

Multivariate Peaks-Over-Threshold (POT) Modelling of Non-stationary Air Pollution Concentration Data

[Dr Leonid Bogachev](#) (SoM), [Dr Haibo Chen](#) (ITS), [Dr Georgios Aivaliotis](#) (SoM), [Prof. Jeanne Houwing-Duistermaat](#) (SoM)

Project partner(s): Leeds City Council; Environment Agency

Contact email: L.V.Bogachev@leeds.ac.uk

Peaks-over-threshold (POT) method is the preferred modern approach to analyse extreme values in a time series. This is due to a better usage of information as compared to the classic block-maxima method (which utilises only one maximum value in each block, e.g. year). Moreover, in many applications the impact of extremes is often implemented through a few moderately large values rather than due to a single highest maximum.

Threshold exceedances approximately follow a generalised Pareto distribution (GPD) with two parameters (scale, shape), which are constant if the data are stationary (i.e. the observed process is in statistical equilibrium). However, in many practical situations including the air pollution, parameters of the system are likely to significantly change with time. Following Davison & Smith (1990), threshold exceedances in non-stationary data should be modelled by treating the GPD parameters as functions of (time-dependent) covariates (e.g. weather and traffic conditions for air pollutants). However, the Davison-Smith regression model is not threshold stable, which means that the model parameters have to be re-estimated with every new threshold (which may need to vary with time). Recently, Gyarmati-Szabo, Bogachev and Chen (2017) proposed a novel model for non-stationary POT which is threshold stable. This has a strong potential to improve dramatically the computational efficiency of the POT model, making it into a versatile and powerful tool for dynamic estimation and prediction of extremes. In particular, this approach may serve as the basis for a semi- or fully automated computational tool designed for efficient on-line estimation and accurate prediction of future extreme events. Due to the property of threshold stability, such methods will work efficiently with variable threshold selection.

Objectives:

The present project will aim to develop a more general methodology of joint modelling of several observables such as different air pollutants, e.g. NO₂, NO, O₃ etc., which are highly correlated due to complex photochemical reactions in the atmosphere in the presence of sunlight. The principal innovation to be achieved is to design a suitable multivariate POT model for non-stationary data that will preserve the property of threshold stability. Data analysis based on such a model will involve the MCMC (Markov Chain Monte Carlo) simulations to obtain posterior distributions of the model parameters. The project is also likely to involve the development of an efficient computer simulator of the (class of) generalised Pareto distributions.

Potential for high impact outcome

Improving air quality is one of the key objectives of the NERC. The project is likely to deliver significant results which may be instrumental for dynamic estimation and prediction of future extreme events in air pollution.

Training

This project will be supervised jointly by the Department of Statistics and the Institute for Transport Studies at Leeds. Also, it has a strong potential to involve collaboration with external organisations, such as the Leeds City Council (LCC) and the Environment Agency. Full training in the related disciplines and skills will be provided through taught courses and hands-on tuition.

Student profile:

The successful PhD candidate should have a solid background in mathematics and statistics, with a strong interest in and a flair for statistical modelling of extreme values. Appreciation of the complexity of modelling air pollution concentrations would be an advantage, as well as a sound grounding in multivariate statistical analysis and Bayesian statistics. Key skills required for the project include competent use of R and experience with programming and statistical computing in general, including MCMC simulations.

CASE Partner

Negotiations with the Leeds City Council (LCC) and the Environment Agency are in progress about their possible involvement as a CASE partner.

References

- Davison, A.C. and Smith, R.L. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society, Ser. B* **52** (1990), 393–442, <http://www.jstor.org/stable/2345667>
- Gyarmati-Szabo, J., Bogachev, L.V. and Chen, H. Nonstationary POT modelling of air pollution concentrations: Statistical analysis of the traffic and meteorological impact. *Environmetrics* **28** (2017), no. 5, Paper e2449, 15 pp, doi:10.1002/env.2449.