



Jun 26th, 5:00 PM - 7:00 PM

Understanding the impact of climate and traffic in air quality: the impact of preprocessing in data science

Gibert Karina

Universitat Politecnica de Catalunya-BarcelonaTech, karina.gibert@upc.edu

Miquel Sànchez Marrè

Gerardo Ezequiel Martin

University Research institute for science and technology of sustainability

Nora Solé-Corcoll

University Research institute for science and technology of sustainability

Laura Vilardell-Magre

University Research institute for science and technology of sustainability

Follow this and additional works at: <https://scholarsarchive.byu.edu/iemssconference>

Karina, Gibert; Marrè, Miquel Sànchez; Martin, Gerardo Ezequiel; Solé-Corcoll, Nora; and Vilardell-Magre, Laura, "Understanding the impact of climate and traffic in air quality: the impact of preprocessing in data science" (2018). *International Congress on Environmental Modelling and Software*. 2.

<https://scholarsarchive.byu.edu/iemssconference/2018/Posters/2>

This Poster Presentation (in exhibition hall) is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Understanding the Impact of Climate and Traffic in Air Quality: the Impact of Preprocessing in Data Science

**Karina Gibert^{a,c,d}, Miquel Sànchez Marrè^{b,c,d},
Gerardo Ezequiel Martín Carreño^d, Nora Sole Corcoll^d, Laura Vilardell Magre^d**

^aDepartment of Statistics and Operations Research

^bDepartment of Computer Science

^cKnowledge Engineering & Machine Learning Group at Intelligent Data Science and Artificial
Intelligent Research Centre (KEMLG at IDEAI-UPC)

^dUniversity Research Institute for Science and Technology of Sustainability
Universitat Politècnica de Catalunya – BarcelonaTech, Barcelona, Catalonia
(karina.gibert@upc.edu, miquel@cs.upc.edu)

Abstract: In the context of smart cities, more and more real time information is available to a better management. The paper shows how to extract relevant decisional knowledge from the open data sources of the city of Barcelona, regarding different aspects of the city behaviour. The main purpose of this work is to get a better understanding of the relationships between climatic conditions, air quality and traffic in the city of Barcelona. A data science approach is followed to combine all these independent data sources into a single working data matrix to be properly exploited.

Detailed preprocessing has involved great efforts as the different opendata sets provide data at different granularities regarding both time (every 5 minutes, hourly, daily) and space (meteorological stations are not necessarily placed near the traffic monitoring points or the air quality stations). Significant efforts have been done to develop a correspondence model between stations of the three data sources and spatio-temporal interpolation has been required to build the reference dataset to be further analyzed through data mining methods. Georeference of the three networks of stations (climatic, air quality and traffic) have been used inside a GIS to establish the proper correspondence model. The work shows how the preprocessing of different kinds of data can be encompassed into a global Data Science process useful to support environmental plans in the city.

In a second step, a clustering model is performed to identify typical climatic-pollution-traffic scenarios in the city and profiling techniques have been used to identify the areas of the city following each scenario at certain time intervals.

Global association among all variables is analyzed by means of factorial analysis methods. Clusters are then projected on the factorial map to confirm profiling synthesis with global associations between variables.

The results show a synthetic view of the city behaviour and the relationships among climatic conditions, air quality and intensity of traffic at different periods of the week.

Keywords: Pre-processing, monitoring, spatio-temporal models, GIS, data science, air quality