Response to interactive comment by P. Nowack

We thank Peer Nowack for his valuable comments. Please find below our comments.

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).
There are many potential applications for machine learning in atmospheric chemistry modeling, and this paper highlights one use case. We chose to focus on the emulation of chemical kinetics as this process is the computational bottleneck in chemistry transport models. However, as pointed out, alternative approaches could be to directly predict air pollution of selected target species, and/or to emulate the entire modeling system (transport, diffusion, emission & deposition, chemistry). We extended the introduction/discussion as follows to better highlight our motivation and other possible approaches:

"Machine learning emulators have also been explored to directly predict air pollution concentration in future time steps (Mallet et al., 2009), as well as for chemistry-climate simulations focusing on model predictions of time-averaged concentrations for selected species such as ozone and OH over time scales of days to months (Nicely et al., 2017; Nowack et al., 2018). In contrast, the algorithm presented here is optimised to capture the small-scale variability of the entire chemical space within a time scale of minutes, with only a small loss of accuracy when used repeatedly over multiple time steps."

"Another potential application area for machine-learning based chemistry emulators are chemistry-climate simulations. Unlike air quality applications, which focus on small-scale variations of air pollutants over comparatively short periods of time of days to weeks, chemistry-climate studies require long simulation windows of the order of decades. Because of this, machine learning models used for these applications need to be optimised such that they accurately reproduce the (long-term) response of selected species - e.g. ozone and OH - to key drivers such as temperature, photolysis rates and NOx (Nicely et al., 2017; Nowack et al., 2018). The here presented method could be optimised for such an application by simplifying the problem set, with the model trained to reproduce daily or even monthly averaged species concentrations."

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).
We indeed ignore here the feedbacks of atmospheric constituents on weather and climate since we use a chemical transport model where the meteorology is fixed. In addition, our focus is on the accurate representation of chemistry at high temporal resolution (hours) but over a relatively short period of time (weeks to months). For chemistry-climate applications, one would need to prioritize the long-term stability and chemical balance of the machine learning system over the short-term accuracy of local surface concentrations. We discuss this now in more detail in the discussions, as already highlighted above. In addition, we added the following sentence to the model description:

"While the GEOS model with GEOS-Chem chemistry can be run as a chemistry-climate model where the chemical constituents (notably ozone and aerosols) directly feed back to the meteorology, we

disable this option here and use prescribed ozone and aerosol concentrations for the meteorology instead. This ensures that any differences between the reference model and the machine learning model can be attributed to imperfections in the emulator, rather than changes in meteorology due to chemistry-climate feedbacks."

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)?
We address the problem of time series correlation in two ways: (1) our emulator is time-agnostic, i.e. we focus on the prediction of concentrations / tendencies due to chemistry at time X based on the chemistry initial conditions at the same time (rather than e.g. predicting concentrations at X + Δt based upon concentrations at time X); (2) by predicting the change in concentrations for long-lived species instead of absolute concentrations, we reduce the temporal correlations between subsequent time steps. Having that said, we fully acknowledge that the training samples are not fully independent, and neither are the input variables. However, one of the advantages of the random forest algorithm is its relative robustness even when using correlated variables as long as the full sampling set is representative of the underlying process, which we consider to be the case for the time period considered in this study.

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?
For our study we consider the impact from stratosphere-troposphere exchange (STE) to be small given the relatively short time window. We address this in more detail in section 3:

"In all simulations the stratospheric chemistry uses a linearised chemistry scheme (Murray et al., 2012). This buffers the impact of the RFR emulator over the long-term since all simulations use the same relaxation scheme in the stratosphere. For the here considered time frame of one month, we consider this impact to be negligible in the lowest 25 model levels."