Orietta Nicolis

orietta.nicolis@unab.cl

Spatial statistics and air quality modelling

Jules Kerckhoffs

j.kerckhoffs@uu.nl

statistical and machine learning methods for air quality mapping.

Andrew Larkin

Andrew.Larkin@oregonstate.edu

Big data and air pollution exposure assessment

Gavin Shaddick

G.Shaddick@exeter.ac.uk

Geostatistics, Bayesian modelling, air quality and environmental health.

ABSTRACT for computers environmental and urban systems

NO2 is a traffic-related air pollutant that is strongly associated with cardiovascular and respiratory diseases. Ground NO2 monitoring stations measure NO2 concentrations at certain locations and statistical predictive methods have been developed to predict NO2 as a continuous surface to inform decision-making. Among them, machine learning methods are the most powerful in capturing non-linear relationships between NO2 measurements and geospatial predictors, but it is unclear if the spatial structure of NO2 is also captured in the response-covariates relationships. In addition, most model comparison studies only compare accuracy in the prediction mean at ground stations, but do not consider prediction intervals and model interpretation and the effects of different model evaluation methods. In this study, we dive into the comparison between spatial and non-spatial data models accounting for the above-mentioned aspects.

Moreover, we implemented a spatial and a non-spatial methods that have not been applied to air pollution mapping before and evaluated stack learning methods with and without modelling the spatial process. We implemented our study using national ground station measurements of NO2 in Germany and Netherlands of the year 2017, predicting NO2 to 100 m resolution grid. Our results indicate the importance of modelling the spatial process especially in areas close to traffic.

The prediction intervals predicted with ensemble tree-based methods are satisfactory but too narrow with the geostatistical methods. Compared to ensemble tree-based methods, the geostatistical methods provide important spatial information for analysing emission sources and the spatial process of observations.

[**joaquin.cavieres@uv.cl**](mailto:joaquin.cavieres@uv.cl)

Universidad de Valparaíso

ABSTRACT for GA

NO2 is a traffic-related air pollutant. Ground NO2 monitoring stations measure NO2 concentrations at certain locations and statistical predictive methods have been developed to predict NO2 as a continuous surface. Among them, ensemble tree-based methods have shown to be powerful in capturing non-linear relationships between NO2 measurements and geospatial predictors but it is unclear if the spatial structure of NO2 is also captured in the response-covariates relationships. We dive into the comparison between spatial and non-spatial data models accounting for prediction accuracy, model interpretation and uncertainty quantification. Moreover, we implement two new spatial and a non-spatial methods that have not been applied to air pollution mapping. We implemented our study using national ground station measurements of NO2 in Germany and Netherlands of 2017.

Our results indicate heterogeneous levels of importance of modelling the spatial process in different areas. The prediction intervals predicted with ensemble tree-based methods are more satisfactory than the geostatistical methods. The two new methods implemented each obtained better prediction accuracy compared to the original ensemble tree-based and stacking methods. The probabilistic distribution of the spatial random field estimated by the geostatistical methods could provide useful information for analysing emission sources and the spatial process of observations.