

THESIS PROJECTS

Time Series Analysis and Spatiotemporal Modeling in Air Quality: A Comprehensive Statistical Perspective

Innovative Approaches to Nonstationary Spatial Modeling and Cluster Estimation in Large Datasets

Description: Contemporary methods of generating digital data, like computer simulations and remote sensing, have tremendously expanded the scale and intricacy of information collected across spatial domains. While the analysis of these extensive spatial datasets typically employs Gaussian processes, the nonstationary nature and computational demands for such large datasets can make infeasible the implementation of Gaussian process models. To make inferences computationally viable for large spatial datasets, a novel approach involves dividing a spatial area into distinct sets using hierarchical clustering of observations and finite differences as a dissimilarity measure capable of identifying locally stationary regions. By segmenting the spatial domain, independent spatial models can be applied to each region, introducing non-stationary characteristics into the overall spatial process. The accurate estimation of local models depends on recognising the underlying "true" clusters. To support this, a case study using observed data and Monte Carlo simulations will be conducted to evaluate the prediction accuracy of the spatially clustered Gaussian process when the estimated clusters differ from the "true" clusters. In the end, the students will be able to handle the non-stationary spatial modelling by multivariate clustering procedure accounting for spatial correlation.

Keywords: Geostatistics and machine learning, non-stationary process, High-performance computing (HPC), Python

Harmonising Data: A Geostatistical Approach to Data Science

Description: The work will mainly focus on the implementation of cutting-edge strategies for geostatistic harmonisation purposes. The techniques used are transversal and helpful for every sector that deals with data characterised by different resolutions, both spatial and temporal. The harmonisation process, involving a change of support techniques, will be developed within a solid statistical framework. The dataset will be developed within the framework of the PNRR project GRINS, aiming to build the AMELIA platform. The dataset is the national version of the well-known "Agrimonia Dataset" available on Zenodo, with 2000+ downloads in just a year. The student will focus on environmental data provided by institutional organisations (mainly the European Union) about meteorology, air quality, emissions, etc. The student will acquire extensive knowledge about how to collect data (including API services), how to manage different formats of high dimensional data, how to

harmonise them from different scales, and how to publish open access datasets along with the code used and write scientifically. The students will meet, work together and be supported by the creator of the Agrimonia dataset. The student will be acknowledged for the publication of the new version of the data set and will be considered for other ongoing projects, especially where a change of support skills is required.

Keywords: Geostatistic harmonisation, Change of support, Python / R / MATLAB

Breath of the Valley: Navigating Air Quality Challenges

Description: The study of air quality, and more in particular, in the Po Valley, has gained extreme importance in the last years, as evidence supports a critical situation. The combination of mountain regions that block air circulation and the extremely high urbanisation and industrialisation of the area leads to the worst air quality among European countries. The non-linear behaviour of air particles arises from intricate chemical reactions, dynamic physical processes like coagulation and dispersion, and complex interactions with atmospheric conditions. These factors contribute to unpredictable and nonlinear responses in the concentration, distribution, and transformation of airborne particles in the atmosphere. Several statistical models, including machine learning, have been developed in air quality modelling in recent years. The stochastic approach to air quality prediction ensures relatively high performance for a much less computational cost concerning the chemical transport model making statistical models perfectly suitable for scenario analysis as several different simulations need to be run. The topic involves different master thesis projects: time series analysis and machine learning techniques to understand the behaviour of PM2.5 (partitioning), and to predict them over continuous maps. At the end of the project, the students will be able to manage advanced statistical models to increase their knowledge about this critical situation and to support local organisations within a data-driven decision framework.

Keywords: Air Quality, Spatio-temporal modelling, Machine learning, Applied statistics

Are you interested? Contact us!

alessandro.fasso@unibg.it

alessandro.fustamoro@unibg.it

jacopo.rodellini@unibg.it

For more information about the AgrImOnIA Project:

GitHub: <https://github.com/AgrImOnIA-project>

Website: <https://agrimonia.net/>

