Geophysical Research Abstracts
Vol. 21, EGU2019-11503, 2019
EGU General Assembly 2019

Atmospheric chemistry modeling using machine learning

Christoph Keller (1,2) and Mat Evans (3,4)

(1) NASA Goddard Space Flight Center, Greenbelt, Maryland, USA, (2) Universities Space Research Association, Columbia, Maryland, USA, (3) Wolfson Atmospheric Chemistry Laboratories, Department of Chemistry, University of York, York, YO10 5DD, UK, (4) National Centre for Atmospheric Sciences, University of York, York, YO10 5DD, UK

Atmospheric chemistry models are a central tool to study the impact of chemical constituents on the environment, vegetation and human health. These models split the atmosphere in a large number of grid-boxes and consider the emission of compounds into these boxes and their subsequent transport, deposition, and chemical processing. The chemistry is represented through a series of simultaneous ordinary differential equations, one for each compound. Given the difference in life-times between the chemical compounds (milli-seconds for O1D to years for CH4) these equations are numerically stiff and solving them consists of a significant fraction of the computational burden of a chemistry model.

We have investigated a machine learning approach to emulate the chemistry instead of solving the differential equations numerically. From a one-month simulation of the GEOS-Chem model we have produced a training dataset consisting of the concentration of compounds before and after the differential equations are solved, together with some key physical parameters for every grid-box and time-step. From this dataset we have trained a machine learning algorithm (regression forest) to be able to predict the concentration of the compounds after the integration step based on the concentrations and physical state at the beginning of the time step. We have then included this algorithm back into the GEOS-Chem model, bypassing the need to integrate the chemistry.

This machine learning approach shows many of the characteristics of the full simulation and has the potential to be substantially faster (100x). There are a wide range of application for such an approach - generating boundary conditions, for use in air quality forecasts, chemical data assimilation systems, etc. We discuss speed and accuracy of our approach, and highlight some potential future directions for improving it.