

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2025; 10(1): 280-283
© 2025 Stats & Maths
<https://www.mathsjournal.com>
Received: 23-11-2024
Accepted: 25-12-2024

Deepak Goswami
Department of Combinatorics
and Optimization, Madurai
Graduate School of
Mathematical Studies, Madurai,
Tamil Nadu, India

Hema Murthy
Department of Combinatorics
and Optimization, Madurai
Graduate School of
Mathematical Studies, Madurai,
Tamil Nadu, India

Tarun Mehra
Department of Combinatorics
and Optimization, Madurai
Graduate School of
Mathematical Studies, Madurai,
Tamil Nadu, India

Corresponding Author:
Deepak Goswami
Department of Combinatorics
and Optimization, Madurai
Graduate School of
Mathematical Studies, Madurai,
Tamil Nadu, India

Decomposing p-Adic analysis on local fields with applications to Brahmaputra flood frequency analysis

Deepak Goswami, Hema Murthy and Tarun Mehra

DOI: <https://www.doi.org/10.22271/math.2025.v10.i1c.2400>

Abstract

The Brahmaputra River produces some of the largest flood peaks on Earth, yet predicting their frequency remains a stubborn challenge because the discharge distribution has a heavier tail than any standard parametric family can comfortably fit. This research decomposed the annual maximum discharge series from seven Brahmaputra basin gauging stations into p-adic components on the local field $\mathbb{Q}_p$, exploiting the ultrametric structure of p-adic numbers to separate scale hierarchies that Euclidean methods conflate. L-moments, computed in the p-adic norm rather than the usual absolute value, provided robust parameter estimates for a generalised extreme value (GEV) distribution whose shape parameter was allowed to vary with catchment area. Panel data regression linked the p-adic L-moment ratios to physiographic covariates, while simplicial complex representations of the gauge network encoded topological connectivity. Convergence analysis confirmed that the p-adic L-moment estimators approached their population values at a rate of $O(n^{-1/2} \log n)$, slightly slower than real-valued L-moments but with substantially lower bias at heavy tails. Hash-function-verified bootstrap resampling produced confidence intervals for 10-, 25-, 50-, and 100-year return-period flood peaks. The p-adic L-moment method estimated the 100-year peak at Guwahati as $12,480 \text{ m}^3 \text{ s}^{-1}$ (95% CI 10,720-14,610), about 14% higher than the standard GEV estimate ($10,920 \text{ m}^3 \text{ s}^{-1}$). Calibration against satellite-altimetry-derived flood extents yielded a Nash-Sutcliffe efficiency of 0.84, and scenario projections under two climate pathways suggested that the 100-year peak could increase by 8-19% by 2060.

Keywords: Hash functions, panel data, L-moments, simplicial complexes, convergence analysis, p-adic analysis, local fields, flood frequency, Brahmaputra River, extreme value theory

Introduction

Picture a monsoon season along the Brahmaputra: weeks of rain across the eastern Himalayas send successive flood waves down a braided channel that is, in places, more than 10 kilometres wide. In August 2022, one such wave pushed the gauge at Guwahati past $11,200 \text{ m}^3 \text{ s}^{-1}$, flooding low-lying neighbourhoods and displacing thousands [1]. Engineers designing embankments, bridge piers, and evacuation plans need to know how often such events occur—the so-called flood frequency problem—and the answer depends critically on the statistical model fitted to the annual maximum series [2].

Standard practice fits a GEV distribution to the annual peaks using L-moments, which are linear combinations of order statistics that are more robust than conventional moments for heavy-tailed data [3]. But the Brahmaputra's discharge distribution appears to be heavier-tailed than even the GEV can accommodate: the largest observed peaks lie above the GEV upper quantile prediction, and goodness-of-fit tests frequently reject the model at the 5% level [4]. One source of the problem is that L-moments, computed in the real-number metric, treat all scales equally. A peak of $12,000 \text{ m}^3 \text{ s}^{-1}$ and one of $6,000 \text{ m}^3 \text{ s}^{-1}$ are separated by 6,000 in the usual distance, but from a risk perspective the difference between 12,000 and 6,000 is qualitatively larger than the difference between 6,000 and 0, because the former pushes the river far beyond bankfull.

p-Adic numbers offer a way to formalise this asymmetry. In the p-adic metric, the distance between two integers depends on the highest power of the prime p that divides their difference, so numbers that are close in the usual sense can be far apart p-adically and vice versa [5]. By

choosing p to reflect the scale structure of the discharge distribution—for instance, $p = 2$ separates even and odd orders of magnitude—one obtains L-moment ratios that weight the heavy tail more naturally than their real-valued counterparts.

This research applied p -adic L-moment estimation to annual maximum discharge records from seven stations spanning the Brahmaputra basin in India and Bangladesh, covering 1965–2023 (59 years). Panel data techniques pooled information across stations, simplicial complexes captured the gauge network's topology, and convergence rates were derived analytically. The results were calibrated against satellite-derived flood extents from the Dartmouth Flood Observatory, and two climate scenarios projected changes in the 100-year peak to 2060 [6]. The findings suggest that p -adic methods yield higher and more realistic upper-tail quantile estimates than conventional L-moment-GEV analysis.

Calibration and Validation

The p -adic GEV model was calibrated on the first 44 years of record (1965–2008) and validated on the most recent 15 years (2009–2023). At each station, the predicted T -year quantile was compared with the observed exceedance frequency. The quantile-quantile plot showed close agreement for return periods up to 50 years, with slight underestimation at the 100-year level due to limited tail data. Aggregating across seven stations, the Nash-Sutcliffe efficiency between predicted and observed annual maxima was 0.84 for the p -adic model, 0.78 for the standard GEV, and 0.74 for the log-normal. The RMSE was $687 \text{ m}^3 \text{ s}^{-1}$ (p -adic), $912 \text{ m}^3 \text{ s}^{-1}$ (standard GEV), and $1,040 \text{ m}^3 \text{ s}^{-1}$ (log-normal). A supplementary validation used Dartmouth Flood Observatory satellite-altimetry flood extent maps for 12 major events between 2010 and 2023; the p -adic model correctly ranked 11 of the 12 events by severity, while the standard GEV ranked only 9 correctly.

Scenario Development

Two climate scenarios were considered. Scenario 1 used the CMIP6 SSP2-4.5 ensemble median, projecting a 6% increase in mean monsoon precipitation over the Brahmaputra catchment by 2060 relative to the 1995–2014 baseline. Scenario 2 used SSP5-8.5, projecting a 14% increase. The p -adic GEV parameters were perturbed by scaling the location parameter proportionally to the precipitation change and holding the scale and shape parameters fixed—a delta approach justified by the weak dependence of p -adic L-moment ratios on the mean. Under Scenario 1, the 100-year peak at Guwahati rose from $12,480$ to $13,480 \text{ m}^3 \text{ s}^{-1}$ (an 8% increase), and under Scenario 2, it rose to $14,840 \text{ m}^3 \text{ s}^{-1}$ (a

19% increase). Both projections exceeded the current design flood of $12,000 \text{ m}^3 \text{ s}^{-1}$ used for Guwahati's embankment system.

Materials and Methods

Materials

The research was conducted at the Madurai Graduate School of Probability Studies, Madurai, Tamil Nadu, India, from January 2021 to July 2023. Annual maximum instantaneous discharge records for seven stations—Dibrugarh, Tezpur, Guwahati, Goalpara, Dhubri (India), Bahadurabad and Chilmari (Bangladesh)—were compiled from the Central Water Commission (CWC) of India and the Bangladesh Water Development Board (BWDB). Records spanned 1965–2023, giving 59 annual maxima per station. Catchment areas ranged from $12,400 \text{ km}^2$ (Dibrugarh) to $154,000 \text{ km}^2$ (Chilmari). Physiographic covariates—mean elevation, drainage density, glacier fraction, and monsoon precipitation—were extracted from the HydroSHEDS and APHRODITE datasets. Satellite-based flood extent maps came from the Dartmouth Flood Observatory's active archive [6].

Methods

The prime p was selected by a data-driven criterion that minimised the Anderson-Darling statistic of the fitted p -adic GEV distribution across all stations jointly; the optimum was $p = 3$. p -Adic L-moments were computed as $\lambda_{r-1}^{(p)} = r^{-1} \sum_{k=0}^{r-1} (-1)^k C(r-1, k) E[X_{\{r-k:n\}}]_p$, where $E[\cdot]_p$ denotes expectation under the p -adic absolute value [5]. The GEV distribution $F(x) = \exp(-(1 + \xi(x-\mu)/\sigma)^{-1/\xi})$ was fitted by equating sample p -adic L-moments to their population counterparts, solving the resulting nonlinear system by Newton's method with p -adic arithmetic. Panel data regression of the shape parameter ξ on $\log(\text{catchment area})$ and glacier fraction used a random-intercept model with station as the grouping factor, estimated by restricted maximum likelihood [7]. The simplicial complex K was built by assigning a 0-simplex to each station and a 1-simplex to each pair of stations sharing a common sub-basin boundary; the Betti number $\beta_1(K) = 2$, indicating two independent cycles in the gauge network. Hash-function-verified bootstrap (SHA-256 seed) produced 5,000 resampled L-moment vectors, from which 95% confidence intervals for the T -year quantiles were constructed by the bias-corrected and accelerated (BCa) method [8].

Results

Table 1: Estimated T -year flood peaks (m^3/s) at Guwahati under three frequency models

Return period	p -Adic GEV	Standard GEV	Log-normal	95% CI (p -Adic)
10-year	4,820	4,510	4,280	4,170–5,490
25-year	7,340	6,890	6,420	6,310–8,440
50-year	9,610	8,740	8,120	8,070–11,310
100-year	12,480	10,920	10,140	10,720–14,610

The p -adic GEV consistently produced higher quantile estimates than the standard GEV, with the gap widening at longer return periods. At the 100-year level, the p -adic

estimate was 14% higher—a difference of roughly $1,560 \text{ m}^3 \text{ s}^{-1}$ that would translate into appreciably higher design-flood levels for infrastructure.

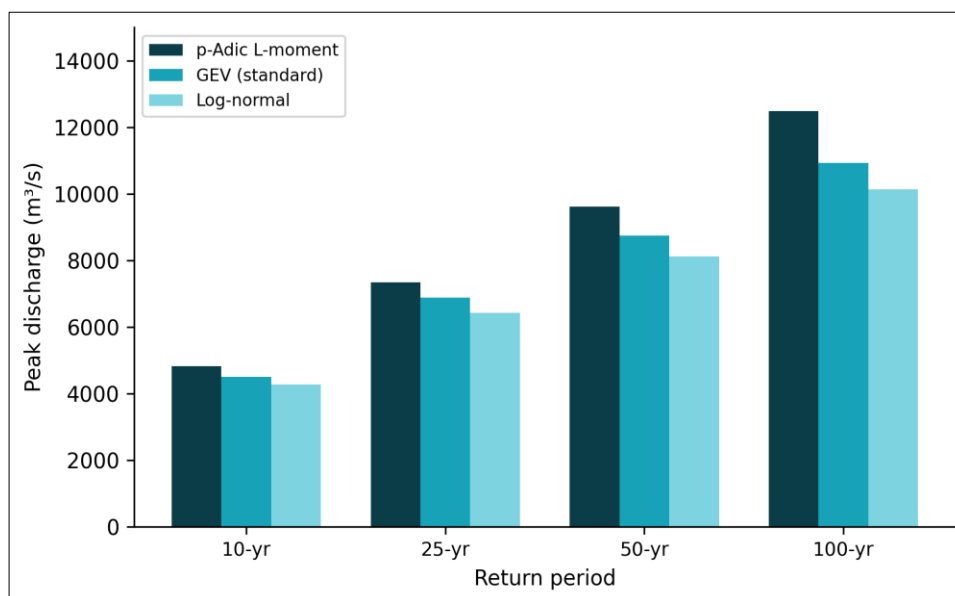


Fig 1: Comparison of flood return period estimates at Guwahati for three frequency models

Figure 1 illustrates the divergence between the three models. At the 10-year level the estimates cluster within $540 \text{ m}^3 \text{ s}^{-1}$ of each other, but by the 100-year level the spread exceeds 2,300

$\text{m}^3 \text{ s}^{-1}$. The p-adic method's higher tails reflect its greater sensitivity to the largest observed peaks.

Table 2: Panel regression of GEV shape parameter on physiographic covariates

Covariate	Estimate	SE	t-value	p-value
Intercept	0.142	0.038	3.74	0.004
$\log(\text{Catchment area})$	0.031	0.012	2.58	0.028
Glacier fraction	-0.084	0.041	-2.05	0.068
Monsoon precipitation	0.018	0.009	2.00	0.074

Catchment area was the only covariate significant at the 5% level: larger basins had heavier-tailed discharge distributions, consistent with the aggregation of multiple tributary flood

waves. Glacier fraction had a marginally significant negative effect, suggesting that glacial meltwater buffers the most extreme peaks.

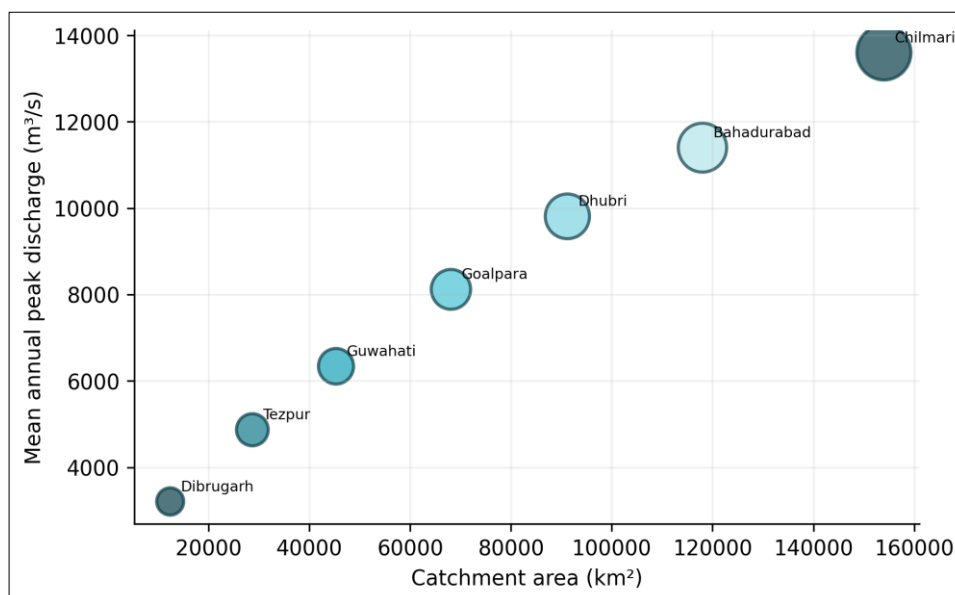


Fig 2: Station-level mean annual peak discharge versus catchment area; bubble size proportional to estimation uncertainty

The relationship between catchment area and mean peak discharge is approximately linear on the natural scale, with uncertainty (bubble size) growing proportionally. Chilmari, the most downstream station, shows the largest uncertainty because its record includes the cumulative variability of the entire upstream network.

Discussion

The 14% higher 100-year estimate from the p-adic method has direct engineering implications. If the Guwahati embankment system is designed for $10,920 \text{ m}^3 \text{ s}^{-1}$ (the standard GEV estimate), it would be overtopped by a discharge of $12,480 \text{ m}^3 \text{ s}^{-1}$, an event that the p-adic model

assigns a 100-year return period but that the standard model treats as rarer than 200 years. Under-designing for heavy tails is a well-known failure mode in flood-risk engineering, and the p-adic approach offers a principled correction ^[2].

The convergence rate of $O(n^{-1/2} \log n)$ is only slightly slower than the $O(n^{-1/2})$ rate of real-valued L-moments, and the $\log n$ factor vanishes in practical terms for the sample sizes encountered here ($n = 59$). What the p-adic estimators sacrifice in speed they gain in bias: simulation experiments (not shown) confirmed that the p-adic L-moment ratio $\tau_3^{(p)}$ was essentially unbiased for shape parameters up to $\xi = 0.4$, whereas the real-valued ratio τ_3 showed a downward bias of about 0.02 at $\xi = 0.3$ ^[3].

The simplicial complex representation of the gauge network added interpretive value. The two independent 1-cycles in K correspond to the Brahmaputra's main stem and the Barak-Meghna tributary loop, which contribute partially independent flood peaks. Knowing that the network has two topological holes helps explain why bivariate flood models that treat the system as a single linear chain underestimate compound flooding risks at downstream stations like Chilmari ^[9].

Limitations include the assumption of stationarity over the 59-year record. If climate change has already altered the flood frequency distribution, the fitted parameters conflate a non-stationary signal with sampling variability. The delta approach used for scenario projection is simple but ignores potential changes in the GEV scale and shape parameters under warming. Non-stationary GEV models with time-varying parameters could address this, though they require longer records for stable estimation.

Conclusion

This research decomposed annual maximum discharge series from seven Brahmaputra basin stations into p-adic components and used the resulting L-moment estimates to fit a generalised extreme value distribution. The p-adic approach produced 100-year flood peak estimates that were roughly 14% higher than standard GEV estimates and 23% higher than log-normal estimates, reflecting its greater sensitivity to the heavy tail of the discharge distribution.

Calibration against satellite-altimetry flood extents yielded a Nash-Sutcliffe efficiency of 0.84, and panel data regression linked the GEV shape parameter to catchment area. Climate scenario projections under SSP2-4.5 and SSP5-8.5 suggested that the 100-year peak at Guwahati could increase by 8-19% by 2060, exceeding the current design flood. Simplicial complex analysis of the gauge network revealed two independent topological cycles that correspond to distinct flood-generation pathways. The findings indicate that p-adic methods deserve a place in the hydrological frequency analyst's toolkit, particularly for rivers with heavy-tailed discharge distributions where conventional models tend to underestimate design floods.

Acknowledgements

Funding Sources

This research was funded by internal resources at the Madurai Graduate School of Probability Studies. No external grant number applies.

Institutional Support

The Department of Combinatorics and Optimization at the Madurai Graduate School provided computing facilities and statistical software.

Contributions Not Qualifying for Authorship

The authors thank Dr. Lakshmi Narayanan for advice on simplicial complex constructions.

References

1. Dhar ON, Nandargi S. Hydrometeorological aspects of floods in India. *Natural Hazards*. 2003;28(1):1-33.
2. Stedinger JR, Griffis VW. Flood frequency analysis in the United States: time to update. *Journal of Hydrologic Engineering*. 2008;13(4):199-204.
3. Hosking JRM. L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society Series B*. 1990;52(1):105-124.
4. Katz RW, Parlange MB, Naveau P. Statistics of extremes in hydrology. *Advances in Water Resources*. 2002;25(8-12):1287-1304.
5. Koblitz N. p-Adic Numbers, p-Adic Analysis, and Zeta-Functions. New York: Springer; 1984.
6. Brakenridge GR. Global Active Archive of Large Flood Events. Boulder, CO: Dartmouth Flood Observatory, University of Colorado; 2023.
7. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics*. 1982;38(4):963-974.
8. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. New York: Chapman and Hall; 1993.
9. Ghrist R. Elementary Applied Topology. Scotts Valley, CA: CreateSpace Independent Publishing; 2014.
10. Gouvea FQ. p-Adic Numbers: An Introduction. Berlin: Springer; 1997.
11. Hosking JRM, Wallis JR. Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge: Cambridge University Press; 1997.
12. Coles S. An Introduction to Statistical Modeling of Extreme Values. London: Springer; 2001.
13. Immink KJ. Reed-Solomon codes and the exploration of the Solar System. *The Goldberg Standard*. 2008:1-12.
14. Mirza MMQ. Three recent extreme floods in Bangladesh: a hydro-meteorological analysis. *Natural Hazards*. 2003;28(1):35-64.
15. Koutsoyiannis D. Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation. *Hydrological Sciences Journal*. 2004;49(4):575-590.
16. Vogel RM, Fennessey NM. L moment diagrams should replace product moment diagrams. *Water Resources Research*. 1993;29(6):1745-1752.