Robust extreme value analysis by semiparametric modelling of the entire distribution range

Frank Kwasniok
Department of Mathematics, University of Exeter, Exeter, United Kingdom (f.kwasniok@exeter.ac.uk)

Traditional extreme value analysis based on the generalised extreme value (GEV) or the generalised Pareto distribution (GPD) suffers from two drawbacks: (i) Both methods are wasteful of data as only block maxima or exceedances over a high threshold are used and the bulk of the data is disregarded, resulting in a large uncertainty in the tail inference. (ii) In the peak-over-threshold approach the choice of the threshold is often difficult in practice as there are no really objective underlying criteria.

Here, two approaches based on maximum likelihood estimation are introduced which simultaneously model the whole distribution range and thus constrain the tail inference by information from the bulk data. Firstly, the bulk matching method models the bulk of the distribution with a flexible exponential family model and the tail with a GPD. The two distributions are linked together at the threshold with appropriate matching conditions. The threshold can be estimated in an outer loop also based on the likelihood function. Secondly, in the extended generalised Pareto distribution (EGPD) model for non-negative variables the whole distribution is modelled with a GPD overlaid with a transition probability density which is again represented by an exponential family. Appropriate conditions ensure that the model is in accordance with extreme value theory both for the lower and upper tail of the distribution. The methods are successfully exemplified on simulated data as well as wind speed and precipitation data.