Lund Institute of Technology
Mathematical Statistics

Centre for Mathematical Sciences

Statistical Modeling of Extreme Values, FMSN55/MASM15

Computer Assignment 3

This assignment is a compulsory part of the course. At the end of the session each group's results will be reviewed and graded as *pass* or *fail*. Note that in order to be able to finish the assignment in the specified time above, you have to read through the whole assignment in addition to the necessary parts in the book and manuals described below **before** you attend the computer session.

Modeling extremes of non-stationary sequences

You should read Chapters 6 and 7 in both `Coles` book before you start this computer session.

1 Getting started

Log in at one of the PCs in the computer room MH:230 or MH:231 using your *STIL*-account. Click on the icon "MClogin" on the desktop and login again with the same user name and password. This will attach the hard drive "`L:`" where your working directory will be saved. Note that you need to do this **before** you start up the software package `R`. Choose the latest version of `R` from the `Start` menu. If you have problems either logging in or starting R, ask for help.

Some tips and hints when typing in `R` code:

- `R` is case sensitive (so `LM()` and `lm()` are not equivalent).
- `R` is tolerant to the use of spaces, so `x <- 1` and `x<-1` are equivalent; though, the former being considered to be more readable.
- You can use the arrow keys to speed things up. The 'up' arrow gives you the previous command that you typed.

- The usual prompt sign for `R` is `>`. If you get a `+` prompt sign instead, it means that `R` is awaiting the completion of the previous command that you typed in. This can happen because you have forgotten to close parentheses, for instance. Just type in the remainder of the command.

2 Maximum sea-level Data

In this assignment we will analyze annual maximum sea-levels at Fremantle, near Perth, Western Australia. The dataset was introduced first in Example 1.3 page 5 of the book and the extreme value analysis of data has been discussed in Section 6.3.1, pages 111-114. To repeat this analysis, start `R` and write `library(in2extRemes)` to attach the related libraries and get access to the dataset and `R` functions. To start the graphical user interface run the command `in2extRemes()` in `R` window.
The dataset on maximum sea levels is stored in `R` dataset `fremantle.R` and can be accessed in the following locations

- `/usr/common/extremvarde/R/datasets/` (from Linux computers in the labs)
- `P:\Rdata` (from Windows computers in the labs)
- `http://www.maths.lth.se/matstat/kurser/fms155mas231/datasetsR.html`.

To read the dataset into `R` choose `File`→`Read Data` in `in2extRemes` window. Note that the dataset has been saved as `R` source so you should choose `R source` under `File Type` in the window which pops up. You do not need to change any other options but do not forget to assign the name `fremantle` to the dataset under `Save As(in R)`.

2.1 Statistical analysis

We will analyze the data by incorporating different trend models in the parameters of GEV according to the six modelsdescribed in Table 1 below ; see also the discussion in Section 6.3.1 of the Coles book. In all models the shape parameter, γ, is a constant. Table 1 shows how the location and scale parameters, μ and σ are modeled.
As part of the analysis we need to model the parameters in GEV as a function of `Year`. For numerical stability reasons, we will transform the `Year` to the interval $[0, 1]$. We will call this variable `time` in the analysis below. Note that this variable is denoted by t in Table 1. The square of `time` will be

Model	location and scale parameters in GEV
M_1	μ and σ are constants.
M_2	$\mu(t) = \mu_0 + \mu_1 t$ and $\sigma(t) = \exp(\phi_0 + \phi_1 t)$.
M_3	$\mu(t) = \mu_0 + \mu_1 t + \mu_2 t^2$ and $\sigma(t) = \exp(\phi_0 + \phi_1 t)$.
M_4	$\mu(t) = \mu_0 + \mu_1 t + \mu_2 SOI(t)$ and $\sigma(t) = \exp(\phi_0 + \phi_1 t + \phi_2 SOI(t))$.
M_5	$\mu(t) = \mu_0 + \mu_1 t + \mu_2 SOI(t)$ and $\sigma(t) = \exp(\phi_0 + \phi_1 SOI(t))$.
M_6	$\mu(t) = \mu_0 + \mu_1 t + \mu_2 SOI(t)$ and σ is constant.

Table 1: Different models which will be fitted to the dataset.

denoted by `timeSq`. The following commands will create these variables and will attach as separate columns to the fremantle data:

```
time=(fremantle$data$Year-min(fremantle$data$Year))/
(max(fremantle$data$Year)-min(fremantle$data$Year))
timeSq=time^2
fremantle$data <- cbind(fremantle$data,time=time,timeSq=time^2)
names(fremantle$data)
```

The last command is just to ensure that the necessary variable have been added to the dataset.

1. Plot the sea level data against year (`Plot`→`Scatter Plot`).

2. Find maximum likelihood estimates of the parameters for generalized extreme value distribution (`Analyze`→`Extreme Value Distributions`) for each of the six models described in Table 1. Note that you can add covariates to the models by simply including the names of the variables in the corresponding fields of the parameters in `in2extRemes`. For instance Figure 1 shows how you can fit model 2 in `extRemes`.

   Recall that a new component with the name `models` will be added to the data object. For instance, writing `names(fremantle$models)` will show you the models which you have fitted at each stage. After first analysis you will see only `fit1`. Write `names(fremantle$models$fit1)` to see which parts are included as the result of the fit. Note that `fremantle$models$fit1$results` will print the parameter estimates, negative log likelihood, Hessian matrix and so on.
   Answer:

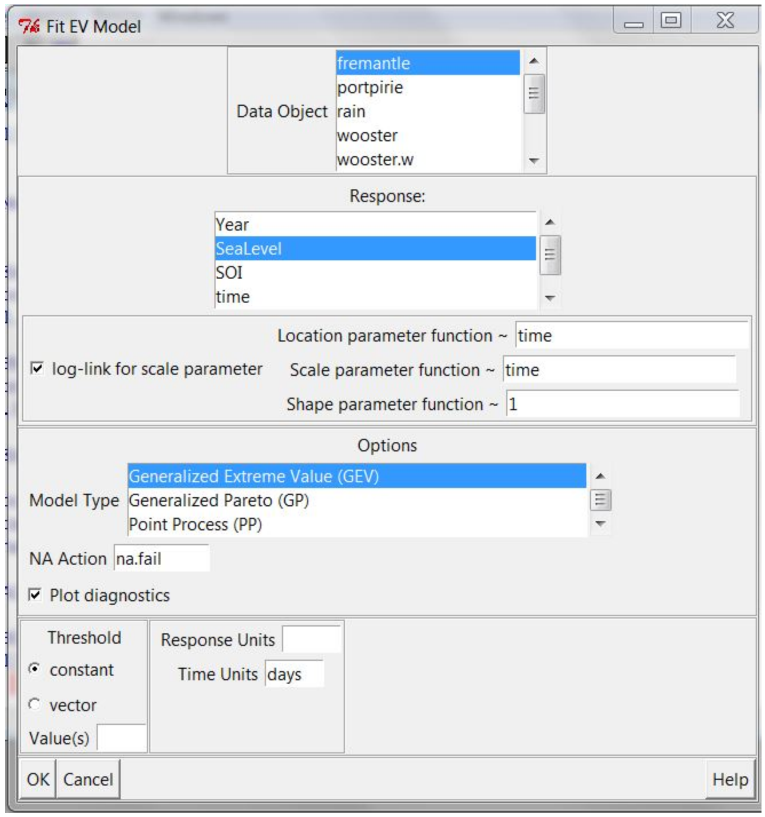

Figure 1: Fitting model M_2 in `extRemes`.

3. Use likelihood ratio test to compare different models and choose a model which seems to fit to the data best (`Analyze`→`Likelihood-ratio test`).
   Answer:

3 Wooster Temperature Series

Daily minimum winter temperature (degrees below 0 F.) are given in the R source file `wooster.R`. If you saved your workspace after the computer assignment 2, the `wooster` dataset will be available in your working directory otherwise you have to read it into `in2extRemes` as usual.
In Computer Assignment 2 in order to obtain an approximately stationary series we considered only the temperatures in winter. As discussed in Chapter 7, it is possible to analyze the whole dataset by modeling exceedances over a time-varying threshold. In Section 7.7, several models have been briefly discussed.
As a first step we need to create the covariates which will be used in the analysis. The following commands shows how this can be done in R. For convenience, these commands are also stored in the file `assignment3.txt` in the following locations:

- `/usr/common/extremvarde/R/datasets/` (from Linux computers in the labs)
- `P:\Rdata` (from Windows computers in the labs)
- `http://www.maths.lth.se/matstat/kurser/fms155mas231/datasetsR.html`.

You can copy and paste from this file to R directly. **Please note the following:** before submitting the following commands you need to read the `wooster.R` dataset to `in2extRemes` first. You need also to save this dataset as `wooster` otherwise the following commands will result in an error.

```
x = 1:length(wooster$data[,2])
usin = function(x, a, b, d)
{
        a + b * sin(((x - d) * 2 * pi)/365.25)
}
wu = usin(x, -30, 25, -75)
winter = c(rep(c(rep(1, 61), rep(0, 273), rep(1, 31)), 5),
        1)
spring = c(rep(c(rep(0, 61), rep(1, 91), rep(0, 365 - 91 -
        61)), 5), 0)
summer = c(rep(c(rep(0, 61 + 91), rep(1, 91), rep(0, 365 -
        91 - 61 - 91)), 5), 0)
fall = c(rep(c(rep(0, 61 + 91 + 91), rep(1, 91), rep(0, 365 -
        91 - 61 - 91 - 91)), 5), 0)
```

```
rescale.covariate  = function(x)
{
r.x = range(x)
x.01 = (x-r.x[1])/diff(r.x)
2*x.01-1
}
ydat = cbind(wu, sin((x * 2 * pi)/365.25), cos((x * 2 * pi)/365.25
        ), rescale.covariate(x), winter, spring, summer, fall)

wooster$data[,"value"] <- - wooster$data[,"value"]
wooster$data <- cbind(wooster$data,ydat)
colnames(wooster$data)[3:6] <- c("wu", "sin","cos","time")
wooster$data <- as.data.frame(wooster$data)
```

3.1 Statistical analysis

1. Plot the temperature data against observation number (`Plot`→`Scatter Plot`).

2. Repeat the analysis in Section 7.7 of the book and compare them by using likelihood ratio test by creating a table similar to Table 7.1 in the Coles book.

   In order to model a parameter as a function of some covariates, you just need to write the name of the corresponding columns in the field for that parameter. For instance Figure 2 shows how you can fit model 6 in `extRemes`. Note that the value of log-likelihood you get in the current version of `R` will slightly differ from those reported in the book.
   Answer:

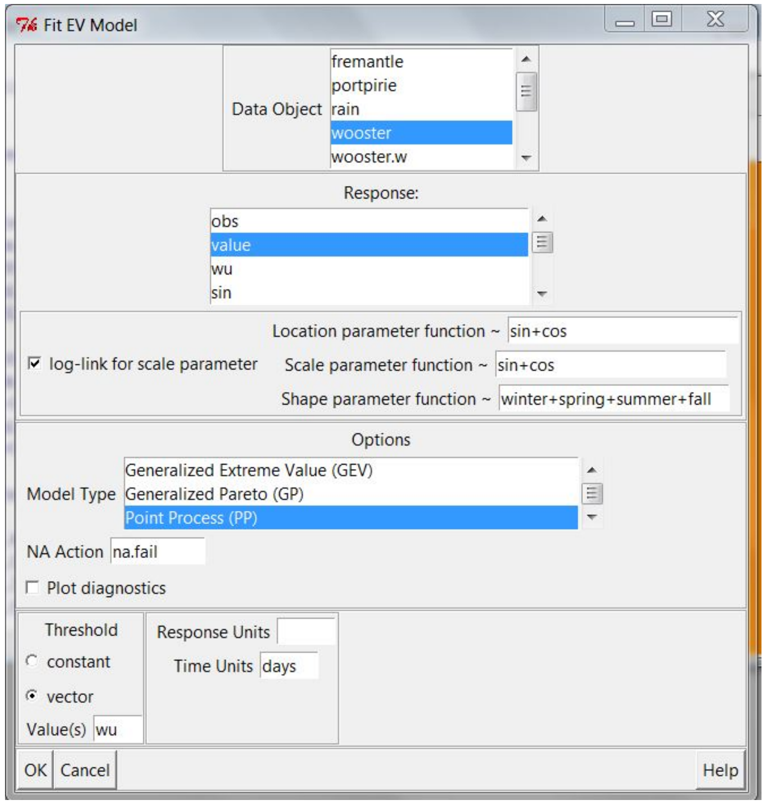

Figure 2: Fitting model 6 in `extRemes`.

3. Include the estimates of the parameters for each model in your report and enclose a number of plots which are useful to understand and compare the models.
   Answer:

4. You should note that by default a constant term is added to each combination of covariates which you specify in a model. This means that if you specify the covariates in model 6 as it is shown in Figure 2, you will get 5 estimates for γ although there are only four seasons in the model. A simple solution to this problem is to put only three of the four seasons as covariates for γ.

   Analyze your results. Which of the six models seems to be adequate for this dataset?

   Answer: