Atmos. Chem. Phys. Discuss.,
https://doi.org/10.5194/acp-2019-753-RC1, 2019

Atmospheric Chemistry and Physics
Discussions
Open Access

ACPD

Interactive comment

Interactive comment on "Improving the prediction of an atmospheric chemistry transport model using gradient boosted regression trees" *by* Peter D. Ivatt and Mathew J. Evans

Anonymous Referee #2

Received and published: 30 October 2019

The paper by Ivatt and Evans applies a machine learning technique for predicting biases in O3 simulations based on observations of O3 collected over previous time periods , and the statistical relationship between these and the model's chemical and physical state. The topic is interesting, timely, and suitable for this journal. The manuscript is generally clear and well written. My main concern is that is a little short / thin, particularly with regards to the relevance of this type of approach for air quality forecasting. The authors mention this numerous times, so it seems to be one of their prime motivating applications. However, from a lack of discussion of the background on this topic, to a lack of depth on exploration of the applicability of their methods to actually air quality forecasting needs, this aspect falls a bit short. The overall paper itself is on

Printer-friendly version

Discussion paper

the shorter side, so it seems with some revisions and substantial additions, the paper could become more applicable in this regard. That is is a “demonstration case” should not be grounds for incomplete context, analysis, or making claims beyond what has been actually shown.

Specific comments: The reference to Gaudel 2018 seems misplaced. The paper is nearly exclusively about trends in observed O3. In support of a statement regarding model biases, the authors are referred to the Young 2018 TOAR paper.

The introduction is too thin on the topic of O3 bias correction in models. There is a long, extensive history of O3 bias correction within the AQ literature. See for example Kang et al., 2010, https://doi.org/10.1016/j.atmosenv.2010.03.017, and half dozen or so papers cited in the introduction therein, and also additional research on the topic for more recent studies is warranted.

The authors bring up AQ forecasting frequently as an application. Concentrations and chemical environments relevant to forecasting seem to me much more highly variable at the scales of most forecasts (10’s of km) compared to the analysis here (100’s of km). How does that impact the authors conclusions regarding the applicability of their results? Would these techniques be expected to capture gradients in O3 biases between urban cores and surrounding areas? I don’t see such issues presently discussed.

The observational dataset seems thin, particularly in Asia, given there is O3 data accessible there, through TOAR itself.

156: I don’t really buy this explanation. The ML approach doesn’t care if there is a true fundamental physical relationship, in reality. It only cares if there is a statistical relationship. The authors thus need to explain why a coastal cite degrades the statistical relationship. Further, I suspect the statistical relationship may be weak here owing to the importance of upwind sources in this region from China, which have a larger association with local O3 than the local model state.

Printer-friendly version

Discussion paper

Did the authors ever think about expanding the physical range of the model state that is included as input for the forecast in any one grid cell? This is done in the field of statistical prediction of PM, for example, since it is known that upwind conditions can drive local PM more than local conditions, especially when forecasting PM at high resolution. I would suspect the situation to be similar for O3. The present study may artificially benefit from the coarse model resolution not really resolving local O3 to begin with, but for future studies with high resolution models, this could become an important consideration.

How does the computational training time scale with the number of grid cells considered? This is an important consideration when considering the applicability of this approach to higher resolution simulations.

Fig 6 is great and left me wanting much more. Can the authors present this as well in terms of diurnal variability? Seasonal variability? Does the overall quantification of bias agree with biases noted in ensembles of air quality models in the Young 2018 TOAR paper (which seemed to have hemispheric N/S patterns, or is GEOS-Chem distinct?

The explanation for why O3 is an important predictor is a bit weak. I'm not sure I believe this is strictly an Antarctic / low-O3 result. But did the authors evaluate the spatial distribution of the importance of these predictors? That would certainly be interesting to see.

For the local model state, the authors didn't include land type or dry deposition velocity, which seem like would be important for correcting model biases associated with O3 loss, which is a known issue with these types of models.

Conclusions regarding the extent of data for training seem potentially biased by the way the authors have designed their performance metric. For AQ forecasting applications where the metric of performances is very short - term, it seems that training based on only the most recent conditions could be of more values, as indicated by the literature in that field.

Printer-friendly version

Discussion paper

It’s a bit scary that removal of training data in areas like Cape Verde or South Africa make the predictions in these locations worse. Granted these locations are distinct from other areas, which performed fine or adequately when corresponding training data was removed. But my question is – how would we know, for a location with no training data – whether or not the bias corrections predicted here would help or degrade the simulation?

250: It seems like there could be regionally specific biases in meteorology that are not necessarily global in nature.

269 - 271: Not sure if I clearly understand the explanation being put forth here – could the authors expand?

277: The authors claim these methods offer a route to significant improvement in the fidelity of forecasts, but I’m not convinced the applicability to the priorities of air quality forecasting has really been demonstrated. We were not presented with any timeseries results. Nor was there evaluation of the extent to which this approach helped with prediction of exceedances or extreme values (or a reduction of false alarms). The results averaged over the entire coarse of the year tend to wash out the features that would be of most interest in a forecasting application.

Corrections: 102: tree, –> tree 132: the the 251: hemisphere(–> hemisphere (255: and so –> so 271: observed –> observations

---

Interactive comment on Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-753, 2019.