Object Oriented Data Analysis for Non-Homogeneous Processes
PhD Candidate: Riccardo Scimone

Advisor: Prof. Piercesare Secchi
Co-Advisro: Prof. Giovanni Azzone

Thesi Abstract:
This doctoral dissertation is the result of a three-years-long research work. The focus of the entire research is the statistical analysis of complex data: such complexity may directly come from the irreducible complexity intrinsic to the data itself or from a complex correlation structure, which could also be exhibited by data as simple as real numbers. Special attention is given to developing novel statistical methodologies devoted to the detection of various kinds of anomalies in such datasets: these methods finds application in a wide variety of fields. In the first part of the work we give attention to data sets whose atoms are complex 3D geometries, proposing in particular a novel statistical framework for quality control and automatic anomaly detection of industrially manufactured complex shapes, for which consolidated monitoring techniques do not exist. The framework is, in particular, perfectly suited for Additive Manufacturing processes, which are becoming increasingly important in the Biomedical and Aerospace sectors. The second part focuses on complex data sets and processes exhibiting spatial non-homogeneity. Partially motivated by the huge impact of COVID-19 on the planet, we develop an original analysis pipeline for the spatio-temporal analysis of curve data, with application to the overall mortality curves in Italian administrative units. The framework, rigorously modeling the spatio-temporal correlation structure of the curves being analyzed, is able to capture spatial anomalies, and can be easily generalized to any family of geo-referenced temporal functions. In the third part of the thesis, the modeling of spatial non-homogeneities is further developed. We focus there on framing an estimation method for a family of statistical models which intrinsically encompass spatial anomalies, i.e., the family of Non Stationary Gaussian Process. While many different proposals have been advanced, they are all affected by limitations involving flexibility or computational limitations, and solid, general methods for parameter estimation for these models are still to be developed. We propose a novel, ensemble-based estimation method designed to overcome said limitations, exploring its performances both on simulated and real spatial data sets where other available methodologies struggle to find reliable solutions.