Session 2. Dependence and non–stationarity

2.1 Extremes of dependent sequences

- 2.1.1 Maxima of stationary sequences
- 2.1.2 Modelling block maxima
- 2.1.3 Modelling threshold exceedances
- 2.1.4 Example

2.2 Trend

2.3 Seasonal variation

2. Dependence and non–stationarity

The asymptotic results introduced in Part 1 have assumed the underlying process to be **independent** and **identically distributed**

They also assume this process is **stationary**.

2. Dependence and non–stationarity

In practice, extreme value data – particularly environmental time series – exhibit some form of departure from this ideal. The most common forms are:

— Local **temporal dependence**, where successive values of the time series are dependent, but values farther apart are independent;

— Long term **trends**, where the underlying distribution changes gradually over time;

— **Seasonal variation**, where the underlying distribution changes periodically through time.

2. Dependence and non–stationarity

These departures can be handled through a combination of extending both the *theory* and the *modelling*.

However, theoretical results have generally been too specific to be of use in modelling data for which the form of non–stationarity is unknown.

Over the last decade or so, it has been more usual for practitioners to employ statistical procedures which allow the existing results to be applied.

In Part 2, we will consider some of these in detail.

2.1 Extremes of dependent sequences

For the types of data to which extreme value models are commonly applied, temporal independence is usually unrealistic:

- Extreme conditions often persist over several consecutive observations...
- ... which brings into question the appropriateness of models such as the GEV
- We haven't got time to consider the mathematics behind this...
- ... but we *will* consider a pivotal result which extends the theory presented in Part 1 to cover processes which display short–term temporal dependence

For the remainder of this section on dependent sequences, we shall assume that our process is *stationary*.

2.1 Extremes of dependent sequences

Dependence in stationary sequences can take many different forms.

With practical applications in mind, it is common to assume a condition that limits the extent of dependence to short–range temporal dependence.

This means that events X_i and X_j, both of which are extreme, are independent provided time points i and j are far enough apart.

2.1 Extremes of dependent sequences

You might see why this property is often satisfactory. For example:

- Suppose we know that it rained heavily today
- That might influence the probability of extreme rainfall tomorrow...
- ... or even the Tuesday...
- ... or maybe even Wednesday!
- But maybe not for a specified day in two months' time!

2.1 Extremes of dependent sequences

Indeed, many real–life sequences satisfy this property.

By excluding the possibility of long–range dependence in this way, we focus our attention on dependence at a much shorter range.

Effects of such short–range dependence, it turns out, can be quantified within the standard extreme value limits discussed in Part 1.

2.1.1 Maxima of stationary sequences

Leadbetter *et al.* (1983) consider, 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.

2.1.1 Maxima of stationary sequences

Theorem

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

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

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

2.1.1 Maxima of stationary sequences

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

- $\theta = 1$ corresponds to a completely independent process
- $\theta \to 0$ with 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 $G^{\theta}(x)$ is also a GEV distribution.

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

2.1.1 Maxima of stationary sequences

So what does this mean in practice?

- If maxima of a stationary sequence converge – and we know they do...
- ... then, provided the $D(u_n)$ condition holds, the limit distribution is related to that of an independent series.
- The effect of dependence is just a replacement of G with G^{θ}.

2.1.1 Maxima of stationary sequences

In fact, if G corresponds to the GEV distribution with parameters (μ, σ, ξ), then

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

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

Thus, if the (approximate) distribution of M_n is GEV with parameters (μ, σ, ξ), then the (approximate) distribution of $\tilde{M}_n$ is GEV with parameters (μ^*, σ^*, ξ)!

2.1.2 Modelling block maxima

Good news!

- Provided long–range dependence is weak, we can model block maxima in the usual way!
- The distribution of block maxima falls within the same family of distributions as would be appropriate if the series were truly independent!
- The only difference is a change in location and scale parameters

However...

- Our implied n (the number we are taking the maxima over) is now effectively reduced due to the dependence
- Thus convergence to the limit distribution will be slower

2.1.3 Modelling threshold exceedances

We have seen that things remain largely unchanged for fitting to a set of block maxima in the presence of short–term temporal dependence

However, 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.

2.1.3 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 post–analysis
3. Explicitly modelling the temporal dependence in the process

Approach 1 is the most widely–used. However, we have focussed on the relative merits of the other two, and have found some surprising results!

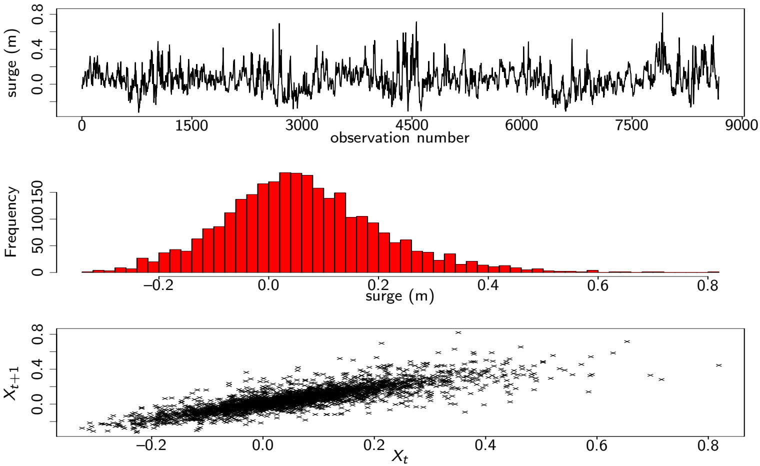
2.1.4 Example: Cluster peaks or all excesses?

Figure 9 shows a series of 3–hourly measurements of sea–surge heights at Newlyn, a coastal town in the southwest of England, collected over a three year period.

The sea–surge is the meteorologically induced non–tidal component of the still–water level of the sea.

The practical motivation for the study of such data is that structural failure — probably a sea–wall in this case — is likely under the condition of extreme surges.

2.1.4 Example: Cluster peaks or all excesses?



2.1.4 Example: Cluster peaks or all excesses?

A natural way of modelling extremes such time series is to use the Generalised Pareto Distribution (GPD) as a model for excesses over a high threshold.

As already discussed in Part 1, this approach might be preferable to the block maxima approach which is highly wasteful of data (and precious extremes!).

Figure 9 also shows the presence of substantial temporal dependence in the sequence of three–hourly surges.

We will now consider approaches **1** and **2**, outlined above, to circumvent this problem.

2.1.4 Example: 'Removing' dependence

The most commonly adopted approach to circumvent the problems caused by temporal dependence is to use 'runs declustering' to filter out a set of independent extremes

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.

2.1.4 Example: 'Removing' dependence

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.

Although this approach is quite easy to implement, there are issues surrounding the choice of κ; 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

2.1.4 Example: 'Removing' dependence

It has also been shown that parameter estimates can be sensitive to the choice of κ.

In this example, we use a separation interval of 60 hours (and so $\kappa = 20$) because

- that's what Coles and Tawn (1991) do
- they suggest that this is large enough to safely assume independence between identified clusters
- this also allows for 'wave propagation time'

We also use a mean residual life plot to identify a suitably high threshold (0.3m).

2.1.4 Example: Cluster peaks or all excesses?

The table below shows maximum likelihood estimates of the GPD scale and shape parameters σ and ξ, along with the associated 95% confidence intervals, fitted to the set of cluster peak excesses using $\kappa = 20$.

Shown for comparison are the corresponding estimates using *all* threshold exceedances, ignoring temporal dependence.

	$\hat{\sigma}$	$\hat{\xi}$
Cluster peaks	0.187	–0.259
95% confidence interval	(0.109, 0.265)	(–0.545, 0.027)
All excesses	0.104	–0.090
95% confidence interval	(0.084, 0.125)	(–0.215, 0.035)

2.1.4 Example: 'Ignoring' dependence

Table 1 shows that, although there is a slight discrepancy in parameter estimation when using (i) cluster peak exceedances and (ii) *all* exceedances, these discrepancies are non–significant.

Therefore, why bother declustering? Surely we're better off using *all* excesses?

- Confidence intervals too narrow – fitting to all exceedances when there is clearly evidence of short–term temporal dependence will result in underestimated standard errors
- But! Smith (1991) suggests a procedure in which we can inflate the standard errors post–analysis to take into account the temporal dependence (sometimes known as the variance "sandwich estimator")
- We haven't got time to go into this in detail today, but a full run–down of this procedure is given in the notes!

2.1.4 Example: 'Ignoring' dependence

Smith's adjustment: technique:

- Replace the critical value used to carry out a likelihood ratio test by a larger value
- The significance levels are then adjusted to take the dependence into account

Overall effect:

- Inflate the standard errors associated with MLEs
- Increase the width of confidence intervals (obtained directly or via profile likelihood)

2.1.4 Example: Cluster peaks or all excesses?

The first table on the next slide is a repeat of the last table, but now the standard errors for the GPD parameters have been inflated according to Smith's variance sandwich estimator procedure.

The second table shows maximum likelihood estimates for return levels for four return periods — $s = 10$, 50, 200 and 1000 years.

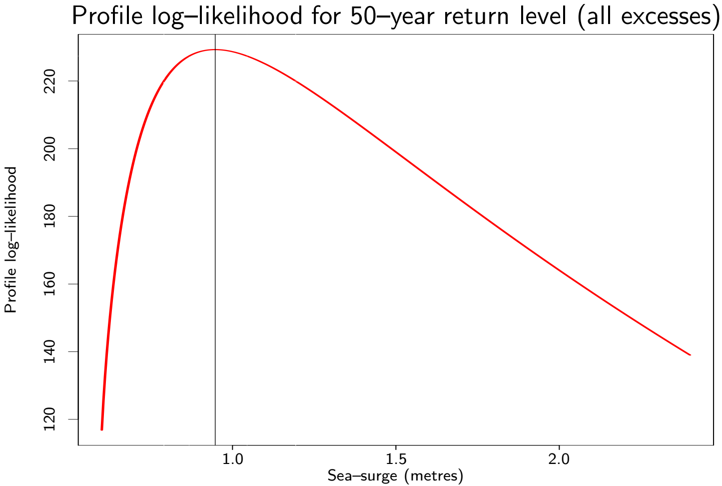
For the return levels, the corresponding 95% confidence intervals have been obtained using the method of profile likelihood, inflated appropriately according to Smith's adjustment

2.1.4 Example: Cluster peaks or all excesses?

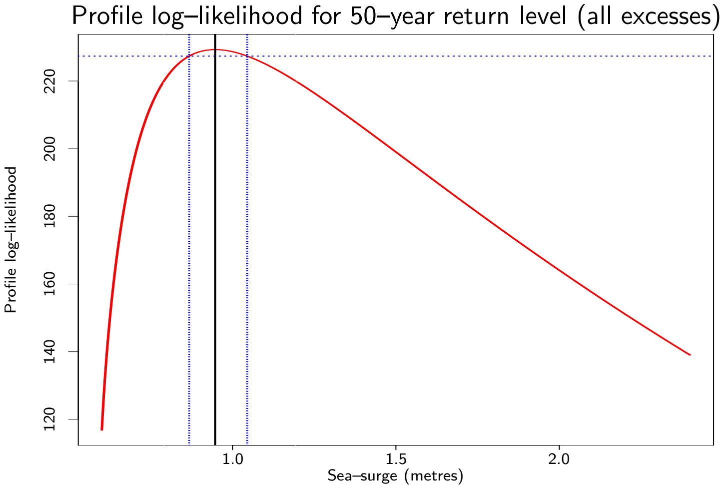
	$\hat{\sigma}$	$\hat{\xi}$
Cluster peaks	0.187	–0.259
95% Confidence Interval	(0.109, 0.265)	(–0.545, 0.027)
All excesses	0.104	–0.090
95% Confidence Interval	(0.082, 0.126)	(–0.217, 0.037)

	$\hat{z}_{10}$	$\hat{z}_{50}$	$\hat{z}_{1000}$
Cluster peaks	0.868	0.920	0.975
95% Confidence Interval	(0.770, 1.031)	(0.813, 1.099)	(0.858, 1.063)
All excesses	0.867	0.947	1.068
95% Confidence Interval	(0.736, 1.067)	(0.790, 1.193)	(0.891, 1.335)

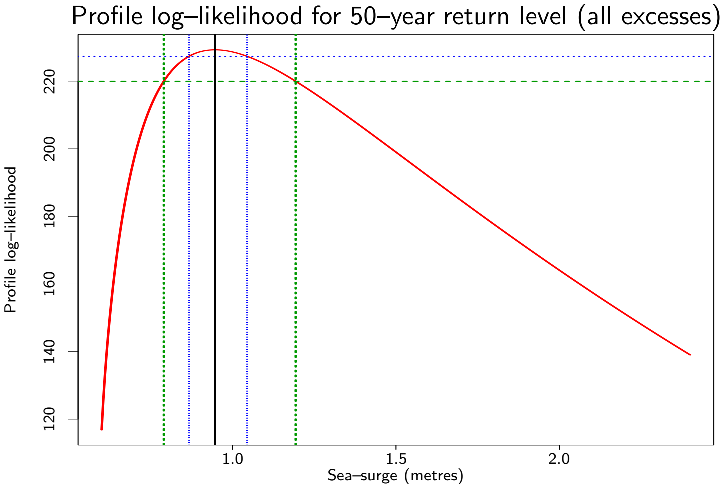
2.1.4 Example: Cluster peaks or all excesses?



2.1.4 Example: Cluster peaks or all excesses?



2.1.4 Example: Cluster peaks or all excesses?

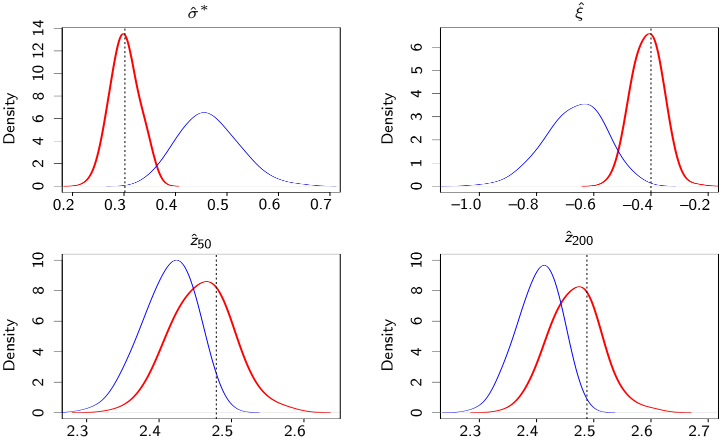


2.1.4 Example: Cluster peaks or all excesses?

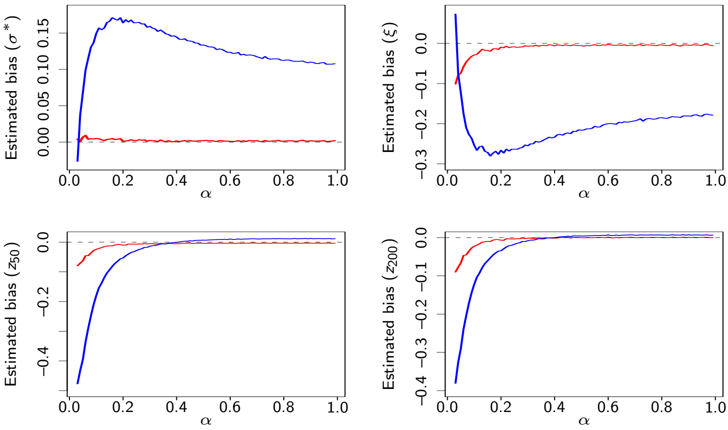
So we know there are differences – some significant – in return level estimation when we use (i) cluster peak excesses and (ii) all threshold excesses. Which approach are we to trust?

- The usual approach is to use cluster peaks, then we have effectively removed temporal dependence
- However, return levels using this approach are underestimated relative to the procedure which uses all threshold excesses
- Using cluster peak excesses could result in substantial under–protection!

2.1.4 Example: Simulation study



2.1.4 Example: Simulation study



2.2 Non–stationarity: trend

In Section 2.1 we demonstrated that the usual extreme value limit models are still applicable in the presence of short–term temporal dependence.

- We can use the results for block maxima directly as they stand
- Some thought is required for threshold modelling

The general theory can not be extended for non–stationary series.

Instead, it is usual to adopt a pragmatic approach of using the standard extreme value models as basic templates that can be augmented by statistical modelling.

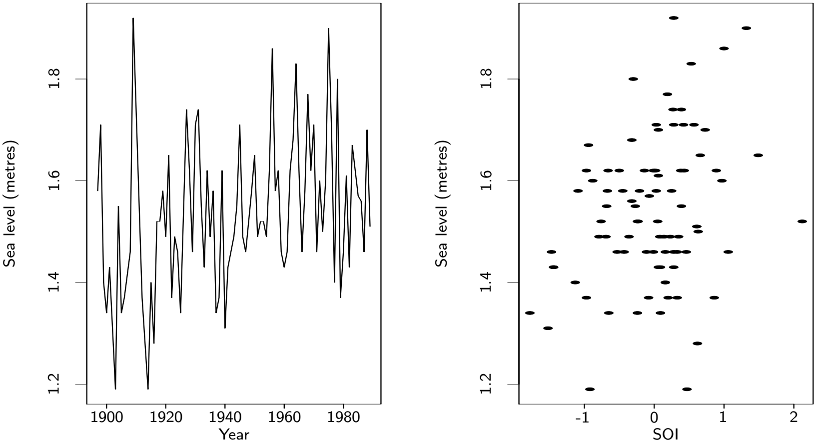
2.2 Non–stationarity: trend

Figure 12 shows a time series plot of annual maximum sea levels observed at Fremantle, Western Australia, between 1900 and 1986.

Also shown are these sea–levels plotted against the annual mean value of the *Southern Oscillation Index* (SOI).

There appears to be an increase in annual maximum sea levels through time, as well as an association between annual maximum sea levels and the mean SOI.

2.2 Non–stationarity: trend



2.2 Non–stationarity: trend

We can accommodate the time–trend shown in the plot on the left–hand–side of Figure 12 by fitting the GEV distribution, but allowing for a linear trend in the underlying level of extreme behaviour.

For example, if Z_t is the annual maximum sea level at Fremantle in year t, then we might use

$$Z_t \sim \text{GEV}\left(\mu(t), \sigma, \xi\right)$$

where

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

2.2 Non–stationarity: trend

In this way, variations through time in the observed process are modelled as a linear trend in the location parameter of the appropriate extreme value model (the GEV in this case).

We might choose to adopt the following model for $\mu(t)$:

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

to allow for a linear association between the maximum sea level in year t and the SOI in year t.

Or perhaps a multiple linear regression model for $\mu(t)$, whereby

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

2.2 Non–stationarity: trend

We can then assess our preferences between:

1. the stationary model: $\mu(t) = \beta_0$
2. the models which allow for a trend in time: $\mu(t) = \beta_0 + \beta_1 t$
3. the model which allows for a dependence on SOI through time: $\mu(t) = \beta_0 + \beta_1 \text{SOI}(t)$
4. the model which allows for a dependence on both time a*and* SOI: $\mu(t) = \beta_0 + \beta_1 t + \beta_2 \text{SOI}(t)$

2.2 Non–stationarity: trend

For example, fitting a stationary GEV distribution to these data, we get:

$$\hat{\mu} = 1.482(0.017) \qquad \hat{\sigma} = 0.141(0.011) \qquad \hat{\xi} = -0.217(0.064),$$

with a maximised log–likelihood of **43.6**.

Fitting the model which allows for a trend in time, we get:

$$\hat{\beta}_0 = 1.387(0.027) \qquad \hat{\beta}_1 = 0.002(0.0005)$$

$$\hat{\sigma} = 0.124(0.010) \qquad \hat{\xi} = -0.128(0.068)$$

with a maximised log–likelihood of **49.79**.

2.2 Non–stationarity: trend

Referring

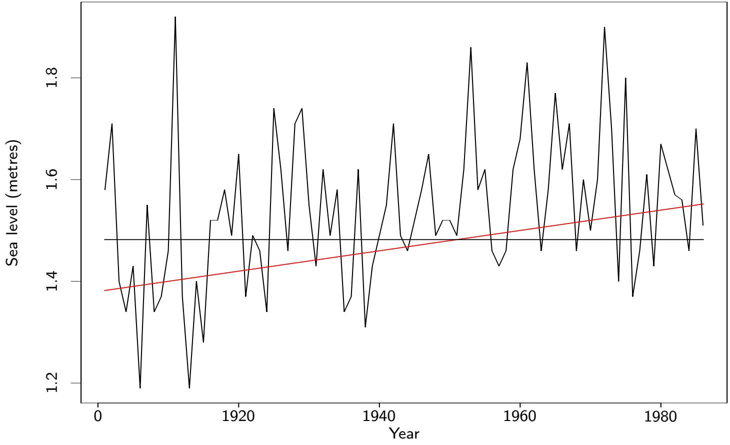
$$
\begin{aligned}
D &= 2\{49.79 - 43.6\} \\
  &= 12.38
\end{aligned}
$$

to χ_1^2 tables, we have a significant result, suggesting that the model which includes a linear trend in time for μ explains substantially more of the variation in the data than the stationary model.

Figure 13 shows the time series plot of the Fremantle sea level data with fitted estimates for μ superimposed.

Also shown, for comparison, is the fitted estimate for μ under the stationary model.

2.2 Non–stationarity: trend



2.2 Non–stationarity: trend

We find that the model which allows for a trend in μ depending on both time and SOI is the best one for our data. In fact, we get:

$$\hat{\beta}_0 = 1.389(0.027) \qquad \hat{\beta}_1 = 0.002(0.0005) \qquad \hat{\beta}_2 = 0.055(0.020)$$

$$\hat{\sigma} = 0.121(0.010) \qquad \hat{\xi} = -0.154(0.064)$$

giving

$$\hat{\mu} = 1.389 + 0.002t + 0.055\mathbf{SOI}(t).$$

Of course, more exotic model structures can be incorporated into this framework, including quadratic models, higher–order polynomial models, and models which allow for non–normal error structures.

Trend can also be incorporated into the other GEV/GPD model parameters.

2.3 Non–stationarity: seasonality

The most widely adopted technique is to partition the data into “seasons”:

- Fit a separate GEV/GPD to data within each season (wherein the data are approximately stationary)
- These seasons might be ‘winter’ and summer’, ’dry’ and ’wet’
- Where seasons are not as clearly defined, we could fit to separate months or years (see, for example, Part 5 this afternoon)

But then how do we ’recombine’ seasonally varying parameter estimates to obtain overall return levels?

2.3 Non–stationarity: seasonality

We could also allow the extremal parameters to vary continuously throughout the period of seasonality.

Fourier forms can be fitted to the parameters, and a model selected based on likelihood ratio tests.

However, Walshaw (1991) suggests that inferences are barely altered in relation to a piecewise seasonality approach (for extreme wind gusts, anyway).

2.3 Non–stationarity: seasonality

Another pragmatic approach is to only consider the season in which the 'most extreme' extremes occur.

Relative to this season, extremes in other seasons wont actually be extreme anyway!