Cross-validatory extreme value threshold selection and uncertainty

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

Joint work with Nicolas Attalides and Philip Jonathan

SuSTaIn Extremes Workshop, University of Bristol
4th July 2014

Introduction

- $d = 1$. Sorry!

Introduction

- $d = 1$. Sorry!
- Selection of a single (best?) threshold
- Threshold uncertainty : average inferences over many thresholds $u_1, u_2, \ldots, u_k$
- Application to ocean storm severity
- Univariate IID case. $X_1, X_2, \ldots, X_n \overset{\text{indep}}{\sim} H$
  - $P(X_i > u) = p_u$,
  - $(X_i - u) \mid X_i > u \sim GP(\sigma_u, \xi)$
- ...but scope to generalize

Threshold diagnostics

Bias-variance trade-off :

- u too low : GP model inappropriate $\rightarrow$ bias
- u too high : fewer excesses $\rightarrow$ unnecessary imprecision

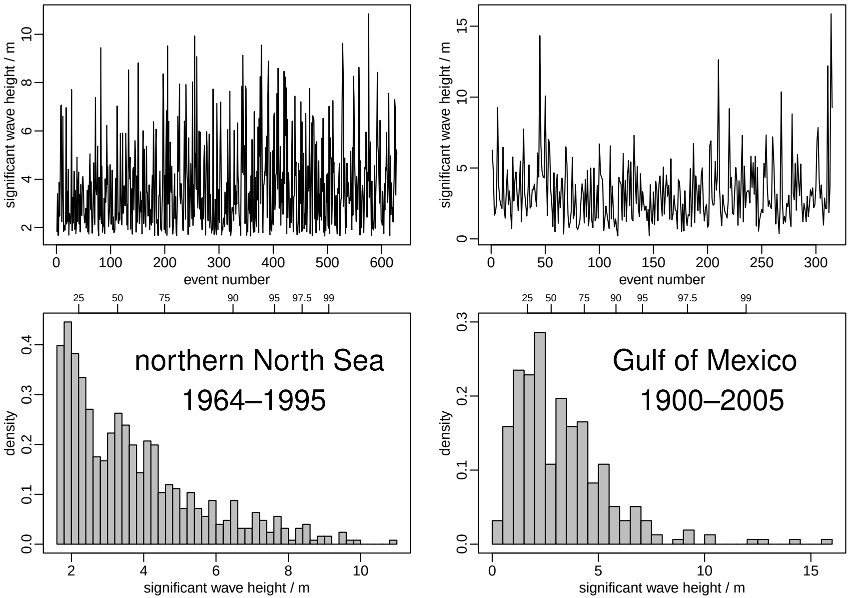
Review paper: Scarrott & MacDonald (2012)

- Estimates of ξ stable above some level of threshold?
  [Drees et al. (2000), Wadsworth and Tawn (2012), Northrop and Coleman (2014)]
- Goodness-of-fit of GP distribution
  [Davison and Smith (1990), Dupuis (1998)]
- Minimize asymptotic MSE of estimates of ξ or extreme quantiles under assumptions about H
  [Ferreira, et al. (2003), Beirlant (2004)]
- Extend EV model below u and make u a model parameter
  [Wadsworth and Tawn (2012), MacDonald et al. (2011)]

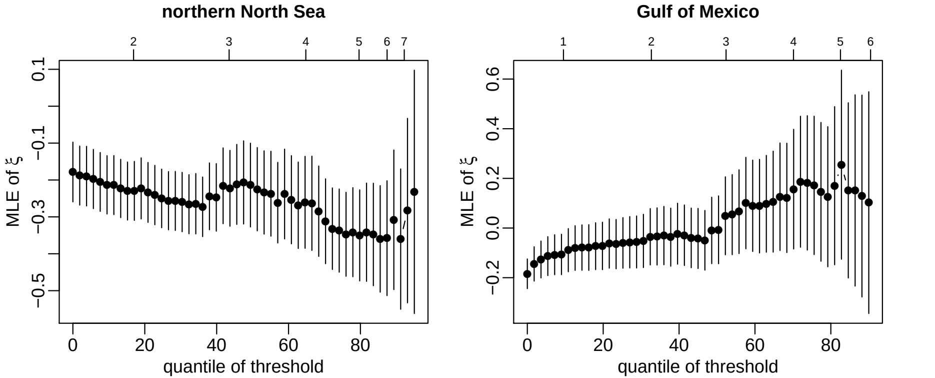
Aims

- address the bias-variance trade-off based on out-of-sample prediction at levels above u
- . . . using the bin-GP model
- produce a simple graphical diagnostic for single threshold selection
- account for uncertainty in threshold
- develop method than can be generalized: e.g. to MV extremes

Hindcast storm peak sig. wave heights



Threshold stability plots



Where to set threshold?

northern North Sea: MLEs of ξ are negative

Gulf of Mexico: MLEs of ξ become positive as u increases

Threshold comparison using CV

- Raw (unthresholded) data $\boldsymbol{x} = (x_1, \ldots, x_n)$
- Training threshold u
- Parameter vector $\theta = (p_u, \sigma_u, \xi)$
- Prior $\pi(\theta)$. Consider reference priors in the first instance
- Validation threshold $v \geqslant u$

Leave-one-out cross-validation

- $\boldsymbol{x}_{-r} = \{x_i, i \neq r\}$
- Posterior

$$\pi_u(\theta \mid \boldsymbol{x}_{-r}) \propto \pi(\theta) \prod_{i \neq r} f_u(\boldsymbol{x}_i \mid \theta)$$

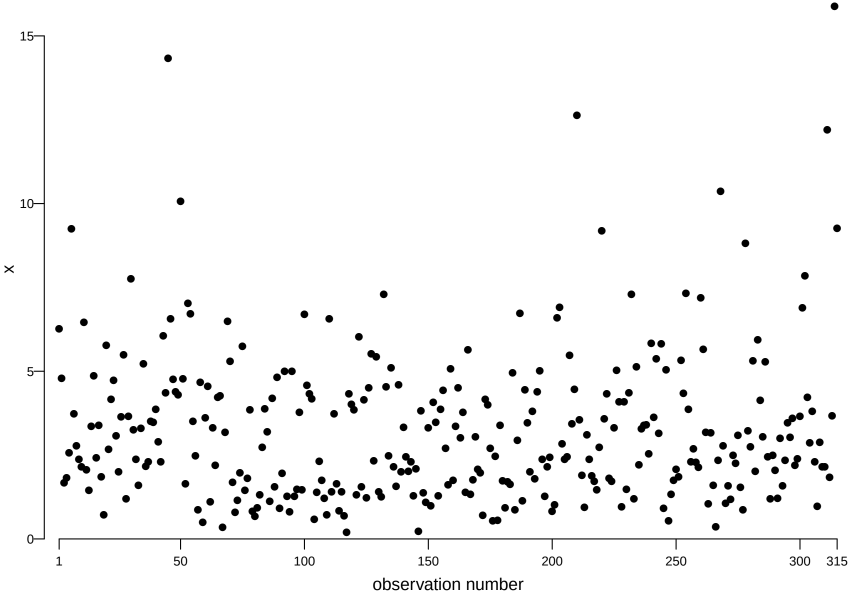
where

$$f_u(x_i \mid \theta) = (1 - p_u)^{I(x_i \leqslant u)} \{p_u \, g(x_i - u : \sigma_u, \xi)\}^{I(x_i > u)}$$

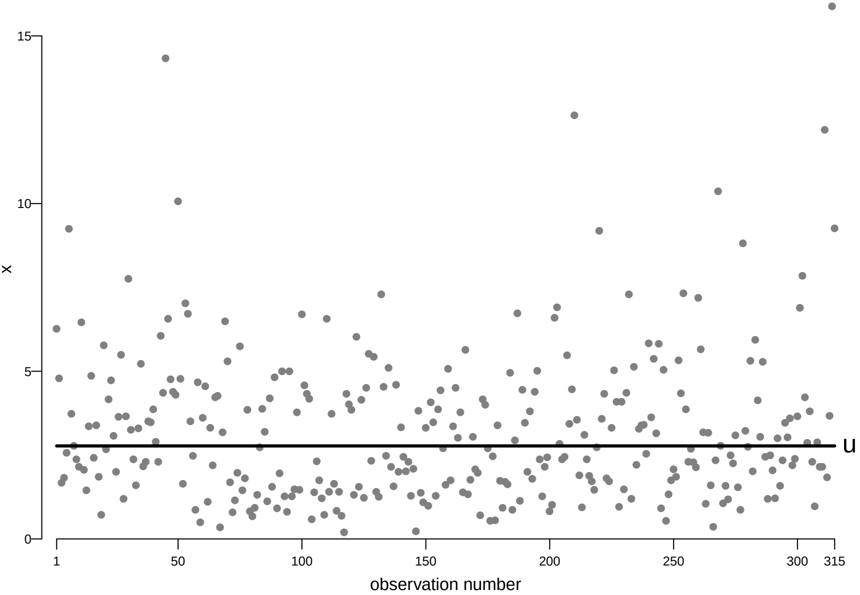
and

$$g(x; \sigma_u, \xi) = \sigma_u^{-1} \left(1 + \frac{\xi x}{\sigma_u}\right)_+^{-(1+1/\xi)}$$

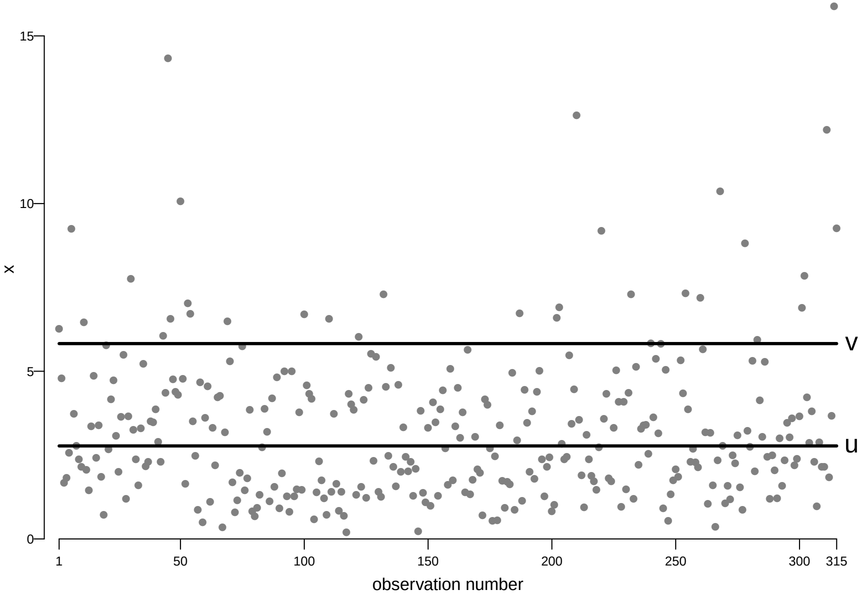
Gulf of Mexico data



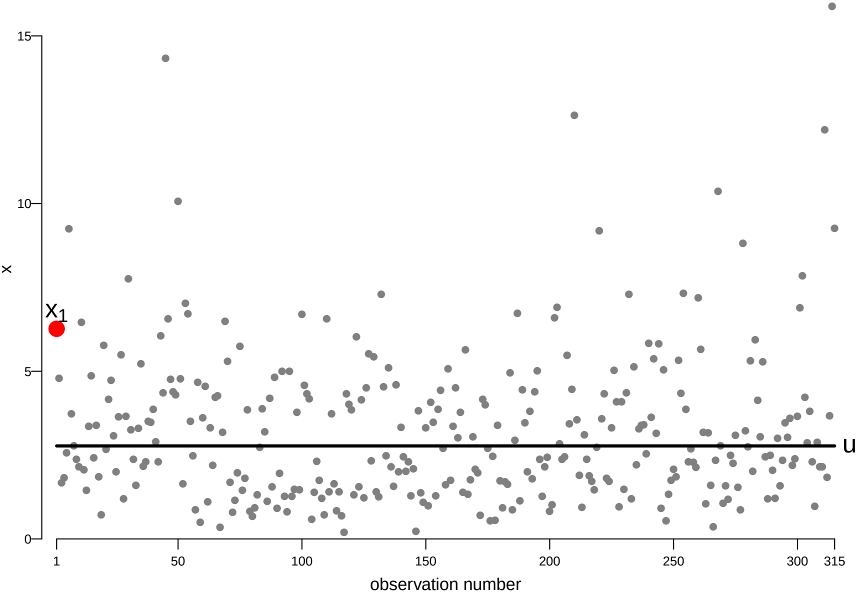
Training threshold *u*



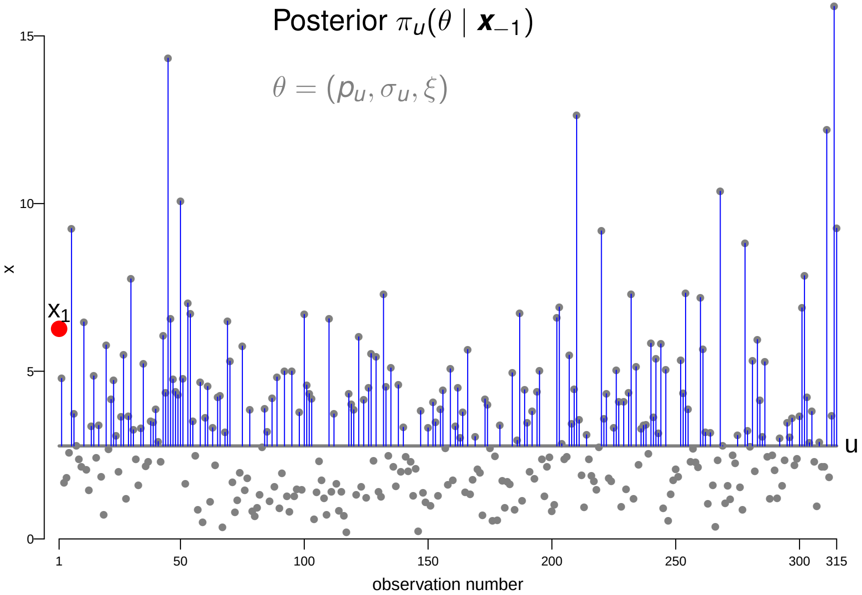
Validation threshold $v \geqslant u$



Leave-one-out cross-validation

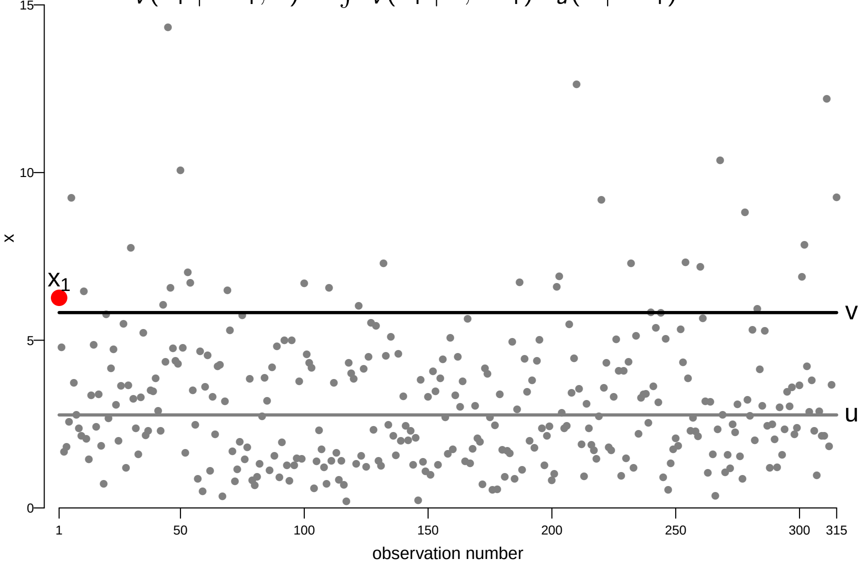


Prediction of non-exceedance of v



Prediction of non-exceedance of v

$$f_v(x_1 \mid \boldsymbol{x}_{-1}, u) = \int f_v(x_1 \mid \theta, \boldsymbol{x}_{-1})\, \pi_u(\theta \mid \boldsymbol{x}_{-1})\, \mathrm{d}\theta$$



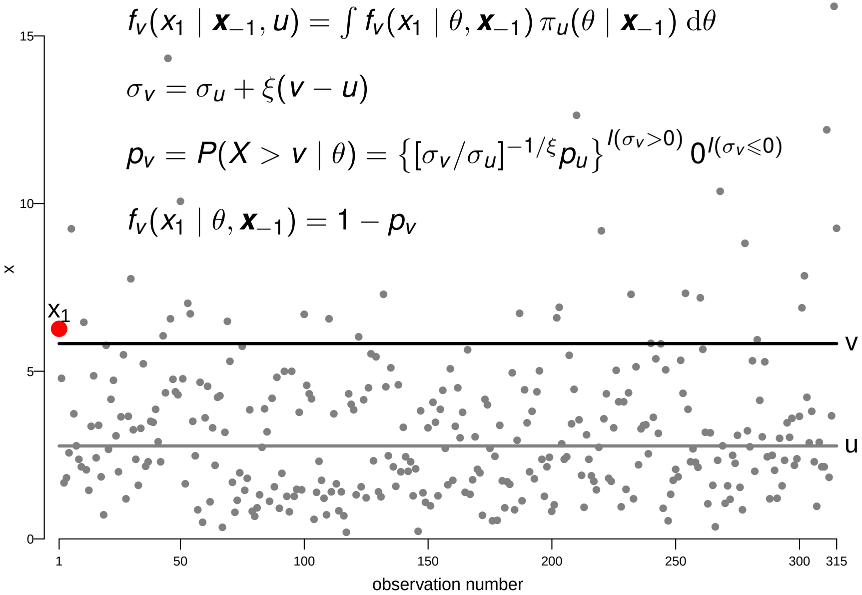
Prediction of non-exceedance of v

$$f_v(x_1 \mid \boldsymbol{x}_{-1}, u) = \int f_v(x_1 \mid \theta, \boldsymbol{x}_{-1})\, \pi_u(\theta \mid \boldsymbol{x}_{-1})\, \mathrm{d}\theta$$

$$\sigma_v = \sigma_u + \xi(v - u)$$

$$p_v = P(X > v \mid \theta) = \left\{[\sigma_v/\sigma_u]^{-1/\xi} p_u\right\}^{I(\sigma_v > 0)}\, 0^{I(\sigma_v \leqslant 0)}$$

$$f_v(x_1 \mid \theta, \boldsymbol{x}_{-1}) = 1 - p_v$$



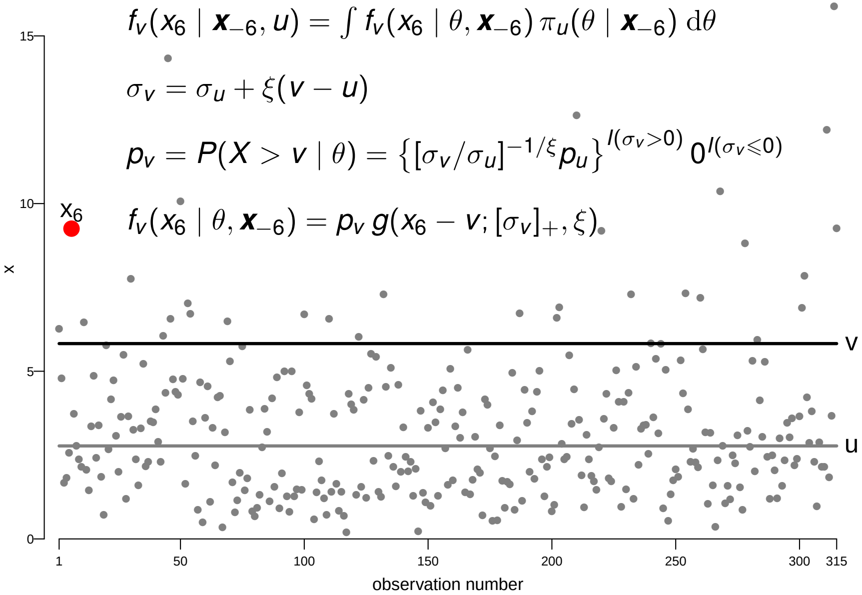
Prediction of exceedance of v

$$f_v(x_6 \mid \boldsymbol{x}_{-6}, u) = \int f_v(x_6 \mid \theta, \boldsymbol{x}_{-6})\, \pi_u(\theta \mid \boldsymbol{x}_{-6})\, \mathrm{d}\theta$$

$$\sigma_v = \sigma_u + \xi(v - u)$$

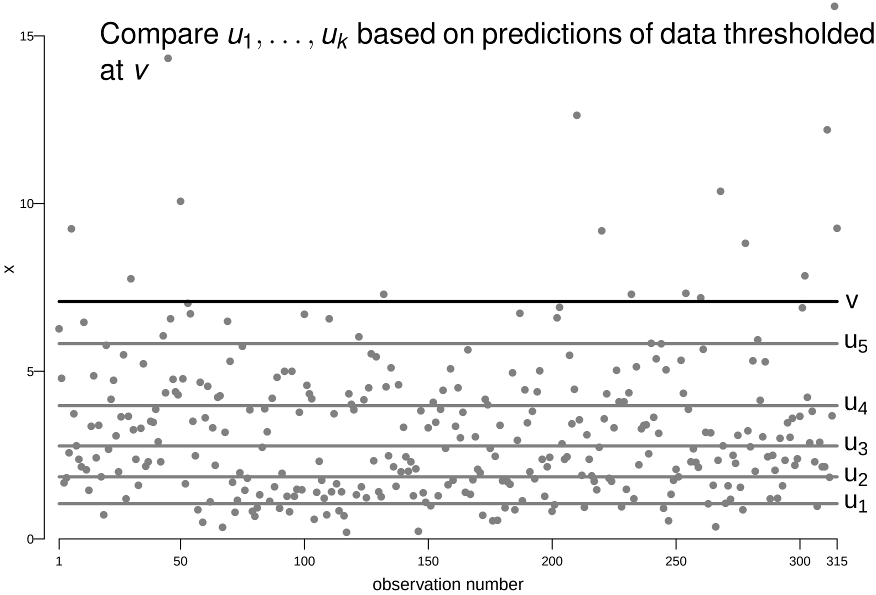
$$p_v = P(X > v \mid \theta) = \left\{[\sigma_v/\sigma_u]^{-1/\xi} p_u\right\}^{I(\sigma_v > 0)} 0^{I(\sigma_v \leqslant 0)}$$

$$f_v(x_6 \mid \theta, \boldsymbol{x}_{-6}) = p_v\, g(x_6 - v; [\sigma_v]_+, \xi)$$



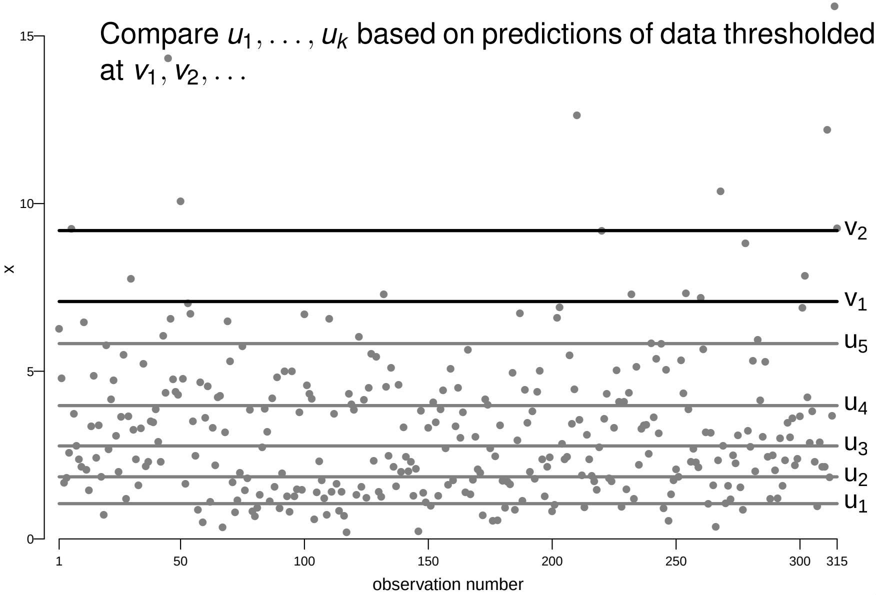
Training and validation thresholds

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 $\pi_u(\theta \mid \boldsymbol{x}_{-r})$ [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 training thresholds $u_1, \ldots, u_k$

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

Threshold weights: $w_1(v), \ldots, w_k(v)$

Importance sampling (IS)

- IS density $h(\theta)$
  [support of $h(\theta)$ must contain support of $\pi(\theta \mid \boldsymbol{x}_{-r})$]
- Let $q_r(\theta) = \pi_u(\theta \mid \boldsymbol{x}_{-r})/h(\theta)$

$$f_v(x_r \mid \boldsymbol{x}_{-r}, u) = \int f_v(x_r \mid \theta, \boldsymbol{x}_{-r})\, q_r(\theta)\, h(\theta)\, \mathrm{d}\theta, \quad r = 1, \ldots, n$$

IS ratio estimator, based on sample $\theta_1, \ldots, \theta_m$ from $h(\theta)$,

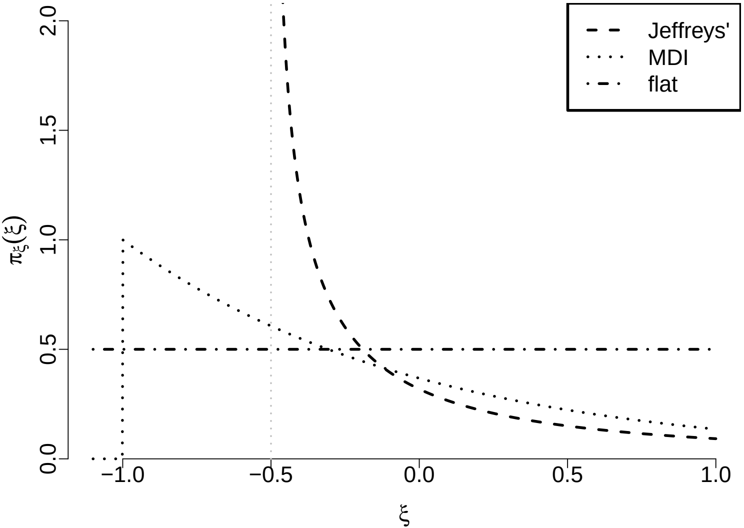
$$\widehat{f}_v(x_r \mid \boldsymbol{x}_{-r}, u) = \frac{\sum_{j=1}^m f_v(x_r \mid \theta_j)\, q_r(\theta_j)}{\sum_{j=1}^m q_r(\theta_j)}$$

Suppose that $x_1 < \cdots < x_n$. Use

$$h(\theta) = \begin{cases} \pi_u(\theta \mid \boldsymbol{x}) & \text{for } r = 1, \ldots, n-1 \\ \pi_u(\theta \mid \boldsymbol{x}_{-n}) & \text{for } r = n \end{cases}$$

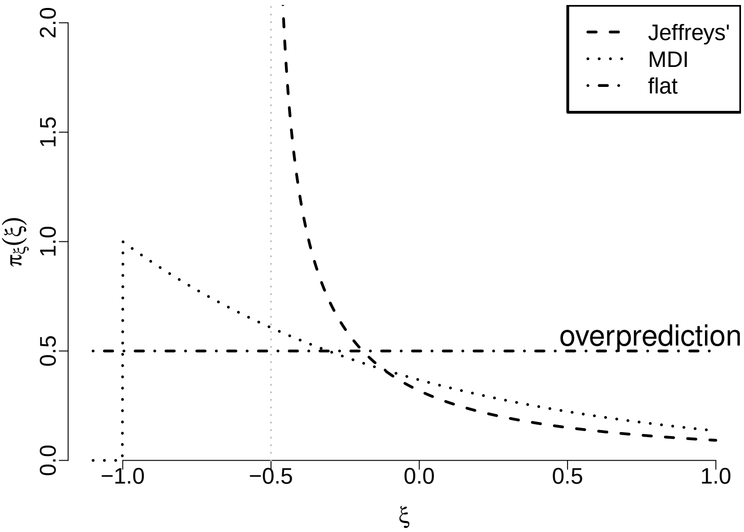
…so only need to send to sample from two posteriors.

Priors: $\pi(\sigma_u, \xi) \propto {\sigma_u}^{-1}\pi_\xi(\xi), \;\; \sigma_u > 0, \xi \in \mathbb{R}$



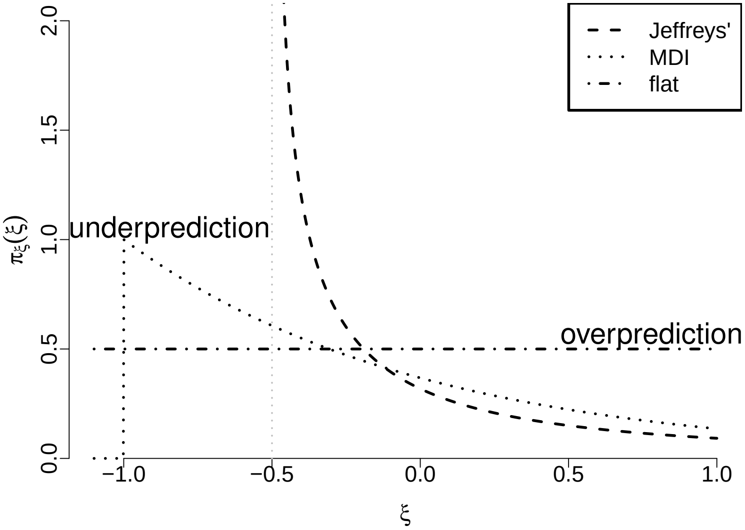
Simulation study with $p_u \in \{0.1, 0.5\}$ and $\xi \in \{-0.2, 0.1\}$

Priors: $\pi(\sigma_u, \xi) \propto \sigma_u^{-1}\pi_\xi(\xi), \;\; \sigma_u > 0, \xi \in \mathbb{R}$



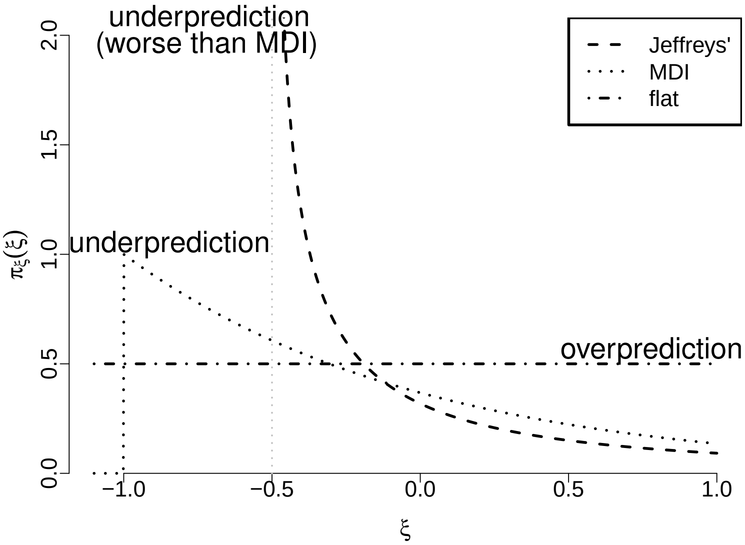
Simulation study with $p_u \in \{0.1, 0.5\}$ and $\xi \in \{-0.2, 0.1\}$

Priors: $\pi(\sigma_u, \xi) \propto \sigma_u^{-1}\pi_\xi(\xi), \;\; \sigma_u > 0, \xi \in \mathbb{R}$



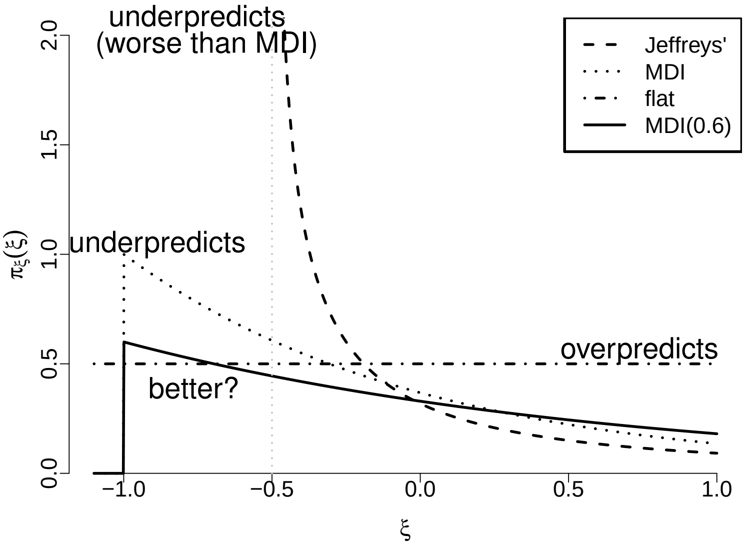
Simulation study with $p_u \in \{0.1, 0.5\}$ and $\xi \in \{-0.2, 0.1\}$

Priors: $\pi(\sigma_u, \xi) \propto {\sigma_u}^{-1}\pi_\xi(\xi), \;\; \sigma_u > 0, \xi \in \mathbb{R}$



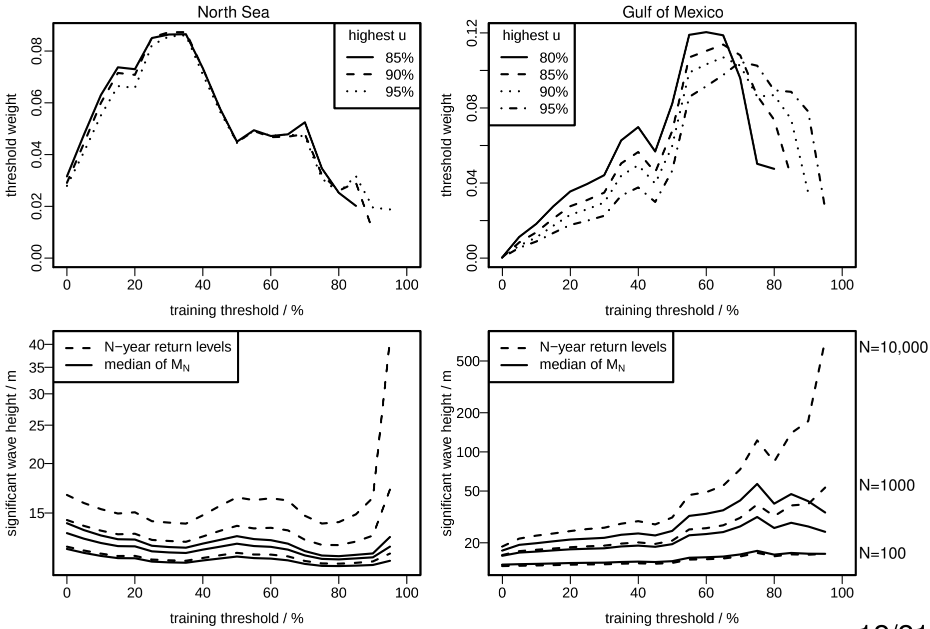
Simulation study with $p_u \in \{0.1, 0.5\}$ and $\xi \in \{-0.2, 0.1\}$

Priors: $\pi(\sigma_u, \xi) \propto {\sigma_u}^{-1}\pi_\xi(\xi)$

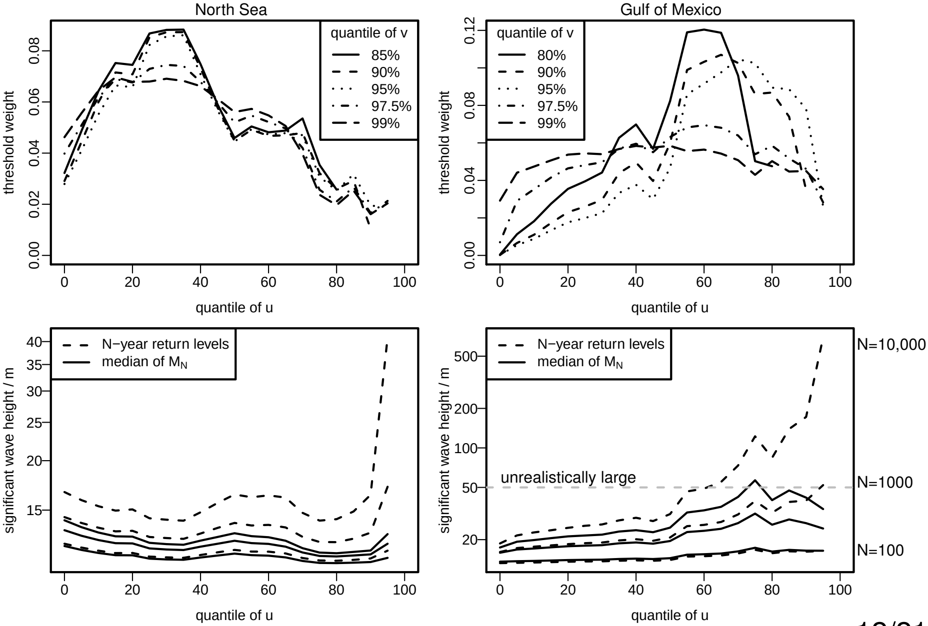


MDI(0.6): $\pi(\sigma_u, \xi) \propto \sigma_u^{-1} 0.6\, \mathrm{e}^{-0.6(\xi+1)}, \sigma_u > 0, \xi \in \mathbb{R}$

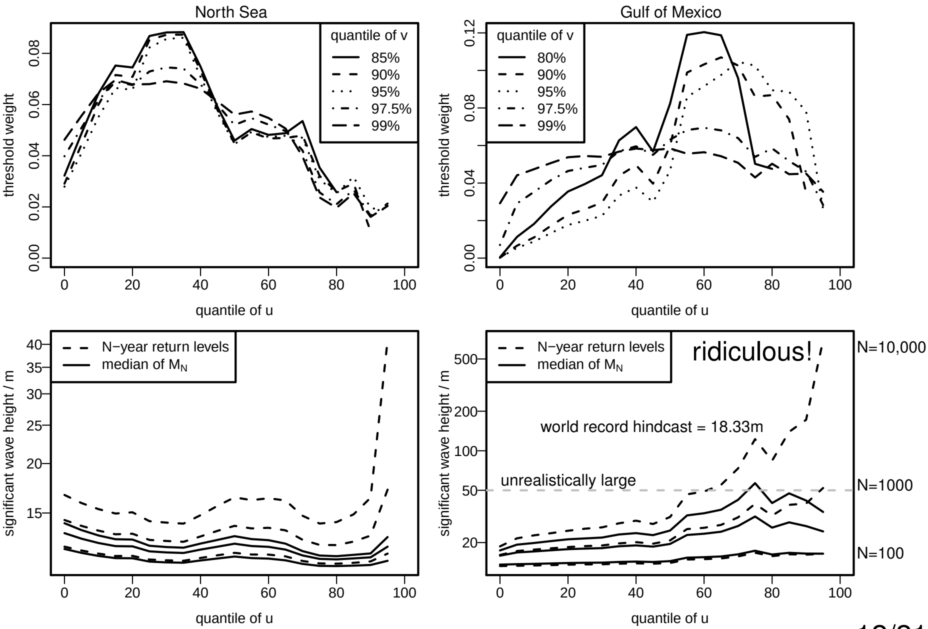
Threshold weights & predictive inference



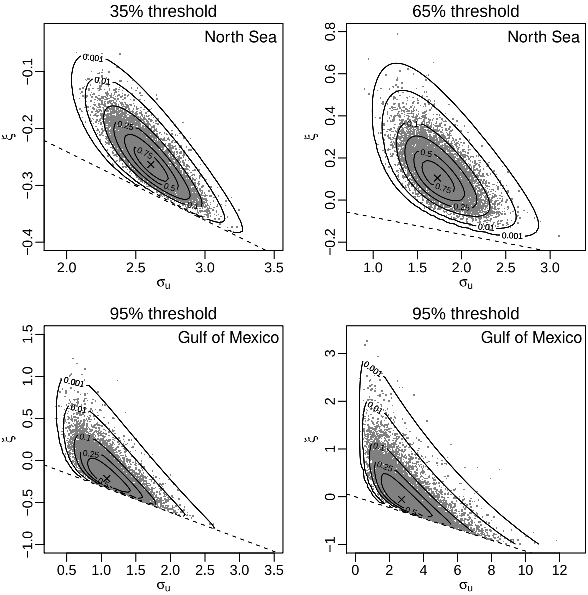
Threshold weights & predictive inference



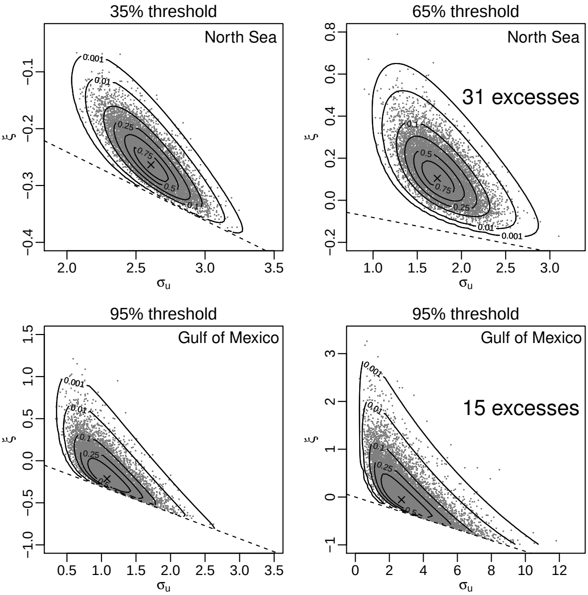
Threshold weights & predictive inference



GP posterior densities



GP posterior densities



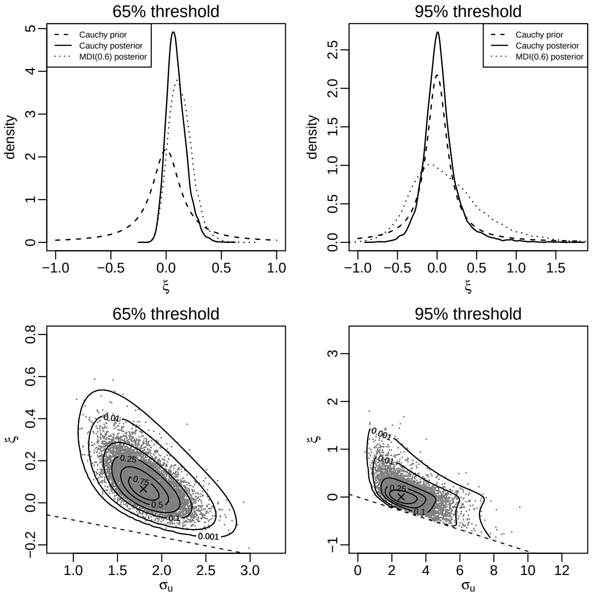
Reference priors appropriate only when dominated by information in the data

Expect sig. wave heights to be bounded above ($\xi < 0$)

GoM: $\widehat{P}(\xi > 1/2 \mid \boldsymbol{x}) \approx 0.2$ and $P(\xi > 1 \mid \boldsymbol{x}) \approx 0.05$

Avoid small samples, or give some information in prior, or don't extrapolate so far into the future

A weakly-informative (Cauchy) prior



Gulf of Mexico data

Downweight large values of ξ *a priori*, but give scope for data to contradict the prior

Cauchy: gentle slope in tails

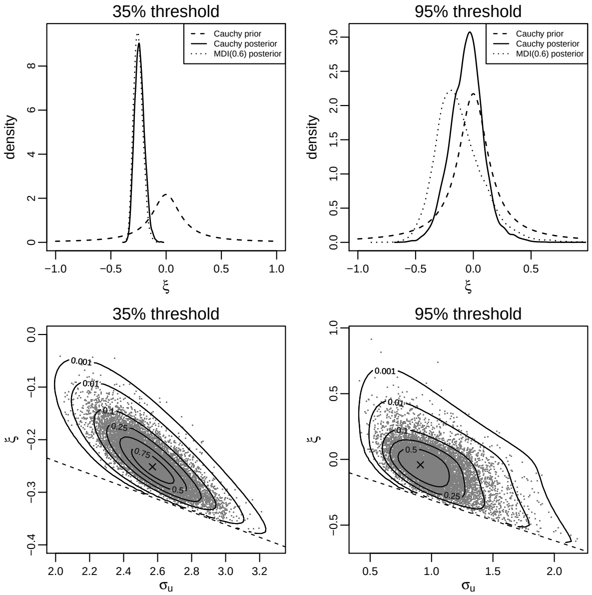
$P(\xi > 1/2) = 0.05$ *a priori*

...based on expert opinion about ratio of 10,000 yr max to 100 yr max

65% threshold : change of prior has little impact

95% threshold : low posterior probability on large ξ

A weakly-informative (Cauchy) prior



North Sea data

35% threshold : change of prior has virtually no impact

95% threshold : Cauchy prior shrinks $\pi(\xi \mid \boldsymbol{x})$ towards 0

Bayesian model averaging (BMA)

Used by Sabourin et al. (2013) for MV EV models

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:

$$P_v(u_i \mid \boldsymbol{x}) = \frac{f_v(\boldsymbol{x} \mid u_i)\, P(u_i)}{\sum_{i=1}^{k} f_v(\boldsymbol{x} \mid u_i)\, P(u_i)},$$

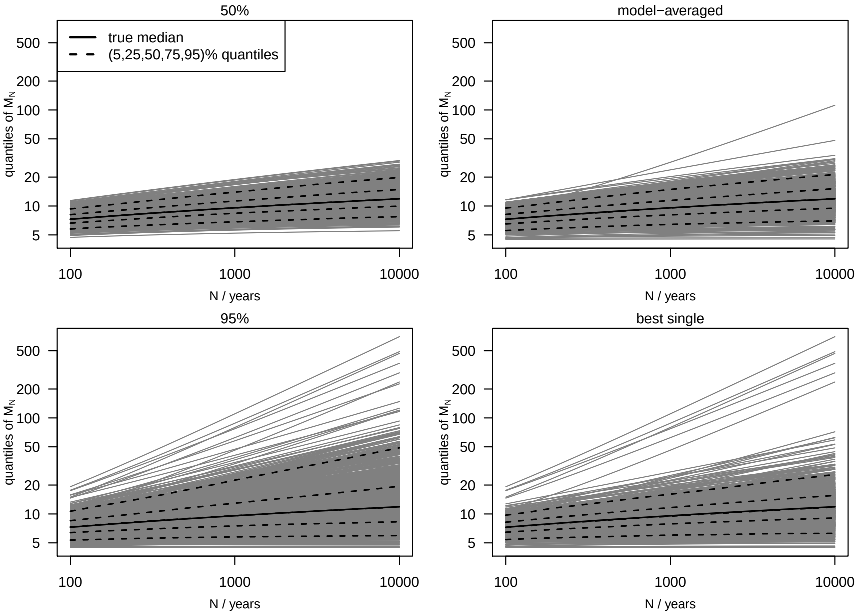
where

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

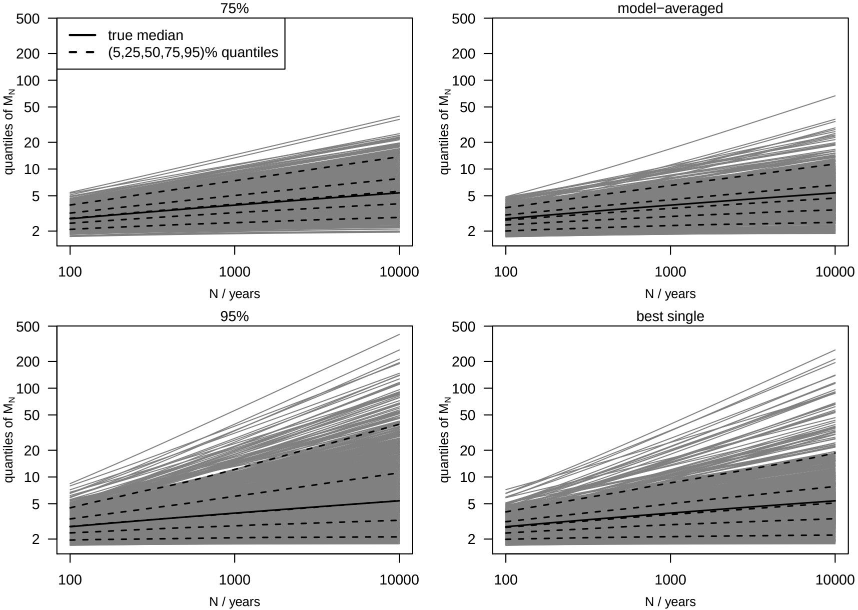
$$\widehat{f}_v(\boldsymbol{x} \mid u_i) = \prod_{r=1}^{n} f_v(x_r \mid \boldsymbol{x}_{-r}, u_i) = \exp\{\widehat{T}_v(u_i)\}$$ [Geisser and Eddy (1979)]

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

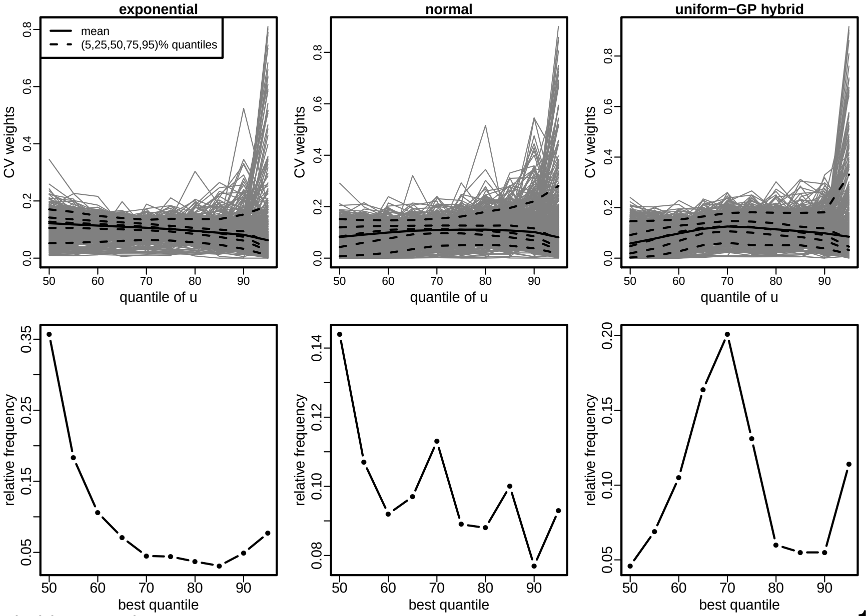
Simulation: exponential (v=95% quantile)



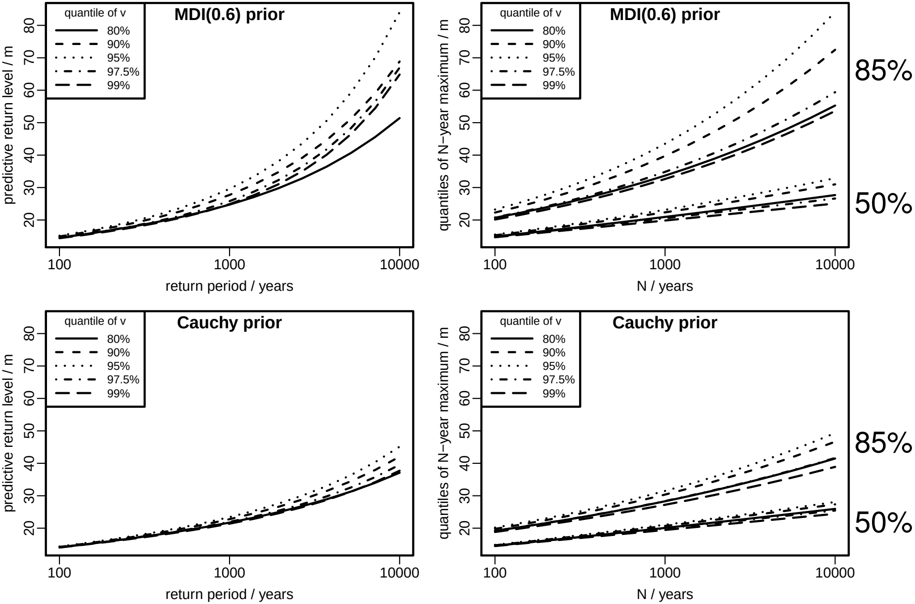
Simulation: uniform-GP(ξ=0.1) hybrid



Threshold weights



Gulf of Mexico : BMA



Concluding remarks

- Cross-validation used to address bias-variance trade-off
  - Could automate: pick ‘best’ threshold
- Threshold uncertainty : Bayesian model averaging
- Subjective inputs
  - Priors: reference, weakly-informative, informative
  - Level of validation threshold v
- On-going ...
  - serial dependence
  - multivariate extremes
  - covariate effects
  - choice of measurement scale

Concluding remarks

- Cross-validation used to address bias-variance trade-off
  - Could automate: pick 'best' threshold
- Threshold uncertainty : Bayesian model averaging
- Subjective inputs
  - Priors: reference, weakly-informative, informative
  - Level of validation threshold v
- 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.