Spatio-temporal Threshold Selection for Induced Seismicity

Conor Murphy[1] c.murphy4@lancaster.ac.uk https://tinyurl.com/cmurphy4

STOR-i Lancaster University

1. Motivation

Context

- Production of oil and gas can cause (shallow, low magnitude) earthquakes ⇒ **substantial damage**. Example: **Groningen gas field** in the Netherlands.
- Potential **impact** of rare high levels of seismicity ⇒ careful modelling of the earthquake magnitudes needed ⇒ forecast hazards under future extraction scenarios.

Challenge: **Partially censored** data.
Cause: Geophone recording network too sparse and insensitive to detect all **low magnitude** events. Fewer low magnitude events **censored** at later times (shown in plot below).

Aim: Use spatial information to further improve the modelling of low magnitude, but potentially high impact, seismic events.
Result: Novel **threshold selection** technique useful for extreme value contexts.

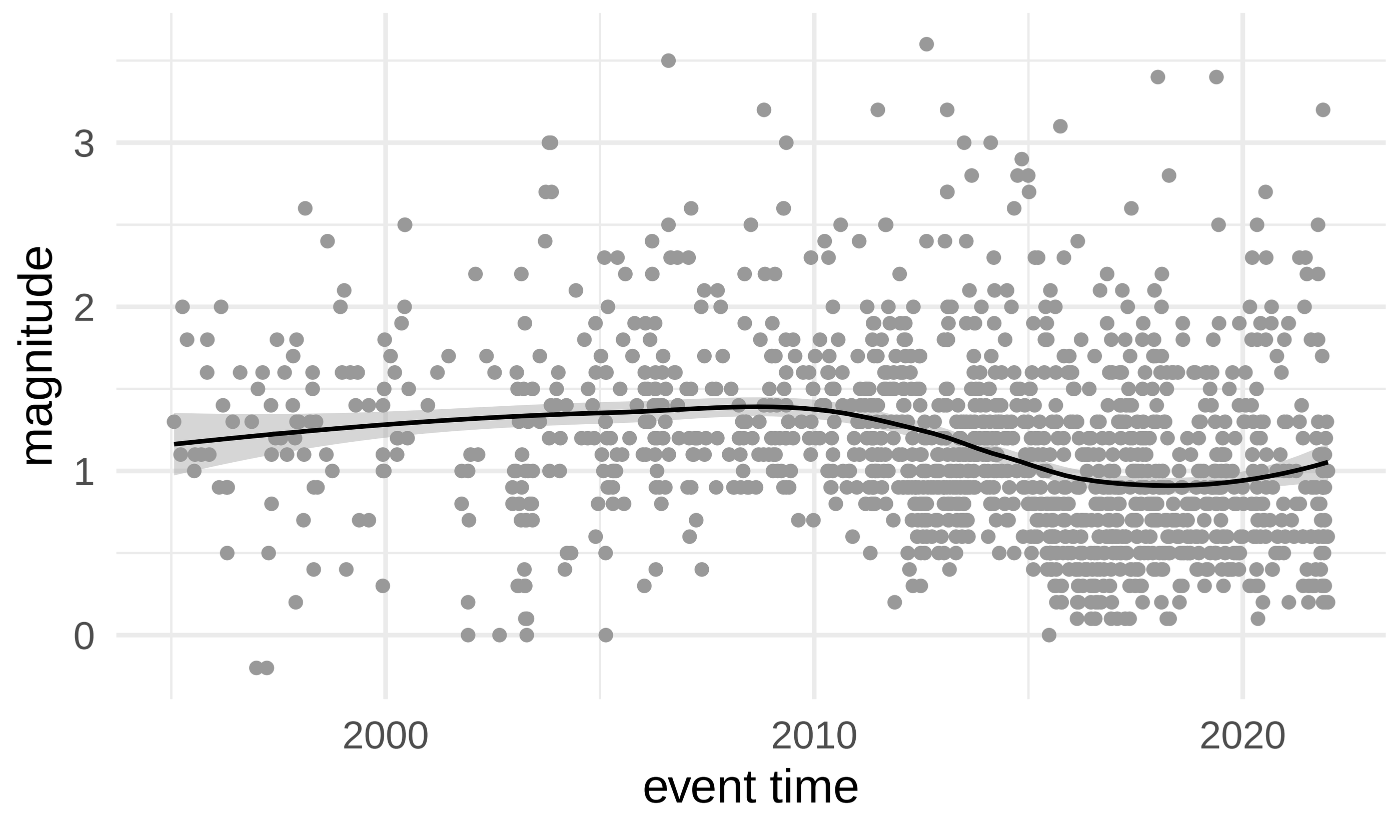


Figure 1: Groningen magnitudes over time with solid line indicating a smoothed estimate for the mean.

2. Method

Background

- **Physical models** suggest magnitudes are approximately **exponential** (Exp).
- **Extreme Value Theory** provides the **generalised Pareto distribution** (GPD) to model excesses of a suitably high threshold u - the Exp distribution is a special case ($\xi = 0$).
- For induced seismicity, physical arguments justify the assumption that magnitudes are i.i.d. and that **partial censoring** is the sole reason for deviation from this assumption.

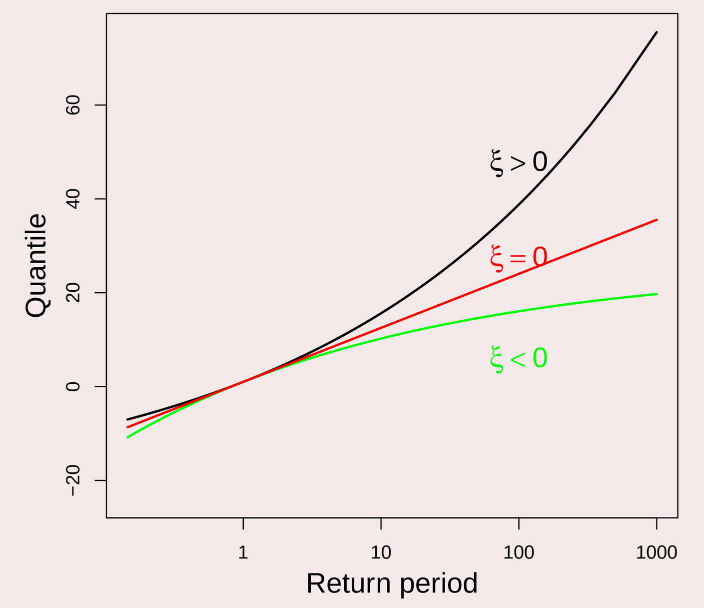


Figure 2: Return level plot showing different cases of GPD.

Previous work

- Estimated a **time-varying threshold** above u such that observed data ∼ GPD.
- *Method*: - **Bootstrap samples** of excesses of each threshold choice.
  - **Transform** magnitudes under fitted GPD onto **shared Exp(1) margins**.
  - Compare **mean-absolute distances** of empirical and true quantiles.
- **The threshold which minimises this overall distance is selected**.

Method directly tackles **bias-variance trade-off**: we want to select a threshold **as low as possible** (to minimise variance) while also achieving a **good fit** of the GPD (minimising bias).

3. Constant Threshold Selection

Developed methods based on Varty et al. (2021) (*https://tinyurl.com/varty21*):

- **Exp measure**: transforms the magnitudes onto standard Exp margins and compares against Exp quantiles with rate 1.
- **GPD measure**: compares quantiles of the sampled excesses of a threshold against GPD quantiles calculated using **maximum likelihood estimates** for (σ, ξ) of GPD .

Mean distances d are calculated over k replications from the following formula with y denoting original/transformed empirical quantiles at m equally-spaced probabilities for GPD/Exp measure:

$$d = \frac{1}{k}\sum_{i=1}^{k} d_i, \quad \text{where} \quad d_i = \frac{1}{m}\sum_{j=1}^{m} |y_j - Q_i(p_j)|,$$

with $Q_i(p) = \frac{\hat{\sigma}_i}{\hat{\xi}_i}\left[(1-p)^{-\hat{\xi}_i} - 1\right]$ for GPD measure and $Q_i(p) = -\log(1-p)$ for Exp measure.

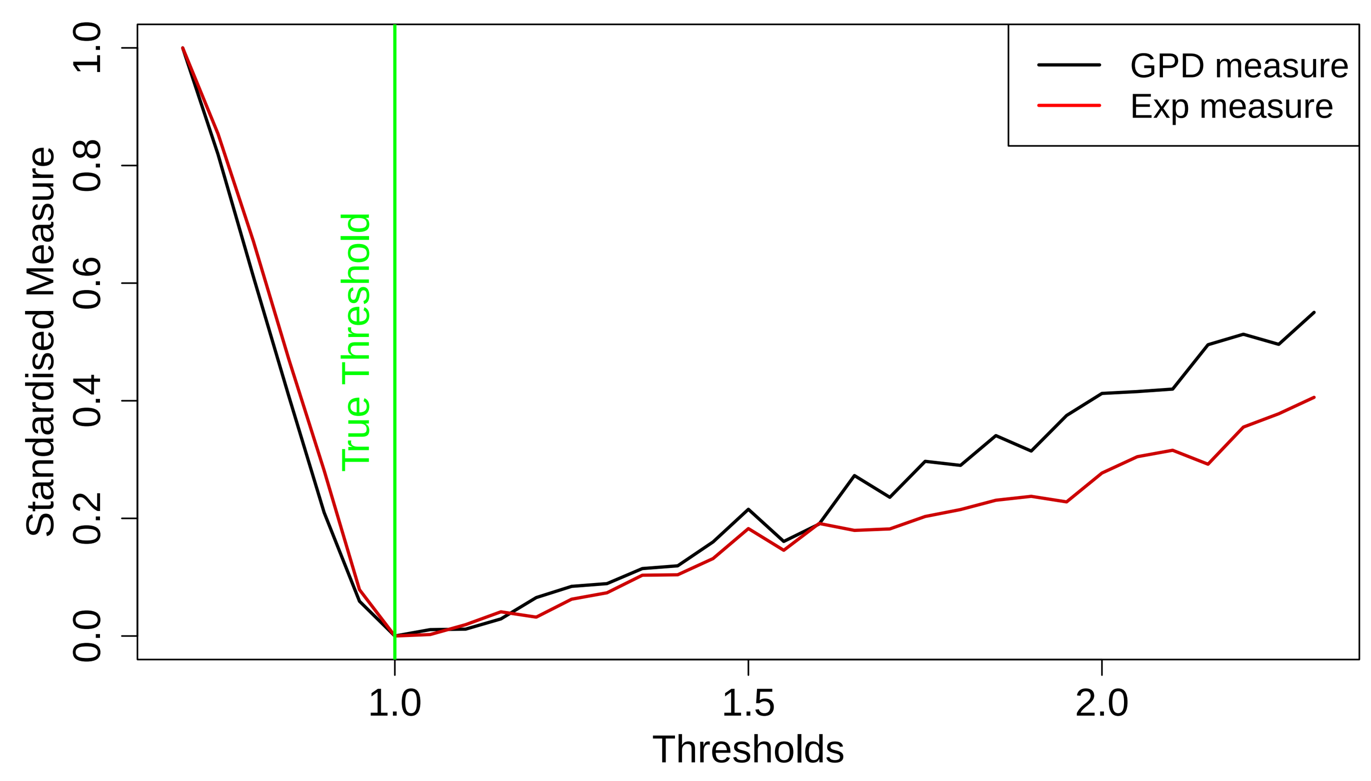
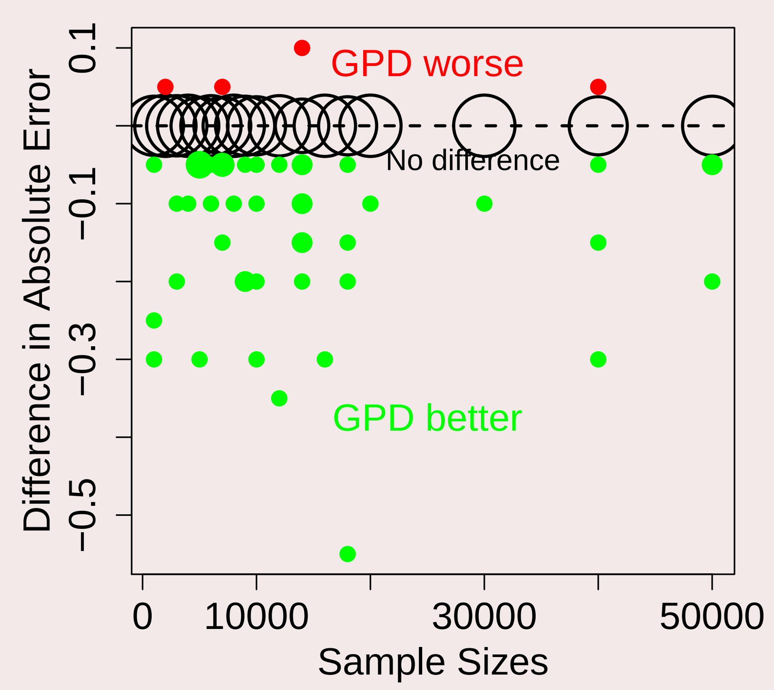
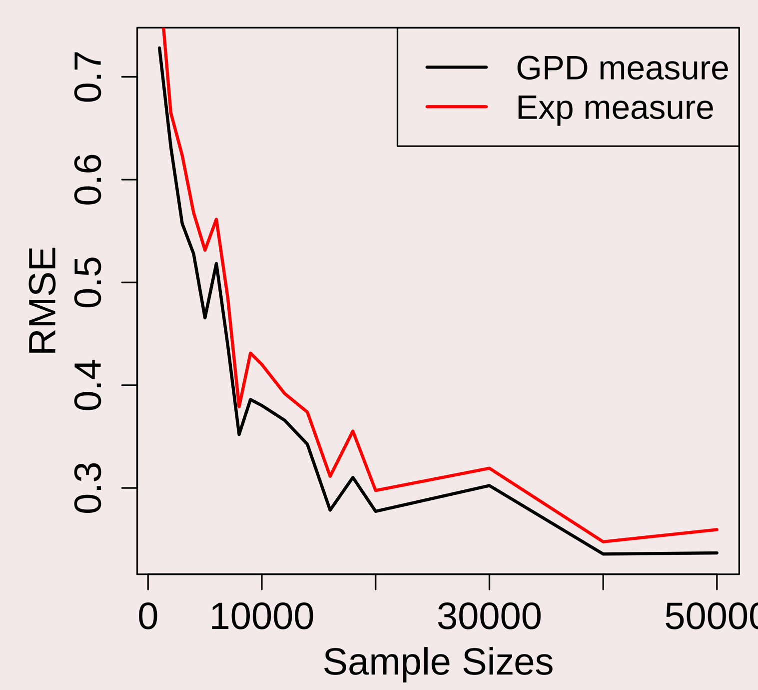


Figure 3: Standardised overall mean distances for each threshold on simulated GPD dataset ($u = 1.0$).

4. I.I.D. Case Studies

- **Known threshold**: Simulated **GPD data with true threshold** $u = 1.0$.
  ⇒ Compare **difference in absolute error from the true threshold** (see left plot below).
- **Unknown threshold**: Simulated **Gaussian data** where **true threshold does not exist**.
  ⇒ Compare RMSEs of **fitted quantiles** for estimated threshold choices (see right plot below).





- GPD measure performs better than Exp measure in threshold selection for all sample sizes.
- RMSE of quantile estimates decreases for both methods as the sample size increases.

Current research: Compare against existing methods for extreme value threshold selection.

5. Spatial Threshold Selection

Challenge: **Spatial variation in the threshold** above which earthquakes are detected.
Cause: **Density of geophone network varies spatially**.

Earthquakes must be **detected by three geophones** ⇒ Link between earthquake detection probability and its distance from the three nearest geophones.

Approach:

- $V(x)$ measures **squared distance of an earthquake location** x **from all geophones** and records the **third smallest**.
- $V(x)$ scaled by a factor θ such that the censoring threshold (red line) $u(x) = \theta V(x)$, where θ represents the geophone sensitivity.
- **Above threshold selection technique adjusted to estimate** θ **given the value of** $V(x)$ **and the magnitude at each observed location.**

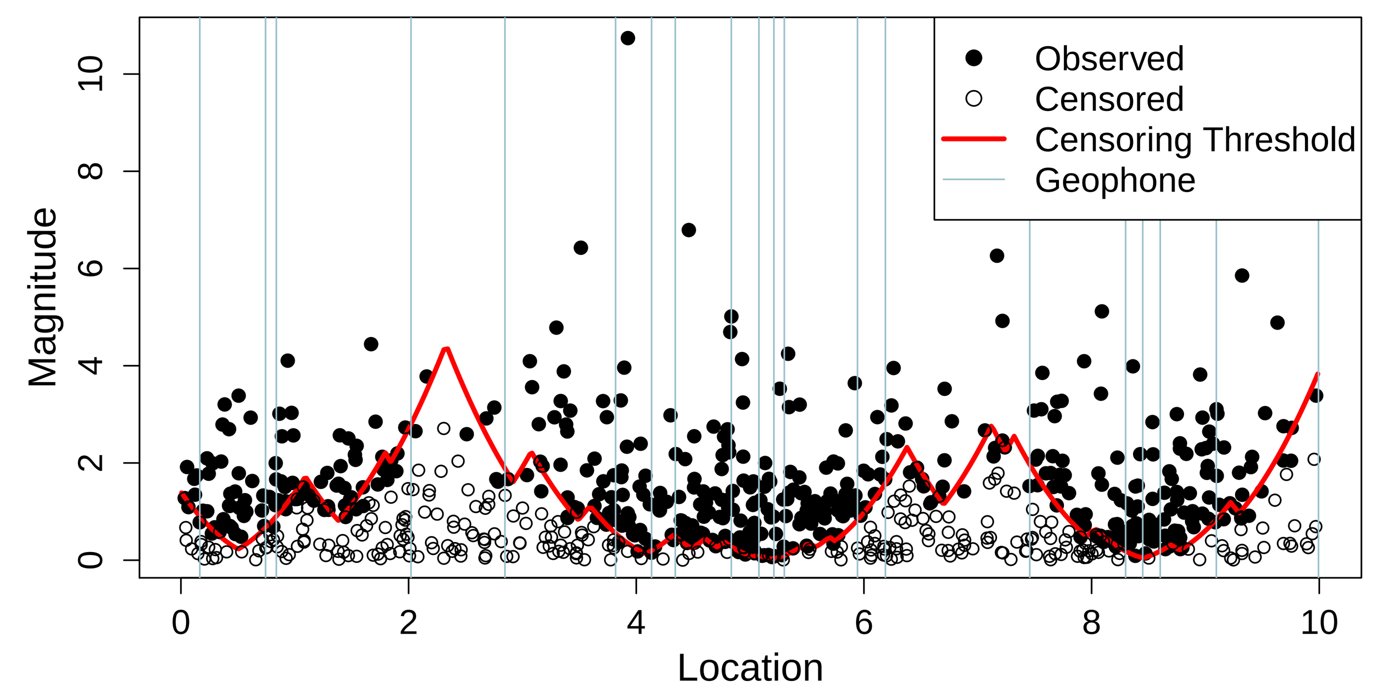
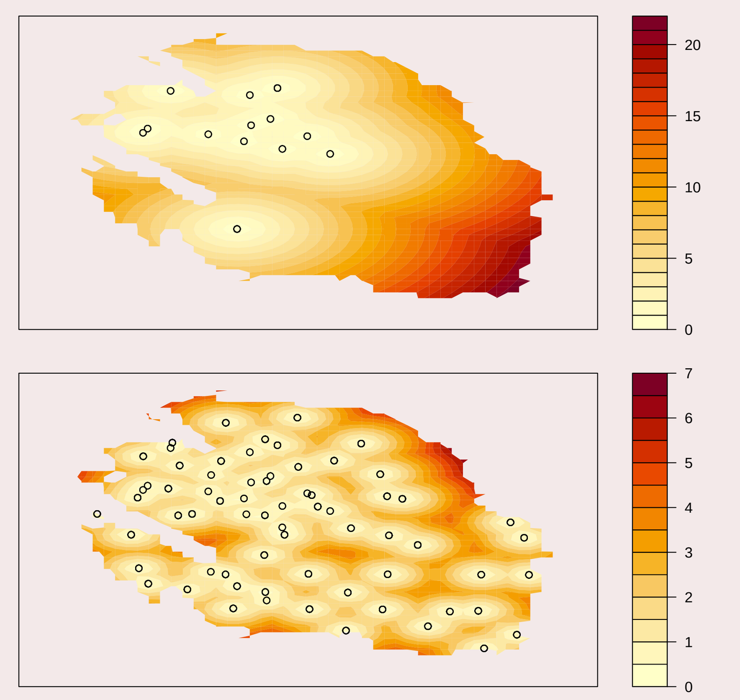


Figure 4: Earthquake/geophone locations simulated uniformly at random across region (0,10); magnitudes generated from a GPD.

6. Spatio-temporal Threshold Selection

Snapshots from 2013 [Top] and 2015 [Bottom] of $V(x)$ surface across gas field with points representing geophones.



Approach:

- Construct a function $u(x,t)$ showing how earthquake detection probability changes over space and time.
- This will give a physical basis for the form of the space-time-varying threshold.

7. Further Research

- Incorporate a probability distribution into the censoring process.
- Account for varying sensitivity of geophones in the network: using measurement error data.
- Allow parameters of GPD to vary spatially according to a spatial covariate, e.g., incremental Coulomb stress.
- Implement improved modelling approaches to inform seismicity forecasting under future extraction strategies to reduce the chance of high impact events occurring.

Supervisors: Jonathan Tawn[1], Peter Atkinson[1], Zak Varty[2] and Ross Towe[3] [1]Lancaster University, [2]Imperial College London, [3]Shell

UKRI Engineering and Physical Sciences Research Council