Machine learning 'De-Weathering' of urban NOx data to quantify meteorological impacts at two traffic sites in Germany

Dominik van Pinxteren, Sebastian Düsing, Alfred Wiedensohler, Hartmut Herrmann

Leibniz Institute for Tropospheric Research (TROPOS), Permoserstr. 15, 04318 Leipzig, Germany

Weather is known to influence observed concentrations of air pollutants in the urban atmosphere either directly, e.g. through reduced ventilation during inversion or indirectly, e.g. through increasing emissions during cold days. Studying the effects of meteorological variables individually is, however, often difficult due to their covariation with other variables.

Within the present study, we trained a machine learning model to learn the complex relationships between NOx concentrations, traffic patterns and meteorological variables at two kerbside sites in Dresden and Leipzig. Once successfully trained, the model predictions allow for a quantitative understanding of the importance of individual variables on the observed pollutant concentrations.

The statistical modelling was performed using the deweather R package1, which provides functions that conveniently wrap the gradient boosting machine (gbm), a tree-based machine learning algorithm. It also facilitates the calculation of partial dependence plots, describing the impact of one predictor on the outcome, while keeping all other predictors at constant levels. Models were trained with 4 years of hourly data at two traffic sites of the air quality monitoring network of the Saxonian Office for the Environment, Agriculture and Geology (LfULG). Predictor variables with a potential impact on NOx included local meteorological measurements, mixing layer height and trajectory length from HYSPLIT back-trajectory calculations, some time variables as proxies for temporal emissions patterns, and traffic densities counted close to the respective monitoring stations as proxies for local traffic emissions.

The results showed that next to traffic densities, wind speed and mixing layer height had the strongest impact on observed NO and NO2. When controlling for all other predictors by setting them to their mean values, NO2 at Dresden-Berstr. decreased nearly linearly from 65 to 35 µg/m3 between 0 and 1 m/s wind speed, and then by another 10 µg/m3 between 1 and 3 m/s. Increasing boundary layer height (from coarse reanalysis data) from 0 to 500 m decreased NO2 at the same site on average by 10 µg/m3. Consistent with the topography at Dresden-Bergstr., the model revealed that at this site only the outbound traffic on steeply rising lanes contributed to NOx pollution, while inbound, descending traffic hardly had an impact due to low engine loads.

Model predictions were also used to calculate 'deweathered' trends of NOx, which allowed for intervention analysis similar to previous suggestions2. The effect of increasing the sampling inlet height at the Dresden-Bergstr. monitoring station from 1.8 to 3.8 m was estimated to be 4-8% lower NOx concentrations. Further results will be shown in the presentation.

Overall, interesting insights into the processes controlling local pollution at urban hotspot sites were obtained, that encourage the further application of modern statistical methods in air quality studies.

References

Carslaw, D., 2018. deweather: Remove the influence of weather on air quality data. http://github.com/davidcarslaw/deweather

Grange, S.K.,Carslaw, D.C., 2019. Using meteorological normalisation to detect interventions in air quality time series. Science of The Total Environment 653, 578-588. doi: 10.1016/j.scitotenv.2018.10.344.