

Investigating the properties of predictive return levels for environmental extremes

Lee Fawcett · David Walshaw

Received: Dec 2015 / Accepted: date

Abstract Keywords Rainfall · sea-surge · wind speeds · extreme value theory · spatial extremes · exceedance probabilities

1 Introduction

Estimating extremes of environmental phenomena such as sea-surge and wind speed plays an important role in structural design. Models from classical extreme value theory, such as the generalised extreme value (GEV) or generalised Pareto (GP) distributions (for block maxima or excesses over a high threshold respectively), give us limiting models for the tail behaviour of such variables and provide a general template for modelling and extrapolation. The aim of most practical applications is the estimation of the event we might expect to see, on average, once every r years: the so-called r -year return level, commonly notated as z_r . Over the last three decades or so, pragmatic solutions to circumvent departures from the ideal of independent and identically distributed observations on extremes have been developed; see, for example, the seminal paper by Davison and Smith (1990). Some of these solutions result in sample size reduction; for example, the use of filtering schemes to avoid issues of serial correlation, or using only those extremes from within a particular calendar unit to avoid problems associated with seasonal variability. However, more recently (e.g. Fawcett and Walshaw, 2012; Eastoe and Tawn, 2012), methods have been proposed to maximise the number of extremes pressed into use which, as Fawcett and Walshaw (2016) illustrate, can substantially reduce return level estimation uncertainty.

More recently, much focus has been given to the extension, and practical application, of the theory of multivariate extremes, often motivated by the need to account for spatial dependence between extremes. Davison *et al.* (2012) provide a comprehensive coverage of the development of models for extremes occurring spatially, including max-stable models: again, the aim is for the estimation of return levels, albeit on maps over a spatial grid. The increasing sophistication of models to address issues such as serial correlation, seasonal variability and trend, coupled with an increase in dimensionality required of an analysis which is – for example – spatial in flavour, often makes inference within a maximum likelihood setting difficult (if not altogether infeasible) and thus researchers might appeal to simulation-based procedures within the Bayesian paradigm to make inferences on model parameters.

Of course, the Bayesian paradigm does not just provide a more convenient and accessible inferential framework; it also allows the incorporation of external sources of information through the prior distribution, an obvious element of appeal for any analyst of extremes working with scarce data. Also appealing is the natural extension within the Bayesian framework to prediction; as Fawcett and Walshaw (2016) discuss, the r -year posterior predictive return level estimate, $\hat{z}_{r,\text{pred}}$, provides practitioners with a single point summary, all sources of variability – including randomness in future observations – being captured within this summary. Given the virtues of a Bayesian analysis of extremes, it is then surprising that more publications in the field do not work within this framework. Certainly, it is our experience that very few practitioners will choose to perform extreme value analyses within a Bayesian setting.

Coles and Powell (1996) provided a solid review of Bayesian inference for extremes up to that date; since then, Coles and

L. Fawcett
School of Mathematics & Statistics, Newcastle University, Newcastle upon Tyne NE1 7RU, U.K.
Tel.: +44-191-2087228
E-mail: lee.fawcett@ncl.ac.uk

Tawn (1996) and Smith and Walshaw (2003) have investigated the merits of expertly-elicited priors whereas Beirlant *et al.* (2004) and Eugenia Castellanos and Cabras (2007) have considered objective priors for the GEV and GP models, and various authors have used the Bayesian paradigm to exploit meteorological structure in their data via hierarchical models for extremes - for example, Fawcett and Walshaw (2006), Sang and Gelfand (2009, 2010) and Davison *et al.* (2012). Smith (1999) compares predictive inference under the Bayesian and frequentist paradigms and Coles and Tawn (1996) give some informal comparisons between predictive return levels and estimates based solely on the posterior distribution for z_r . Fawcett and Walshaw (2006; 2006a; 2008; 2016) demonstrate predictive inference for return levels of wind speed and sea-surge extremes, recommending $\hat{z}_{r,\text{pred}}$ as the most convenient, and useful, representation of a return level for practitioners. However (as far as we are aware) no published work supports this through an investigation into the performance of this statistic. The main contribution of this paper, then, is to explore the properties of $\hat{z}_{r,\text{pred}}$. In particular, we focus on a comparison of the exceedance probabilities of $\hat{z}_{r,\text{pred}}$ to their intended values, and the general performance of $\hat{z}_{r,\text{pred}}$ to other estimative summaries obtained directly from the posterior distribution for $\hat{z}_r$. **More here...**

This paper is organised as follows...

2 Practical motivation

In this Section, we introduce the sea-surge and wind speed data we will refer to throughout the paper. We then give a brief overview of the basic methods for modelling extremes on these processes, including some general background on Bayesian sampling and specific details relating to the posterior predictive return level. Some results are then presented comparing estimative return levels to their predictive counterparts.

2.1 Data

Figure 1 shows time series plots and plots of each observation against its lag 1 counterpart, for a section of three-hourly sea-surge data observed at Newlyn, southwest England, and a section of hourly wind speed data observed at High Bradfield in the Peak District of England. On close inspection, the wind speed data exhibit clear seasonal variability and, as Fawcett and Walshaw (2015) discuss, a partition into seasonal units of one month allows this variation to be captured through a monthly piecewise approach to modelling, the assumption of stationarity *within* months being reasonable. Although there are no obvious signs of non-stationarity within the sea-surge series, both series clearly

Figure 1 about here

Fig. 1 Figure 1's caption here

display significant first-order autocorrelation, even above a threshold beyond which we might consider observation to be extreme. Estimates of z_r or $z_{r,\text{pred}}$ based on fitting an appropriate model to the sea-surge extremes might be used to inform the design of a new sea wall; the British Standards Institute (BSI) use estimates of the 50-year wind speed, from fits to wind speed extremes such as those shown for High Bradfield, to produce contour maps displaying the strength requirements for new structures. In either case, we argue that a Bayesian approach to return level inference can improve the estimation procedure, with the potential to reduce estimation uncertainty and, in practical terms, to ability to provide practitioners with a single point summary having all sources of uncertainty 'built in' through the predictive return level.

2.2 Statistical modelling

2.2.1 The basics

Let $\{X_n\}$ denote a stationary sequence of random variables with common distribution function F , and let $M_n = \max\{X_n\}$. It is typically the case that, as $n \rightarrow \infty$,

$$\Pr(M_n \leq x) \approx F^{n\theta}(x), \quad (1)$$

where $\theta \in (0, 1)$ is known as the *extremal index*; e.g. Leadbetter and Rootzén (1988). As $\theta \rightarrow 0$ there is increasing dependence in the extremes of the process; for an independent process, $\theta = 1$. In practice, F is unknown, and very small discrepancies in estimates of F obtained from observed data can lead to rather substantial discrepancies for F^n (and thus $F^{n\theta}$). Initially concerned with the independent case (i.e. $\theta = 1$), classical extreme value theory sought families of limiting models for F^n for large n . This leads to the GEV distribution (e.g. Jenkinson, 1955), with distribution function

$$\mathcal{G}(y) = \begin{cases} \exp\left[-(1 + \xi(y - \mu)/\varsigma)^{-1/\xi}\right], & \xi \neq 0 \\ \exp[-\exp(-(y - \mu)/\varsigma)] & \xi = 0, \end{cases} \quad (2)$$

defined on $\{y : 1 + \xi(y - \mu)/\varsigma > 0\}$, where $-\infty < \mu < \infty$, $\varsigma > 0$ and $-\infty < \xi < \infty$ are location, scale and shape parameters (respectively). The GEV can be used to model a set of

block maxima $\{M_\tau\}$ with block length τ ; the calendar year is often used for τ , giving the set of *annual maxima*.

Pickands (1975) showed that for large enough u the distribution of $(X - u)$, conditional on $X > u$, is approximately GP with distribution function

$$\mathcal{H}(y) = \begin{cases} 1 - (1 + \xi y / \sigma)^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp[-y/\sigma] & \xi = 0, \end{cases} \quad (3)$$

defined on $\{y : y > 0 \text{ and } (1 + \xi y / \sigma) > 0\}$, where $\sigma = \zeta + \xi(u - \mu)$ and ξ are the GP scale and shape parameters (respectively). The GP distribution, being the limiting distribution for excesses over a high threshold u , provides a natural way of modelling extremes of time series such as those shown in Figure 1. Modelling extremes in this way can be less wasteful than a block maxima approach using the GEV, since more extremes are usually pressed into use.

2.2.2 Practicalities

Using the GP distribution to model threshold excesses, the linearity of $E[X - u | X > u]$ in u can be exploited in a *mean residual life plot* (MRL plot; see Coles, 2001, Ch. 4) to help find a suitably high threshold u for the classification of extremes. To maximise estimation precision, Fawcett and Walshaw (2015) suggest making use of *all* excesses over u , despite the obvious serial correlation present. Specifically, they propose fitting (3) to all excesses over u , by adopting one of the following strategies:

- (i) Where appropriate, assume a first-order Markov structure for the temporal evolution of extremes; that is, assume the following likelihood for ψ is:

$$L(\psi) = \prod_{i=1}^{n-1} f(x_i, x_{i+1}; \psi) \bigg/ \prod_{i=2}^{n-1} f(x_i; \psi), \quad (4)$$

where ψ is a parameter vector containing marginal and dependence parameter(s). Appealing to bivariate extreme value theory, transformation from GP to standard Fréchet margins gives a range of models to use for the serial correlation between consecutive extremes, the most commonly used being the logstic family with dependence parameter α and distribution function:

$$G(x_i, x_{i+1}) = \exp \left\{ - \left(x_i^{-1/\alpha} + x_{i+1}^{-1/\alpha} \right)^\alpha \right\}. \quad (5)$$

Here, independence and complete dependence are attained when $\alpha = 1$ and $\alpha \rightarrow 0$ respectively. Differentiation of (5), with careful censoring when either one or both of (x_i, x_{i+1}) lies sub-threshold, gives pairwise contributions to the numerator in (4); univariate contributions to the denominator are given through (3). The polynomial relationship:

$$\theta \approx 0.013 - 0.092\alpha + 1.833\alpha^2 - 0.756\alpha^3, \quad (6)$$

as discussed in Fawcett and Walshaw (2012), can then be used – after estimation of $\psi = (\sigma, \xi, \alpha)^T$ (through maximum likelihood or Bayesian inference, for example) – to provide the fitted distribution for the right-hand-side of (1). Within this class of models for asymptotic dependence, other models can be used. For example, the bilogistic model, which allows for asymmetry in the dependence structure between (x_i, x_{i+1}) through the inclusion of an additional dependence parameter β ($0 < \alpha, \beta < 1$); see, for example, Coles (2001, Ch. 8). Indeed, Fawcett and Walshaw (2012) suggest polynomial expressions for the extremal index based on the fitted values of (α, β) here, too. Of course it might be that, for the dependence structure, asymptotic *independence* is more appropriate and here, standard time series models such as $AR(p)$ processes can be used in place of the models we have outlined for consecutive variables which are asymptotically dependent. In this case, $\theta = 1$, although Ancona-Navarrete and Tawn (2000) derive penultimate approximations for $\theta(u_p)$, a threshold-dependent extremal index with a threshold set at the p -% quantile u_p ; for example, for an $AR(1)$ process, $\theta(u_{0.95}) \approx 0.711$ and $\theta(u_{0.99}) \approx 0.855$.

- (ii) Initially ignore dependence and proceed by fitting the GP distribution to all excesses over u to approximate F^n in (1). Then use a suitable extremal index estimator to complete the right-hand-side in (1) and adjust for extremal dependence. Fawcett and Walshaw (2015) make various recommendations for the extremal index estimator that should be used, but a simulation study shows that the estimator of Ferro and Segers (2003), given by

$$\hat{\theta} = \min \left(1, \frac{2 \{ \sum_{i=1}^{K-1} (T_i - a) \}^2}{(K-1) \sum_{i=1}^{K-1} (T_i - b)(T_i - c)} \right), \quad (7)$$

where T_i are the inter-arrival times between threshold excesses ($a = b = c = 0$ if the largest inter-arrival time is no greater than 2, and $a = b = 1$ and $c = 2$ if the largest inter-arrival time is greater than 2), strikes a good balance between optimising accuracy and precision *and* providing the practitioner with an estimator which is easy to use.

In the presence of seasonal variability, Fawcett and Walshaw (2015) recommend a seasonal piecewise approach to modelling with a unique GP model for extremes within each season. For example, assuming either (i) or (ii) above to capture temporal dependence, estimates of $(\sigma_m, \xi_m, \theta_m)$, $m = 1, \dots$ might be obtained for each season m , the analysis being simplified by assuming a common shape parameter ξ or extremal index θ across all seasons where appropriate. Anecdotal evidence in, for example, Walshaw (1994) and Fawcett and Walshaw (2006) indicates there are no real gains to be made, in terms of return level inference, by allowing the GP

parameters to vary smoothly through time (e.g. by assuming Fourier forms).

Though not a feature of our data in Figure 1, long term trends in extremes can be modelled by imposing a time dependence on the GPD scale parameter, i.e. $\sigma = \exp\{\beta_0 + \beta_1 t\}$, $t = 1, 2, \dots$, where t is an indicator of time. More generally, a dependence on covariates can be induced by writing the parameters in the form $h(\mathbf{X}^T \boldsymbol{\beta})$, where h is a specified function, $\boldsymbol{\beta}$ is a vector of parameters and $\mathbf{X}$ is a model vector.

2.2.3 Return levels

Inversion of the right-hand-side of (1), assuming a GP model for threshold excesses, gives the following expression for the r -year return level:

$$z_r = \begin{cases} u + \sigma \xi^{-1} \left[(\lambda_u^{-1} w_r)^{-\xi} - 1 \right] & \xi \neq 0 \\ u - \sigma \log(\lambda_u^{-1} w_r) & \xi = 0, \end{cases} \quad (8)$$

where $w_r = 1 - [1 - (rn_y)^{-1}]^{1/\theta}$, λ_u is the rate of threshold excess and n_y is the (average) number of observations per year. Under approach (i) for dependence, as outlined in the previous Section, (σ, ξ, θ) can be replaced with their maximum likelihood estimates or Bayesian estimates to obtain estimates of z_r , with an assessment of uncertainty being made through standard errors or posterior standard deviations. Under approach (ii), (σ, ξ) can be replaced with their maximum likelihood or Bayesian estimates, with a bootstrap procedure as proposed in Fawcett and Walshaw (2012) enabling the incorporation of uncertainty in estimates of θ through (7) into estimates of z_r .

To recombine seasonally-varying parameters when estimating return levels, assuming independence between seasons we can solve the following for $x = z_r$:

$$\prod_{m=1}^M H(x)^{n_m \theta_m} = 1 - r^{-1}, \quad (9)$$

where H_m is the GP distribution function in season m with parameter set $(\lambda_{u_m}, \sigma_m, \xi_m)$, θ_m / n_m are the extremal index / number of observations in season m , and M is the number of seasonal units over which we partition our data. Of course, as discussed earlier, inference can be simplified if we assume a constant shape or dependence across all seasons. Uncertainty can be quantified through the standard error or posterior standard deviation for z_r , depending on the estimation framework being used. Similarly, confidence intervals (perhaps based on profile likelihood) or credible intervals can also be obtained.

2.3 Bayesian inference for sea-surge and wind speed extremes

Given investigations into the clustering of our sea-surge and wind speed extremes in Fawcett and Walshaw (2006a), we adopt approach (i) as outlined in Section 2.2.2 for handling serial correlation in both cases, the logistic model (Equation 5) under an assumption of first-order Markovianity seemingly being appropriate (comparisons with other models, and with models of higher orders, did not improve over a fit to the first-order logistic model, and diagnostics such as the $\chi / \bar{\chi}$ plots discussed in Coles (2001, Ch. 8) imply asymptotic dependence in both cases). Following the recommendations of Fawcett and Walshaw (2015) we adopt a fully Bayesian approach to inference, using Markov chain Monte Carlo (MCMC) techniques to draw approximate samples from the marginal posteriors for (σ, ξ, θ) and hence z_r through Equations (8) or (9). Details on MCMC techniques are now extensively published (e.g. Smith and Roberts, 1993), so we omit such detail here; suffice to say, all inferences are performed using a Hastings-Metropolis procedure and interested readers can re-produce the results shown in this Section by downloading our R code, and the wind speed and sea-surge data, at:

www.mas.ncl.ac.uk/~nlf8/predict

In both sea-surge and wind speed analyses, MRL plots have been used for the selection of suitable thresholds (monthly-varying for the wind speed data; see Section 2.3.1).

2.3.1 Prior specification

Generally, we work with a re-parameterised GP scale parameter:

$$\eta = \log(\sigma - \xi u).$$

This re-parameterisation gives a scale which is threshold-independent (unlike σ), allowing the specification of an objective prior for η across all threshold levels; working with the natural logarithm of this retains the positivity of this parameter in our MCMC sampling scheme. In the absence of any prior knowledge regarding the GP parameters in the sea-surge analysis, we adopt largely non-informative, independent priors for (η, ξ, α) ; based on work in Fawcett (2005), we adopt informative prior distributions for these parameters in the wind speed analysis based on observations on wind speed extremes made at a nearby location. Specifically, for the sea-surge analysis, we use:

$$\eta \sim N(0, 10^4), \quad \xi \sim N(0, 10^2) \quad \text{and} \quad \alpha \sim U(0, 1).$$

For the analysis of wind speeds at High Bradfield, we use:

$$(\eta_m, \xi_m) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{and} \quad \alpha_m \sim \text{Beta}(10, 19),$$

$m = 1, \dots, 12$. Given the seasonal variability observed in the wind speed extremes, and discussions about the U.K. wind climate in Walshaw (1994), we use a seasonal unit of one month, assuming stationarity of wind speed extremes *within* each month. The components of μ are chosen to closely match our beliefs about what are the most likely values of (η_m, ξ_m) based on our study of wind speeds at the nearby location. We specify values for $\text{cov}(\eta_m, \xi_m)$ according to our beliefs about the covariances between these parameters at the nearby location, scaled (although rather crudely) to reflect our uncertainties about differences between monthly wind speeds at the two locations. We choose our priors for α_m for similar reasons and, as is often the case, we assume independence between the marginal and dependence parameters.

2.3.2 MCMC sampling

We set initial values for all parameters to their prior means, updating using a simple Metropolis step to give successive draws

$$(\eta^{[j]}, \xi^{[j]}, \alpha^{[j]}) \quad \text{and} \quad (\eta_m^{[j]}, \xi_m^{[j]}, \alpha_m^{[j]}), \quad j = 1, \dots, 10^5,$$

in the sea-surge and wind speed analyses, respectively, with $m = 1, \dots, 12$ in the latter. Within each Metropolis step, we use a random walk update to generate candidate values for each of the parameters, tuning the innovation variances to optimise the efficiency of our sampler. Convergence was assessed by starting each chain at multiple new initial values and observing the trace plots.

Equation (6) is used to obtain a sample from the posterior for the extremal index, after which a posterior sample for z_r is obtained on substitution of successive draws for the GP parameters and the extremal index into (8) / (9) in the sea-surge / wind speed analyses respectively.

2.3.3 Prediction

As discussed earlier, one of the advantages of a Bayesian analysis of extremes is the natural extension to prediction via the posterior predictive distribution. If Y denotes a future extreme of our sea-surge or wind speed data, then allowing for uncertainty in parameter estimation and randomness in future observations we can write

$$\Pr\{Y \leq y | \mathbf{x}\} = \int_{\Psi} \Pr\{Y \leq y | \psi\} \pi(\psi | \mathbf{x}) d\psi \quad (10)$$

for the *predictive distribution* of our extremes, where $\mathbf{x}$ represents past observations. Solving

$$\Pr\{Y \leq z_{r,\text{pred}} | \mathbf{x}\} = 1 - r^{-1} \quad (11)$$

for $z_{r,\text{pred}}$ therefore gives an estimate of the r -year return level that incorporates uncertainty due to model estimation and

randomness in future observations. Although (10) is analytically intractable, it can be approximated since we have estimated the posterior distribution using MCMC. Regarding our sample $\psi^{(1)}, \dots, \psi^{(S)}$ as realisations from the stationary distribution $\pi(\psi | \mathbf{x})$, we have

$$\Pr\{Y \leq z_{r,\text{pred}} | \mathbf{x}\} \approx \frac{1}{S} \sum_{j=1}^S \Pr\{Y \leq z_{r,\text{pred}} | \psi^{[j]}\}, \quad (12)$$

which we can set equal to $1 - r^{-1}$ and solve for $z_{r,\text{pred}}$ using a numerical solver.

2.3.4 Some results

Table 1 shows some estimative return levels for the sea-surge and wind speed extremes; that is, some point summaries from the posterior distributions for z_r , for $r = 10, 50$ and 1000. Also shown are summaries of the spread of these posteriors, including the posterior standard deviations and 95% credible intervals. Accompanying these estimative return levels are their predictive counterparts $z_{r,\text{pred}}$. Figure 2 shows plots of both z_r and $z_{r,\text{pred}}$ over a range of values for r (on the usual logarithmic scale for these plots to magnify results for long-range return periods; the posterior mean, median and modes are shown for z_r , along with the 95% credible intervals). It is clear from both Table 1 and Figure 2 that designing a structure to withstand the extremes of sea-surge or wind-speed as suggested by the estimative return levels could result in substantial under-protection, relative to the predictive estimates, even if the upper end-point of the 95% credible interval were used for longer-range return periods.

As discussed in Fawcett and Walshaw (2015), the predictive return level estimate might be preferred owing to its inclusion of all sources of uncertainty *and* because it provides the practitioner with a single point summary which encapsulates this uncertainty. However, open questions remain about the quantity $z_{r,\text{pred}}$: For example, how do exceedance probabilities of $z_{r,\text{pred}}$ compare to the intended values r^{-1} (on an annual scale)? How do these probabilities compare to those under an estimative approach for z_r ? We might expect $z_{r,\text{pred}}$ to give exceedance probabilities much lower than r^{-1} , but can these discrepancies be quantified and could they result in *over-protection*? At the very least, practitioners should be aware of these discrepancies, should they choose to work with $z_{r,\text{pred}}$ (and we recommend that they do). Are there any advantages of using $z_{r,\text{pred}}$ as opposed to using some other point summary from the posterior for z_r , such as the upper end-point of the 95% credible interval or perhaps some other quantile? We aim to answer these questions, and more, in the simulation study in the next Section.

Figure 2 about here

Fig. 2 Figure 2's caption here

Table 1 about here

Table 1 Table 1's caption here

3 Simulation study

Given the results discussed in Section 2.3.4, we believe there is a need to investigate the properties of $z_{r,\text{pred}}$ relative to estimative return levels obtained as point summaries from the posterior distribution for z_r . In particular, of interest to practitioners might be the actual return probabilities associated with $z_{r,\text{pred}}$ and other estimative return levels, and how these compare to r^{-1} .

3.1 Study design

The first stage of our simulation study requires the simulation of a reference series of stationary extremes $\mathbf{y}^{(\text{Ref})}$ of length N years with N_y observations per year, where N is very large, assuming a first-order Markov structure. We consider a range of models for the dependence between neighbouring extremes in $\mathbf{y}^{(\text{Ref})}$. Specifically, under the assumption of asymptotic dependence we use the logistic and bilogistic models as discussed in Section 2.2.2, covering

a range of serial correlations through specific choices for $\alpha / (\alpha, \beta)$. To account for scenarios in which dependence is observed above levels of practical interest, but in which an assumption of independence asymptotically might be more realistic, we also simulate $\mathbf{y}^{(\text{Ref})}$ from an AR(1) process with autocorrelation A , again covering a range of serial correlations through specific choices for A . Marginally, our reference series are GP-distributed with scale and shape (σ, ξ) giving scale and shape $(\sigma^* = \sigma + \xi u, \xi^* = \xi)$ for excesses over some threshold u owing to the threshold stability property of the GP distribution.

Now, at each replication ℓ in our simulation study, $\ell = 1, \dots, L$, we simulate a stationary series of extremes $\mathbf{y}^{(\ell)}$ of length n years, with n_y observations per year, n perhaps being typical of what we might usually work with in terms of environmental extremes. As with the reference dataset, we assume a first-order Markov structure according to models for asymptotic dependence and asymptotic independence, with the same GP marginal structure as before. For each series $\mathbf{y}^{(\ell)}$ of length n years, $\ell = 1, \dots, L$, we perform a full MCMC procedure, using objective priors akin to those used for the sea-surge data in Section 2.3.1. This yields S iterations after burn-in, on threshold exceedances above a range of thresholds u , to obtain approximate samples $\sigma^{*(\ell)}$ and $\xi^{*(\ell)}$ of length S from the posterior distributions of the GPD scale and shape σ^* and ξ^* , as well as approximate samples from the posteriors of the dependence parameters (i.e. α , (α, β) or A) and hence samples $\theta^{(\ell)}$ from the posterior of the extremal index; see Section 2.2.2. At each replication ℓ , via Equation (8) we also obtain posterior draws $\mathbf{z}_r^{(\ell)}$ from the r -year return levels for a range of return periods r ; from these draws we can obtain the posterior mean $\bar{z}_r^{(\ell)}$, the posterior mode $\dot{z}_r^{(\ell)}$ and the posterior 95% confidence interval upper bound $z_{r,\text{upper}}^{(\ell)}$; we also obtain the predictive return level $z_{r,\text{pred}}^{(\ell)}$ via Equation (12), essentially giving sampling distributions of length L for each of these return level summaries. Defining p to be the proportion of annual maxima in $\mathbf{y}^{(\text{Ref})}$ exceeding each of $\bar{z}_r^{(\ell)}$, $\dot{z}_r^{(\ell)}$, $z_{r,\text{upper}}^{(\ell)}$ and $z_{r,\text{pred}}^{(\ell)}$, gives sampling distributions for $p_{\bar{z}_r}$, $p_{\dot{z}_r}$, $p_{z_{r,\text{upper}}}$ and $p_{z_{r,\text{pred}}}$, respectively. Other than the sampling distributions for each of the return level summaries themselves, of particular interest might be comparisons between each of the proportions p and the intended exceedance proportions r^{-1} .

3.2 Parameters in our study

In our study, we use $N = 10^5$ and $N_y = N \times 365.25 \times 24$, in line with having hourly measurements on our variable. We use $\alpha = \{0.1, 0.2, \dots, 0.9\}$ for the logistic model; for the bilogistic model we fix α at 0.5 and use $\beta = \{0.1, 0.2, \dots, 0.9\}$; for the AR(1) process we use $A = \{0.1, 0.2, \dots, 0.9\}$.

Marginally, we hold σ unit constant but consider a range of tail behaviours through the GP shape parameter ξ , where $\xi = \{-0.4, -0.1, 0, 0.1, 0.3\}$, yielding $\sigma^* = \xi u + 1$ and $\xi^* = \xi$, where we use $u = \{u_{0.9}, u_{0.95}, u_{0.99}\}$. We perform $L = 1000$ replications, within which we simulate chains $y^{(\ell)}$ of length $n = 10$ years with $n_y = N_y$, in keeping with the length of our wind speed series at High Bradfield; for each chain, we perform MCMC with $S = 10,000$ iterations after an appropriate burn-in discard. For each combination of $(\sigma^*, \xi^*, \theta)$ across all dependence models considered, we perform relatively small MCMC pilot runs in a bid to select suitable values for the MCMC tuning parameters before running the full simulation study, aiming for acceptance rates of between 20-30%.

3.3 Results

3.3.1 Asymptotic dependence

3.3.2 Asymptotic independence