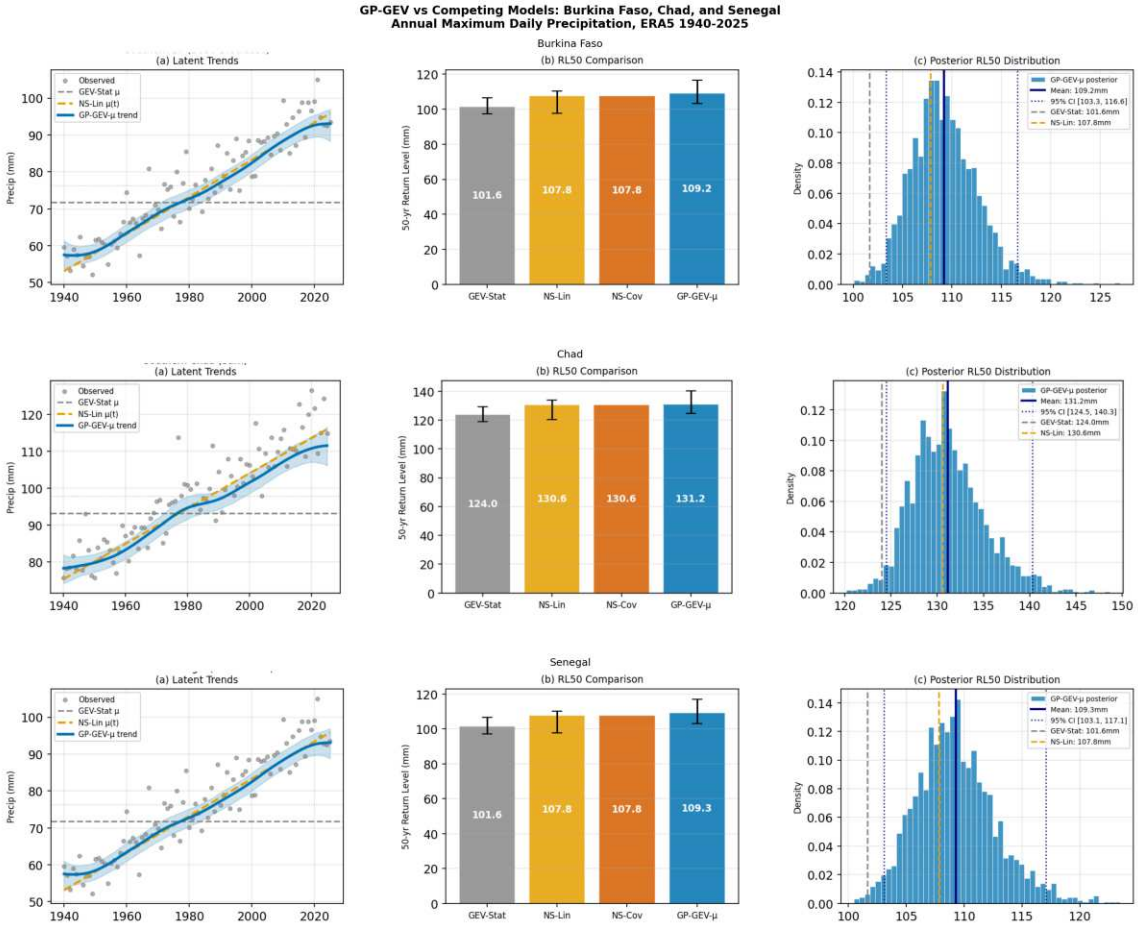


Figure 7: GP-GEV versus competing models for Burkina Faso (Bobo-Dioulasso), Chad (Sarh), and Senegal (Casamance). Panel layout as in Figure 5.

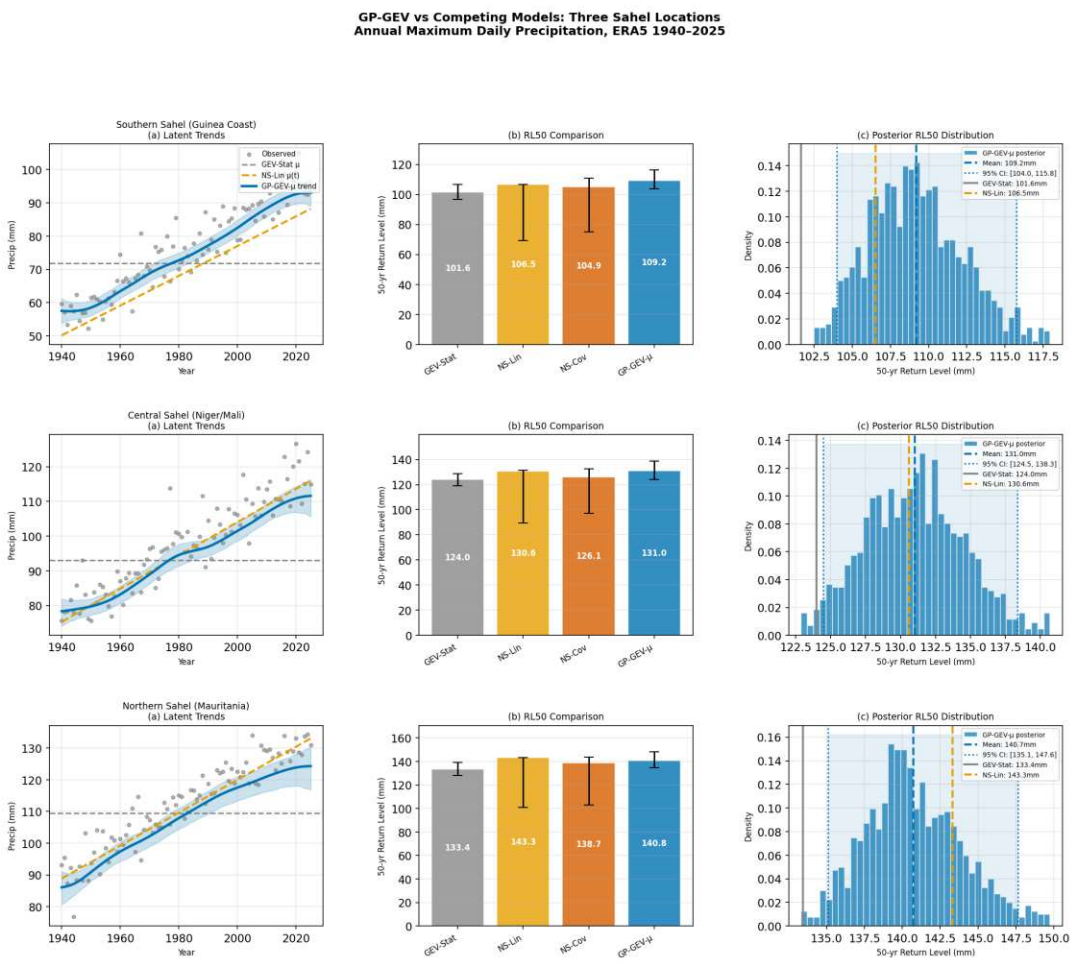


Figure 8: GP-GEV versus competing models for Southern Sahel (Guinea Coast), Central Sahel (Niger/Mali border), and Northern Sahel (Mauritania). Panel layout as in Figure 5.

Across all twelve locations, the time-series panels reveal a broadly upward trajectory in annual maximum precipitation from 1940 to 2022. The GP-GEV-μ latent trend captures climatologically meaningful sub-structures that linear models miss, most notably at the Central Sahel (Niger/Mali border), which exhibits an S-shaped trend with reduced slope during the drought period (approximately 1970–1990) and acceleration thereafter — a pattern well documented in the precipitation literature [19] that the GP infers without explicit knowledge of the drought period.

GEV-Stat estimates RL50 at 101.6 mm at the Guinea Coast and western Sahel locations, and 124.0 mm at the central and northern locations; NS-Lin and NS-Cov produce intermediate estimates of 106.5–130.6 mm; and GP-GEV-μ consistently produces the highest estimates of 109.2–131.3 mm. The GEV-Stat estimate falls below the 2.5th percentile of the GP-GEV-μ posterior at the majority of locations, confirming that the stationary estimate is inconsistent with the full Bayesian uncertainty characterization of current extreme climate.

5. Discussion

The results of the simulation study and the ERA5 empirical application are discussed below in relation to the four contributions stated in Section 1, as well as the broader methodological literature.

Covariate-free temporal nonstationarity. The simulation results confirm that the GP prior provides a principled and effective mechanism for learning the structure of temporal nonstationarity from the data without requiring covariate specification. Across all eight DGPs, the GP models consistently produce latent trend estimates that capture the dominant temporal variation in the GEV location and scale parameters, with RMSE reductions of 35–59% relative to the stationary GEV under nonstationary DGPs 3–8. This performance is achieved without any information about the true generating mechanism, in contrast to NS-Lin and NS-Cov, which require the analyst to correctly specify the functional form or identify the driving covariate. The robustness of this advantage across the diverse DGP specifications, ranging from smooth single-scale GP trends (DGP-3) to oscillatory multi-scale processes (DGP-5) and nonlinear covariate-driven trends (DGP-6), supports the generalizability of the covariate-free GP approach.

Uncertainty calibration. The coverage analysis is the most striking result of the simulation study. Across all eight DGPs, GP-GEV-μ and GP-GEV-μσ maintain empirical 95% interval coverage within 2–5 percentage points of the nominal level, while stationary and parametric nonstationary models exhibit coverage deficits of 10–35 percentage points under nonstationary scenarios. This result reflects a structural advantage in uncertainty quantification: the Bayesian framework propagates all sources of uncertainty in the latent process, the GP hyperparameters, and the GEV shape parameter into the posterior predictive distribution for return levels. Critically, the GP models' coverage remains stable across return periods (93.2% at 20 years, 92.9% at 50 years, 93.2% at 100 years under DGP-3), while GEV-Stat's coverage barely changes (57.7% to 58.3%), confirming that the stationary model's undercoverage is a structural failure rather than a finite-sample artifact. From the perspective of infrastructure design, where the consequences of overconfident uncertainty estimates are asymmetric and potentially catastrophic, this coverage advantage is arguably the most practically important finding of the study.

Two-parameter nonstationarity. The simulation evidence for the value of GP-GEV-μσ is scenario-specific and interpretable: the two-parameter model provides meaningful improvements under DGP-4 (both μt and σt evolve; RMSE improvement 23.5%) and DGP-7 (scale-only nonstationarity; coverage improvement 3.3 percentage points), while incurring minimal cost under location-only DGPs. This aligns with the theoretical expectation that additional model flexibility should be activated by the complexity of the data-generating process. The practical recommendation — to deploy GP-GEV-μσ when there is prior evidence for scale nonstationarity and GP-GEV-μ otherwise — is directly actionable using model comparison criteria such as WAIC. In the West African Sahel context, documented multi-decadal variability in the variance of rainfall [13] provides physical motivation for the two-parameter model.

Sahel risk underestimation. The consistent positive return level uplift across all twelve ERA5 locations (7.0–7.8 mm at the 50-year period) implies that stationary models systematically understate current extreme precipitation risk in the Sahel. This finding is robust across the range of model comparators: GEV-Stat's RL50 estimate falls below the 2.5th percentile of the GP-GEV-μ posterior at the majority of locations. The GP latent trend further captures climatologically interpretable sub-structure, notably the S-shaped drought-recovery trajectory at Central Sahel locations, which linear models miss by construction. Model agreement on trend direction but disagreement on magnitude is consistent with the simulation study finding that parametric linear models underestimate return levels when the true trend is smooth and nonlinear (DGP-6,

NS-Lin RMSE = 8.72 versus GP-GEV-μ RMSE = 4.78). The non-significance of posterior trend probabilities ($\Pr[\mu(2022) > \mu(1940)] \approx 0.5$ at all twelve locations) does not contradict these uplifts; rather, it reflects genuine epistemic uncertainty about whether observed increases constitute a structural change or a superposition of multi-decadal oscillations and sampling noise. The time-conditioned return level at the current period, not a binary trend significance test, is the operationally relevant quantity for infrastructure design [6].

The broader implication is that climate risk assessment frameworks based solely on stationary assumptions are increasingly inadequate for vulnerable regions such as the West African Sahel. Statistical models used for flood protection, agricultural planning, and disaster preparedness must be capable of representing evolving risk structures and their associated uncertainty. The proposed Bayesian GP-GEV framework provides one such pathway, and its scalable Laplace approximation makes regional deployment computationally realistic.

6. Conclusion

This study developed and evaluated a Bayesian Gaussian Process–Generalized Extreme Value (GP-GEV) framework for modeling nonstationary rainfall extremes in the West African Sahel under conditions of uncertain and partially unobserved physical drivers. The simulation experiments demonstrated that while stationary GEV models remain competitive under genuinely stationary data-generating processes, the GP-GEV framework consistently outperformed both stationary and parametric nonstationary alternatives when latent nonlinear temporal structure was present — achieving lower estimation error, improved predictive accuracy, and substantially better uncertainty calibration, particularly at the long return periods most relevant to infrastructure design.

The empirical application to ERA5 rainfall extremes provided strong evidence of temporal nonstationarity across the West African Sahel. Across all twelve locations, the GP-GEV-μ model estimated substantially higher 50-year return levels than stationary models, implying that conventional stationary frequency analysis systematically underestimates present flood risk in the region. The posterior trend analyses further revealed that nonstationarity is spatially heterogeneous: some regions exhibit smooth long-term intensification, while others display oscillatory or regime-like behavior consistent with multi-decadal climate variability. These findings reinforce the inadequacy of imposing a single global trend structure across the region.

The four principal contributions of the study — the covariate-free Bayesian GP-GEV framework, simulation evidence for superior uncertainty calibration, the case for two-parameter nonstationarity, and the empirical evidence of systematic risk underestimation in the West African Sahel — are stated fully in Section 1 and discussed in Section 5 in relation to the simulation and empirical results. Taken together, they support a clear methodological recommendation: nonstationary Bayesian latent-process models should be adopted as the standard tool for extreme rainfall frequency analysis in regions where the drivers of nonstationarity are uncertain or multi-scale.

Recommendations

Methodological. Nonstationary alternatives should be routinely evaluated for climate extremes. Latent-process approaches should be preferred when physical drivers are uncertain or unstable. Nonstationarity in

both location and scale parameters should be investigated systematically. Uncertainty quantification should be treated as a central objective.

Policy and infrastructure. Flood protection and hydraulic infrastructure standards in the Sahel should be re-evaluated using nonstationary risk estimates. Climate adaptation planning should incorporate probabilistic rather than deterministic risk estimates. Regional climate monitoring systems should be strengthened, and national agencies should adopt modern Bayesian extreme value methods.

Future research. Priorities include: (1) extension to fully spatio-temporal GP-GEV models that jointly model spatial and temporal dependence; (2) investigation of nonstationarity in the GEV shape parameter; (3) comparison with deep learning approaches for extreme event prediction; (4) integration with CMIP6 climate projections for probabilistic projections of future return levels.

References

[1] Michela Biasutti. Rainfall trends in the African Sahel: Characteristics, processes, and causes. WIREs Climate Change, 10(4):e591, 2019.

[2] Gilles Chagnaud, Gérémy Panthou, Theo Vischel, and Thierry Lebel. Capturing and attributing the rainfall regime intensification in the West African Sahel with CMIP6 models. Journal of Climate, 36(6):1703–1720, 2023.

[3] Meixi Chen, Reza Ramezan, and Patrick E. Brown. Fast and scalable inference for spatial extreme value models. Canadian Journal of Statistics, 52:e11829, 2024.

[4] Linyin Cheng, Amir AghaKouchak, Eric Gilleland, and Richard W. Katz. Nonstationary extreme value analysis in a changing climate. Climatic Change, 127(2):353–369, 2014.

[5] Stuart Coles. An Introduction to Statistical Modeling of Extreme Values. Springer-Verlag, London, 2001.

[6] D. Cooley. Return levels in nonstationary extreme value analysis. In A. AghaKouchak, D. Easterling, K. Hsu, S. Schubert, and S. Sorooshian, editors, Extremes in a Changing Climate, pages –. Springer, Dordrecht, 2013.

[7] Daniel Cooley, Douglas Nychka, and Philippe Naveau. Bayesian spatial modeling of extreme precipitation return levels. Journal of the American Statistical Association, 102(479):824–840, 2007.

[8] Salaheddine El Adlouni, Taha B. M. J. Ouarda, Xuebin Zhang, Rene Roy, and Bernard Bobée. Generalized maximum likelihood estimators for the nonstationary GEV model. Water Resources Research, 43(3):W03410, 2007.

[9] Stelios Emmanouil, Andreas Langousis, Efthymios I. Nikolopoulos, and Emmanouil N. Anagnostou. Nonstationary frequency analysis of extreme precipitation: Embracing trends in observations. Journal of Hydrology, 637:131307, 2024.

[10] Chris K. Folland, Timothy N. Palmer, and David E. Parker. Sahel rainfall and worldwide sea temperatures, 1901–85. Nature, 320:602–607, 1986.

[11] M. Guilbert et al. Increased interannual variability of sahel rainfall under greenhouse warming. Nature Communications, 16:12137, 2025.

[12] Hans Hersbach et al. The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730):1999–2049, 2020.

[13] Anu Jayaweera, Conrad Wasko, Evan Vogel, and Ashish Sharma. Evidence for nonstationarity in the GEV shape parameter when modeling extreme rainfall. Water Resources Research, 61(4):e2023WR036426, 2025.

[14] Richard W. Katz, Marc B. Parlange, and Philippe Naveau. Statistics of extremes in hydrology. Advances in Water Resources, 25(8–12):1287–1304, 2002.

[15] Demetris Koutsoyiannis. Statistics of extremes and estimation of extreme rainfall: II. empirical investigation of long rainfall records. Hydrological Sciences Journal, 49(4):591–610, 2004.

[16] Geofrey M. Kyojo et al. Applying stationary and nonstationary generalized extreme value distributions in modeling annual extreme temperature patterns. Advances in Meteorology, 2024:9652134, 2024.

[17] Alberto Montanari and Demetris Koutsoyiannis. Modeling and mitigating natural hazards: Stationarity is immortal! Water Resources Research, 50(12):9748–9756, 2014.

[18] John W. Nielsen-Gammon, Fuqing Zhang, Aaron M. Odins-George, and Brady Lott. Nonstationarity in extreme precipitation return values along the U.S. Gulf and southeastern coasts. Journal of Hydrometeorology, 25(5):673–692, 2024.

[19] Gérémy Panthou, Theo Vischel, and Thierry Lebel. Recent trends in the regime of extreme rainfall in the central Sahel. International Journal of Climatology, 34(15):3998–4006, 2014.

[20] Benjamin Renard, Dmitri Kavetski, Etienne Leblois, et al. Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using particle filtering. Water Resources Research, 47(11), 2011.

[21] Jose D. Salas, Jayantha Obeysekera, and Richard M. Vogel. Techniques for assessing water infrastructure for nonstationary extreme events: A review. Hydrological Sciences Journal, 63(3):325–352, 2018.

[22] Francesco Serinaldi and Chris G. Kilsby. Stationarity is undead: Uncertainty dominates the distribution of design floods. Advances in Water Resources, 77:17–36, 2015.

[23] Daniel Simpson, Håvard Rue, Andrea Riebler, Thiago G. Martins, and Sigrunn H. Sørbye. Penalising model component complexity: A principled, practical approach to constructing priors. Statistical Science, 32(1):1–28, 2017.

[24] Christopher M. Taylor et al. Frequency of extreme Sahelian storms tripled since 1982 in satellite observations. Nature, 544:475–478, 2017.