Chapter 5

Non–stationary extremes

5.1 Introduction

In the context of environmental processes, it is common to observe **non–stationarity**:

- different **seasons** having different climate patterns
- **long term trends** owing to climate change

The models from chapters 2 and 3 assume that the observations used are IID!

5.1 Introduction

We examined the effects of **dependence** in Chapter 4:

- Likely that consecutive extremes of wind or rain will be **correlated**
- Block maxima: Doesn't matter – provided we can assume long–range independence (Leadbetter's $D(u_n)$ condition), then the GEV still appropriate
- Threshold exceedances: Not quite so straightforward, although we can **decluster**, and fit the GPD to **cluster peak excesses**? Problems?

5.1 Introduction

To date, no general theory for **non–stationary extremes** has been established.

In practice, it is common to adopt pragmatic ‘workarounds’ based on the type of non–stationarity observed.

For this reason, in this Chapter we will give some specific examples of how practitioners have dealt with non–stationarity in recent work and publications.

Recall **question 4** in **problems sheet 2**, which investigated the use of the r–largest order statistics model for extreme sea levels at Venice (1961–2011).

These data are available to download from the course webpage.

Demonstration in R

One way of capturing the trend observed in the Venice annual maxima is to allow the GEV location parameter μ to vary across time.

From Figure 5.1, a **simple linear trend** in time seems plausible for our annual maximum sea levels X, and so we could use the model

$$X_t \sim \text{GEV}(\mu(t), \sigma, \xi),$$

where

$$\mu(t) = \beta_0 + \beta_1 t \tag{5.1}$$

and t is an indicator of year.

5.2 Annual maximum sea levels in Venice

$$\mu(t) = \beta_0 + \beta_1 t :$$

- Variations over time modelled as a linear trend in the **location parameter**
- β_1 represents the **slope** – in this case, the annual rate of change in sea–level maxima at Venice
- The time–homogeneous model is a special case of this time dependent model, with $\beta_1 = 0$...
- ...since this is **nested** within the model which allows for a time dependence, the deviance statistic can be used to formally compare models (see later)

Recall the log–likelihood function for the GEV:

$$\ell(\mu, \sigma, \xi; \boldsymbol{x}) = -m\log\sigma - (1/\xi + 1)\sum_{i=1}^{m} \log\left[1 + \xi\left(\frac{x_i - \mu}{\sigma}\right)\right]_+ - \sum_{i=1}^{m}\left[1 + \xi\left(\frac{x_i - \mu}{\sigma}\right)\right]_+^{-1/\xi},$$

where m is the number of block maxima $x_1, x_2, \ldots, x_m$.

We simply replace μ in the above expression with equation (5.1), giving

$$\begin{aligned}\ell(\beta_0, \beta_1, \sigma, \xi; \boldsymbol{x}, t) &= -m\log\sigma \\ &\quad -(1/\xi + 1)\sum_{i=1}^{m} \log\left[1 + \xi\left(\frac{x_i - (\beta_0 + \beta_1 t_i)}{\sigma}\right)\right]_+ \\ &\quad -\sum_{i=1}^{m}\left[1 + \xi\left(\frac{x_i - (\beta_0 + \beta_1 t_i)}{\sigma}\right)\right]_+^{-1/\xi},\end{aligned}$$

with the usual replacement when $\xi = 0$.

- *Could* maximise $\ell(\beta_0, \beta_1, \sigma, \xi; \boldsymbol{x}, t)$ 'from first principles' using `nlm`
- However, easier to do in the ismev package!

Demonstration in R

5.2.1 Parameter estimation

Thus, we have

$$\hat{\beta}_0 = 96.986(4.249) \quad \hat{\beta}_1 = 0.564(0.139)$$

$$\hat{\sigma} = 14.584(1.578) \quad \hat{\xi} = -0.027(0.083)$$

(with standard errors in parentheses).

This gives:

$$\hat{\mu}(t) = 96.986 + 0.564t,$$

giving an estimated increase in maximum sea levels at Venice of about 0.564cm per year.

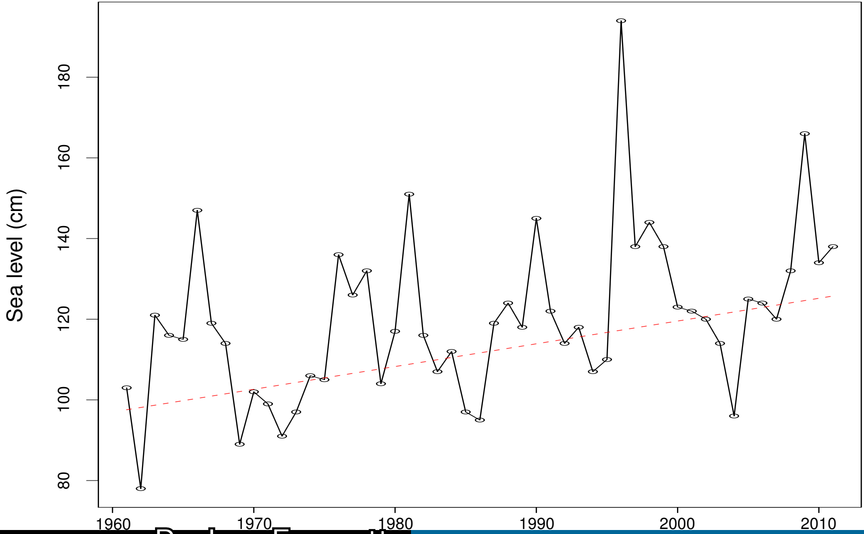
For example, the estimated value for μ in the year 2013 would be

$$\hat{\mu}(53) = 96.986 + 0.564 \times 53 = 126.878;$$

we could, of course, use the delta method to obtain the corresponding standard error.

Using this simple linear model, we can estimate μ for $t = 1, 2, \ldots, 51$ to cover the years for which we have data (i.e. 1961, 1962, . . . , 2011).

5.2.1 Parameter estimation



5.2.2 Model choice

Fitting a completely stationary model to the set of annual maxima, as you did in question 4(a) of problems sheet 2, gives:

```
gev.fit(venice.anmax)

$conv
[1] 0

$nllh
[1] 222.7145

$mle
[1] 111.09925486 17.17548761 -0.07673265

$se
[1] 2.6280071 1.8033672 0.0735214
```

5.2.2 Model choice

Questions:

- Is the non–stationary model **worthwhile?**
- Is the trend we observe in Figure 5.1 **significant**?
- In other words, does the non–stationary model provide an **improvement** in fit over the simpler model shown here?

5.2.2 Model choice

We can use a version of the result in Section 2.4 (page 41) to address this question.

Generally, maximum likelihood estimation of nested models leads to a simple test procedure of one model against the other.

With models $\mathcal{M}_0 \subset \mathcal{M}_1$, we define the deviance statistic as

$$D = 2\left\{\ell_1(\mathcal{M}_1) - \ell_0(\mathcal{M}_0)\right\},$$

where $\ell_1(\mathcal{M}_1)$ and $\ell_0(\mathcal{M}_0)$ are the maximised log–likelihood under models $\mathcal{M}_1$ and $\mathcal{M}_0$ respectively.

Asymptotically, $D \sim \chi^2_k$, where $k = \dim(\mathcal{M}_1) - \dim(\mathcal{M}_0)$.

5.2.2 Model choice

Formally,

$$H_0 : \mathcal{M}_1 = \mathcal{M}_0.$$

Thus, large values of D suggest that the increase in model size has been worthwhile, and that model $\mathcal{M}_1$ explains substantially more of the variation in the data than $\mathcal{M}_0$.

5.2.2 Model choice

Suppose

- $\mathcal{M}_1$: model allowing for **linear trend** in μ
- $\mathcal{M}_0$: **stationary** model

Thus,

$$D = 2\left\{-216.0626 - (-222.7145)\right\} = 13.3038.$$

The difference in dimensionality is 1, so the critical value for our test is

$$\chi_1^2(0.05) = 3.841.$$

Relative to this, our deviance statistic is large: the model which allows for a linear trend in μ explains substantially more of the variation in the data than does the simpler stationary model.

We could, of course, use this method to check for a more complex association through time.

For example, to check for a **quadratic trend** we might use a model $\mathcal{M}_2$ with the following form for μ:

$$\mu(t) = \beta_0 + \beta_1 t + \beta_2 t^2.$$

Demonstration in R

Comparing model $\mathcal{M}_2$ with model $\mathcal{M}_1$ gives

$$D = 2\left\{-216.0555 - (-216.0626)\right\} = 0.0142,$$

which is *small* compared to $\chi_1^2(0.05) = 3.841$.

Thus, allowing for a quadratic dependence in time does not improve on our model which allows for a linear trend through time, and so we would reject model $\mathcal{M}_2$.

5.2.3 Model diagnostics

Before estimating return levels, we should check the goodness–of–fit of our model which allows for a linear trend in μ.

The lack of homogeneity in the distributional assumptions for each observation, however, mean some modification of the standard procedures (e.g. **probability plots** and **quantile plots**) is required.

For example, for the Venice annual maximum sea levels, we have

$$X_t \sim \text{GEV}(\mu(t), \sigma, \xi), \quad t = 1, 2, \ldots, 51,$$

giving a different GEV in each year indicated by t.

What we need to do is standardise so that we can assume the X_t *are* IID across all years t.

Usually, the set of non–stationary annual maxima are transformed to a common Gumbel distribution with distribution function $F(y) = \exp\{-e^{-y}\}$.

We can obtain the required transformation by equating $F(y_t)$ to $\text{GEV}(x_t; \hat{\mu}(t), \hat{\sigma}, \hat{\xi})$ and solving for y_t:

$$\begin{aligned}
\exp\left\{-e^{-y_t}\right\} &= \exp\left\{-\left[1+\hat{\xi}\left(\frac{x_t-(\hat{\beta}_0+\hat{\beta}_1 t)}{\hat{\sigma}}\right)\right]^{-1/\hat{\xi}}\right\} \\
e^{-y_t} &= \left[1+\hat{\xi}\left(\frac{x_t-(\hat{\beta}_0+\hat{\beta}_1 t)}{\hat{\sigma}}\right)\right]^{-1/\hat{\xi}} \\
y_t &= \frac{1}{\hat{\xi}}\log\left[1+\hat{\xi}\left(\frac{x_t-(\hat{\beta}_0+\hat{\beta}_1 t)}{\hat{\sigma}}\right)\right],
\end{aligned}$$

This gives gives the following transformation to common Gumbel margins for our Venice annual maxima:

$$y_t = -\frac{1}{0.027}\log\left[1 - 0.027\left(\frac{x_t - (96.986 + 0.564t)}{14.584}\right)\right].$$

Demonstration in R

Now that we have transformed the original data to a single common distribution, we can apply the standard graphical diagnostics.

For example, we can compare the **empirical probabilities** and **quantiles** of y_t to their **theoretical** counterparts from the Gumbel distribution.

Fortunately, the usual command in ismev can be used to produce these plots for us.

Demonstration in R

5.2.4 Return level estimation

Recall Equation (2.10) (**memorise!**) from Chapter 2 for estimating return levels from the GEV:

$$\hat{z}_r = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}\left[\left(-\log\left(1 - r^{-1}\right)\right)^{-\hat{\xi}} - 1\right].$$

Since we have a time–varying location parameter $\hat{\mu}(t) = \hat{\beta}_0 + \hat{\beta}_1 t$, we will clearly have time–varying estimates of return levels $\hat{z}_r(t)$.

For example, an estimate of the sea level we might expect to see in Venice once every 100 years is given by

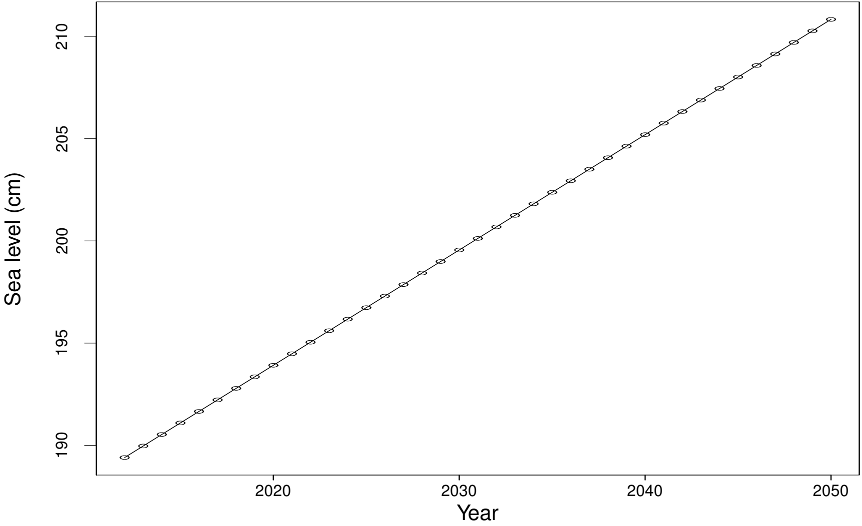
$$\hat{z}_r(t) = (96.986 + 0.564t) - \frac{14.584}{0.027}\left[\left(-\log\left(1 - 100^{-1}\right)\right)^{0.027} - 1\right].$$

Figure 5.4 shows how we might expect this estimate to vary for $t = 52, 53, \ldots$, i.e. for the years 2012, 2013, ….

We could treat these as forecasts of the 100–year return levels as we move through time.

Obviously, such forecasts will assume the linear trend for μ continues beyond the range of data we have observed and will, of course be subject to error (which we can estimate by constructing point–wise 95% confidence intervals using the profile log–likelihood, for example).

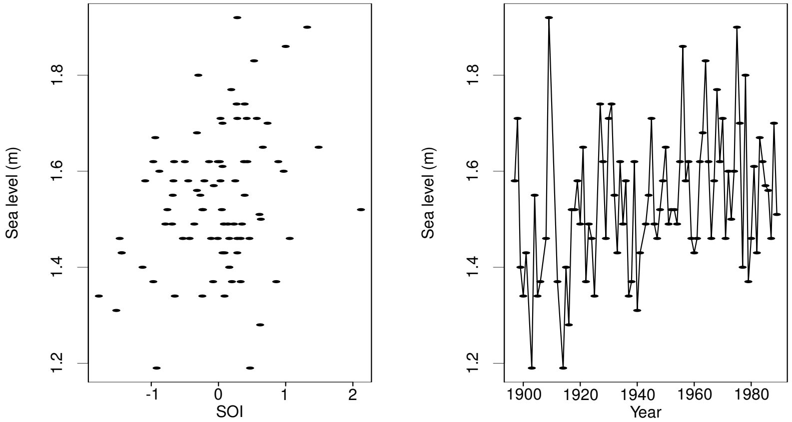
5.2.4 Return level estimation



A different situation which could use the same approach as that in the previous section is where the extremal behaviour of a series is related to another variable, rather than time.

For example, studies have revealed a link between annual maximum sea levels at Fremantle, Australia, and the mean value of the **Southern Oscillation Index** (SOI, an indicator of meteorological volatility due to effects such as El Niño).

5.3 Sea levels and the SOI



Thus, the following model for X_t, the annual maximum sea level at Fremantle in year t, might be suitable:

$$X_t \sim \text{GEV}(\mu(t), \sigma, \xi),$$

where

$$\mu(t) = \beta_0 + \beta_1 \text{SOI}(t), \qquad (5.3)$$

where SOI(t) denotes the mean value of SOI in year t.

However, the plot in the right–hand–side of Figure 5.5 also reveals a possible trend in sea levels through time, suggesting

$$\mu(t) = \beta_0 + \beta_1 t, \tag{5.4}$$

where $t = 1, 2, \ldots,$ as in Example 5.2.

We can combine Equations (5.3) and (5.4) to allow for a dependence on time *and* SOI by letting

$$\mu(t) = \beta_0 + \beta_1 \text{SOI}(t) + \beta_2 t; \tag{5.5}$$

however, a technique of **forward selection** should be used to check whether or not any of Equations (5.3), (5.4) or (5.5) give significant improvement over the stationary model.

Class demonstration in R

Recall Section 3.2 in which we modelled rainfall extremes in New York using a threshold–based approach.

The Generalised Pareto distribution was applied to rainfall exceedances over a threshold of $u = 30$mm, giving estimates of the scale and shape as

$$\hat{\sigma} = 7.44(0.958) \quad \hat{\xi} = 0.184(0.101)$$

(respectively).

Recall that the GPD(σ, ξ) arises from the GEV(μ, σ, ξ), where the GPD scale parameter is a function of the GEV location and shape parameters.

Thus, attempting to model any trend in our threshold exceedances is usually done through linear modelling of the scale parameter σ (the GPD doesn't have a location parameter *per se*).

Since the scale parameter σ must be positive, we might choose to model a trend through time as

$$\sigma(t) = \exp\{\beta_0 + \beta_1 t\}, \tag{5.6}$$

where t is, once again, an indicator of time.

Class demonstration in R

5.5 Generalisation

With reference to the examples discussed so far, we could model non–stationarity through *any* of the parameters in our extremal model.

For example, take a non–stationary GEV model to describe the distribution of X_t, for $t = 1, 2, \ldots, m$:

$$X_t \sim \text{GEV}(\mu(t), \sigma(t), \xi(t)),$$

where each of the model parameters have an expression in terms of a parameter vector $\boldsymbol{\beta}$ and some covariates.

The likelihood is then

$$L(x_t; \boldsymbol{\beta}) = \prod_{i=1}^{m} g(x_t; \mu(t), \sigma(t), \xi(t)),$$

where g is the GEV density function.

From this, we can form the log–likelihood, and then maximise in the usual way (for example, using `nlm` in R).

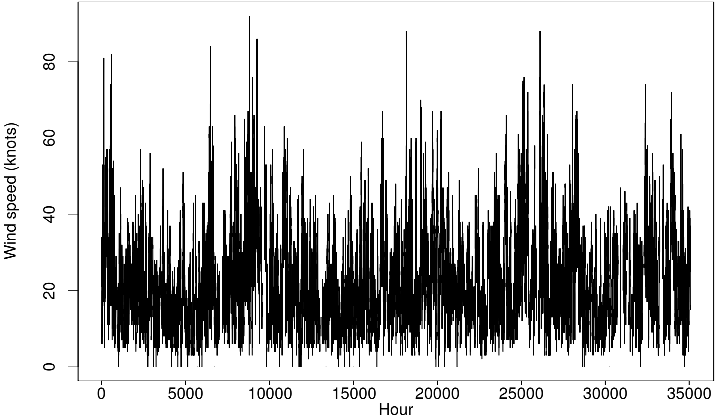
In terms of threshold exceedances Y_t, $t = 1, 2, \ldots, k$, we could replace the GEV with the GPD:

$$Y_t \sim \text{GPD}(\sigma(t), \xi(t)),$$

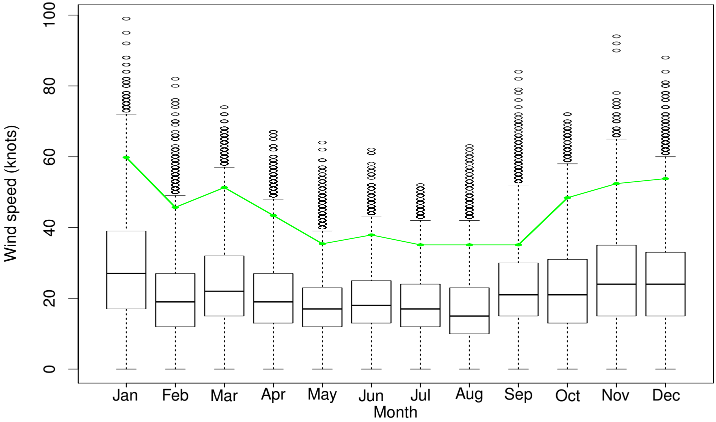
with σ being defined as in Equation (**5.6**) to retain the positivity of the GPD scale.

- **Temporal dependence:** block maxima approach – still OK; threshold approach – decluster using POT
- **Trend:** Use the deviance statistic to check for trends in GEV location parameter/GPD scale parameter
- **Covariates:** Use the deviance statistic, as above
- **Seasonal variability**

5.6 Wind speed extremes at High Bradfield



5.6 Wind speed extremes at High Bradfield



5.6.1 Single season approach

Under the single season approach, an extremal model is fitted to the extremes of an environmental process from the season which gives rise to the 'most extreme' extremes.

- Bradfield wind speeds – use January only
- Some appeal: **easy to implement**, focuses on the **most extreme** extremes
- But: **wasteful** of data – especially if we decluster as well!

5.6.2 Seasonal piecewise approach

It is usual in strongly seasonal climates for the occurrence of extreme winds to be confined to a certain part of the yearly cycle.

Example: U.K. $\longrightarrow$ unusual for wind damage to occur outside the period **October–March**.

5.6.2 Seasonal piecewise approach

The seasonal variation observed at Bradfield might be expected: Prolonged, **anticyclonic** periods more prevalent during June, July and August.

There is only a point to modelling the extremes which occur in summer months if we believe that they can help us understand what happens in winter months, where genuinely large events *can* occur.

5.6.2 Seasonal piecewise approach

For this to be realistic, we must assume that the **same mechanism** is responsible for the generation of large gusts throughout the year $\longrightarrow$ it is just the **scale** of this mechanism which changes.

Indeed, in temperate climates (such as that of the U.K.), the same alternating sequence of anticyclones and depressions leads to most of the storms which occur throughout the year.

The seasonal variability comes from the **severity** of these systems.

5.6.2 Seasonal piecewise approach

Taking the **calendar month** as our seasonal unit:

- For the U.K., experience suggests this might be OK
- Strikes a good balance between:
  - reflecting the continuous nature of seasonal changes in climate
  - retaining a substantial amount of data for analysis within each season

5.6.2 Seasonal piecewise approach

Month (m)	u_m	n_m	$\hat{\sigma}_m$	$\hat{\xi}_m$
1	55.341	28	21.373 (5.358)	−0.420 (0.183)
2	41.531	24	15.130 (4.635)	−0.226 (0.233)
3	48.100	29	23.277 (6.316)	−0.894 (0.259)
4	39.910	29	14.853 (4.448)	−0.440 (0.249)
5	31.943	46	9.456 (1.990)	−0.158 (0.147)
6	35.670	35	12.329 (2.592)	−0.409 (0.143)
7	32.290	36	12.517 (2.609)	−0.605 (0.161)
8	32.639	34	10.199 (2.361)	−0.203 (0.159)
9	33.232	49	18.772 (3.668)	−0.255 (0.138)
10	44.914	34	11.669 (3.533)	−0.274 (0.254)
11	48.394	33	14.991 (3.381)	−0.225 (0.149)
12	49.341	35	18.681 (4.229)	−0.416 (0.166)

In terms of **return level inference**, it would not make practical sense to have monthly varying estimates of the r–year return level z_r.

To include information from *all* months in our return level estimation procedure, we solve

$$\prod_{m=1}^{12} H(\hat{z}_r; \hat{\lambda}_{u_m}, \hat{\sigma}_m, \hat{\xi}_m) \quad = \quad 1 - \frac{1}{rn_y}$$

for $\hat{z}_r$, where H is the GPD distribution function and n_y is the (average) number of observations per year.

Thus, we need to solve

$$\prod_{m=1}^{12}\left\{1-\hat{\lambda}_u\left[1+\hat{\xi}\left(\frac{\hat{z}_r-u_m}{\hat{\sigma}_m}\right)\right]_+^{-1/\hat{\xi}_m}\right\} \;=\; 1-\frac{1}{rn_y}$$

for $\hat{z}_r$, that is

$$\prod_{m=1}^{12}\left\{1-\hat{\lambda}_u\left[1+\hat{\xi}\left(\frac{\hat{z}_r-u_m}{\hat{\sigma}_m}\right)\right]_+^{-1/\hat{\xi}_m}\right\}-\left(1-\frac{1}{rn_y}\right) \;=\; 0.$$

This equation cannot be solved analytically.

Rather, a numerical procedure must be used.

This is easy to do in R using the `uniroot` procedure.

5.6.2 Seasonal piecewise approach

```
f = function(z){
    r = 10
    ny = 365.25*24
    sigma = c(21.373,15.130,23.277,..., 18.681)
    xi = c(-0.420,-0.226,-0.894,-0.440,..., -0.416)
    u = c(55.341,41.531,48.1,...,49.341)
    nobs = c(31*24*10,28*24*7+29*24*3,...,31*24*10)
    lambda = c(28/nobs[1],24/nobs[2],...,35/nobs[12])
    component = vector('numeric',12)
    inner = vector('numeric',12)
      for(m in 1:12){
          inner[m] = max((1+xi[m]*((z-u[m])/sigma[m])),0)
          component[m] = 1-lambda[m]*((inner[m])**(-1/xi[m]))}
    answer = prod(component)-(1-1/(r*ny))
    return (answer)}
```

5.6.2 Seasonal piecewise approach

```
> uniroot(f,lower=0,upper=200)
$root
[1] 102.3291
$f.root
[1] -6.242917e-11
$iter
[1] 12
$estim.prec
[1] 6.103516e-05
```

5.6.2 Seasonal piecewise approach

	Return period (r years)			
	10	50	200	1000
$\hat{z}_r$ (st. err.)	102.33 (3.970)	104.59 (15.951)	106.21 (23.793)	108.89 (44.865)

For the standard errors, we use the delta method. However, we now have:

$$V = \begin{pmatrix} \frac{\hat{\lambda}_{u_1}(1-\hat{\lambda}_{u_1})}{N_1} & 0 & \dots & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots & \ddots & \vdots \\ \vdots & \ddots & \frac{\hat{\lambda}_{u_{12}}(1-\hat{\lambda}_{u_{12}})}{N_{12}} & 0 & \dots & 0 \\ 0 & \dots & 0 & v_{1,1} & \dots & v_{1,24} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & v_{24,1} & \dots & v_{24,24} \end{pmatrix},$$

where $v_{i,j}$ denotes the (i,j)–th term of the variance–covariance matrix of $\hat{\sigma}_m$ and $\hat{\xi}_m$, $m = 1, \dots, 12$.

Hence, by the delta method,

$$\text{Var}(\hat{z}_r) \approx \nabla z_r^T V \nabla z_r,$$

where

$$\nabla z_r^T = \left[\frac{\partial z_r}{\partial \lambda_{u_1}}, \ldots, \frac{\partial z_r}{\partial \lambda_{u_{12}}}, \frac{\partial z_r}{\partial \sigma_1}, \ldots, \frac{\partial z_r}{\partial \sigma_{12}}, \frac{\partial z_r}{\partial \xi_1}, \ldots, \frac{\partial z_r}{\partial \xi_{12}}\right],$$

evaluated at $(\hat{\lambda}_{u_1}, \hat{\sigma}_1, \hat{\xi}_1, \ldots, \hat{\lambda}_{u_{12}}, \hat{\sigma}_{12}, \hat{\xi}_{12})$.

5.6.2 Seasonal piecewise approach

As an aside, notice how the estimate of the 50–year return level wind speed here differs to that when we assumed stationarity in Section 4.3.2:

$$\text{Assuming stationarity} \quad : \quad \hat{z}_{50} = 101.53 \text{ knots}$$

$$\text{Seasonal piecewise} \quad : \quad \hat{z}_{50} = 104.59 \text{ knots}$$

Such discrepancies become more pronounced for longer–period return levels.

5.6.3 Smoothly varying seasonal parameters

Various authors (e.g. Fawcett and Walshaw (2006)) have investigated the use of continuously varying parameters for the GPD when seasonal variation is present.

For example, **Fourier forms** can be used to allow the GPD scale and shape to vary smoothly through time.

However, such analyses for the Bradfield wind speed data yielded little, if any, improvement over the seasonal piecewise approach.