EGU21-12146
https://doi.org/10.5194/egusphere-egu21-12146
EGU General Assembly 2021


# Representing chemical history for ozone time-series predictions - a method development study for deep learning models

**Felix Kleinert**[1,2], Lukas H. Leufen[1,2], Aurelia Lupascu[3], Tim Butler[3], and Martin G. Schultz[1]
[1]Forschungszentrum Jülich GmbH, Jülich Supercomputing Centre, Jülich, Germany
[2]Rheinische Friedrich-Wilhelms-Universität Bonn, Institute of Geosciences, Bonn, Germany
[3]Institute for Advanced Sustainability Studies, Potsdam, Germany

Machine learning techniques like deep learning gained enormous momentum in recent years. This was mainly caused by the success story of the main drivers like image and speech recognition, video prediction and autonomous driving, to name just a few.
Air pollutant forecasting models are an example, where earth system scientists start picking up deep learning models to enhance the forecast quality of time series. Almost all previous air pollution forecasts with machine learning rely solely on analysing temporal features in the observed time series of the target compound(s) and additional variables describing precursor concentrations and meteorological conditions. These studies, therefore, neglect the "chemical history" of air masses, i.e. the fact that air pollutant concentrations at a given observation site are a result of emission and sink processes, mixing and chemical transformations along the transport pathways of air.
This study develops a concept of how such factors can be represented in the recently published deep learning model IntelliO3. The concept is demonstrated with numerical model data from the WRF-Chem model because the gridded model data provides an internally consistent dataset with complete spatial coverage and no missing values.
Furthermore, using model data allows for attributing changes of the forecasting performance to specific conceptual aspects. For example, we use the 8 wind sectors (N, NE, E, SE, etc.) and circles with predefined radii around our target locations to aggregate meteorological and chemical data from the intersections. Afterwards, we feed this aggregated data into a deep neural network while using the ozone concentration of the central point's next timesteps as targets. By analysing the change of forecast quality when moving from 4-dimensional (x, y, z, t) to 3-dimensional (x, y, t or r, φ, t) sectors and thinning out the underlying model data, we can deliver first estimates of expected performance gains or losses when applying our concept to station based surface observations in future studies.