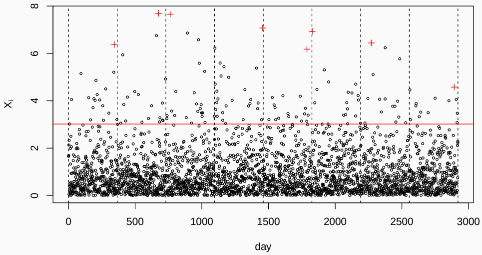
Novel threshold selection for univariate extremes

Conor Murphy
Jonathan Tawn, Zak Varty

Lancaster University, Shell

Background

- Number of methods for extreme value inference
- Block maxima, peaks over threshold, mixture models
- Focus on POT



Challenge:

- Selection of appropriate threshold
- Bias-variance trade-off

Threshold modelling

Suppose, $X_1, \ldots, X_n$ is a sequence of iid random variables, with common distribution function F.

For $X > u$, the distribution of $Y = X - u$ converges to the generalised Pareto distribution (GPD) as $u \to x^F$.

In practice, a suitably high threshold u is chosen, and the excesses Y are modelled by a GPD(σ_u, ξ) with distribution function

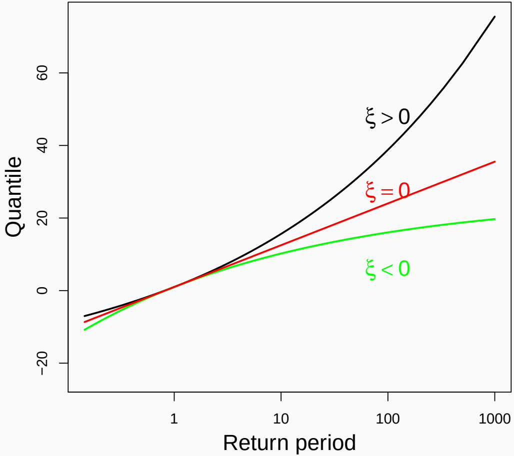
$$H(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma_u}\right)_+^{-1/\xi}, & \xi \neq 0, \\ 1 - \exp\left(-\frac{y}{\sigma_u}\right), & \xi = 0, \end{cases} \tag{1}$$

with $y > 0$, $w_+ = \max(w, 0)$, shape parameter $\xi \in \mathbb{R}$ and the threshold-dependent scale parameter $\sigma_u > 0$.

Note: if excesses of u are GPD(σ_u, ξ), then excesses of $v > u$ are also GPD(σ_v, ξ) with $\sigma_v = \sigma_u + \xi(v - u)$.

Threshold modelling

- For $\xi < 0$, the distribution of X has a finite upper end-point at $u - \sigma_u/\xi$.
- The distribution is unbounded above for $\xi \geq 0$.



Threshold selection

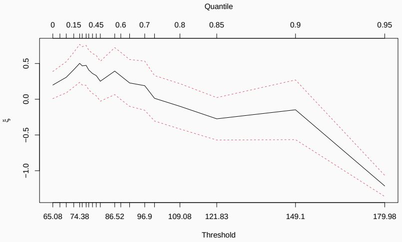
The most fundamental part of a threshold-based model!

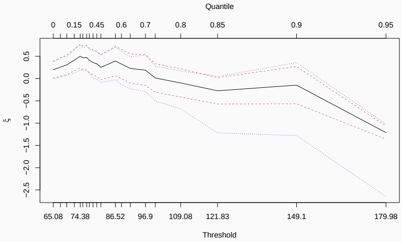
Standard methods:

- MRL plots
- Rule of thumb methods
- Parameter stability plots (most widely used!)

River Nidd dataset:

- The River Nidd is a tributary in North Yorkshire.
- 154 observations of river flow above a threshold of 65 m^3/s.





-> **Clearly automated methods are needed!**

Automated methods

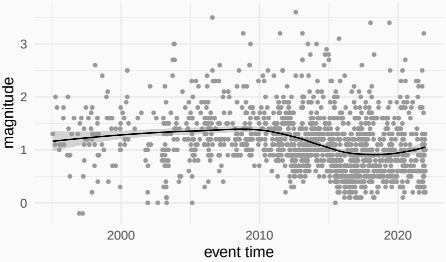
Lancaster University

- Wadsworth (2016) utilises the asymptotic joint distribution of MLEs:
  - Main result: $(\hat{\xi}_1^*, \ldots, \hat{\xi}_{k-1}^*)^T \to \boldsymbol{Z}$ where $\boldsymbol{Z} \sim N_{k-1}(\mathbf{0}, \mathbf{1}_{k-1})$.
  - Simple changepoint model.
  - Likelihood ratio test if $\hat{\xi}_i^* \sim N(\beta, \gamma)$.

- Northrop et al. (2017) use leave-one-out cross-validation in a Bayesian framework:
  - Compare predictive ability.
  - Average inferences over posterior distribution of parameters.
  - Importance sampling.

Motivation:

- Production of oil/gas can cause earthquakes
- Largest 3.6 M_L but shallow
- Substantial damage

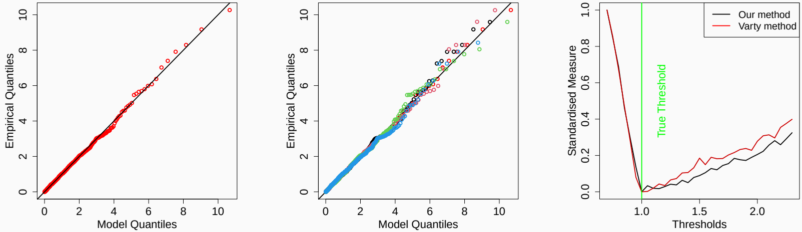


Challenge:

- Partial censoring due to development of geophone network
- Network too sparse/insensitive to detect low magnitude events

Our method

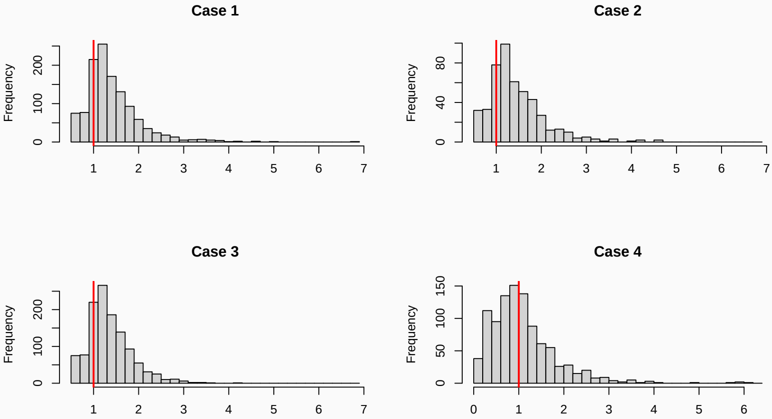
- Compares the deviation from the line of equality on a QQ-plot across replications for each threshold.
- Result: A set of metric values corresponding to each proposed threshold.



$$d_{(i)} = \frac{1}{m}\sum_{j=1}^{m} |M_{(i)}(p_j) - Q_{(i)}(p_j)| \qquad d = \frac{1}{k}\sum_{i=1}^{k} d_{(i)}. \qquad (2)$$

Simulation study

Examples of simulated datasets:



Results

	Our method			*Wadsworth**			*Northrop*		
Case	RMSE	Bias	Variance	RMSE	Bias	Variance	RMSE	Bias	Variance
Case 1	**5.3**	**3.4**	**0.2**	41.3	15.1	14.8	52.7	25.7	21.1
Case 2	**5.5**	**3.0**	**0.2**	43.9	18.8	15.8	54.5	26.9	22.5
Case 3	**7.2**	4.6	**0.3**	13.7	**3.9**	1.7	42.7	22.9	12.9
Case 4	**10.2**	**6.8**	**0.6**	38.5	7.2	14.3	48.9	15.0	21.7

-> Our method achieves RMSEs **between 1.9 and 8 times smaller** than the Wadsworth (2016) method, always with **lower variance** and in 3 out of 4 cases, is **the least biased**.

Tables have been scaled by a factor of 100

*Results for Wadsworth are calculated only on the samples where a threshold was estimated. The method failed to estimate a threshold for 2%, 28%, 0.2%, 4% of the simulated datasets in Cases 1-4.

Quantile estimation

True quantiles from the simulated distributions can be calculated as follows:

$$x_p = 1 + \frac{\sigma_1}{\xi}\left[\left(\frac{6p}{5}\right)^{-\xi} - 1\right] \qquad y_p = 1 + \frac{\sigma_1}{\xi}\left[\left(\frac{p}{1-q}\right)^{-\xi} - 1\right]$$

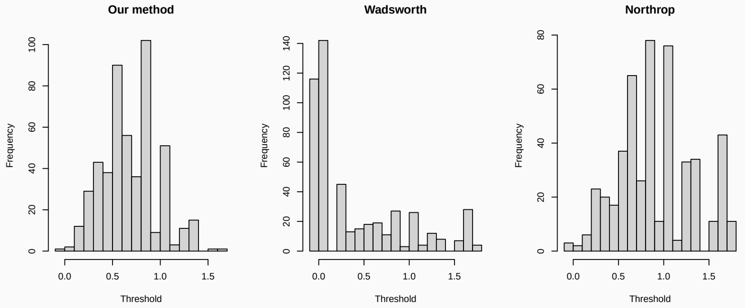
p	*Our method*	*Wadsworth**	*Northrop*	*Our method*	*Wadsworth**	*Northrop*
	Case 1			**Case 2**		
$1/n$	**5.8**	6.1	7.4	**6.2**	**6.2**	7.4
$1/10n$	**13.3**	14.7	20.8	**15.3**	15.8	26.4
$1/100n$	**26.2**	28.9	52.9	**32.2**	33.9	93.6
	Case 3			**Case 4**		
$1/n$	**2.0**	2.0	2.5	**7.0**	7.7	8.5
$1/10n$	**3.3**	**3.4**	4.8	**16.5**	19.4	26.6
$1/100n$	**4.9**	**5.0**	8.2	**33.3**	40.1	84.9

Tables have been scaled by a factor of 10

*Results for Wadsworth are calculated only on the samples where a threshold was estimated.The method failed to estimate a threshold for 2%, 28%, 0.2%, 4% of the simulated datasets in Cases 1-4.

Gaussian data

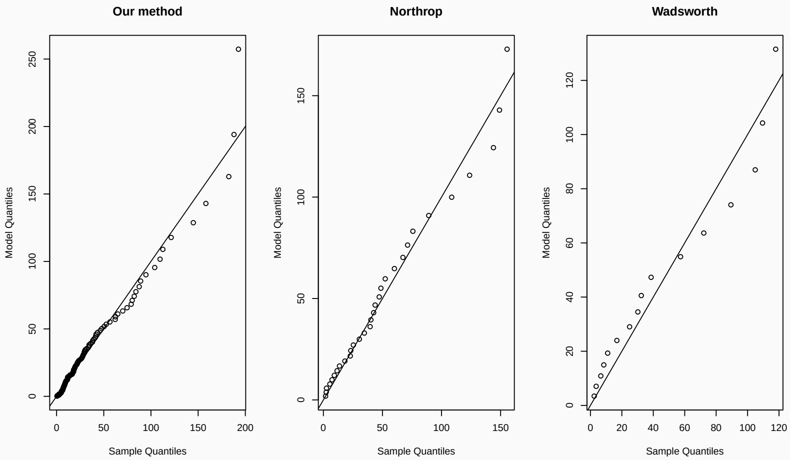
Gaussian Case			
p	*Our method*	*Wadsworth**	*Northrop*
$1/n$	**2.1**	2.5	**2.3**
$1/10n$	**4.3**	5.4	**4.6**
$1/100n$	**7.0**	9.0	7.7



Tables have been scaled by a factor of 10

*Results for Wadsworth are calculated only on the samples where a threshold was estimated. In this case, the method failed to obtain an estimate for 0.4% of the samples.

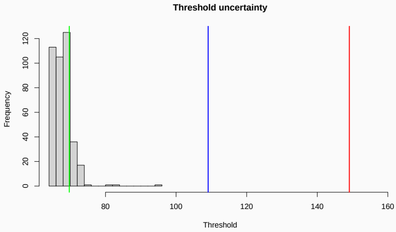
River Nidd



Our method
Northrop
Wadsworth
Model Quantiles
Sample Quantiles
0
50
100
150
200
250
0
50
100
150
0
20
40
60
80
100
120

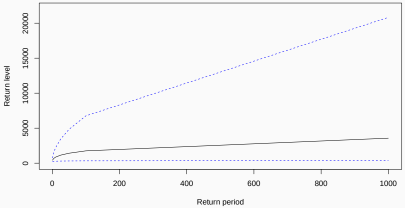
Uncertainty

- Sensitivity to proposed set of thresholds.
- Low threshold uncertainty for our method.



Further work:

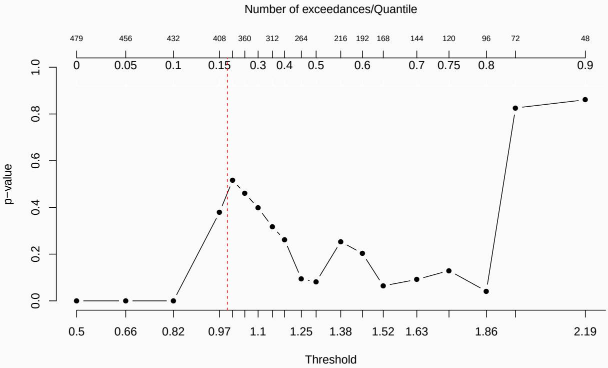
- Comparison between methods.
- Include other methods.
- Extensions of method.



Thanks for listening!

Developments from parameter stability plots

- Northrop and Coleman (2014) developed a multiple-threshold GPD model -> likelihood ratio and score tests to assess stability.



-> **Clearly automated methods are needed!**

Simulated from two distributions:

$$F_1(x) = \begin{cases} \frac{x-0.5}{3}, & 0.5 \leq x \leq 1 \\ \frac{1}{6} + \frac{5}{6}\left[H(x-1; 0.5, 0.1)\right], & x > 1. \end{cases}$$

$$F_2(x) = \begin{cases} \int_0^x h(x; 0.5, 0.1)\mathbb{P}(B < x)\mathrm{d}x, & 0 \leq x \leq 1 \\ q + (1-q)\left[H(x-1; 0.5, 0.1)\right], & x > 1. \end{cases}$$

where $q = \int_0^1 h(x; 0.5, 0.1)\mathbb{P}(B < x)\mathrm{d}x$.

Breakdown of RMSE:

- Bias and variance of threshold choice for GPD data.

	Our method			*Varty method*		
n	RMSE	Bias	Variance	RMSE	Bias	Variance
1000	**9.4**	4.7	0.7	10.7	5.0	0.9
10000	**13.2**	3.5	1.6	13.3	3.8	1.6
40000	**5.8**	2.7	0.2	8.1	3.3	0.5

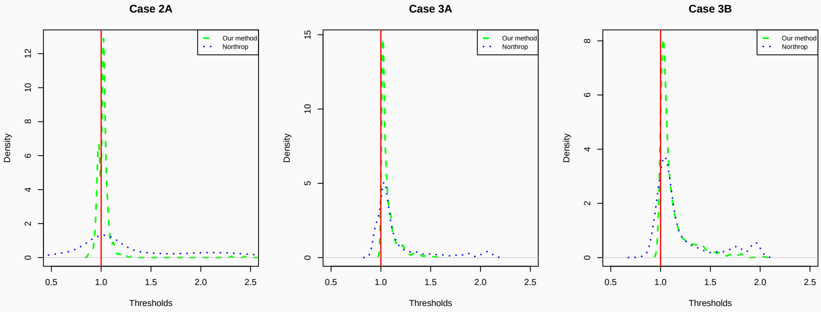
- Bias and variance of quantile estimation for Gaussian data.

	Our method			*Varty method*		
n	RMSE	Bias	Variance	RMSE	Bias	Variance
1000	**72.8**	62.6	13.9	79.3	70.3	13.5
10000	**38.0**	25.2	8.1	42.0	30.5	8.3
40000	**23.6**	16.6	2.8	24.8	18.1	2.9

Table values have been scaled by a factor of 100

Results

Comparison in cases where Wadsworth (2016) broke down:



- Small sample of 120
- Same number of thresholds
- Case 3A: $\xi = -0.2$
- Case 3B: $\xi = -0.3$

-> **Our method achieves accurate results in all cases!**

References

Northrop, P. J., Attalides, N., and Jonathan, P. (2017). Cross-validatory extreme value threshold selection and uncertainty with application to ocean storm severity. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 66(1):93–120.

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

Varty, Z., Tawn, J. A., Atkinson, P. M., and Bierman, S. (2021). Inference for extreme earthquake magnitudes accounting for a time-varying measurement process. *arXiv preprint arXiv:2102.00884*.

Wadsworth, J. L. (2016). Exploiting structure of maximum likelihood estimators for extreme value threshold selection. *Technometrics*, 58(1):116–126.