Chapter 4

Dependent extremes

The threshold–based approach to modelling extremes has obvious practical advantages over the more traditional ‘block maxima’ approach.

Since extremes are (by their very nature) scarce, a modelling procedure which allows the inclusion of more data in the analysis has got to be a good thing.

Indeed, the whole point of a threshold–based analysis is that we include *all* extremes in the analysis — extreme in the sense that the observations used have all exceeded some pre–determined threshold u.

Advantages

- Data pre–processing dead easy
- Estimates with **smaller standard errors** (including return levels!)
- **Narrower confidence intervals** (standard or profile–likelihood based)

Disadvantages

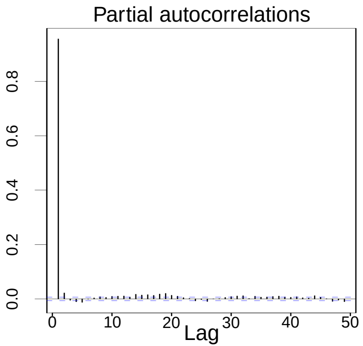
- **Temporal dependence** (e.g. temperature extremes, wind gusts,...)
- **Non–stationarity** (e.g. trends and seasonal effects)

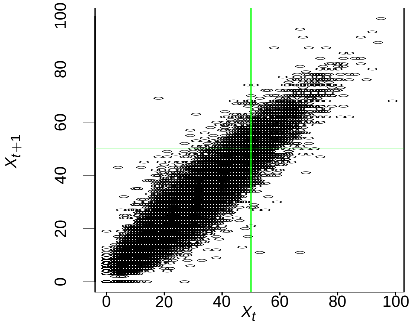
Hourly maximum wind gusts (in **knots**) were collected at High Bradfield, in the Peak District, over a period of 10 years from January 1st 2003 to December 31st 2012.

This gives total of 87,672 observations (including any missing values: 81,835 after the missing values have been removed).

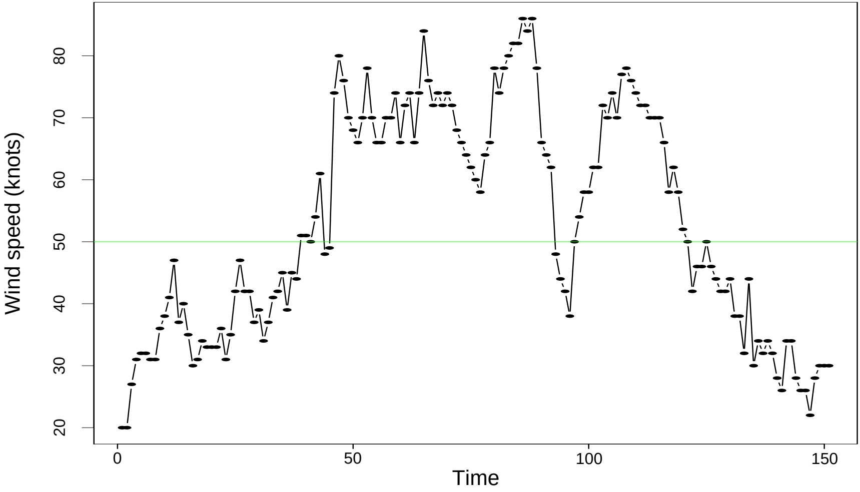
Suppose these hourly observations are in the vector `brad` in R; the series without the missing values has been stored in the vector `brad2`.

4.2 Example: Hourly wind speeds at High Bradfield





4.2 Example: Hourly wind speeds at High Bradfield



4.2 Example: Hourly wind speeds at High Bradfield

It is clear that we have (fairly strong) dependence between consecutive wind speeds recordings – even at extreme levels.

How can we proceed?

- **Ignore** this dependence – proceed as normal?

  Possible consequence: Standard errors under–estimated as there are fewer independence observations than the likelihood assumes

- **Model** this dependence – see Chapter 6
- **Filter out** this dependence

The book by **Leadbetter *et al.* (1983)** considers, in great detail, properties of extremes of dependent processes.

A key result often used is '**Leadbetter's $D(u_n)$ condition**', which ensures that long–range dependence is sufficiently weak so as not to affect the asymptotics of an extreme value analysis.

This condition is stated more formally in the definition below.

Definition *(Leadbetter's $D(u_n)$ condition)*

A stationary series $\tilde{X}_1, \tilde{X}_2, \ldots$ is said to satisfy the $D(u_n)$ condition if, for all $i_1 < \ldots < i_p < j_1 < \ldots < j_q$ with $j_1 - i_p > l$,

$$\Big| \Pr\left\{\tilde{X}_{i_1} \leq u_n, \ldots, \tilde{X}_{i_p} \leq u_n, \tilde{X}_{j_1} \leq u_n, \ldots, \tilde{X}_{j_q} \leq u_n\right\} \\ -\Pr\left\{\tilde{X}_{i_1} \leq u_n, \ldots, \tilde{X}_{i_p} \leq u_n\right\} \Pr\left\{\tilde{X}_{j_1} \leq u_n, \ldots, \tilde{X}_{j_q} \leq u_n\right\} \Big| \leq \alpha(n, l),$$

where $\alpha(n, l) \to 0$ for some sequence l_n such that $l_n/n \to 0$ as $n \to \infty$.

For sequences of independent variables, the difference in probabilities in the above expression is exactly zero for *any* sequence u_n.

More generally, we will require that the $D(u_n)$ condition holds only for a specific sequence of thresholds u_n that increases with n.

For such a sequence, the $D(u_n)$ condition ensures that, for sets of variables that are far enough apart, the difference in probabilities expressed in (4.1), while not zero, is sufficiently close to zero to have no effect on the limit laws for extremes.

Theorem *(Extremes of dependent sequences)*

Let $\tilde{X}_1, \tilde{X}_2, \ldots$ be a stationary series satisfying Leadbetter's $D(u_n)$ condition, and let $\tilde{M}_n = \max\{\tilde{X}_1, \ldots, \tilde{X}_n\}$.

Now let $X_1, X_2, \ldots$ be an **independent** series with X having the same distribution as $\tilde{X}$, and let $M_n = \max\{X_1, \ldots, X_n\}$. Then if M_n has a non–degenerate limit law given by $\Pr\{(M_n - b_n)/a_n \leq x\} \to G(x)$, it follows that

$$\Pr\left\{(\tilde{M}_n - b_n)/a_n \leq x\right\} \to G^{\theta}(x) \qquad (4.2)$$

for some $0 \leq \theta \leq 1$.

The parameter θ is known as the **extremal index**, and quantifies the extent of extremal dependence

- $\theta = 1$: completely independent process
- $\theta \to 0$: increasing levels of (extremal) dependence

Since *G* in the above theorem is necessarily an extreme value distribution, and due to the **max–stability** property (see Leadbetter *et al.*, 1983), then the distribution of maxima in processes displaying short–range temporal dependence (characterised by the extremal index θ) is also a GEV distribution.

The powering of the limit distribution by θ only affects the location and scale parameters of this distribution.

What does this all mean in **practical terms**?

- If maxima of a stationary series converge in distribution (and we know they do – the GEV), and
- if Leadbetter's $D(u_n)$ condition holds (i.e. long–range dependence is negligible), then
- the limit distribution is related to that of the independent series...
- ...in fact, it's $G^{\theta}(x)$, where G is the GEV for the independent series (with parameters μ, σ and ξ)

Further, it can be shown that

$$\begin{aligned} G^{\theta}(x) &= \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}^{\theta} \\ &= \exp\left\{-\left[1+\xi\left(\frac{x-\mu^*}{\sigma^*}\right)\right]^{-1/\xi}\right\}, \end{aligned}$$

where $\mu^* = \mu - \frac{\sigma}{\xi}\left(1-\theta^{-\xi}\right)$ and $\sigma^* = \sigma\theta^{\xi}$.

We can proceed as we did in Chapter 2 – even in the presence of dependence...

... provided long range dependence is weak (i.e. **Leadbetter's $D(u_n)$ condition** can be assumed)

The only difference is the parameters of the GEV:
$(\mu, \sigma, \xi) \longrightarrow (\mu^*, \sigma^*, \xi)$.

Since we estimates these parameters anyway, who cares?!

4.3.2 Modelling threshold exceedances

Though the modelling procedure for fitting the GEV to a set of annual maxima is unchanged for series which display short–term temporal dependence, some revision is needed of the threshold exceedance approach.

If all threshold exceedances are used in our analysis, and the GPD fitted to the set of threshold excesses, the likelihoods we use will be incorrect since they assume independence of sample observations.

4.3.2 Modelling threshold exceedances

In practice, several techniques have been developed to circumvent this problem, including:

1. filtering out an (approximately) independent set of threshold exceedances
2. fitting the GPD to *all* exceedances, ignoring dependence, but then appropriately adjusting the inference (usually an inflation of standard errors) to take into account the reduction in information
3. Explicitly modelling the temporal dependence in the process

4.3.2 Declustering

Since the mid–1990s, various methods for **declustering** a series of extremes, to extract a set of independent extremes, have been discussed in the literature.

The most natural, commonly–used method of declustering is that of **runs declustering**. This is how it works:

1. Choose an auxiliary ‘declustering parameter’ (which we call κ)
2. A cluster of threshold excesses is then deemed to have terminated as soon as at least κ consecutive observations fall below the threshold
3. Go through the entire series identifying clusters in this way
4. The maximum (or ‘peak’) observation from each cluster is then extracted, and the GPD fitted to the set of cluster peak excesses.

This approach is often referred to as the **peaks over threshold** approach (POT, **Davison and Smith, 1990**) and is widely accepted as the main pragmatic approach for dealing with clustered extremes.

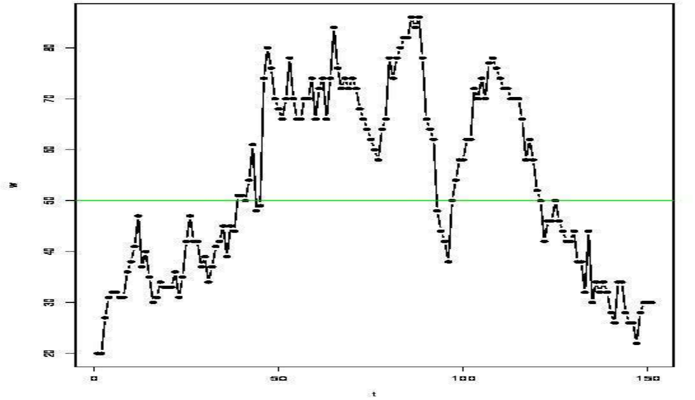
Easy to implement, but if:

- κ is too small, the cluster peaks will not be far enough apart to safely assume independence
- κ is too large, there will be too few cluster exceedances on which to form our inference

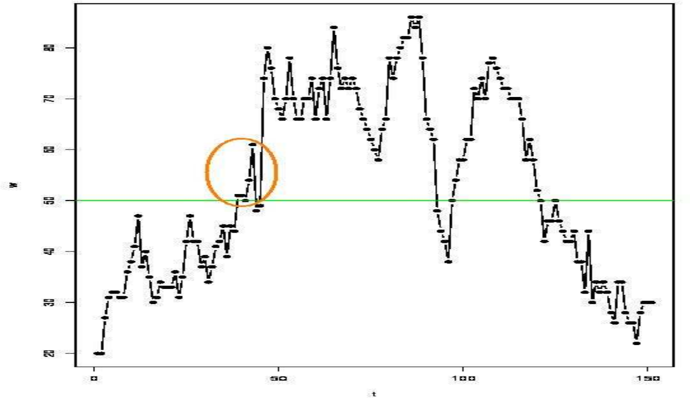
It has also been shown (**Fawcett and Walshaw, 2012**)[1] that parameter estimates can be sensitive to the choice of κ, and κ is all too often chosen arbitrarily.

[1] Estimating return levels from serially dependent extremes, *Environmetrics* **23**(3), pp 272–283

Using $\kappa = 2$:

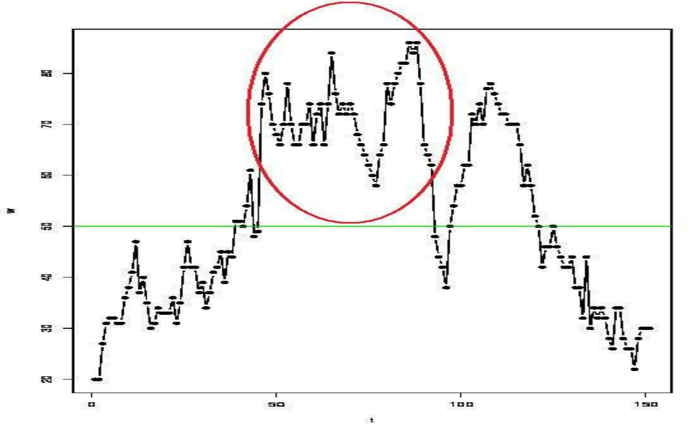


Using $\kappa = 2$:



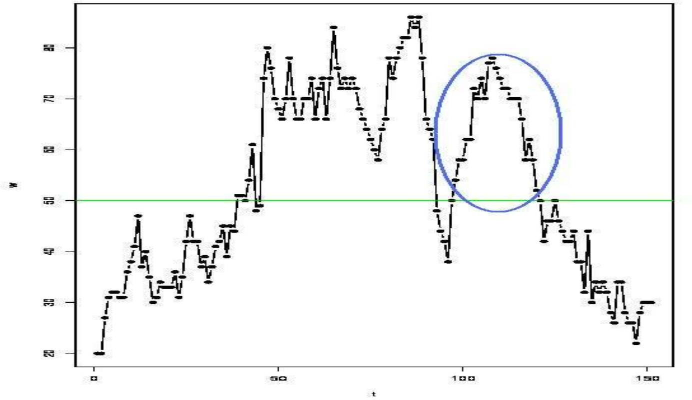
4.3.2 Declustering

Using $\kappa = 2$:



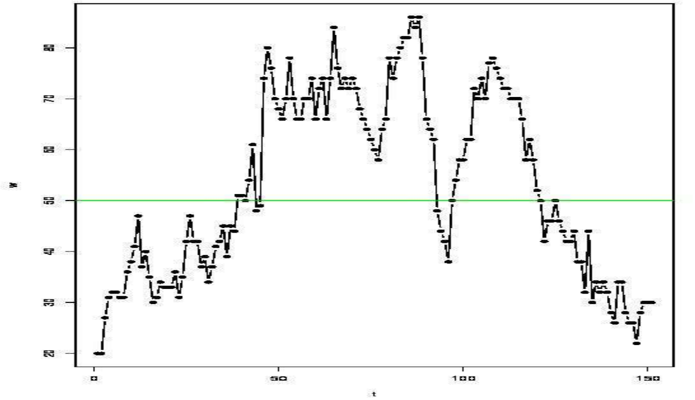
4.3.2 Declustering

Using $\kappa = 2$:

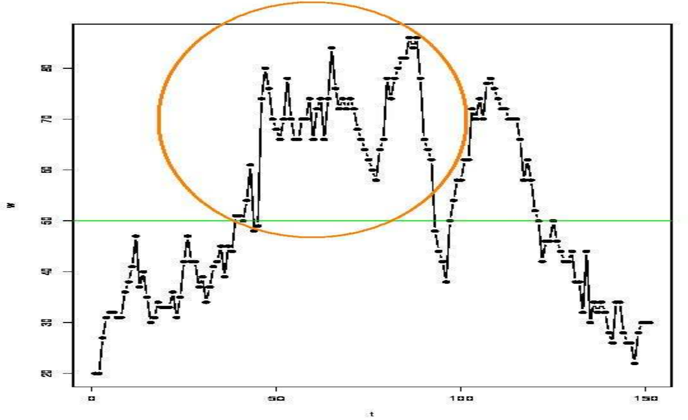


4.3.2 Declustering

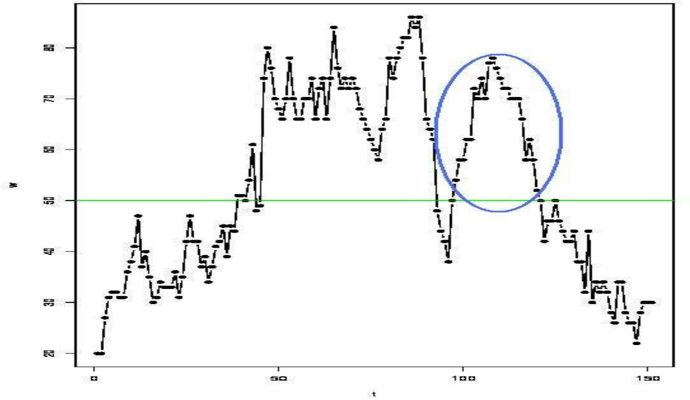
Using $\kappa = 4$:



Using $\kappa = 4$:

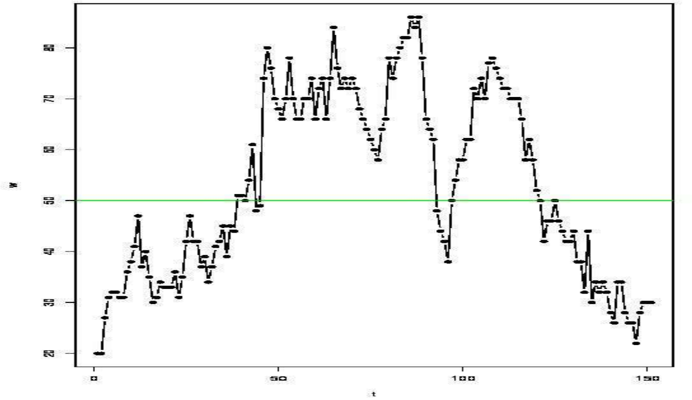


Using $\kappa = 4$:



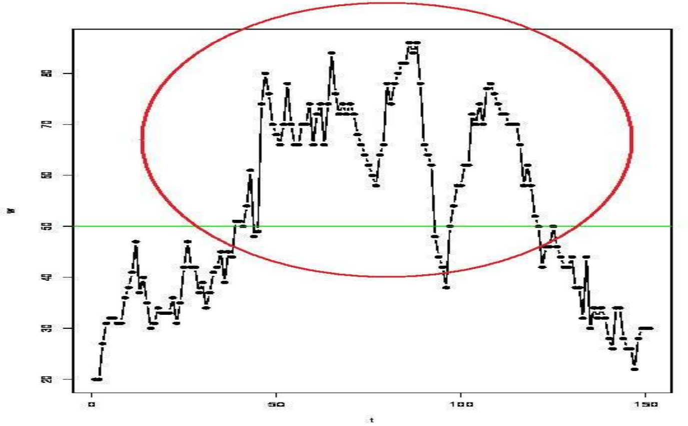
4.3.2 Declustering

Using $\kappa = 10$:



4.3.2 Declustering

Using $\kappa = 10$:



Suppose we use $\kappa = 10$.

Obviously, we wouldn't want to identify clusters by hand for the full Bradfield wind speed series (recall that we have 10 years of hourly observations!).

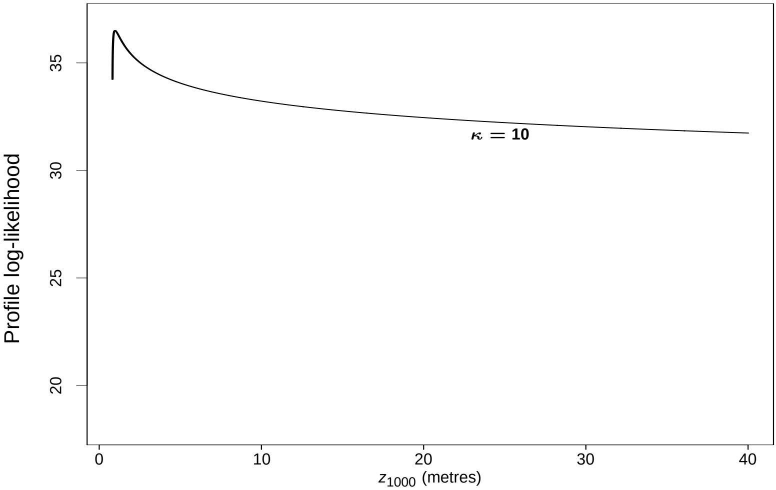
Unfortunately, there is no function in ismev to perform runs declustering.

I have written the following R code to perform this declustering, for $\kappa = 10$.

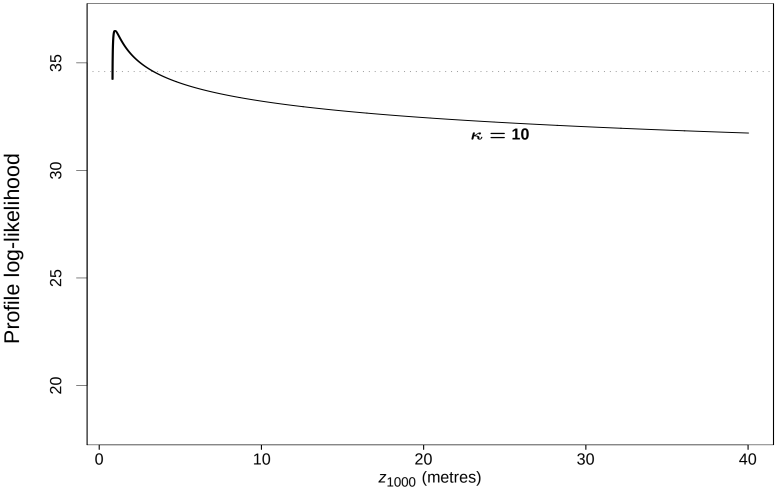
4.3.2 Declustering

Class demonstration in R

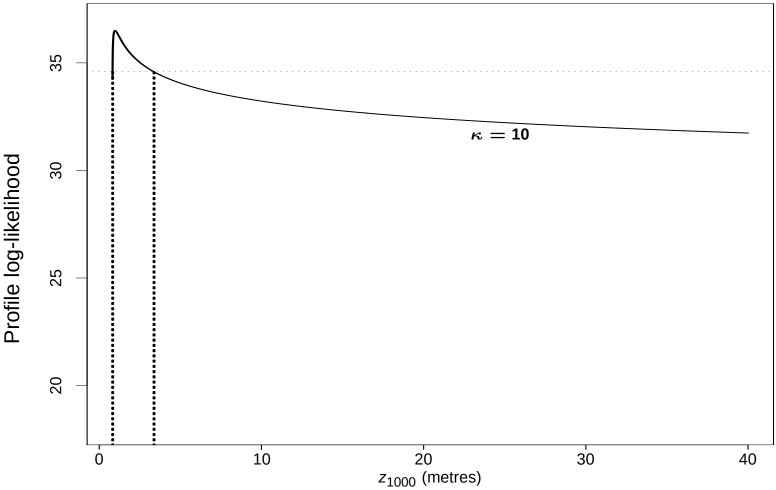
4.4 Words of warning (*Fawcett & Walshaw, 2012*)



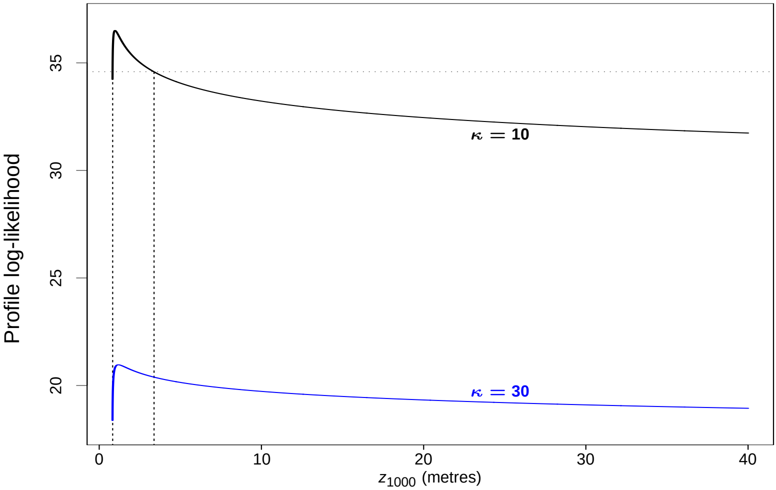
4.4 Words of warning (*Fawcett & Walshaw, 2012*)



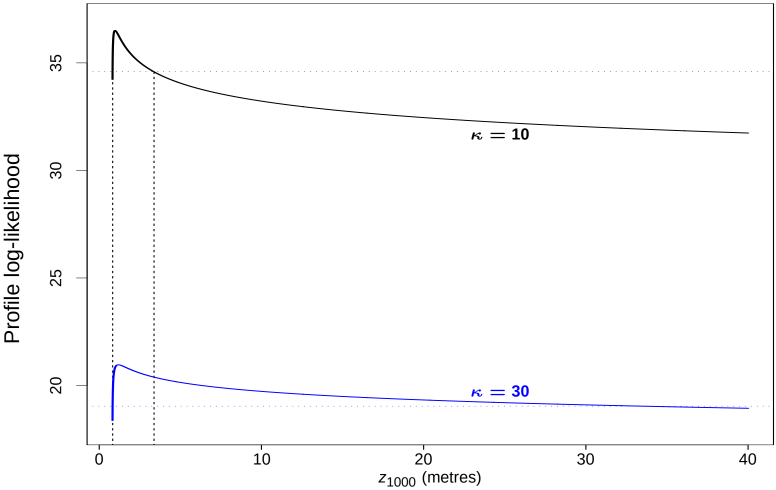
4.4 Words of warning (*Fawcett & Walshaw, 2012*)



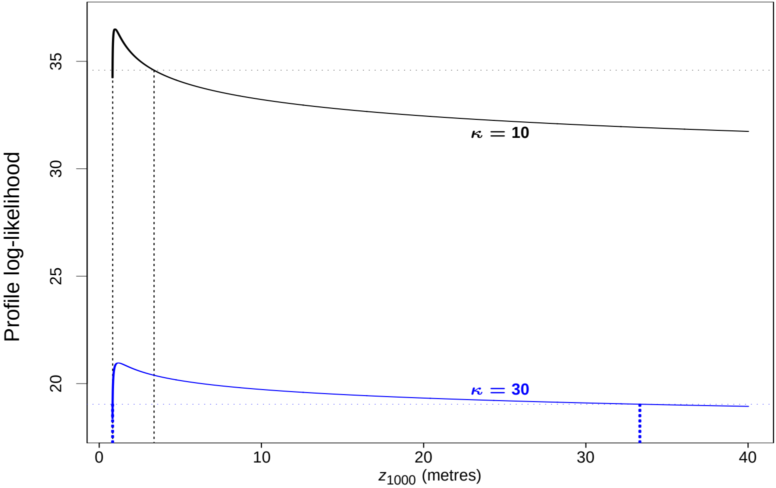
4.4 Words of warning (*Fawcett & Walshaw, 2012*)



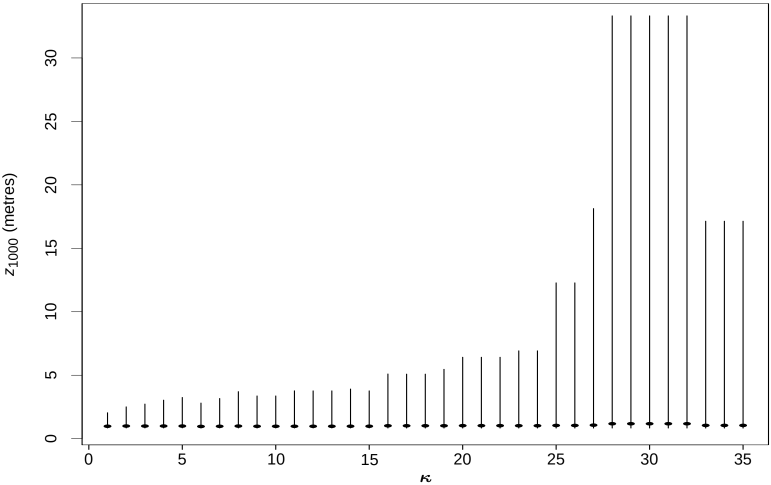
4.4 Words of warning (*Fawcett & Walshaw, 2012*)



4.4 Words of warning (*Fawcett & Walshaw, 2012*)



4.4 Words of warning (*Fawcett & Walshaw, 2012*)



4.4 Words of warning

- Similar to the block maxima approach, this is **wasteful of data**! We are throwing away observations we have identified as extreme!
- How do we choose κ? Estimates of GPD parameters – and return levels/confidence intervals for return levels – can be sensitive to the choice of κ (often hugely so)
- What if we are interested in the dependence?