Received xxxxxx; | Revised xxxxxx; | Accepted xxxxxx

DOI: xxx/xxxx

DISCUSSION PAPER

Discussion on Spatio-Temporal Modelling of Extreme Induced Seismicity in the Presence of An Evolving Measurement Network

Dáire Healy[1] | Ilaria Prosdocimi*[2]

[1]School of Mathematics and Statistics, University College Dublin, Dublin, Ireland
[2]Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Venice, Italy

Correspondence
*Ilaria Prosdocimi. Email: ilaria.prosdocimi@unive.it

Summary

Murphy et al. present a methodology to provide updated and improved estimates of the upper tail of induced earthquake magnitude distributions for the Groningen gas field, Netherlands. In particular, they propose regression models to characterise the evolving nature of both the magnitude of completion (which corresponds to the threshold in the classical peaks over threshold extreme value model) and the exceedance magnitude. These regression approaches are routinely used to analyse environmental extremes. Our discussion focuses on some aspects of their suitability and applicability to investigate changes in extremes. We consider the risk of confounding physical and observational non-stationarity, the difficulty of attributing changes in high quantiles to the threshold model or the exceedance model when both vary, and the role of threshold stability in estimating and interpreting the upper end-point of the Generalised Pareto distribution.

We congratulate the authors for providing a novel, interesting and timely article dealing with the estimation of the risks connected to earthquakes, with an application to the highly relevant case study of the Groningen gas field in the Netherlands. The proposed methodology builds upon a number of publications by some of the authors, demonstrating how several iterations are often needed to improve the statistical analysis of a dataset to provide a refined answer to a practical question of interest. The paper builds upon the classical peaks over threshold (POT) approach within the extreme value analysis context (Coles, 2001). One of the challenges in applying the POT approach for the specific case study is that the data collection effort has changed over time: while it would be possible to use a unique high threshold, appropriate for both older and more recent records, this would be wasteful and would not exploit the improvements in the data collection. The use of a varying threshold (called *magnitude of completion*) has already been proposed by some of the authors of the paper Varty et al. (2021); Murphy et al. (2025). In this additional work, the authors refine their efforts to estimate the varying threshold, assess the uncertainty of this estimation and propagate the uncertainty in the threshold estimation to the final risk assessment.

The authors' contribution is likely to be relevant not only for the analysis of earthquakes in the Groningen field, but for the quantification of risk linked to other natural hazards. Sampling variations and instrument limitations are ubiquitous in environmental observations. Evolving observational networks and instrument degradation represent structural features of many environmental datasets. The issue of having censored or incomplete historical records can be linked to changes in both the presence of working measuring stations and to the precision of the instrument in the station (see, for example, for precipitation Prakash et al. 2019; Su et al. 2026, and for temperature Easterling et al. 2000). The usefulness of including information on past events, even if measured imprecisely, has been widely recognised (see for example Kjeldsen et al., 2014, for a review in the area of flood frequency estimation), and ad-hoc approaches that account for the changes in recording effort have already been proposed (for example Coles and Sparks, 2006, in the context of volcanic eruptions). Extending the methodology to other environmental applications raises a fundamental challenge, namely, how to disentangle non-stationarity generated by the observation network from non-stationarity in the underlying physical process. Any trend detected in the extremal observations reflects composite (and potentially confounding) physical and observational mechanisms. Failure to distinguish between these distinct sources leads to misattributing the true source of non-stationarity. This is particularly troublesome in applications where causal interpretation is of primary interest, for example, in the detection of climate-change signals. Assuming that $(Y - u \mid Y > u) \sim \mathrm{GP}(\sigma_u, \xi)$, where Y is the process of interest, changes in properties of the distribution are typically modelled by means of regression. For example, consider the scale parameter of the Generalised Pareto (GP) distribution, modelled using environmental covariates. Ignoring network evolution forces the coefficients of the environmental covariates to absorb the underlying trend. Alternatively, if network development coincides spatially or temporally with environmental changes in the observed process, incorporating a network covariate directly into an additive scale parameter risks collinearity, compromising parameter identifiability in a fashion similar to spatial confounding (Dupont et al., 2025). We note that the authors' proposed model alleviates this identifiability concern in practice. The network covariate, $V_i(\boldsymbol{x}, t)$, enters only the threshold function, $u(\boldsymbol{x}, t)$, while the physical stress covariate, $s(\boldsymbol{x}, t)$, enters only the baseline scale parameter, $\sigma_0(\boldsymbol{x}, t)$. The scale parameter $\sigma_u(\boldsymbol{x}, t)$ is then a function of both covariates through a structural combination, $\sigma_u(\boldsymbol{x}, t) = \sigma_0(\boldsymbol{x}, t) + \xi u(\boldsymbol{x}, t)$, rather than a shared regression. This should therefore not be read as a critique of the present work, but as a caution for applying similar methodology in settings with less separable covariates. There is a suite of available techniques to mitigate statistical issues of collinearity where they may arise in other applications. For example, orthogonalising the network development covariate by regressing it against physical drivers and retaining the residuals (Hodges and Reich, 2010). Determining if a covariate-dependent GP distribution scale parameter σ_u can isolate environmental non-stationarity from sampling variations, and, if the proposed structural separation of network and environmetnal covariates suffices to alleviate issues of collinearity, certainly warrants further investigation.

The authors assess changes in extremes by using regression models to represent the evolving nature of both the threshold and exceedance distribution. A potential drawback of this approach is that the estimated high quantiles of interest (the so-called *effective design levels* in this case) vary as functions of the predictors. This variation occurs regardless of whether the threshold and exceedance distribution models share identical covariates. When this occurs, it can be difficult to interpret how the different varying components of the model affect the final estimate, with the threshold regression model potentially compensating for, or amplifying, the effects of the GP regression model (Northrop and Jonathan, 2011). The threshold model is often treated as a preprocessing step, with its uncertainty typically ignored and not propagated to the final estimate (Scarrott and MacDonald, 2012). We commend the authors for developing an approach which clearly allows this propagation of uncertainty. Nevertheless, the issue of separating out the impact of the different regression models on the final estimates for risk measures remains something we feel could be further explored in the literature.

Classical threshold stability requires the scale parameter to satisfy the linear relation $\sigma_v = \sigma_u + \xi(v - u)$ when evaluating exceedances above a higher threshold $v > u$, with the shape parameter, ξ, being threshold-invariant. This condition guarantees that derived physical limits remain invariant to the choice of threshold, provided it is sufficiently high. For a GP distribution, we define the upper endpoint as $\zeta = u - \sigma_u/\xi$. When the GP model is evaluated for a higher threshold value, v, and the threshold stability property holds, we find

$$\zeta_v = v - \frac{\sigma_v}{\xi} = v - \frac{\sigma_u + (v-u)\xi}{\xi} = \frac{\xi v - \sigma_u - \xi v + \xi u}{\xi} = u - \frac{\sigma_u}{\xi} = \zeta_u$$

Effectively, the choice of a sufficiently high threshold does not affect the upper endpoint of the GP distribution, ζ. Notice also for $\xi < 1$, we have that $\mathbb{E}[Y] = \mu = \sigma_u/(1-\xi)$, and for $\xi < 1/2$, we have $\mathrm{Var}[Y] = \sigma_u^2/[(1-\xi)^2(1-2\xi)]$. If we have threshold stability for σ_u, we also obtain threshold stability for the mean and standard deviation of the distribution. In this sense, the mean-max parametrization discussed in Tendijck et al. (2024) retains some interesting threshold-stability properties. In assessing the usefulness of other parameterisations which are proposed in the literature (Huet and Prosdocimi, 2026), threshold stability should also play a role. The authors' model implicitly preserves threshold stability for the mean exceedance magnitude and its variance across space and time, ensuring estimated summary statistics for seismic activity remain mathematically consistent regardless of changes in geophone completeness thresholds.

As mentioned by the authors, in practice, regression models for GP parameters frequently fail to retain threshold stability (Eastoe and Tawn, 2009). Standard extreme value modeling software routinely applies a non-linear log-link function to covariates in the scale parameter (Belzile et al., 2023) to enforce positivity and avoid optimization issues. This non-linear formulation violates threshold stability, causing physical quantities (such as the upper endpoint, ζ) to vary depending on the chosen threshold u or v, and become mathematically inconsistent. This mathematical inconsistency compounds when thresholds u and v are specified as functions rather than constants. We commend the authors for making use of an identity link for the scale parameter to preserve threshold stability, and we hope to see the issue of threshold stability discussed more transparently in future POT non-stationary analysis.

Estimating finite upper bounds in extreme value distributions introduces distinct inferential challenges. Namely, when the shape parameter is estimated to be negative, the limit of the return level asymptotes to the upper endpoint of the distribution. For long return periods (such as the one reported in the study), the estimated return level is very close to the estimated upper endpoint, $\hat{\zeta}$. Interpreting the physical meaning of the upper endpoint of the distribution is very challenging when the data is limited. Unconstrained maximum likelihood estimation risks producing unreasonably wide uncertainty intervals. Imposing hard constraints (i.e., ensuring $-c < \xi < 0$, with c a suitable small value specified by the analyst) fails to alleviate parameter instability (Coles and Dixon, 1999). Building upon these structural considerations and the earlier work of Yue et al. (2025), a potential avenue of future work to address both parameter instability and the physical interpretability of the upper endpoint is to parameterise ζ directly within the model structure. This could be achieved by expressing the scale parameter as a function of the physical limit, $\sigma_u = (u - \zeta)\xi$ or equivalently, $\xi = -\sigma_u/(\zeta - u)$. This reparameterization provides a natural foundation for a Bayesian framework, affording practitioners a principled approach to provide and incorporate prior information for an upper limit derived from geophysical domain knowledge and help resolve parameter instability (Coles and Tawn, 1996).

For a number of natural hazards, it is of key importance to accurately model changes in extremes in continuously evolving contexts. Peaks over threshold approaches, tailored to the specific application, are commonly used for this. The work offers an exemplar for propagating threshold uncertainty while preserving classical threshold stability, within a complex data setting. The modeling approach in the paper is specifically tailored to the context of the data. We congratulate the authors on their effort to provide statistically sound estimates to enhance risk management in the area of study. The contribution has promising applications for broader analyses of observed environmental extremes, and we hope it will inspire more researchers to go beyond out-of-the-box approaches to fully exploit the information available in observational data.

References

L. R. Belzile, C. Dutang, P. J. Northrop, and T. Opitz. A modeler's guide to extreme value software. *Extremes*, 26(4):595–638, 2023.

S. G. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, London, 2001.

S. G. Coles and M. J. Dixon. Likelihood-based inference for extreme value models. *Extremes*, 2(1):5–23, 1999.

S. G. Coles and R. S. J. Sparks. Extreme value methods for modelling historical series of large volcanic magnitudes. In *Statistics in Vulcanogy, edited by H. M. Mader, S. G. Coles, C. B. Connor, L. J. Connor*. Geological Society of London, 2006.

S. G. Coles and J. A. Tawn. A Bayesian analysis of extreme rainfall data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 45(4):463–478, 1996.

E. Dupont, I. Marques, and T. Kneib. Demystifying spatial confounding. *arXiv preprint*, arXiv:2309.16861, 2025.

D. R. Easterling, J. L. Evans, P. Y. Groisman, T. R. Karl, K. E. Kunkel, and P. Ambenje. Observed variability and trends in extreme climate events: a brief review. *Bulletin of the American Meteorological Society*, 81(3):417–426, 2000.

E. F. Eastoe and J. A. Tawn. Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 58(1):25–45, 2009.

J. S. Hodges and B. J. Reich. Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician*, 64(4):325–334, 2010.

N. Huet and I. Prosdocimi. Orthogonal parametrisations of extreme-value distributions. *Statistics & Probability Letters*, 237: 110830, 2026.

T. Kjeldsen, N. Macdonald, M. Lang, L. Mediero, T. Albuquerque, E. Bogdanowicz, R. Brázdil, A. Castellarin, V. David, A. Fleig, G. Gül, J. Kriauciuniene, S. Kohnová, B. Merz, O. Nicholson, L. Roald, J. Salinas, D. Sarauskiene, M. Šraj, W. Strupczewski, J. Szolgay, A. Toumazis, W. Vanneuville, N. Veijalainen, and D. Wilson. Documentary evidence of past floods in Europe and their utility in flood frequency estimation. *Journal of Hydrology*, 517:963–973, 2014.

C. Murphy, J. A. Tawn, and Z. Varty. Automated threshold selection and associated inference uncertainty for univariate extremes. *Technometrics*, 67(2):215–224, 2025.

P. J. Northrop and P. Jonathan. Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics*, 22(7):799–809, 2011.

S. Prakash, A. Seshadri, J. Srinivasan, and D. Pai. A new parameter to assess impact of rain gauge density on uncertainty in the estimate of monthly rainfall over India. *Journal of Hydrometeorology*, 20(5):821–832, 2019.

C. Scarrott and A. MacDonald. A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT-Statistical Journal*, 10(1):33–60, 2012.

J. Su, C. Miao, F. Zwiers, H. Beck, P. Jones, Q. Sun, L. J. Slater, W. R. Berghuijs, Y. Wada, D. Rosenfeld, J. Gou, Y. Wu, P. Tarolli, P. Borrelli, P. Panagos, L. V. Alexander, Q. Zhang, J. Hu, S.-K. Min, L. Samaniego, Q. Duan, G. Destouni, J. A. Marengo, R. Modarres, and S. Sorooshian. Precipitation observing network gaps limit climate change impact assessment. *Nature*, 652:119–125, 2026.

S. Tendijck, D. Randell, G. Feld, and P. Jonathan. Practical non-stationary extreme value analysis of peaks over threshold using the generalised Pareto distribution: Estimating uncertainties in return values. *Ocean Engineering*, 312(3):119247, 2024.

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

W. Yue, J. Tawn, R. Towe, and Z. Varty. Integrating experts' belief in upper tail inference for modelling of human-induced earthquake magnitudes. *Journal of Agricultural, Biological and Environmental Statistics*, pages 1–25, 2025.