Student projects of the course
"Statistical extreme-value analysis of climate data"

Lecturers: Thomas Opitz, Florian Lasgorceux (BioSP, INRAE Avignon)

General instructions:

- Projects must be conducted by groups of two students.
- A summary of important project results has to be presented orally and in written form:
  - A written report, for instance a well-structured `R Markdown` output document (pdf or html) containing well-documented code, R output and discussion of results. The name of the file is "Project[N].pdf" with "[N]" replaced by the number of your project. Please also indicate the names of your team members on the first page of the report. There should be a maximum of 4 pages of text (without counting figures, code and `R` output).
  - A 10-minute oral presentation (with slides to projected). Your talk will be followed by approximately 5 minutes of discussion with questions.
- The code file (`R` or `Rmd`) of your work must also be joined with the report.
- The grading of the project will be based on the above elements.
- **Deadline for the written report: March 16th, 2022, before noon (= 12h)**.
- Written reports, presentation slides and the oral examination can be in French or English language.

Project description and tasks

1 Data

The Rdata file `data.RData` contains the data for all projects in two data frames. The first one is called `data_df` and contains the data for Tasks 1 to 4. It has columns `date` (with the observation date) and `var1` to `var11` with daily observations of various meteorological variables. The second one is called `data2_df` and contains the data for Task 5. It has columns `date` (with the observation date) and `var1` to `var15` with daily observations of a meteorological variable at various locations. You can load the data into `R` with the following command: `load( "data.RData")`.

2 Observations to be used for the different projects

- For Tasks 1 to 5 and `data_df`, the project with number $N \in \{1, 2, \ldots, 11\}$ will work with the variable called `varN` (but NOT with any of the other variables).
- For Task 6 and `data2_df`, the attribution ”project number → variables to use” is as follows:
   1. → (1,2)
   2. → (3,4)
   3. → (5,6)
   4. → (7,8)
   5. → (9,10)
   6. → (11,12)
   7. → (13,14)
   8. → (15,1)
   9. → (2,4)
   10. → (6,8)
   11. → (10, 12)

3 Tasks

1. **General exploratory analysis**
   Perform standard visual and exploratory statistical analyses of the time series and discuss the results. Examples: histograms, time series plots, autocorrelation (function `acf` in `R`), number and occurrence structure of missing values, range of values, positive/negative/zero values, light/heavy tails and outliers, perceptible seasonal trends or long-term trends.

2. **Analysis using blockwise maxima**
   (a) Extract annual maxima. Remove years with more than 30 missing values.
   (b) Estimate a stationary GEV model for annual maxima.
   (c) Propose and estimate one or several time-nonstationary GEV model(s) for annual maxima (for instance, nonstationarity in the scale parameter, or in the position parameter).

(d) Compare stationary and nonstationary models. Could there be a climate change effect? Remark: you can compare AIC values of different models, and you can also check if (approximate) symmetric confidence intervals of covariate coefficients do not overlap 0.

(e) Calculate a 100-year return level using (i) the stationary model (ii) the nonstationary model for 2020. Compare.

3. **Analysis using Peaks-over-Threshold**

   (a) Based on threshold diagnostic tools (for instance, mean excess plots, parameter stability plots) and other considerations such as sample size, choose a sensible value for a stationary threshold and motivate your choice.

   (b) To study the temporal dependence of extremes, estimate the tail autocorrelation function and comment.

   (c) Decluster threshold exceedances using the runs method with at least $k = 1$ non-exceeding observations between consecutive clusters.

   (d) Estimate a stationary GPD for the threshold exceedances (that is, threshold excesses of cluster maxima). Compare estimated parameters with those from the analysis of maxima data.

4. $\star\star$**Bonus question**$\star\star$
   To which daily variable from the following list corresponds your observation series? And why do you think so?

   (a) Precipitation, Berlin, Germany
   (b) Average temperature, Riyad, Saudi Arabia
   (c) Maximum temperature, Madrid, Spain
   (d) Precipitation, Brest, France
   (e) Maximum temperature, Kathmandu, Nepal
   (f) Minimum temperature, Brest, France
   (g) Snow depth, Zugspitze, Germany
   (h) Precipitation, Madrid, Spain
   (i) Average temperature, Irkutsk, Russia
   (j) Precipitation, Edinburgh, UK
   (k) Maximum temperature, Berlin, Germany

5. **Analysis of bivariate dependence in extremes**

   (a) Compute and visualize the bivariate dependence measure $\chi(u)$ and the bivariate Pickands function, and comment the result.

   (b) Fit several bivariate max-stable models (for instance, logistic, Huesler–Reiss) to componentwise annual maxima, compare them and select one of them (for instance, using the AIC where "lower is better").

4 Reminders

The confidence interval with type-1 error of α and confidence level $1 - \alpha$ (for instance, $\alpha = 0.05$) is given as follows:

$$(\text{parameter estimate}) \pm (\text{standard error}) \times z_{1-\alpha/2}$$

where z_p is the p-quantile of the standard normal distribution. For $\alpha = 0.05$, it is $z_{0.975} = 1.96$.