

A land-use regression-based confidence predictor for modeling of Munich air pollution data

Olga Ivina^{1*}

1. Postdoctoral researcher, Institute of Epidemiology I, Helmholtz Zentrum München

*Contact author: lyolya@gmail.com

Keywords: Air pollution, Land-use regression, Conformal predictors, Spatial modeling

This work provides valid predictions with confidence estimation for air pollution levels in Munich region, Germany. Data from the ESCAPE study have been used, and they include annual mean concentrations of traffic-related air pollutants observed at the monitoring sites together with their geographical positioning, as well as land-use covariates obtained from the geographic information system (GIS).

This research takes as a basis a well-established statistical method for air pollution modeling, land-use regression (LUR), and transforms it into a machine learning method, creating a conformal predictor around it. This helps provide confidence to the prediction, guaranteed by the definition of a conformal predictor. Also, any conformal predictor always yields valid predictions, and very high levels of confidence can be guaranteed starting from very small-sized datasets. Different spatial covariance functions for the data can be considered in a LUR-based conformal predictor employing the “kernel trick”.

R programming language has been used to perform the task. LUR models have been fitted with the use of the **SpatioTemporal** package. R functions allowing to create conformal predictors that have LUR as the underlying method have been written, and they provide the possibility to implement several spatial covariance models. These functions use the framework set by the function `iidpred` from the **PredictiveRegression** package.

References

- [1] Kees de Hoogh et al. (2013). Development of Land Use Regression Models for Particle Composition in Twenty Study Areas in Europe. *Environmental Science and Technology* 47 (11), 5778–5786.
- [2] Bergen S., Lindström J (2013), SpatioTemporal: An R Package for Spatio-Temporal Modelling of Air-Pollution. Comprehensive Tutorial for the Spatio-Temporal R-package. <http://cran.r-project.org/web/packages/SpatioTemporal/index.html>
- [3] Vovk V., Nouretdinov I., Gammerman A. (2009), On-line predictive linear regression, *Annals of Statistics* 37 (3), 1566–1590.
- [4] Ivina, O., Nouretdinov, I., Gammerman, A. (2012) Valid predictions with confidence estimation in air pollution problem. *Progress in Artificial Intelligence* 1 (3), 235-243.