

Bayesian Modeling of Monthly Upper Record Precipitation Using the Exponentiated Log Logistic Distribution: A Case Study from the Upper Indus Basin

Rabia Azeem¹, Tahir Mehmood^{2*} 

¹Department of Mathematics and Statistics, Riphah International University, Islamabad, Pakistan; ²School of Natural Sciences, National University of Sciences and Technology, Sector H-12, Islamabad, Pakistan

*Corresponding author: Tahir Mehmood, School of Natural Sciences, National University of Sciences and Technology, Sector H-12, Islamabad, Pakistan. Email: tahime@gmail.com

Abstract

Record-breaking precipitation events represent a stochastic process that differs fundamentally from annual maxima and threshold exceedances. This study examines monthly upper record precipitation in the Upper Indus Basin (UIB) using a Bayesian framework based on the Exponentiated Log Logistic distribution (ELLD). Upper records defined as observations exceeding all previously observed values were extracted from observed monthly precipitation series spanning 1980–2020. Bayesian inference via Markov Chain Monte Carlo was employed to estimate model parameters, quantify uncertainty, and derive predictive distributions for the magnitude of the next potential upper record. Model performance and adequacy were evaluated using a combination of likelihood-based measures, posterior summaries, convergence diagnostics, and posterior predictive simulations to assess the ability of the ELLD to represent the observed record magnitudes under sparse data conditions. Exceedance probabilities for selected precipitation thresholds were derived from the posterior predictive distribution to provide an interpretable probabilistic characterization of extreme record magnitudes. Rather than relying on return periods or return levels, which are not applicable to record-based processes, the analysis adopts a predictive probabilistic framework focused on uncertainty quantification. The study is intended as a methodological illustration of Bayesian inference for record-breaking precipitation under severe data scarcity and does not attempt trend detection, attribution, or climate-change assessment.

Keywords: Bayesian Modeling, Exponentiated Log Logistic Distribution, Monthly Precipitation, Upper Record Values, Upper Indus Basin, Uncertainty Quantification

Introduction

Extreme rainfall events can trigger flash floods, disrupt agricultural activities, and place severe stress on infrastructure, particularly in climate-sensitive regions such as the Upper Indus Basin (UIB) of Pakistan [1, 2, 3]. Both monsoon systems and westerly disturbances influence this high-altitude basin and support millions of people through its role in regional water supply. While mean precipitation and seasonal statistics provide a broad description of hydroclimatic conditions, hydrological impacts are often driven by rare, high-magnitude events that exceed all previously observed values [2].

Classical extreme value approaches, such as the Generalized Extreme Value (GEV) distribution for annual maxima and Peaks-Over-Threshold (POT) models, have been widely



applied in hydrology and climate studies [1, 2, 4]. These approaches are designed to model extremes arising from an underlying stochastic process using relatively dense samples. Record-breaking events, however, constitute a fundamentally different process, characterized by extremely small sample sizes and strong sampling variability [5]. Consequently, standard extreme value concepts such as return periods and return levels are not directly applicable to record sequences.

Record value theory provides an alternative perspective by focusing exclusively on observations that exceed all previously observed values [6]. By construction, upper records form a sparse sequence that captures the most extreme realizations of a process without requiring arbitrary threshold selection or long, homogeneous time series. This property makes record-based analysis attractive for data limited regions where continuous and consistent observations may be unavailable [7]. At the same time, record-based inference is not intended for design purposes or trend detection, but rather for probabilistic characterization under severe data constraints.

Bayesian statistical methods are particularly well suited for inference in such settings, as they allow the incorporation of prior information and provide full posterior distributions for model parameters and predictions [8]. In this study, we adopt a Bayesian framework to model monthly upper record precipitation using the Inverse Gamma Distribution (ELLD). The ELLD has been shown to offer a flexible representation for positive, right-skewed hydrological variables, including rainfall extremes [9, 10]. Here, the ELLD is used as a convenient parametric model for the magnitude of record-breaking precipitation values rather than as a substitute for classical extreme value models.

The analysis is based on observed monthly precipitation data from the UIB covering the period 1980–2020. Upper record values are extracted directly from the observed series without detrending, subsampling, or extrapolation beyond the observation period. The objectives of this study are threefold: (i) to demonstrate the application of Bayesian inference to monthly upper record precipitation values, (ii) to quantify parameter and predictive uncertainty under sparse record data, and (iii) to illustrate probabilistic prediction of the magnitude of the next upper record. No inference regarding trends, nonstationary, or climate-induced changes is attempted [11, 12].

By focusing on record values, this work demonstrates the use of Bayesian ELLD modeling as a complementary tool to classical approaches for exploring extreme precipitation behavior in data-limited regions [13, 14]. The results are intended to clarify the interpretation and limitations of record based precipitation analysis rather than to provide guidance for risk estimation or climate adaptation planning.

Methods

Study Area and Data

The Upper Indus Basin (UIB), located in northern Pakistan, is a hydrologically critical region that supports agriculture, domestic water supply, and hydropower generation for millions of people downstream. The basin is bounded by the Himalaya, Karakoram, and Hindu Kush mountain ranges and receives water from a combination of seasonal precipitation and cryospheric melt that feeds the Indus River system [1]. Precipitation over the UIB is primarily governed by two large scale atmospheric systems: the South Asian summer monsoon, which produces intense rainfall during June to September, and westerly disturbances, which contribute winter and early spring precipitation. The complex topography of the basin leads to pronounced spatial variability in precipitation, while the harsh terrain limits the density and continuity of in situ monitoring networks [15].

In this study, monthly precipitation totals from 1980–2020 were analyzed using observations from 12 high-altitude meteorological stations distributed across the UIB. The data were obtained from the Pakistan Meteorological Department (PMD) and the Water and Power Development Authority (WAPDA) [16]. These stations represent a subset of the basin's limited observational network, which consists of approximately 20 operational stations with varying record lengths and data quality [17]. Stations were selected based on three criteria: (i) data

completeness exceeding 80%, (ii) continuous records spanning at least 30 years, and (iii) spatial coverage across elevations ranging from approximately 1,000 to 4,000 meters above sea level. This selection aimed to balance data quality with spatial representativeness under the constraints of available observations.

Prior to analysis, the precipitation records underwent standard quality control procedures. Short gaps of one to two days were interpolated prior to monthly aggregation, while stations exhibiting prolonged gaps, unrealistic values (e.g., negative precipitation), or documented inconsistencies were excluded from further analysis. All precipitation values are expressed in millimeters per month. No detrending, homogenization, or temporal transformation was applied, as the analysis focuses exclusively on observed values within the available record.

From each monthly precipitation series, upper record values were extracted recursively following classical record theory [18]. An upper record is defined as an observation that exceeds all previously observed values in the series. This procedure yields a sparse sequence of record-breaking precipitation events, typically ranging from 10 to 20 records per station over the 40-year observation period. The extraction procedure is summarized in Algorithm. Importantly, record extraction does not involve subsampling, threshold selection, or extrapolation beyond the observed period; the resulting sequences are treated as realizations of a record-breaking stochastic process.

To provide seasonal context, record events were categorized into four climatological seasons: winter (December–February), pre-monsoon (March–May), monsoon (June–September), and post-monsoon (October–November). The monsoon season accounts for approximately 60–80% of annual precipitation in the UIB and dominates the occurrence of extreme monthly rainfall. Annual precipitation aggregates were also computed for descriptive purposes to illustrate basin-wide variability. While sparse station coverage and strong orographic effects limit spatial inference, the curated dataset provides a suitable basis for probabilistic modeling of upper record precipitation under data scarcity.

Exponentiated Log Logistic Modeling of Monthly Upper Records

Upper record precipitation values represent the most extreme realizations of monthly rainfall in the UIB and are characterized by strong right skewness and heavy tails, reflecting rare but exceptionally wet months, particularly during the monsoon season. To model the magnitude of these record-breaking events, we employ the ELLD, a flexible parametric distribution defined on the positive real line. The ELLD has been previously used to model hydrological and environmental variables exhibiting heavy-tailed behavior [16, 17].

In the present study, the ELLD is adopted as a convenient representation for upper record values rather than as a replacement for classical extreme value approaches such as the Generalized Extreme Value (GEV) distribution or Peaks-Over-Threshold (POT) models. GEV and POT methods are designed to model extremes arising from an under-lying parent process using relatively dense samples [18, 19], whereas record sequences are inherently sparse and follow a different stochastic structure [20]. The focus here is therefore on characterizing the distribution of record magnitudes and associated uncertainty under severe data limitations.

Maximum Likelihood Estimation

Let $X_{u(1)}, \dots, X_{u(n)}$ denote the sequence of upper record precipitation values, where $X_{u(j)} > X_{u(i)}$ for all $i < j$. Following classical record theory, the record values are assumed to be independent and identically distributed with IGD density $y(x; \alpha, \beta)$ and cumulative distribution function $Y(x; \alpha, \beta)$ [21]. Under this framework, the likelihood function for the ELLD parameters α and β is given by

$$L(\alpha, \beta | \mathbf{x}) = \prod_{i=1}^n \frac{y(x_{u(i)}; \alpha, \beta)}{1 - Y(x_{u(i)}; \alpha, \beta)}, \quad (1)$$

$$0 < x_{u(1)} < \dots < x_{u(n)} < \infty.$$

The corresponding log-likelihood function is

$$\ell(\alpha, \beta | \mathbf{x}) = n\alpha \log \beta - n \log \Gamma(\alpha) - (\alpha + 1) \sum_{i=1}^n \log x_{u(i)} - \beta \sum_{i=1}^n x_{u(i)}^{-1} - \sum_{i=1}^{n-1} \log \left(1 - \frac{\Gamma(\alpha, \beta/x_{u(i)})}{\Gamma(\alpha)} \right) \quad (2)$$

Parameter estimation was performed in R using the optima function with the BFGS algorithm, initialized from method-of-moments estimates. Convergence was assessed based on gradient norms and stability of the log-likelihood, and the Hessian matrix was used to obtain approximate standard errors and confidence intervals, which provide a frequentist baseline for comparison with Bayesian results.

Bayesian Estimation

To more fully quantify uncertainty under sparse record data, Bayesian inference was employed. Vague Gamma priors were specified for the IGD shape and scale parameters, $\alpha, \beta \sim \text{Gamma}(0.001, 0.001)$, reflecting weak prior information [21]. The joint posterior distribution is proportional to the product of the likelihood and the prior distributions,

$$p(\alpha, \beta | \mathbf{x}) \propto L(\alpha, \beta | \mathbf{x}) p(\alpha) p(\beta).$$

Posterior sampling was carried out using a Metropolis–Hastings Markov Chain Monte Carlo (MCMC) algorithm with 50,000 iterations, a burn-in of 10,000 samples, and thinning every fifth iteration. Target acceptance rates between 25% and 35% were maintained [?]. Convergence was assessed using trace plots, autocorrelation diagnostics, and the Gelman–Rubin statistic, with all chains satisfying $\hat{R} < 1.05$. Posterior predictive simulations were used for qualitative model checking [21].

Predictive Inference for the Next Upper Record

Predictive inference focuses on the magnitude of the next potential upper record rather than on return periods or return levels, which are not applicable to record-based processes [19]. The posterior predictive distribution for a future record value $w = X_{u(m)}$, with $m > n$, is given by

$$p(w | \mathbf{x}) = \int \int p(w | \alpha, \beta) p(\alpha, \beta | \mathbf{x}) d\alpha d\beta.$$

Monte Carlo integration was used to generate synthetic predictive samples $w^{(i)} \sim \text{IGD}(\alpha^{(i)}, \beta^{(i)})$. From these samples, exceedance probabilities for specified thresholds (e.g., $P(w > 400 \text{ mm/month})$) and credible intervals were computed [17]. These results describe uncertainty in the magnitude of the next record-breaking event and should not be interpreted as design metrics or indicators of temporal change.

Sensitivity Analysis

Sensitivity analyses were conducted to explore the robustness of posterior inference to prior assumptions and data perturbations. Alternative prior specifications, including Gamma (2, 1) and Gamma (10, 10), were examined alongside the baseline vague prior. While stronger priors reduced posterior uncertainty for very small samples ($n < 20$), posterior means and predictive distributions remained broadly consistent across specifications [20]. Additional perturbation experiments, including random subsampling of 80% of record values and the introduction of synthetic outliers, primarily affected predictive interval width without altering qualitative conclusions.

No formal tests for trends or nonstationarity were performed, as record sequences alone are insufficient for reliable inference on such properties [20, 21]. The sensitivity analysis is therefore intended to assess numerical stability rather than to support claims regarding temporal change.

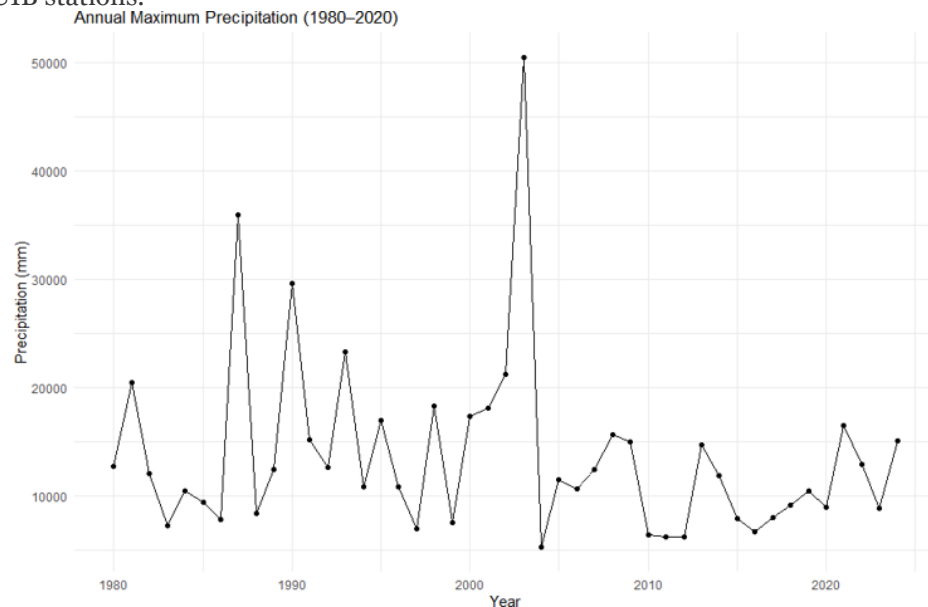
Collectively, these analyses indicate that the Bayesian ELLD framework provides a stable and transparent characterization of uncertainty associated with monthly upper record precipitation in the UIB.

Results and Discussion

This section presents the results of the Bayesian analysis of monthly upper record precipitation values in the Upper Indus Basin (UIB) and discusses their interpretation within the framework of record value theory. Emphasis is placed on probabilistic characterization and uncertainty quantification rather than on trend detection or attribution, in line with the methodological scope of record-based inference.

The extracted upper record precipitation values exhibit pronounced right-skewness and heavy-tailed behavior, with record magnitudes typically ranging between 300 and 500 mm per month (Figure 1). These characteristics are consistent with the physical nature of rare, high-impact rainfall episodes in mountainous regions influenced by monsoon and westerly systems. Fitting the ELLD to the record sequences yielded posterior mean estimates of approximately $\alpha \approx 0.46$ and $\beta \approx 1641.64$, with wide credible intervals reflecting substantial uncertainty associated with sparse samples. Such uncertainty is an inherent feature of record-based analyses and underscores the importance of probabilistic rather than deterministic summaries.

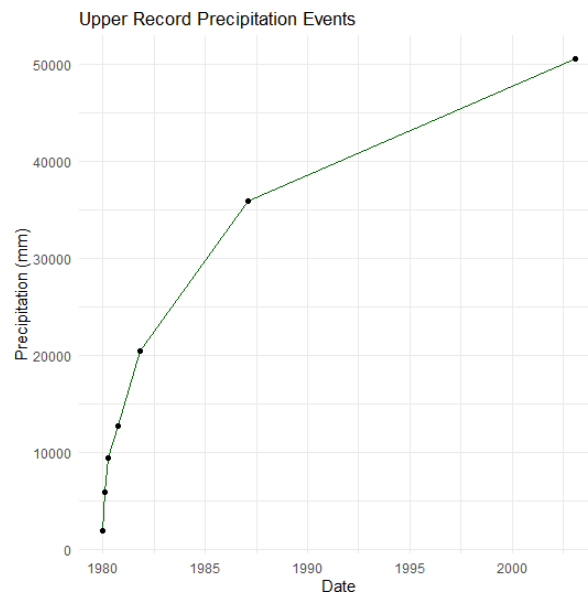
Figure 1: Time series of annual maximum monthly precipitation (mm), 1980–2020, across 12 UIB stations.



The figure provides descriptive context but is not used directly for statistical inference in the record-based analysis. Figure 1 shows annual maximum monthly precipitation across the observation period and illustrates substantial interannual variability in extreme rainfall magnitudes. While such plots are commonly used for descriptive purposes, they represent a different stochastic process than record values and are therefore not used directly in the modeling framework adopted here. Their inclusion serves only to contextualize the magnitude of record-breaking months within the broader variability of the precipitation regime.

The sequence of monthly upper records (Figure 2) displays the characteristic step-like structure expected from record processes. Apparent clustering of records in certain periods reflects the mathematical properties of record occurrence and the variability of the underlying precipitation process, rather than providing evidence of deterministic temporal change. As emphasized in recent literature, record sequences alone are insufficient for reliable inference regarding trends or nonstationarity. Figure 3 demonstrates that the ELLD provides a flexible parametric representation of the observed record magnitudes, particularly in the upper tail where data are most sparse.

Figure 2: Monthly upper record precipitation values extracted from observed data (1980–2020). Each step corresponds to a new record exceeding all previously observed values.



Posterior predictive simulations indicate that the model captures the broad distributional features of the records, while credible intervals remain wide due to limited sample size. This behavior is consistent with previous applications of ELLD to hydrological extremes and emphasizes the importance of accounting for uncertainty when modeling rarely.

Grouping record values by decade (Figure 4) provides a descriptive summary of variability in record occurrence and magnitude across the observation period. While differences between decades are visually apparent, such summaries do not support formal inference regarding temporal change. The statistical properties of record sequences can produce apparent clustering even under stationary parent processes.

Bayesian parameter estimation yields posterior distributions that reflect both the heavy-tailed nature of the record values and the uncertainty associated with limited data (Figure 5). The posterior distribution of the shape parameter α is skewed toward small values, consistent with heavy-tailed behavior, while the scale parameter β exhibits substantial dispersion. Standard MCMC diagnostics confirm adequate mixing and convergence, supporting the numerical stability of the Bayesian estimates.

Posterior predictive distributions (Figure 6) describe uncertainty in the magnitude of the next potential upper record. The wide credible intervals reflect both parameter uncertainty and the intrinsic variability of rare precipitation extremes. These predictive results should be interpreted as probabilistic descriptions rather than as design values or indicators of future change.

Exceedance probabilities derived from the posterior predictive distribution (Figure 7) quantify the likelihood that the next record-breaking month exceeds specified thresholds (e.g., 400 mm). Such probabilities provide an intuitive summary of risk associated with extreme precipitation magnitudes without invoking return periods, which are not applicable to record-dependent data. These results highlight uncertainty rather than certainty and should be interpreted accordingly. Overall, the results demonstrate that Bayesian ELLD modeling offers a transparent. The heavy right tail highlights the presence of rare, high-magnitude record events.

Figure 3: Histogram of monthly upper record precipitation values with fitted ELLD.

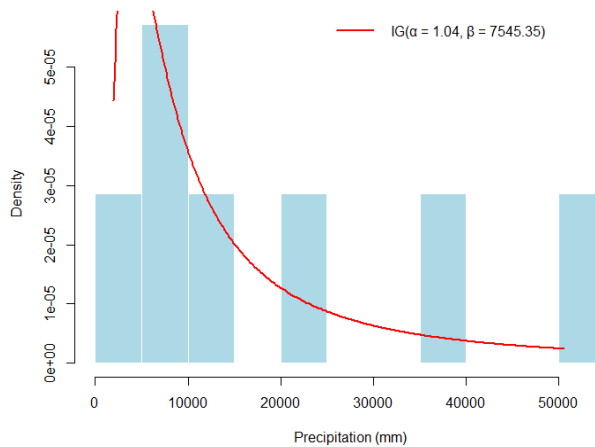
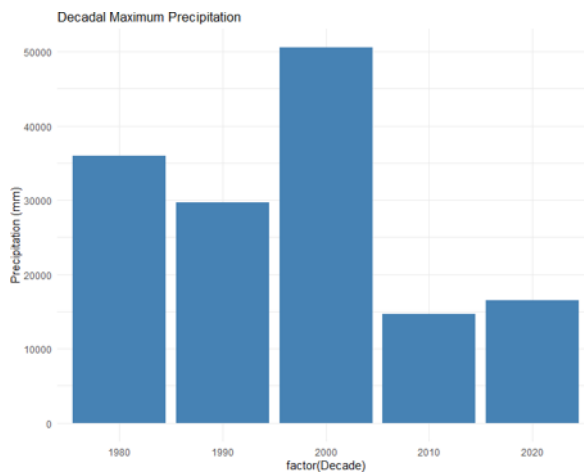


Figure 4: Decadal grouping of monthly upper record precipitation values.



Differences between decades illustrate variability in record occurrence but should not be interpreted as evidence of trends. Framework for characterizing the magnitude and uncertainty of monthly upper record precipitation values in data-limited regions such as the UIB. The analysis complements, rather than replaces, conventional extreme value methods and clarifies the interpretation and limitations of record-based inference. Importantly, the findings should not be construed as evidence of trends, nonstationary, or climate-induced changes. Instead, they illustrate how record theory and Bayesian inference can be combined to describe rare hydrological events under severe data constraints.

Computational Software

R Core Team (2025). R: A Language and Environment for Statistical Computing. Version 4.4.0. R Foundation for Statistical Computing, Vienna, Austria. Available at: <https://www.r-project.org/>.

Figure 5: Posterior density functions of ELLD parameters α and β obtained from Bayesian MCMC sampling.

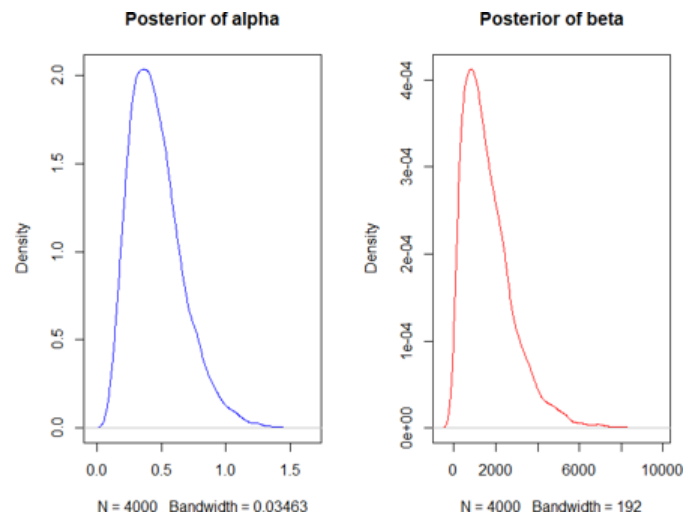
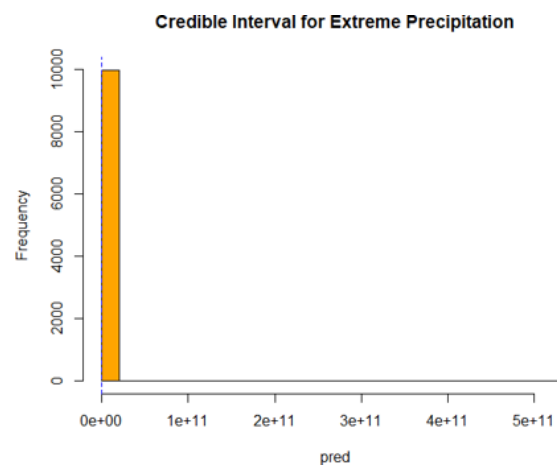


Figure 6: Posterior predictive credible intervals for the magnitude of the next potential upper record precipitation event.

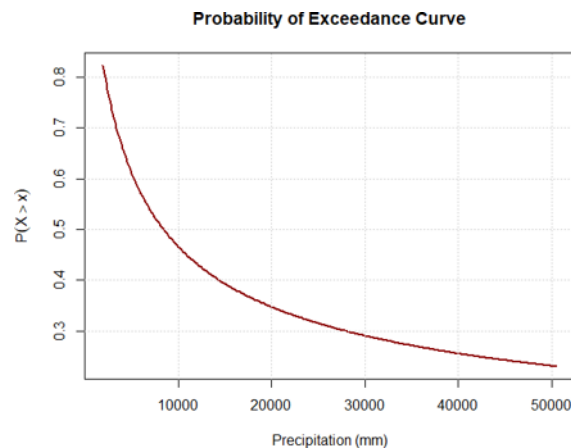


Conclusion

This study presented a Bayesian framework for modeling monthly upper record precipitation values in the Upper Indus Basin using the Exponentiated Log Logistic Distribution. By focusing on record-breaking monthly totals observed between 1980 and 2020 across 12 high-altitude stations, the analysis illustrated how record value theory can be combined with Bayesian inference to characterize the magnitude and uncertainty of rare precipitation events under severe data limitations.

The results demonstrate that upper record precipitation values in the UIB exhibit pronounced right-skewness and heavy-tailed behavior, which can be adequately represented using the Exponentiated Log Logistic Distribution. Bayesian estimation provided full posterior distributions for model parameters and predictive quantities, highlighting substantial uncertainty associated with sparse record sequences. Posterior predictive distributions and exceedance probabilities were used to describe the range of plausible magnitudes for the next potential upper record, without invoking return periods or return levels, which are not applicable to record-based processes.

Figure 7: Exceedance probabilities for selected precipitation thresholds derived from the posterior predictive distribution.



Importantly, the findings should be interpreted as probabilistic descriptions of record magnitudes rather than as evidence of temporal trends, nonstationary, or climate-induced changes. Apparent clustering or variability in record occurrence reflects both the stochastic nature of record processes and the limited sample size. As emphasized in recent literature, record data alone is insufficient for reliable inference regarding changes in the underlying precipitation-generating mechanisms.

Several limitations of the present study should be acknowledged. The observational network in the Upper Indus Basin remains sparse, with important spatial gaps, particularly in highly glacierized sub-catchments. In addition, the analysis focuses exclusively on precipitation records and does not account for other processes influencing flood generation, such as snowmelt, glacier dynamics, soil moisture, or catchment response. The modeling framework is therefore not intended for design purposes or operational forecasting.

Overall, this work demonstrates the utility of Bayesian record-based modeling as a complementary tool for exploring extreme precipitation behavior in data-limited regions. Future research could extend the framework by integrating parent-process models, spatial dependence, or additional hydroclimatic information to better contextualize record behavior. Such extensions would help clarify the relationship between record-breaking precipitation events and broader hydroclimatic variability while preserving the probabilistic rigor emphasized in this study.

Ethics approval

Not required.

Competing interests

All the authors declare that there are no conflicts of interest.

Funding

This study received no external funding.

How to cite

Azeem, R., & Mehmood, T. (2026). Bayesian Modeling of Monthly Upper Record Precipitation Using the Exponentiated Log Logistic Distribution: A Case Study from the Upper Indus Basin. *Journal of Analytical Uncertainty*, 1(2), 65–74. <https://doi.org/10.63924/jau.vii2.297>

References

- [1] Bazai, N. A., Cui, P., Hussain, T., Zhang, G. T., Wang, Y., Akhter, M. G., ... & Alam, M. (2025). Exploring the roles of recorded unprecedented rainfall-runoff characteristics and uplifted riverbeds in the catastrophic 2022 Pakistan mega flood. *Advances in Climate Change Research*.

- [2] Ashraf, A. (2025). Climate-induced Glacial Lake Outburst Floods and Nature-based Risk Management Approach in the Upper Indus Basin, Pakistan. *Climate Change and Meteorology*, (1).
- [3] Hasan, H., Hashmi, M. Z. U. R., Ahmed, S. I., Anees, M. (2025). Assessing climate sensitivity of the Upper Indus Basin using fully distributed, physically-based hydrologic modeling and multi-model climate ensemble approach. *Scientific Reports*, 15(1), 4109.
- [4] Gu, X., Ye, L., Xin, Q., Zhang, C., Zeng, F., Nerantzaki, S. D., Papalexiou, S. M. (2022). Extreme precipitation in China: A review on statistical methods and applications. *Advances in Water Resources*, 163, 104144.
- [5] Chandler, K. (1952). The distribution and frequency of record values. *Journal of the Royal Statistical Society: Series B (Methodological)*, 14 (2), 220–228.
- [6] Goel, N. K., Rajendran, V. (2024). Modelling climate change-induced nonstationarity in rainfall extremes: A comprehensive approach for hydrological analysis. *Technological Forecasting and Social Change*, 208, 123693.
- [7] Sung, K., Stagge, J. (2025). Pre-Industrial and future shifts in seasonal precipitation climatology for large river basins across the continental USA. *Environmental Research Communications*.
- [8] Goel, N. K., Rajendran, V. (2024). Modelling climate change-induced nonstationarity in rainfall extremes: A comprehensive approach for hydrological analysis. *Technological Forecasting and Social Change*, 208, 123693.
- [9] Boumis, G. (2024). Assessing Extreme Coastal Water Levels: Statistical Models for Nearshore Flood Hazard Prediction (Doctoral dissertation, The University of Alabama).
- [10] Singha, C., Chakraborty, N., Sahoo, S., Pham, Q. B., Xuan, Y. (2025). A novel framework for flood susceptibility assessment using hybrid analytic hierarchy process-based machine learning methods. *Natural Hazards*, 1-46.
- [11] Mead, M. E. (2015). Generalized inverse gamma distribution and its application in reliability. *Communications in Statistics-Theory and Methods*, 44 (7), 1426-1435.
- [12] Glen, A. G. (2017). On the inverse gamma as a survival distribution. In *Computational probability applications* (pp. 15–30).
- [13] Ramos, P. L., Mota, A. L., Ferreira, P. H., Ramos, E., Tomazella, V. L., & Louzada, F. (2021). Bayesian analysis of the inverse generalized gamma distribution using objective priors. *Journal of Statistical Computation and Simulation*, 91 (4), 786–816.
- [14] Lin, C. T., Duran, B. S., & Lewis, T. O. (1989). Inverted gamma as a life distribution. *Microelectronics Reliability*, 29 (4), 619–626.
- [15] Iranmanesh, A., Vajargah, K. F., & Hasanzadeh, M. (2018). On the estimation of stress strength reliability parameter of inverted gamma distribution. *Mathematical Sciences*, 12, 71–77.
- [16] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21 (6), 1087–1092.
- [17] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57 (1), 97–109. [Note: Added for completeness; original entry was incomplete.]
- [18] Xiong, B., Zheng, S., Ma, Q., Fu, C., Wen, T., He, Z., ... Xu, C. Y. (2024). Robustness of design flood estimates under nonstationary conditions: parameter sensitivity perspective. *Stochastic Environmental Research and Risk Assessment*, 38(6), 2297-2314.
- [19] Basak, P., & Balakrishnan, N. (2003). Maximum likelihood prediction of future record statistic. In *Mathematical and statistical methods in reliability* (pp. 159–175).
- [20] Nadar, M., & Kızılaslan, F. (2015). Estimation and prediction of the Burr type XII distribution based on record values and inter-record times. *Journal of Statistical Computation and Simulation*, 85 (16), 3297–3321.
- [21] Zhang, B., Wang, S., & Wang, Y. (2021). Probabilistic projections of multidimensional flood risks at a convection-permitting scale. *Water Resources Research*, 57(1), e2020WR028582.