Estimating return levels from serially dependent extremes

Lee Fawcett*and David Walshaw,
Newcastle University, Newcastle upon Tyne, UK

November 25, 2010

Summary

In this paper, we investigate the relationship between *return levels* of a process and the strength of serial correlation present in the extremes of that process. Estimates of long range return levels are often used as design requirements, and *peaks over thresholds* (POT) analyses have, in the past, been used to obtain such estimates. Using real–life data examples involving wind speeds and sea–surges, we demonstrate the sensitivity of return level estimates obtained via POT analyses to the choice of scheme for identifying *clusters* of extremes. Furthermore, with simulated data we show that using all threshold excesses, instead of a filtered set of independent cluster peak excesses, can give more accurate estimates of return levels – provided the dependence on serial correlation is properly quantified.

As well as avoiding the sometimes arbitrary process of cluster identification, and providing more accurate estimates of return levels, we demonstrate that using all extremes greatly increases precision relative to the usual POT approach.

Key words: Extreme value theory, temporal dependence, clusters, peaks over thresholds, return levels, sea–surges, wind speeds.

Address for correspondence: Dr Lee Fawcett, School of Mathematics and Statistics, Herschel Building, Newcastle University, Newcastle upon Tyne NE1 7RU, UK. Email: `lee.fawcett@ncl.ac.uk`

1 Introduction

Statistical modelling of environmental extremes has a very practical motivation – reliability: anything that is built needs to have a good chance of withstanding the weather/environment it is exposed to for the duration of its working life. This has obvious implications for civil engineers and planners. For example, they need to know how strong to make buildings; how high to build sea–walls; how tall to build reservoir dams; motivating the need to estimate the strongest wind; the highest tide; the heaviest rainfall. Thus, estimating such *return levels* is often the primary objective in an analysis of extreme values. A commonly employed estimation procedure is the peaks over threshold (POT) approach (see, for example, Davison and Smith, 1990). Here, an appropriate limiting distribution for independent excesses over a high threshold u is fitted to the largest exceedance selected from each 'cluster' of values above u – these *cluster peak* excesses are used instead of *all* excesses to avoid the issue of serial correlation, a phenomenon usually present in environmental extremes. Return levels are then estimated using high quantiles of the limiting distribution, fitted values of which are obtained by inversion of the corresponding distribution function.

Data on extremes are, by their very nature, scarce. Although methods such as the peaks over threshold approach do make use of more information on extremes than the classical annual maxima approach (see, for example Coles (2001), Ch. 3), discarding all but the cluster maxima is still extremely wasteful of precious data. This usually manifests itself when we come to quantify our uncertainty in parameter estimates. For example, standard errors attached to estimates of return levels as a result of a POT analysis are often so large that confidence intervals constructed in the usual way (parameter estimate $\pm$ 1.96 $\times$ standard error) are completely uninformative (see, for example, the analysis of river flow extremes in the River Nidd by Davison and Smith, 1990). Making use of the profile log–likelihood (Venzon and Moolgavkar, 1988) can improve inference for return levels, but resulting 95% confidence intervals, for example, can still be very wide.

Not only does a POT analysis result in a loss of precision – extensive simulations in Fawcett and Walshaw (2007) show that using cluster peak excesses results in significantly biased estimates of model parameters and associated quantiles. Findings from their simulations are supported by an analysis of sea–surge extremes in the same paper, as well as analyses of: wind speed extremes in Fawcett (2005), river flow extremes in Eastoe and Tawn (2010) and hurricane–induced wave heights in Northrop and Jonathan (2010). Fawcett and Walshaw (2007) show that, by initially ignoring any serial correlation between threshold excesses, bias in parameter estimates is virtually eliminated. Obviously, estimated standard errors will be misleading since, owing to the presence of (often quite strong) serial correlation, there will be less information in the data than the likelihood requires if the data were independent; however, they implement a well–known procedure for modifying the usual variance–covariance matrix, resulting in standard errors that are inflated to account for temporal dependence.

Ignoring temporal dependence and making use of all extremes, as suggested by Fawcett and Walshaw (2007), can improve over typical POT analyses when making inferences for model parameters. However, the presence of serial correlation cannot simply be ignored when estimating return levels. In this paper, we investigate the relationship between return levels of extremes of a stationary process and the strength of serial correlation present in

that process. We argue that, provided this dependence can be quantified in an appropriate way, inference for return levels using all extremes can improve over a typical POT analysis: Not only can estimation precision be increased, but using all extremes completely avoids having to identify independent clusters of extremes. Indeed, we also highlight the sensitivity of return level estimates obtained from a POT analysis to the method used to identify the clusters which, in practice, is often chosen arbitrarily.

2 Background

2.1 Motivating examples

Figure 1 (top left) shows a series of 3–hourly measurements of sea–surge heights at Newlyn, a coastal town in the southwest of England, collected over a 3–year period. The sea–surge is the meteorologically induced non–tidal component of the still–water level of the sea. Figure 1 (top right) shows the first three years of a series of hourly gust maximum wind speeds recorded at High Bradfield, a high altitude site in the Pennines. The practical motivation for the study of such data is that structural failure of, perhaps, a sea–wall or a building, is possible if extreme surges or extreme wind speeds (respectively) are observed. The presence of strong short–term serial correlation is also apparent from the plots given directly underneath each time series, which show each value against their lag 1 counterpart. The green lines in these plots represent high thresholds above which events are classified as extreme (see Section 2.2); above these thresholds serial correlation still seems to exist, indicating the presence of *extremal* dependence between successive observations. Close inspection of the wind speed data also reveals non–stationarity in the form of seasonal variation.

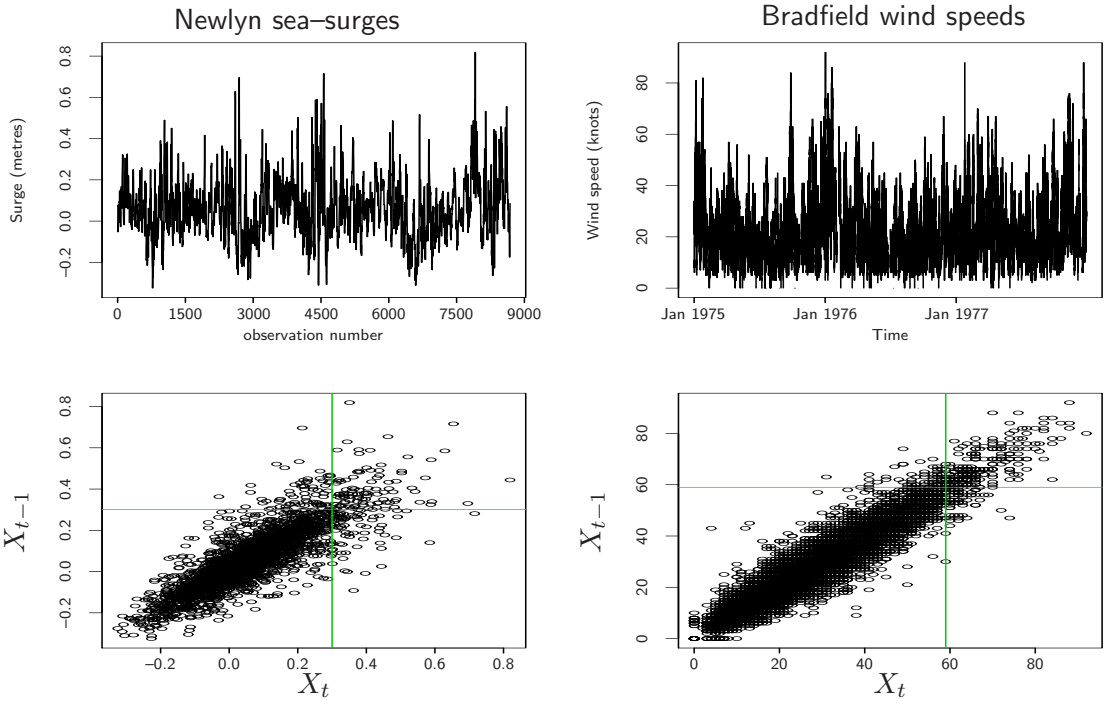

FIGURE 1: Left–hand–side: Newlyn sea–surge data; right–hand–side: Bradfield wind speed data. Top: time series plots; bottom: plots of time series against series at lag 1.

2.2 A model for extremes: the generalised Pareto distribution

A natural way of modelling extremes of time series such as those shown in Figure 1 is to use the Generalised Pareto Distribution (GPD) as a model for excesses over a high threshold. For example, suppose $X_1, \ldots, X_n$ is our series of independent and identically distributed (i.i.d.) random variables with common distribution function F. Then for a high threshold u, the conditional distribution of the exceedances of u, i.e. $X - u$, given that $X > u$, will be approximately GPD with distribution function

$$H(y) \;=\; 1 - \left(1 + \frac{\xi y}{\sigma}\right)_+^{-1/\xi}, \tag{1}$$

where $a_+ = \max(0, a)$ and σ and ξ are scale and shape parameters (respectively). This arises because, if a limiting distribution exists for any normalisation of $(X - u|X > u)$, then this limit will be exactly of GPD form (Davison and Smith, 1990). The limit is taken as $u \nearrow u_F$, where u_F is the finite upper endpoint of F if it exists, or ∞ otherwise.

In practice, a *mean residual life plot* (see for example Coles, 2001) can be employed to select a suitably high threshold u, and maximum likelihood can be used to estimate the GPD scale and shape parameters for excesses over u. The initial choice of threshold can then be assessed by fitting the GPD to other thresholds $u^* > u$, and checking for stability in estimates of σ^*[†] and ξ. Estimated standard errors for the GPD parameters can be obtained by referring to the usual asymptotic theory for maximum likelihood estimators, and confidence intervals constructed in the usual way.

2.3 Temporal dependence and non–stationarity

The arguments leading to the GPD (1) as the limiting distribution for excesses over u have assumed they are i.i.d.; environmental data often show considerable departures from this ideal. The plots in Figure 1 reveal substantial short–term serial correlation in both the Newlyn sea–surges and Bradfield wind speeds, even between successive observations exceeding u; the lag 1 autocorrelations for the sea–surge data and the wind speed data are 0.836 and 0.957 (respectively). The most commonly employed procedure is to filter out a set of independent extremes using 'runs declustering', whereby a cluster of extremes is deemed to have terminated as soon as κ consecutive observations fall below u. From each cluster, the maximum is then extracted and the GPD (1) is fitted to the set of independent cluster peak excesses (giving the usual POT analysis). Various approaches have been developed to deal with issues of seasonal variability, as shown in Figure 1 for the Bradfield wind speed data. Walshaw (1991) investigates the use of Fourier forms for the GPD parameters allowing them to vary continuously; Fawcett and Walshaw (2006) use a piecewise seasonality approach whereby a different GPD is fitted to sets of excesses over monthly varying thresholds u; often, attention is restricted to the season within which extremes appear largest (and are stationary within; see, for example, Smith *et al.*, 1997).

[†] $\sigma^* = \sigma + \xi u^*$; then (σ^*, ξ) are *threshold invariant* – if exceedances over a threshold u are generalised Pareto distributed, then exceedances over all $u^* > u$ are also GPD with the same parameters (Coles, 2001)

2.4 Return level estimation

Suppose the GPD(σ, ξ) is a suitable model for threshold exceedances ($X - u$) of a threshold u by a variable X. Then, from equation (1), and for $x > u$, it can easily be shown that

$$\Pr(X \leq x) \;=\; 1 - \lambda_u \left[1 + \xi \left(\frac{x-u}{\sigma}\right)\right]_+^{-1/\xi}, \tag{2}$$

where $\lambda_u = \Pr(X > u)$ is the threshold exceedance rate and can be estimated empirically as the proportion of observations above u. Estimates of an extreme quantile z_s can then be obtained by defining z_s by $\Pr(X \leq z_s) = 1 - s^{-1}$ and then solving (2) for z_s, where z_s is the *s–observation return level* associated with *return period* s; this can be thought of as the level that is exceeded once every s observations. It is usually more convenient to work with return levels on an annual scale; the *r–year* return level, z_r, is then given by

$$z_r \;=\; u + \frac{\sigma}{\xi}\left[(\lambda_u r n_y)^{\xi} - 1\right], \tag{3}$$

where n_y is the number of observations per year. In practice, r–year return level estimates $\hat{z}_r$ are commonly obtained by replacing (λ_u, σ, ξ) in Equation (3) with their maximum likelihood estimates $(\hat{\lambda}_u, \hat{\sigma}, \hat{\xi})$.

Confidence intervals for estimated return levels are usually constructed via profile likelihood, owing to the severe asymmetry of the likelihood surface often encountered for return levels, i.e., we obtain the set of values z_0 for which $2\{\ell(\hat{z}_r) - \ell(\hat{z}_0)\}$ is not significant when compared to a χ^2_1 distribution, where ℓ is the associated log–likelihood.

2.5 Oceanographic example: Newlyn sea–surges

We use a threshold of $u = 0.3$ metres to identify extreme sea–surges at Newlyn. In an analysis of the same dataset, Coles and Tawn (1991) suggest runs declustering using a cluster separation interval of 60 hours – or $\kappa = 20$ observations – for the sea surges at this location, to allow for wave propagation time. Using $\kappa = 20$ gives maximum likelihood estimates of $\hat{\lambda} = 0.013$ (0.0005), $\hat{\sigma} = 0.187$ (0.040) and $\hat{\xi} = -0.259$ (0.146), with estimated standard errors shown in parentheses. Figure 2 shows a plot of estimated r–year return levels found by substituting these maximum likelihood estimates into Equation (3) and using a range of values for r. This plot has been constructed on the usual logarithmic scale to compress the tail of the distribution so that return level estimates for long return periods are displayed. The dotted lines give the corresponding 95% confidence bands obtained by referring to profile log–likelihood curves at each value of r used. These intervals are heavily asymmetric, with the interval widths increasing dramatically as r increases. For example, the 95% confidence interval for the 10–year return level is (0.765, 1.569) metres, whereas the 1000–year return level has a range of (0.835, 3.365) metres.

Although Coles and Tawn (1991) suggest $\kappa = 20$ for this dataset, κ is often chosen arbitrarily or, at best, simply by visual inspection of the data. As an illustration of the sensitivity of return level estimation to the choice of cluster separation interval, we have reproduced the return level curve obtained using $\kappa = 20$ for $\kappa = 5$, 30 and 50 (Figure 2). We see that using $\kappa = 30$ observations over–estimates relative to return levels obtained using $\kappa = 20$; conversely, using $\kappa = 50$ observations produces relatively *under*–estimated return levels. From

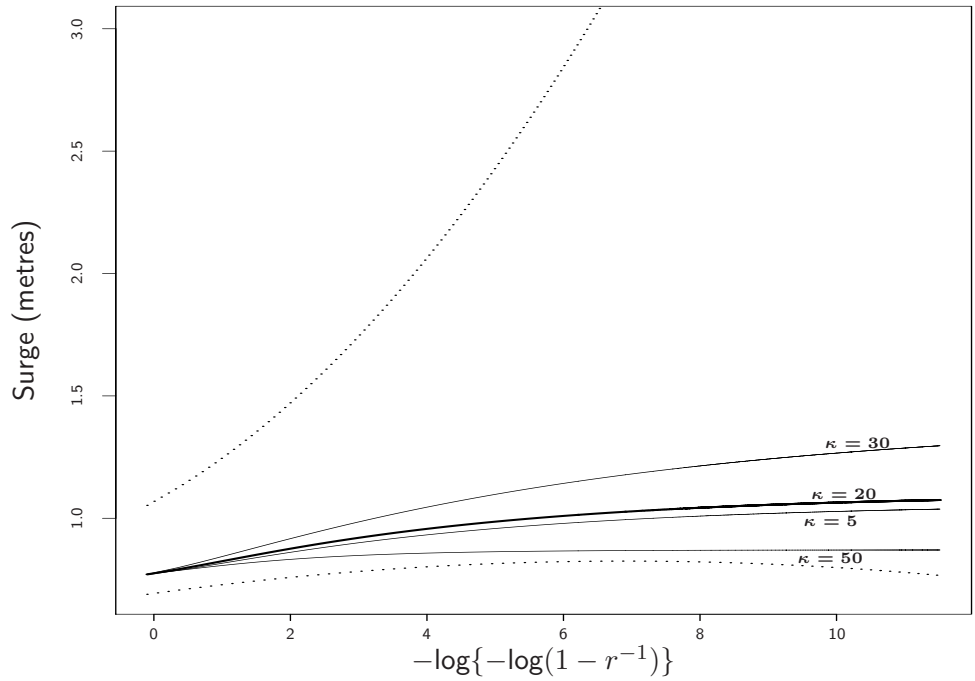

FIGURE 2: Bold line: return level curve for the Newlyn sea surge data, with associated 95% profile log–likelihood confidence bands (dotted lines). Also shown are return level curves obtained using three other values of cluster separation interval: $\kappa = 5$, 30 and 50.

a practitioner's point of view, this is worrying: designing a sea–wall, for example, to a height specified by POT analysis using $\kappa = 50$ could result in substantial under–protection. To avoid an over–complicated graph, the 95% confidence bands for the other values of κ used are not displayed, although the level of variability for each of these is similar to that shown for return levels obtained using $\kappa = 20$, with very high upper bounds for long–period return levels. A study of the Bradfield wind speed extremes reveals a similar sensitivity of return level estimation to the choice of κ.

3 Simulation study

In Section 2.5 we illustrated the difficulties facing return level estimation when attempting to use a filtered set of independent extremes. Firstly, estimates can be sensitive to the choice of cluster separation interval, κ, which is often chosen arbitrarily. Indeed, results can be sensitive to the arbitrary choices made in cluster determination more generally; there are other methods available for cluster identification, such as 'blocks declustering' (see, for example, Smith and Weissman, 1994), and so it is often the case that the declustering scheme, as well as the 'declustering parameter' within that scheme, are *both* chosen arbitrarily. Secondly, there is a wastage of information in discarding all data except the cluster maxima, often resulting in very wide confidence intervals for return levels.

Fawcett and Walshaw (2007) advocate the use of *all* threshold excesses when estimating GPD parameters and associated quantiles, initially ignoring serial correlation but inflating standard errors, accordingly, post–analysis. However, when it comes to estimating return

levels, temporal dependence cannot be ignored. In fact, as we will discuss, there is a dependence between return levels and the strength of correlation present between successive extremes, with return levels decreasing for increasing levels of temporal dependence.

In this Section, we use simulated data to investigate the apparent sensitivity of return level estimation to the choice of cluster separation interval, κ, used in POT analyses. For the same sets of simulated data, we also estimate return levels using *all* threshold excesses, properly accounting for the strength of extremal dependence present. We will: introduce the theory of bivariate extremes, and outline how this theory can be invoked to simulate Markov chains with first–order temporal dependence; give some design details for our simulation study; investigate the relationship between return levels and the strength of serial correlation present in the extremes of our simulated chains; present some results which highlight the main conclusions of our investigation.

3.1 Bivariate threshold excess models

Suppose $(x_1, y_1), \ldots, (x_n, y_n)$ are independent realisations of a random variable (X, Y) with joint distribution function F. For suitably large thresholds u_x and u_y, the marginal distributions of F each have an approximation of the form given by Equation (2), with respective parameter sets $(\lambda_{u_x}, \sigma_x, \xi_x)$ and $(\lambda_{u_y}, \sigma_y, \xi_y)$. After transforming the margins for (X, Y) to standard Fréchet (see Coles, 2001) it can be shown (Pickands, 1981) that the joint distribution function $G(x, y)$ for a bivariate extreme value distribution has the representation

$$G(x, y) = \exp\{-V(x, y)\}, \tag{4}$$

for $x > 0$, $y > 0$, where

$$V(x, y) = 2\int_0^1 \max(q/x, (1-q)/y) dH(q), \tag{5}$$

and H is a distribution function on $[0, 1]$ which satisfies the mean constraint

$$\int_0^1 q dH(q) = \frac{1}{2}.$$

There is no characterisation of the complete family of distributions specified by (4), and so model choice involves specifying an appropriate sub–family through the choice of H in (5).

There are various possibilities for the dependence function H; two commonly used *symmetric* models are the logistic and negative logistic models, where

$$V(x, y) = \left(x^{-1/\alpha} + y^{-1/\alpha}\right)^{\alpha}, \qquad 0 < \alpha \leq 1, \qquad \text{and} \tag{6}$$

$$V(x, y) = -x - y + \left(x^{-\rho} + y^{-\rho}\right)^{-1/\rho}, \qquad \rho > 0, \tag{7}$$

respectively. Independence is obtained when $\alpha = 1$ and $\rho \searrow 0$ for the logistic and negative logistic models respectively, while complete dependence is obtained when $\alpha \searrow 0$ and $\rho \to \infty$. The *asymmetric* bilogistic model has

$$V(x, y) = -x\gamma^{1-\alpha} - y(1-\gamma)^{1-\beta}, \qquad 0 < \alpha, \beta < 1, \tag{8}$$

where $\gamma = \gamma(x, y; \alpha, \beta)$ is the solution of $(1-\alpha)x(1-\gamma)^\beta = (1-\beta)y\gamma^\alpha$. When $\alpha = \beta$, this model reduces to the (symmetric) logistic model. The value $\alpha - \beta$ determines the extent of asymmetry in the dependence structure. Independence is obtained when $\alpha = \beta \to 1$, or when one of α or β is fixed and the other approaches 1. Different limits occur when one of α or β is fixed and the other approaches zero.

Replacing x and y in (6), (7) or (8) with x_i and x_{i+1}, respectively, puts the models for the dependence function H in a time series context; assuming a first–order Markov structure for our process gives

$$f(x_1, \ldots, x_n; \boldsymbol{\psi}) \;=\; \prod_{i=1}^{n} f(x_i, x_{i+1}; \boldsymbol{\psi}) \Big/ \prod_{i=2}^{n-1} f(x_i; \boldsymbol{\psi}) \tag{9}$$

as the joint density function for $x_1, \ldots, x_n$, where $\boldsymbol{\psi}$ is a specified parameter vector for a model for $f(x_i, x_{i+1}; \boldsymbol{\psi})$, $i = 1, \ldots, n-1$. A limiting model for $f(x_i; \boldsymbol{\psi})$ for the region (u, ∞) is the density of the GPD function given in Equation (1); differentiation of Equation (4), using the logistic, negative logistic or bilogistic models (for example), gives contributions to the numerator in (9) on the region $(u, \infty) \times (u, \infty)$. An example of a simulation procedure is an envelope rejection scheme, whereby x_1 is simulated from a standard Fréchet, then we simulate from the conditionals $x_i | x_{i-1}$ for $i = 2, \ldots, n$ (see Fawcett (2005) for full details of such a scheme as applied to the logistic model; Stephenson (2003) outlines various other (computationally quicker) methods).

3.2 Simulation study details

We simulate stationary first–order Markov chains of extreme value type according to the three models given by Equations (6), (7) and (8). Within each dependence model, we simulate chains with varying degrees of temporal dependence according to the parameter(s) for that model. Specifically: for the logistic model (6), we use $\alpha = 0.10, 0.11, \ldots, 1$ to cover almost the entire range of serial correlation, from very strong dependence to independence; for the negative logistic model (7), we use $\rho = 0.10, 0.15, \ldots, 1, 1.1, 1.2, \ldots, 7.9, 8$; and for the bilogistic model (8), we fix α at 0.6 and use $\beta = 0.10, 0.11, \ldots, 0.99$. The marginals of these simulated chains are transformed from standard Fréchet to GPD to assess the effects of declustering on return level estimation. We use five different values of GPD shape parameter ξ to transform the chains: -0.4, -0.1, 0, 0.3 and 0.8, to reflect the various tails which might be observed in a real–life dataset; the GPD scale parameter σ is held unit constant. We set u at the 95% marginal quantile for different values of ξ, giving $\sigma^* = 1 + \xi u^\dagger$.

For all three dependence models used, we simulate N series of length n for each strength of temporal dependence used (i.e. for each value of α, ρ and β for the logistic, negative logistic and bilogistic models respectively). For each series, the GPD is fitted, via maximum likelihood, to the set of cluster peak excesses identified through runs declustering. Equation (3) is then used on each triple of estimates $(\hat{\lambda}_u^{(i)}, \hat{\sigma}^{*(i)}, \hat{\xi}^{(i)})$ to obtain estimates of return levels $\hat{z}_r^{(i)}$ at each iteration $i = 1, \ldots, N$. In line with the Newlyn sea–surge data, we use $n_y = 365.25 \times (24/3) = 2922$. To assess the sensitivity of return level estimates on the cluster identification procedure, we use a range of cluster separation intervals in the runs declustering scheme: $\kappa = 5$, 20, 30, 50 and 60. We simulate chains of length $n = 10,000$ at each iteration $i = 1, \ldots, 1000$.

3.3 Relationship between return levels and extremal dependence

The procedure for estimating return levels from the simulated chains, as outlined above, can only be applied to i.i.d. threshold exceedances. This is fine if we have used a declustering procedure to identify an independent set of extremes. However, we would like to compare return level estimates from standard POT analyses to those obtained from an analysis using *all* extremes and, as already discussed, this requires careful consideration of the dependence of return levels on the strength of serial correlation present in the series.

3.3.1 The extremal index

Let $X'_1, \ldots, X'_n$ be the first n observations of a stationary series satisfying Leadbetter's $D(u_n)$ condition (a mixing condition which ensures long range dependence is negligible; see Leadbetter *et al.*, 1983, for more details), and let $M'_n = \max\{X'_1, \ldots, X'_n\}$. Now let $X_1, \ldots, X_n$ be an independent series, with X having the same distribution as X', and let $M_n = \max\{X_1, \ldots, X_n\}$. Then standard arguments in Leadbetter *et al.* (1983, Ch. 3) give

$$\Pr\{M'_n \leq x\} \approx G^{n\theta}(x), \tag{10}$$

where G is the marginal tail of X given by the GPD in Equation (2), and $\theta \in (0, 1)$ is known as the *extremal index*, the concept of which was developed in a series of papers including Newell (1964), Loynes (1965), O'Brien (1974) and Leadbetter *et al.* (1983), with a review by Leadbetter and Rootzén (1988). As $\theta \to 0$ there is increasing dependence in the extremes of the process; for an independent process, $\theta = 1$. The value z_r of x which equates (10) to $1 - r^{-1}$ (and where $n = n_y$, the number of observations per year), is the r–year return level of the process. Thus, in the presence of short–term temporal dependence, the value z_r will be related to the strength of that dependence defined through θ. When using the set of independent cluster peak excesses, $\theta = 1$ and so the solution of (10), on equating to $1 - r^{-1}$, reduces to Equation (3).

Figure 3 shows how the 50– and 200–year return level can vary with θ for one arm of the simulation study. Here, $\xi = -0.4$; setting the threshold at the 95% marginal quantile, i.e. using $\lambda_u = 0.05$, gives $u = 1.746$ and $\sigma^* = 0.302$ (recall that σ is held unit constant). Equation (10) (using the GPD in Equation (2) for G and $n = n_y = 2922$ in line with the Newlyn sea–surge data) has been set equal to $1 - 1/50$ and solved for x for $\theta = 0.01, 0.02, \ldots, 1$.

3.3.2 Connection with bivariate threshold excess models

Define x_m such that $G^m(x_m) = 1/2$. Then, using Equation (10), we can define

$$\theta_m \;=\; -\frac{\log \Pr\left(\max\{X'_1, \ldots, X'_m\} \leq x_m\right)}{\log 2}, \tag{11}$$

and so $\theta_m \to \theta$ as $m \to \infty$. We can use (11) to investigate the relationship between the extremal index θ, and the dependence parameter(s) in a bivariate threshold excess model, via simulation. The extremal index is deterministically related to the dependence parameter(s) for any given model for dependence – the need for simulation arises because the relationship is analytically intractable. For example, if the logistic model (6) is assumed for the joint distribution of consecutive pairs, we can simulate M first–order Markov chains each of length

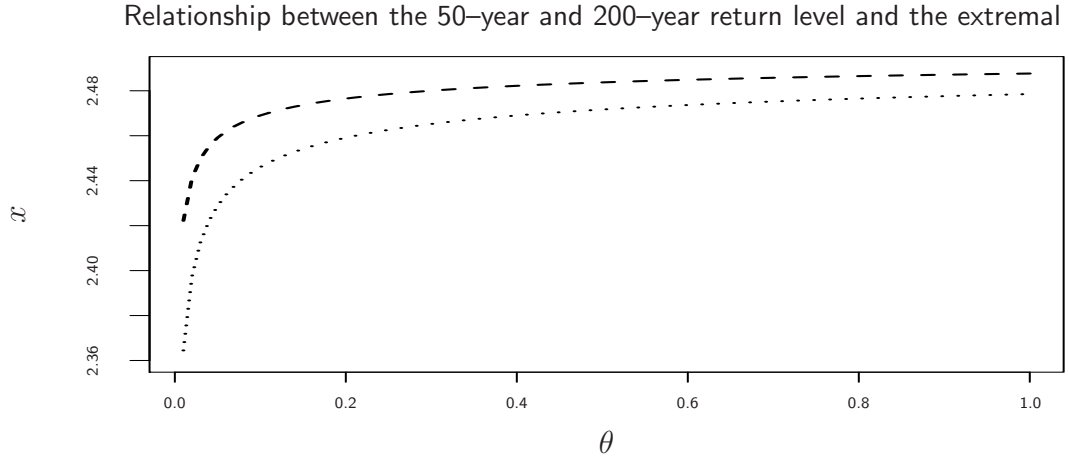

FIGURE 3: Plot of return level against extremal index θ: the dotted line corresponds to the 50–year return level and the dashed line corresponds to the 200–year return level.

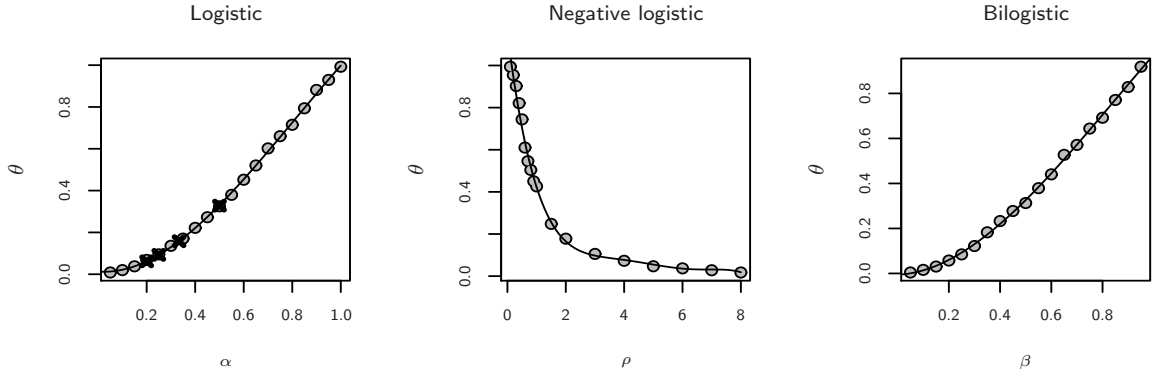

FIGURE 4: Simulated values of the extremal index θ for α (logistic), ρ (negative logistic) and β (bilogistic with $\alpha = 0.6$). The solid lines correspond to fitted polynomials: $\theta = 0.013 - 0.092\alpha + 1.833\alpha^2 - 0.756\alpha^3$; $\theta = 1.153 - 1.107\rho + 0.463\rho^2 - 0.096\rho^3 + 0.010\rho^4 - 0.0004\rho^5$; $\theta = -0.005 + 0.045\beta + 1.539\beta^2 - 0.607\beta^3$. The crosses in the first plot show limiting values of θ for the some values of α in the logistic model, as derived in Smith (1992).

m with dependence parameter $\alpha \in (0, 1)$ controlling the strength of serial correlation; then the probability in the numerator of Equation (11) can be estimated as the proportion of simulated chains whose maximum value does not exceed x_m. The left–hand plot in Figure 4 shows the results of such simulations for $\alpha = 0.05, 0.10, \ldots, 1$ in the logistic model. Here, we use $M = m = 10,000$. The dashed line shows the fit from a cubic regression of θ on α; substituting $\alpha = \frac{1}{2}, \frac{1}{3}, \frac{1}{4}$ and $\frac{1}{5}$ into this cubic gives $\theta \approx 0.331$, 0.158, 0.0925 and 0.0616 (respectively), which match the limiting values for θ reported in Smith (1992) almost perfectly. Smith uses a numerical procedure to find these limiting values; this procedure is computationally expensive, and the results differ negligibly to the fitted values obtained directly from our cubic. Similar polynomial relationships between the extremal index and the dependence parameter(s) for the negative logistic and bilogistic models are obtained in the same way (see the middle and right–hand plots in Figure 4, respectively).

3.3.3 Consequences for simulation study

The dependence of return levels on the strength of serial correlation, as illustrated by Figure 3, will be used in two ways in the simulation study. Firstly, we intend to compare the sampling distributions $\hat{z}_r^{(i)}$, $i = 1, \ldots, 1000$, obtained from POT analyses (see Section 3.2) to those obtained from analyses making use of *all* threshold excesses. To account for temporal dependence in the latter, at each iteration in the simulation study we now use Equation (10) instead of Equation (3) to cater for the dependence on serial correlation through θ. Thus, Equation (10) is now solved for $x = \hat{z}_r^{(i)}$, $i = 1, \ldots, 1000$, and the GPD parameter estimates $(\hat{\lambda}_u^{(i)}, \hat{\sigma}^{*(i)}, \hat{\xi}^{(i)})$ found by using *all* threshold excesses are used in G; as before, we use $n = n_y = 2922$. When the logistic model is used for the temporal structure, we obtain the extremal index θ for each value of α used $(0.10, 0.11, \ldots, 1)$ to simulate the chain via the cubic relationship shown in Figure 4 (left); the relationships between θ and the dependence parameter(s) in the negative logistic and bilogistic models (Figure 4 (middle) and Figure 4 (right), respectively), are exploited in the same way.

Secondly, the sampling distributions for z_r need to be compared to a true value that has taken the dependence on serial correlation into account. For example, when using the logistic model in our simulations, we label this true value $z_r(\alpha)$. For each value of α used to simulate a series, the corresponding extremal index θ is obtained via the cubic relationship shown in Figure 4 (left) – this is substituted into Equation (10), which is then set equal to $1 - r^{-1}$ and solved for $x = z_r(\alpha)$, where G is now the GPD distribution function with *known* parameters $(\lambda_u = 0.05, \sigma^* = 1 + \xi u, \xi; \xi = -0.4, -0.1, 0, 0.3, 0.8)$.

3.4 Results

Plots down the left–hand–side of Figure 5 display means of the sampling distributions $\hat{z}_{50}^{(i)}$, $i = 1, \ldots, 1000$, for the 50–year return level. In all three plots, $\xi = -0.4$. Using all threshold excesses, properly accounting for extremal dependence, seems to give (on average) fairly accurate estimates of the 50–year return level with the heavy lines following the dashed lines quite closely. The POT analyses using $\kappa = 5$ also perform reasonably well; however, using $\kappa = 20$, 30, 50 or 60 seems to result in the 50–year return level being under–estimated across all levels of serial correlation. Similar results (though not shown here) were obtained for $\xi = -0.1$ and $\xi = 0$; for $\xi = 0.3$ and, particularly, $\xi = 0.8$, all analyses produced sampling distribution means for z_{50} that appeared substantially higher than the true values, probably owing to the fact that the GPD is unbounded for $\xi > 0$. Using the sampling distribution median produces more well–behaved plots, similar to those shown here, for $\xi = 0.3$.

Dotted vertical lines have been drawn at $\alpha = 0.577$, $\rho = 1.022$ and $\beta = 0.544$ in the plots down the left–hand–side in Figure 5, which correspond to the fitted values for these dependence parameters when the logistic, negative logistic and bilogistic[‡] models are applied to consecutive extremes in the Newlyn dataset (we will return to the viability of these models for the Newlyn sea–surge extremes in Section 4.2.1). Focussing on these particular levels of dependence, the plots down the right–hand–side of Figure 5 show means of the sampling distribution $\hat{z}_r^{(i)}$ for a range of return periods r, given on the usual scale for such plots.

[‡]Fitting the bilogistic model gives $(\hat{\alpha}, \hat{\beta}) = (0.608, 0.544)$; the fitted value for α is close to the fixed value used in the simulation study

Vertical lines have been drawn at $r = 50$ for direct correspondence to the plots down the left–hand–side. These plots show that extrapolating beyond 50 years gives similar findings to the plots for the 50–year return level, with analyses using all threshold excesses, having properly accounted for dependence, performing best. Return level estimates are shown to be sensitive to the choice of κ when using cluster peaks, with $\kappa = 20$, 30, 50 and 60 giving estimated return levels that fall substantially short of the true values. As can be seen, this is observed for all three dependence models used in the simulation study.

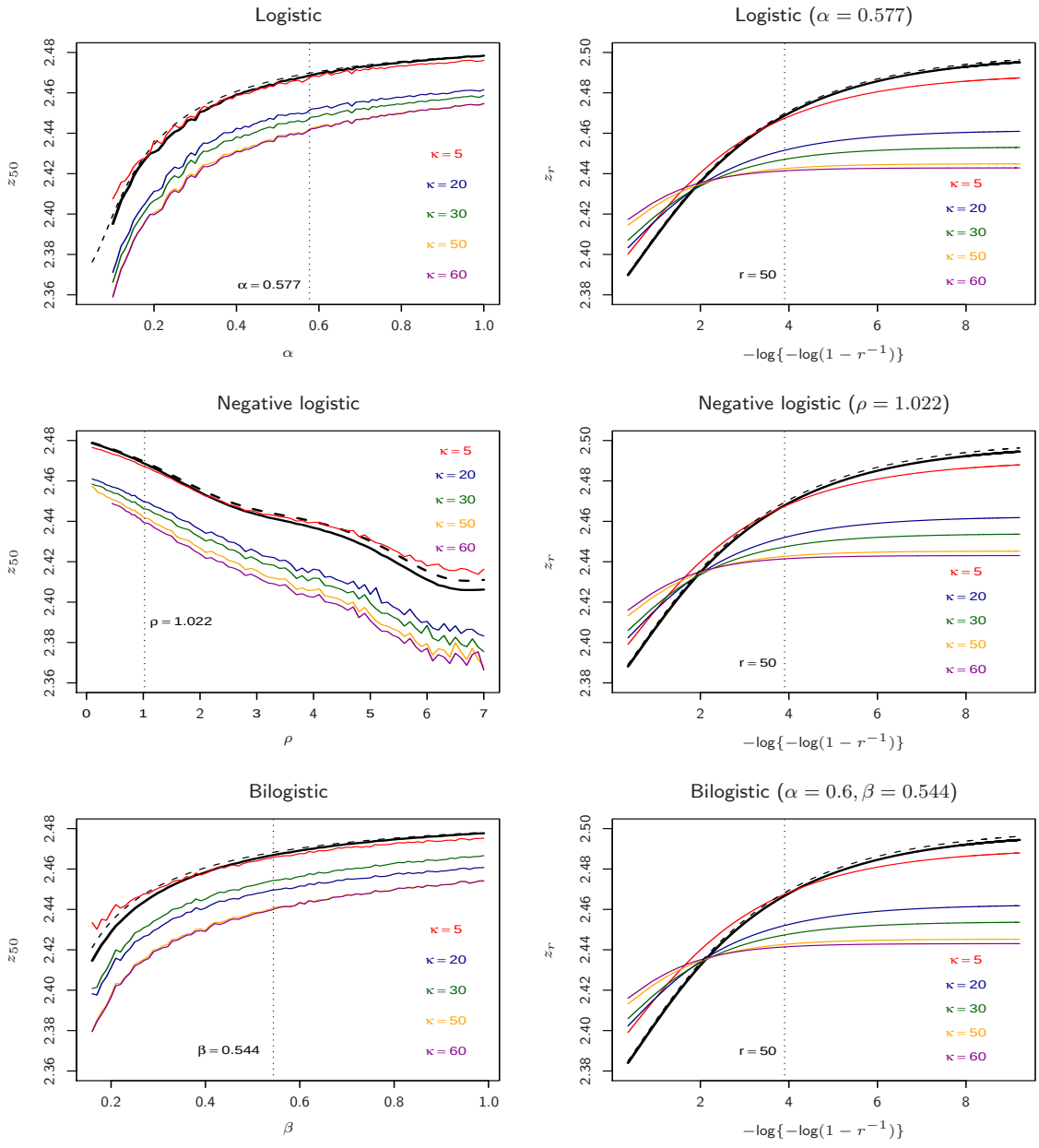

Figure 5: Plots of sampling distribution means for: the 50–year return level, for different strengths of temporal dependence (left); r–year return levels for fixed strengths of temporal dependence (right). Successive extremes have been simulated from the logistic (top), negative logistic (middle) and bilogistic (bottom) models. The curved dashed lines represent the true return levels; the bold lines indicate estimated return levels using all threshold excesses; the thin lines indicate the use of cluster peak excesses using cluster separation intervals κ.

3.5 Discussion

In this Section we have shown, through simulations, that return levels obtained as a result of a standard POT analysis can fall short of the true return level. In the results shown in Figure 5, POT analyses using $\kappa = 5$ seem to perform well, but it is often the case that the cluster separation interval is chosen arbitrarily – all other values of κ used give under–estimated return levels for return periods commonly of interest, and this under–estimation seems to get worse as r increases (see Figure 5, right–hand–side). Both the sensitivity of return level estimates to the choice of κ, and the lack of accuracy of estimates for some values of κ, cast doubt on the reliability of POT analyses for obtaining return levels. We have shown this to be the case for different types of temporal structure in the extremes (both symmetric and asymmetric), as well as for a range of tail behaviours. In practice, using all threshold exceedances would avoid the need to decluster, could give more accurate estimates of return levels, as well as increase precision owing to the inclusion of more data.

However, using all threshold excesses requires the inclusion of the extremal index θ in the estimation procedure for return levels. In real–life applications, we might no longer be able to rely on the relationship between the extremal index and the parameter(s) of a dependence model for extremes, since the precise form of temporal structure might no longer known. Indeed, in this paper we have examined only three models for the dependence structure between successive extremes – there are many more, and basing our estimate of θ on a particular model would require us to first check the appropriateness of that model. Where models are nested (e.g. logistic within the bilogistic), likelihood ratio tests can be performed to choose between models; otherwise, more informal *ad hoc* procedures have to be used. Though this can be done (e.g. for the Bradfield wind speed data in Fawcett and Walshaw (2006)), this can be rather subjective. We now turn our attention to the sensitivity of return level estimation to different methods for *estimating* the extremal index when attempting to use all threshold excesses.

4 Extremal index estimation

In this Section, we investigate the sensitivity of return level estimation to the choice of extremal index estimator. We consider five methods for estimating θ:

1. One method for estimating the extremal index is to fit a bivariate extreme value model (e.g. the logistic, negative logistic or bilogistic) to successive pairs in our series; then we can use the polynomial relationships shown in Figure 4 to obtain our estimate of θ. We call the extremal index estimator obtained in this way a *polynomial estimator*, and we label it $\hat{\theta}^{[1]}$ (with subscripts $_{\log}$, $_{\text{neglog}}$ or $_{\text{bilog}}$ for each of the three dependence models considered in this paper).

2. Section 3.3.1 gives a formal definition of the extremal index; an alternative characterisation, provided by Hsing *et al.* (1988), is that θ^{-1} is the limiting mean cluster size in the point process of exceedance times over a high threshold. This suggests that a suitable way to estimate the extremal index can be found through methods which identify clusters of extremes, the estimate itself being found as the reciprocal of the mean cluster size. We call such an estimator a *cluster size method*; when runs declustering is used to identify clusters of extremes, we label this estimator $\hat{\theta}^{[2]}$.

3. Another cluster size method uses 'blocks declustering' to identify clusters of extremes, and again estimates θ as the reciprocal of the mean cluster size. This method of cluster identification partitions the data into approximately k blocks of length τ and the threshold exceedances within each block are treated as a single cluster of extremes. We label this estimator $\hat{\theta}^{[3]}$.

4. Gomes (1993) proposes a method for estimating θ based on separate model fits to the two sets of block maxima $M_{\tau,i}$ and $M'_{\tau,i}$, $i = i, \ldots, k$, where $M'_{\tau,i}$ are the block maxima from a stationary series and $M_{\tau,i}$ are the block maxima from an *independent* series with the same marginal distribution as the $M'_{\tau,i}$, having been obtained after randomising the index of the original observations. The Generalised Extreme Value distribution (GEV) is used to model both sets of block maxima, with distribution function

$$G(x; \mu, \varsigma, \xi) \;=\; \exp\left\{ -\left[1 + \xi\left(\frac{x-\mu}{\varsigma}\right)\right]_+^{-1/\xi} \right\},$$

where μ, ς and ξ are location, scale and shape parameters respectively; the shape parameter ξ in the GEV takes exactly the same value as that for the corresponding GPD in Equation (1), and the scale parameter $\varsigma = \sigma - \xi(u - \mu)$ where σ is the GPD scale parameter. If the extremal index for the block maxima from the stationary series $M'_{\tau,i}$ is equal to θ then, from Equation (10), the $M'_{\tau,i}$ have distribution function $G^{\theta}(x; \mu, \varsigma, \xi)$, which is easily seen to be a GEV distribution $G(x; \mu_\theta, \varsigma_\theta, \xi_\theta)$, where

$$\begin{aligned} \mu_\theta &= \mu - \varsigma(1 - \theta^{\xi})/\xi, \\ \varsigma_\theta &= \varsigma\theta^{\xi} \qquad \text{and} \\ \xi_\theta &= \xi. \end{aligned}$$

Gomes (1993) suggests estimating θ from estimates $(\hat{\mu}, \hat{\varsigma}, \hat{\xi})$ and $(\hat{\mu}_\theta, \hat{\varsigma}_\theta, \hat{\xi}_\theta)$ obtained from the separate fits to the two sets of block maxima. A pooled estimate of ξ is calculated as

$$\tilde{\xi} \;=\; \frac{\hat{\varsigma} - \hat{\varsigma}_\theta}{\hat{\mu} - \hat{\mu}_\theta}.$$

Then an estimate of θ, which we label $\hat{\theta}^{[4]}$, is given by

$$\hat{\theta}^{[4]} \;=\; \left(\frac{\hat{\varsigma}}{\hat{\varsigma}_\theta}\right)^{-1/\tilde{\xi}}. \tag{12}$$

We call this approach the *maxima method*.

5. We also consider an *intervals estimator* for θ based on the inter–arrival times of threshold exceedances, proposed by Ferro and Segers (2003). Suppose N exceedance times are observed: $S_1 < S_2 < \cdots < S_N$, and $T_i = S_{i+1} - S_i$ for for $i = 1, \ldots, N-1$ are the inter–arrival times. Ferro and Segers (2003) derive the distribution of the T_i and show that if the second moment of this distribution is equated to the empirical second moment, replacing θ with $\hat{\theta}^{[5]}$ and solving for $\hat{\theta}^{[5]}$ gives

$$\hat{\theta}^{[5]} \;=\; \frac{2\left\{\sum_{i=1}^{N-1}(T_i - 1)\right\}^2}{(N-1)\sum_{i=1}^{N-1}(T_i - 1)(T_i - 2)}.$$

4.1 Simulated data

We now return to the analyses using all threshold exceedances in the simulation study of Section 3. At each iteration i, $i = 1, \ldots, 1000$, as well as obtaining $(\hat{\lambda}_u^{(i)}, \hat{\sigma}^{*(i)}, \hat{\xi}^{(i)})$ for each set of threshold excesses, we also estimate the extremal index, giving $\hat{\theta}^{[1]^{(i)}}, \ldots, \hat{\theta}^{[5]^{(i)}}$. Each set of estimates $(\hat{\lambda}_u^{(i)}, \hat{\sigma}^{*(i)}, \hat{\xi}^{(i)}, \hat{\theta}^{[1]^{(i)}}), \ldots, (\hat{\lambda}_u^{(i)}, \hat{\sigma}^{*(i)}, \hat{\xi}^{(i)}, \hat{\theta}^{[5]^{(i)}})$ is then used to estimate the r–year return level by equating the right–hand–side of (10) to $1 - r^{-1}$ and solving for $x = z_r$. As before, we consider the logistic, negative logistic and bilogistic models for consecutive extremes in our simulations, using the polynomial relationships shown in Figure 4 to obtain $\hat{\theta}^{[1]}_{\text{log}}$, $\hat{\theta}^{[1]}_{\text{neglog}}$ and $\hat{\theta}^{[1]}_{\text{bilog}}$ for $\hat{\theta}^{[1]}$ (respectively). We also use the same values of ξ as before.

The left–hand–side of Figure 6 shows sampling distribution means for each of the five extremal index estimators, for the different strengths of temporal dependence. As might be expected, the polynomial estimators $\hat{\theta}^{[1]}_{\text{log}}$, $\hat{\theta}^{[1]}_{\text{neglog}}$ and $\hat{\theta}^{[1]}_{\text{bilog}}$ perform very well. The only source of error here is the estimation of the dependence parameters – once we have the maximum likelihood estimate for α, ρ or β, we use the fitted polynomials themselves to find the corresponding estimate of θ. A first–order Markov structure is assumed with dependence structure given by Equations (6), (7) or (8) (top, middle and bottom, respectively), which is exactly how the simulated data have been constructed! Of course, as we shall see in the next Section, we might not expect $\hat{\theta}^{[1]}$ to perform quite so well in practice since, with real–life data, there will be deviations from the ideal: both the model, and the model–order, might be inappropriate for the data being analysed.

The performance of both cluster size methods $\hat{\theta}^{[2]}$ and $\hat{\theta}^{[3]}$ is variable, with substantial deviations from the fitted polynomials for some levels of dependence in all three plots. Here, the cluster separation interval was set at $\kappa = 20$ observations for $\hat{\theta}^{[2]}$; the block length for $\hat{\theta}^{[3]}$ was set at $\tau = 100$ so that $k = \tau$. In practice, $\hat{\theta}^{[2]}$ and $\hat{\theta}^{[3]}$ can be sensitive to the choice of κ/τ respectively (see Fawcett and Walshaw, 2008); without good physical grounds leading to an 'optimal' choice for κ or τ, estimating θ in this way can be prone to severe bias. One of the advantages of using all threshold exceedances as opposed to working with cluster *peak* exceedances is that we can completely avoid having to decluster – estimating θ using a cluster size method once again introduces the problem of declustering.

Both the maxima method ($\hat{\theta}^{[4]}$) and the intervals estimator ($\hat{\theta}^{[5]}$) perform very well, showing little deviation from the fitted polynomials in all three dependence models. This is promising. In particular, the intervals estimator does not rely on any choice of declustering parameter (as do the cluster size methods); neither does it rely on any assumptions regarding the form of dependence structure in the extremes (as do the polynomial estimators). This could make $\hat{\theta}^{[5]}$ a candidate for use with real–life data.

The right–hand–side of Figure 6 shows the various extremal index estimators applied to return level estimation. As might be expected, those estimators that perform well then give the the best estimates of the 50–year return level. The conclusion here is that, if we can find a suitable extremal index estimator for real–life data (perhaps $\hat{\theta}^{[5]}$ might be the most robust) then, as far as return level estimation is concerned, we should analyse *all* threshold excesses: doing so avoids the complications of declustering and will lend greater precision to our inferences owing to the inclusion of more data.

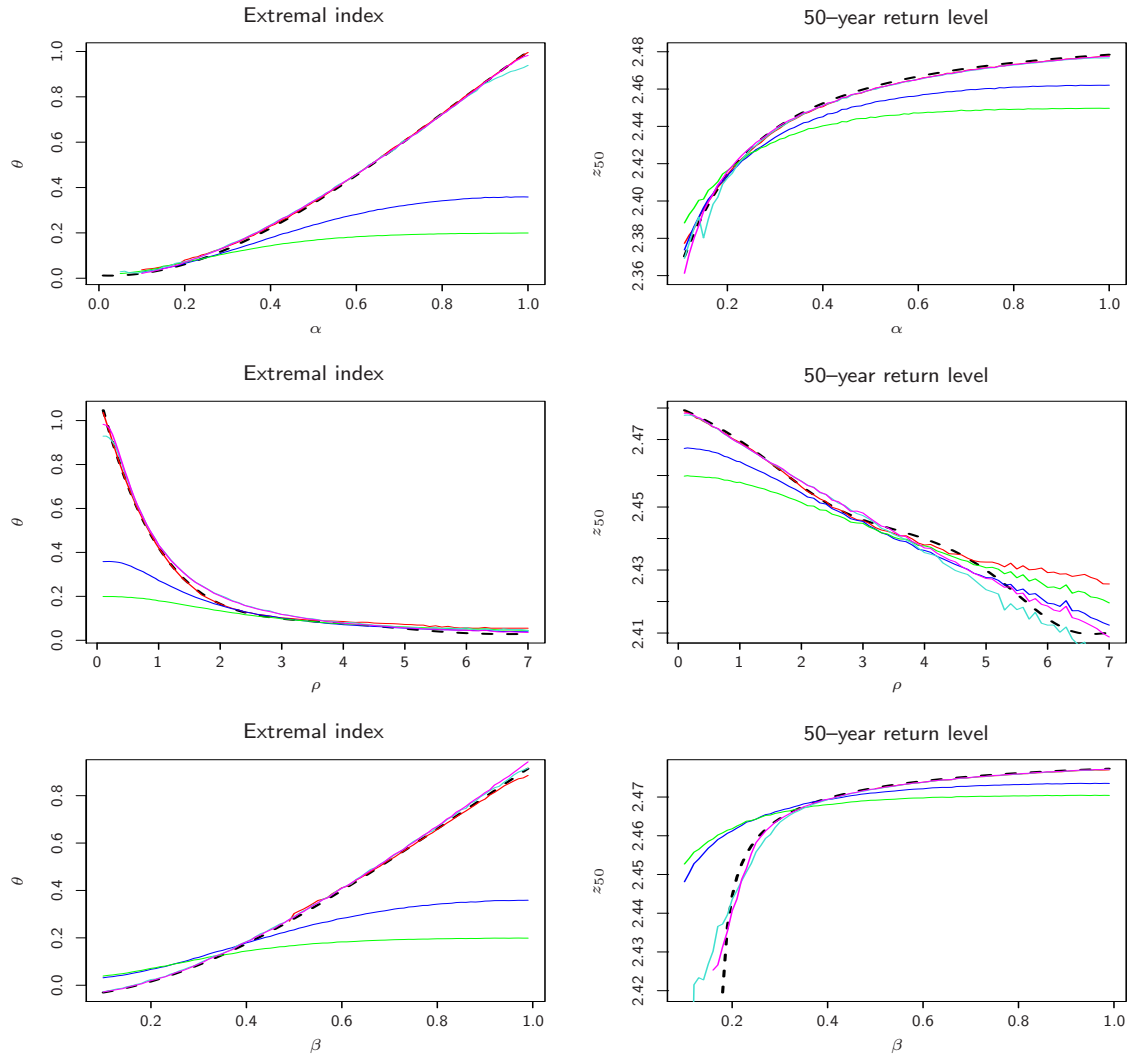

Figure 6: Sampling distribution means for estimates of the extremal index θ (left) and the 50–year return level (right) for a range of strengths of temporal dependence for the logistic model (top), the negative logistic model (middle) and the bilogistic model (bottom). Colours indicate the use of the: **polynomial estimator** $\hat{\theta}^{[1]}$; **cluster size estimator** $\hat{\theta}^{[2]}$; **cluster size estimator** $\hat{\theta}^{[3]}$; **maxima method** $\hat{\theta}^{[4]}$; **intervals estimator** $\hat{\theta}^{[5]}$. The broken lines indicate the true values (i.e. obtained using the fitted polynomials shown in Figure 4).

4.2 Data applications: sea–surges and wind speeds

4.2.1 Newlyn sea–surges

Figure 2 shows the sensitivity of return level estimation to the choice of cluster separation interval κ when applying the POT method to the Newlyn sea–surge data. We now estimate return levels using *all* excesses above a threshold of $u = 0.3$ metres, accounting for the dependence on serial correlation by estimating the extremal index. Table 1 shows estimates of the extremal index for the Newlyn sea–surge data using the five estimators $\hat{\theta}^{[1]}, \ldots, \hat{\theta}^{[5]}$. Both $\hat{\theta}^{[1]}$ and $\hat{\theta}^{[4]}$ are functions of other parameters, and so we have used the delta method (Coles, 2001, Ch. 2) to obtain estimated standard errors here. The standard error for $\hat{\theta}^{[5]}$ has been estimated using a bootstrap procedure; here, the inter–arrival times T_i between

successive exceedances of u were resampled with replacement B times, yielding a collection of estimates $\{\hat{\theta}^{[5]}_{(1)}, \ldots, \hat{\theta}^{[5]}_{(B)}\}$ for which the standard deviation was calculated. This procedure was repeated for increasing B, which showed convergence to about 0.040. The sensitivity of extremal index estimation to the choice of estimation procedure is clear, with estimates ranging from 0.106 to 0.425.

The simulation study revealed that both cluster size estimators ($\hat{\theta}^{[2]}$ and $\hat{\theta}^{[3]}$) perform poorly at some strengths of temporal dependence, which casts doubt on their reliability here. Fawcett and Walshaw (2008) also show that both these methods are sensitive to the choice of κ or τ (for $\hat{\theta}^{[2]}$ and $\hat{\theta}^{[3]}$ respectively) and, as already mentioned, we would rather avoid having to decluster altogether when using all threshold excesses. Having been obtained completely empirically, we do not have standard errors for the cluster size estimators.

Using a polynomial estimator ($\hat{\theta}^{[1]}$) relies on us being able to confirm a first–order Markov structure for the sea–surge extremes, as well as the suitability of the chosen model (e.g. logistic, negative logistic or bilogistic). Using a diagnostic plot to assess the validity of a first–order temporal dependence, relative to higher–order dependencies (see the "simplex plots" in Coles and Tawn, 1991), reveals that a second–order dependence structure might be more plausible for the Newlyn extremes. For the logistic and bilogistic models, referring $2\{\ell(\hat{\alpha}, \hat{\beta}; \text{bilogistic}) - \ell(\hat{\alpha}; \text{logistic})\}$ to χ^2_1 tables shows no significant improvement by including the asymmetry parameter β. However, comparing other symmetric models to the logistic model (e.g. the negative logistic) is not so straightforward and would rely on various *ad hoc* checks such as those used in Fawcett and Walshaw (2006). Although the logistic model might be appropriate here, the assumption of a first–order temporal structure is questionable, possibly casting doubt on the reliability of any of our polynomial estimators.

Figure 6 showed that, regardless of the structure of temporal dependence, both the maxima method ($\hat{\theta}^{[4]}$) and the intervals estimator ($\hat{\theta}^{[5]}$) were robust across all strengths of serial correlation. The similarity of both $\hat{\theta}^{[4]}$ and $\hat{\theta}^{[5]}$ (notwithstanding the large standard error for the former) might lend support to either of these being used to obtain estimates of return levels. Of the two, the intervals estimator is probably preferable – unlike the maxima method, which requires a block length τ to be chosen, $\hat{\theta}^{[5]}$ is completely automatic.

The five extremal index estimators $\hat{\theta}^{[1]}, \ldots, \hat{\theta}^{[5]}$, along with the GPD parameter estimates for all threshold excesses, were then used to estimate return levels by equating the right–hand–side of (10) to $1 - r^{-1}$ and solving for $x = z_r$. Table 1 shows some numerical results for $r = 10$, 50 and 1000, along with estimates from a POT analysis using $\kappa = 20$ observations (corresponding to the bold line in Figure 2). Though we can see that, when attempting to make use of all extremes, return level estimates are sensitive to the choice of extremal index estimator, we have already discussed that we might prefer to use $\hat{\theta}^{[5]}$. The simulations show that using all excesses, having appropriately accounted for the dependence on serial correlation, gives more accurate return level estimates than those from a standard POT analysis; we also see an increase in precision here, with smaller standard errors attached to estimates for the periods considered owing to the inclusion of more data (e.g. the standard error for $\hat{z}_{10}$ using cluster peaks is almost twice that from the analysis using all excesses with $\hat{\theta}^{[5]}$). By using all excesses with $\hat{\theta}^{[5]}$, we completely avoid any arbitrary choice of cluster identification procedure giving rise to return level estimation sensitivity as shown in Figure 2.

	$\hat{\theta}$	$\hat{z}_{10}$	$\hat{z}_{50}$	$\hat{z}_{1000}$
All excesses				
using $\hat{\theta}^{[1]}_{\text{log}}$	0.425 (0.045)	0.817 (0.067)	0.903 (0.104)	1.034 (0.178)
using $\hat{\theta}^{[1]}_{\text{neglog}}$	0.413 (0.037)	0.816 (0.072)	0.902 (0.106)	1.033 (0.178)
using $\hat{\theta}^{[1]}_{\text{bilog}}$	0.377 (0.020)	0.810 (0.070)	0.897 (0.104)	1.029 (0.175)
using $\hat{\theta}^{[2]}$	0.182	0.767 (0.091)	0.860 (0.127)	1.000 (0.201)
using $\hat{\theta}^{[3]}$	0.106	0.732 (0.091)	0.830 (0.127)	0.978 (0.201)
using $\hat{\theta}^{[4]}$	0.282 (0.206)	0.793 (0.078)	0.883 (0.131)	1.024 (0.171)
using $\hat{\theta}^{[5]}$	0.223 (0.040)	0.779 (0.052)	0.870 (0.093)	1.018 (0.163)
Cluster peak excesses	—	0.868 (0.106)	0.920 (0.144)	0.975 (0.202)

TABLE 1: Maximum likelihood estimates for the extremal index and three return levels for the Newlyn sea–surge data (units are in metres). Standard errors, where available, are given in parentheses.

4.2.2 Bradfield wind speeds

An extensive study of the wind speeds observed at Bradfield (Figure 1, right–hand–side) in Fawcett and Walshaw (2006) suggests that a first–order Markov structure, with logistic dependence, is a suitable assumption for the temporal evolution of the process. In fact, inferences for return levels barely changed when a first–order assumption was relaxed to allow higher–order temporal dependencies, and it was concluded that the added complexity (and associated computational expense) of fitting higher–order models was not worthwhile. Model–based estimates of various quantities, including the *mean storm length* (average length of a cluster) and the *mean inter–storm duration* (average time between clusters), compared quite well to their empirical counterparts when assuming a first–order Markov structure with logistic dependence. Thus, we might trust our polynomial estimator of the extremal index here ($\hat{\theta}^{[1]}_{\text{log}}$) to obtain estimates of return levels.

As discussed in Section 2.3, a piecewise seasonality approach, whereby a different GPD is fitted to sets of excesses over monthly varying thresholds, can be adopted to circumvent the problem of seasonal variability in the Bradfield wind speed data. The monthly varying GPD parameters can be recombined to estimate overall return levels by considering the *annual* exceedance rate of x, given by

$$\sum_{j=1}^{12} \left\{1 - G_j(x)^{n_j \theta_j}\right\}, \qquad j = 1, \ldots, 12, \tag{13}$$

where $\{1 - G_j(x)^{n_j\theta_j}\}$ is the annual exceedance rate of x in month j, G_j is the GPD distribution function in month j with parameters $(\lambda_{u_j}, \sigma_j, \xi_j)$, n_j is the number of observations in month j and θ_j is the extremal index in month j. Equating (13) to r^{-1} and solving for $x = z_r$ gives the overall r–year return level for the process. Table 2 shows return level estimates from a standard POT analysis (where *reclustered excess plots* (Walshaw, 1994) have been used to simultaneously choose monthly varying thresholds and a cluster separation interval, u_j and κ, respectively) along with estimates obtained using all threshold exceedances – this time using only $\hat{\theta}^{[1]}_{\text{log},j}$ and $\hat{\theta}^{[5]}_j$.

	$\hat{z}_{10}$	$\hat{z}_{50}$	$\hat{z}_{1000}$
All excesses			
using $\hat{\theta}^{[1]}_{\log,j}$	88.611 (5.519)	96.093 (9.964)	107.645 (22.432)
using $\hat{\theta}^{[5]}_{j}$	87.054 (5.604)	94.908 (8.615)	107.004 (19.665)
Cluster peak excesses	96.556 (13.527)	102.537 (22.776)	107.143 (43.052)

TABLE 2: Maximum likelihood estimates for three return levels for the Bradfield wind speed data (units are in knots). Standard errors are given in parentheses.

Some discrepancies in point estimates for z_r are noted between the approaches which use all threshold excesses and the POT analysis (for the 10– and 50–year return periods). Indeed, results from the simulation study suggest we might rather trust the analyses using all exceedances, and the fact that there is close agreement between estimates obtained using $\hat{\theta}^{[1]}_{\log,j}$ and $\hat{\theta}^{[5]}_{j}$ lends support to this. Although results are not shown here, return level estimates from a POT analysis are sensitive to the choice of cluster separation interval κ, with estimates of the 50–year return level varying between 95.303 and 106.100 when κ varies between 5 and 50. Avoiding declustering, and pressing all extremes into use, obviously increases the precision of our estimates; as with the sea–surge analyses, we see much smaller standard errors attached to return level estimates obtained using all threshold excesses than we do for those found under the standard POT approach.

References

Coles, S.G. (2001). *An introduction to statistical modeling of extreme values.* Springer, London.

Coles, S.G. and Tawn, J.A. (1991). Modelling Extreme Multivariate Events. *J. R. Statist. Soc., B*, **53**, pp. 377—392.

Davison, A.C. and Smith, R.L. (1990). Models for Exceedances over High Thresholds (with discussion). *J. R. Statist. Soc., B*, **52**, pp. 393—442.

Eastoe, E.F. and Tawn, J.A. (2010). The distribution for the cluster maxima of exceedances of sub–asymptotic thresholds. *Biometrika*, **xx**, x, pp. 1—16.

Fawcett, L. (2005). Statistical Methodology for the Estimation of Environmental Extremes. PhD Thesis, University of Newcastle–upon–Tyne.

Fawcett, L. and Walshaw, D. (2008). Bayesian inference for clustered extremes. *Extremes*, **11**, pp. 217—233.

Fawcett, L. and Walshaw, D. (2007). Improved Estimation for Temporally Clustered Extremes. *Environmetrics*, **18**, 2, pp. 173—188.

Fawcett, L. and Walshaw, D. (2006). Markov Chain Models for Extreme Wind Speeds. *Environmetrics*, **17**, 8, pp. 795—809.

Ferro, C.A.T. and Segers, J. (2003). Inference for clusters of extreme values. *J. R. Statist. Soc., B*, **65**, pp. 545—556.

Gomes, M.I., (1993). On the estimation of parameters of rare events in environmental time series. In *Statistics for the environment 2: Water Related Issues* (V. Barnett and K.F. Turkman), pp. 225—241.

Hsing, T., Hüsler, J. and Leadbetter, M.R. (1988). On the exceedance point process for a stationary sequence. *Prob. Theory Rel. Fields* **78**, pp. 97—112.

Leadbetter, M.R. and Rootzén, H. (1988). Extremal theory for stochastic processes. *Ann. Probab.*, **16**, pp. 431—476.

Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Series*. Springer–Verlag, New York.

Loynes, R.M. (1965). Extreme values in uniformly mixing stationary stochastic processes. *Ann. Math. Statist.*, **36**, pp. 993—999.

Newell, G.F. (1964). Asymptotic extremes for m–dependent random variables. *Ann. Math Statist.*, **35**, pp. 1322—1325.

Northrop, P. and Jonathan, P. (2010). Modelling spatially–dependent non–stationary extremes with application to hurricane–induced wave heights. *Environmetrics*, submitted.

O'Brien, G.L. (1974). The maximum term of uniformly mixing stationary processes. *Z. Wahrscheinlichkeitsth*, **30**, pp. 57—63.

Pickands, J. (1981). Multivariate extreme value distributions. *Bullettin International Statistical Institute* **XLXI** (Book 2): pp. 859—878.

Smith, R.L. (1992). The Extremal Index for a Markov Chain. *J. Appl. Prob*, **29**, pp. 37—45.

Smith, R.L. (1991). Regional Estimation from Spatially Dependent Data. Preprint.

Smith, R.L. and Weissman, I. (1994). Estimating the Extremal Index. *J. R. Statist. Soc., B*, **56**, 515—528.

Stephenson, A. (2003). Simulating Multivariate Extreme Value Distributions of Logistic Type. *Extremes*, **6**, pp. 49—59.

Venzon, D.J. and Moolgavkar, S.H. (1988). Profile-likelihood–based confidence intervals. *Applications of Statistics*, **37**, pp. 87—94.

Walshaw, D. (1994). Getting the Most From Your Extreme Wind Data: A Step by Step Guide. *J. Res. Natl. Inst. Stand. Technol.*, **99**, pp. 399—411.

Walshaw, D. (1991). Statistical Analysis of Extreme Wind Speeds. PhD Thesis, University of Sheffield.