MAS8306: Problems sheet 3

Estimating Environmental Extremes

Students can choose to attempt Section A *or* Section B. Both Sections carry equal marks. Do not submit solutions to both parts!

Answers should be placed in the homework submission box in the foyer of the Maths & Stats General Office (Herschel Building 3rd floor) by no later than **4pm, Thursday 10th May**.

Section A

Each student has been assigned a unique dataset for this part of the assignment. This dataset will either be:

- daily snow depths;
- maximum daily temperatures;
- maximum daily wind gusts, or
- daily precipitation accumulations,

for a location in Europe. Go to the "Problems" section of the MAS8306 webpage, click on Personal datasets for Section A, and then click on your name. A file will open containing a description about your data – including the site at which the data were collected and some other important information. Make sure you read this material carefully! This file will also have a table containing – amongst other things – the data themselves.

Save this file as "`mas8306-data-master.txt`" on your `H:` drive, or somewhere else convenient. Now make a copy of this file for editing, and open this copy. In this copy of the master file you should make the data ready for reading into R. For example, all text at the top of the file should be removed; you might also want to remove all commas in the data table itself. The aim is to be able to read the data table into R using the command `read.table`.

Using methods from the course you think are most appropriate (Chapters 2–5 of the lecture notes), estimate your 100–year return level, and write a short report. Some things to consider:

- At the start of your report, briefly describe your data: include the type of data you have, the geographic location, and the practical significance of estimating the 100–year return level
- Include some simple exploratory analyses
- Look out for missing values!
- Think carefully about any trends or seasonal effects that could impact the validity of your analyses

Section B

Questions in this section are based on the paper by Fawcett & Walshaw (2012):

Estimating return levels from serially dependent extremes, *Environmetrics*, **23**, pp.272—283.

This article is available to download from the MAS8306 course webpage. You should carefully read through the paper, and then work through the following questions.

1. In your own words (and in plain English), discuss the main aims of the paper. *[Max 200 words]*

2. What do the plots in Figure 1 reveal? Outline the various approaches the authors mention in Section 2.3 to circumvent the problems illustrated in Figure 1. Again, make sure you use your own words (and explain things in plain English). *[Max 200 words]*

3. In Equation (5), the authors give a formula for the r–year return level z_r.

(a) *How*, and *why*, does this differ from Equation (3.9) in the lecture notes?

(b) Derive the result given in Equation (5).

4. The authors demonstrate the standard *peaks over threshold* approach in Section 2.3.4. What are their main concerns with this method? *[Max 200 words]*

5. In their simulation study in Section 3, the authors simulate data which are *first order Markovian*.

(a) Very briefly, explain what this means, and show how we get from the left-hand-side of Equation (6) to the right-hand-side.

(b) Does the assumption of first order Markovianity seem plausible for real–life extremes?

6. In the simulation study, the authors make use of three models for the distribution of consecutive pairs of extremes. Each of these models has at east one dependence parameter.

(a) Briefly explain the role of the dependence parameter(s) in each of the three models.

(b) Briefly explain the main similarities, and differences, between the three models.

7. In your own words,

(a) briefly explain what Figure 2 shows;

(b) briefly explain the purpose of Figure 3.

(c) Why are we so interested in the *extremal index* here?

8. Figure 4 summarises the main findings of the simulation study. *In your own words*, explain what the graphs in this figure show, with particular reference to the column of plots down the right-hand-side. *[Max 500 words]*

9. Tables 1 and 2 summarise results using *all* threshold excesses for sea–surge extremes at Newlyn and wind speed extremes at High Bradfield (respectively). Briefly explain how the *best* of these approaches compares to the standard approach using *cluster peak* excesses. Also explain how the 95% confidence intervals for the Newlyn sea–surge extremes, given in Table 3, compare to those discussed in Section 2.3.4. *[Max 200 words]*

10. Why are return levels for the Bradfield wind speed extremes more awkward to estimate than those for the Newlyn sea surge extremes?