Discussion of the papers by Dankers and Feyen, Cooley, and Keef

A. C. Davison
Ecole Polytechnique Fédérale de Lausanne
EPFL-FSB-MATHAA-STAT, Station 8
1015 Lausanne, Switzerland
E-mail: Anthony.Davison@epfl.ch

The classical paradigm

These very interesting papers give an excellent springboard for the discussion of some issues in contemporary extremal analysis, and I am grateful to the organisers for the chance to contribute this.

It seems worthwhile first to make some general points about the statistics of extremes, which differs from much of mainstream statistics in the sense that its primary goal is extrapolation into the tail of a distribution, typically based on limited relevant data. Any extrapolation is fraught with risks, and hence one seeks a principled basis for it. The classical paradigm of extremal statistics is based on showing that any non-degenerate limiting distribution H for the maximum of a random sample must be max-stable, i.e., for any $m \in \mathbb{N}$ there exist real numbers $a_m > 0$ and b_m such that

$$H^m(b_m + a_m x) = H(x), \quad x \in \mathbb{R}. \tag{1}$$

The only distribution that satisfies (1) is the generalized extreme-value distribution (GEV)

$$H(y) = \begin{cases} \exp\left[-\left\{1 + \xi(y-\eta)/\tau\right\}_+^{-1/\xi}\right], & \xi \neq 0, \\ \exp\left[-\exp\left\{-(y-\eta)/\tau\right\}\right], & \xi = 0, \end{cases} \tag{2}$$

where $a_+ = \max(a, 0)$ and η and τ are respectively a real location parameter and a positive scale parameter. The shape parameter ξ determines the weight of the upper tail of the density, with $\xi < 0$ corresponding to the reverse Weibull case in which the support of the density has a finite upper bound, $\xi = 0$ corresponding to the light-tailed Gumbel distribution, and $\xi > 0$ corresponding to the heavy-tailed Fréchet distribution.

As the only possible solution to (1), expression (2) has strong mathematical support as a suitable distribution for the fitting of maxima for scalar random variables. Similar arguments apply to exceedances over high thresholds, for which there is an analogous notion of threshold-stability, leading to the use of the generalized Pareto distribution (GPD) in that context, again as the only possible solution to a functional equation. In multivariate and infinite-dimensional settings (1) is interpreted in terms of vector or functional x, leading to analogous semiparametric limiting models.

In all cases there is *large uncertainty* surrounding the extrapolations, and it seems to me important to say so, however unpalatable this may be in applications. Even in the scalar case the parameter ξ is generally difficult to estimate but has a strong impact on estimated quantiles. For this reason I was dismayed to see that Dankers and Feyen restricted their fitting for floods to the Gumbel distribution; this would greatly reduce the uncertainty of quantile estimates, probably to an unrealistic degree. As they make clear, their graphs don't have an uncertainty assessment in the statistical sense, but rather compare point estimates from 2^3 possible scenarios, so the effects of this restriction are not directly apparent, but it seems important to know whether suitable confidence sets for the estimates in their graphs contain the possibility of a 0% change, for example.

The large uncertainty suggests the use of strategies to *pool information*, through inclusion of covariates, through the use of hierarchical models (Cooley) or through multivariate modelling (Keef). This seems essential in situations where modelling of events is needed, but is also helpful even in cases where the goal is to model features of marginal distributions (Dankers and Feyen).

There is an important caveat: in order for solutions to (1) to be relevant to applications, we must *assume that a limiting distribution exists*, and this demands careful consideration in any given context. Are the elegant mathematical arguments leading to (2) and related models relevant to the real world? Even when (2) is thought to be suitable, one might wonder whether sub-asymptotic distributions, so-called penultimate approximations, may be more appropriate for use with finite samples. In practice, the GEV and GPD are often found to provide adequate fits to data, both because they are rather flexible and because the sample is often too small to provide much power to detect lack of fit.

Another generic concern is *regime change in the distribution tails*. Statistical extrapolation will typically be based on distributions fitted to observed data, but will be useful only if past data are relevant to the future. For example, data on extreme temperatures observed in conditions where there is reasonable soil moisture, which will absorb energy through evapotranspiration, may severely underestimate temperatures in a future climate once that moisture has all evaporated away, for example during a (now) exceptionally long heatwave. For this reason one has to hope that careful long-range analyses such as that of Dankers and Feyen are based on 'data' that have captured the relevant physical processes accurately enough to be useful for tail extrapolation. There is a literature on so-called 'dragon kings' also known as 'black swans', viz., monstrously extreme observations that seem to arise from processes different from those leading to 'ordinary' extremes and which perhaps could not have been anticipated from analysis of available data.

In view of worries about its relevance, it seems wise to view the classical paradigm as a *guide to sensible practice*, rather than as a series of recipes to be followed slavishly. For example, as Keef points out, the GEV, which can be generated as the distribution of a maximum of a Poisson number of independent GPD variates, can be broadened to a four-parameter model by using a negative binomial rather than a Poisson distribution. Likewise Cooley uses a hierarchical mixture of max-stable models in the spatial context. Not surprisingly, in both cases there is a better fit. But this raises the question: *how far can we go from the classical paradigm before jettisoning it altogether, and how can we decide when to do so?*

Some more specific points

The three papers under discussion raise several further issues for extremal analysis and its application, and I'd like to highlight some of these.

One perennial issue is *data quality* and its impact on extremal modelling. Statistics of extremes by definition concerns observations that might be downweighted or even rejected as aberrant when focusing on measures of central tendency, and these values have a strong impact on the subsequent inferences. Systematic downweighting of them can lead to a radical loss of information on tail parameters (Davison and Smith, 1990), yet in many cases they may be subject to measurement error or may arise from a heterogeneous population. In either case blind use of standard fitting methods may be misleading. Little research seems to have been performed on measurement error in extremes, though the fitting of mixtures of generalised extreme-value or generalised Pareto distributions, when there are enough data, is becoming more common (e.g., Süveges and Davison, 2012); of course Bayesian modelling will typically use infinite mixture distributions obtained by averaging out prior uncertainty. Either of these, or the use of robust approaches (Dupuis and Field, 1998; Dupuis and Morgenthaler, 2002; Dupuis, 2005), seems more satisfactory than the ad hoc approach adopted by Keef to deal with the river flow outliers.

Statistics of extremes is increasingly applied to, or used in conjunction with, output from sophisticated *physical models*. Examples are the hydraulic and hydrological models applied by Keef and by Dankers and Feyen, and the regional climate models (RCMs) used by Cooley and by Dankers and Feyen. However the agreement between climate models, with their myriad parameters, and observed data seems to be less close than one might naively suppose. As Dankers and Feyen mention, there seems to be a consensus that extremes are not yet represented well by RCMs, and clearly their analysis greatly benefits from calibration based on observed extreme river discharges during the years 1960–1991, which provides an essential reality check. An associated issue is that of up- and downscaling: RCM output typically smears rainfall uniformly across grid cells, and how this should be compared with observations gathered at individual sites seems to be open, at least with regard to extremes. Radar data are also sometimes available and their integration poses additional challenges. Another issue in physical modelling is raised by the increasing use of fast cheap statistical emulators of slow expensive numerical models, often based on Gaussian process approximations. There has been a good deal of work on emulation models for the average output, but almost none for extreme output; presumably suitable emulators would be based on max-stable processes.

Another generic issue is the *inappropriate application* of standard extreme-value analysis; one suspects that this is widespread in the financial industry. The paper of Dankers and Feyen seems to me to illustrate this in at least two ways: first, they appear to fit the GEV to non-stationary annual maxima, treated as though they were independent and identically distributed, and this is bound to lead to concerns about the usefulness of the quantile estimates from the fitted models. Moreover they appear to analyse the individual series of 30 years of annual maxima, on a 5×5 km grid derived from a 50×50 km grid, as though these were independent series, when clearly strong dependencies, both spatial and temporal and with nested levels of variation, might be expected. So far as I can understand, their analysis makes no attempt either to model these dependences or to make any allowance for them, but they are likely to weaken any conclusions that might be drawn from their analysis. A more appropriate approach could be to apply the tools used in the other two papers, in order to pool information from their very short individual series of 30 annual maxima, and thereby to allow some discussion of uncertainty. This comment may sound critical of Dankers and Feyen, but it should rather be taken as criticism of the statistical community, which has only recently begun serious development of models for spatial extremes, despite an obvious lacuna in the extreme-value literature. A spatial analysis of extremes was unavailable when their main papers were written, and would be close to or at the frontier of statistical methods development even now (Padoan *et al.*, 2010; Blanchet and Davison, 2011; Davison and Gholamrezaee, 2012; Davison *et al.*, 2012; Ribatet *et al.*, 2012).

The paper of Keef raises another generic theme of modern extremal analysis, namely when and whether max-stable models are appropriate. Many phenomena show an attenuation of dependence as events become rarer, and this is inconsistent with max-stability. The diagnosis of this so-called *near-independence* and its modelling have been important themes of research over the past 15 years (Ledford and Tawn, 1996, 1997, 2003; Heffernan and Tawn, 2004; Ramos and Ledford, 2009), but much remains to be done in order to achieve a unified paradigm for these cases analogous to the classical one. Although the Heffernan–Tawn approach adopted by Keef does seem to be well-adapted to her particular problem, it is far from obvious how it would be applied in the other two papers, and we await its extension to truly spatial settings.

A final comment concerns the inclusion of time in extremal analysis. Clearly time plays a crucial role both in flooding and in rainfall modelling, and it seems important that the type of ad hoc approaches adopted in these papers, *faut de mieux*, be replaced by more appropriate *space-time models*, perhaps analogous to those of Cox and Isham (1988) and Rodríguez-Iturbe *et al.* (1987, 1988). Some progress on this is being made by R. Huser at EPFL, based on space-time max-stable models involving random sets corresponding to storm cells, but heavy computational demands mean that in realistic cases it might be appropriate to apply the ideas suggested by Cooley.

Acknowledgements

The work was supported by the Swiss National Science Foundation.

References

Blanchet, J. and Davison, A. C. (2011) Spatial modelling of extreme snow depth. *Annals of Applied Statistics* **5**, 1699–1725.

Cox, D. R. and Isham, V. S. (1988) A simple spatial-temporal model of rainfall. *Proceedings of the Royal Society of London, series A* **415**, 317–328.

Davison, A. C. and Gholamrezaee, M. M. (2012) Geostatistics of extremes. *Proceedings of the Royal Society of London, series A* **468**, 581–608.

Davison, A. C., Padoan, S. and Ribatet, M. (2012) Statistical modelling of spatial extremes (with Discussion). *Statistical Science* **to appear**.

Davison, A. C. and Smith, R. L. (1990) Models for exceedances over high thresholds (with Discussion). *Journal of the Royal Statistical Society series B* **52**, 393–442.

Dupuis, D. J. (2005) Ozone concentrations: A robust analysis of multivariate extremes. *Technometrics* **47**, 191–201.

Dupuis, D. J. and Field, C. A. (1998) Robust estimation of extremes. *Canadian Journal of Statistics* **26**, 199–215.

Dupuis, D. J. and Morgenthaler, S. (2002) Robust weighted likelihood estimators with an application to bivariate extreme value problems. *Canadian Journal of Statistics* **30**, 17–36.

Heffernan, J. E. and Tawn, J. A. (2004) A conditional approach for multivariate extreme values (with Discussion). *Journal of the Royal Statistical Society series B* **66**, 497–546.

Ledford, A. W. and Tawn, J. A. (1996) Statistics for near independence in multivariate extreme values. *Biometrika* **83**, 169–187.

Ledford, A. W. and Tawn, J. A. (1997) Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society series B* **59**, 475–499.

Ledford, A. W. and Tawn, J. A. (2003) Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society series B* **65**, 521–543.

Padoan, S., Ribatet, M. and Sisson, S. (2010) Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association* **105**, 263–277.

Ramos, A. and Ledford, A. W. (2009) A new class of models for bivariate joint extremes. *Journal of the Royal Statistical Society series B* **71**, 219–241.

Ribatet, M., Cooley, D. and Davison, A. C. (2012) Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica* **to appear**.

Rodríguez-Iturbe, I., Cox, D. R. and Isham, V. S. (1987) Some models for rainfall based on stochastic point processes. *Proceedings of the Royal Society of London, series A* **410**, 269–288.

Rodríguez-Iturbe, I., Cox, D. R. and Isham, V. S. (1988) A point process model for rainfall: further developments. *Proceedings of the Royal Society of London, series A* **417**, 283–298.

Süveges, M. and Davison, A. C. (2012) A case study of a “Dragon King”: The 1999 Venezuelan catastrophe. *European Physical Journal, series E* **to appear**.