Statistics for Environmental and Climate Analysis with R

To understand the water cycle and climate changes impacts in urban environments, whether from a hydrological, ecological or climatological point of view, it is essential to have statistical skills to deal with highly dynamic spatiotemporal variables. This course covers from statistical concepts to applied methods applied to environmental and climactic variables while helping the participants develop their skills in R programming and data processing. The R codes and key references to master the techniques and tools included in this course will be available for the participants through a GitHub repository. In the end, a workshop will allow participants to use their data and a method of their choice with the instructors' support.

Duration: 1.5 ECTS (15x45 min) Date: April 19-23
19/04 (Monday) 13:30 - 15:30: Introduction to R (non-compulsory)
20/04 13:30 - 17:00: Data science overview (30 min break)
21/04 13:30 - 17:00: Core concepts of statistics (30 min break)
22/04 13:30 - 17:00: Regression models (30 min break)
23/04 (Friday) 13:30 - 16:00: Modelling climate events and time-series data (30 min break)

Introduction to R (non-compulsory*)

- R studio and R base: *R environment and essential functions*
- R packages: *finding packages that suit your needs*
- R markdown: *working with markdown files throughout the course*

Data science overview

- Data structure: *determining statistics based on types of variables*
- Data visualization: *plotting cool graphs according to the aim and variable type*
- Exploratory Analysis: *finding relationships and identifying variable's behaviour*
- Modelling approaches: *explaining, predicting or forecasting?*

Core concepts of statistics

- Random variables and probability distribution: *beyond Normal*
- Confidence intervals and hypothesis testing: *samples, parameters and estimates*
- Statistical Inference: w*here all statistical conclusions rely on, classical or Bayesian*
- Data modelling: *general assumptions, model selection and accuracy*

Modelling climate events and time-series data

- Time-series data: working with dates and time in R
- Extreme Values Analysis: *changes and intensity of extreme (or rare) events*
- Time-series modelling: dealing with *seasonality and trends to forecast with ARIMA models*

Regression models

- Generalized Additive Model (GAM): *smoothing your issues, study case application*
- Machine Learning algorithms: *tuning parameters to control model complexity*
- Support Vector Machine (SVM): *tuning cost and epsilon, study case application*

Other statistical methods

- Multivariate Analysis: *Principal Components (PCA/PLSR), Cluster and Factorial Analysis*
- What is still missing: *an overview of uncovering subsects such as spatial Analysis and DOE*

Workshop

- Hands-on: *working on your data and interest*

*non-compulsory module with a duration of 2x45 minutes offers to help the participants who are not familiar with R and R studio.