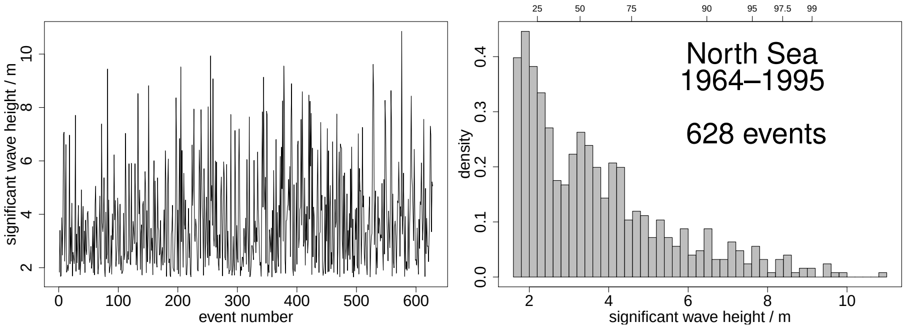
Cross-validatory extreme value threshold selection and uncertainty with application to off-shore engineering

Paul Northrop and Nicolas Attalides
University College London
p.northrop@ucl.ac.uk

Philip Jonathan
Shell Projects & Technology, Manchester

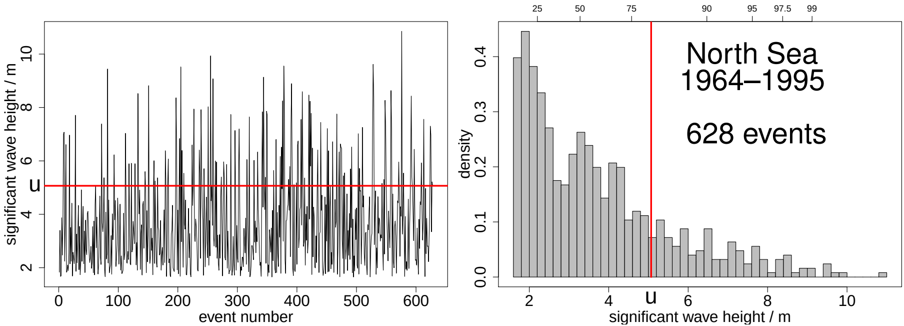
RSS International Conference 2014, Sheffield
2nd September 2014

Storm peak significant wave heights



- EV analysis: extapolate beyond the range of the data
- EV models: limiting models for extreme values
  (akin to normal model for sample means)

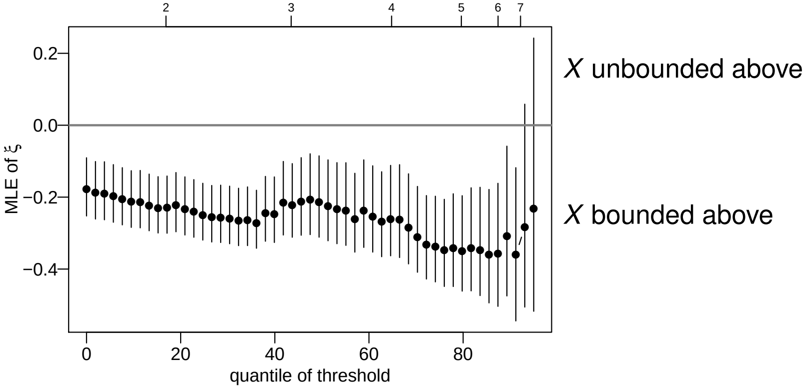
Storm peak significant wave heights



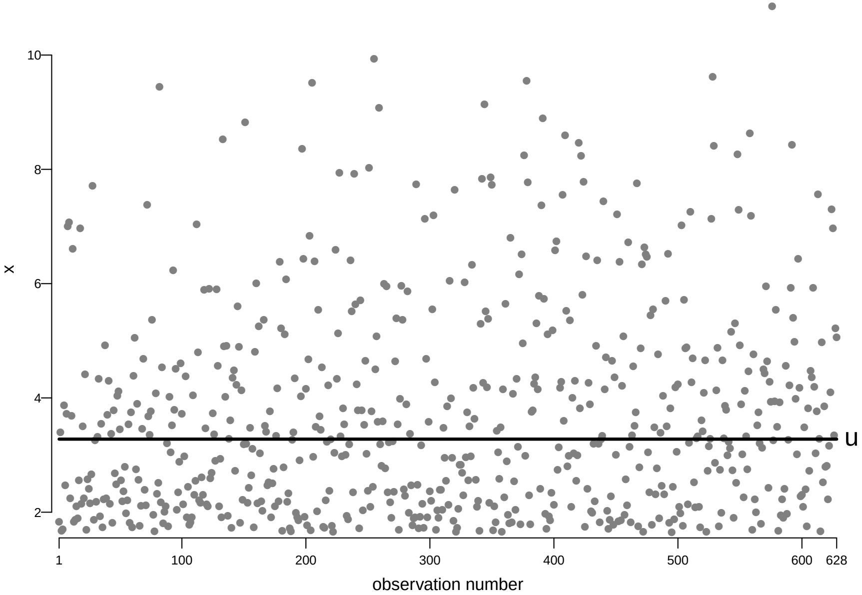
- EV analysis: extrapolate beyond the range of the data
- EV models: limiting models for extreme values (akin to normal model for sample means)
- $X_1, X_2, \ldots, X_n \overset{\text{indep}}{\sim} H$. Set a threshold u.
  - $P(X_i > u) = p_u$,
  - $(X_i - u) \mid X_i > u \;\dot{\sim}\; GP(\sigma_u, \xi)$, for sufficiently large u …

Threshold diagnostics

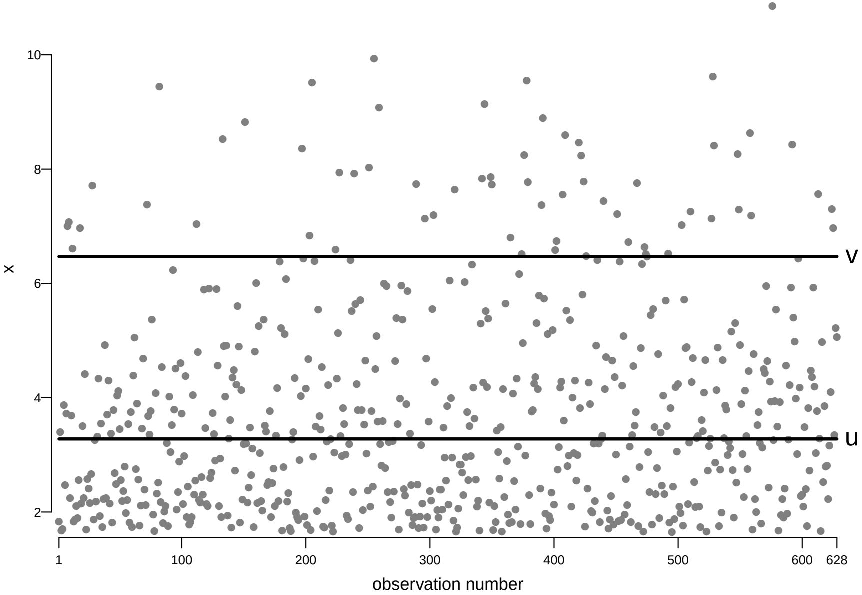
- Bias-variance trade-off
  - u too low : GP model inappropriate $\rightarrow$ bias
  - u too high : fewer excesses $\rightarrow$ unnecessary imprecision
- Many methods. Generalizability is desirable.
- E.g. look for stability in $\widehat{\xi}$ as u is increased . . .



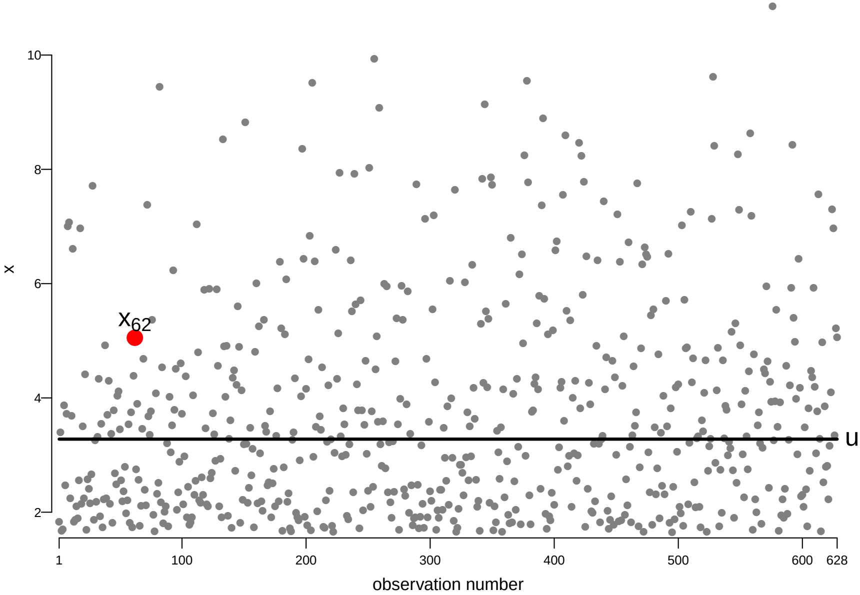
Training threshold *u*



Validation threshold $v \geqslant u$

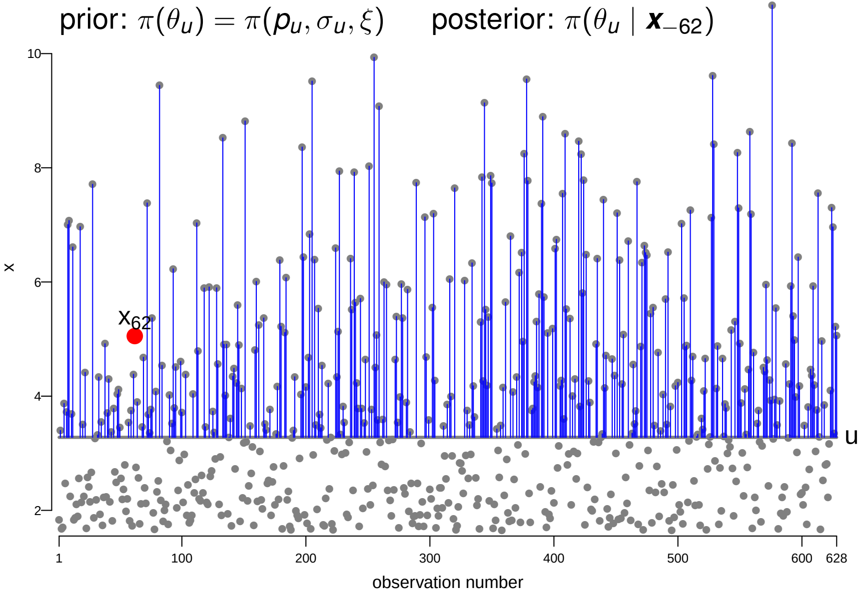


Leave-one-out cross-validation



Infer θ_u using $\boldsymbol{x}_{-62}$

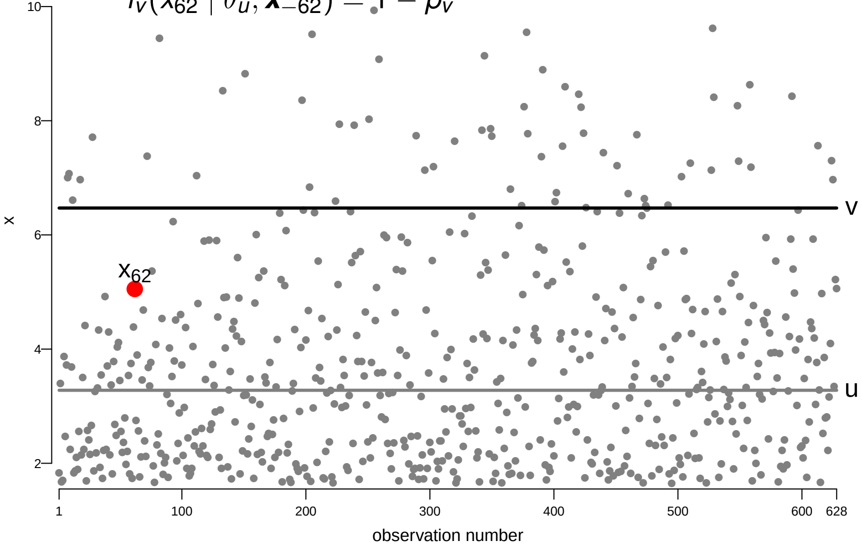
prior: $\pi(\theta_u) = \pi(p_u, \sigma_u, \xi)$ posterior: $\pi(\theta_u \mid \boldsymbol{x}_{-62})$



Prediction of non-exceedance of v

$$f_v(x_{62} \mid \boldsymbol{x}_{-62}, u) = \int f_v(x_{62} \mid \theta_u, \boldsymbol{x}_{-62})\, \pi(\theta_u \mid \boldsymbol{x}_{-62})\, \mathrm{d}\theta_u$$

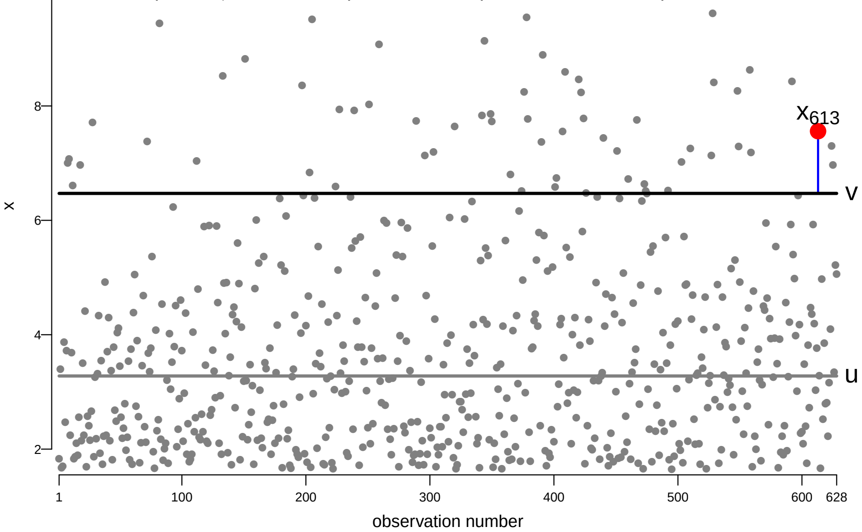
$$f_v(x_{62} \mid \theta_u, \boldsymbol{x}_{-62}) = 1 - p_v$$



Prediction of exceedance of v

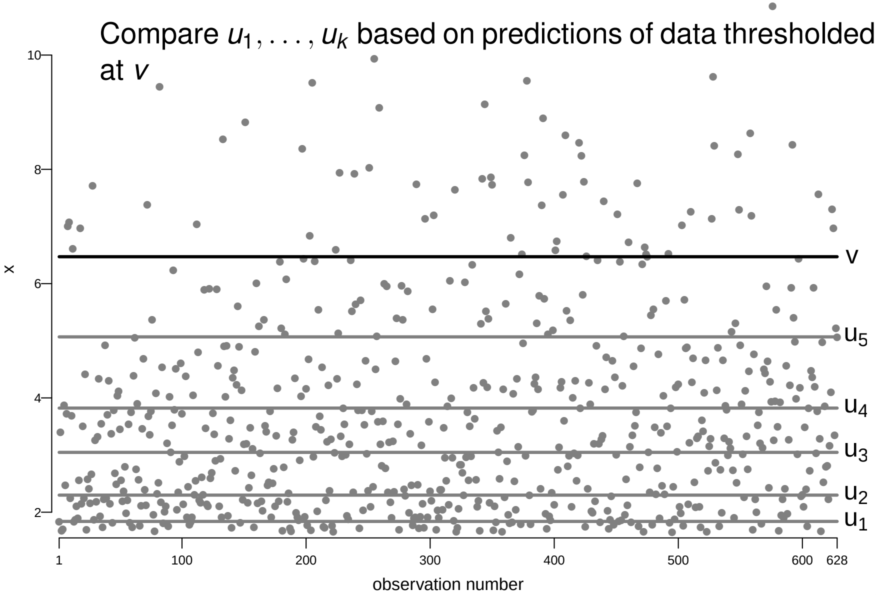
$$f_v(x_{613} \mid \boldsymbol{x}_{-613}, u) = \int f_v(x_{613} \mid \theta_u, \boldsymbol{x}_{-613})\, \pi(\theta_u \mid \boldsymbol{x}_{-613})\, \mathrm{d}\theta_u$$

$$f_v(x_{613} \mid \theta_u, \boldsymbol{x}_{-613}) = p_v\, f_{GP}(x_{613} - v; \sigma_v, \xi)$$



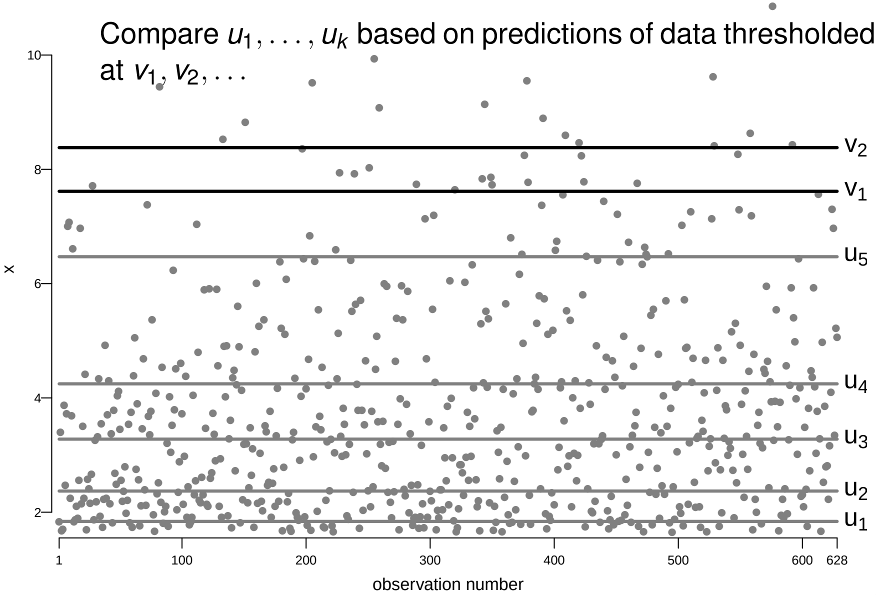
Training and validation thresholds

Compare $u_1, \ldots, u_k$ based on predictions of data thresholded at v



Training and validation thresholds

Compare $u_1, \ldots, u_k$ based on predictions of data thresholded at $v_1, v_2, \ldots$



Comparing thresholds

Sample $\theta_1^{(r)}, \ldots, \theta_m^{(r)}$ from $\pi(\theta_u \mid \boldsymbol{x}_{-r})$ [R-o-U or MCMC]

$$\widehat{f}_v(x_r \mid \boldsymbol{x}_{-r}, u) = \frac{1}{m} \sum_{j=1}^{m} f_v(x_r \mid \theta_j^{(r)})$$

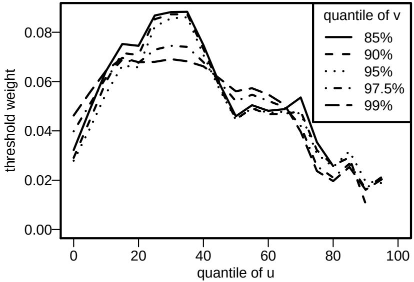
Measure of predictive performance at v when training at u

$$\widehat{T}_v(u) = \sum_{r=1}^{n} \log \widehat{f}_v(x_r \mid \boldsymbol{x}_{-r}, u)$$

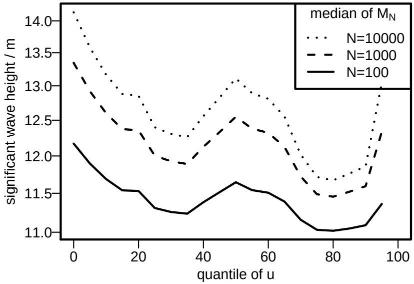
Normalize over thresholds $u_1, \ldots, u_k$ to give threshold weights

$$w_i(v) = \exp\{\widehat{T}_v(u_i)\} / \sum_{j=1}^{k} \exp\{\widehat{T}_v(u_j)\}, \qquad i = 1, \ldots, k$$

Importance sampling: 2 posterior samples (per u) rather than n



- Reference prior: dominated by likelihood?
- $u = 95\%$ quantile: 31 excesses
- $v = 99\%$ quantile: 6 excesses
- Good agreement over different v
- 'Best' threshold $\approx$ 35% quantile



- n_y=mean number of events per year
- $M_N = \max(X_1^{new}, \ldots, X_{n_y N}^{new})$,

$$F(z;\theta_u) = P(X \leqslant z \mid \theta_u)$$

$$P(M_N \leqslant z \mid \boldsymbol{x}) = \int F(z;\theta_u)^{n_y N} \pi(\theta_u \mid \boldsymbol{x}) \mathrm{d}\theta_u$$

Averaging inferences over thresholds

View k thresholds $u_1, \ldots, u_k$ as defining k competing models

- Prior probabilities: $P(u_i) = 1/k, i = 1, \ldots, k$ [... or something else]
- $\theta_i = (p_i, \sigma_i, \xi_i)$ under model u_i, with prior $\pi(\theta_i \mid u_i)$

Posterior threshold weights:

$$P_v(u_i \mid \boldsymbol{x}) \propto f_v(\boldsymbol{x} \mid u_i)\, P(u_i),$$

where

$$f_v(\boldsymbol{x} \mid u_i) = \int f_v(\boldsymbol{x} \mid \theta_i, u_i)\pi(\theta_i \mid u_i)\, \mathrm{d}\theta_i$$

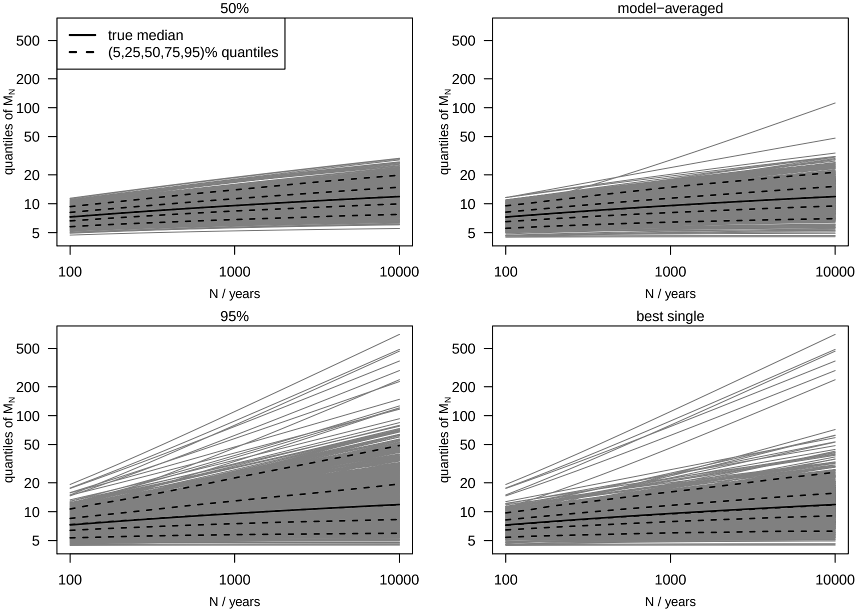
$$\widehat{f}_v(\boldsymbol{x} \mid u_i) = \prod_{r=1}^{n} f_v(x_r \mid \boldsymbol{x}_{-r}, u_i) = \exp\{\widehat{T}_v(u_i)\}$$ [Geisser and Eddy (1979)]

$$\widehat{P}_v(u_i \mid \boldsymbol{x}) = \frac{\exp\{\widehat{T}_v(u_i)\}\, P(u_i)}{\sum_{j=1}^{k} \exp\{\widehat{T}_v(u_j)\}\, P(u_j)} = w_i(v)$$

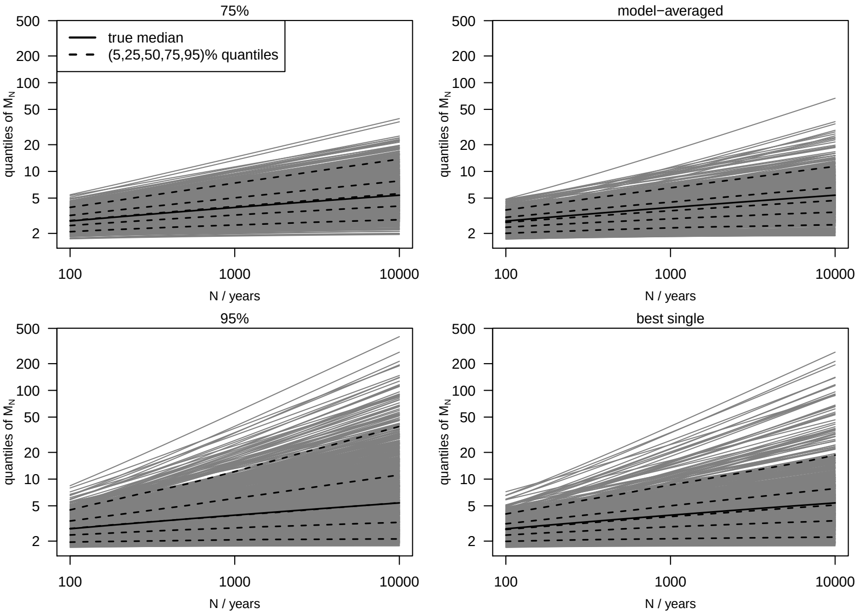
Model-averaged predictive inferences:

$$\widehat{P}_v(M_N \leqslant z \mid \boldsymbol{x}) = \sum_{i=1}^{k} \widehat{P}(M_N \leqslant z \mid \boldsymbol{x}, u_i)\, \widehat{P}_v(u_i \mid \boldsymbol{x}),$$

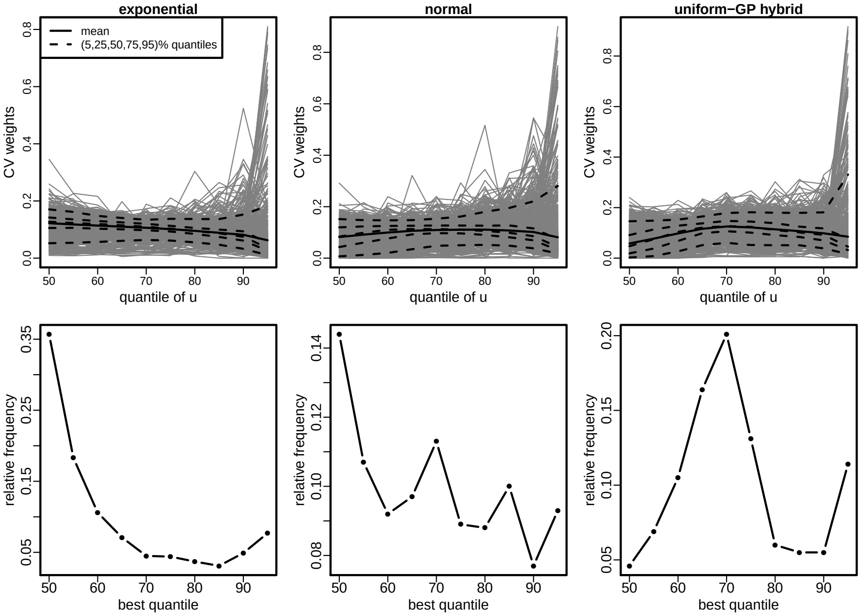
Simulation: exponential (v=95% quantile)



Simulation: uniform-GP(ξ=0.1) hybrid

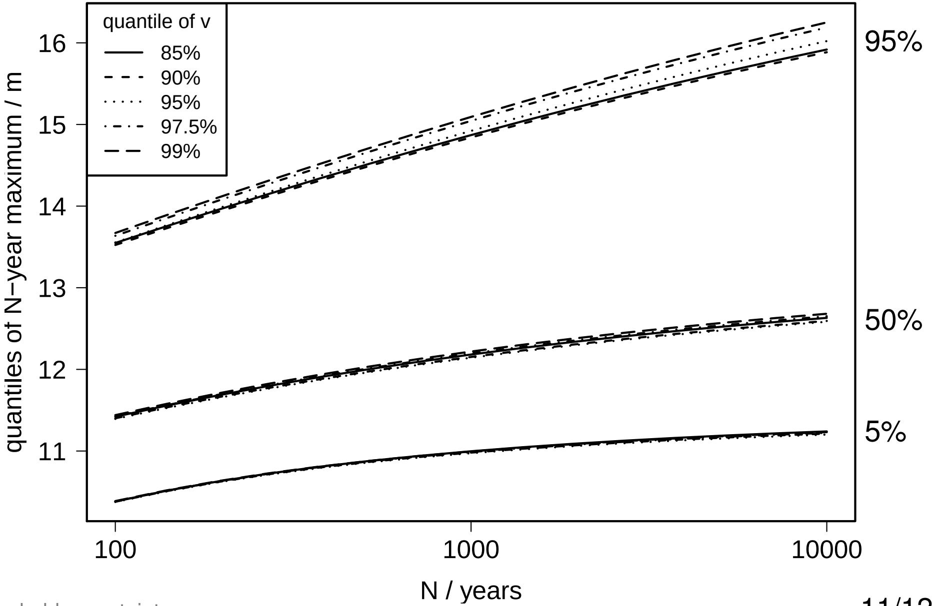


Threshold weights



North Sea : model-averaged predictions

UCL



Concluding remarks

Pros:

- simple graphical diagnostic
- facilitates averaging over different estimated tail behaviours
- based on standard model: easier to generalize

Cons: need to choose

- level of validation threshold v
- priors

Priors:

- Reference priors: need to avoid small sample sizes
- Use (weakly-)informative priors to avoid ridiculous answers?

On-going: serial dependence, multivariate extremes, covariate effects, choice of measurement scale

Concluding remarks

Pros:

- simple graphical diagnostic
- facilitates averaging over different estimated tail behaviours
- based on standard model: easier to generalize

Cons: need to choose

- level of validation threshold v
- priors

Priors:

- Reference priors: need to avoid small sample sizes
- Use (weakly-)informative priors to avoid ridiculous answers?

On-going: serial dependence, multivariate extremes, covariate effects, choice of measurement scale

Thank you for your attention.

References

Beirlant, J., Y. Goegebeur, J. Teugels, and J. Segers (2004). *Statistics of Extremes : Theory and Applications*. London: Oxford University Press.

Drees, H., L. de Haan, and S. Resnick (2000). How to make a Hill plot. *The Annals of Statistics* **28(1)**, 254–274.

Ferreira, A., L. de Haan, and L. Peng (2003). On optimising the estimation of high quantiles of a probability distribution. *Statistics* **37(5)**, 401–434.

Geisser, S. and W. F. Eddy (1979). A predictive approach to model selection. *Journal of the American Statistical Association* **74(365)**, 153–160.

MacDonald, A., C. Scarrott, D. Lee, B. Darlow, M. Reale, and G. Russell (2011). A flexible extreme value mixture model. *Comp. Statist. Data Anal.* **55**, 2137–2157.

Northrop, P. J. and Coleman, C. L. (2014) Improved threshold diagnostic plots for extreme value analyses. *Extremes* **17(2)**, 289–303.

Sabourin, A., P. Naveau, and A.-L. Fougres (2013). Bayesian model averaging for multivariate extremes. *Extremes* **16(3)**, 325–350.

Scarrott C, MacDonald A (2012) A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT - Statistical Journal* **10(1)**, 33–60.

Wadsworth, J. and J. Tawn (2012). Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling. *J. Royal Statist. Soc.* **B 74(3)**, 543–567.