Multivariate Declustering Techniques

by

S. Nadarajah

University of Plymouth*

Keywords: Declustering techniques, Exceedances, Multivariate extreme value models

One important assumption underlying the currently known methods for fitting multivariate extreme events (Nadarajah, 1998) is that data from different events are independent. For this we need techniques that identify independent events in the data record and filter out extreme values from them. We refer to these techniques as declustering techniques. In this paper we develop a new declustering technique for multivariate data that has wider applicability than any of the existing ones.

In modelling of univariate extremes, techniques for declustering a series of observations to obtain independent values are well established (National Environment Research Council, 1975). Essentially three approaches have been proposed. Smith (1989) identifies events as clusters of high–level exceedances and filters out the cluster peaks. Specifically a *model threshold* and a *separation interval* are chosen such that if two exceedances over the model threshold are closer together than the separation interval then they are deemed part of the same cluster, but if the time interval between successive exceedances is longer than the separation interval it is considered that the old cluster has finished and a new one begun. Tawn (1988) presents an alternative approach to extract the r largest 'independent' annual events. This approach is based on the assumption that each independent event has a standard length, and only the maximum value within each event is filtered out. Both these approaches are *ad hoc* and avoid specification of a precise model. Smith (1984) and Davison and Smith (1990) propose parametric models for the full point process of exceedance times over a model threshold based on the doubly stochastic process. The models are fitted to the process of observed inter exceedance times and an estimate of inter exceedance time between independent exceedances is obtained. The identification of events and filtering out of extreme values from them is performed as in Smith (1989) with the separation interval determined by the parametric estimate.

The notion of multivariate declustering introduces new complications. Consider the simple situation where events in each marginal process occur simultaneously. Univariate declustering applied to the marginal processes separately above model thresholds t_i, say, will filter out only the events that are extreme in each margin. However the known multivariate extreme value models require the number of independent events in total. Hence even in this simple situation univariate declustering is not adequate. A solution is to apply univariate declustering to the marginal processes at levels far lower than t_i, which we term *declustering* thresholds, and to concatenate independent extremes in each margin with the concurrent values in the other margins to produce independent events.

Coles and Tawn (1991) consider equivalent situations where there is a time lag between events in different marginal processes. They achieve multivariate declustering by assuming that each event has a standard length. To protect against lack of knowledge of time lag and to ensure marginal independence the event length is chosen large enough, and independent events are filtered out by concatenating the maxima of the processes within each event length. For illustration of this approach consider Figure 1 where rainfall and wind speed series for some 350 hours are plotted. Also shown in the figure are some model thresholds for the two margins. An event length of 100 hours produces the three independent events marked by different symbols. The weakness of the declustering approach in this illustration is that the rainfall component of the first event is identified with the same cluster of exceedances as one of other events.

The phenomenon that events in different processes occur simultaneously or with a time lag between them does not often occur in environmental processes. Hence there is a real need for a multivariate declustering

**Address for Correspondence:* School of Mathematics and Statistics, University of Plymouth, Drake Circus, Plymouth PL4 8AA, United Kingdom; E-mail: s.nadarajah@plymouth.ac.uk

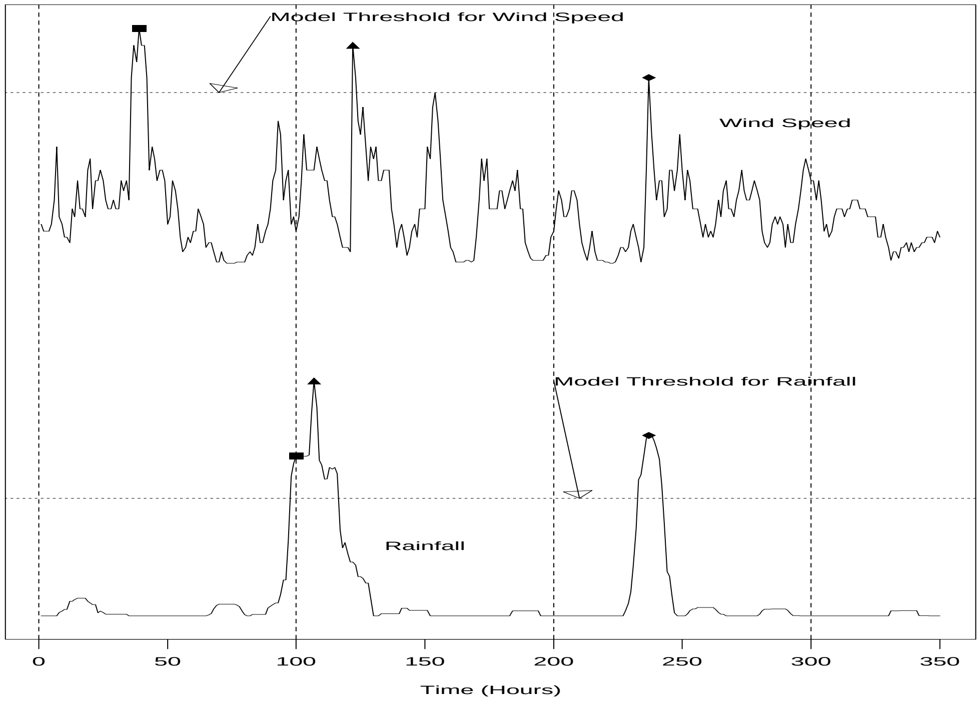

Figure 1: Coles and Tawn (1991)'s Approach

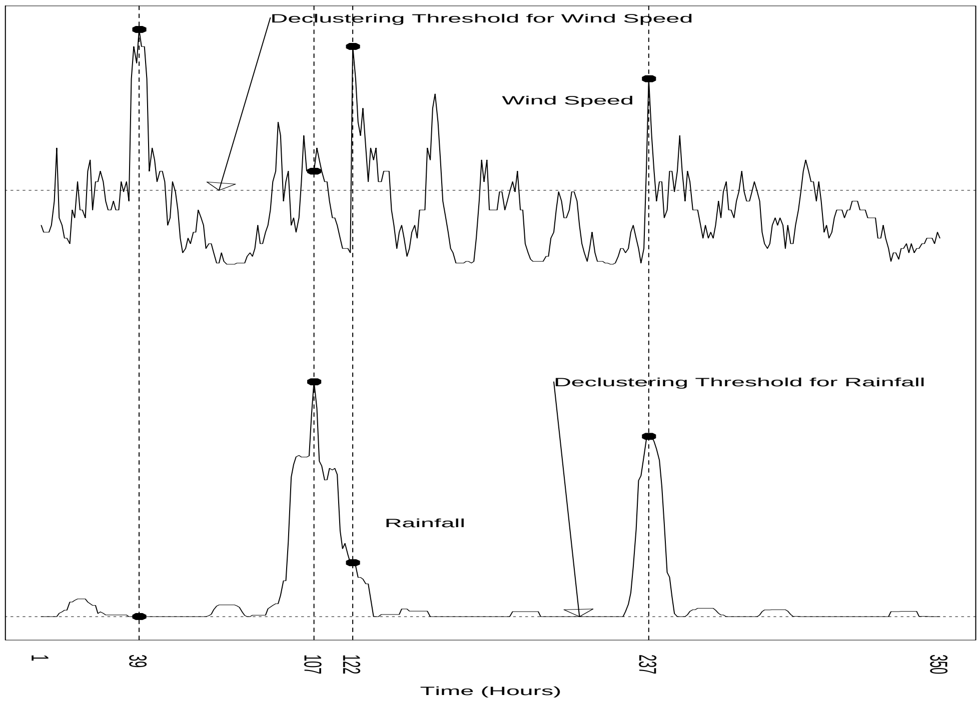

Figure 2: Our Approach

technique that has wider applicability. The technique we propose in this paper meets that need.

Consider the case where events in each marginal process are identified as clusters of exceedances over a declustering threshold. Clusters of exceedances in each marginal process may be separated by a large enough separation interval, but there may be clusters in different margins that are so close in time that any multivariate events filtered out from them cannot plausibly be assumed independent. For illustration consider Figure 2 where exceedances of the rainfall and wind speed series over some declustering thresholds are plotted. An application of the univariate declustering approach in Smith (1989) with a separation interval of 24 hours filters out from the rainfall series the two cluster peaks occurring at times 107 and 237 hours. Similar application for the wind speed series filters out the three cluster peaks occurring at times 39, 122 and 237 hours. The vertical lines represent bivariate events formed by concatenating the cluster peaks with concurrent values in the other margin. The events at 107 and 122 hours cannot plausibly be assumed independent since their rainfall components belong to the same cluster of exceedances. However, generalising the idea in Smith (1989), we can choose a separation interval to ensure that any two such events closer together in time than the separation interval are deemed not independent. The declustering technique of the paper is based on this idea.

In addition to proposing the new technique, the paper shows how the technique could be extended to apply to a range of specialised (multivariate) data structures of the following type:

- Data with different levels of temporal dependence: Consider for example a bivariate process in which the second margin has so high a degree of temporal dependence that the number of clusters of exceedances there is far lower than in the first margin. In whatever way one declusters the process into independent bivariate events, many of the apparently independent univariate events in the first marginal process will have to be discarded because they are associated with dependent values in the other margin. The paper develops a method that circumvents this difficulty.
- Aggregated and lagged data: It is known in hydrology that extreme values of water level at a reservoir's dam are generated by storms of a particular duration, d, where d may differ from reservoir to reservoir. Total rainfall during a storm is therefore the aggregate of hourly rainfall totals over d hours. There is some delay between the fall of rain over a reservoir's catchment and its peak effect on the water level, and this lag, l, again is a characteristic of the reservoir and catchment under study. Extreme values of water level are therefore generated by d–hour aggregated values of rain, and the values of other variables such as wind at a time l hours later. The declustering procedure must therefore filter out values of variables linked by an appropriate time l (and in the case of total rainfall, aggregated over d hours) from episodes when these variables are extreme; and it must ensure that the assumption of independence of the resulting vectors is credible. The paper develops a method to do this.
- Seasonal data: Modelling of seasonal and other nonstationary effects in a dataset entails filtering out of independent events from different time periods of the data (e.g. months or years). An obvious procedure for this is to decluster the data for each time period separately. This can however be computationally time–consuming and since there can be clusters of exceedances which overlap two neighbouring time periods we may end up filtering out more than one event from the same cluster, leading to events in different time periods which are not plausibly independent. The paper provides a solution to this.

References

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of Royal Statistical Society: Series B*, **53**, 377-392.

Davison, A. C. and Smith, R. L. (1990). Model for exceedances over high thresholds. *Journal of Royal Statistical Society: Series B*, **52**, 393-442.

Nadarajah, S. (1998). Simulation and modelling of extreme values. To appear in *Handbook of Statistics*, editors C. R. Rao and D. N. Shanbhag.

Natural Environment Research Council (1975). Floods studies report (5 volumes). NERC, London.

Smith, R. L. (1984). Threshold methods for sample extremes. In *Statistical Extremes and Applications*, 621–638, editor J. Tiago de Oliveira, Reidel, Dordrecht.

Smith, R. L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground level ozone. *Statistical Science*, **4**, 367-393.

Tawn, J. A. (1988). An extreme value theory model for dependent observations. *Journal of Hydrology*, **101**, 227–250.