Geoscientific
Model Development
Discussions

Interactive comment on "Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-Chem chemistry model v10" *by* Christoph A. Keller and Mat J. Evans

Nowack

p.nowack@imperial.ac.uk

Received and published: 27 November 2018

This is a highly interesting paper, which I enjoyed reading. Just a few comments that I would recommend considering/discussing in a revised version:

1. Using random forests, the authors here focus on emulating the chemical system of an air pollution forecasting model for the troposphere. However, machine learning has also been used to forecast air pollution itself (e.g. Mallet et al. 2008).

2. Composition changes such as ozone are also important for climate, which is a topic not considered here, but worth mentioning in the Discussions (cf. Nowack et al. 2018). I assume due to the use of a chemistry-transport model you did not consider feedbacks of ozone on meteorology/climate? It would be good to briefly contrast the challenges in air pollution modelling as compared to climate modelling where stratospheric ozone changes are particularly important and still expensive to calculate (cf. Nowack et al. 2015).

3. Concerning the selection of the cross-validation method: since you predict time series of chemical species concentrations/concentration changes, the samples for longer-lived species are not independent. From the current description in the paper, it seems that these time-dependencies were not taken into account in the cross-validation. Are the authors using a sequential cross-validation method, e.g.

   http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.TimeSeriesSplit.html

   which was also used and described in Nowack et al. (2018)?

4. How is the boundary between the troposphere and stratosphere (where chemistry is interactive but linearized) handled? Is there any effect of the tropospheric machine learning predictions on the stratosphere? If yes, could this in turn affect some of the tropospheric results, for example, due to changes in the photochemical environment, or transport?

References:

1. Mallet V., Stoltz G., Mauricette B. Ozone ensemble forecast with machine learning algorithms. Journal of Geophysical Research: Atmospheres 114, D05307, doi:10.1029/2008JD009978, (2009).

2. Nowack P., Braesicke P., Haigh J., Abraham N. L., Pyle J., Voulgarakis A. Using machine learning to build temperature-based ozone parameterizations for climate sensitivity simulations. Environmental Research Letters 13, 104016, doi:10.1088/1748-9326/aae2b, (2018).

3. Nowack P. J., Abraham N. L., Maycock A. C., Braesicke P., Gregory J. M., Joshi M. M., Osprey A., Pyle J. A. A large ozone-circulation feedback and its implications for global warming assessments. Nature Climate Change, 5(1), 41–45. doi:10.1038/nclimate2451, (2015).