Geophysical Research Abstracts
Vol. 21, EGU2019-3294, 2019
EGU General Assembly 2019

Predicting in-vehicle concentrations of air pollutants using machine learning approaches – Implications for daily exposure

Vasileios Matthaios (1), Luke Knibbs (2), Louisa Kramer (1), Leigh Crilley (1), and William Bloss (1)

(1) University of Birmingham, Geography Earth and Environmental Science, United Kingdom (vxm668@bham.ac.uk), (2) School of Public Health, University of Queensland, Australia

Modelling concentrations of air pollutants within vehicles in an essential step to estimate our overall daily exposure to air pollution, and to provide practical advice for citizens to effectively reduce their exposure. This is a challenging issue however, since the processes that affect the concentration levels and hence exposure in-vehicles change not only with different on-road conditions but also with ventilation settings. The latter can alter processes such as air exchange, deposition, and inhalation that can happen in-vehicles. In this study, on-road mobile driving mode and mobile driving in-vehicle mode measurements of PM10, PM2.5, PM1, NO_2, NO_x, LSDA (lung surface deposited area) and UFP under different ventilation settings in four vehicles were used to build a model to estimate in-vehicle concentrations. The model was build by taking into account the basic atmospheric processes that occur inside vehicles (e.g. in-flows deposition, filtration etc). The dataset was divided into 80:20 of which 80% was used to develop the model and 20% to validate it. A machine learning (ML) algorithm (k-nearest neighbor, kNN) was applied to train the model (in the 80% of the data) and the new ML predictions were validated against unseen observations (20% of the data). The ML trained model, showed skill in predicting the random measurements with root mean squared error lower than the standard deviation. Good index of agreement (IOA > 0.69) and Pearson correlation coefficient (r > 0.80) for all the species and with good statistical measures for all the PM classes (PM10, PM2.5 and PM1) and NO_2 were found, however, the model under-predicted the LSDA, UFP and NO_x. The use of ambient monitoring data (recorded at a central site) instead of on-road data (measured on a mobile driving mode) in the ML algorithm showed that the ML trained model can predict the general pattern of the in-vehicle concentrations, thus giving promising applicability of the method based upon routine monitoring observations. The method presented here can be used in health assessment studies to estimate other in-vehicle air pollution species and hence provide a better understanding about passenger's exposure.