EGU2020-13063
https://doi.org/10.5194/egusphere-egu2020-13063
EGU General Assembly 2020

Supervised regression learning for predictions of aerosol particle size distributions from PM2.5, total particle number and meteorological parameters at Helsinki SMEAR3 station

Juha Kangasluoma[1], Yusheng Wu[1], Runlong Cai[1], Joel Kuula[2], Hilkka Timonen[2], Pasi Aalto[1], Markku Kulmala[1], and Tuukka Petäjä[1]
[1]Institute for Atmospheric and Earth System Research / Physics, Faculty of Science, University of Helsinki, Finland
[2]Finnish meteorological institute, Helsinki, Finland

Supervised regression learning for predictions of aerosol particle size distributions from PM2.5, total particle number and meteorological parameters at Helsinki SMEAR3 station

J. Kangasluoma[1], Y. Wu[1], R. Cai[1], J. Kuula[2], H. Timonen[2], P. P. Aalto[1], M. Kulmala[1], T. Petäjä[1]

[1] Institute for Atmospheric and Earth System Research / Physics, Faculty of Science, University of Helsinki, Finland

[2] Finnish Meteorological Institute, Erik Palménin aukio 1, 00560 Helsinki, Finland

Atmospheric particulate material is a significant pollutant and causes millions premature deaths yearly especially in urban city environments. To conduct epidemiological studies and quantify of the role of sub-micron particles, especially role of the ultrafine particles (<100 nm), in mortality caused by the particulate matter, long-term monitoring of the particle number, surface area, mass and chemical composition are needed. Such monitoring on large scale is currently done only for particulate mass, namely PM2.5 (mass of particulates smaller than 2.5 μm), while large body of evidence suggests that ultrafine particles, which dominate the number of the aerosol distribution, cause significant health effects that do not originate from particle mass.

The chicken-egg-problem here is that monitoring of particle number or surface area is not required from the authorities due to lack of epidemiological evidence showing the harm and suitable instrumentation (although car industry already voluntarily limits the ultrafine particle number emissions), while these epidemiological studies are lacking because of the suitable lack of

data. Here we present the first step in solving this “lack of data issue” by predicting aerosol particle size distributions based on PM2.5, particle total number and meteorological measurements, from which particle size distribution, and subsequently number, surface area and mass exposure can be calculated.

We use baggedtree supervised regression learning (from Matlab toolbox) to train an algorithm with one full year data from SMEAR3 station at 10 min time resolution in Helsinki during 2018. The response variable is the particle size distribution (each bin separately) and the training variables are PM2.5, particle number and meteorological parameters. The trained algorithm is then used with the same training variables data, but from 2019 to predict size distributions, which are directly compared to the measured size distributions by a differential mobility particle sizer.

To check the model performance, we divide the predicted distributions to three size bins, 3-25, 25-100 and 100-1000 nm, and calculate the coefficient of determination (r^2) between the measured and predicted number concentration at 10 min time resolution, which are 0.79, 0.60 and 0.50 respectively. We also calculate r^2 between the measured and predicted number, surface area and mass exposure, which are 0.87, 0.79 and 0.74, respectively. Uncertainties in the prediction are mostly random, thus the r^2 values will increase at longer averaging times.

Our results show that an algorithm that is trained with particle size distribution data, and particle number, PM2.5 and meteorological data can predict particle size distributions and number, surface area and mass exposures. In practice, these predictions can be realized e.g. in air pollution monitoring networks by implementing a condensation particle counter at each site, and circulating a differential mobility size spectrometer around the sites.