

Analysis of non-linear air pollution time series

Laboratory:

Swiss Data Science Center

Type:

Semester Project

Description:

Air pollution is one of the leading causes of mortality globally. A thorough understanding of time-varying statistics of pollutant concentration in the air, such as their entire probability density function, is critical to support decision-makers concerned with the definition of policies to reduce citizens' overall exposure to air pollution.

This project aims at investigating the statistical patterns of the concentration of different pollutants in the air. To this aim, a real-world dataset including time-series observations from hundreds of monitoring stations throughout all Europe and covering a period of five years will be mined using a combination of advanced analytical tools including complexity measures and time-series clustering algorithms.

Results are expected to confirm the relationships between anthropogenic activities e.g., land use, and pollutant concentration patterns.

Goals/benefits:

- Working with advanced time series analysis tools
- Hands-on experience with real world data and problems
- Possibility to publish a research paper

Prerequisites:

- Applied statistics (advanced or intermediate skills)
- Machine learning (intermediate skills)
- Python (advanced skills) ; basic knowledge of R is a plus
- Interested in interdisciplinary applications, with specific focus on environmental sciences

Deliverables:

- Well-documented code
- Written report and oral presentation

References:

[1] Donner, Reik V., and Susana M. Barbosa. "Nonlinear time series analysis in the geosciences." Lecture Notes in Earth Sciences 112 (2008): 37.

[2] He, Hankun, Benjamin Schäfer, and Christian Beck. "Spatial heterogeneity of air pollution statistics in Europe." *Scientific Reports* 12.1 (2022): 1-12.

[3] Wile, Christopher K., Andrew Zammit-Mangion, and Noel Cressie. *Spatio-temporal Statistics with R*. Chapman and Hall/CRC, 2019.

Contact:

federico.amato@epfl.ch