Statistical Modelling for Earthquakes Accounting for Measurement Error

Kushagra Agarwal[1]

Supervised by: Wanchen Yue[2]

[1]University of Warwick [2]Lancaster University

30/08/24

Introduction & Background

- Groningen gas field was founded in 1959 and is one of the largest gas fields in the world.
- Extraction of gas from the field began in 1963 by a joint venture by Shell and Exxon.
- The extraction of gas has led to a number of **human-induced earthquakes** (Hi-Quakes) since 1991 in the province of Groningen, and this remains a problem to this day.
- The induced seismic activity has caused damage to buildings and infrastructure in the region, leading to financial losses and safety concerns.

Figure 1: Impact of Hi-Quakes on the city of Groningen.

Hawkes Processes

A Hawkes process is a counting process $(N(t) : t \geq 0)$ with associated history $(\mathcal{H}(t) : t \geq 0)$ that satisfies:

$$\mathbb{P}(N(t+h) - N(t) = m \mid \mathcal{H}(t)) = \begin{cases} \lambda^*(t)h + \mathrm{o}(h), & m = 1 \\ \mathrm{o}(h), & m > 1 \\ 1 - \lambda^*(t)h + \mathrm{o}(h), & m = 0 \end{cases}$$

where $\lambda^*(t)$ is the intensity function of the Hawkes process and is of the form:

$$\lambda(t \mid \mathcal{H}_t) =: \lambda^*(t) = \mu + \sum_{i:t_i<t} \gamma(t - t_i) \quad (1)$$

The most common choice for the excitation function is the exponential decay, i.e., $\gamma(t) = \alpha e^{-\beta t}$, where α, $\beta > 0$ are parameters.

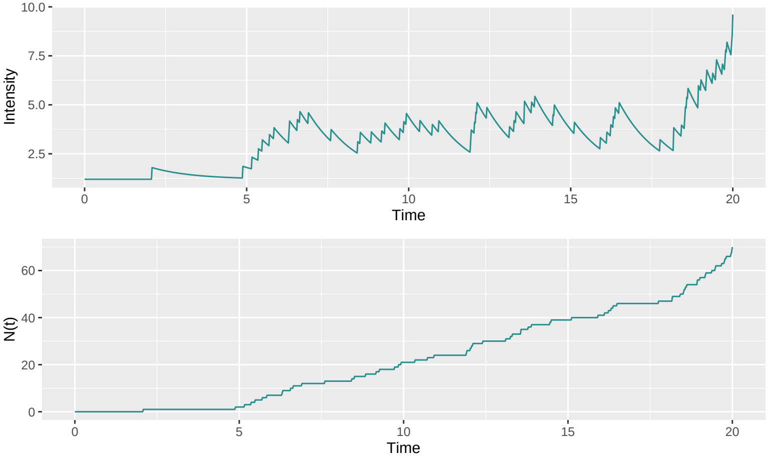

Figure 2: Realisation of a Hawkes process with conditional intensity function (above) and the corresponding counting process (below) with parameters $\mu = 1.2$, $\alpha = 0.6$, and $\beta = 0.8$.

ETAS Model

Epidemic Type Aftershock Sequence (ETAS) model is an extension of Hawkes process that is mainly used to model seismic occurrences and their aftershocks.

The modified Omori formula describes the rate of aftershocks as a power law decay function. So, the intensity function for ETAS model is given by

$$\lambda^*(t) = \mu + \sum_{i:t_i<t} \frac{K_0}{(t - t_i + c)^p} \cdot e^{\alpha(m_i - M_r)}$$

where K_0, c, p, and α are parameters, μ is the background intensity, m_i are magnitudes of earthquakes, and M_c is the pre-determined cut-off magnitude.

Extreme Value Theory

Block Maxima

- If $X_1, \ldots, X_n$ are i.i.d random variables, then there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that $\frac{M_n - b_n}{a_n} \dot{\sim}$ GEV(μ, σ, ξ) as $n \to \infty$.

- $$G(x) = \begin{cases} \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\xi}\right\}, & \xi \neq 0 \\ \exp\left\{-\exp\left[-\left(\frac{x-\mu}{\sigma}\right)\right]\right\}, & \xi = 0 \end{cases}$$

  where $x_+ = \max(0, x)$.

Threshold Excess Modelling

- But, using only the maximum is a waste of data.
- Using the same setting as above, for large enough u and conditional on $X_i > u$, $X_i - u \dot{\sim} \text{GPD}(u, \sigma, \xi)$ as $n \to \infty$.
- $$H(x) = \begin{cases} 1 - \left[1 + \xi\left(\frac{x-u}{\sigma}\right)\right]_+^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp\left[-\left(\frac{x-u}{\sigma}\right)\right], & \xi = 0 \end{cases}$$

Probability Density Functions

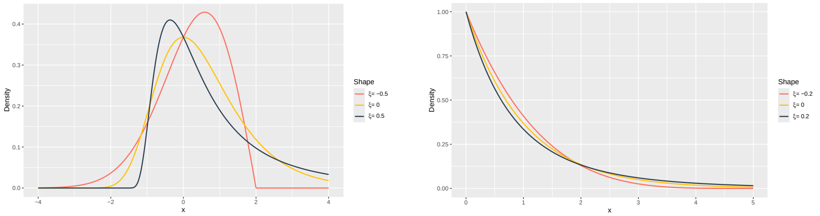

Figure 3: Probability Density Functions for GEV and GPD Distributions with $\mu = 0$, $u = 0$, $\sigma = 1$, and different values for ξ.

Measurement Errors

Assumptions

- The errors are i.i.d $\mathcal{N}(0, \sigma_N^2)$, where σ_N is known and do not depend on the magnitudes of earthquakes.
- The threshold u is pre-determined and fixed.
- We model the magnitudes of earthquakes as a GPD distribution. Assume $X \sim GPD(u, \sigma, \xi)$ where σ and ξ are unknown. Let $Y := X + \epsilon$, where ϵ are the errors.

Method 1: Estimation using Likelihood for $Y = X + \epsilon$

By convolution, the PDF for Y is given by:

$$f_Y(y \mid \sigma, \xi) = \int_u^{\infty} f_X(x \mid \sigma, \xi) \cdot \frac{1}{\sqrt{2\pi\sigma_N^2}} e^{\frac{-(y-x)^2}{2\sigma_N^2}} \mathrm{d}x \qquad (2)$$

and so, the log-likelihood is given by:

$$\ell(y_1, \ldots, y_n \mid \sigma, \xi) = \sum_{i=1}^{n} \log(f_Y(y_i \mid \sigma, \xi)) \qquad (3)$$

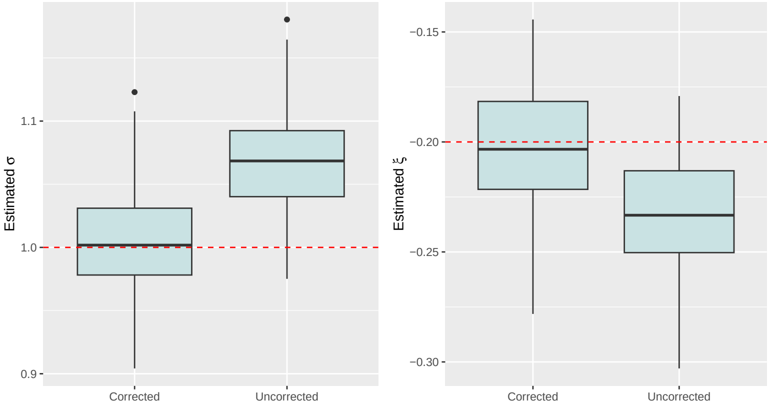

Figure 4: Likelihood for Y against likelihood for X.

Method 2: Monte Carlo-Based Estimation

Input: Initial parameters σ, ξ, noise standard deviation σ_N, observed data Y, threshold u, maximum iterations M, tolerance ϵ'
Output: Estimated parameters σ^* and ξ^*
Initialize: Set initial guesses for σ and ξ
for *iteration* $i = 1$ *to* M **do**
  E-Step:
  for *each observed* Y_i **do**
    Generate $X_{ij} \sim \text{GPD}(\sigma, \xi, u)$ for $j = 1$ to N samples
    Compute weights $w_{ij} = \frac{1}{\sqrt{2\pi\sigma_N^2}} \exp\left(-\frac{(Y_i - X_{ij})^2}{2\sigma_N^2}\right)$
    Calculate expected value $E[X_i \mid Y_i] = \frac{\sum_j X_{ij} w_{ij}}{\sum_j w_{ij}}$
  end
  M-Step:
  Define negative log-likelihood function using $E[X_i \mid Y_i]$
  Optimize parameters σ and ξ by minimizing the negative log-likelihood
  Convergence Check:
  if *change in parameters* $< \epsilon'$ **then**
    Break
  end
end
return Estimated parameters σ^* and ξ^*

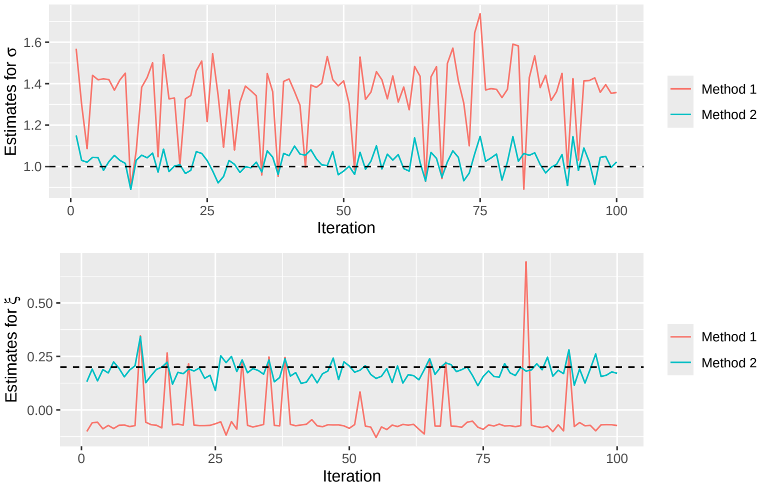

Figure 5: Estimates for σ and ξ by methods 1 and 2.

Further Research

Unknown σ_N and its covariates

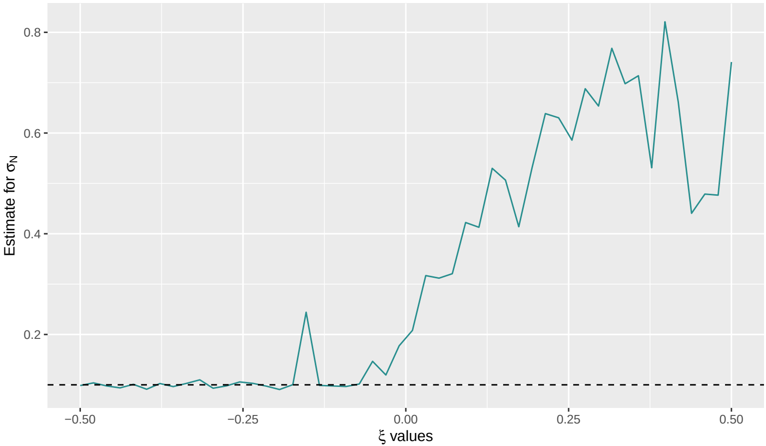

Figure 6: Estimates for σ_N with different values for ξ.

ETAS Model Parameters

- Inference on ETAS model parameters K_0, α, c, and p can be computationally intensive and costly, and better techniques using Bayesian analysis are being developed.
- A spatial parameter can be added in the ETAS model to take into account the space coordinates of earthquakes with different regions having different functions.
- Similar measurement errors might be present in time and space measurements, and one can research further to correct the errors.

References

[1] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*. Springer, 2001.

[2] Y. Ogata, "Space-time point process models for earthquake occurrences," *The Institute of Statistical Mathematics*, 1997.

[3] Y. Ogata, "Statistical models for earthquake occurrences and residual analysis for point processes," *Journal of the American Statistical Association*, 1988.

[4] Y. Chen, "Thinning algorithms for simulating point processes," *Florida State University, Tallahassee*, 2016.

Thank you for listening!

Thank you for listening!

Any questions?