Quantifying uncertainty in Probability of Exceedence (PE) curves

Jonathan Rougier

Department of Mathematics
University of Bristol, UK

Catastrophe modelling workshop
Lloyd's of London, Tue 6 October

Climate, weather, and flooding

Climate is the distribution of weather, and a natural hazard like a flood is an extreme weather event.

- *A priori:* The distribution of 'weather events' (e.g. storms) over the coming year is uncertain; hence we imagine a collection of possible storms, Ω, and a matching collection of probabilities, $P := \{p_\omega : \omega \in \Omega\}$.

$$\text{Climate} = \big\{\Omega, P\big\}.$$

Climate, weather, and flooding

Climate is the distribution of weather, and a natural hazard like a flood is an extreme weather event.

- *A priori:* The distribution of 'weather events' (e.g. storms) over the coming year is uncertain; hence we imagine a collection of possible storms, Ω, and a matching collection of probabilities, $P := \{p_\omega : \omega \in \Omega\}$.

$$\text{Climate} = \big\{\Omega, P\big\}.$$

- *A posteriori:* For a given storm ω, we can compute a single scalar summary, v_ω:

$$\overbrace{\omega}^{\text{storm}} \longrightarrow \overbrace{h_\omega(\mathbf{x}, t)}^{\text{footprint}} \longrightarrow \overbrace{\max_t \big|h_\omega(\mathbf{x}_0, t)\big|}^{\text{summary}} =: v_\omega$$

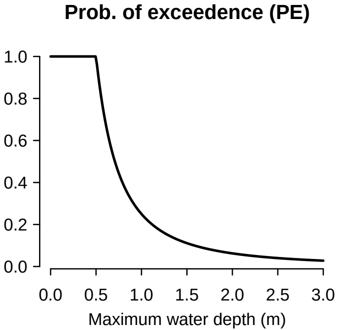
where v_ω might be the maximum depth of water upstream of a bridge in a town centre (location $\mathbf{x}_0$).

Probability of exceedence (PE) curves

Treating ω as uncertain, v_ω becomes an uncertain quantity which we label as $\tilde{v}$. The distribution of $\tilde{v}$ is a summary of climate, in terms of the impact of extreme weather events.

Probability of exceedence (PE) curves

Treating ω as uncertain, v_ω becomes an uncertain quantity which we label as $\tilde{v}$. The distribution of $\tilde{v}$ is a summary of climate, in terms of the impact of extreme weather events.



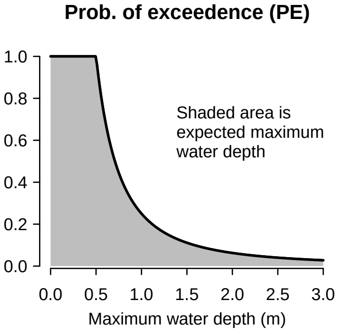
The probability distribution function of $\tilde{v}$ is

$$F_{\tilde{v}}(v) := \Pr\{\tilde{v} \leq v\} = \sum_{\omega} 1[v_\omega \leq v] p_\omega .$$

The PE curve is the plot of v (x-axis) against $1 - F_{\tilde{v}}(v)$.

Probability of exceedence (PE) curves

Treating ω as uncertain, v_ω becomes an uncertain quantity which we label as $\tilde{v}$. The distribution of $\tilde{v}$ is a summary of climate, in terms of the impact of extreme weather events.



The probability distribution function of $\tilde{v}$ is

$$\begin{aligned} F_{\tilde{v}}(v) &:= \Pr\{\tilde{v} \le v\} \\ &= \sum_\omega 1[v_\omega \le v] p_\omega . \end{aligned}$$

The PE curve is the plot of v (x-axis) against $1 - F_{\tilde{v}}(v)$.

Uncertainty from sampling

Often, we cannot evaluate the probability distribution function of $\tilde{v}$ exactly, and have to approximate by sampling.

In other words, we generate a large number N of independent realizations of $\tilde{v}$, and use the resulting empirical distribution function as an estimate of the true distribution function:

$$\tilde{v}^{(1)}, \ldots, \tilde{v}^{(N)} \overset{\text{iid}}{\sim} F_{\tilde{v}} \quad \text{and} \quad \hat{F}_{\tilde{v}}(v) = N^{-1} \sum_{i=1}^{N} 1[\tilde{v}^{(i)} \leq v]$$

Uncertainty from sampling

Often, we cannot evaluate the probability distribution function of $\tilde{v}$ exactly, and have to approximate by sampling.

In other words, we generate a large number N of independent realizations of $\tilde{v}$, and use the resulting empirical distribution function as an estimate of the true distribution function:

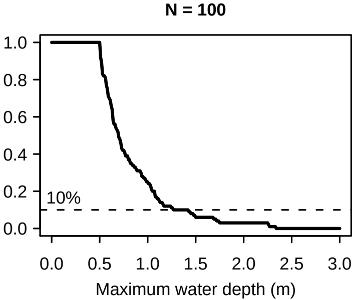
$$\tilde{v}^{(1)}, \ldots, \tilde{v}^{(N)} \overset{\text{iid}}{\sim} F_{\tilde{v}} \quad \text{and} \quad \hat{F}_{\tilde{v}}(v) = N^{-1} \sum_{i=1}^{N} 1[\tilde{v}^{(i)} \leq v]$$

- Because this is only an estimate, we should quantify our uncertainty about the true PE curve in terms of $1-\alpha$ confidence bands, such that

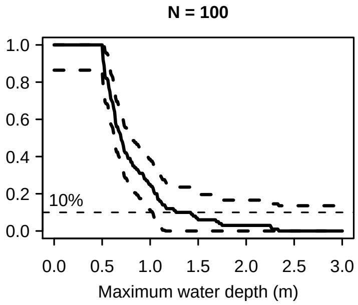
  $$\Pr\left\{L(x) \leq F_{\tilde{v}}(x) \leq U(x) \text{ for all } x\right\} \geq 1-\alpha.$$

  The width of $U(x) - L(x)$ will depend on N and α. The functions L and U can be specified using the Dvoretzky- Kiefer-Wolfowitz inequality.

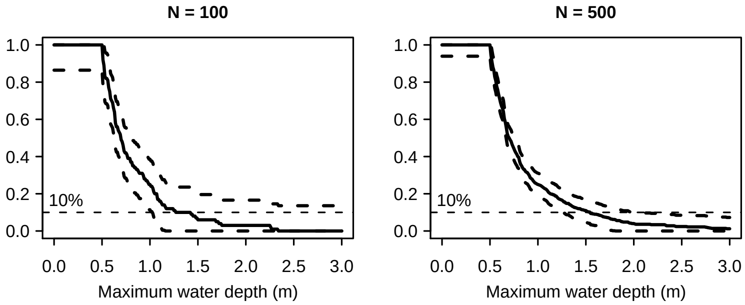
PE curve uncertainty from sampling, $\alpha = 5\%$



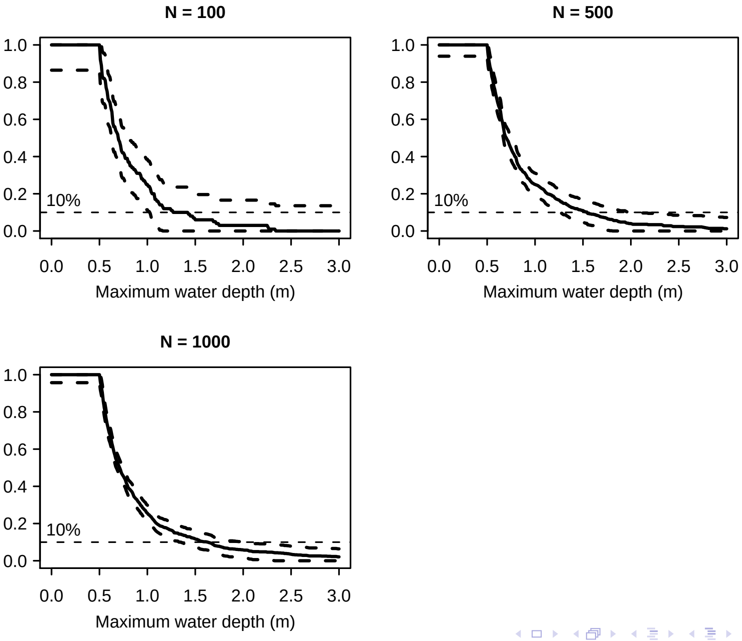
PE curve uncertainty from sampling, $\alpha = 5\%$



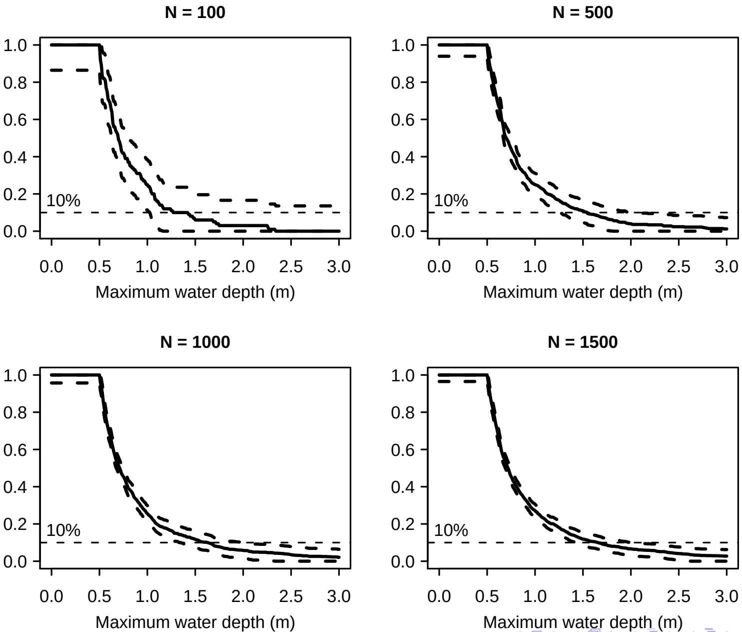
PE curve uncertainty from sampling, $\alpha = 5\%$



PE curve uncertainty from sampling, $\alpha = 5\%$



PE curve uncertainty from sampling, $\alpha = 5\%$

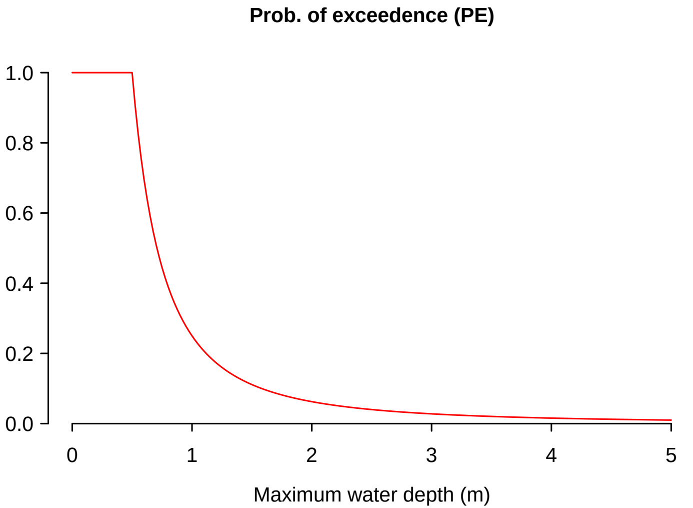


Uncertainty about climate

Uncertainty about climate means multiple candidates for climate, which means multiple PE curves.

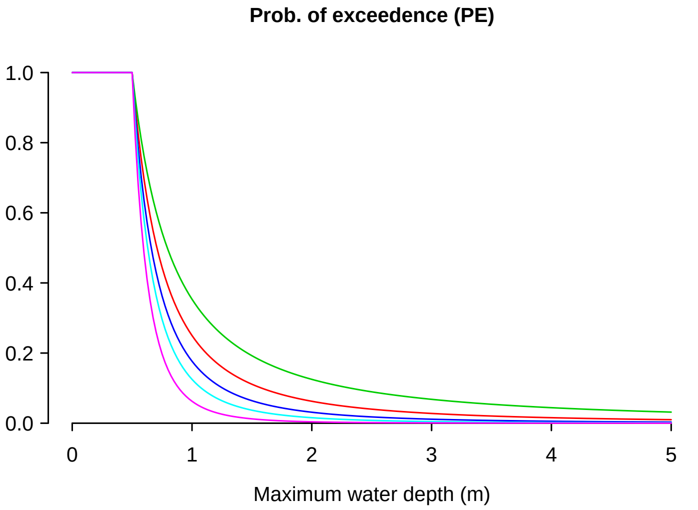
Uncertainty about climate

Uncertainty about climate means multiple candidates for climate, which means multiple PE curves.



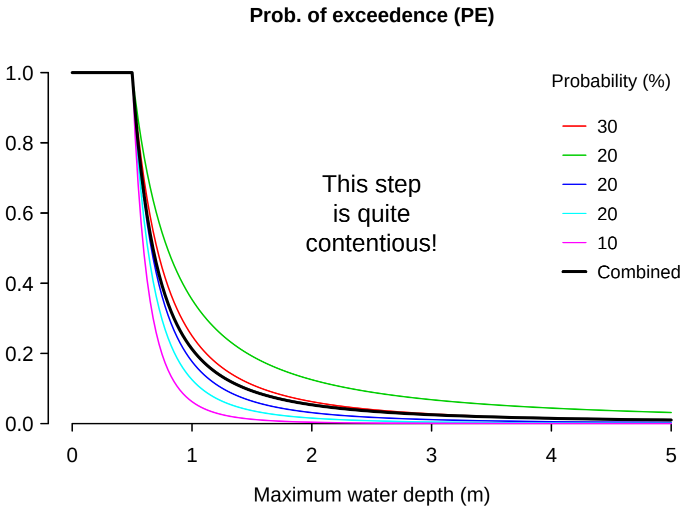
Uncertainty about climate

Uncertainty about climate means multiple candidates for climate, which means multiple PE curves.



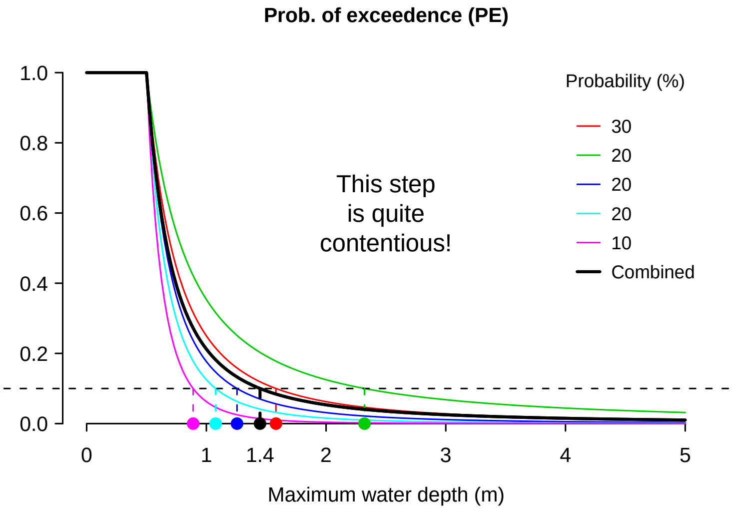
Uncertainty about climate

Uncertainty about climate means multiple candidates for climate, which means multiple PE curves.



Uncertainty about climate

Uncertainty about climate means multiple candidates for climate, which means multiple PE curves.



Summary

‘Aleatory’ uncertainty

The inherent uncertainty in a natural hazard, e.g. the uncertainty of extreme weather events in a given climate. A PE curve is a summary of aleatory uncertainty.

Summary

'Aleatory' uncertainty

The inherent uncertainty in a natural hazard, e.g. the uncertainty of extreme weather events in a given climate. A PE curve is a summary of aleatory uncertainty.

'Epistemic' uncertainty

'Other' uncertainty, notably that which arises from our incomplete knowledge. Obvious sources:

1. Limited resources for sampling, in cases where distribution functions cannot be evaluated pointwise.
2. Uncertainty about the hazard process, e.g. uncertainty about future climate.

Summary

'Aleatory' uncertainty

The inherent uncertainty in a natural hazard, e.g. the uncertainty of extreme weather events in a given climate. A PE curve is a summary of aleatory uncertainty.

'Epistemic' uncertainty

'Other' uncertainty, notably that which arises from our incomplete knowledge. Obvious sources:

1. Limited resources for sampling, in cases where distribution functions cannot be evaluated pointwise.
2. Uncertainty about the hazard process, e.g. uncertainty about future climate.
3. Uncertainty about the physical processes, e.g. the footprint function. *Is this the largest uncertainty of all?*