

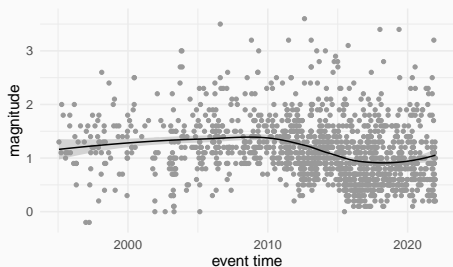
Automated threshold selection and associated inference uncertainty

Conor Murphy

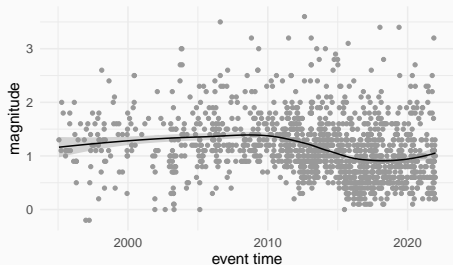
Jonathan Tawn, Zak Varty, Peter Atkinson, Ross Towe

Lancaster University, Shell

- Production of oil/gas can cause earthquakes.
- Low magnitude events at shallow depths.
- Similar characteristics at CO₂ storage sites.



- Missed observations due to development of geophone network.
- Network too sparse/insensitive to detect low magnitude events.



⇒ Improved forecasting of seismic hazards under future extraction scenarios...

Constant threshold selection

Why is threshold selection important?

- Parameter estimates
- Quantiles/Return levels
- Uncertainty

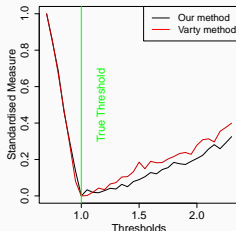
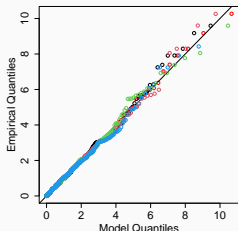
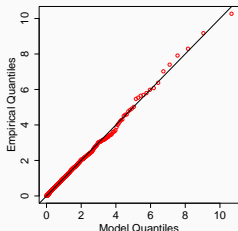
Often chosen by rule of thumb or subjective methods!

Challenge:

⇒ Bias-variance trade-off

Expected Quantile Discrepancy (EQD)

- **Input:** Data, Set of candidate thresholds.
- **Method:** Expected deviation between model and sample quantiles.
- **Output:** EQD for each candidate.



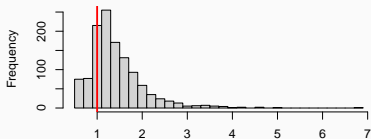
$$d_b(u) = \frac{1}{m} \sum_{j=1}^m |M_b(p_j) - Q(p_j, \mathbf{x}_u^b)|$$

$$d_E(u) = \frac{1}{k} \sum_{b=1}^k d_b(u).$$

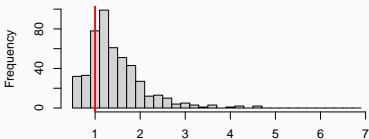
- Wadsworth (2016) utilises the asymptotic joint distribution of MLEs:
 - Consider $\hat{\xi}_i^* = \frac{\hat{\xi}_i - \hat{\xi}_{i+1}}{\nu_i}$ the standardised increments.
 - Main result: $(\hat{\xi}_1^*, \dots, \hat{\xi}_{k-1}^*)^T \rightarrow \mathbf{Z}$ where $\mathbf{Z} \sim N_{k-1}(\mathbf{0}, \mathbf{1}_{k-1})$ above u^* .
 - Changepoint model and likelihood ratio test.
- Northrop et al. (2017) use leave-one-out cross-validation in a Bayesian framework:
 - Assess predictive ability using candidate u at $v > u$.
 - Average inferences over posterior distribution of parameters.
 - Maximise measure of predictive performance.

Examples of simulated datasets:

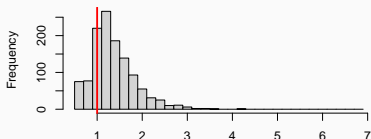
Case 1



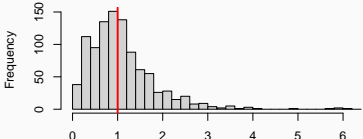
Case 2



Case 3



Case 4



	<i>EQD</i>	<i>Wadsworth*</i>	<i>Northrop</i>
Case 1	5.3	41.3	52.7
Case 2	5.5	43.9	54.5
Case 3	7.2	13.7	42.7
Case 4	10.2	38.5	48.9

- ⇒ EQD achieves RMSEs **between 1.90 and 7.98 times smaller** than the Wadsworth (2016) method.
- ⇒ Always with **lower variance**.
- ⇒ In 3 out of 4 cases, EQD 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

	Case 1			Case 2		
p	<i>EQD</i>	<i>Wadsworth*</i>	<i>Northrop</i>	<i>EQD</i>	<i>Wadsworth*</i>	<i>Northrop</i>
$1/n$	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		
p	<i>EQD</i>	<i>Wadsworth</i>	<i>Northrop</i>	<i>EQD</i>	<i>Wadsworth*</i>	<i>Northrop</i>
$1/n$	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

⇒ EQD achieves smallest RMSEs in all cases again!

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.

Results for Gaussian data

Gaussian Case			
p	<i>EQD</i>	<i>Wadsworth</i> [*]	<i>Northrop</i>
$1/n$	2.1	2.5	2.3
$1/10n$	4.3	5.4	4.6
$1/100n$	7.0	9.0	7.7

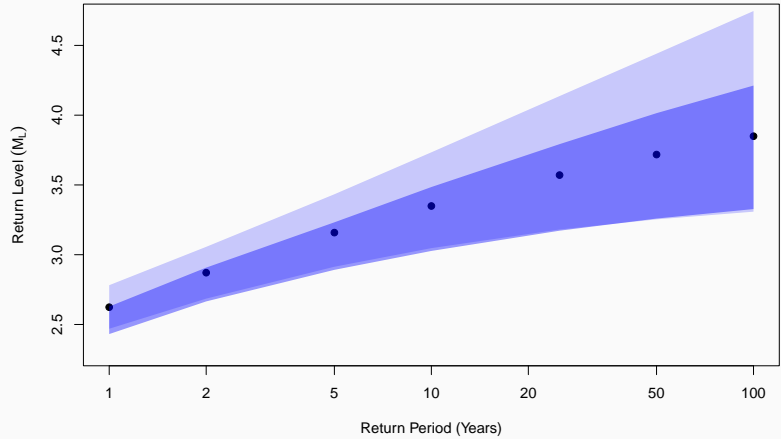
- ⇒ EQD outperforms existing threshold selection methods.
- ⇒ Enables more accurate quantile estimation.

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.

Uncertainty

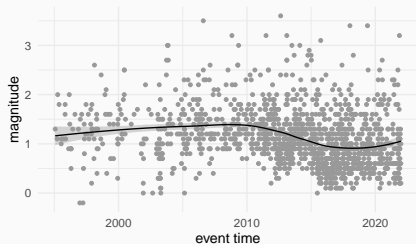
- Reliance on point estimates can be dangerous.
- Threshold uncertainty often omitted!
- Double-bootstrap procedure to incorporate different uncertainties.



What we have so far

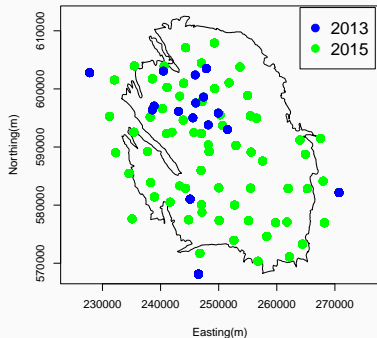
- Working method for constant threshold selection.
- Dataset with missing observations.
- Varty et al. (2021) incorporated time-varying data quality into threshold.

⇒ What now?



Covariate inclusion

- Missed observations caused by geophone network.
- Spatial variability also present.
- Earthquake needs to be detected by three or more geophones.
- Can we use this as covariate?
- $V_{\text{geo}}(x, t)$ = distance to third-nearest geophone.



Model given by:

$$u(x, t) = \theta V_{\text{geo}}(x, t)$$

$$Y - u(x, t) | Y > u(x, t) \sim \text{GPD}(\sigma_0 + \xi u(x, t), \xi)$$

- Covariate known for each seismic event.
- Given $V_{\text{geo}}(x, t)$, can estimate θ using same method.

Assumption: (σ_0, ξ) are constant!

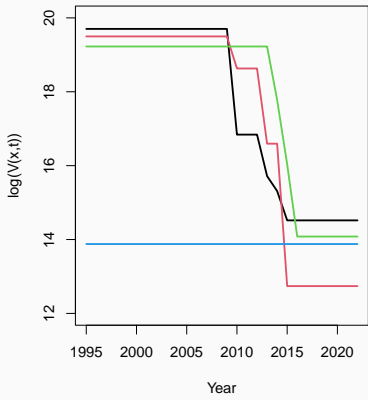
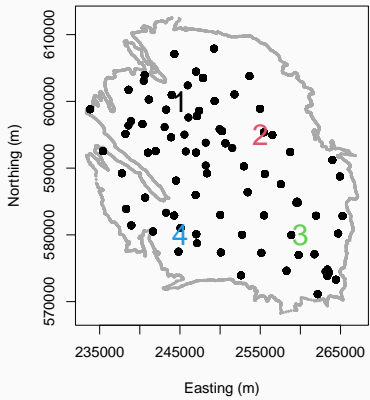
Spatio-temporal modelling

- Assess variability in (σ_0, ξ) with spatio-temporal threshold.
- Utilise other relevant covariates for GPD model.
- Explore more complex relationships between $u(x, t)$ and $V_{\text{geo}}(x, t)$.

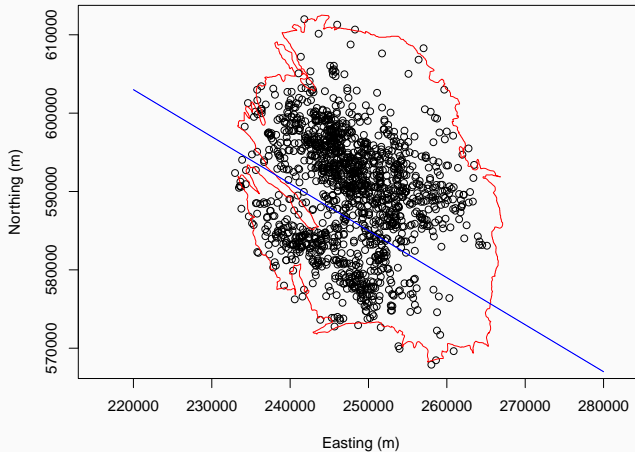
Combined threshold & model selection

1. Adjust threshold selection method for all desired GPD parameterisations.
2. Transform to common margins and record $\min(EQD)$.
3. Select model which minimises $\min(EQD)$ values.

Thanks for listening!



Should the parameters (σ_0, ξ) vary spatially?



Likelihood ratio tests

Thresholds	Models	GPD(σ, ξ)	GPD(σ_R, ξ)
$u = 1.07$	GPD(σ_R, ξ)	0.000	NA
	GPD(σ_R, ξ_R)	0.000	0.036 ↑
$u = 1.318$	GPD(σ_R, ξ)	0.158	NA
	GPD(σ_R, ξ_R)	0.357	0.797
$(u_U, u_L) = (1.2, 0.876)$	GPD(σ_R, ξ)	0.001	NA
	GPD(σ_R, ξ_R)	0.001	0.064 ↑

⇒ Evidence to suggest GPD scale parameter varies over region.

Next steps:

- Compare above models using appropriate thresholds for all cases... How?

Simulated from two distributions:

$$F_1(x) = \begin{cases} \frac{x-0.5}{3}, & 0.5 \leq x \leq 1 \\ \frac{1}{6} + \frac{5}{6} [H(x-1; 0.5, 0.1)], & x > 1. \end{cases}$$
$$F_2(x) = \begin{cases} \int_0^x h(x; 0.5, 0.1) \mathbb{P}(B < x) dx, & 0 \leq x \leq 1 \\ q + (1-q) [H(x-1; 0.5, 0.1)], & x > 1. \end{cases}$$

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

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

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

Breakdown of RMSE:

- Bias and variance of threshold choice for GPD data.

n	<i>Our method</i>			<i>Varty method</i>		
	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.

n	<i>Our method</i>			<i>Varty method</i>		
	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

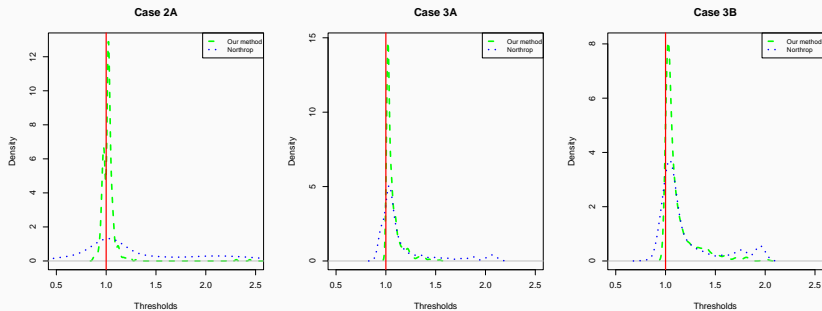
	<i>Our method</i>			<i>Wadsworth*</i>			<i>Northrop</i>		
	RMSE	Bias	Variance	RMSE	Bias	Variance	RMSE	Bias	Variance
Case 1	5.3	3.4	0.2	41.3	15.1	14.8	52.7	25.7	21.1
Case 2	5.5	3.0	0.2	43.9	18.8	15.8	54.5	26.9	22.5
Case 3	7.2	4.6	0.3	13.7	3.9	1.7	42.7	22.9	12.9
Case 4	10.2	6.8	0.6	38.5	7.2	14.3	48.9	15.0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

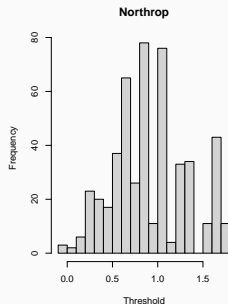
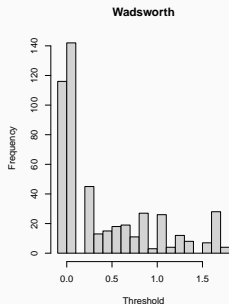
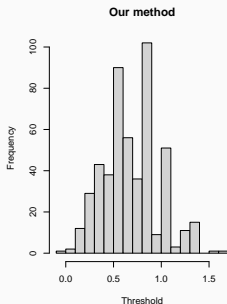
*Results for Wadsworth are calculated only on the samples where a threshold was estimated. The method failed to estimate a threshold for 2%, 28%, 0.2%, 4% of the simulated datasets in Cases 1-4.

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!**

Gaussian Case			
p	<i>EQD</i>	<i>Wadsworth</i> *	<i>Northrop</i>
$1/n$	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.



Further work

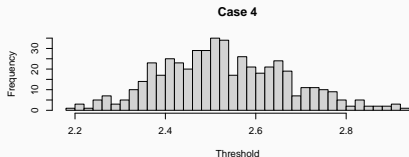
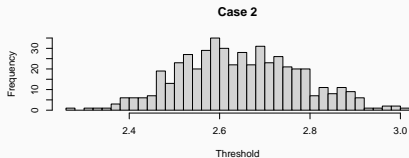
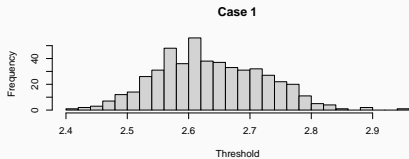
Compare against other existing methods in fixed threshold selection.

Danielsson et al. (2019):

- Quantile-driven approach.
- Maximum distance between empirical and model quantiles.

Applied to River Nidd dataset:

⇒ $u = 189.02$.



- Danielsson, J., Ergun, L., de Haan, L., and de Vries, C. G. (2019). Tail Index Estimation: Quantile-Driven Threshold Selection. Staff Working Papers 19-28, Bank of Canada.
- Northrop, P. J., Attalides, N., and Jonathan, P. (2017). Cross-validators 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.
- 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.