

Levee Overtopping Reliability Under Flood and Hydraulic Uncertainty

Masoud Salar*¹, Jamshid Piri*², Carlo De Michele¹

¹ Department of Civil Engineering, Chabahar Maritime University, Chabahar, Iran

² Department of Water Engineering, Faculty of Water and Soil, University of Zabol, Zabol, Iran

³ Department of Civil and Environmental Engineering, Politecnico di Milano, Milan, Italy

* Corresponding authors. E-mail: masoud.salar@cmu.ac.ir; j.piri@uoz.ac.ir

Received: 06 September 2025

Accepted: 08 October 2025

Available online: 10 December 2025

Abstract: Flood levees are usually designed for a discharge of a given return period, whereas the uncertainty of the hydraulic parameters that convert discharge into water level is rarely included in the design. This study assessed the annual overtopping probability of hypothetical levees on three rivers with contrasting flood regimes: the regulated Platte River (Nebraska, USA; 51 years), the Choptank River (Maryland, USA; 32 years) and the Ngaruroro River (New Zealand; 31 years). Monthly maximum daily discharges were extracted from open daily records, and annual maxima were fitted with the generalized extreme value (GEV) distribution by L-moments. A seasonal model based on monthly GEV distributions was also tested. Each levee crest was set at the Manning depth of the nominal 100-year flood plus 0.5 m freeboard. The limit state included the annual maximum discharge, Manning roughness, bed slope, channel width, a model-error factor and the crest level. The failure probability was computed by the first- and second-order reliability methods (FORM, SORM), crude Monte Carlo simulation (MCS, 5×10^6 samples) and importance sampling (IS, 2×10^4 samples). The MCS failure probabilities were 4.45×10^{-3} to 8.16×10^{-3} . FORM was within 2–15% of MCS, and the Hohenbichler–Rackwitz SORM within 0.7%, while needing only 92–279 limit-state evaluations. Hydraulic uncertainty raised the failure probability by 1.15–1.58 times at the heavy-tailed Platte and Choptank. At the Ngaruroro, whose fitted GEV is bounded above, overtopping was impossible with discharge as the only random variable, and roughness and model error contributed 48% of the FORM importance. Parametric bootstrap gave 90% intervals of the failure probability spanning more than an order of magnitude. A significant upward trend in the Choptank annual maxima ($p = 0.012$) tripled the failure probability between 1980 and 2011 under a non-stationary GEV. The seasonal model overestimated risk because monthly maxima were serially correlated. FORM and SORM are efficient and accurate for levee overtopping, but parameter uncertainty, the choice of the distribution and non-stationarity affect the failure probability more than the choice of the reliability method.

Keywords: Flood levee, Overtopping, FORM, SORM, Monte Carlo simulation, Generalized extreme value distribution

1. Introduction

Levees are the most common structural measure for flood protection along rivers. Conventional design sets the crest level at the water level of a design flood, typically the 100-year flood, plus a freeboard that is meant to cover uncertainties in the hydraulic calculation. This approach treats the design discharge and the stage–discharge relationship as known. In reality, both are uncertain, and the probability that the levee is overtopped can differ considerably from the nominal exceedance probability (Wood, 1977; Tung and Mays, 1981). Reliability-based design makes these uncertainties explicit and expresses the safety of the levee as a failure probability or reliability index (Apel et al., 2004; Vorogushyn et al., 2009).

Structural reliability theory offers several methods to compute failure probabilities. The first-order reliability method (FORM) linearises the limit-state function at the most probable failure point in standard normal space (Hasofer and Lind, 1974; Rackwitz and Fiessler, 1978). Second-order methods (SORM) correct the FORM estimate using the curvatures of the limit-state surface (Breitung, 1984; Hohenbichler and Rackwitz, 1988). Crude Monte Carlo simulation (MCS) is unbiased but

needs many samples for small probabilities, and importance sampling (IS) reduces this cost by sampling around the design point (Melchers and Beck, 2018). The accuracy of FORM depends on the non-linearity of the limit state and on the shape of the input distributions, particularly heavy-tailed or bounded extreme value distributions. It should therefore be checked for each application.

The discharge distribution is the key input to levee reliability. Flood frequency analysis is commonly based on annual maximum series (AMS) fitted with the generalized extreme value (GEV) distribution. L-moment estimation is robust for the short samples typical in hydrology (Hosking et al., 1985; Hosking and Wallis, 1997). Monthly maxima provide twelve times more values and can describe the seasonality of flood risk, but combining monthly distributions into an annual one requires an assumption about the dependence between months. Parameter uncertainty in short records is large and should be separated from natural variability (Merz and Thieken, 2005). In addition, trends in flood series challenge the stationarity assumption on which return periods are based (Milly et al., 2008; Salas and Obeysekera, 2014; Read and Vogel, 2015).

The objective of this study was to assess the overtopping reliability of hypothetical levees on three rivers with long, openly available daily discharge records. The specific aims were to (i) characterise the monthly and annual maximum discharges and fit stationary, seasonal and non-stationary GEV models; (ii) compare FORM, SORM, MCS and IS for a levee overtopping limit state; (iii) identify the most important random variables; and (iv) quantify how freeboard, hydraulic uncertainty, GEV parameter uncertainty, distribution choice and non-stationarity affect the failure probability.

2. Materials and Methods

2.1. Study Rivers and Data

Three daily discharge records distributed with open-source R packages were used (Figure 1; Table 1). The Platte River at Brady, Nebraska (USGS gauge

06766000; March 1939 to September 1991) is a regulated river on the Great Plains, where upstream reservoirs and diversions have changed the flow regime (Williams, 1978); its record is distributed with the *lmomco* package (Asquith, 2026). The Choptank River near Greensboro, Maryland (USGS 01491000; October 1979 to September 2011) is a Coastal Plain river whose daily discharge is distributed with the *EGRET* package (Hirsch and De Cicco, 2015). The Ngaruroro River in New Zealand (September 1963 to December 2000) drains a mountainous catchment; its record is one of the example data sets of the WMO Manual on Low-flow Estimation and Prediction (Gustard and Demuth, 2009) and is distributed with the *lfstat* package (Koffler et al., 2025). Platte discharges were converted from $\text{ft}^3 \text{s}^{-1}$ to $\text{m}^3 \text{s}^{-1}$.

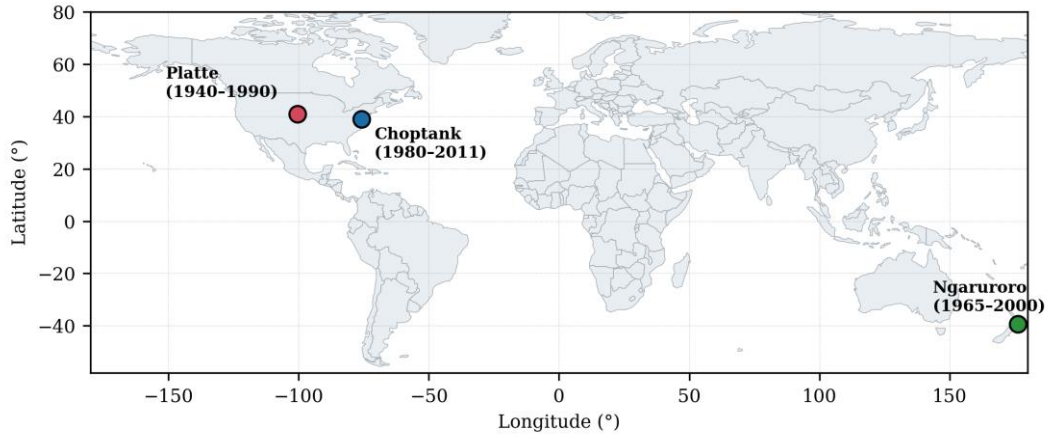


Figure 1. Approximate locations of the three gauging stations and the periods of the annual maximum series

The monthly maximum daily discharge was extracted for every month with at least 20 valid days. The water year of each river was defined to start in the month with the lowest mean monthly maximum: November for the Platte, October for the Choptank and February for the Ngaruroro. The annual maximum was

taken as the largest monthly maximum in water years with 12 valid months, which gave 51, 32 and 31 annual maxima, respectively. Trends were tested with the Mann–Kendall test with the Hamed and Rao (1998) variance correction and the Sen slope, and change points with the Pettitt (1979) test.

Table 1. Discharge records and statistics of the annual maximum series (AMS)

River	Country	Daily record	AMS water years (n)	Mean Q ($\text{m}^3 \text{s}^{-1}$)	Mean AMS ($\text{m}^3 \text{s}^{-1}$)	C_v	τ_3	τ_4	MK Z (p)	Sen slope ($\text{m}^3 \text{s}^{-1} \text{yr}^{-1}$)	Pettitt year (p)	r_1 monthly max.
Platte	USA	1939-03 – 1991-09	1940–1990 (51)	22.0	145	0.99	0.461	0.199	2.06 (0.040)	+0.91	1965 (0.049)	0.72
Choptank	USA	1979-10 – 2011-09	1980–2011 (32)	4.1	66	0.73	0.280	0.230	3.41 (0.001)	+2.22	1994 (0.003)	0.48
Ngaruroro	New Zealand	1963-09 – 2000-12	1965–2000 (31)	17.2	177	0.37	0.041	0.084	0.58 (0.563)	+0.58	1986 (0.574)	0.09

τ_3 , τ_4 = sample L-skewness and L-kurtosis; MK = Mann–Kendall test with Hamed–Rao correction; r_1 monthly max. = lag-one autocorrelation of log-transformed monthly maxima standardised by calendar month.

2.2. Flood Frequency Models

The GEV distribution has the cumulative distribution function

$$F(x) = \exp \left\{ - \left[1 - \frac{k(x-\xi)}{\alpha} \right]^{1/k} \right\} \quad (1)$$

where ξ , α and k are the location, scale and shape parameters (Hosking et al., 1985). A negative k gives a

heavy upper tail, and a positive k gives an upper bound at $\xi + \alpha/k$. The parameters were estimated from the sample L-moments with the approximation of Hosking et al. (1985):

$$k \approx 7.8590 c + 2.9554 c^2, \quad c = \frac{2}{3+\tau_3} - \frac{\ln 2}{\ln 3} \quad (2)$$

$$\alpha = \frac{\lambda_2 k}{(1-2^{-k})\Gamma(1+k)}, \quad \xi = \lambda_1 + \frac{\alpha}{k} [\Gamma(1+k) - 1] \quad (3)$$

where λ_1 , λ_2 and τ_3 are the sample L-mean, L-scale and L-skewness (Hosking, 1990). The two-parameter Gumbel distribution ($k = 0$) was also fitted by L-moments. Fit was checked with the L-moment ratio diagram, the Anderson–Darling (AD) statistic and the probability-plot correlation coefficient (PPCC) with Gringorten (1963) plotting positions. Parameter uncertainty was quantified with 1000 parametric bootstrap samples of the same length as the record.

Seasonal model. A GEV distribution was fitted by L-moments to the maxima of each calendar month m , with k limited to $[-0.5, 0.5]$, the range of the approximation. If monthly maxima are independent, the annual maximum distribution is the product of the monthly distributions:

$$F_A(x) = \prod_{m=1}^{12} F_m(x) \quad (4)$$

Non-stationary model. To test for a trend, a GEV with a location parameter linear in time, $\mu(t) = \mu_0 + \mu_1 t$, was fitted by generalized maximum likelihood (GMLE) with a beta(6, 9) prior on k , which keeps the shape within a physically reasonable range (Martins and Stedinger, 2000; El Adlouni et al., 2007). It was compared with a stationary GMLE fit by the deviance statistic (χ^2 with one degree of freedom) and the Akaike information criterion (Coles, 2001).

2.3. Levee Overtopping Limit State

A hypothetical levee reach was defined for each river (Figure 2; Table 2). The channel was treated as wide and rectangular, so the Manning equation gives the flow depth h for a discharge Q :

$$h = \left(\frac{n Q}{B \sqrt{S}} \right)^{3/5} \quad (5)$$

where n is the Manning roughness, B the channel width and S the bed slope (Chow, 1959). The nominal width was estimated from the regime relationship $B_0 = 4.8 Q_2^{0.5}$ (Lacey, 1930), where Q_2 is the median annual maximum. The nominal slopes and roughness coefficients were chosen to represent a braided sand-bed plains river (Platte), a low-gradient vegetated Coastal Plain channel (Choptank) and a steep gravel-bed river (Ngaruroro). The nominal crest height above the bed was set to the Manning depth of the 100-year GEV discharge plus a freeboard of 0.5 m. The limit-state function is:

$$g(\mathbf{X}) = Z_c - \theta \left(\frac{n Q}{B \sqrt{S}} \right)^{3/5} \quad (6)$$

where θ is a multiplicative model-error factor of the stage calculation and Z_c is the actual crest height. Failure (overtopping) occurs when $g \leq 0$. The six random variables and their distributions are listed in Table 2. Roughness, slope and model error were lognormal, with coefficients of variation (CoV) of 0.20, 0.15 and 0.10. Width was normal with a CoV of 0.05, and the crest height was normal with a standard deviation of 0.10 m to represent settlement and survey error. The annual maximum discharge followed the fitted GEV. All variables were assumed independent.

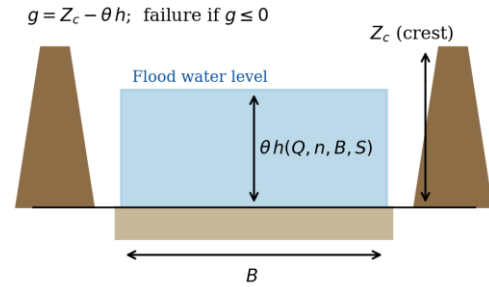


Figure 2. Definition sketch of the levee overtopping limit state

Table 2. Random variables (means for n , S , B , θ and Z_c) and nominal design values of the hypothetical levees

Variable	Distribution	Platte	Choptank	Ngaruroro
Q ($\text{m}^3 \text{s}^{-1}$)	GEV (L-moments)	$\xi = 73.2, \alpha = 57.6, k = -0.409$	$\xi = 43.2, \alpha = 29.5, k = -0.165$	$\xi = 151.3, \alpha = 63.4, k = 0.212$
n ($\text{s m}^{-1/3}$)	Lognormal, CoV 0.20	0.030	0.040	0.035
S (–)	Lognormal, CoV 0.15	0.0012	0.0005	0.0040
B (m)	Normal, CoV 0.05	47.0	35.4	63.3
θ (–)	Lognormal, CoV 0.10	1.00	1.00	1.00
Z_c (m)	Normal, SD 0.10 m	5.73	5.04	2.42
Q_{100} ($\text{m}^3 \text{s}^{-1}$)	Design discharge	855	246	338
h_{100} (m)	Nominal design depth	5.23	4.54	1.92
Q_{cap} ($\text{m}^3 \text{s}^{-1}$)	Nominal overtopping discharge	996	293	497

2.4. Reliability Methods

The random vector $\mathbf{X}$ was mapped to independent standard normal variables $\mathbf{U}$ with the marginal transformation $X_i = F_i^{-1}[\Phi(U_i)]$. The reliability index β is the distance from the origin to the most probable failure point $\mathbf{u}^*$ on the surface $G(\mathbf{u}) = 0$ (Hasofer and Lind, 1974):

$$\beta = \min \|\mathbf{u}\| \quad \text{subject to} \quad G(\mathbf{u}) = 0 \quad (7)$$

It was found with the HL-RF algorithm (Rackwitz and Fiessler, 1978), using central-difference gradients and a merit-function line search to ensure convergence:

$$\mathbf{u}_{k+1} = \frac{\nabla G_k^T \mathbf{u}_k - G_k}{\|\nabla G_k\|^2} \nabla G_k \quad (8)$$

The FORM estimate is $P_f \approx \Phi(-\beta)$, and the importance factor of variable i is α_i^2 , where $\alpha = -\nabla G / \|\nabla G\|$ at the design point. For SORM, the Hessian of G at $\mathbf{u}^*$ was computed by finite differences and rotated

so that the last axis points to the design point. The main curvatures κ_i were the eigenvalues of the remaining $(n - 1)$ -dimensional block divided by $\nabla^2 G$. Two SORM estimates were computed, those of Breitung (1984) and Hohenbichler and Rackwitz (1988):

$$P_f^B = \Phi(-\beta) \prod_{i=1}^{n-1} (1 + \beta \kappa_i)^{-1/2} \quad (9)$$

$$P_f^{HR} = \Phi(-\beta) \prod_{i=1}^{n-1} \left(1 + \frac{\phi(\beta)}{\Phi(-\beta)} \kappa_i\right)^{-1/2} \quad (10)$$

Crude MCS used 5×10^6 samples, with a coefficient of variation of the estimate of

$$\text{CoV}(\hat{P}_f) = \sqrt{\frac{1 - \hat{P}_f}{N \hat{P}_f}} \quad (11)$$

Importance sampling drew $N = 2 \times 10^4$ samples from a unit-variance normal density centred on $\mathbf{u}^*$ and weighted each failure by $\phi(\mathbf{u})/\phi(\mathbf{u} - \mathbf{u}^*)$ (Melchers and Beck, 2018). The algorithms were first checked against cases with exact solutions: a linear limit state with normal variables ($\beta = 2$), a lognormal ratio ($\beta = 2.834$) and a parabolic limit state with known curvature. FORM reproduced the exact β to machine precision in the linear

cases, and SORM-HR was within 1% of the exact probability of the parabolic case. The lifetime failure probability for a 50-year service life was $1 - (1 - P_f)^{50}$, assuming independent years.

2.5. Sensitivity and Uncertainty Analyses

Five further analyses were made: (i) the freeboard was varied from 0 to 1.5 m (FORM, SORM and MCS with 2×10^6 samples); (ii) the CoVs of n , S , B and θ and the SD of Z_c were scaled by factors from 0 to 2; (iii) FORM was repeated for 300 bootstrap GEV parameter sets to obtain the distribution of P_f due to parameter uncertainty and the predictive (mean) failure probability; (iv) the GEV was replaced by the Gumbel distribution; and (v) the failure probability was computed with the non-stationary GMLE model for each year of record. Monthly failure probabilities were also computed with the monthly GEV distributions. All calculations were coded in Python with NumPy, SciPy and lmoments3.

3. Results and Discussion

3.1. Monthly and Annual Maximum Discharges

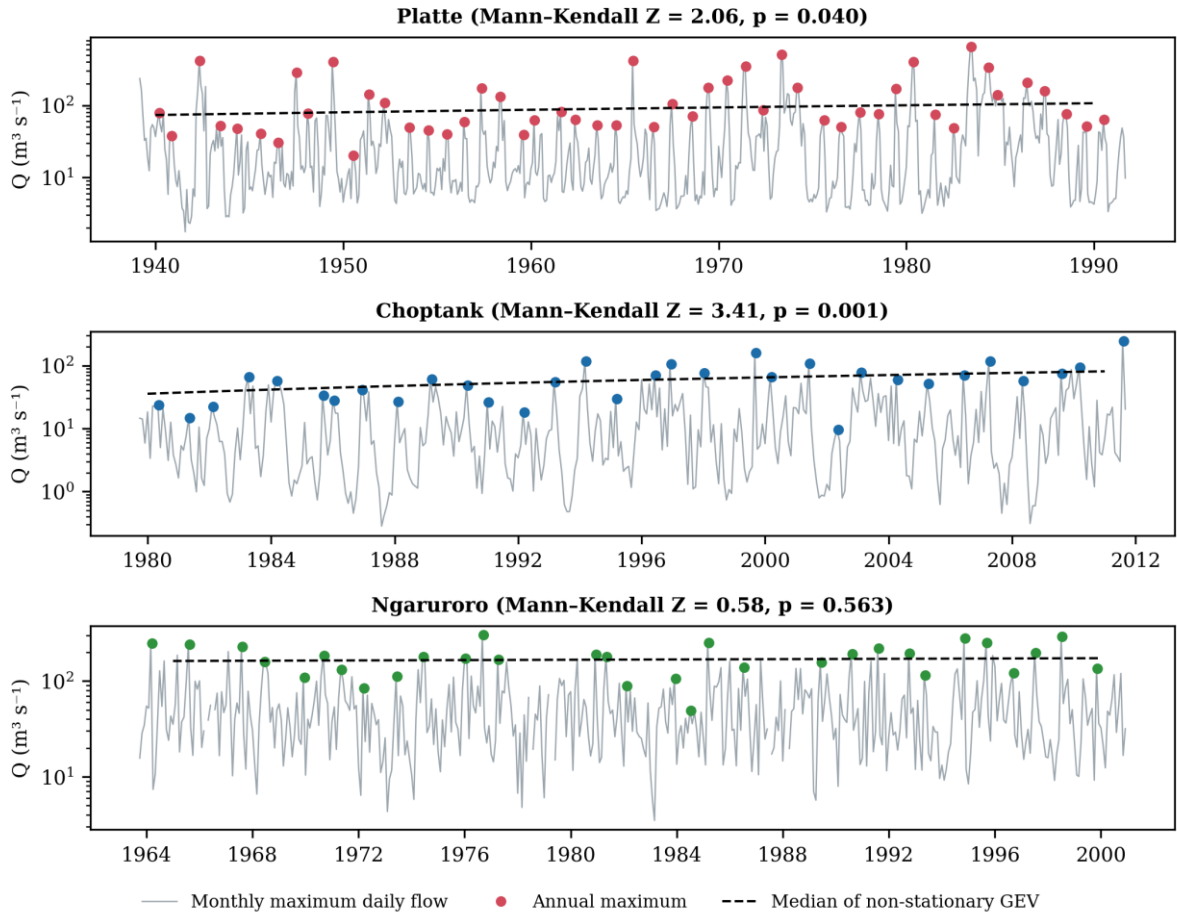


Figure 3. Monthly maximum daily discharges, annual maxima and the median of the non-stationary GEV model

The mean annual maximum was $145 \text{ m}^3 \text{ s}^{-1}$ at the Platte, $66 \text{ m}^3 \text{ s}^{-1}$ at the Choptank and $177 \text{ m}^3 \text{ s}^{-1}$ at the Ngaruroro (Table 1; Figure 3). The Platte and Choptank series were highly variable ($C_v = 0.99$ and 0.73) and strongly skewed ($\tau_3 = 0.46$ and 0.28). The Ngaruroro

series was less variable ($C_v = 0.37$) and almost symmetric ($\tau_3 = 0.04$).

Floods had a clear season at the Platte, where 38 of the 51 annual maxima occurred in May–July, with snowmelt and convective rainfall on the Great Plains

(Figure 4). Choptank maxima occurred mostly in late winter and spring (8 of 32 in March). Ngaruroro maxima were spread over the year. The monthly maxima were

serially correlated at the Platte ($r_1 = 0.72$) and Choptank (0.48) but hardly at the Ngaruroro (0.09).

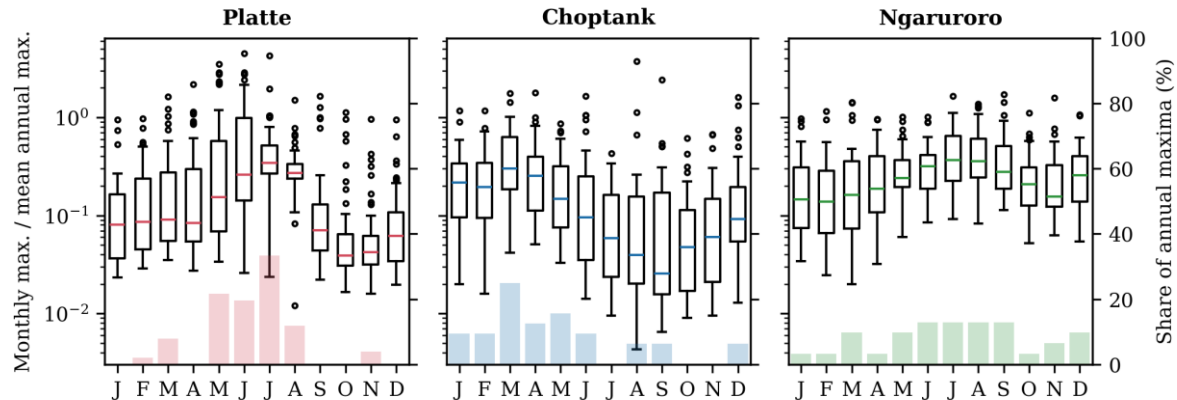


Figure 4. Distribution of monthly maxima by calendar month (boxes, relative to the mean annual maximum) and share of annual maxima in each month (bars)

The Choptank annual maxima had a significant upward trend ($Z = 3.41$, $p = 0.001$; Sen slope $+2.2 \text{ m}^3 \text{ s}^{-1} \text{ yr}^{-1}$) with a change point in 1994 ($p = 0.003$). The Platte trend was weakly significant ($p = 0.040$), and the Ngaruroro showed no trend. The non-stationary GMLE model confirmed these results. The deviance test was significant for the Choptank ($D = 6.38$, $p = 0.012$; $\mu_1 = 14.6 \text{ m}^3 \text{ s}^{-1}$ per decade) and marginal for the Platte ($p = 0.048$), and the AIC favoured the non-stationary model at both rivers (see Table 5).

3.2. Flood Frequency Analysis

In the L-moment ratio diagram, the Choptank lay close to the GLO curve and slightly above the GEV curve, the Ngaruroro below all three-parameter curves (light tail) and the Platte well below the GEV curve and close to the PE3 curve (Figure 5). The GEV fitted the Choptank and Ngaruroro well ($AD = 0.27$ and 0.17 ; $PPCC \geq 0.988$) and better than the Gumbel at all rivers (Table 3). The Platte fit was poorer ($AD = 1.20$, $PPCC = 0.956$), which is consistent with a mixed flood

population in a regulated river. The fitted shape parameters were -0.409 (Platte), -0.165 (Choptank) and 0.212 (Ngaruroro). The positive Ngaruroro value implies an upper bound of $450 \text{ m}^3 \text{ s}^{-1}$.

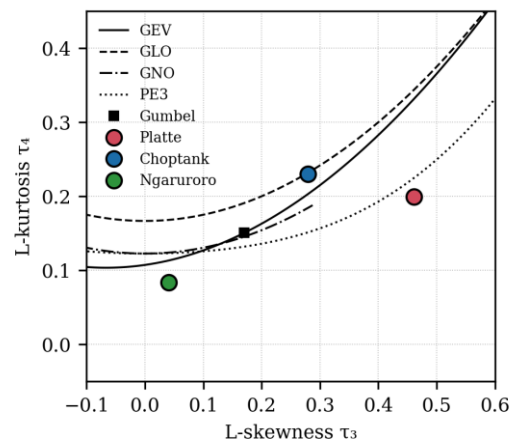


Figure 5. L-moment ratio diagram of the annual maximum series

Table 3. Goodness of fit and flood quantiles ($\text{m}^3 \text{ s}^{-1}$)

River	AD GEV/Gumbel	PPCC GEV/Gumbel	Q_{10} GEV	Q_{100} GEV (90% CI)	Q_{500} GEV	Q_{100} Gumbel	Q_{100} seasonal	Q_{100} GMLE
Platte	1.20 / 3.22	0.956 / 0.947	286	855 (445–1413)	1717	549	1179	602
Choptank	0.27 / 0.38	0.988 / 0.968	124	246 (149–404)	363	208	273	233
Ngaruroro	0.17 / 0.36	0.994 / 0.975	265	338 (279–406)	370	394	700	392

GEV and Gumbel fitted by L-moments; CI from 1000 parametric bootstrap samples; seasonal = product of 12 monthly GEV distributions; GMLE = stationary generalized maximum likelihood GEV.

The 100-year GEV discharge was 855, 246 and $338 \text{ m}^3 \text{ s}^{-1}$ (Figure 6). The 90% bootstrap intervals were wide for the heavy-tailed rivers: $445\text{--}1413 \text{ m}^3 \text{ s}^{-1}$ at the Platte and $149\text{--}404 \text{ m}^3 \text{ s}^{-1}$ at the Choptank. The estimator also mattered: the stationary GMLE gave a 100-year discharge of $602 \text{ m}^3 \text{ s}^{-1}$ at the Platte, because the prior limited the shape parameter to -0.34 .

The seasonal product model gave larger upper quantiles than the AMS model at all rivers, and much larger at the Ngaruroro (700 against $338 \text{ m}^3 \text{ s}^{-1}$). Two

reasons explain this. First, the monthly GEVs were fitted to all monthly maxima, most of which are not floods, so their upper tails were strongly skewed and several shape parameters reached the lower limit of -0.5 (supplementary data file `monthly_gev.csv`). Second, the product formula assumes independence between months, which the positive correlation of monthly maxima at the Platte and Choptank contradicts. The seasonal model is therefore useful to describe when floods occur, but not to estimate annual design quantiles.

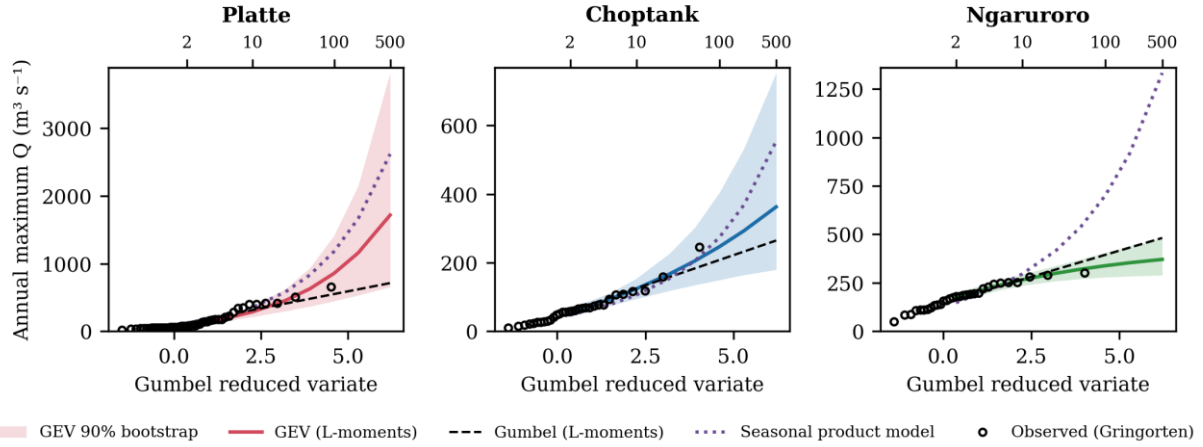


Figure 6. Flood frequency curves: GEV with 90% bootstrap band, Gumbel and seasonal product model (top axis: return period in years)

3.3. Comparison of Reliability Methods

Table 4. Annual overtopping probability by the reliability methods (freeboard 0.5 m)

River	Method	g-calls	P_f ($\times 10^{-3}$)	β	Error vs MCS (%)	Time (s)
Platte	Nominal (Q only)	—	7.08	2.453	—	—
	FORM	120	7.99	2.409	-2.2	0.13
	SORM-Breitung	204	8.13	2.403	-0.4	0.20
	SORM-HR	204	8.15	2.402	-0.2	0.20
	IS (CoV 0.012)	20,120	8.13	2.403	-0.4	0.14
	MCS (CoV 0.005)	5,000,000	8.16	2.401	ref.	2.92
Choptank	Nominal (Q only)	—	4.97	2.578	—	—
	FORM	92	7.68	2.424	-2.1	0.09
	SORM-Breitung	176	7.82	2.417	-0.4	0.16
	SORM-HR	176	7.84	2.416	-0.1	0.16
	IS (CoV 0.012)	20,092	7.93	2.412	+1.0	0.10
	MCS (CoV 0.005)	5,000,000	7.85	2.416	ref.	2.76
Ngaruroro	Nominal (Q only)	—	0.00	∞	—	—
	FORM	195	5.12	2.568	+15.1	0.19
	SORM-Breitung	279	4.53	2.610	+2.0	0.26
	SORM-HR	279	4.48	2.614	+0.7	0.26
	IS (CoV 0.013)	20,195	4.39	2.620	-1.2	0.20
	MCS (CoV 0.007)	5,000,000	4.45	2.616	ref.	2.77

All methods gave annual failure probabilities between 4×10^{-3} and 9×10^{-3} , i.e. reliability indices of 2.40–2.62 (Table 4; Figure 7). For the Platte and Choptank, FORM underestimated the MCS probability by only 2.2% and 2.1%, and the SORM estimates were within the MCS confidence interval. The curvatures of the limit state were close to zero, so the surface was almost flat in standard normal space. For the Ngaruroro, FORM overestimated the MCS probability by 15%. The SORM analysis revealed a main curvature of 0.125, associated with the bounded GEV tail, and the SORM-HR estimate was within 0.7% of MCS. Importance sampling with 2×10^4 samples reached a CoV of 0.012–0.013, and its estimates differed from MCS by less than 2%.

The computational effort differed by four orders of magnitude. FORM converged in 6–14 iterations with 92–195 limit-state evaluations, and SORM needed 84 further evaluations. MCS needed 5×10^6 evaluations to reach a CoV of 0.005–0.007. With the vectorised limit state used here, even MCS took only about 3 s. For levee systems with numerical hydraulic models, however, each evaluation may take minutes, and the efficiency of FORM, SORM and IS becomes decisive.

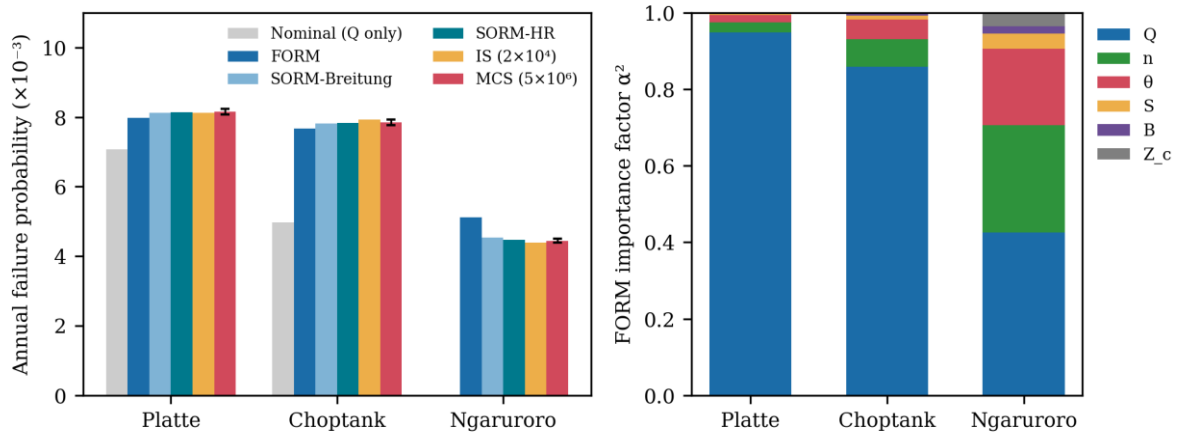


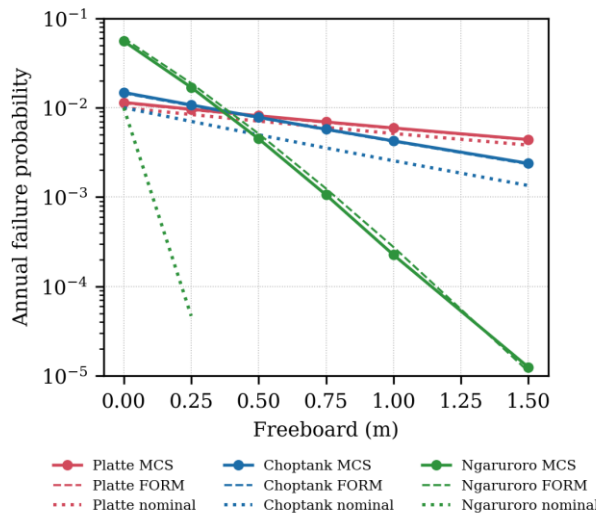
Figure 7. Annual failure probability by method (MCS with 95% confidence interval) and FORM importance factors

3.4. Design Point and Importance Factors

Discharge dominated the failure probability at the Platte ($\alpha^2 = 0.95$) and Choptank (0.86) (Figure 7). Roughness ranked second (0.026 and 0.073), followed by the model error. At the Ngaruroro, discharge explained only 43% of the importance; roughness (28%), model error (20%), slope (4%) and crest level (4%) accounted for the rest. The design-point discharge at the Ngaruroro ($293 \text{ m}^3 \text{ s}^{-1}$) was below the nominal overtopping discharge ($497 \text{ m}^3 \text{ s}^{-1}$), which itself exceeded the upper bound of the fitted GEV. Failure at this site therefore requires a combination of a large flood with higher-than-expected roughness ($n^* = 0.045$) and a positive model error ($\theta^* = 1.12$). With discharge as the only random variable, the nominal failure probability was zero.

At the Platte and Choptank, the nominal overtopping discharge corresponded to return periods of 141 and 201 years. Including hydraulic uncertainty increased the failure probability from 7.08×10^{-3} to 8.16×10^{-3} (Platte) and from 4.97×10^{-3} to 7.85×10^{-3} (Choptank). The 0.5 m freeboard therefore did not fully offset the hydraulic uncertainty. Over a 50-year service life, the probability of at least one overtopping was 0.20–0.34.

3.5. Sensitivity to Freeboard and Hydraulic Uncertainty



Increasing the freeboard reduced the failure probability at very different rates (Figure 8). At the Ngaruroro, the MCS probability fell from 0.055 without freeboard to 2.3×10^{-4} with 1 m, because the stage–discharge curve of a steep river is flat and the discharge has an upper bound. At the Platte, 1 m of freeboard reduced the probability only from 0.0115 to 0.0059, because the heavy GEV tail makes large discharges relatively frequent. Without freeboard, the failure probability exceeded the nominal 0.01 at all rivers: 0.0115 (Platte), 0.0148 (Choptank) and 0.0555 (Ngaruroro). A uniform freeboard thus provides very different safety levels on different rivers. FORM followed the MCS results closely over the whole range, and the largest relative deviations occurred at the Ngaruroro.

Scaling the hydraulic uncertainties had the opposite pattern. Doubling all CoVs increased the FORM probability by a factor of 1.4 at the Platte, 2.3 at the Choptank and 8.1 at the Ngaruroro. Halving them reduced the Ngaruroro probability to 1.9×10^{-4} . The reliability of levees on rivers with light-tailed floods is therefore controlled by the hydraulic parameters, and field calibration of roughness is the most effective way to reduce uncertainty there.

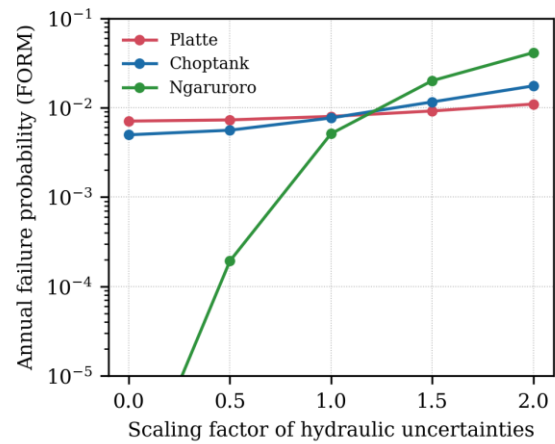


Figure 8. Failure probability versus freeboard (left) and versus a scaling factor applied to the hydraulic uncertainties (right)

3.6. Parameter Uncertainty, Distribution Choice and Non-Stationarity

Table 5. Effect of GEV parameter uncertainty, distribution choice and non-stationarity on the annual failure probability ($\times 10^{-3}$)

River	Point (FORM)	Bootstrap median (90% interval)	Predictive mean	Gumbel MCS (α^2)	μ , per decade (p)	AIC stat./non-stat.	Non-stat. first / last year	Stationary GMLE
Platte	7.99	5.71 (0.55–18.59)	7.19	0.78 (0.74)	+6.7 (0.048)	604.5 / 602.5	2.68 / 2.98	3.13
Choptank	7.68	5.99 (0.31–23.69)	8.41	3.50 (0.73)	+14.6 (0.012)	325.9 / 321.6	1.91 / 5.62	6.26
Ngaruroro	5.12	5.27 (1.69–12.20)	5.85	8.33 (0.63)	+3.2 (0.778)	351.5 / 353.4	7.75 / 9.50	8.69

Bootstrap: FORM for 300 parametric bootstrap GEV parameter sets. Non-stationary and stationary GMLE results use the levee designed with the L-moment GEV. First/last years: 1940/1990 (Platte), 1980/2011 (Choptank), 1965/2000 (Ngaruroro).

Parameter uncertainty had a much larger effect on the failure probability than the choice of reliability method (Table 5; Figure 9). The 90% bootstrap intervals

spanned $0.55\text{--}18.59 \times 10^{-3}$ at the Platte and $0.31\text{--}23.69 \times 10^{-3}$ at the Choptank, i.e. factors of 34 and 78. The Ngaruroro interval was narrower (1.69--

12.20×10^{-3}). The predictive failure probability, which integrates over parameter uncertainty, was close to the point estimate at all rivers. The bootstrap distributions were skewed, and their medians were below the point estimates at the Platte and Choptank. This agrees with the need to report epistemic uncertainty separately from natural variability (Merz and Thielen, 2005).

Replacing the GEV by the Gumbel distribution changed the results in both directions. At the heavy-tailed Platte and Choptank, the Gumbel failure probability was 10 and 2.2 times smaller. At the Ngaruroro it was 1.9 times larger, because the Gumbel has no upper bound. Distribution choice is therefore a major source of uncertainty in levee reliability, particularly when short records suggest a bounded tail.

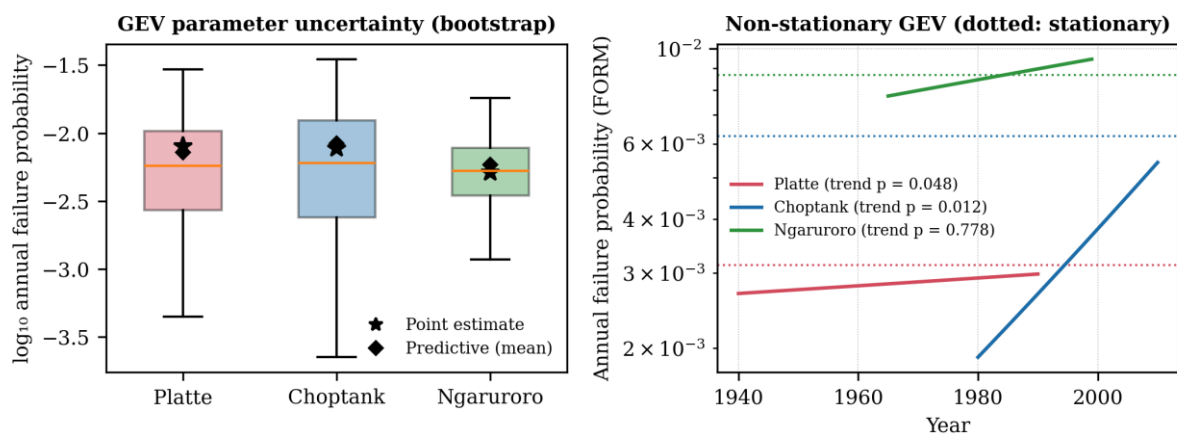


Figure 9. Failure probability under GEV parameter uncertainty (left; box = interquartile range, whiskers = 5–95%) and under the non-stationary GEV model (right)

3.7. Seasonal Distribution of Failure Probability

The monthly GEV models located the months of highest overtopping risk (Figure 10). The largest monthly probability occurred in June at the Platte (7.01×10^{-3} , 46% of the sum of the monthly probabilities), in June at the Choptank and in September at the Ngaruroro. At the Platte, the high-risk months matched the observed flood season. At the Choptank, the high June value results from the heavy upper tail fitted to the June maxima (shape parameter at the lower limit) and does not match the usual timing of annual maxima in March. The annual failure probabilities combined from the monthly values (15.27 , 9.94 and 30.86×10^{-3}) were 1.3–6 times the AMS-based FORM values. This confirms that the seasonal product model overestimates risk. It is best used to schedule inspection and emergency preparedness, not to estimate annual reliability.

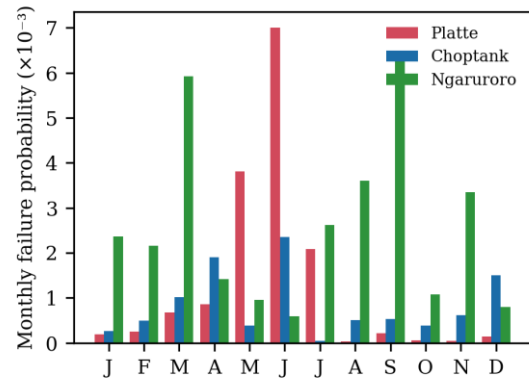


Figure 10. Monthly overtopping probability (FORM) computed with calendar-month GEV distributions

3.8. Limitations

The levees and channel geometries are hypothetical, and a wide rectangular channel with uniform flow is a strong simplification of real river reaches. Daily mean discharges underestimate instantaneous peaks, particularly in small, flashy catchments, and the Platte record reflects changing regulation. Only overtopping was considered; breaching by piping, slope instability or erosion can dominate levee failure (Vorogushyn et al., 2009). The random variables were assumed

independent, whereas roughness and discharge may be correlated. Finally, the annual failure probability assumes independent years. Despite these limitations, the framework and data are open and can be transferred to real levee systems with site-specific hydraulic models.

4. Conclusions

The overtopping reliability of hypothetical levees on three rivers was assessed with FORM, SORM, importance sampling and Monte Carlo simulation, using GEV models of annual and monthly maximum discharges. The main conclusions are:

(1) FORM was within 2% of MCS for heavy-tailed discharge distributions and within 15% for a bounded one. SORM-HR reduced the error to less than 1% with fewer than 300 limit-state evaluations. Importance sampling gave unbiased estimates with a CoV of about 0.01 using 2×10^4 samples.

(2) Hydraulic uncertainty increased the failure probability of levees designed for the 100-year flood with 0.5 m freeboard. Over 50 years, the probability of at least one overtopping was 0.20–0.34.

(3) Discharge dominated the reliability of levees on rivers with heavy-tailed floods ($\alpha^2 = 0.86$ –0.95), whereas roughness and model error dominated on the river with a bounded flood distribution. A uniform freeboard therefore gives unequal safety.

(4) GEV parameter uncertainty (90% intervals spanning more than an order of magnitude at two rivers), distribution choice and non-stationarity (a threefold increase at the Choptank over 32 years) affected the failure probability far more than the choice of reliability method.

(5) Monthly maxima describe the seasonality of flood risk, but a product of monthly distributions overestimates annual risk when monthly maxima are correlated.

Data Availability Statement

The daily discharge data are openly available in the R packages *lmomco* (Platte River, USGS 06766000; <https://CRAN.R-project.org/package=lmomco>), *EGRET* (Choptank River, USGS 01491000; <https://CRAN.R-project.org/package=EGRET>) and *lfstat* (Ngaruroro River; <https://CRAN.R-project.org/package=lfstat>). Scripts are available from the corresponding author on request.

Conflict of Interest

The authors declare no conflict of interest.

References

- Apel, H., Thielen, A.H., Merz, B., & Blöschl, G. (2004). Flood risk assessment and associated uncertainty. *Natural Hazards and Earth System Sciences*, 4, 295–308. <https://doi.org/10.5194/nhess-4-295-2004>
- Asquith, W.H. (2026). *lmomco*: L-moments, censored L-moments, trimmed L-moments, L-comoments, and many distributions. R package version 2.5.7. <https://CRAN.R-project.org/package=lmomco>
- Breitung, K. (1984). Asymptotic approximations for multinormal integrals. *Journal of Engineering Mechanics*, 110(3), 357–366. [https://doi.org/10.1061/\(ASCE\)0733-9399\(1984\)110:3\(357\)](https://doi.org/10.1061/(ASCE)0733-9399(1984)110:3(357))
- Chow, V.T. (1959). *Open-Channel Hydraulics*. McGraw-Hill, New York.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.
- El Adlouni, S., Ouarda, T.B.M.J., Zhang, X., Roy, R., & Bobée, B. (2007). Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resources Research*, 43, W03410. <https://doi.org/10.1029/2005WR004545>
- Gringorten, I.I. (1963). A plotting rule for extreme probability paper. *Journal of Geophysical Research*, 68(3), 813–814.
- Gustard, A., & Demuth, S. (Eds.) (2009). *Manual on Low-flow Estimation and Prediction*. Operational Hydrology Report No. 50, WMO-No. 1029, World Meteorological Organization, Geneva.
- Hamed, K.H., & Rao, A.R. (1998). A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology*, 204, 182–196.
- Hasofer, A.M., & Lind, N.C. (1974). Exact and invariant second-moment code format. *Journal of the Engineering Mechanics Division*, 100(1), 111–121.
- Hirsch, R.M., & De Cicco, L.A. (2015). User guide to *Exploration and Graphics for RivEr Trends (EGRET)* and *dataRetrieval*: R packages for hydrologic data. U.S. Geological Survey Techniques and Methods, book 4, chap. A10. <https://doi.org/10.3133/tm4A10>
- Hohenbichler, M., & Rackwitz, R. (1988). Improvement of second-order reliability estimates by importance sampling. *Journal of Engineering Mechanics*, 114(12), 2195–2199.
- Hosking, J.R.M. (1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B*, 52(1), 105–124.
- Hosking, J.R.M., & Wallis, J.R. (1997). *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, Cambridge.
- Hosking, J.R.M., Wallis, J.R., & Wood, E.F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3), 251–261.
- Koffler, D., Gauster, T., & Laaha, G. (2025). *lfstat*: Calculation of low flow statistics for daily stream flow data. R package version 0.9.13. <https://CRAN.R-project.org/package=lfstat>
- Lacey, G. (1930). Stable channels in alluvium. *Minutes of the Proceedings of the Institution of Civil Engineers*, 229, 259–292.
- Martins, E.S., & Stedinger, J.R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, 36(3), 737–744. <https://doi.org/10.1029/1999WR000330>

- Melchers, R.E., & Beck, A.T. (2018). *Structural Reliability Analysis and Prediction* (3rd ed.). John Wiley & Sons, Chichester.
- Merz, B., & Thieken, A.H. (2005). Separating natural and epistemic uncertainty in flood frequency analysis. *Journal of Hydrology*, 309, 114–132. <https://doi.org/10.1016/j.jhydrol.2004.11.015>
- Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., & Stouffer, R.J. (2008). Stationarity is dead: whither water management? *Science*, 319, 573–574. <https://doi.org/10.1126/science.1151915>
- Pettitt, A.N. (1979). A non-parametric approach to the change-point problem. *Journal of the Royal Statistical Society, Series C*, 28(2), 126–135.
- Rackwitz, R., & Fiessler, B. (1978). Structural reliability under combined random load sequences. *Computers & Structures*, 9(5), 489–494. [https://doi.org/10.1016/0045-7949\(78\)90046-9](https://doi.org/10.1016/0045-7949(78)90046-9)
- Read, L.K., & Vogel, R.M. (2015). Reliability, return periods, and risk under nonstationarity. *Water Resources Research*, 51, 6381–6398. <https://doi.org/10.1002/2015WR017089>
- Salas, J.D., & Obeysekera, J. (2014). Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. *Journal of Hydrologic Engineering*, 19(3), 554–568. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000820](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000820)
- Tung, Y.-K., & Mays, L.W. (1981). Risk models for flood levee design. *Water Resources Research*, 17(4), 833–841. <https://doi.org/10.1029/WR017i004p00833>
- Vorogushyn, S., Merz, B., & Apel, H. (2009). Development of dike fragility curves for piping and micro-instability breach mechanisms. *Natural Hazards and Earth System Sciences*, 9, 1383–1401. <https://doi.org/10.5194/nhess-9-1383-2009>
- Williams, G.P. (1978). The case of the shrinking channels—the North Platte and Platte Rivers in Nebraska. U.S. Geological Survey Circular 781. <https://doi.org/10.3133/cir781>
- Wood, E.F. (1977). An analysis of flood levee reliability. *Water Resources Research*, 13(3), 665–671. <https://doi.org/10.1029/WR013i003p00665>