On the interpretation of return values

David Randell, Jenny Wadsworth, Jonathan Tawn, Philip Jonathan
Shell & Lancaster University

Overview

- Return value
- Problem: incorporating estimation (epistemic) uncertainty
- Possible approaches
- Simulation study
- Theoretical properties
- Conclusions

What is a return value?

- Random variable A represents the maximum value of some physical quantity X per annum
- The N-year return value x_N of X is then defined by the equation

$$F_A(x_N) = \Pr(A \leq x_N) = 1 - \frac{1}{N}$$

- Typically $N \in [10^2, 10^8]$ years

What is a return value?

- Random variable A_N represents the N-year maximum value of X
- The N-year return value x'_N of X can be found from F_{A_N} for large N since

$$F_A(x_N) = 1 - \frac{1}{N} \Rightarrow F_{A_N}(x_N) = \left(1 - \frac{1}{N}\right)^N \approx \exp(-1)$$

- Use $F_{A_N}(x'_N) = \exp(-1)$ to define an alternative return value x'_N

Estimating a return value

- To estimate x_N, we need knowledge of the distribution function F_A of the annual maximum
- We might estimate F_A using extreme value analysis on a sample of independent observations of A
- Typically more efficient to estimate the distribution $F_{X|X>\psi}$ of threshold exceedances of X above some high threshold ψ using a sample of independent observations of X, and use this in turn to estimate F_A and x_N
- How is this done?

Estimating a return value

- Asymptotic theory suggests for large ψ that

$$F_{X|X>\psi}(x|\psi,\sigma,\xi) = 1 - \left(1 + \frac{\xi}{\sigma}(x-\psi)\right)_+^{-1/\xi}$$

  for $x > \psi$, threshold $\psi \in (-\infty,\infty)$, shape $\xi \in (-\infty,\infty)$ and scale $\sigma \in (0,\infty)$
- The full distribution of X is $F_X(x) = \tau + (1-\tau)F_{X|X>\psi}(x)$ where $\Pr(X \leq \psi) = \tau$
- Thus

$$F_A(x) = \Pr(A \leq x) = \sum_{k=0}^{\infty} f_C(k) F_X^k(x)$$

  where C is the number of occurrences of X per annum, with probability mass function f_C to be estimated (say with a Poisson model with parameter λ)

What's the problem?

What's the issue?

- x_N (or x'_N) can be estimated easily in the absence of uncertainty

- In reality, we estimate parameters λ, ψ, σ and ξ from a sample of data
- How does epistemic uncertainty affect return value estimates?

- A number of plausible estimators for return values under uncertainty
- Different estimators perform differently (bias and variance)
- Which estimators are likely to perform reasonably in fairly general circumstances?
- Is it sensible even to estimate return values?

Incorporating uncertainty

- If a distribution $F_{Y|\mathbf{Z}}$ of random variable Y is known conditional on random variables $\mathbf{Z}$, and the joint density $f_{\mathbf{Z}}$ of $\mathbf{Z}$ is also known, the unconditional distribution F_Y can be evaluated using

$$F_Y(y) = \int_{\zeta} F_{Y|\mathbf{Z}}(x|\zeta) f_{\mathbf{Z}}(\zeta)\, d\zeta$$

- Expected value of deterministic function g of parameters $\mathbf{Z}$ given $f_{\mathbf{Z}}$

$$E[g(\mathbf{Z})] = \int_{\zeta} g(\zeta) f_{\mathbf{Z}}(\zeta)\, d\zeta$$

- $\zeta = (\lambda, \psi, \sigma, \xi)$, $Y = A$ or $Y = A_N$

Return value estimated using expected values of parameters, $x_N(E[Z])$

- Motivated by the widespread approach of ignoring uncertainty in parameters ζ for estimation of return values

$$x_{N1} = x_N(E[\mathbf{Z}])$$

- $E[\mathbf{Z}] = \int_\zeta \zeta f_{\mathbf{Z}}(\zeta)\, d\zeta$
- A related estimator converging to x_{N1} with increasing N, would be $x'_N(E[Z])$
- similar choices of estimator here could be the MLE, MAP, median parameter values

Expected quantile of distribution of A with NEP $1 - 1/N$, $E[x_N(Z)]$

$$x_{N2} = E[x_N(Z)] = \int_\zeta x_N(\zeta) f_Z(\zeta)\, d\zeta$$

- Solve for quantile $x_N(\zeta)$ of the distribution of A with NEP $1 - 1/N$ for a large number of parameter choices ζ, and then integrate
- A related estimator $E[x'_N(Z)]$ is the expected quantile of distribution of A_N with NEP $\exp(-1)$ (converges to x_{N2} as N increases)

Quantile of predictive distribution of A with NEP $1-1/N$, $Q_A(1-1/N)$

$$F_A(x_{N3}) = 1 - \frac{1}{N}$$

- $F_A(x) = \int_\zeta F_{A|\mathbf{Z}}(x|\zeta) f_{\mathbf{Z}}(\zeta)\, d\zeta$
- Write briefly as $x_{N3} = Q_A(1-1/N)$, where Q_A is the quantile function corresponding to cumulative distribution function F_A

Quantile of predictive distribution of A_N with NEP $\exp(-1)$, $Q_{A_N}(\exp(-1))$

$$F_{A_N}(x_{N4}) = \exp(-1)$$

- $F_{A_N}(x) = \int_\zeta F_{A_N|\mathbf{Z}}(x|\zeta) f_{\mathbf{Z}}(\zeta)\, d\zeta$
- Write briefly as $x_{N4} = Q_{A_N}(\exp(-1))$, where Q_{A_N} is the quantile function corresponding to F_{A_N}

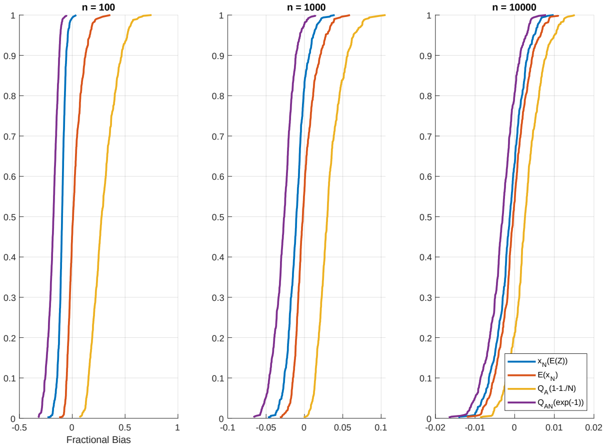
Simulation study

- $n_P = 100$ random (latin hypercube) pairs of ξ and σ on $[-0.2, 0.2] \times [1, 3]$. Then for each pair ξ, σ
  - $n_R = 500$ realisations of sample (size n) generated from generalised Pareto tail
  - Estimates for ξ, σ obtained using maximum likelihood
  - Return value estimated using one of 4 approaches above (incorporating uncertainty from all n_R realisations)
  - Compute 1000-year return value assuming 10 storms per annum.
- Distribution of fractional bias estimated (using all n_P selections)

$$\text{fractional bias} = \frac{\text{estimated return value} - \text{true return value}}{\text{true return value}}$$

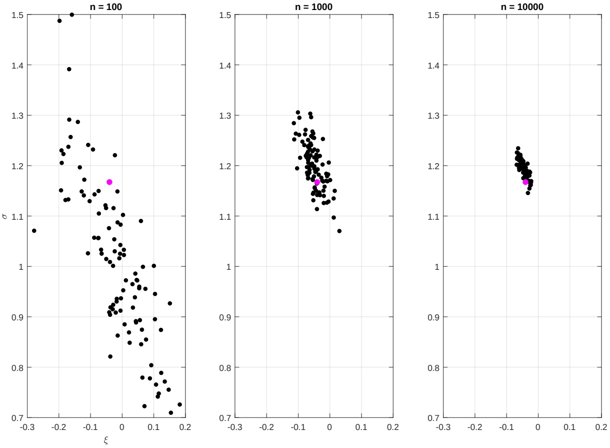
Distribution of fractional bias

- 500 realisations of sample size n, truth known per sample
- ξ, σ and x_N estimated per sample using MLE
- Empirical distribution of fractional bias accumulated



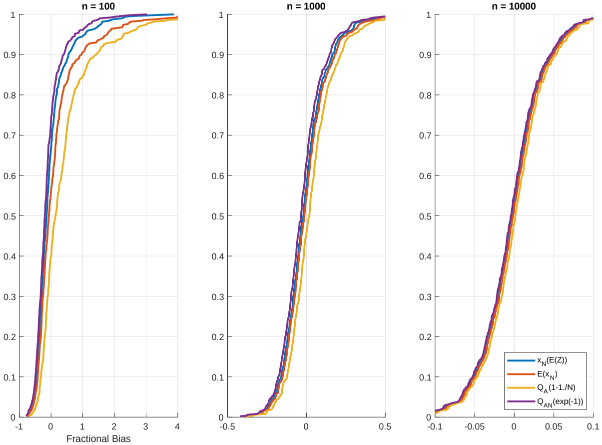
Bootstrap uncertainty estimation

- Illustrative single sample (size n)
- ξ, σ and x_N estimates shows for each of 100 bootstrap resamples



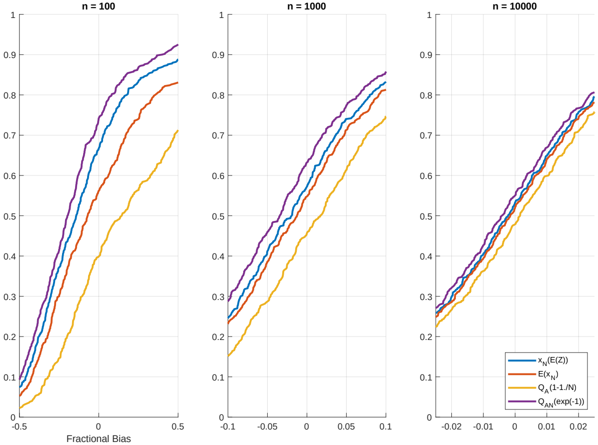
Bootstrap uncertainty estimation

- Empirical distribution accumulated:
- over all 100 bootstrap resamples and
- over all 500 sample realisations



Bootstrap uncertainty estimation

- Empirical distribution accumulated:
- over all 100 bootstrap resamples and
- over all 500 sample realisations



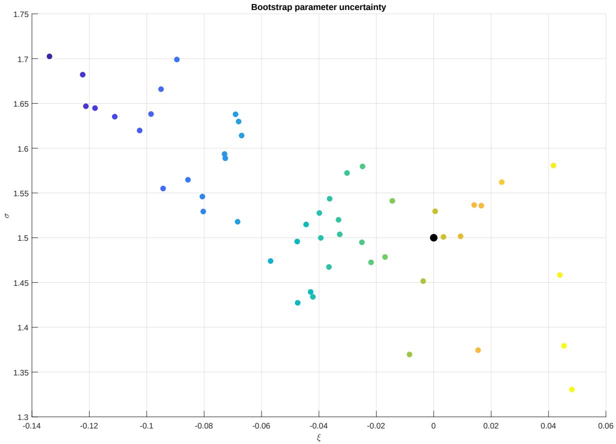
Theoretical properties

- In simple thought experiment, can show that
  - Quantile of predictive distribution $Q_A(1-1/N)$ will have positive bias
  - Quantile of predictive distribution $Q_{A_N}(\exp(-1))$ will have negative bias
  - Expected return value $E[x_N(\boldsymbol{Z})]$ is unbiased
- When the true value ξ_0 of ξ is negative, the form of the far tail of $Q_A(1-1/N)$ is dictated by values $\xi > \xi_0$. Hence likely that $Q_A(1-1/N)$ is biased high
- When $Q_A(1-1/N)$ is large, and the maximum observed ξ from n_R realisations is ξ^+, and n_R is large, that

$$\frac{E[x_N(\boldsymbol{Z})]}{Q_A(1-1/N)} \approx n_R^{(\xi^+-1)} \quad \Rightarrow \quad E[x_N(\boldsymbol{Z})] < Q_A(1-1/N) \text{ when } \xi^+ < 1$$

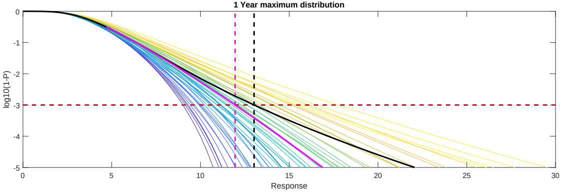
Theoretical properties

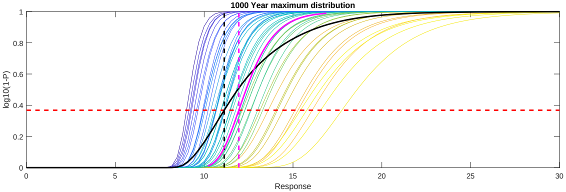
- Single sample
- Bootstrap estimates for ξ, σ
- Coloured by estimated ξ

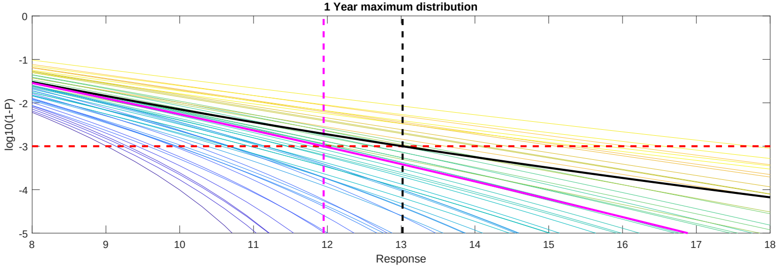


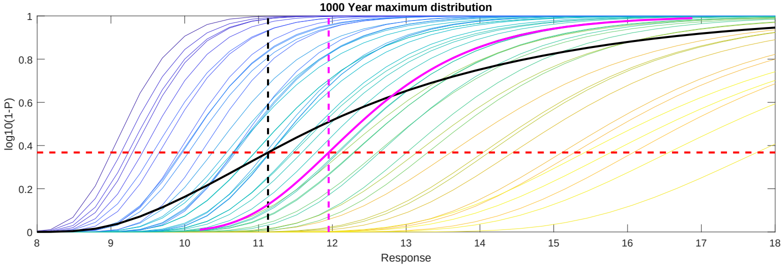
Theoretical properties

- Curves coloured by estimated ξ
- Vertical magenta: $E(x_N)$
- Vertical black
  - Top: $Q_A(1 - 1/N)$
  - Bot: $Q_{A_N}(\exp(-1))$






1 Year maximum distribution
0
-1
-2
-3
-4
-5
log10(1-P)
8
9
10
11
12
13
14
15
16
17
18
Response


1000 Year maximum distribution
1
0.8
0.6
0.4
0.2
0
log10(1-P)
8
9
10
11
12
13
14
15
16
17
18
Response

Findings

- Return value estimators yield different estimates under uncertainty
- $E[x_N(\boldsymbol{Z})]$ less biased in current simulations, estimated from F_A or F_{A_N}
- $Q_A(1-1/N)$ statistically preferable, has given exceedance probability

Are return values necessary?

- Why estimate a return value? What question are we trying to answer?
- Do safety factors elsewhere in the design process require return values with assumed characteristics?
- Does the framework used for inference affect estimated return values?

Recommendations

- Take great care in estimating and interpreting return values, when model parameters are uncertain
- Propagate full sample $\{\lambda_k, \psi_k, \sigma_k, \xi_k\}_{k=1}^{n_R}$ and "integrate out parameter uncertainty" as late as possible in inference
- Use decision theory: structure the decision problem, and estimate risk