Statistical Analysis Plan

Average flows predicting based on generalized additive models and frequent model averaging

Mengran Li

Population

Daily river flows in Scotland.

Objectives

- Do average daily flows in Scotland vary over space?

- Do average daily flows in Scotland vary over time?

- Are the covariates effective to predict average flows?

Data Collection

The daily flow data and gauge station information are provided by the Scottish Environment Protection Agency (SEPA) and the National River Flow Archive(NRFA). The data contains observations of 64 gauge stations during 01/01/1989 - 31/12/2015.

Variables Under Consideration

- Flow: average daily flow (m^3/s); **main response variables**

- Year: temporal effect, 1989 - 2015; **covariate**

- Day: seasonal effect, days within a year, 1 - 365; **covariate**

- Easting and Northing: spatial effect, longitude and latitude; **covariates**

- Catchment.Area: catchment area of the measured river (km^2); **covariate**

- Max.Altitude: max altitude inside a catchment; **covariate**

Missing Data Procedures

The proportion of missing values will be checked and the missing values will be imputed using linear interpolation separately for months.

Summaries to be presented

- For each gauge station, the river main characteristics will be displayed: catchment area, max elevation, mean flow (se), min flow, max flow, mean flow/catchment area etc.

- A kernel smoothed method will be employed to detect the trend of mean flows and smoothed lines will present the data patterns.

- A map will show the positions of gauge stations and the colors represent the mean flows which indicate the spatio patterns.

Models to be fitted

- Candidate models

  - Linear models

  - GAMs

  - Gaussian Process

- Variable sets

  - Catchment.Area + Max.Altitude

  - Trend + Catchment.Area + Max.Altitude

  - Trend + Season + Catchment.Area + Max.Altitude

  - Trend + Season + Spatio + Catchment.Area + Max.Altitude

  - Trend + Season + Spatio + interaction + Catchment.Area + Max.Altitude

- Model Selection

  - AIC/BIC/SIC etc. to compare different nested models.

  - Likelihood ratio to test the significance of covariates under nested models.

  - A best model is selected in each type of models.

  - The estimated parameters and mean flows with 95% confidence intervals of best models of different types will be plotted and compared.

- Prediction

  The predictions of mean flows in 1, 10, 100 years are illustraed and compared.

  Root mean square prediction error: leave-one-out cross validation to choose the best model.

  The final full predictions of whole scottish mean flows in 1/10/100 years will be drawn in a map.