Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI:10.1002/env.1125

V. Chavez-Demoulin[a], A. C. Davison[b]* and L. Frossard[b]

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

a *Faculty of Business and Economics, University of Lausanne, 1015 Lausanne, Switzerland*

b *Ecole Polytechnique Fédérale de Lausanne, EPFL-FSB-MATHAA-STAT, Station 8, 1015 Lausanne, Switzerland*

1. Introduction

We congratulate the authors on their very stimulating paper. There is much in it worthy of comment, but we shall focus on two issues that we wish to highlight, parametrization and efficiency, and on an alternative simpler approach to high quantile estimation which we think has advantages over that proposed in the paper.

2. Parametrization

Although the block maximum and peaks over threshold approaches to modelling extremes are mathematically equivalent, their statistical merits depend on the use to which they are put. Threshold stability is a key property of the generalized Pareto distribution (GPD): when the threshold increases from u to $v > u$, the shape parameter ξ_u is unchanged but the GPD scale parameter τ_u changes to $\tau_v = \tau_u + (v-u)\xi$. Thus, if these parameters are respectively modelled using regression formulations with covariates x and z at threshold u, then their values $\tau_u(x)$ and $\xi_u(z)$ change to $\tau_v(x,z) = \tau_u(x) + (v-u)\xi_u(z)$ and $\xi_v(z) = \xi_u(z)$ at the higher threshold v: the way in which the covariates enter τ depends on the threshold. In many applications the shape is taken to be constant, or, even if not, the numerical value of $(v-u)\xi_u(z)$ may be relatively small, but the statistically unnatural behaviour of τ suggests that it is usually preferable to insert covariates into the parameters of the generalized extreme-value distribution or the corresponding point process model, which do not depend on the threshold. The authors of the present paper are clearly aware of this, but it ought to be more widely appreciated.

3. Efficiency

Peaks over threshold modelling is often said to be more efficient than the block maximum approach. Madsen *et al*. (1997a,b) simulated independent data to compare these approaches in estimating a high quantile, and concluded that with maximum likelihood estimation the peaks over threshold method does indeed have lower root mean squared error, though the efficiency gain is less cut-and-dried for other estimators popular among hydrologists. The gain when using maximum likelihood depends on the mean number of annual exceedances, and in the cases considered in this paper, the mean squared error for threshold-based estimation of a 100-year return period based on 30 years of data, a value typical in hydrological studies, seems likely to be around 0.6 of that of using annual maximum series. For independent data, there is thus a substantial gain in efficiency through using peaks over thresholds.

Real data usually exhibit clustering. Fawcett and Walshaw (2007) show that declustering of dependent exceedances over a threshold can systematically bias the estimators of σ and ξ, leading to severe underestimation of extreme quantiles for very dependent data. They recommend that if the sole interest is the estimation of extreme quantiles, inference should be based on all exceedances, with appropriate adjustment for the within-cluster dependence. Guidance on how clustering affects estimation efficiencies does not yet seem to be available.

4. Alternative approach

Efficient estimation is desirable, but more importantly in many applications, the peaks over threshold approach enables detailed modelling of clusters of rare events. The focus of the present paper is estimation of high quantiles, however, and then one may question the need for any threshold model. A third way uses neither annual maxima nor peaks over thresholds, but the r-largest values $y_s^{(1)} \geq \cdots \geq y_s^{(r)}$ at the sth site $(s = 1, \ldots, S)$. If covariates x_s affect the location parameter and the data from different sites are supposed to be independent, then the likelihood function is

$$\prod_{s=1}^{S}\prod_{j=1}^{r}\sigma^{-1}\left(1+\xi z_s^{(j)}\right)_+^{-1/\xi-1} \times \exp\left\{-n_y\left(1+\xi z_s^{(r)}\right)_+^{-1/\xi}\right\}, \tag{1}$$

where $a_+ = \max(a, 0)$, n_y is the number of years of data, and for compactness we write $z_s^{(j)} = \{y_s^{(j)} - \mu(x_s)\}/\sigma$, for $j = 1, \ldots, r$. Regression models for the scale and shape parameters σ and ξ may be added in (1), and correction for dependence between the different sites may be performed as in the paper, with only minor changes. Use of (1) has the following advantages: (a) there is no need for a quantile regression model, with the additional effort and complexity that it entails. In particular, there is no need to introduce a further component of uncertainty

(neglected in the paper) to allow for the estimated threshold; (b) unlike with using the classical GPD, the parametrization is threshold-invariant; and (c) choice of the threshold is replaced by choice of r. This poses the same bias/variance tradeoff as does threshold choice, but since the threshold does not appear in (1), there is no need to model it. Moreover, the number of observations contributing to (1) might be allowed to vary from site to site, i.e., we might replace r with r_s.

An equivalent formulation is that the threshold at the sth site is chosen to equal $y_s^{(r)}$, and that, conditional on $y_s^{(r)}$, there are $r-1$ independent exceedances each having the GPD with scale and shape parameters respectively $\tau_s^{(r)} = \sigma + \xi\{y_s^{(r)} - \mu(x_s)\}$ and ξ. The likelihood contribution from the data at the sth site is then proportional to

$$\frac{1}{\sigma}\left(1+\xi z_s^{(r)}\right)_+^{-r/\xi-1} \exp\left\{-n_y\left(1+\xi z_s^{(r)}\right)_+^{-1/\xi}\right\} \times \prod_{j=1}^{r-1} \frac{1}{\tau_s^{(r)}}\left(1+\xi\frac{y_s^{(j)}-y_s^{(r)}}{\tau_s^{(r)}}\right)^{-1/\xi-1}. \tag{2}$$

Apart from constants of proportionality, the first term in (2) is the density of $y_s^{(r)}$, and the second is the joint density of $y_s^{(1)}, \cdots, y_s^{(r-1)}$, conditional on $y_s^{(r)}$. Expression (2) is just a re-expression of (1), but it shows that, unlike with quantile regression, the effect of using a random threshold is directly incorporated into the likelihood.

Possible drawbacks of our proposal are that: (d) the extrema $y_s^{(1)} \geq \cdots \geq y_s^{(r)}$ are supposed to be based on independent identically distributed variables, and in practice some declustering of an underlying time series will be needed, in order to obtain 'independent' cluster maxima. This may introduce bias (Fawcett and Walshaw, 2007); (e) strongly seasonal data must be divided into roughly stationary blocks (e.g., months), whose r largest values are taken. Typically r will then be small, and covariates representing the seasonal variation must be introduced (Frossard, 2010); and (f) detailed modelling of clusters themselves cannot be undertaken.

In the present case, the prior declustering and the apparent lack of seasonality imply that these drawbacks are unimportant.

5. Application

Figures 1 and 2 show results from using (1) to fit the extremal model to the data from the paper.

The r-largest estimates for $r = 130$ shown in Figure 1 are very similar to those in Table 2 of the paper, but the standard errors for the parameters contributing to μ, obtained using the analogue of the variance estimate in expression (2) of the paper, are much smaller. We suspect that this reflects a difficulty with the matrix $\widehat{V}$, which is often close to singular; for $r = 130$, the ratio of its largest and smallest singular values is 4.6×10^7, and this ratio is never smaller than 7.7×10^5 for $r = 20, 30, \ldots, 130$. Further signs of trouble are the counter-intuitive finding in §3.3 of the paper that adjustment for dependence increases the significance of the regression coefficients, and that for large r the

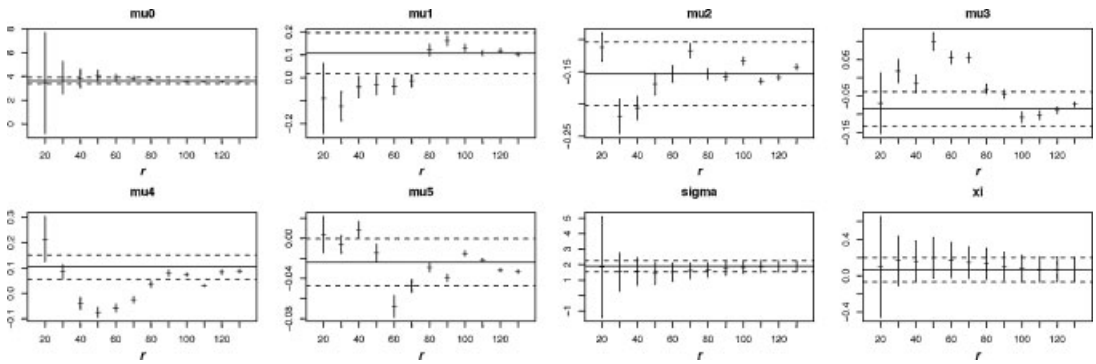

Figure 1. Parameter estimates and pointwise 95% confidence intervals based on adjusted standard errors, from Northrop and Jonathan (horizontal lines) and using the r-largest method with $r = 20, 30, \ldots, 130$

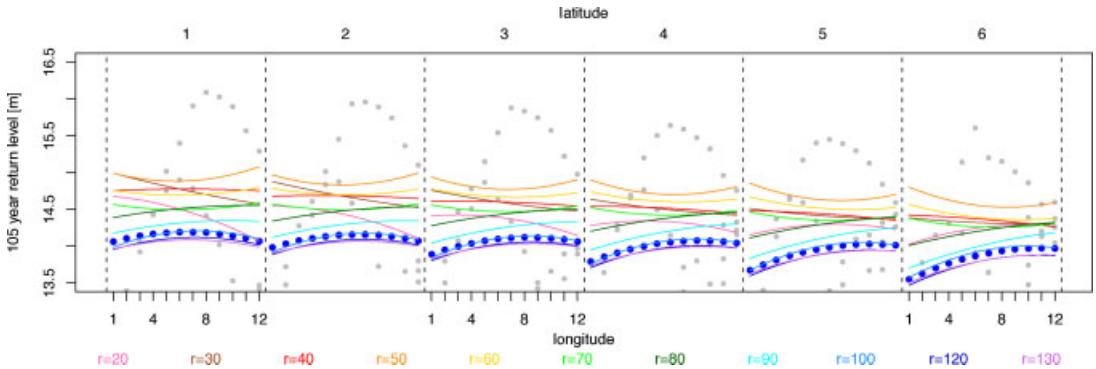

Figure 2. Estimated 105-year return level from Northrop and Jonathan (heavy blobs) and using the r-largest method with $r = 20, 30, \ldots, 130$. The lighter grey blobs are data points

adjusted standard errors for $\widehat{\mu}_0$ are smaller than the unadjusted ones, though the opposite is true for small r. These difficulties with $\widehat{V}$ may be due to the relatively high correlations among the extreme wave heights (Figure 3 of the paper). In our experience, such matrices can be quite unstable in extremal contexts, and it seems worthwhile to investigate how they can be made more reliable, or, perhaps better, avoided.

Figure 2 compares the 105-year return levels from the paper with those using (1). As one might anticipate from Figure 1, there is strong dependence on r, but the estimates using the two approaches are quite similar for $r \geq 80$. In view of the uncertainty shown in Figure 7 of the paper, the differences between the approaches seem unimportant. More surprisingly, the choice of model for μ essentially removes the marked spatial pattern seen in Figure 1 of the paper, and the shapes of the return levels depend strongly on r, and, by implication, on the threshold. The reason for this undesirable behaviour is the changing signs of the estimated coefficients of the Legendre polynomials (cf. Figure 1); curiously, taking smaller r leads to fits that clash even more strongly with Figure 1 of the paper.

It would be interesting to see if explicit modelling of the spatial dependence could remove some of the difficulties with both the approach proposed in the paper and our alternative.

Acknowledgements

We thank the authors for providing the wave height data, the editor for the opportunity to contribute to this discussion, and Peter Craigmile for comments on a draft. The work was supported by the Swiss National Science Foundation and the CCES project EXTREMES.

(wileyonlinelibrary.com) DOI:10.1002/env.1125

Daniel Cooley[a*]

* *Correspondence to: D. Cooley, Department of Statistics, Colorado State University, Fort Collins, Colorado. E-mail: cooleyd@stat.colostate.edu*

Let me begin by congratulating Northrop and Jonathan (2011) on a very sensible approach to a real and important application. This work's ultimate goal is to provide engineers with accurate assessments of wave heights so that safe marine structures can be constructed. As recent history shows, failure of marine structures (whether due to human error or extreme natural events) can have immense consequences. The authors' approach, while accounting for spatial variation in the extremal behavior, is very accessible and does not require fitting of complicated multivariate/spatial models. The difference between the estimates for the shape parameter and between the current analysis and the earlier, more simple analysis (Section 3.3) illustrates the need for accounting how extremes can vary with location and would doubtlessly have a noticeable effect on estimates of long return levels for wave heights.

Two approaches for describing the upper tail have seen increased interest over the last couple of decades: quantile regression and extreme value theory (EVT). Each approach has its advantages. Quantile regression allows the practitioner to model high quantiles on observed covariates without requiring one to make assumptions about the underlying distribution. However, one cannot use quantile regression to extrapolate beyond the range of the observed data. The advantages of EVT are almost the opposite of quantile regression. Asymptotic theory results in a known parametric form for the tail, which allows one to extrapolate. Importantly, the parametric form for the tail does not require the practitioner to make overarching assumptions about the particular distribution which generated the data. To my knowledge, no one has previously tried to employ both techniques in tandem.

When modeling threshold exceedances, selecting a threshold is often tricky. The usual diagnostics such as mean-residual-life plots are open to interpretation and can demonstrate unrealistic variability between spatial locations which should share similar characteristics. As the authors demonstrate, it is often inappropriate to set a common threshold over a study region. The quantile regression approach for threshold selection offers several advantages, not the least of which is that it borrows strength across locations when setting thresholds. One still must check that the threshold is set so that EVT applies. Equation (5) which relates the threshold to the covariates for a given spatially-varying extreme value model is a straightforward calculation, but indicates that the authors have taken care that the quantile regression for the threshold does in fact work in tandem with the extreme value model for the exceedances.

The authors also appropriately handle the dependent observations. Here, like in the longitudinal analyses found in medical applications, dependence in the observations is a nuisance, and one is only interested in the marginal behavior and its relation to the covariates (location). Although a hurricane affects multiple locations at once, a marine structure is only struck by waves at its location. In other applications such as flood prediction, the effects are some function (e.g., the sum) of multiple locations and the dependence must be explicitly modeled. Modeling multivariate or spatial extremes requires much more work, and the available models are limited, some references are Schlather (2002);

Kabluchko *et al*. (2009); Cooley *et al*. (2010). Spatial models for threshold exceedances are still being formulated. Rather than explicitly modeling the dependence between sites, in this application the authors are able to circumvent the issue by fitting an independence likelihood and then using the Godambe information approach to account for the mis-specified model. This approach could readily be applied by researchers who might find modeling dependent extremes challenging. However, this route cannot be chosen solely for convenience. One must keep in mind what quantities are critical for one's application.

The study region is relatively simple which allows the authors to fit a relatively simple model based on EVT; let me make a few comments about other situations in which the model might need to be more complex. First, the authors are able to model the marginal parameters with a trend surface; in fact they need only to model the location parameter μ, and do so by regressing on Legendre polynomials of latitude and longitude. In many situations, available covariates are not rich enough to characterize how extremes vary over a study region and it becomes necessary to employ a hierarchical approach (Cooley *et al*., 2007; Sang and Gelfand, 2009). Ribatet *et al*. (2010) show a trend surface approach is unable to capture the extremal behavior of precipitation over Switzerland. Fitting a hierarchical model is done via MCMC methods and requires both a significant amount of human and computational time; it should only be done when the study region is sufficiently complex to necessitate the approach. Conversely, a trend surface approach should only be used when the study region is sufficiently simple, such as in this example.

More related to this article's threshold modeling procedure, the authors presume that both ξ and p, the probability of exceeding the threshold, are constant over the study region. The authors assume a constant ξ when developing equation (5), but I see no reason that the argument leading to this equation could not be adapted to the case where ξ varies with the location $\boldsymbol{x}_{i,j}$. Because the study region is relatively simple, assuming a constant ξ is appropriate, but there are situations where assuming a constant shape parameter is inappropriate. In our study of modeled precipitation over the Western US, Cooley and Sain (2010), found the shape parameter to vary over the region. This is important because it shows that the distribution's tail differs over the region in a fundamental way that cannot be captured by a shift or scaling. This has implications for p, for if the distribution is fundamentally different, then its rate of convergence to the extremal type is likely to also differ over the study region. Consequently using a quantile regression approach for a common p across the study region might be inappropriate. I imagine the quantile regression approach could be adapted to allow p to vary over a more complex region.

Discussion

(wileyonlinelibrary.com) DOI:10.1002/env.1125

I. Papastathopoulos[a], J. A. Tawn[a]* and J. L. Wadsworth[a]

* *Correspondence to: J. A. Tawn, Department of Mathematics and Statistics, Lancaster University, Lancaster, Lancashire, UK. E-mail: j.tawn@lancaster.ac.uk*

a *Department of Mathematics and Statistics, Lancaster University, Lancaster, Lancashire, UK*

We would like to congratulate the authors on proposing what promises to be a useful new approach for threshold selection in extreme value problems with covariates. The major development of the paper is the use of quantile regression for threshold selection. However, other aspects of the paper that we particularly like are the use of spatial methods for marginal extreme value analysis, their illustration through a major application, and the careful use of diagnostics.

The use of data arising only from hurricane events is interesting. Is this because the underlying mechanisms suggest that wave heights in hurricanes arise from a different population to other extreme wave heights? If this is so, then is maximization over three hurricane events per year sufficient to justify the fit of the GEV (as described in Section 3.2)? If this is not the case, then are potentially informative data being excluded from the analysis?

We suspect that the analysis of the hindcast data in the paper does not illustrate the full benefits of quantile regression for the threshold selection. As all sites have the same amount of data over the same period of time, site-wise threshold selection using the same fixed empirical quantile could produce similar findings to the current analysis, but with substantially reduced effort. However in other applications with non-equal data lengths at different sites we can see that the proposal may have substantial benefits. An unsettling feature of quantile regression is the lack of ordering of estimated conditional quantiles for different percentiles, with model mis-specification and heterogeneity across the design space being two of the most common causes for the violation of the monotonicity of quantile estimates. In the present analysis the authors have not addressed this issue. Were different exceedance probabilities examined with respect to the crossing effect?

Section 3.2.2 raises an interesting point on the connection between the parameterization of covariate effects in $\mu(\boldsymbol{x}_{ij})$, $\sigma(\boldsymbol{x}_{ij})$ and $u(\boldsymbol{x}_{ij})$ under an assumption of constant exceedance probability. Typically, the quantity on the right hand side of equation (4) is estimated rather well in extremal analyses. Thus, if a constant exceedance probability is attained by a linear quantile regression for $u(\boldsymbol{x}_{ij})$, then we would indeed

anticipate that any further model selection would implicate linear models for $\mu(\boldsymbol{x}_{ij})$ and $\sigma(\boldsymbol{x}_{ij})$. This highlights the importance of attaining an appropriate regression form for $u(\boldsymbol{x}_{ij})$. The authors' current suggestion is to inform this process entirely by a GEV analysis of maxima. This seems a sensible idea, but given the influence of this stage, could it be augmented by alternative procedures?

The parameter stability plots in Figure 4 appear to show that covariate trends may not be independent of threshold. Of course, the power to detect genuine trends decreases with sample size, but the change of sign of point estimates for some parameters suggests the possibility that some covariate effects change or disappear at higher thresholds. Could it be possible that different regression forms for $u(\boldsymbol{x}_{ij})$ are implied for different exceedance probabilities? A simple, informal, way to assess this using the authors' current approach could be to consider GEV fits using different block lengths to derive maxima, block length having analogies with threshold level. However, the aforementioned difficulties of reduced power in relation to smaller sample sizes may be an issue here.

Table 1 displays compelling evidence for taking the quadratic form for the linear model in $\mu(\boldsymbol{x}_{ij})$ when $\sigma(\boldsymbol{x}_{ij})$ is held fixed; this is unsurprising in light of the preceding discussion surrounding Section 3.2.2. Were threshold models with non-constant $\sigma(\boldsymbol{x}_{ij})$ fitted in conjunction with each of the entertained models for $\mu(\boldsymbol{x}_{ij})$? Given that equation (5) effectively dictates the form of $\mu(\boldsymbol{x}_{ij}) + c\sigma(\boldsymbol{x}_{ij})$, it could be interesting to analyze which of the covariate effects fitted to $u(\boldsymbol{x}_{ij})$ are naturally assigned to $\mu(\boldsymbol{x}_{ij})$, and which to $\sigma(\boldsymbol{x}_{ij})$. An intriguing question is whether it is possible to assess the choice of covariate fit to $\mu(\boldsymbol{x}_{ij})$, $\sigma(\boldsymbol{x}_{ij})$ and $u(\boldsymbol{x}_{ij})$ simultaneously.

Figure 7 seems to suggest that there are limited covariate effects on the fitted return levels relative to the uncertainty in the estimates, and that potential structure in the at-site estimates has been smoothed out. We wonder if there is covariate structure for $\sigma(\boldsymbol{x}_{ij})$ which has been omitted. In particular, your assessment of the spatial model fit using site-by-site Q–Q plots provides relatively weak diagnostic ability. Heffernan and Tawn (2001) were able to draw stronger conclusions by constructing pooled Q–Q plots by combining data over independent but not identically distributed units. Here that approach may be helpful, though the spatial dependence in the data will need accounting for in developing tolerance intervals for such plots.

The authors acknowledge that no element of uncertainty is carried forth from the selection of the exceedance probability for the quantile regression or the covariate fit to $u(\boldsymbol{x}_{ij})$ into the subsequent analysis. This is a ubiquitous problem in almost all (stationary and non-stationary) extreme value analyses. There is a growing body of literature surrounding the issue of threshold uncertainty under stationarity (e.g. Frigessi *et al*., 2002; Tancredi *et al*., 2006). To our knowledge, this problem has yet to be considered under non-stationarity. This paper plays a helpful role in clarifying the need for such a development.

Rejoinder

(wileyonlinelibrary.com) DOI:10.1002/env.1125

Paul Northrop[a*] and Philip Jonathan[b]

* *Correspondence to: P. J. Northrop, Department of Statistical Science, University College London, Gower Street, London, WC1E 6BT, UK. E-mail: paul@stats.ucl.ac.uk*

a *Department of Statistical Science, University College London, Gower Street, London, WC1E 6BT, UK*

b *Shell Technology Centre Thornton, P.O. Box 1, Chester, CH1 3SH, UK*

We would like to thank all the discussants for their comments. We have found them to be extremely helpful and we expect others will too. There are some interesting suggestions for methodological developments.

Alternative approaches. Chavez-Demoulin *et al.* suggest, and implement, an alternative approach based on the r largest values at each site. We agree that for data of the type considered in our paper—pre-declustered balanced response data on a spatial grid—that this approach has advantages over using quantile regression to set a threshold. Tawn *et al.* made a similar suggestion and we agree with their comment that the data in our paper are not best suited for illustrating the potential of quantile regression to set thresholds. In more general situations, for example, where continuous covariates are involved, then quantile regression can be used to set a threshold without the need to discretize covariates. Daniel Cooley gives an excellent summary of various issues, including the importance of tailoring modelling approaches based on the needs of the practical application, the complexity of the problem and the availability of data. We appreciate his point that our approach is relatively accessible for practical application by engineers.

Threshold selection using quantile regression. Selecting an appropriate level of threshold is non-trivial because threshold diagnostic plots, such as those in Figure 4 of our paper, can be expected to depend on the model used to produce them. This necessitates an iterative process in which threshold diagnostic plots based on a sensible working model are first used to suggest a suitable level for the threshold. The threshold diagnostics are to be revisited if the working model has been modified subsequently. We suspect that we did not make this iterative process clear.

We used a GEV fit to annual maxima merely to provide the working model. As Tawn *et al.* suspect the physics of hurricane events is different to that of other storms. The hurricane events in the data represent the most severe occurrences of specific weather conditions giving rise to tropical storms and hurricanes. Although the GEV model appeared to fit the data well in this example Tawn *et al.* are quite right to question whether this is an appropriate strategy more generally. In many cases simple exploratory plots should be sufficient to suggest a working model.

We think that Tawn *et al.* overstate the link between the quantile regression fit used to set the threshold and the resulting extreme value fit. Of course, there is some link, because they are based on the same raw data. However, they use different aspects of these data. Provided that there are no data points with high leverage (as will be the case for the gridded data considered in our paper) the quantile regression fit is relatively insensitive to the values of the largest threshold exceedances. In contrast, these values will be influential in the fit of the extreme value model.

Tawn *et al.* point out that quantile regression lines for different quantiles can cross. We confess that we had not previously investigated this. However, it turns out that the regression lines for different exceedance probabilities do not cross for our data. Bondell *et al*. (2010) have developed methodology for avoiding the crossing problem. However, it may be preferable to check for crossing of thresholds as this will highlight model mis-specification and thus provide an additional threshold diagnostic.

Chavez-Demoulin *et al.* note that the estimates of regression parameters μ_i, $i = 1, \ldots, 5$ are much less stable over different levels of threshold than the estimates of the marginal parameters μ_0, σ, ξ. We believe this is at least partly due to the effects of the particular positioning of the larger individual hurricanes relative to the data grid. This can be seen in Figure 7 in our paper. In future work we will consider whether it is beneficial to incorporate the positioning of hurricanes into the modelling.

We like the idea (Tawn *et al.*) of trying to disentangle the relative contributions of μ and σ to the threshold u. If μ and σ are both linear in the same covariates then it cannot be possible to do this based only on a single quantile regression line. One would need at least two quantile regression lines in order to identify all the parameters involved. If, for example, μ and $\log\sigma$ are linear in the covariates the fitting the implied non-linear quantile regression model would estimate separately the covariate effects on μ and $\log\sigma$. The question of estimating simultaneously the threshold and the parameters of the extreme value model is one we have considered. One possibility is to fix the number of exceedances and maximize the log-likelihood over the parameters of the extreme value model based on a threshold implied by equation (5) of our paper. This is a challenging optimization problem: as the parameters of the extreme value model are changed, the threshold changes and data points move across the threshold. We have recently found a discussion of this on page 227 of Beirlant *et al*. (2004).

Modelling of wave data. Tawn *et al.* ask about covariate effects on σ. We did report on this briefly in section 3.3 of the paper, finding that spatial variation in σ was not strong. We have also fitted a model in which μ is quadratic in space and σ is proportional to μ, finding that the maximized log-likelihood for this model is much smaller than for the corresponding model with σ constant. We think that the spatial effects appear over-smoothed at least partly because, in hindsight, Figure 1 is a too simplistic a summary of the spatial effects. It overstates the spatial effect of location as it is dominated by two hurricanes.

Tawn *et al.* suggest pooling suitably-defined residuals from different sites in a Q–Q plot, acknowledging that this is not a trivial exercise due to the spatial dependence in the data. If dependent data are pooled then the plotting positions of the Q–Q plot will also need to be adjusted for spatial dependence since, for example, the largest value from a dependent sample is stochastically smaller than the largest value from an independent sample. A simple (but less useful) alternative is to superimpose separate Q–Q plots of residuals from each of the sites.

Parameterization. Daniel Cooley considers situations where it is necessary to allow ξ to vary with covariates. We agree that in this event equation (5) in our paper can be used to estimate a threshold: using non-linear quantile regression. Daniel Cooley also points out that it may then be more appropriate to allow the probability of exceedance p to vary with the covariates. We agree that it should be possible to do this for given values of p. However, choosing p would require knowledge of the distribution of the response variable or empirical justification. In the spatial case perhaps separate threshold diagnostic plots could be produced for different regions.

We agree with Chavez-Demoulin *et al.* that it is often useful or necessary to build covariate effects into the GEV parameters (μ, σ, ξ). An argument sometimes made for the alternative parameterization—covariate effects in the probability p of exceedance and/or the scale parameter σ_u of the GP distribution of exceedances of u—is the orthogonality of the p and σ_u parameters. This can improve the speed and ease of model fitting in comparison to the non-orthogonal (μ, σ, ξ) parameterization. We suggest the following hybrid approach. Suppose that covariates have been mean-centred, u_0 is the threshold when all covariates are zero and μ_0 and σ_0 are the intercepts of μ and σ. We parameterize the covariate effects in terms of μ and σ. Then, for the purposes of fitting the model only, we reparameterize (μ_0, σ_0) to $(p_0, \sigma_{u0}(1+\xi))$ where $\sigma_{u0} = \sigma_0 + \xi(u_0 - \mu_0)$ and $p_0 = (1/\lambda)[1 + \xi(u_0 - \mu_0)/\sigma_0]^{-1/\xi}$. Thus (Davison, 2003, page 688) p_0, $\sigma_{u0}(1+\xi)$ and ξ are orthogonal. We have implemented this for the model fitted to the wave data in our paper and have found that convergence is faster and initial parameter values are less critical.

Estimation of adjusted standard errors. Chavez-Demoulin *et al.* raise issues surrounding the adjustment of standard errors. That the adjustment for dependence can reduce the standard errors of the regression coefficients is not counter-intuitive. Consider as an example the estimation of the difference μ_d in GEV location μ between two sites. This difference is more precisely estimated when there is positive dependence between the data at the sites than when the series sites are independent: the differences between the responses at the two sites will be less variable in the former case than in the latter. In the extreme case where the two series are completely dependent, μ_d is estimated with perfect precision because pairs of values from the two sites always differ by exactly μ_d. Chandler and Bate (2007) also observe this phenomenon for a bivariate extreme value model.

The adjusted variance matrix of the parameter estimates is estimated by $\widehat{H}^{-1}\widehat{V}\widehat{H}^{-1}$. The extent of the adjustment is governed by how different $\widehat{V}$ is from $-\widehat{H}$. Adjustment for strong spatial dependence may increase the standard errors of the marginal parameters substantially and decrease the standard errors of the marginal parameters substantially. Thus we can expect $\widehat{V}$ to have large and small elements and thus have a large condition number (ratio of the largest and smallest singular values). A large condition number does not necessarily mean that the adjusted standard errors are incorrect. In the simulation study of section 5 we created very strong spatial dependence and found that the adjusted standard errors were of the correct magnitudes.

Of course it is undesirable for calculations to involve matrices with large condition numbers. We could alter the condition number of $\widehat{V}$ quite substantially by a simple scaling of the parameters involved. Chavez-Demoulin *et al.* ask whether we can avoid estimating *V*. We note that

$$\widehat{H}^{-1}\widehat{V}\widehat{H}^{-1} = \widehat{H}^{-1}\left(\sum_{i=1}^{k} U_i^T(\widehat{\theta})U_i(\widehat{\theta})\right)\widehat{H}^{-1} = \sum_{i=1}^{k}\widehat{H}^{-1}U_i(\widehat{\theta})^T U_i(\widehat{\theta})\widehat{H}^{-1} = (\widehat{U}\widehat{H}^{-1})^T\widehat{U}\widehat{H}^{-1},$$

where $\widehat{U}$ is a matrix of score contributions evaluated at $\widehat{\theta}$ and k is the number of clusters. Instead of calculating $\widehat{V}$ we work with transformed scores $\widehat{U}\widehat{H}^{-1}$. For the final model in our paper the condition numbers of $\widehat{V}$ and $\widehat{H}^{-1}$ are 9150 and 28, respectively. The condition numbers of $\widehat{U}$ and $\widehat{U}\widehat{H}^{-1}$ are 96 and 18, respectively. Nevertheless, we agree that research is needed on how best to estimate adjusted standard errors for extreme value models.

REFERENCES TO THE DISCUSSIONS AND REJOINDER

Beirlant J, Goegebeur Y, Segers J, Teugels J. 2004. Statistics of Extremes: Theory and Applications. Wiley: Chichester, England.

Bondell HD, Reich BJ, Wang H. 2010. Non-crossing quantile regression curve estimation. *Biometrika* **97**(4): 825–838.

Chandler RE, Bate SB. 2007. Inference for clustered data using the independence loglikelihood. *Biometrika* **94**(1): 167–183.

Cooley D, Davis RA, Naveau P. 2010. The pairwise beta: a flexible parametric multivariate model for extremes. *Journal of Multivariate Analysis* **101**: 2103–2117. http://www.stat.colostate.edu/∼cooleyd/Papers/pairwiseBeta.pdf.

Cooley D, Nychka D, Naveau P. 2007. Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association* **102**: 824–840.

Cooley D, Sain SR. 2010. Spatial hierarchical modeling of precipiation extremes from a regional climate model. *Journal of Agricultural, Biological, and Envrionmental Statistics* **15**: 381–402.

Davison AC. 2003. Statistical Models. Cambridge University Press: Cambridge.

Fawcett L, Walshaw D. 2007. Improved estimation for temporally clustered extremes. *Environmetrics* **18**: 173–188.

Frossard L. 2010. *Modelling Minima of Ozone Data in the Southern Mid-Latitudes*. Master's thesis, Ecole Polytechnique Fédérale de Lausanne, Section de Mathématiques.

Frigessi A, Haug O, Rue H. 2002. A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes* **5**: 219–235.

Heffernan JE, Tawn JA. 2001. Extreme value analysis of a large designed experiment: a case study in bulk carrier safety. *Extremes* **4**: 359–378.

Kabluchko Z, Schlather M, de Haan L. 2009. Stationary max-stable fields associated to negative definite functions. *The Annals of Probability* **37**: 20422065.

Madsen H, Rasmussen PF, Rosbjerg D. 1997a. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 1. At-site modeling. *Water Resources Research* **33**: 747–757.

Madsen H, Rasmussen PF, Rosbjerg D. 1997b. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 2. Regional modeling. *Water Resources Research* **33**: 759–769.

Northrop P, Jonathan P. 2011. Threshold modeling of spatially-dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics*

Ribatet M, Cooley D, Davison AC. 2010. Bayesian inference for composite likelihood models and an application to spatial extremes. Submitted. http://arxiv.org/abs/0911.5357.

Sang H, Gelfand AE. 2009. Hierarchical modeling for extreme values observed over space and time. *Environmental and Ecological Statistics* **16**: 407–426.

Schlather M. 2002. Models for stationary max-stable random fields. *Extremes* **5**(1): 33–44.

Tancredi A, Anderson C, O'Hagan A. 2006. Accounting for threshold uncertainty in extreme value estimation. *Extremes* **9**: 87–106.

