

Geosci. Model Dev. Discuss., referee comment RC2  
<https://doi.org/10.5194/gmd-2021-347-RC2>, 2022  
© Author(s) 2022. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## **Comment on gmd-2021-347**

Anonymous Referee #2

---

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-RC2>, 2022

---

In this paper, deep neural network based model is used to calculate nitrous acid (HONO) mixing ratios based on the analysis using HONO measurement data from Seoul between 2016 and 2019. Since I am not an expert in atmospheric sciences, but in data and computer science, I will in my review focus on the computational method used and its validity based on the size and type of the data.

The paper is generally well written and takes action to document the use of the suggested model. The citation to code availability is missing DOI (and one has to go over to Zenodo to locate the code)

The approach taken is motivated by the success of deep learning based methods in various areas. However, here (as often elsewhere) it is not taken into account, that deep learning is most useful in situations in which there are massive amounts of training data — which is not the case here. There are nine input features and there are 1636 data items (1122 for training and 514 for validation). Hence, the data is not really massive and because the amount of interactions is limited (only nine input variables), it is quite likely that more traditional machine learning methods would work well (e.g., ordinary linear regression could be used to provide a baseline (and could even suffice), then one could see how e.g., support vector machine or random forest would work). In the paper, the use of deep neural networks is argued by them being more useful than traditional models, because they are able to handle large amounts of data. For the data used, there is no reason to assume that it could not be handled using also some of the traditional methods, in particular, when the data is small, more complicated models are quite prone to overfitting.

*Suggestion for improvement 1:* Test different ML learning models to be able to evaluate properly the usability of the suggested model.

My second concern is the feature selection or the lack of it. The model blindly uses the nine input variables from the data. This kind of "taking an ML model off-the-shelf" very rarely produces the best possible results and can seriously affect the performance of the model. In addition to feature selection, it might be also possible to compute some surrogate features, e.g., provide information about dependencies in the modelling domain, reducing the need for the ML models to explicitly model these dependencies.

*Suggestion for improvement 2:* Use feature selection (for all the models) to search for a best possible set of input features.

Finally, the testing of the model using data from April 2019, shows some of the limitations of the developed model. It seems that there is an occurrence of concept drift (when the distribution of data changes, the model does not work well anymore). Also, the error might increase due to overfitting of the model. This aspect should be studied further, in particular it would be important to be able to provide the region in which the model's accuracy is on an acceptable level. There is a rich body of literature in detecting concept drift (for a survey, e.g., see Zliobaite I., Pechenizkiy M., Gama J. (2016) An Overview of Concept Drift Applications. In: Japkowicz N., Stefanowski J. (eds) Big Data Analysis: New Algorithms for a New Society. Studies in Big Data, vol 16. Springer, Cham. [https://doi.org/10.1007/978-3-319-26989-4\\_4](https://doi.org/10.1007/978-3-319-26989-4_4)).

*Suggestion for improvement 3:* Analyse the region in which the proposed model can be expected to work, at least provide some discussion on the effect of overfitting and concept drift and how these affect the usability of the model.

Based on these observations, I would reject the paper in its current form, with the encouragement to resubmit, taking the suggestions for improvement into account.