The New England Journal of Statistics in Data Science Volume 0, 1–6 (2026)
DOI: https://doi.org/0000

Supplementary Material to "Extreme Value Modeling with Generalized Pareto Distributions for Rounded Data"

Sai MA, Jun YAN*, and Xuebin ZHANG

1. ADDITIONAL SIMULATION STUDIES

This section provides supplementary simulation results to complement those presented in the main paper. The goal is to evaluate the robustness of the proposed MLE-IC method in estimation and the power of the Anderson–Darlin (AD) test under more extreme conditions.

1.1 Estimation of an extreme setting

Figure 1 compares the bias and variability of three estimators when the rounding interval is large ($\delta = 0.5$, $\sigma = 0.5$, $\xi = 0.2$). The proposed MLE-IC method shows negligible bias, while MLE-N (ignoring rounding) underestimates the scale and overestimates the shape parameter. The jittering approach exhibits biases in the opposite directions for two parameters. Their biases are clearly visible in comparison with the MLE-IC method. These results demonstrate that robust behavior MLE-IC in this extreme scenario.

The empirical and estimated standard errors of MLE-IC for σ are 0.0368 and 0.0367, and for ξ are 0.0572 and 0.0555, respectively. The close agreement between them validates the observed-information approximation, confirming its reliability for uncertainty quantification even in this challenging scenario.

1.2 Power comparison under an extreme setting

Figure 2 and Figure 3 extend the goodness-of-fit power comparison to lower-tail and higher-tail regimes. The HUGPD for Figure 2 has $\mu = 0.3$, $\sigma = 0.075$, $\xi = 0.1$, and $p = 0.8$, and the HUGPD for Figure 3 has $\mu = 0.3$, $\sigma = 1.2$, $\xi = 0.1$, and $p = 0.8$. Sample size for both plots is 250 and 500.

The AD and CvM tests consistently display the strongest sensitivity to deviations, while KS and CS tests show lower power. Doubling the sample size from 250 to 500 markedly increases power, confirming that the ranking of test strengths remains stable under more extreme scenarios. Note that when $p = 0.8$, $\delta = 0.1$, and the truncate point is $a = 0.3$, the AD, CvM and KS tests have lower power than 0.05, which is probably due to too many points mess at 0.3.

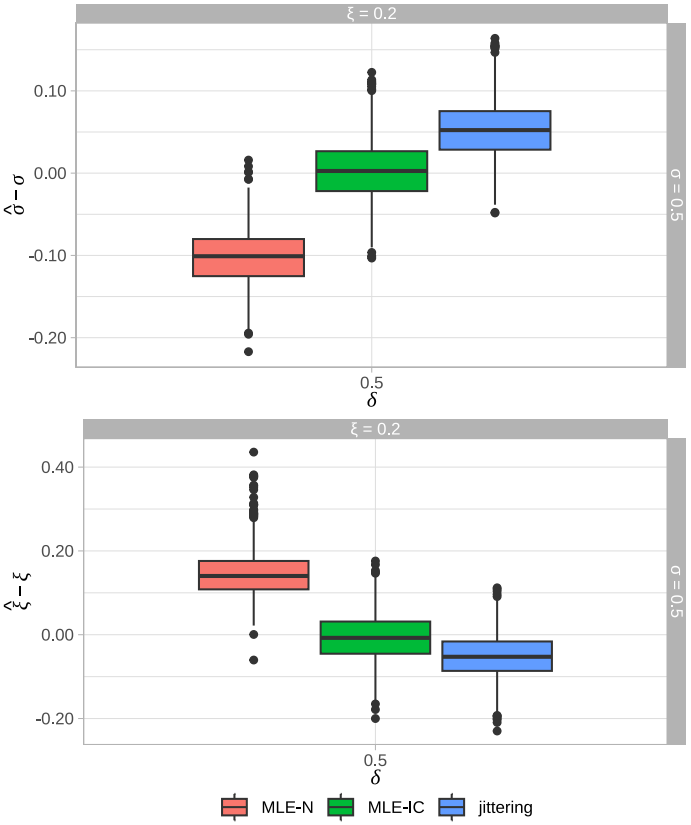

Figure 1: Boxplots of MLE-N, MLE-IC and jittering method. The upper panels show the boxplots of estimated scale parameter $\hat{\sigma}$ and the lower panels show the boxplots of estimated shape parameter $\hat{\xi}$.

*Corresponding author.

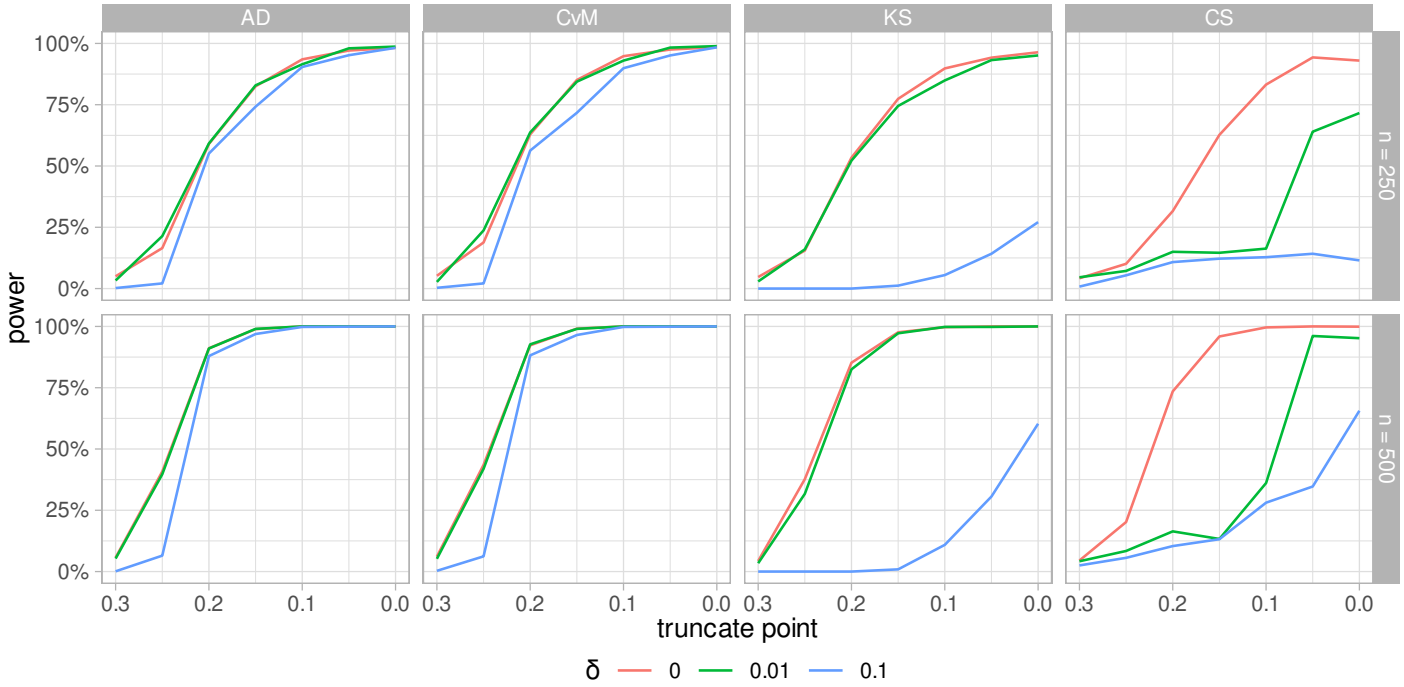

Figure 2: Power of $g(x|x > a)$ for four tests applied to the HUGPD, with $\mu = 0.3$, $\sigma = 0.075$, $\xi = 0.1$, and $p = 0.8$. Sample size is 250 and 500.

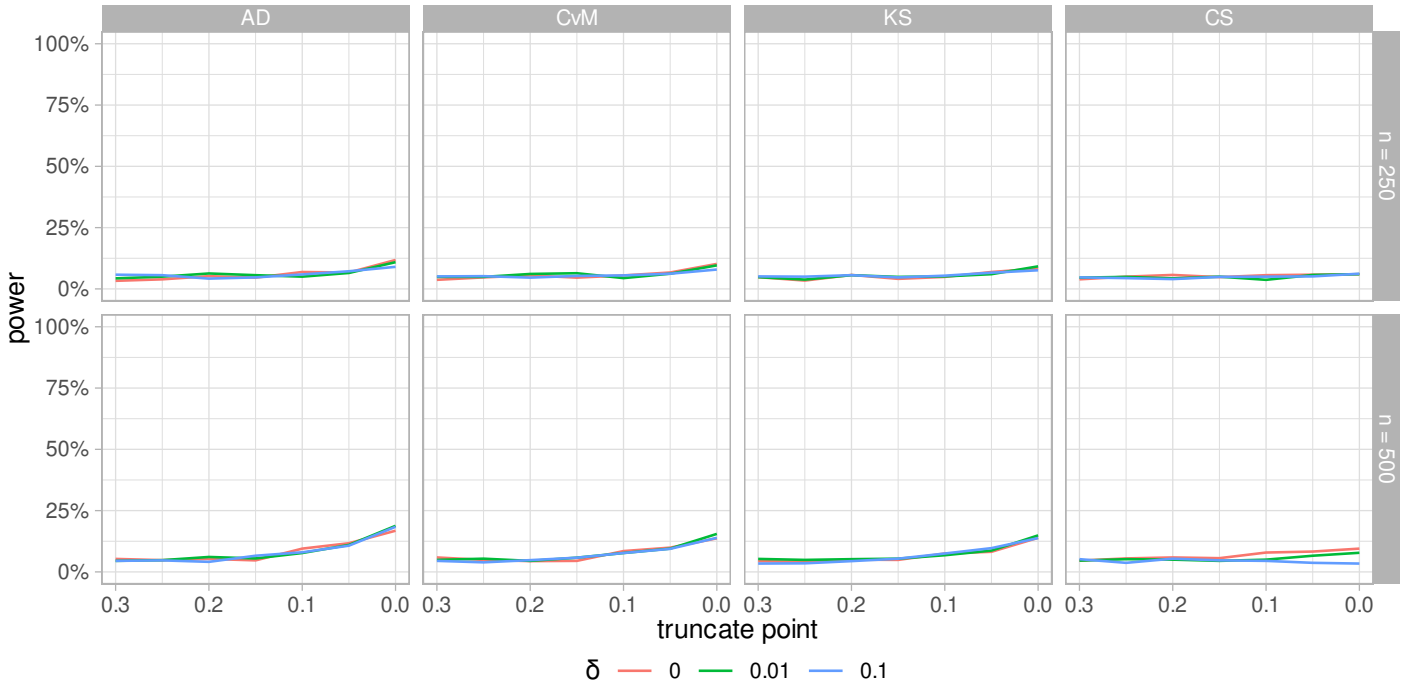

Figure 3: Power of $g(x|x > a)$ for four tests applied to the HUGPD, with $\mu = 0.3$, $\sigma = 1.2$, $\xi = 0.1$, and $p = 0.2$. Sample size is 250 and 500.

Table 1. The summary of 18 *monitoring stations from 1969 to 2018 in the eastern part of the Washington State.*

Station	MLE-N			MLE-IC		
	threshold	percentile (%)	number of exceedances	threshold	percentile (%)	number of exceedances
Chewelah	1.75	98	131	0.25	76	1596
Coulee Dam 1 SW						
Davenport						
Harrington						
Ice Harbor Dam				0.51	94	487
Lacrosse						
Mill Creek Dam				0.08	70	2190
Newport				1.30	94	411
Odessa						
Pomeroy				1.40	98	141
Pullman 2 NW				0.79	92	625
Republic						
Ritzville 1 SSE				0.79	96	310
Rosalia						
St. John						
Whitman Mission	1.12	98	143	0.66	94	425
Wilbur						
Spokane Intl AP				0.28	80	1515

2. ADDITIONAL DETAILS OF THE DATA ANALYSIS

This section provides additional results that complement the precipitation analysis in the main paper. We begin with a regional summary of threshold selection across all 18 monitoring stations and then focus on two representative sites—Chewelah and Ice Harbor Dam—where thresholds were identified by both MLE-N and MLE-IC. The detailed results for these two stations illustrate how modeling discretization affects threshold selection, model fit, and estimation of return levels.

Table 1 ummarizes the threshold selection results for 18 precipitation monitoring stations in eastern Washington from 1969 to 2018 under both MLE-N and MLE-IC estimation. The contrast between the two methods is striking. MLE-IC successfully identifies thresholds at 10 stations, typically corresponding to lower percentiles (ranging from about the 70th to the 94th) and producing substantially more exceedances. In contrast, MLE-N finds valid thresholds at only two stations, both at the very upper tail (98th percentile), leaving insufficient data for reliable tail estimation. The larger number of exceedances retained by MLE-IC enhances estimation precision and stabilizes regional extreme-value analysis, while the higher thresholds selected by MLE-N risk excessive variance and bias. The overall pattern demonstrates that accounting for data discretization is critical for automated threshold detection, enabling more robust inference of spatial patterns in precipitation extremes across the monitoring network.

Figure 4 displays the precipitation time series for the Chewelah and Ice Harbor Dam stations, with the thresholds selected by MLE-N and MLE-IC overlaid as horizontal

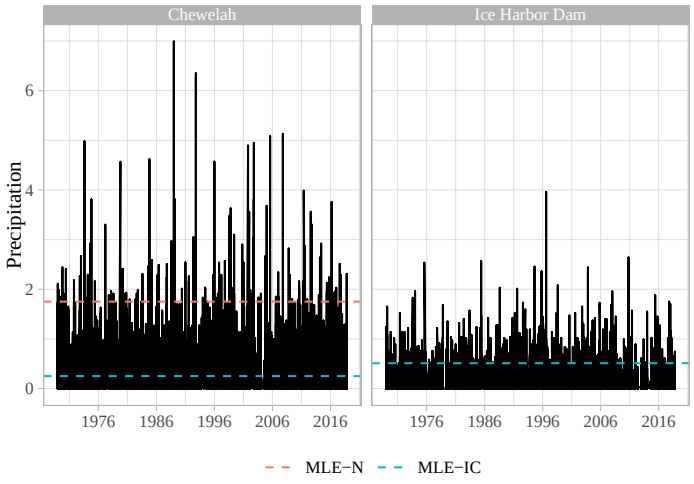

Figure 4: Time series plot of Chewelah station and Ice Harbor Dam station with thresholds.

Table 2. The candidate thresholds and number of exceedances of Chewelah station and Ice Harbor Dam station.

	Chewelah		Ice Harbor Dam	
percentile (%)	candidate thresholds (inches)	number of exceedances	candidate thresholds (inches)	number of exceedances
70	0.13	1984	0.03	2141
72	0.15	1868	0.05	1831
74	0.20	1722	0.05	1831
76	0.25	1596	0.08	1626
78	0.28	1491	0.10	1496
80	0.36	1331	0.13	1320
82	0.41	1218	0.15	1236
84	0.51	1067	0.20	1089
86	0.56	939	0.25	920
88	0.64	828	0.28	858
90	0.76	677	0.36	718
92	0.91	526	0.43	559
94	1.07	412	0.51	430
96	1.27	289	0.69	281
98	1.75	135	0.94	141

lines. At Chewelah, the MLE-N threshold of 1.75 in (98th percentile) lies far above most observed precipitation events, resulting in very few exceedances, while the MLE-IC threshold of 0.25 in (76th percentile) captures a richer set of heavy rainfall occurrences that are still representative of extreme behavior. A similar contrast is seen at Ice Harbor Dam, where MLE-IC identifies a valid threshold of 0.51 in (94th percentile), but MLE-N fails to select any threshold, implying that the rounding-induced discretization caused the standard procedure to reject all candidates. The time series plots clearly illustrate that modeling rounding via induced censoring allows MLE-IC to utilize more informative exceedances, producing stable and realistic threshold choices that preserve the continuity of extreme events over time.

Table 2 lists the candidate thresholds expressed both as percentiles and as precipitation amounts (in inches), together with the corresponding numbers of exceedances for the Chewelah and Ice Harbor Dam stations. As expected, higher percentiles correspond to larger thresholds and fewer exceedances, reflecting the typical bias–variance trade-off in threshold selection. For instance, at Chewelah, the 70th percentile (0.13 in) yields 1,984 exceedances, while the 98th percentile (1.75 in) leaves only 135 exceedances. A similar monotone decline occurs at Ice Harbor Dam, where exceedances drop from 2,141 at the 70th percentile (0.03 in) to 141 at the 98th percentile (0.94 in).

The comparison between the two estimation methods shows that MLE-IC generally selects lower percentile thresholds than MLE-N. Specifically, MLE-IC identifies 0.25 in (76th percentile) for Chewelah and 0.51 in (94th percentile) for Ice Harbor Dam, whereas MLE-N selects a much higher 1.75 in (98th percentile) threshold at Chewelah and fails to select any valid threshold at Ice Harbor Dam. By choosing lower percentiles, MLE-IC includes more exceedances in the model fitting, thereby stabilizing parameter estimation and reducing sampling variability. These results illustrate how accounting for data discretization avoids spurious over-rejection of candidate thresholds and yields a more informative sample for tail inference. lists percentile-based thresholds and corresponding exceedance counts for Chewelah and Ice Harbor Dam. As expected, higher percentiles lead to larger thresholds and fewer exceedances. MLE-IC favors lower thresholds (0.25 in at Chewelah and 0.51 in at Ice Harbor Dam) compared with MLE-N, resulting in more exceedances and a richer tail sample. MLE-N fails to identify a valid threshold at Ice Harbor Dam, illustrating its over-rejection tendency when rounding effects are ignored.

Figure 5 shows the p-values from the Anderson–Darling goodness-of-fit tests across the 15 candidate thresholds for the Chewelah and Ice Harbor Dam stations, both before and after the ForwardStop adjustment. At both sites, the p-values obtained from MLE-IC are consistently higher than those from MLE-N, indicating that explicitly modeling discretization improves the goodness of fit of the generalized Pareto model. After applying the ForwardStop rule, the higher p-values under MLE-IC cause the adjusted sequence to cross the decision boundary earlier, leading to smaller selected thresholds and, consequently, more exceedances available for tail modeling. For example, MLE-IC identifies 0.25 in (76th percentile) as the threshold at Chewelah and 0.51 in (94th percentile) at Ice Harbor Dam, whereas MLE-N either selects a much higher threshold or fails to identify one altogether. The figure illustrates that accounting for rounding effects prevents spurious rejections of plausible thresholds and supports more stable, data-rich extreme-value inference.

Figure 6 shows Q–Q plots comparing the empirical quantiles of exceedances above the selected thresholds with the fitted generalized Pareto quantiles for the Chewelah and Ice

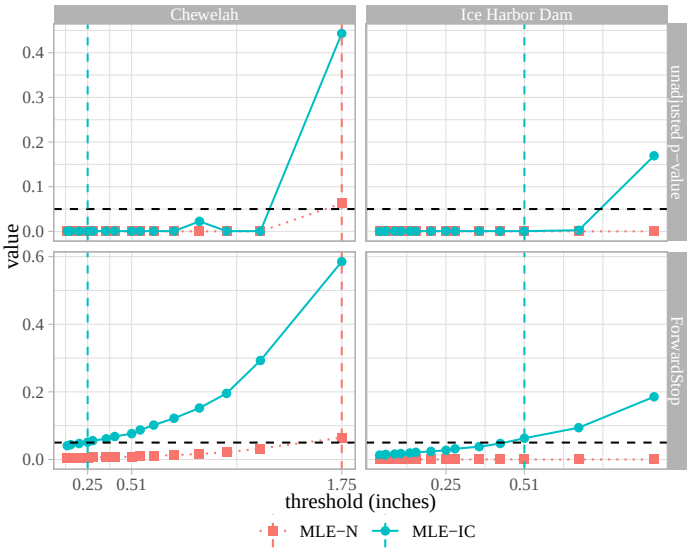

Figure 5: P-values before and after the ForwardStop adjustment. Those based on MLE-IC are higher than those based on MLE-N, so the selected thresholds from MLE-IC are smaller.

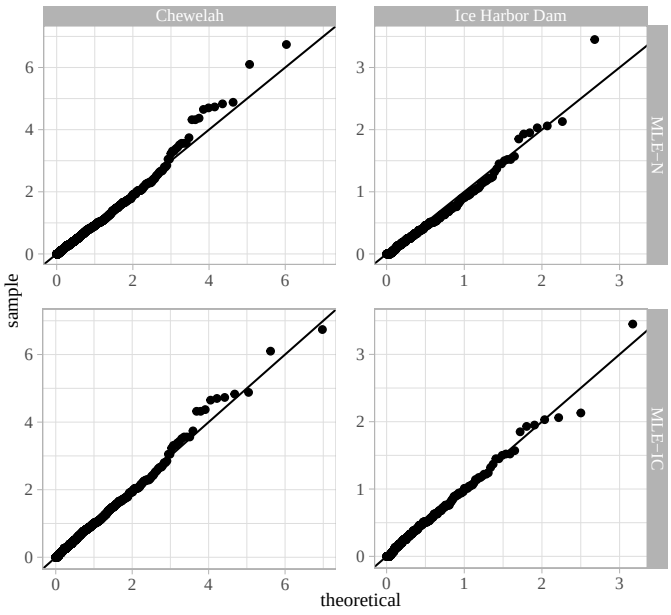

Figure 6: Q-Q plots of the observation above threshold versus fitted GPD.

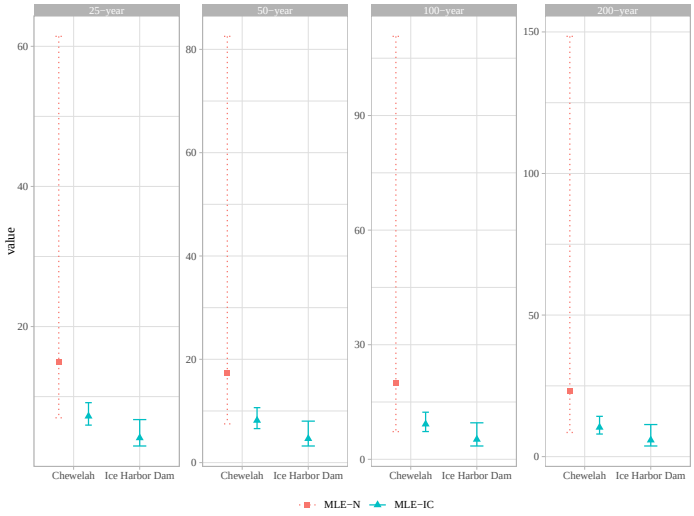

Figure 7: 25-, 50-, 100-, 200-year return level and confidence interval at the threshold of each station.

Harbor Dam stations. The results reveal that the MLE-IC fits align closely with the 45-degree reference line across nearly the entire range of observed values, indicating that the induced-censoring model provides an excellent representation of the upper-tail behavior. Conversely, the MLE-N fits exhibit noticeable deviations in the upper quantiles, particularly for Ice Harbor Dam, where ignoring rounding leads to systematic underestimation of extremes. The improvement from MLE-IC is most apparent in the tail region, where accurate modeling is critical for return-level estimation. Overall, these Q–Q plots confirm that accounting for quantization substantially improves model fit and ensures that extrapolation to rare events is based on a statistically coherent representation of the data's tail distribution.

Figure 7 presents the estimated 25-, 50-, 100-, and 200-year return levels and their confidence intervals for the Chewelah and Ice Harbor Dam stations under both MLE-N and MLE-IC estimation. Across all return periods, MLE-IC yields higher return levels than MLE-N, reflecting its smaller estimated shape parameter and larger estimated scale parameter, which together imply a heavier and more realistic upper tail. The difference between the two methods grows with the return period, with MLE-N severely underestimating extreme quantiles due to its bias in parameter estimation when rounding is ignored. Moreover, the confidence intervals from MLE-IC are much narrower, showing improved efficiency and numerical stability in tail extrapolation. These results confirm that properly accounting for quantization not only corrects bias in return level estimation but also enhances precision, leading to more credible assessments of extreme precipitation risk.

Received August 2024

Sai Ma. Department of Statistics, University of Connecticut, Storrs, CT, USA.
E-mail address: sai.ma@uconn.edu

Jun Yan. Department of Statistics, University of Connecticut, Storrs, CT, USA.
E-mail address: jun.yan@uconn.edu

Xuebin Zhang. Pacific Climate Impacts Consortium, Univesity of Victoria, Victoria, BC, Canada.
E-mail address: xuebinzhang23@uvic.ca