Machine Learning Application to Atmospheric Chemistry Modeling

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

Introduction

Atmospheric chemistry models are computationally expensive. For example, running the GEOS-Chem model at 0.25 degree horizontal resolution requires 3500 CPU's. **We developed an alternative approach to simulate atmospheric chemistry based on machine learning that has the potential to be much faster.**

Chemistry Model

Our reference is the GEOS-Chem model v10 embedded into the NASA GEOS-5 model. The training data consists of hourly model output (July 2013) of the chemical species' concentrations immediately before and after chemistry, the photolysis rates, and 7 meteorological variables. The full training data set consists of 2.4 x 10^8 data points.

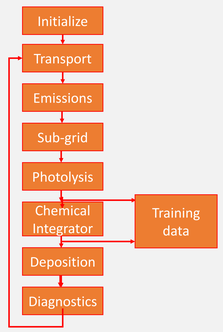
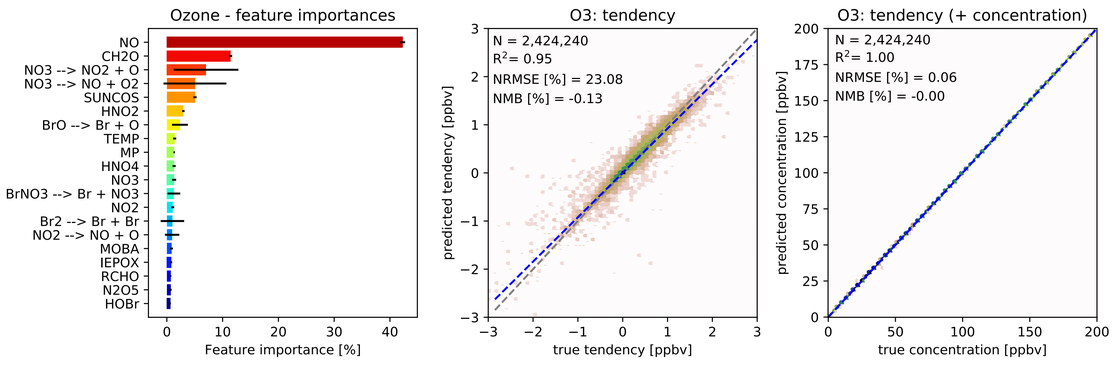


Figure 1: Characteristics of ozone prediction model based on random forest regression



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.

Machine Learning

The random forest regression algorithm is used to generate a prediction model for each chemical species (30 trees, 10,000 leaves). The models predict change in concentration for long-lived species and absolute concentration for short-lived species. NO and NO_2 are predicted as a family (NO_x).

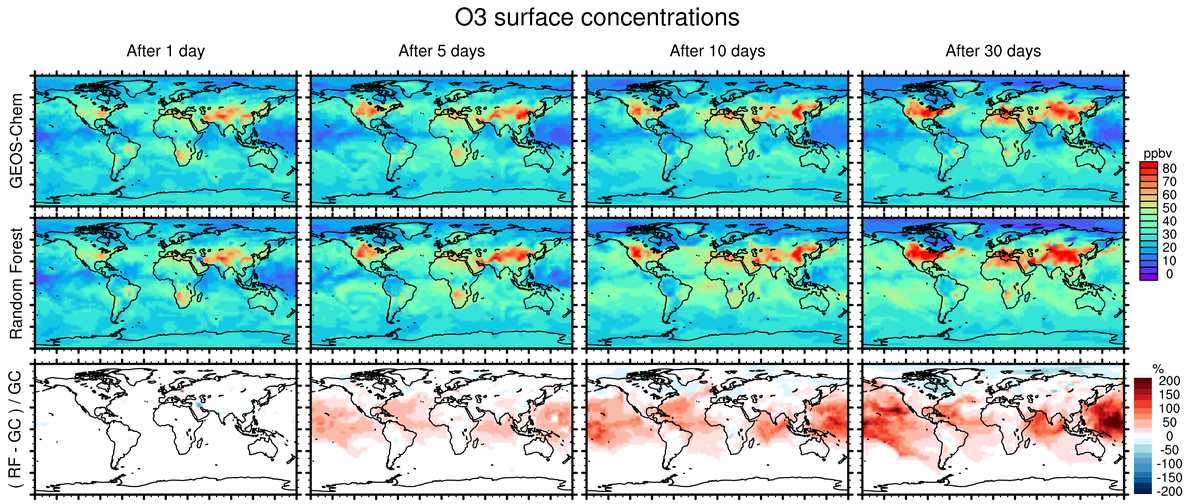
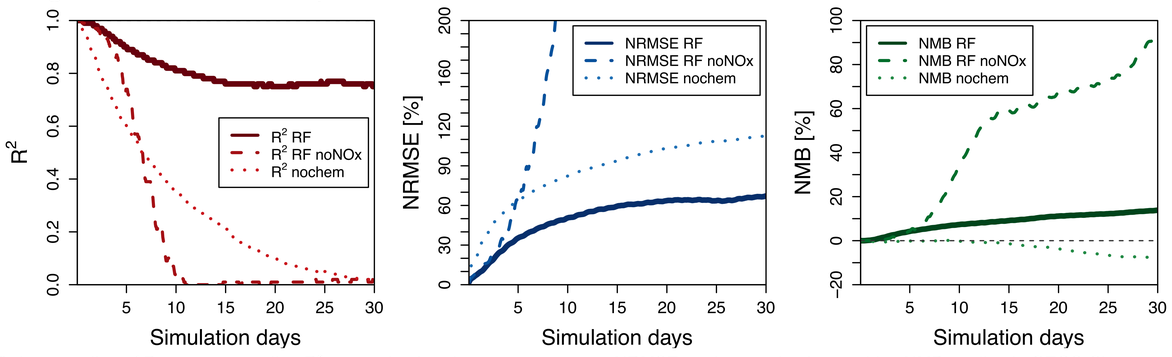


Figure 2: Comparison of surface ozone simulated by GEOS-Chem (top row) and predicted by the random forest model (middle row) after 1 day, 5 days, 10 days, and 30 days.

One-Month Simulation Using Random Forest

The random forest simulation is able to realistically reproduce ozone concentrations across the globe, albeit it overpredicts ozone over remote regions, caused by an overprediction of NO_x in those areas.

Figure 3: Statistical comparison of random forest emulator vs. GEOS-Chem reference simulation



30-day evolution of Pearson correlation R^2, normalized root mean square error NRMSE, and normalized mean bias NMB relative to GEOS-Chem reference simulation for a run using the random forest emulator (RF), a RF simulation without NO_x family prediction, and a simulation with no chemistry.

Conclusions

We show that machine learning methods have the potential to emulate atmospheric chemistry at time scales of days to weeks. A single prediction is ~1000 times faster than the numerical integrator.

Reference: Keller and Evans, GMDD, https://doi.org/10.5194/gmd-2018-229, in review.

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