Bayesian Inference for Clustered Extremes

Lee Fawcett and Dave Walshaw

Newcastle University, Newcastle-upon-Tyne, U.K.

lee.fawcett@ncl.ac.uk

20th TIES Conference: Bologna, Italy, July 2009

Structure of this talk

1. **Motivation and background**
2. **Inference for the extremal index**
   - Review of existing methods
   - Limitations/difficulties
   - A Bayesian sampling scheme
3. **Implementation**
   - Simulated data – just to check!
   - Extreme wind speeds observed at High Bradfield

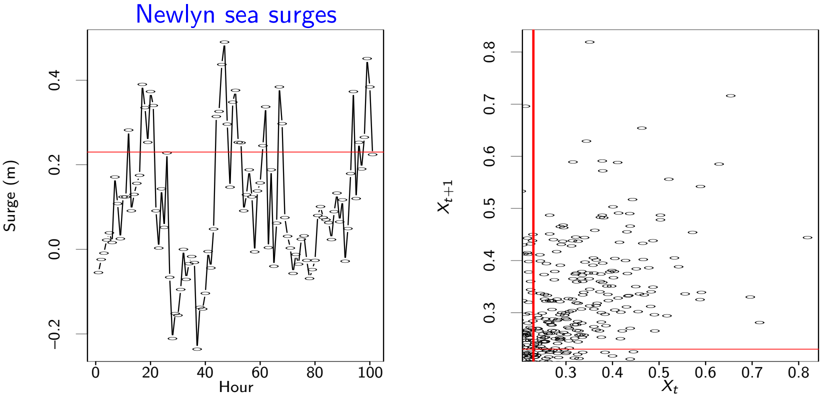
Motivation

Extremes of many observed environmental processes often occur in **clusters** due to short–term temporal dependence.

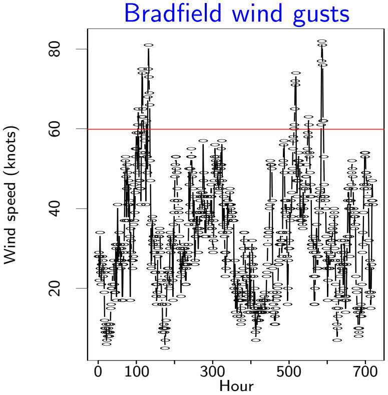
Such clusters of extremes often correspond to storms.

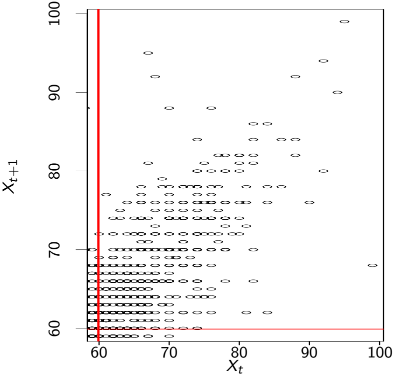
Being able to quantify this extremal dependence, and any other storm characteristics induced by this, can be of interest to meteorologists and/or engineers.

Examples


Newlyn sea surges
Surge (m)
−0.2
0.0
0.2
0.4
0
20
40
60
80
100
Hour
X_{t+1}
0.3
0.4
0.5
0.6
0.7
0.8
0.3
0.4
0.5
0.6
0.7
0.8
X_t

Examples


Bradfield wind gusts
Wind speed (knots)
80
60
40
20
0
100
300
500
700
Hour


X_{t+1}
100
90
80
70
60
60
70
80
90
100
X_t

The extremal index

Let $\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n$ be the first n observations of a stationary series satisfying Leadbetter's $D(u_n)$ condition (Leadbetter *et al.*, 1983), and let $\tilde{M}_n = \max\{\tilde{X}_1, \tilde{X}_2, \ldots \tilde{X}_n\}$.

Now let $X_1, X_2, \ldots, X_n$ be an *independent* series, with X having the same distribution as $\tilde{X}$, and let $M_n = \max\{X_1, X_2, \ldots, X_n\}$.

Then if M_n has a non–degenerate limit law given by $\Pr\{(M_n - b_n)/a_n \leq x\} \rightarrow G(x)$, it follows that

$$\Pr\left\{\left(\tilde{M}_n - b_n\right)/a_n \leq x\right\} \quad \rightarrow \quad G^{\theta}(x) \qquad (1)$$

for some $0 \leq \theta \leq 1$ (Leadbetter *et al.*, 1983).

The parameter θ is known as the **extremal index** and is a key parameter which quantifies the extent of extremal clustering.

The extremal index

A more convenient way of interpreting the extremal index, due to Hsing *et al.* (1988), is in terms of the propensity of the series to cluster at extreme levels.

Loosely,

$$\theta = (\text{limiting mean cluster size})^{-1}.$$

- For independent series, $\theta = 1$ (though the converse is not necessarily true)
- As $\theta \to 0$ we have increasing levels of extremal dependence

Thus, identifying **clusters** can be important in estimating θ.

Inference for the extremal index

We consider the following methods for estimating the extremal index:

- Cluster size methods;
- Maxima methods;
- Simulation–based methods;
- A method based on arrival times of extremes,

as well as a hybrid of two of these in a Bayesian setting.

We will also consider the viability of these approaches for estimating other cluster characteristics, often referred to as **cluster functionals**.

Simulating a dependent series

We generate a sequence of artificial data X_i, $i = 1, \ldots, n$, where the joint distribution of consecutive observations is given by

$$G(x_i, x_{i+1}; \alpha) = \exp\left[-\left\{\exp(-x_i/\alpha) + \exp(-x_{i+1}/\alpha)\right\}^{\alpha}\right] \quad (2)$$

for $i = 1, \ldots n-1$, $x_i, x_{i+1} > 0$ and $\alpha \in (0, 1]$.

- This is the (symmetric) **logistic** model
- Independence corresponds to $\alpha = 1$
- We have complete dependence as $\alpha \to 0$
- The margins are of **Gumbel** form, i.e. $F(x) = \exp\left\{\exp(-x)\right\}$

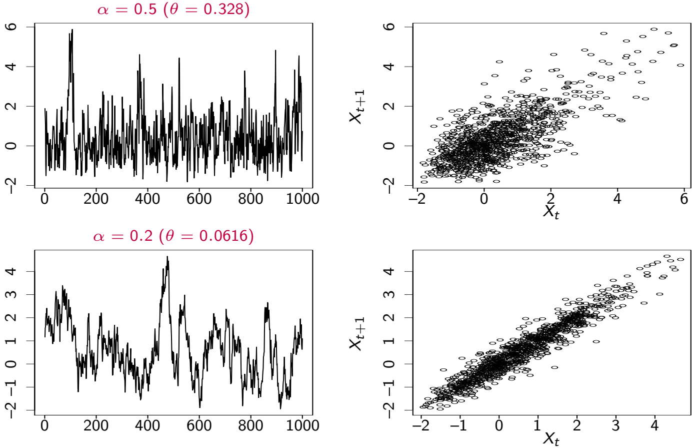
Simulating a dependent series

Work by Smith (1992) shows the following:

α	$\frac{1}{5}$	$\frac{1}{3}$	$\frac{1}{2}$
θ	0.0616	0.158	0.328
r_1 $(= 1 - \alpha^2)$	0.96	0.89	0.75

Thus, if we can simulate successive values from the logistic model in (2), then we can compare methods for estimating θ.

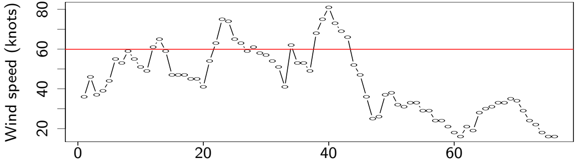
Simulating a dependent series



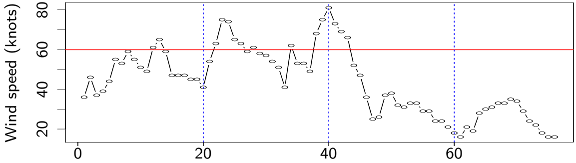
Cluster size methods

Cluster size methods are based on estimating θ as the reciprocal of the mean cluster size... but how do we identify clusters?

Runs declustering (separation interval $= \kappa$)



Blocks declustering (block length $= \tau$)



Cluster size methods

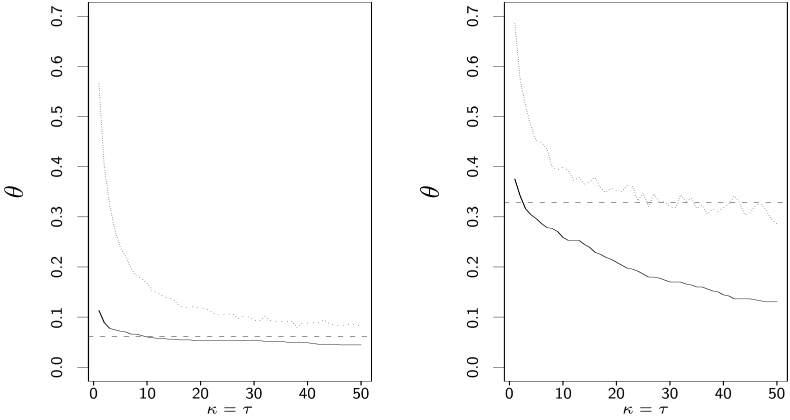
Then

$$\hat{\theta} \;=\; (\text{mean cluster size})^{-1}$$

Problems

- What value of κ/τ do we use?
- The choice of κ/τ will influence cluster size
- This will mean estimates of θ are sensitive to the choice of declustering scheme/choice of "declustering parameter"
- We get a point estimate of θ without any natural way of quantifying its uncertainty
- Other cluster functionals also sensitive to the choice of κ/τ!

Cluster size methods



Maxima methods

Recall that if M_n has a non–degenerate limit law given by $\Pr\{(M_n - b_n)/a_n \leq x\} \rightarrow G(x)$, it follows that

$$\Pr\left\{\left(\tilde{M}_n - b_n\right)/a_n \leq x\right\} \quad \rightarrow \quad G^{\theta}(x).$$

Denote the maximum of the ith "block" $M_{\tau,i}$.

It follows that, for large enough τ, the $M_{\tau,i}$ are approximately independent observations from a **generalised extreme value distribution** (GEV), where

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

Maxima methods

What about the *dependent* series $\tilde{M}_{\tau,i}$?

This is also GEV, but with d.f. $G^{\theta}(x) = G(x; \mu_\theta, \sigma_\theta, \xi_\theta)$, where

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

Ancona–Navarrete and Tawn (2000) suggest simultaneous estimation of the parameter vector $(\mu, \sigma, \xi, \theta)$ by treating components of the vector $(\mathbf{M}_\tau, \tilde{\mathbf{M}}_\tau)$ as independent GEV random variables.

Maxima methods

Results

Using maximum likelihood estimation, we obtain the following results for the simulated datasets:

- $\theta = 0.328$: $\hat{\theta} = 0.297$ (0.047)
- $\theta = 0.0616$: $\hat{\theta} = 0.063$ (0.052)

Limitations

- Joint vector not independent, though Ancona–Navarrete and Tawn (2000) show that the impact of this approximation is asymptotically zero
- No clusters are identified! And so we cannot estimate cluster characteristics that are not direct functions of θ

Simulation methods

Smith *et al.* (1997) propose a simulation framework for estimating the extremal index and other cluster functionals.

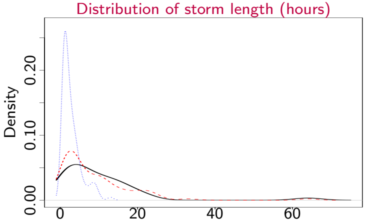
- Assume a first–order Markov structure for extremes
- Model the distribution of consecutive pairs using a bivariate extreme value distribution (such as the logistic model), estimating an appropriate dependence parameter (such as α)
- Repeatedly simulate clusters of extremes using the fitted model, and observe what happens!

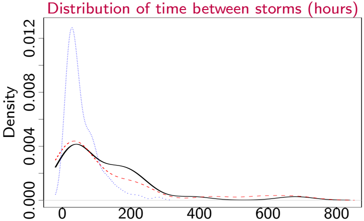
Simulation methods

Limitations

- Is a first–order Markov assumption appropriate?
- Which bivariate extreme value model for consecutive pairs?
- How can we check the above two points? Ad–hoc procedures?

Simulation methods

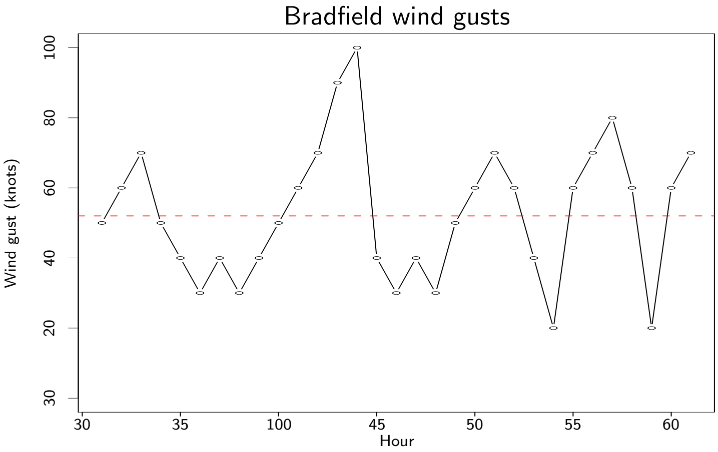




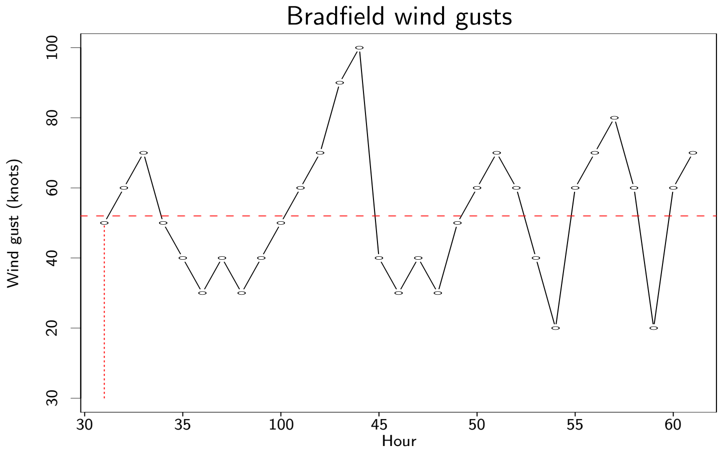
Automatic declustering (Ferro and Segers, 2003)

- Let T_i, $i = 1, \ldots, N-1$ be the **inter–arrival times** between threshold exceedances $S_1, \ldots, S_N$;
- Assume that the largest $C - 1 = \lfloor \theta N \rfloor$ inter–arrival times are approximately independent *inter*–cluster times that divide the series into independent sets of *intra*–cluster times;
- Equivalent to runs declustering with $\kappa = T_{(C)}$, where $T_{(C)}$ is the C–th largest inter–arrival time;
- No need to specify an arbitrary value for κ now – let this be governed by the level of extremal dependence in the process (via θ);
- What about θ? Likelihood based on inter–arrival times performs poorly, so they use a non–parametric approach.

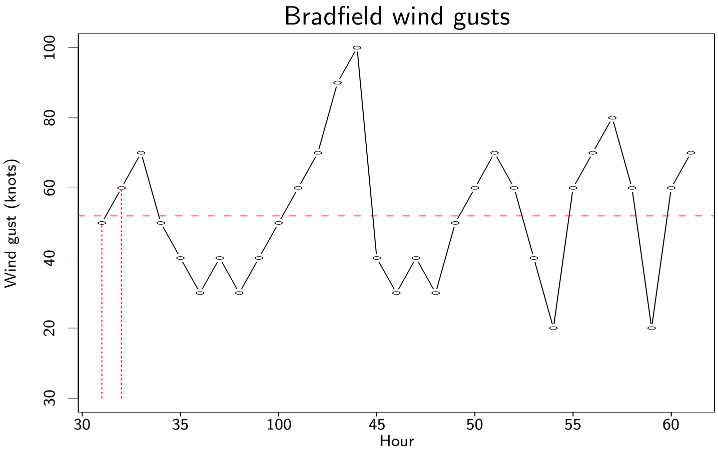
Automatic declustering (Ferro and Segers, 2003)



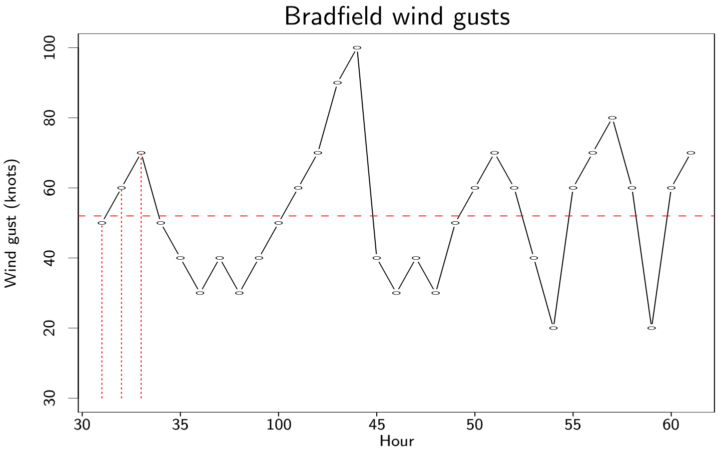
Automatic declustering (Ferro and Segers, 2003)



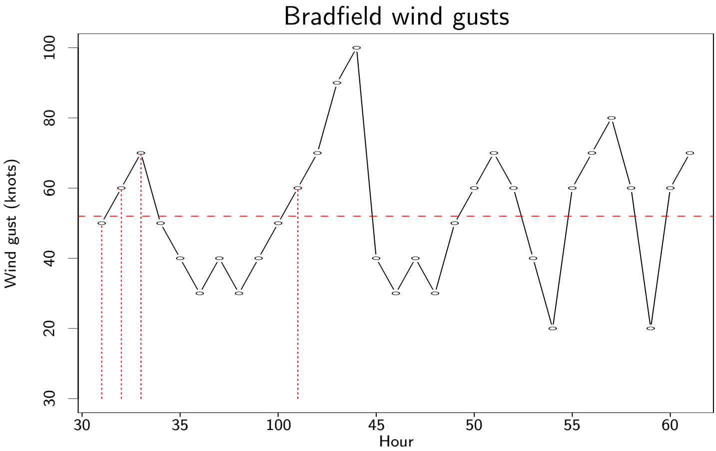
Automatic declustering (Ferro and Segers, 2003)



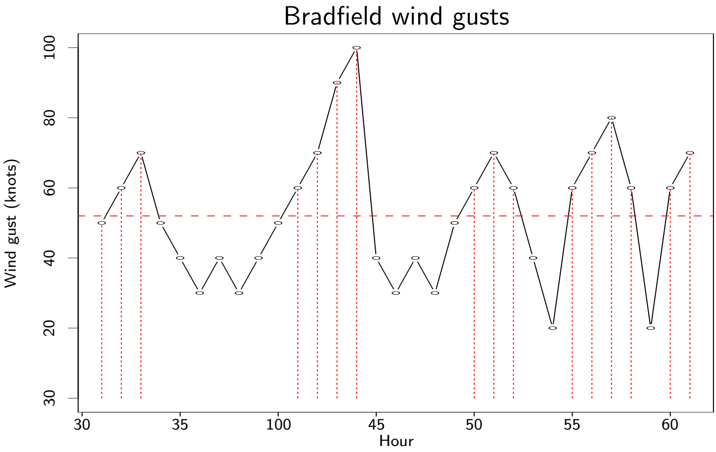
Automatic declustering (Ferro and Segers, 2003)



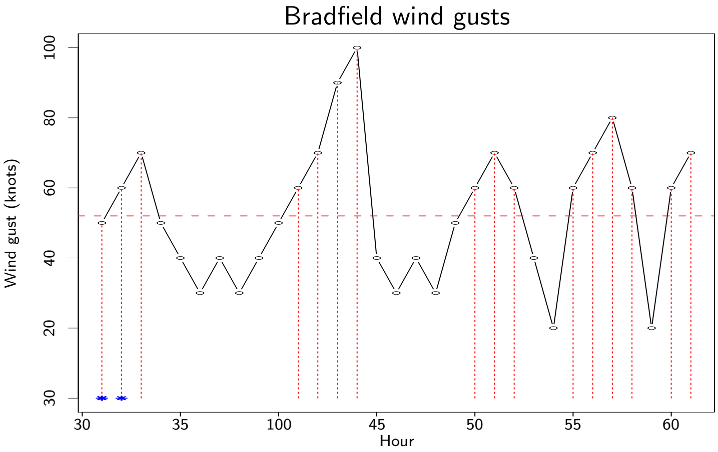
Automatic declustering (Ferro and Segers, 2003)



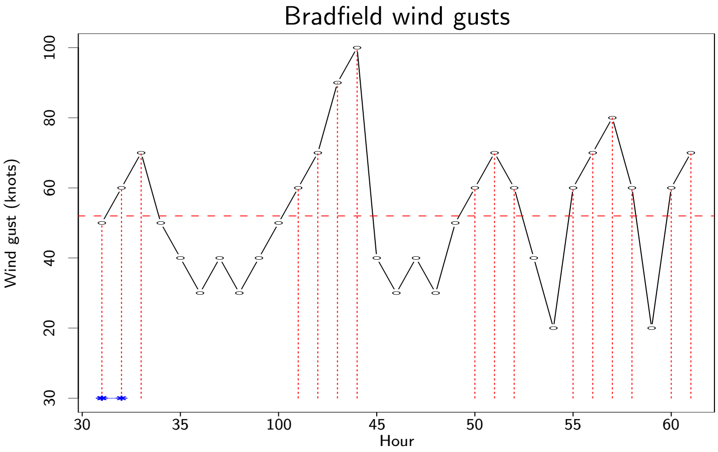
Automatic declustering (Ferro and Segers, 2003)



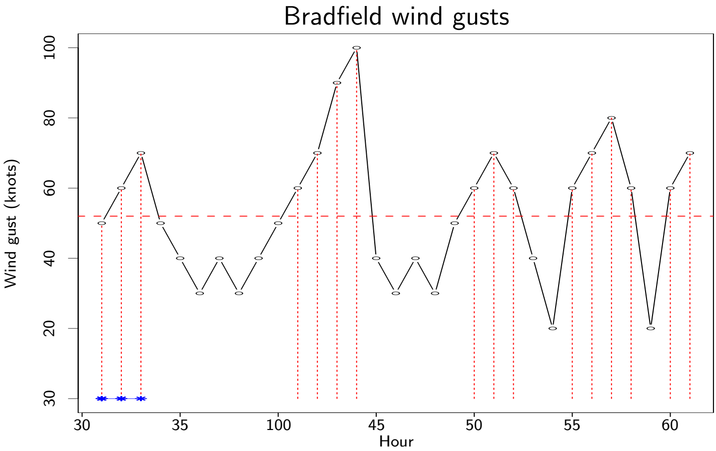
Automatic declustering (Ferro and Segers, 2003)



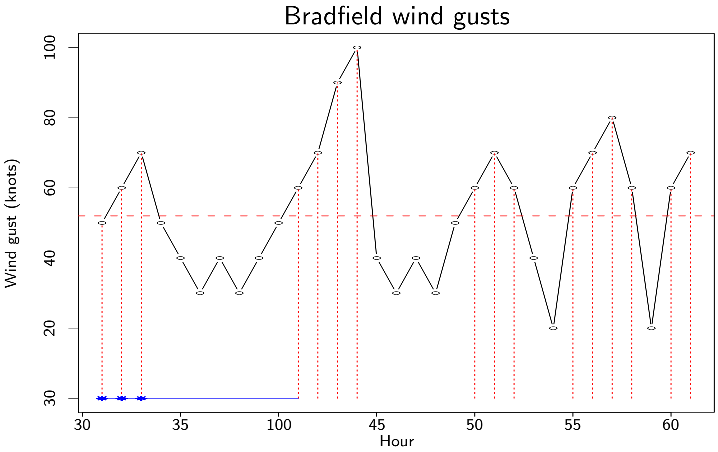
Automatic declustering (Ferro and Segers, 2003)



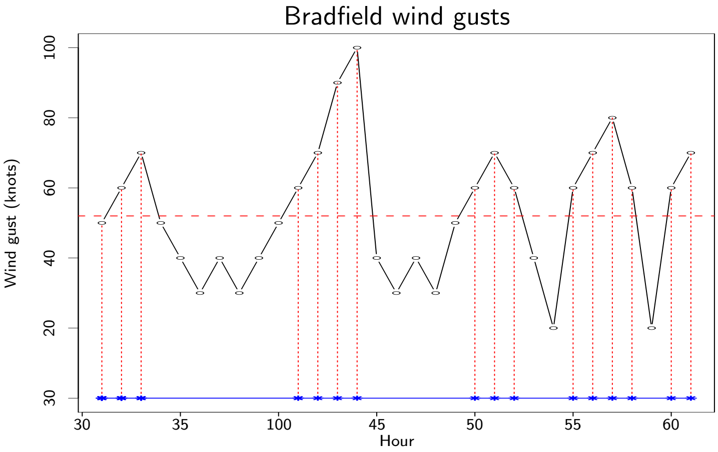
Automatic declustering (Ferro and Segers, 2003)



Automatic declustering (Ferro and Segers, 2003)



Automatic declustering (Ferro and Segers, 2003)



Bayesian inference

We now combine the likelihood from the maxima method (Ancona–Navarrete and Tawn, 2000), with the 'automatic' declustering procedure (Ferro and Segers, 2003), to implement a Bayesian sampling scheme for θ and any other cluster functional of interest:

1. Obtain a posterior sample $\boldsymbol{\psi}^{(1)}, \ldots, \boldsymbol{\psi}^{(R)}$, where $\boldsymbol{\psi} = (\mu, \sigma, \xi, \theta)$, using the log–likelihood ℓ from the maxima method;
2. calculate $C^{(r)} = \lfloor \theta^{(r)} N \rfloor + 1$, $r = 1, \ldots, R$;
3. find $\kappa^{(r)}$, the $C^{(r)}$–th largest inter–exceedance time;

Bayesian inference

4. use each $\kappa^{(r)}$, $r = 1, \ldots, R$, as the declustering interval to implement a full cluster identification procedure based on runs declustering;
5. use each set of identified clusters found using $\kappa^{(r)}$, $r = 1, \ldots, R$, to estimate any other cluster characteristic, say $H^{(r)}$, and so obtain draws from the (approximate) posterior distribution for that functional also.

Implementation

We specify independent, highly uninformative prior distributions for the GEV parameters; specifically, we use

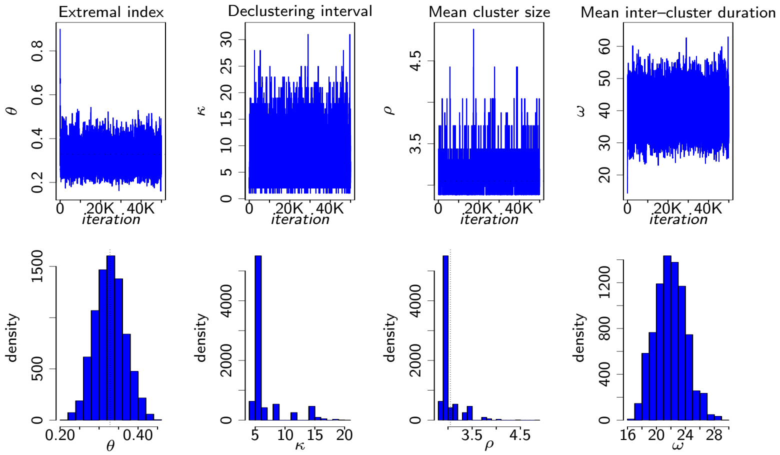
$$\begin{aligned} \mu &\sim N\left(0, 10^4\right), \\ \log(\sigma) &\sim N\left(0, 10^4\right) \qquad \text{and} \\ \xi &\sim N\left(0, 10^2\right). \end{aligned}$$

In the absence of any useful prior information about the extremal index, we use:

$$\theta \sim U(0,1).$$

We use each posterior draw for θ to obtain a corresponding draw for κ; each κ implements a full declustering procedure from which we can observe the (posterior) distribution for *any* cluster characteristic!

Simulated data



Simulated data

Some numerical summaries (after burn–in)

	θ (= 0.062)	κ	ρ	ω
Posterior mean (s.d.)	0.065 (0.050)	14.035 (4.002)	14.648 (0.707)	53.343 (10.546)
95% credible interval	(0.031, 0.165)	(6, 23)	(12.203, 18.321)	(32.222, 77.372)
m.l.e. (asymp. s.e.)	0.058 (0.052)	—	13.317 (1.193)	—
	θ (= 0.328)	κ	ρ	ω
Posterior mean (s.d.)	0.319 (0.048)	5.974 (3.678)	3.048 (0.237)	39.662 (5.440)
95% credible interval	(0.225, 0.416)	(2, 16)	(2.618, 3.420)	(29.715, 50.867)
m.l.e. (asymp. s.e.)	0.297 (0.047)	—	3.361 (0.233)	—

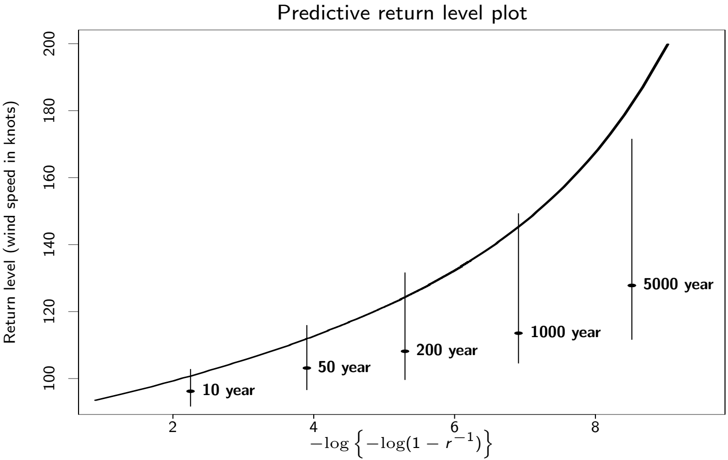
Bradfield wind speed data

Some numerical summaries for January (after burn–in)

	θ	κ	Mean storm length	Mean time between storms
Posterior mean (s.d.)	0.243 (0.047)	5.266 (5.840)	4.924 (0.637)	82.747 (15.271)
95% credible interval	(0.162, 0.347)	(2, 24)	(4.289, 6.246)	(56.435, 117.259)
m.l.e. (asymp. s.e.)	0.207 (0.042)	—	4.833 (0.578)	—

Bradfield wind speed data

Return level inference



References

Ancona–Navarrete, M.A. and Tawn, J.A. (2000). A Comparison of Methods for Estimating the Extremal Index. *Extremes*, **3**, pp. 5—38.

Fawcett, L. and Walshaw, D. (2008). Bayesian Inference for Clustered Extremes. *Extremes*, **11**, pp. 217—233.

Fawcett, L. and Walshaw, D. (2006). Markov chain models for extreme wind speeds. *Environmetrics*, **17**, 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.

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

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

Smith, R.L., Coles, S.G. and Tawn, J.A. (1997). Markov chain models for threshold exceedances. *Biometrika*, **84**, pp. 249—268.