

STATISTICAL LEARNING FROM COMPLEX DATA FOR INDUSTRY AND SOCIETY

ALESSANDRA MENAFOGLIO^{*}, PIERCESARE SECCHI^{*}
AND SIMONE VANTINI^{*}

^{*} MOX – Department of Mathematics, Politecnico di Milano
Piazza Leonardo da Vinci 32, 20133, Milan, Italy
alessandra.menafoglio@polimi.it
piercesare.secchi@polimi.it
simone.vantini@polimi.it (Corresponding organizer)

ABSTRACT

This session will focus on one of the most impelling current research areas of investigation in Data Science, that is the development of algorithms and mathematical models for the statistical analysis of complex data. Consistently with the notion of complexity used in the Object-oriented data analysis literature [1], complexity of data is here intended as related to two possible aspects of the data analysis:

- **Complexity of the mathematical space where data points are embedded.** The session will indeed focus on the analysis of data in which sample units are not associated to standard Euclidean vectors but to more elaborated mathematical structures like: data sets made of mathematical functions which are used for modelling curves, surfaces, and fields (e.g., functional data); or data sets made of networks/graphs which are used for modelling pairwise relationships between anonymous entities (e.g., network/graph data).
- **Complexity of the stochastic spatial dependence between data points.** The session will focus also on the analysis of data for which a standard stationary dependence based on the Euclidean distance between the locations where the data have been measured is strongly not realistic and for which more sophisticated types of dependence are needed like: data sets made of data points collected on Riemannian manifolds with no stationary dependence; data sets whose dependence is possibly related to an underlying physical phenomenon for which some constitutional laws are partially known.

The talks will explore different perspectives on the topic: data exploration (which is used in real applications for current scenario assessment and inspection), model inference (which is used for supporting policy design and long-term strategy development), and spatial and temporal prediction (which is used for real-time monitoring and short-term strategy development). The session talks will range from being more methodological (focusing on mathematical and algorithmic challenges) to more applicative (focusing on the challenges of deploying the developed tools in real projects with public institutions and private companies).

REFERENCES (Not mandatory, maximum 2 references)

- [1] Marron, J.S. and Alonso, A.M. (2014), Overview of object oriented data analysis. *Biom. J.*, 56: 732-753 and the rejoinders therein.