

Getting started hints - for students

This section provides a few brief hints for the student in how to begin thinking about analysing the data.

Project 1 - Average daily temperatures in the US

Because of the big dimensionality of data, it is imperative to perform a thorough exploratory data analysis. Make sure you identify if there are any missing values: you then may have to input an appropriate value for such missing entries. To start understanding the data, first look at some individual representative cities, for which the value of the covariates are different. I would then recommend analysing a single year to start with and to build an appropriate statistical model for such a subset of the data. You may expect a simple linear model not to be appropriate and some more flexible model, for instance a GAM, may be a better option.

Project 2 - Average daily river flows in Scotland

Because of the big dimensionality of data, it is imperative to perform a thorough exploratory data analysis. To start understanding the data, first look at some individual representative rivers, for which the value of the covariates are different. I would then recommend analysing a single year to start with and to build an appropriate statistical model for such a subset of the data. You may expect a simple linear model not to be appropriate and some more flexible model, for instance a GAM, may be a better option.

Project 3 - Maxima daily temperatures in under canopy vs. open field stations in Switzerland

Because of the big dimensionality of data, it is imperative to perform a thorough exploratory data analysis. Make sure you identify if there are any missing values and to code them properly. You then may have to input an appropriate value for such missing entries. To start understanding the data, first look at some individual representative stations and check if there are differences between open field and under canopy temperatures. I would then recommend analysing a single year to start with and to build an appropriate statistical model for such a subset of the data. You may expect a simple linear model not to be appropriate and some more flexible model, for instance a GAM, may be a better option.