

1 **Technical note: Graphical Methods for Screening Annual Maxima and Peaks over Threshold flood data before**
2 **Flood Frequency Analysis**

3 Paul H. Whitfield¹ and Donald H. Burn²

4 ¹Centre for Hydrology, University of Saskatchewan, Canmore, AB, Canada

5 ²Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, ON, Canada;

6 [Supplementary material:](#)

7 The supplementary material provides these three visualization approaches to flood screening for four different rivers across Canada.
8 Each visualization is provided for the annual maxima and the second for POT. In plots where POT data were used, they are so
9 labelled. The four cases span Canada from east to west and are provided so that the reader has access to much broader range of cases
10 than could be covered in the main paper.

11

12 Table S1. Flood Frequency Analysis Screening checklist with examples.

Time series plot	Description	Example
Entire or part years	Solid points indicate maximum of a complete year, open symbols indicate less than a complete year. Are partial years representative?	S1a, S4a, S7a,
Flagged data points	Are flagged data going to be influential in the analysis?	S1a, S1b,S4a, S4b, S7a, S7b,
Missing years or observations	Large gaps that might affect the analysis?	S1b, S10a, S10b
Trends	Trends in the magnitude that might compromise flood frequency analysis?	S1a, S7b, S10b,
Change points	Change point that indicates a methodological change, or a structural change related to changes in the seasonal distribution of peaks?	S1a, S4b, S7b, S10b
Return period plot		
Low outliers	Low outliers may be influential in selecting an incorrect distribution.	S2a, S2b, S8a,
High outliers	The statistical distribution is fat tailed, or there are many high outliers there may be mixed populations.	S5a, S5b, S11a, S11b
Timing outliers	Timing outliers, particularly of high magnitude events, may indicate separate rare “Rogue” events generated by an entirely different process.	S2b, S5b, S8a, S8b, S11a, S11b
Polar plot		
High / low outliers	Where do high and low outliers occur seasonally	S3a, S3b, S6a, S6b, S12a, S12b
Timing outliers	Do timing outliers indicate Rogue events?	S3b, S6b, S9a, S9b
Kernel density	Is there a single population of events or multiple clusters indicating a single or multiple generating processes?	S12a, S12b
Rogue events	Floods that are outliers in magnitude, timing, and/or density	S12a, S12b
Fuzzy clustering	Is there a single population of events or multiple clusters indicating a single or multiple generating processes?	S3b, S9a, S9b, S12a, S12b

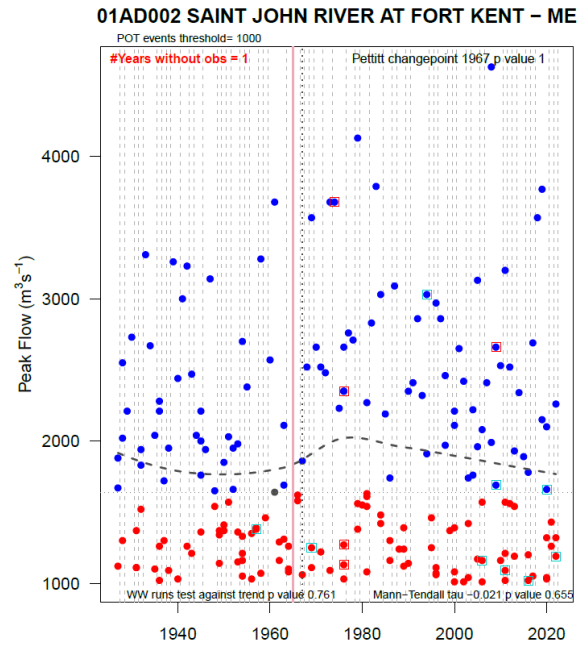
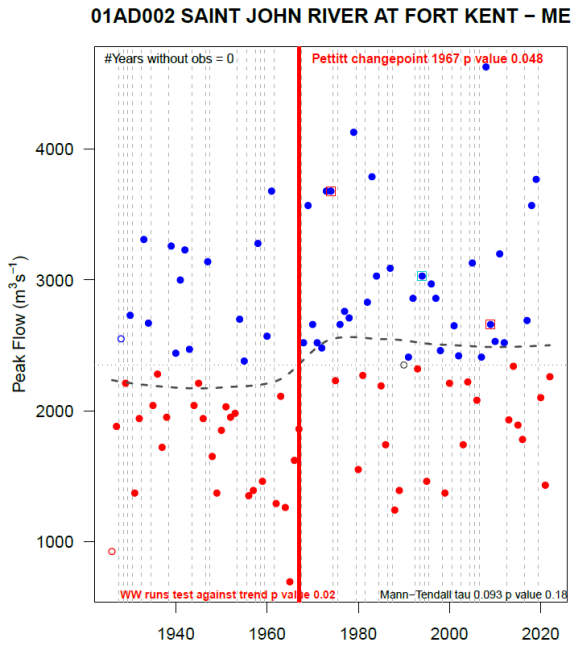
13

14

15

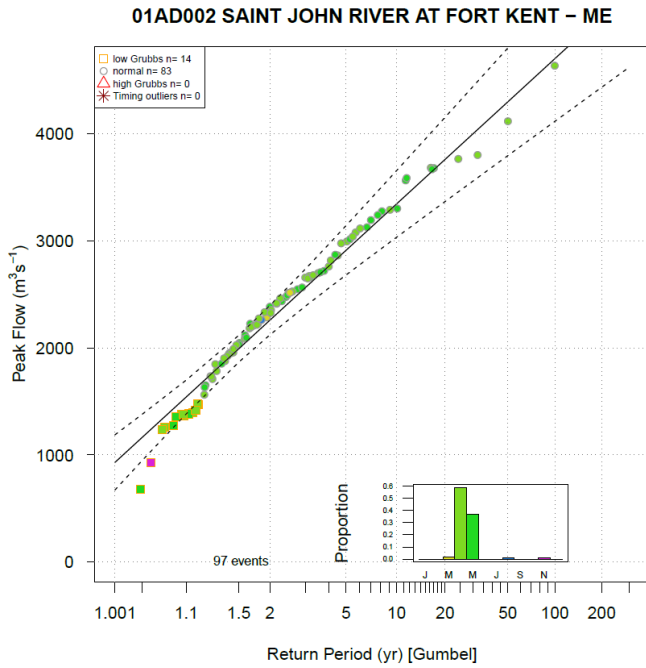
a)

b)

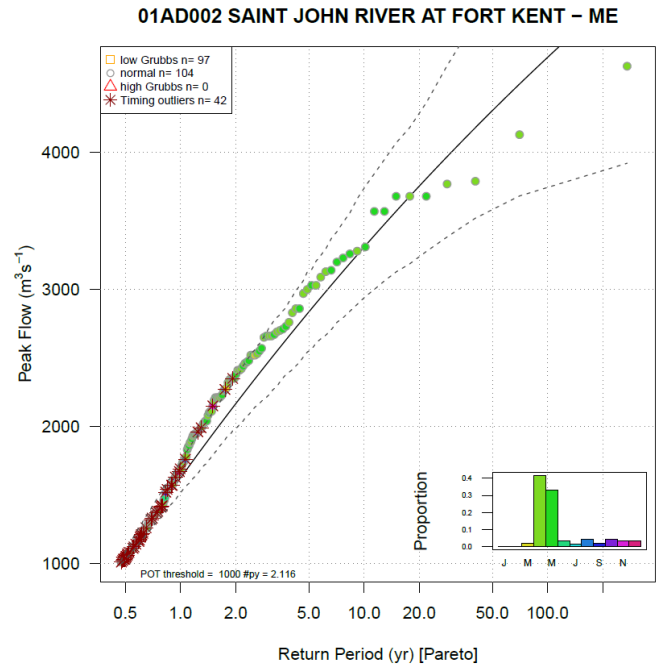


16 Figure S1. Annotated time series plots for (a) annual maximum series, and (b) peaks-over-threshold from 01AD002 St John River at
 17 Fort Kent, ME. Amax series with partial years have open symbols. Years without observations are shown as pink vertical lines.
 18 Coloured squares mark points that are flagged by Water Survey of Canada as partial days, backwater, dry, or estimated. Points
 19 above/below the median are shown in blue/red following Wald-Wolfowitz test against trend. The loess line indicates tendency, and
 20 shows significance of Mann-Kendall trend test (NS – dotted gray, $p < 0.10$ - dashed dark red, or $p < 0.05$ – solid red line). A vertical
 21 line shows the change point for the Pettitt test following the same scheme. Text indicates the number of years without observations if $>$
 22 0 (upper left), Pettitt test (upper right), Wald-Wolfowitz test (lower left), and Mann-Kendall (lower right).

a)



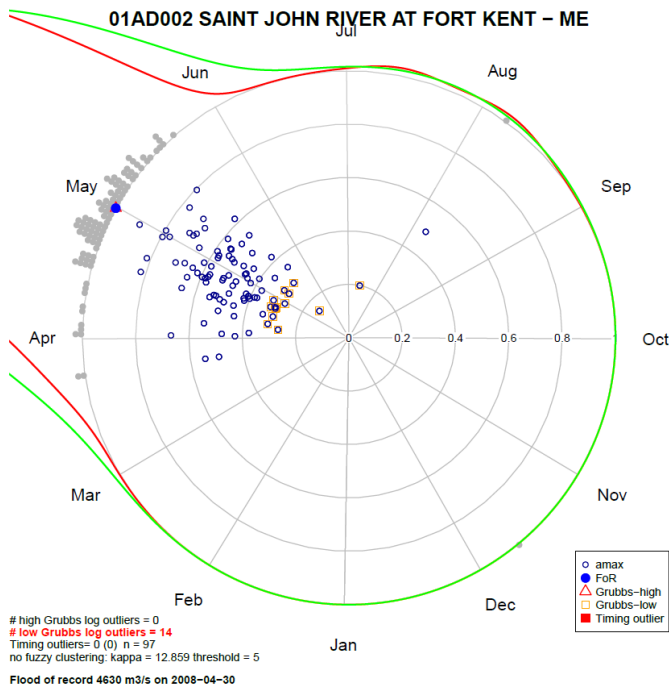
b)



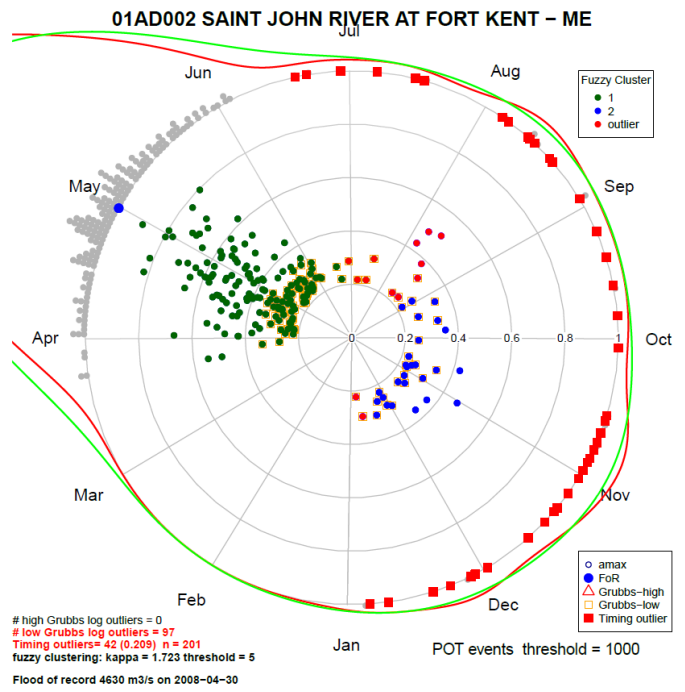
23 Figure S2. Return period plots showing the magnitude of annual maximum flood peaks for (a) annual maximum series, and (b) peaks-
24 over-threshold series for 01AD002 St John River at Fort Kent, ME. The annual maxima series is shown using Gumbel, and the peaks-
25 over-threshold using a Pareto. Commonly used plots in flood frequency analysis, here include tests for high, low, and timing outliers.
26 Symbols indicate whether an observation is a high (triangle), low (square), or timing (asterisk) outliers. The colored symbols indicated
27 the months of the year thus indicating seasonal distributions of floods and a histogram of these is placed in a corner of the plot to show
28 the distribution of peaks across months. The legend includes the counts of high, normal, low, and timing outliers.

29

a)



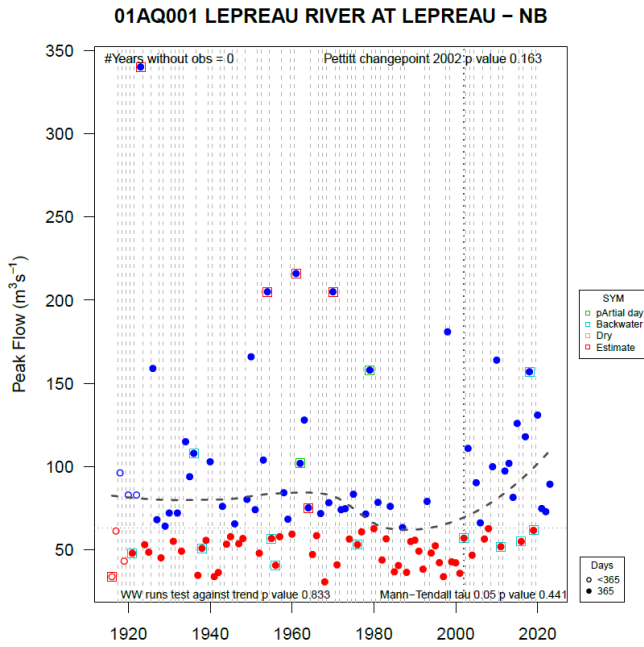
b)



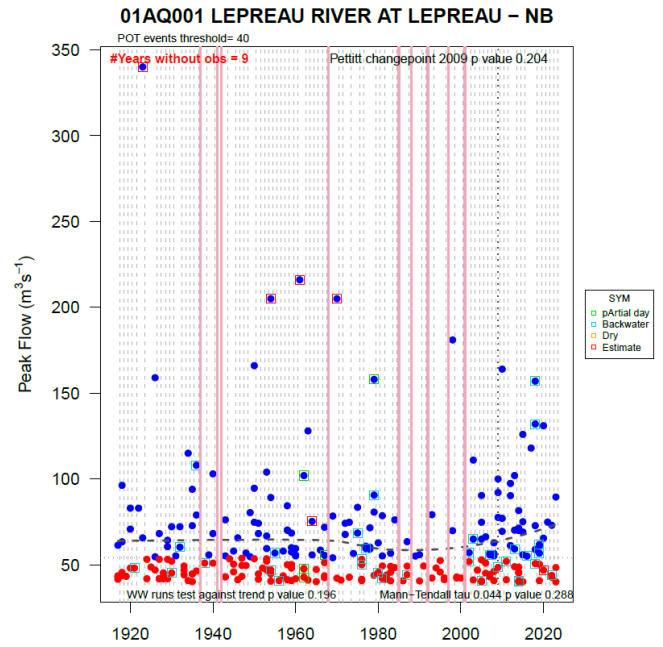
30 Figure S3. Flood series as a polar plot for (a) annual maximum series, and (b) peaks-over-threshold series for 01AD002 St John River
 31 at Fort Kent, ME. The plots show the seasonal distribution of floods and their relative magnitude on the radial axis. High (red
 32 triangle) and low (orange square) outliers are marked with surrounding symbols on the main plot. Timing outliers are shown (red solid
 33 squares) on the unit circle. The flood of record (FoR) is highlighted as a blue dot, and a histogram of individual events is on the
 34 outside of the circle. Two kernel density estimates (bandwidths of 15 and 50, red and green respectively) as a visual indication of the
 35 seasonality. Fuzzy clustering is performed when κ is less than 5, and individual maxima are colored based on cluster membership
 36 shown in the legend in the upper right and may include cluster outliers. Details about the flood of record with the counts of the outliers
 37 are shown below the plot.

38

a)



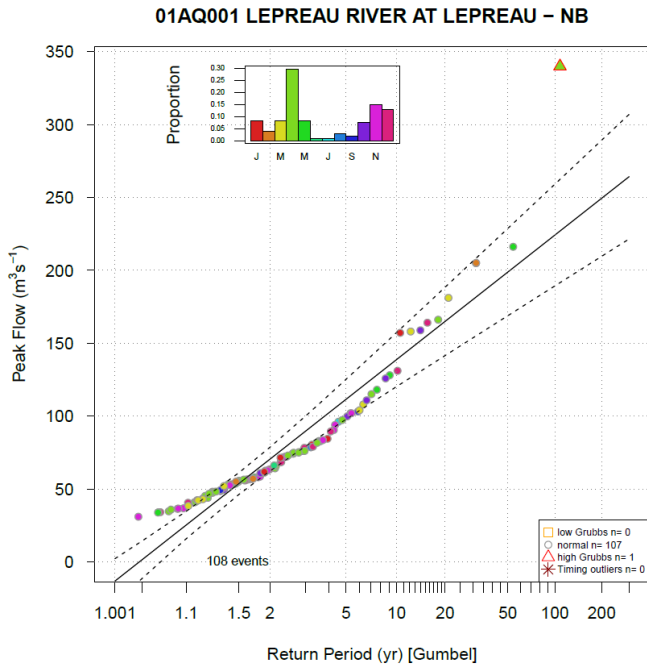
b)



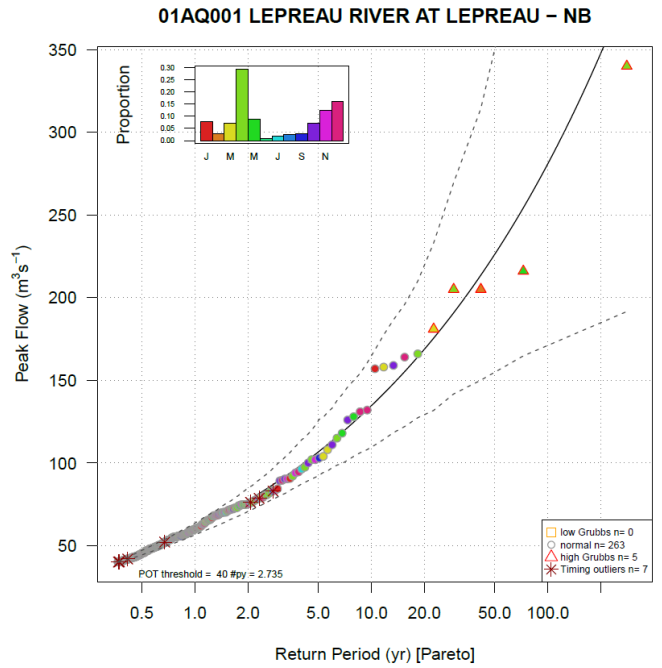
39 Figure S4. Annotated time series plots for (a) annual maximum series, and (b) peaks-over-threshold from 01AQ001 Lepreau River at
 40 Lepreau, NB. Amax series with partial years have open symbols. Years without observations are shown as pink vertical lines.
 41 Coloured squares mark points that are flagged by Water Survey of Canada as partial days, backwater, dry, or estimated. Points
 42 above/below the median are shown in blue/red following Wald-Wolfowitz test against trend. The loess line indicates tendency, and
 43 shows significance of Mann-Kendall trend test (NS – dotted gray, $p < 0.10$ - dashed dark red, or $p < 0.05$ – solid red line). A vertical
 44 line shows the change point for the Pettitt test following the same scheme. Text indicates the number of years without observations if >
 45 0 (upper left), Pettitt test (upper right), Wald-Wolfowitz test (lower left), and Mann-Kendall (lower right).

46

a)



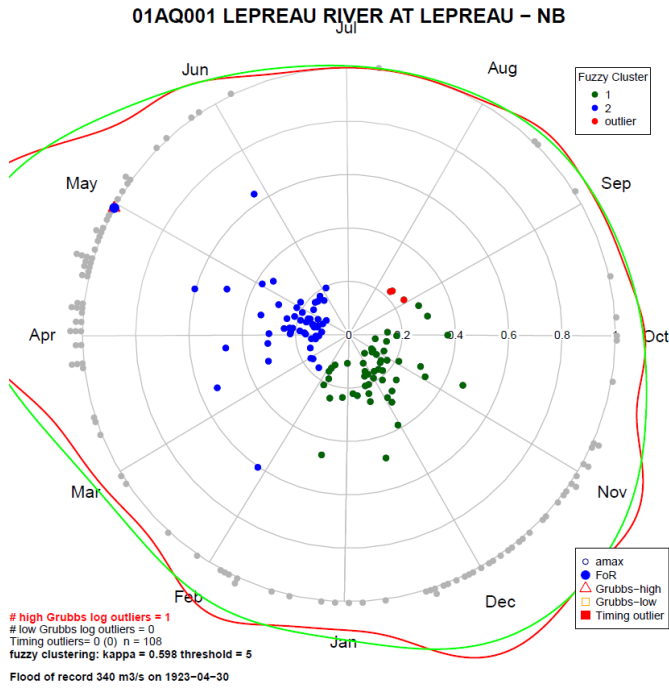
b)



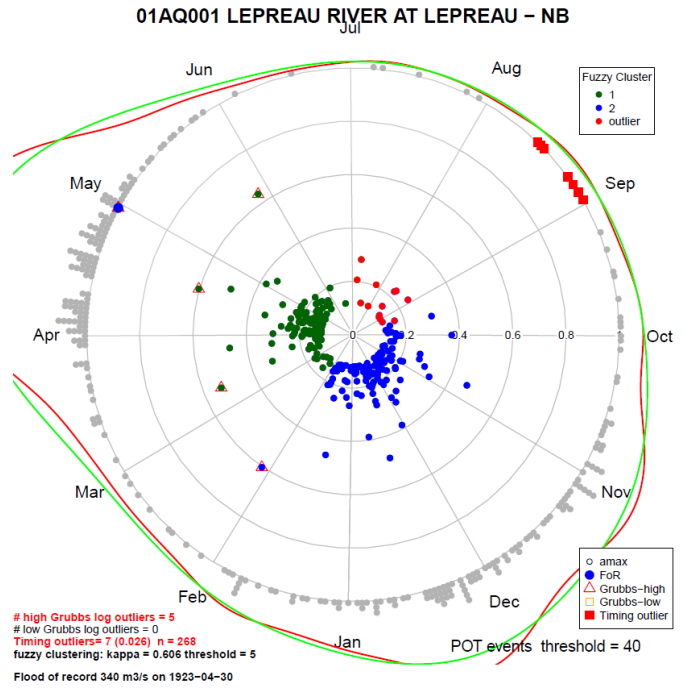
47 Figure S5. Return period plots showing the magnitude of annual maximum flood peaks for (a) annual maximum series, and (b) peaks-
 48 over-threshold series for 01AQ001 Lepreau River at Lepreau, NB. The annual maxima series is shown using Gumbel, and the peaks-
 49 over-threshold using a Pareto. Commonly used plots in flood frequency analysis, here include tests for high, low, and timing outliers.
 50 Symbols indicate whether an observation is a high (triangle), low (square), or timing (asterisk) outliers. The colored symbols indicated
 51 the months of the year thus indicating seasonal distributions of floods and a histogram of these is placed in a corner of the plot to show
 52 the distribution of peaks across months. The legend includes the counts of high, normal, low, and timing outliers.

53

a)



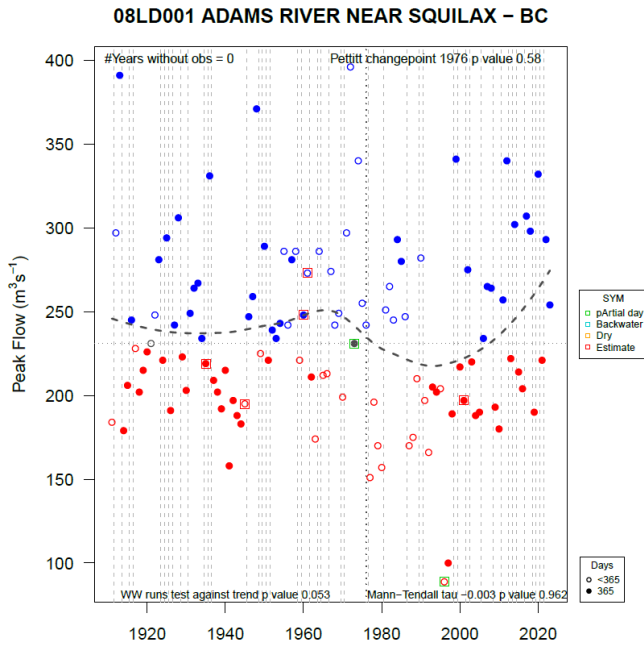
b)



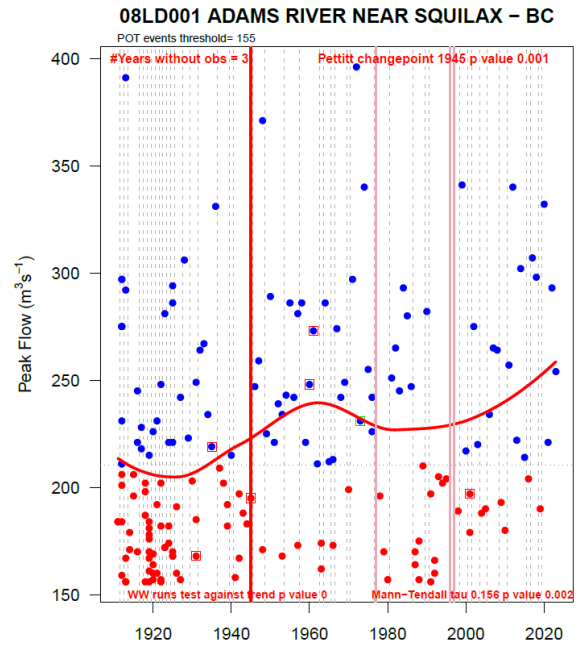
54 Figure S6. Flood series as a polar plot for (a) annual maximum series, and (b) peaks-over-threshold series for 01AQ001 Lepreau River
 55 at Lepreau, NB. The plots show the seasonal distribution of floods and their relative magnitude on the radial axis. High (red triangle)
 56 and low (orange square) outliers are marked with surrounding symbols on the main plot. Timing outliers are shown (red solid squares)
 57 on the unit circle. The flood of record (FoR) is highlighted as a blue dot, and a histogram of individual events is on the outside of the
 58 circle. Two kernel density estimates (bandwidths of 15 and 50, red and green respectively) as a visual indication of the seasonality.
 59 Fuzzy clustering is performed when κ is less than 5, and individual maxima are colored based on cluster membership shown in the
 60 legend in the upper right and may include cluster outliers. Details about the flood of record with the counts of the outliers are shown
 61 below the plot.

62

a)



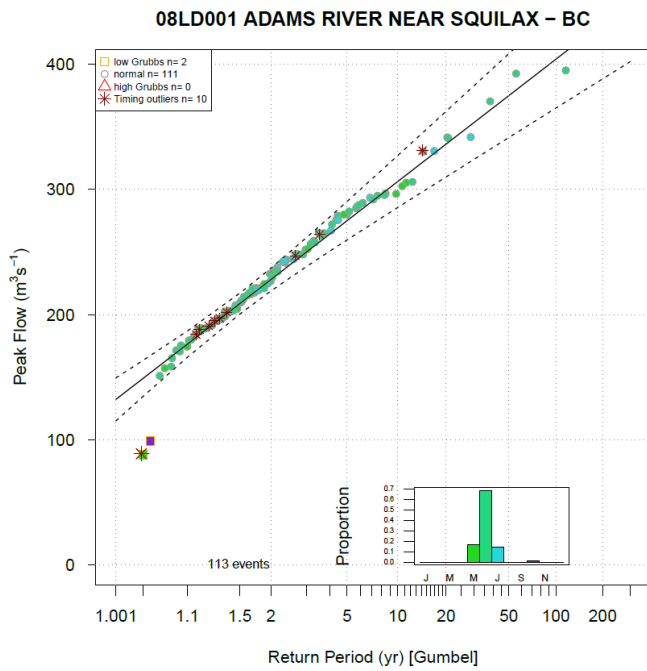
b)



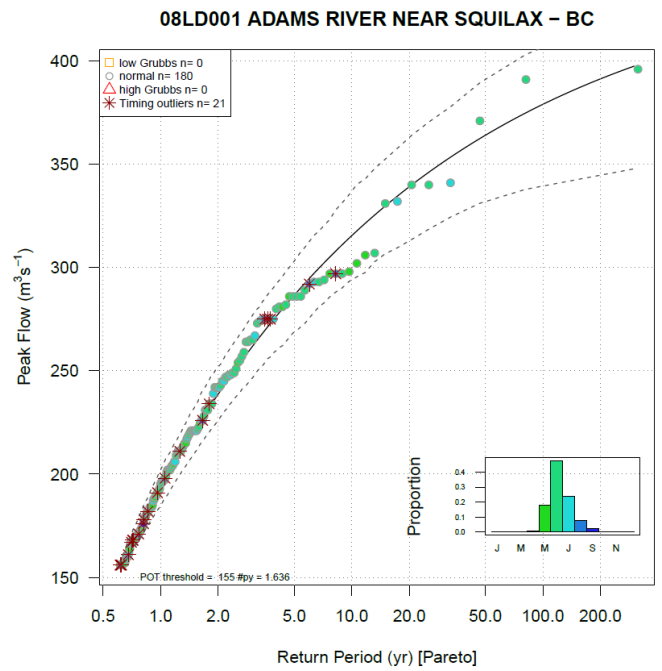
63 Figure S7. Annotated time series plots for (a) annual maximum series, and (b) peaks-over-threshold from 08LD001 Adams River near
 64 Squilax, BC. Amax series with partial years have open symbols. Years without observations are shown as pink vertical lines. Coloured
 65 squares mark points that are flagged by Water Survey of Canada as partial days, backwater, dry, or estimated. Points above/below the
 66 median are shown in blue/red following Wald-Wolfowitz test against trend. The loess line indicates tendency, and shows significance
 67 of Mann-Kendall trend test (NS – dotted gray, $p < 0.10$ - dashed dark red, or $p < 0.05$ – solid red line). A vertical line shows the change
 68 point for the Pettitt test following the same scheme. Text indicates the number of years without observations if > 0 (upper left), Pettitt
 69 test (upper right), Wald-Wolfowitz test (lower left), and Mann-Kendall (lower right).

70

a)

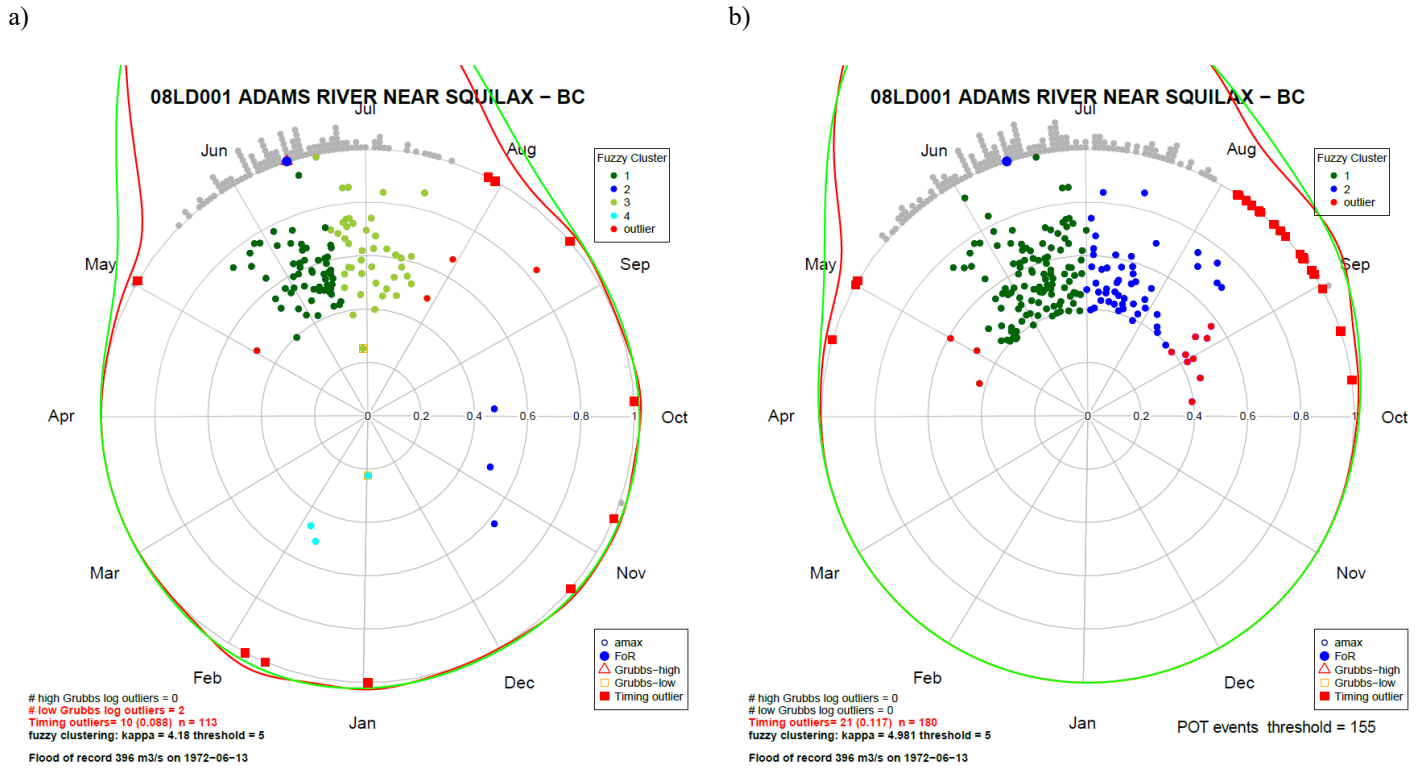


b)



71 Figure S8. Return period plots showing the magnitude of annual maximum flood peaks for (a) annual maximum series, and (b) peaks-
 72 over-threshold series for 08LD001 Adams River near Squilax, BC. The annual maxima series is shown using Gumbel, and the peaks-
 73 over-threshold using a Pareto. Commonly used plots in flood frequency analysis, here include tests for high, low, and timing outliers.
 74 Symbols indicate whether an observation is a high (triangle), low (square), or timing (asterisk) outliers. The colored symbols indicated
 75 the months of the year thus indicating seasonal distributions of floods and a histogram of these is placed in a corner of the plot to show
 76 the distribution of peaks across months. The legend includes the counts of high, normal, low, and timing outliers.

77

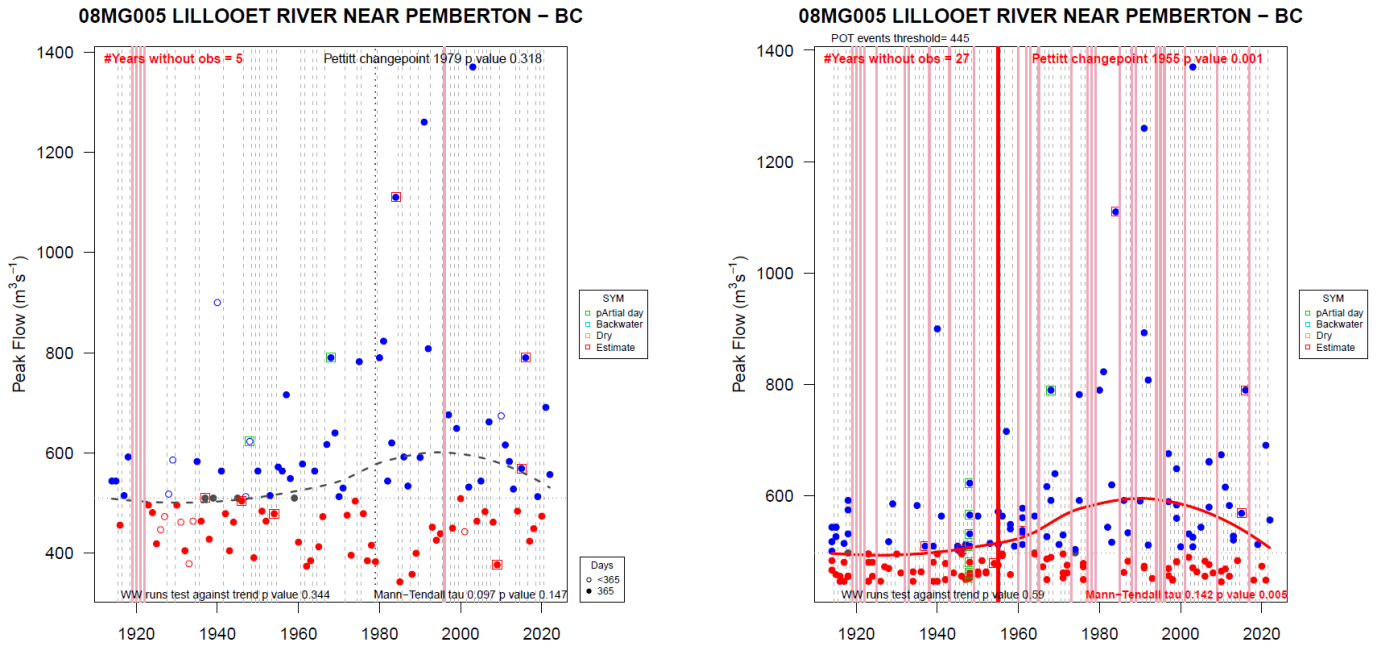


78 Figure S9. Flood series as a polar plot for (a) annual maximum series, and (b) peaks-over-threshold series for 08LD001 Adams River
 79 near Squilax, BC. The plots show the seasonal distribution of floods and their relative magnitude on the radial axis. High (red triangle)
 80 and low (orange square) outliers are marked with surrounding symbols on the main plot. Timing outliers are shown (red solid squares)
 81 on the unit circle. The flood of record (FoR) is highlighted as a blue dot, and a histogram of individual events is on the outside of the
 82 circle. Two kernel density estimates (bandwidths of 15 and 50, red and green respectively) as a visual indication of the seasonality.
 83 Fuzzy clustering is performed when κ is less than 5, and individual maxima are colored based on cluster membership shown in the
 84 legend in the upper right and may include cluster outliers. Details about the flood of record with the counts of the outliers are shown
 85 below the plot.

86

a)

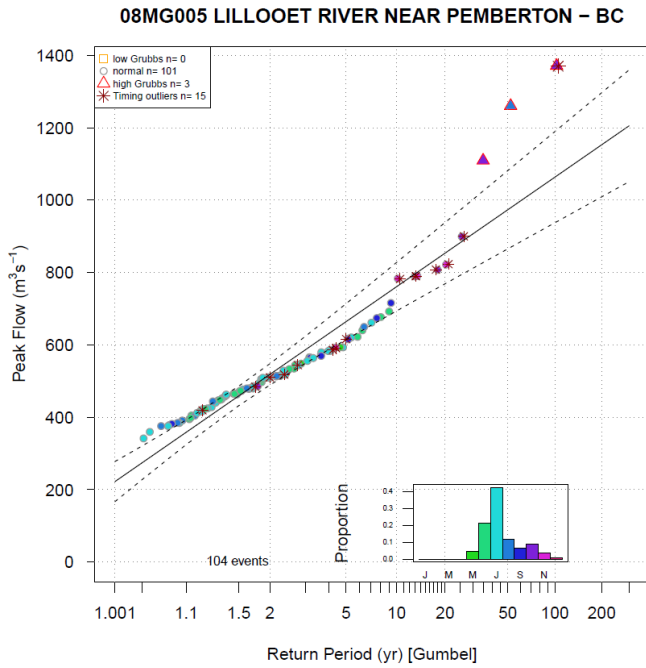
b)



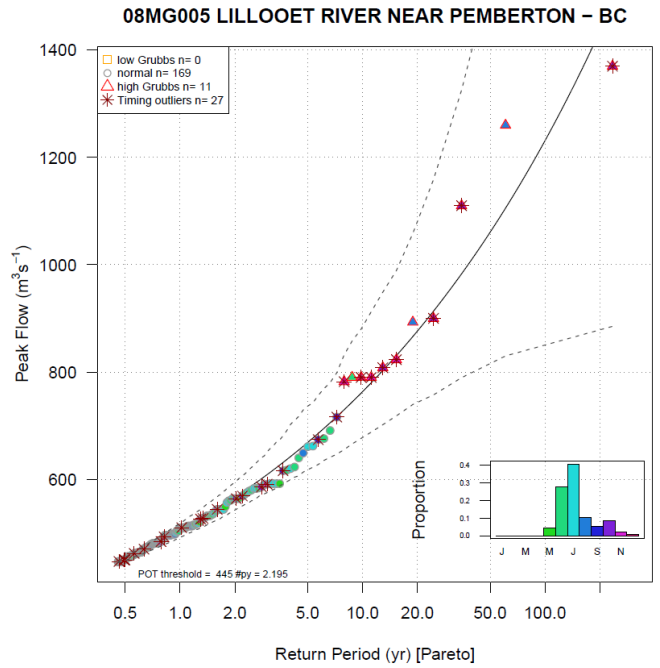
87 Figure S10. Annotated time series plots for (a) annual maximum series, and (b) peaks-over-threshold from 08MG005 Lillooet River
 88 near Pemberton, BC. Amax series with partial years have open symbols. Years without observations are shown as pink vertical lines.
 89 Coloured squares mark points that are flagged by Water Survey of Canada as partial days, backwater, dry, or estimated. Points
 90 above/below the median are shown in blue/red following Wald-Wolfowitz test against trend. The loess line indicates tendency, and
 91 shows significance of Mann-Kendall trend test (NS – dotted gray, $p < 0.10$ - dashed dark red, or $p < 0.05$ – solid red line). A vertical
 92 line shows the change point for the Pettitt test following the same scheme. Text indicates the number of years without observations if >
 93 0 (upper left), Pettitt test (upper right), Wald-Wolfowitz test (lower left), and Mann-Kendall (lower right).

94

a)



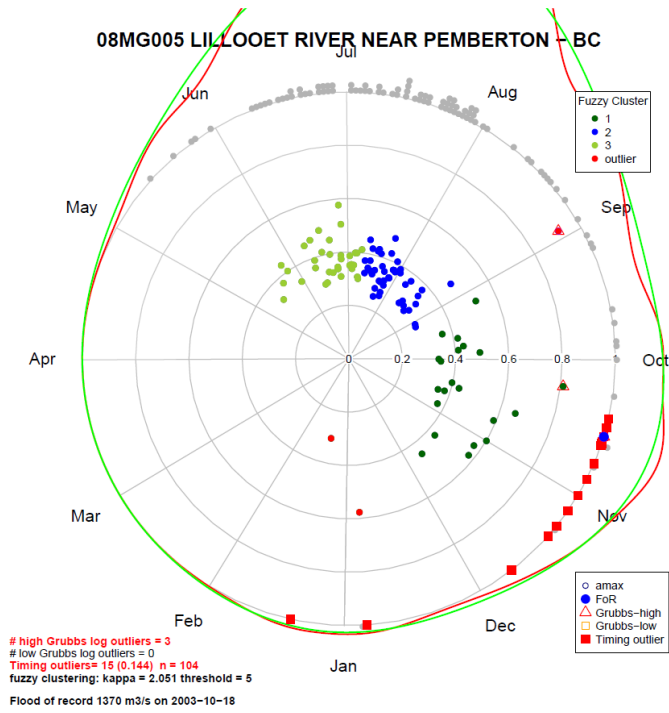
b)



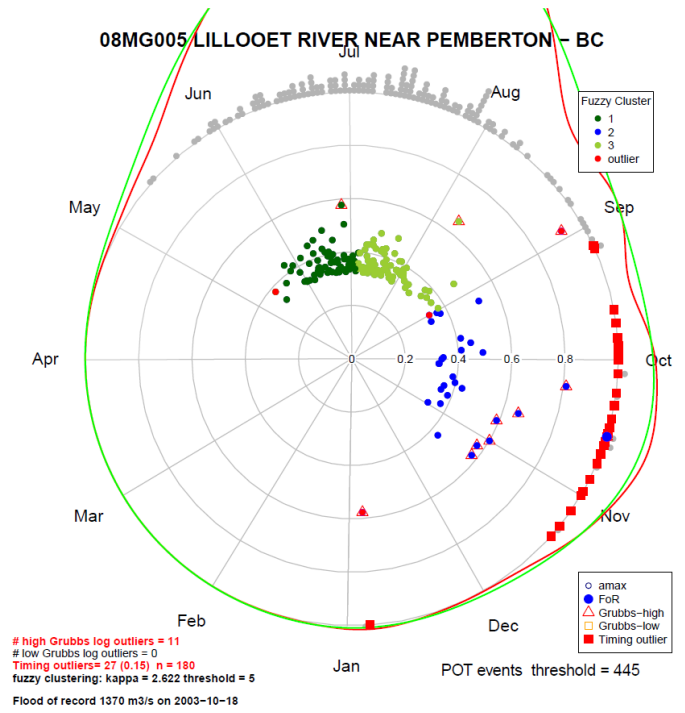
95 Figure S11. Return period plots showing the magnitude of annual maximum flood peaks for (a) annual maximum series, and (b)
96 peaks-over-threshold series for 08MG005 Lillooet River near Pemberton, BC. The annual maxima series is shown using Gumbel, and
97 the peaks-over-threshold using a Pareto. Commonly used plots in flood frequency analysis, here include tests for high, low, and timing
98 outliers. Symbols indicate whether an observation is a high (triangle), low (square), or timing (asterisk) outliers. The colored symbols
99 indicated the months of the year thus indicating seasonal distributions of floods and a histogram of these is placed in a corner of the
100 plot to show the distribution of peaks across months. The legend includes the counts of high, normal, low, and timing outliers.

101

a)



b)



102 Figure S12. Flood series as a polar plot for (a) annual maximum series, and (b) peaks-over-threshold series for 08MG005 Lillooet
 103 River near Pemberton, BC. The plots show the seasonal distribution of floods and their relative magnitude on the radial axis. High (red
 104 triangle) and low (orange square) outliers are marked with surrounding symbols on the main plot. Timing outliers are shown (red solid
 105 squares) on the unit circle. The flood of record (FoR) is highlighted as a blue dot, and a histogram of individual events is on the
 106 outside of the circle. Two kernel density estimates (bandwidths of 15 and 50, red and green respectively) as a visual indication of the
 107 seasonality. Fuzzy clustering is performed when κ is less than 5, and individual maxima are colored based on cluster membership
 108 shown in the legend in the upper right and may include cluster outliers. Details about the flood of record with the counts of the outliers
 109 are shown below the plot.