EGU22-7113, updated on 04 Dec 2022
https://doi.org/10.5194/egusphere-egu22-7113
EGU General Assembly 2022

Coupling regional air quality simulations of EURAD-IM with street canyon observations - a machine learning approach

Charlotte Neubacher[1], Philipp Franke[1], Alexander Heinlein[2], Axel Klawonn[3], Astrid Kiendler-Scharr[1], and Anne-Caroline Lange[1]

[1]Institute of Energy and Climate Research - Troposphere (IEK - 8), Forschungszentrum Jülich GmbH, Jülich, Germany
[2]Institute of Applied Mathematics, Delft Universtity of Technology, Delft, Netherlands
[3]Institute for Mathematics, Universität zu Köln, Köln, Germany

State of the art atmospheric chemistry transport models on regional scales as the EURAD-IM (EURopean Air pollution Dispersion-Inverse Model) simulate physical and chemical processes in the atmosphere to predict the dispersion of air pollutants. With EURAD-IM's 4D-var data assimilation application, detailed analyses of the air quality can be conducted. These analyses allow for improvements of atmospheric chemistry forecast as well as emission source strength assessments. Simulations of EURAD-IM can be nested to a spatial resolution of 1 km, which does not correspond to the urban scale. Thus, inner city street canyon observations cannot be exploited since here, anthropogenic pollution vary vastly over scales of 100 m or less.

We address this issue by implementing a machine learning (ML) module into EURAD-IM, forming a hybrid model that enable bridging the representativeness gap between model resolution and inner-city observations. Thus, the data assimilation of EURAD-IM is strengthened by additional observations in urban regions. Our approach of the ML module is based on a neural network (NN) with relevant environmental information of street architecture, traffic density, meteorology, and atmospheric pollutant concentrations from EURAD-IM as well as the street canyon observation of pollutants as input features. The NN then maps the observed concentration from street canyon scale to larger spatial scales.

We are currently working with a fully controllable test environment created from EURAD-IM forecasts of the years 2020 and 2021 at different spatial resolutions. Here, the ML model maps the high-resolution hourly NO_2 concentration to the concentration of the low resolution model grid. It turns out that it is very difficult for NNs to learn the hourly concentrations with equal accuracy using diurnal cycles of pollutant concentrations. Thus, we develop a model that uses an independent NN for each hour to support time-of-day learning. This allows to reduce the training error by a factor of 10^2. As a proof of concept, we trained the ML model in an overfitting regime where the mean squared training error reduce to 0.001% for each hour. Furthermore, by optimizing the hyperparameters and introducing regularization terms to reduce the overfitting, we achieved a validation error of 9−12% during night and 9−16% during day.