Christoph A. Keller[1,2], Mat J. Evans[3,4]

[1] NASA GMAO, NASA GSFC, Greenbelt MD [2] Universities Space Research Association, Columbia MD [3] Dept. Chemistry, University of York, UK. [4] National Centre for Atmospheric Science, UK

GMAO
Global Modeling & Assimilation Office

Atmospheric Chemistry Modeling Using Machine Learning:
A Case Study Using GEOS-Chem

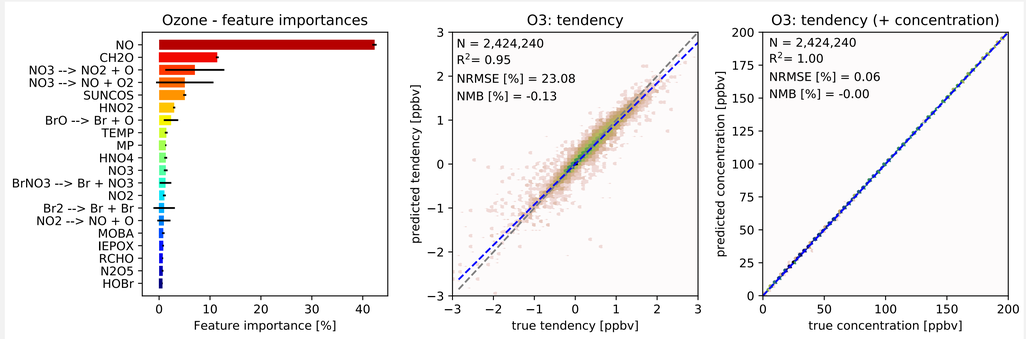
SUMMARY

We developed an alternative approach to simulate atmospheric chemistry based on machine learning. This method has the potential to be much faster than conventional atmospheric chemistry models.

Our method replaces the numerical gas-phase chemistry integrator with a machine learning emulator, trained from model data produced by the reference model. Our method was tested using training data consisting of one month of model output of chemical conditions along with the model physical state, produced from the GEOS-Chem model v10.

Several machine learning algorithms were tested to reproduce the concentration of each chemical species after the integrator, based on the physical and chemical conditions before integration. We find that tree-based methods (random forest, XGBoost) perform best and identify key drivers of chemical formation and destruction that are consistent with our scientific understanding of atmospheric chemistry *(Figure 1).*

Figure 1: Characteristics of ozone prediction model based on random forest regression



Left panel shows the relative importance of input features for the prediction of ozone. Middle panel shows random forest regression skill to predict chemical tendencies using a validation data set not used for training. Right panel shows prediction skill relative to absolute species concentration.

The random forest machine learning emulator was implemented back into GEOS-Chem. A one-month test simulation with this model shows that it compares well against the reference simulation *(Figure 2).* However, efficient implementation of the machine learning emulator (written in Python) into GEOS-Chem (written in Fortran) is a software challenge that needs to be addressed in order for this model to be a viable alternative.

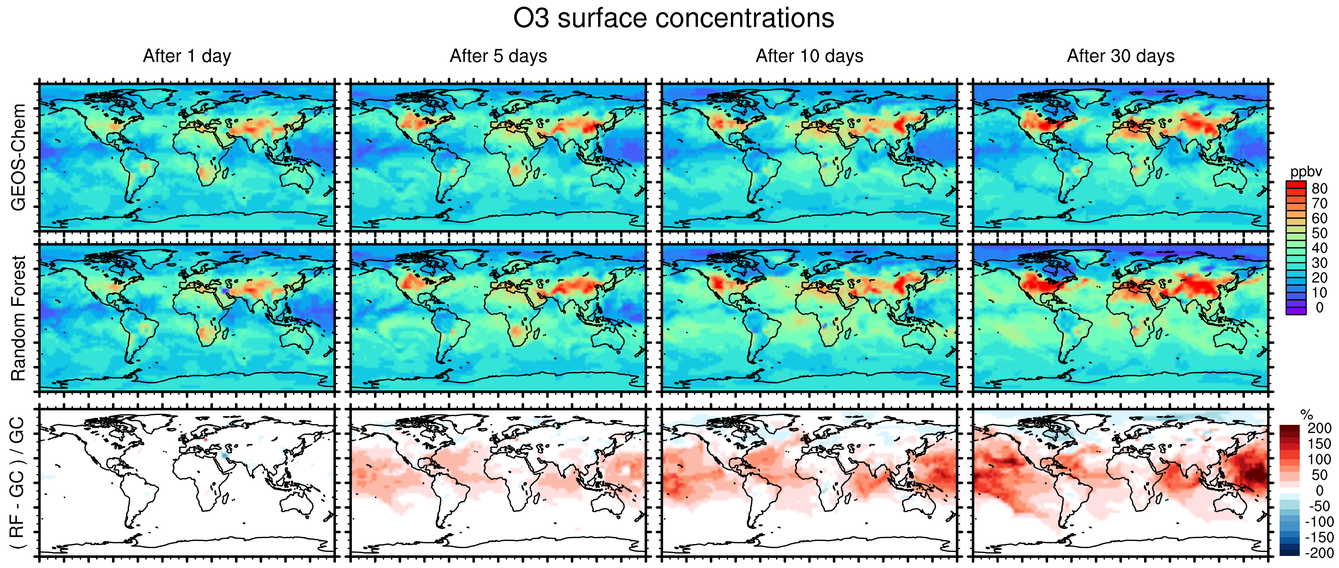


Figure 2: Comparison of surface ozone simulated by GEOS-Chem (top row) and predicted by the random forest model (middle row) after 1 day, 5 days, 10 days, and 30 days.

REFERENCE

Keller and Evans, Geosci. Model Dev., 12, 1209–1225, https://doi.org/10.5194/gmd-12-1209-2019, 2019.

E-mail: christoph.a.keller@nasa.gov
Web: gmao.gsfc.nasa.gov

USRA

National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

University of York

National Aeronautics and Space Administration
NASA