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

Sai Ma[1], Jun Yan[1,2], Xuebin Zhang[3]

[1]Department of Statistics, University of Connecticut, Storrs, CT, USA.
[2]Center for Environmental Sciences and Engineering, University of Connecticut, Storrs, CT, USA.
[3]Climate Research Division, Environment and Climate Change Canada, Downsview, Ontario, Canada.

1 Return Level Plots in the Simulation

Figure 1 shows the biases of the estimates of the 25-, 50-, 100-, and 200-year return levels from the MLE-IC and MLE-N. The corresponding MSEs are shown in Figure 2.

2 Supplementary Materials to Application

Table 1 summarizes the candidate thresholds and the corresponding number of exceedances at the two stations, one at Chewelah and the other at Ice Harbor Dam.

Figure 3 shows the p-values at the 15 candidate thresholds at the two sites before and after the ForwardStop adjustment using the AD test as in Bader et al. (2018).

Figure 4 shows the 25-, 50-, 100-, 200-year return levels and confidence intervals at the thresholds. The return levels are totally different between the two methods. Since

Corresponding author: Jun Yan, jun.yan@uconn.edu

Table 1. The candidate thresholds and number of exceedances of Chewelah station and Ice Harbor Dam station.

	Chewelah		Ice Harbor Dam	
percentiles (%)	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

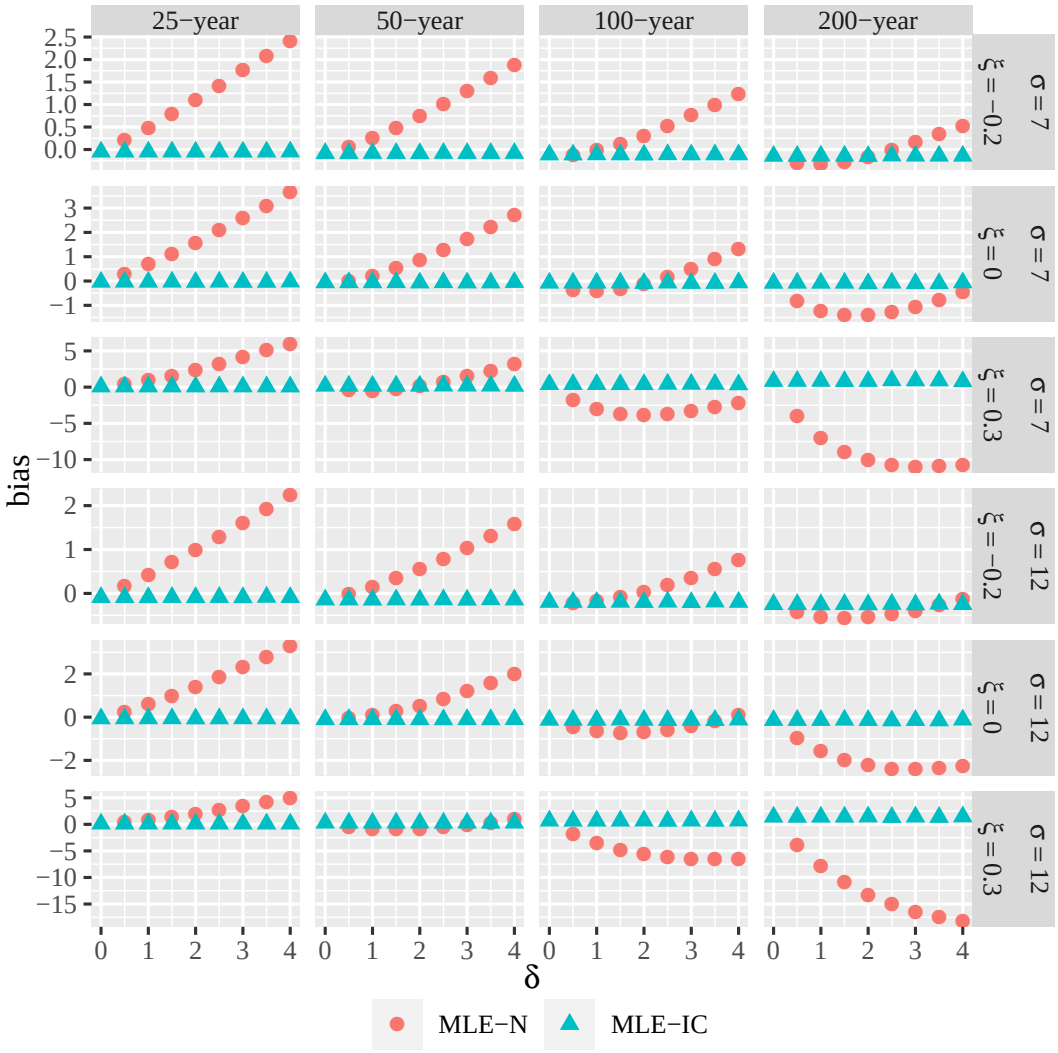

Figure 1. Bias of 25-, 50-, 100-, 200-year return level. The continuous scenarios were obtained by regular MLE. The rounded-off scenarios were obtained by naive MLE and MLE-IC.

MLE-IC has smaller estimated shape parameter and larger estimated scale parameter than MLE-N, it always has larger return level than MLE-N. Also, MLE-IC has smaller confidence interval than MLE-N.

Table 2 summarizes the selected thresholds and the corresponding numbers of exceedances.

References

Bader, B., Yan, J., & Zhang, X. (2018). Automated threshold selection for extreme value analysis via ordered goodness-of-fit tests with adjustment for false discovery rate. *The Annals of Applied Statistics*, *12*(1), 310–329.

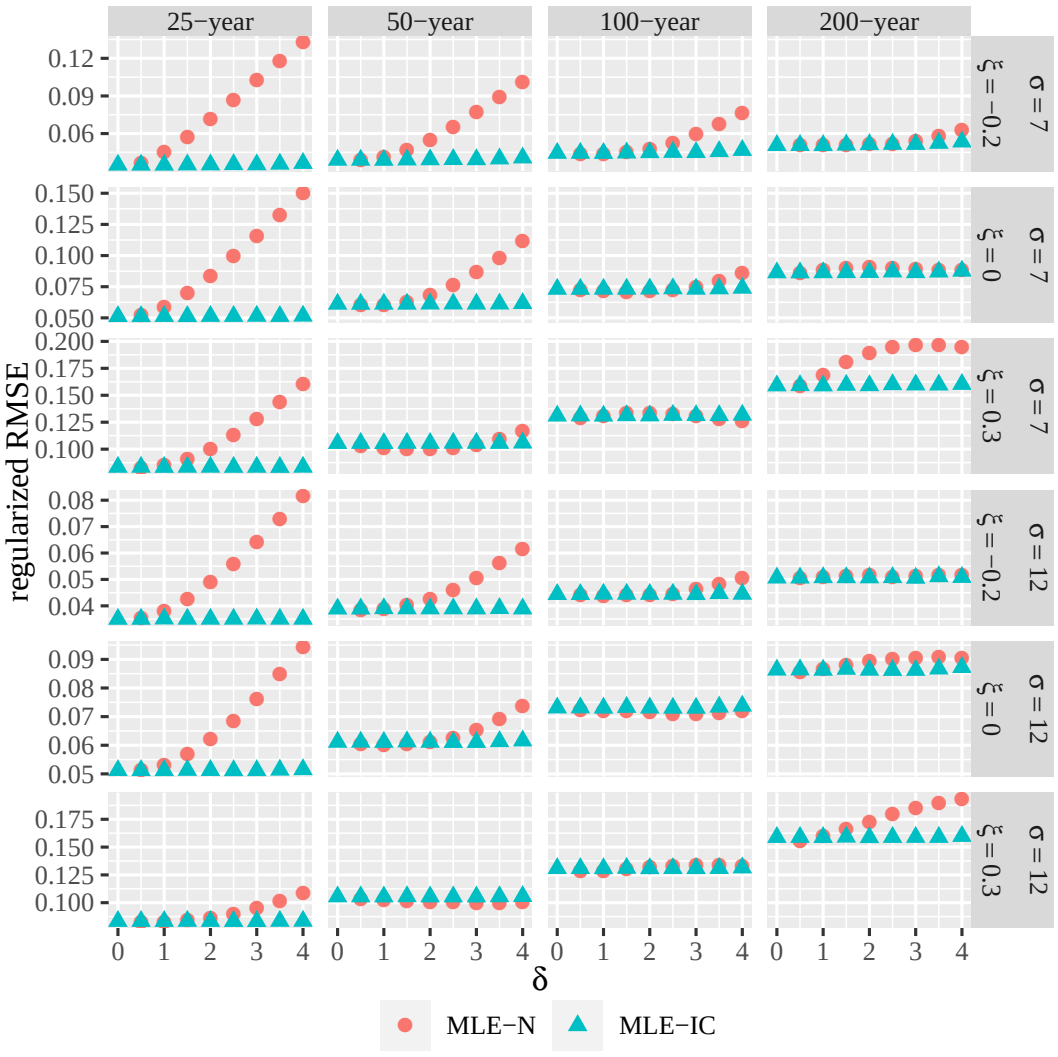

Figure 2. Regularized RMSE of 25-, 50-, 100-, 200-year return level. The continuous scenarios were obtained by regular MLE. The rounded-off scenarios were obtained by MLE-N and MLE-IC.

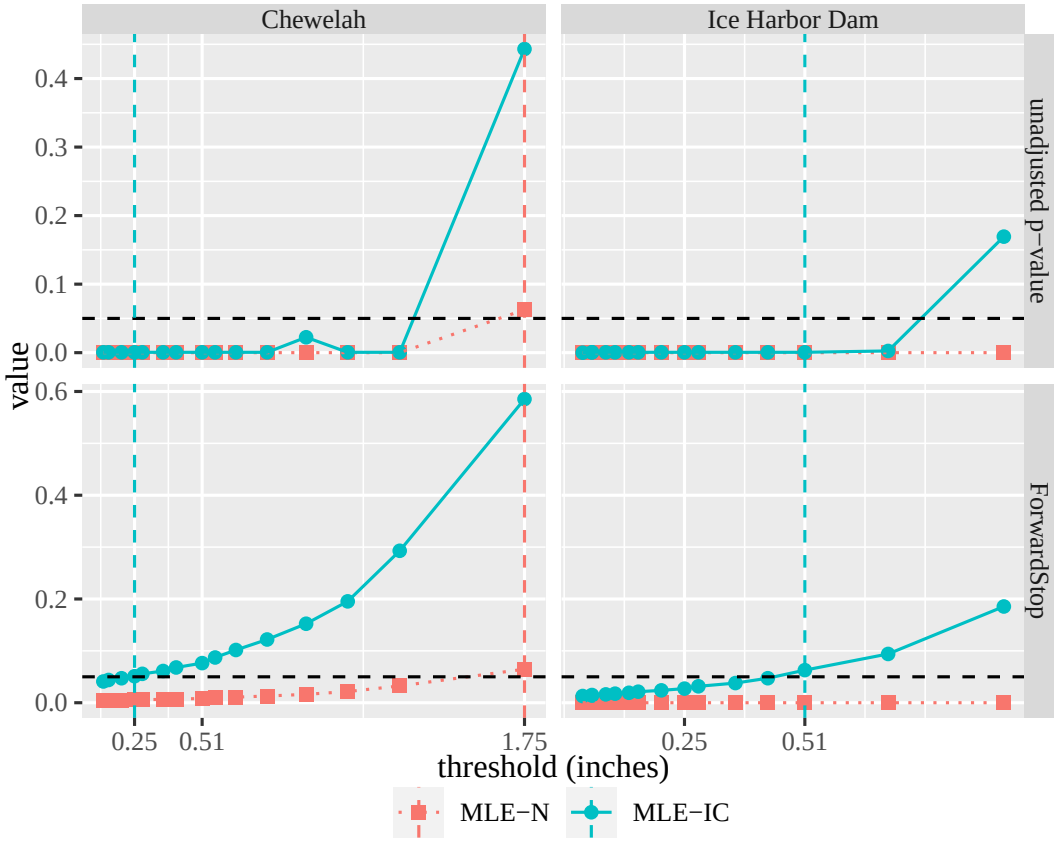

Figure 3. P-values before and after the ForwardStop adjustment. Those based on MLE-IC are higher than those based on MLE-N, so the selected thresholds from MLE-IC are smaller.

Table 2. The summary of 18 monitoring stations from 1969 to 2018 in the eastern part of the Washington State.

Station	MLE-N		MLE-IC	
	threshold	number of exceedances	threshold	number of exceedances
Chewelah	1.75	131	0.25	1596
Coulee Dam 1 SW				
Davenport				
Harrington				
Ice Harbor Dam			0.51	430
Lacrosse				
Mill Creek Dam			0.08	2190
Newport			1.30	411
Odessa				
Pomeroy			1.40	141
Pullman 2 NW			0.79	625
Republic				
Ritzville 1 SSE			0.79	310
Rosalia				
St. John				
Whitman Mission	1.12	143	0.66	425
Wilbur				
Spokane Intl AP			0.28	1515

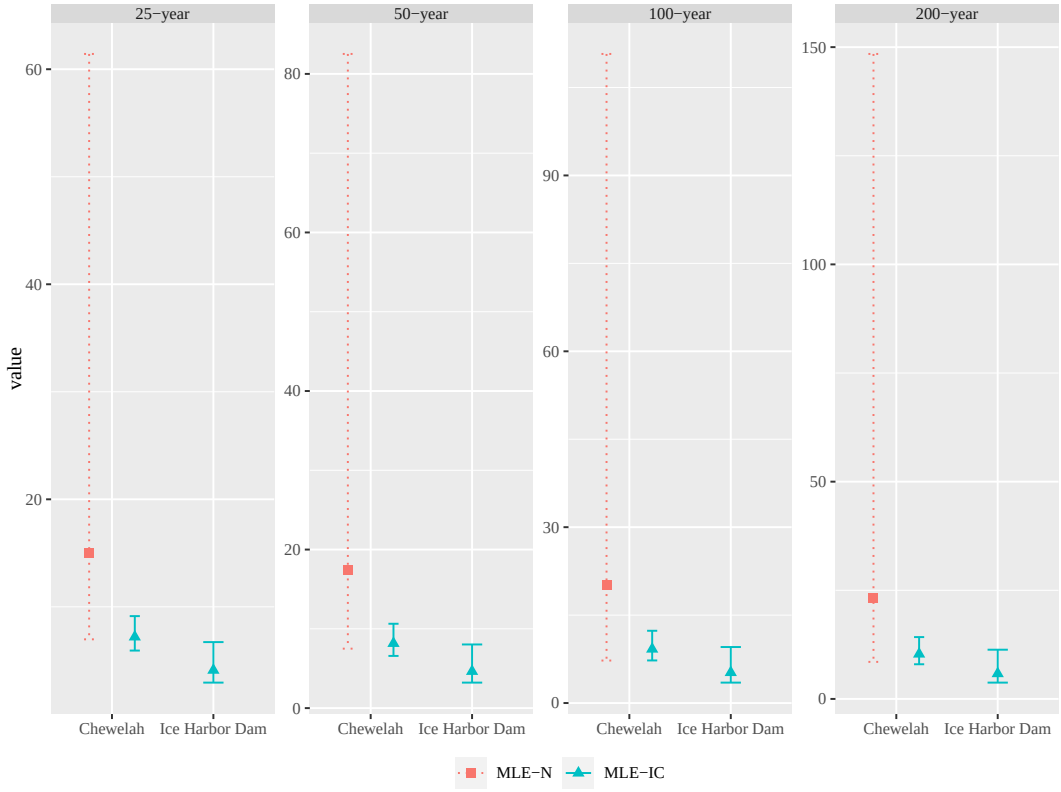

Figure 4. 25-, 50-, 100-, 200-year return level and confidence interval at the threshold of each station.