

EGU2020-11461

<https://doi.org/10.5194/egusphere-egu2020-11461>

EGU General Assembly 2020

© Author(s) 2023. This work is distributed under the Creative Commons Attribution 4.0 License.



Advanced Exploratory Analysis of Air Pollution Multivariate Spatio-Temporal Data

Mikhail Kanevski¹, Federico Amato², and Fabian Guignard³

¹University of Lausanne, Institute of Earth Surface Dynamics, Lausanne, Switzerland (Mikhail.Kanevski@unil.ch)

²University of Lausanne, Institute of Earth Surface Dynamics, Lausanne, Switzerland (Federico.Amato@unil.ch)

³University of Lausanne, Institute of Earth Surface Dynamics, Lausanne, Switzerland (Fabian.Guignard@unil.ch)

The research deals with an application of advanced exploratory tools to study hourly spatio-temporal air pollution data collected by NABEL monitoring network in Switzerland. Data analyzed consist of several pollutants, mainly NO₂, O₃, PM_{2.5}, measured during last two years at 16 stations distributed over the country. The data are considered in two different ways: 1) as multivariate time series measured at the same station (different pollutants and environmental variables, like temperature), 2) as a spatially distributed time series of the same pollutant. In the first case, it is interesting to study both univariate and multivariate time series and their complexity. In the second case, similarity between time series distributed in space can signify the similar underlying phenomena and environmental conditions giving rise to the pollution. An important aspect of the data is that they are collected at the places of different land use classes – urban, suburban, rural etc., which helps in understanding and interpretation of the results.

Nowadays, unsupervised learning algorithms are widely applied in intelligent exploratory data analysis. Well known tasks of unsupervised learning include manifold learning, dimensionality reduction and clustering. In the present research, intrinsic and fractal dimensions, measures characterizing the similarity and redundancy in data and machine learning clustering algorithms were adapted and applied. The results obtained give a new and important information on the air pollution spatio-temporal patterns. The following results, between others, can be mentioned: 1) some measures of similarity (e.g., complexity-independent distance) are efficient in discriminating between time series; 2) intrinsic dimension, characterizing the ensemble of monitoring data, is pollutant dependent; 3) clustering of time series observed can be interpreted using the available information on land use.