

Geosci. Model Dev. Discuss., referee comment RC1
<https://doi.org/10.5194/gmd-2021-347-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on gmd-2021-347

Anonymous Referee #1

Referee comment on "Simulation Model of Reactive Nitrogen Species in an Urban Atmosphere using a Deep Neural Network: RND v1.0" by Junsu Gil et al., Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2021-347-RC1>, 2022

Review of gmd-2021-347

Simulation Model of Reactive Nitrogen Species in an Urban Atmosphere using a Deep Neural Network: RND v1.0

by Junsu Gil et al.

General comments:

This manuscript describes a new application of a simple feed forward neural network model to calculate HONO mixing ratios based on a set of other measured variables. While this is an interesting and worthwhile application, the paper lacks the necessary details in the description of the deep learning model and contains no ablation studies which are needed to provide the credibility in the results. I also question the validity of the cross validation and test cases that are discussed, because I doubt that these test cases are truly independent data samples. There is no proof of the generalisation capability of the model, so it may well be that this model fails completely if it were applied to measurement data obtained under different conditions.

In summary, this manuscript falls between "major revisions" and "reject". In computer science conferences it would be ranked "weak reject", which means the paper could be saved if the authors invest substantial work in rerunning their model several times and improving the text.

Specific comments:

Abstract: Confusing sentence after "In this study,". After reading 3 times I understood that you are resolving the acronym RND here, but this is well hidden. Suggestion: "In this study, a new simulation approach to calculate HONO mixing ratios using a deep learning technique based on measured variables was developed. The 'Reactive Nitrogen species Deep neural network' (RND) has been implemented in Python. It was trained, ..."

Abstract: Why should RND be called a *supplementary* model? What does it supplement?

I.35: too vague "observational constraints on individual species". Does this refer to NO_y compounds or any species involved in the tropospheric ozone production cycle?

I.40: NO_y has been the focus of attention already in the 1990s. See for example papers by Sandy Sillman et al. You may say "renewed attention".

I.43: to the uninitiated reader it might not be clear what heterogeneous reactions have to do with NO_y and ozone chemistry. This would merit one or a few more general sentence(s) to describe NO_y chemistry. If this text will get a little longer, please consider summarizing the HONO/NO_y chemistry in a supplement and refer to it. Nevertheless, one or two sentences will be needed here.

I.44: you could add <https://doi.org/10.5194/acp-18-3147-2018> to the list of references here.

I.52/53: it would be useful to know if there is general agreement among these different measurement methods or if they haven't reached a satisfactory level of consistency yet. In the following sentence, please provide some order of magnitude numbers of observed versus simulated HONO levels (or a value range).

I.57: the recent adaptation of machine learning techniques in atmospheric sciences is more general than "multi layer artificial neural network". In this context, it suffices to say that "machine learning" has been adopted. Then, in a following sentence you can narrow this down to the employment of deep (artificial) neural networks, which have a capability to learn more complex non-linear relations in data, but also require larger amounts of data for training." The selection of references appears a bit arbitrary. For example, there is a whole special issue in Philosophical Transactions A () on machine learning for weather and climate. Indeed, you may want to first provide two or three general references for ML in atmospheric science (with cf.), then write a sentence which refers specifically to atmospheric chemistry/atmospheric composition and provide some more references there.

I.59-62: the description why deep learning might be useful for the analysis of atmospheric chemical measurements remains vague and superficial. You should state explicitly that neural networks learn relations in data (similar to function fitting) and you should state in what way NNs may improve on numerical simulations (I guess you refer to the fact that they are inherently bias-free?).

I.62/63: introduction of the model acronym: difficult to disentangle the sentence - see comment on abstract above.

I.67: as this is supposed to be a manuscript for the special issue on "machine learning methods and benchmark datasets", you should add a statement here that the code and training data can be downloaded from ..." (you can of course also refer to the code and data availability section here). Re-usability of your model is a key aspect for this special issue (and for GMD in general).

I.70: the steps which are described don't guide the development of RND, but describe the typical machine learning workflow.

I.77: similar issue - this reads as if every user of RND will first have to perform measurements for her/himself. Please separate the dataset preparation from the model development. The model should be generalizable, i.e. be independent of the specific set of measurements which you describe in the paper.

I.105: "wind direction should be converted..." - please describe what you did, not what should be done.

I.106: "missing values" same as above. Did you filter or interpolate?

I.107: what is an "array of measurement data"? Also, what is missing is a description of the time resolution of the measurements and how many independent samples were prepared for the machine learning. How was the train-test-val split done? Have you checked the frequency distributions of the (normalized) variables? Have you considered log transform for non Gaussian variables? How many time steps are included in each sample?

Section 2.3: there is a lot of information missing from the network description: how many nodes per layer? What is the learning rate? How many epochs were trained? Did the learning rate change during training? Did you try out different numbers of layers and

nodes per layer to determine the optimum model? Did you perform a hyperparameter search? Also, what exactly is the input data and what exactly is the target output? Loss function... Those things are standard in the machine learning literature and should be adhered to. I see some of this information appears in the figures and the following section (varying the number of nodes), but this belongs in the model description text.

I.136: if June 2018 has been used in the training already, then this month is not an ondependent test dataset any more.

I.154: does this mean that you always used the same number of nodes in each layer? And you did not try to reduce the number of layers? 1600 samples appears rather small for a network with 5 layers.

I.160: I don't understand this. First you train the network for 2016 to 2019, then you run it again to obtain HONO results? You already have them from the training.(?)

I.167: I doubt that the inability fo the model to capture minima and maxima is due to the limited amount of data. This is a general aspect of regression models and extensively discussed in Kleinert et al (2021): <https://doi.org/10.5194/gmd-14-1-2021>

I.205 and following: this discussion of atmospheric chemsistry doesn't belong into a section describing the application of the model. Is this supposed to be a general discussion section, comparing RND to other (CTM) models?

I.235 Finally, here is a list of the input variables. But is has not been discussed, which variable has which influence on the results. I have a suspicion that the network really makes use only of 3 or 4 of the 9 variables it is given. See Kleinert et al. (2021) for a way how this can be tested with bootstrapping.

I.250/251: the ML model doesn't gain any physical understanding of the HONO chemistry, so it cannot be used to test the existing knowledge. You could use such a tool to forecast HONO levels, for example to determine if it might be worthwhile conducting HONO measurements at a specific location or during a specific time period. You may also be able to use the tool in the context of quality controlling the measurements: any strong disagreement would raise a warning that measurements should be checked with extra care.

Also, you can of course use it to estimate HONO concentrations when these were not measured in order to then perform 0D model runs, as you show in Figure 8. And in this light, I would agree with the statement that RND is a "supplementary tool".

I.262: please provide an explicit URL here (you can still add the reference)

Technical corrections:

I.55: related to the comment on I.43: you presume that the reader is familiar with the basics of HONO chemistry, but this cannot be taken for granted.

I.30 play instead of plays

I.34 and *it* determines...