EGU21-7596
https://doi.org/10.5194/egusphere-egu21-7596
EGU General Assembly 2021

Global fine resolution mapping of ozone metrics through explainable machine learning

Clara Betancourt[1], Scarlet Stadtler[1], Timo Stomberg[2], Ann-Kathrin Edrich[3], Ankit Patnala[1], Ribana Roscher[2], Julia Kowalski[3], and Martin G. Schultz[1]

[1]Jülich Supercomputing Centre, Jülich Research Centre, Jülich, Germany
[2]Institute of Geodesy and Geoinformation, Bonn University, Bonn, Germany
[3]Aachen Institute for Advanced Study in Computational Engineering Science, RWTH Aachen University, Aachen, Germany

Through the availability of multi-year ground based ozone observations on a global scale, substantial geospatial meta data, and high performance computing capacities, it is now possible to use machine learning for a global data-driven ozone assessment. In this presentation, we will show a novel, completely data-driven approach to map tropospheric ozone globally.

Our goal is to interpolate ozone metrics and aggregated statistics from the database of the Tropospheric Ozone Assessment Report (TOAR) onto a global 0.1° x 0.1° resolution grid. It is challenging to interpolate ozone, a toxic greenhouse gas because its formation depends on many interconnected environmental factors on small scales. We conduct the interpolation with various machine learning methods trained on aggregated hourly ozone data from five years at more than 5500 locations worldwide. We use several geospatial datasets as training inputs to provide proxy input for environmental factors controlling ozone formation, such as precursor emissions and climate. The resulting maps contain different ozone metrics, i.e. statistical aggregations which are widely used to assess air pollution impacts on health, vegetation, and climate.

The key aspects of this contribution are twofold: First, we apply explainable machine learning methods to the data-driven ozone assessment. Second, we discuss dominant uncertainties relevant to the ozone mapping and quantify their impact whenever possible. Our methods include a thorough a-priori uncertainty estimation of the various data and methods, assessment of scientific consistency, finding critical model parameters, using ensemble methods, and performing error modeling.

Our work aims to increase the reliability and integrity of the derived ozone maps through the provision of scientific robustness to a data-centric machine learning task. This study hence represents a blueprint for how to formulate an environmental machine learning task scientifically, gather the necessary data, and develop a data-driven workflow that focuses on optimizing transparency and applicability of its product to maximize its scientific knowledge return.