0. 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.

- Thank you for your constructive comments and helpful advice. The point-by-point responses are given below, along with relevant parts of the revised manuscript, where all changes are marked in blue.

1. 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 DOI of reference is updated (Gil, J.: RNDv1.0 and example, https://doi.org/10.5281/zenodo.5540180, in, Zenodo, 2021)

2. 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), its 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.

- You are absolutely right. In general, the performance of deep learning (DL) is better than or at least similar to traditional machine learning (ML) such as support vector machine or random forest (Sumathi et al., 2020; Baek et al., 2021). This advantage would be greater with larger data set and even small data set can benefit from it (Dang et al., 2020). DL is also known to be better than general liner regression for data in non-linear relationship.

The test result of RNDv1.0 demonstrates that it reasonably represents ambient HONO levels and captures well the averaged variation. In comparison, it tends to underestimate high concentrations. This is a weakness of our model but indicates that our model does not overfit the training dataset.

In the revised manuscript, introduction is fully revised with background information on HONO and the application of DNN to HONO simulation.

Line 85-104: Among ML methods, the Neural Network (NN) architecture is widely used owing to its powerful ability to process large amounts of data, allowing improvement in the performance of conventional models through being integrated with physical equations (Reichstein et al., 2019;Schultz et al., 2021). As a NN architecture, a multi-layer artificial neural network, referred to as a Deep Neural Network (DNN), employs a statistical method that learn non-linear relations in data and obtain the optimum solution for the target species without prior information on the physicochemical processes. DNN has advantages over other NN architecture such as Convolution NN (CNN) or Long-Short Term Memory (LSTM) because it works well for discrete spatiotemporal data. In general, the performance of DNN is similar to or better than other ML methods for small number of data as well as large data set (Baek and Jung, 2021;Dang et al., 2021;Sumathi and Pugalendhi, 2021).

When the DNN method is applied to atmospheric chemical constituents, it requires large amount of data for training and thus, the size of measurement data becomes a limiting factor for trace species such as HONO, which are not routinely measured such as O3 or PM2.5. In this regard, the daily average HONO mixing ratio was attempted to be estimated using ensemble ML models with satellite measurements (Cui and Wang, 2021). In comparison, the hourly HONO mixing ratio was calculated using a simple NN architecture with measured variables, which were thought to be closely linked with HONO formation (Gil et al., 2021). The accuracy of the hourly HONO estimated from input variables such as aerosol surface areas and mixed layer height was better than the daily HONO estimate.

3. 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.

- The OH produced from HONO photolysis will fuel the photochemical formation of O3 and PM2.5, which are target species of 0-dimensional photochemical models and chemical transport models (CTM). It is demonstrated in section 3 that the presence of HONO has a significant contribution to the performance of photochemical model.

  In this regard, the purpose of this study is to construct a model for estimating the HONO mixing ratio using atmospheric variables that are continuously and routinely measured, but not to improve the performance of model in which the accuracy matters. We hope that our recent observations will be incorporated into the RND model, and the model will be able to provide robust HONO concentrations for operational forecasting models in the future.

  In a previous study, we built a simple Neural Network model that estimated HONO mixing ratio, and we know that selecting the appropriate variables can increase the accuracy of the model (Gil et al., 2021). In this study, we aim to construct a kind of universal and cheap model to estimate HONO concentration in urban areas using atmospheric variables provided through measurement networks. These input variables that were used for model construction did not show any meaningful correlations (Figure S2)

  In addition, bootstrap test similar to what was done in Kleinert et al. (2021), was conducted by setting each variable to zero with keeping other values and the results were compared with measurements. Among nine input variables, NO2 was found to have the most significant influence on

HONO concentration, followed by RH, temperature, and solar zenith angle (Table S2). This result is in good agreement with our previous study (Gil et al., 2021), implying that the input feature used for the model are suitable for estimating HONO concentrations.

In the revised manuscript, the detailed feature selection process is stated in Section1 and Section2.

- Line 105-107: In this study, we aimed to construct a user-friendly 'Reactive Nitrogen species simulation using DNN' (RND) model and estimate HONO mixing ratio using routinely measured atmospheric variables in a highly polluted urban area.
- Line 151-154: As input variables, hourly measurements of chemical and meteorological parameters are used, including the mixing ratios of O3, NO2, CO, and SO2, along with temperature (T), relative humidity (RH), wind speed (WS), wind direction (WD), and solar zenith angle (SZA) to estimate the target species, HONO, as the output.
- Line 241-253: 2.5. Influence of input variables to HONO concentration

  Additionally, a simple bootstrapping test was conducted by setting each variable to zero with keeping other variables (Kleinert et al., 2021). Then, the importance of each input variable to HONO concentration was evaluated using MAE and root mean square error (RMSE). Of nine input variables, NO2 was found to have the most significant influence on HONO concentration, followed by RH, temperature, and solar zenith angle (Table S2). The result of bootstrap test is in good agreement with those of our previous study (Gil et al., 2021), where more detailed information such as aerosol surface area and mixing layer height were incorporated into the model and highlighted the role of precursor gases and heterogeneous conversion in HONO formation. Therefore, these results demonstrate that the RND model constructed using routinely observed variables, reasonably traced the level of HONO in urban atmosphere.

4. 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).

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 theses affect the usability of the model.

- Atmospheric parameters including meteorological factors and chemical constituents show clear diurnal variations, especially in urban areas with high anthropogenic emissions. For example, NO2 reached the maximum during the morning rush hour, decreased down to the minimum in the afternoon, and increased at nighttime. This type of variation remained nearly constant through the year with changes in seasonal amplitude depending on emissions and meteorological factors determining the dilution and transport of air pollutants. The variation in O3 is just opposite to NO2.
- Our model was constructed for urban applications. When the model was tested against data obtained April, model uncertainty was increased. Although our model was trained and validated with data obtained during May-June, the variations in input variables for test period were similar to those of train-validation periods. Considering the result of previous study about HONO formation mechanism, the increased model uncertainty could be due to some factors that were not constrained in the model such as aerosol surface areas.

  Therefore, it is quite likely that the increased model uncertainty is not associated with the occurrence of concept drift.

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.

Citation: https://doi.org/10.5194/gmd-2021-347-RC2

Reference in answers

Baek, W.-K., and Jung, H.-S.: Performance Comparison of Oil Spill and Ship Classification from X-Band Dual-and Single-Polarized SAR Image Using Support Vector Machine, Random Forest, and Deep Neural Network, Remote Sensing, 13, 3203, 20

Sumathi, S., and Pugalendhi, G. K.: Cognition based spam mail text analysis using combined approach of deep neural network classifier and random forest, Journal of Ambient Intelligence and Humanized Computing, 12, 5721-5731, 2021.

Dang, C., Liu, Y., Yue, H., Qian, J., and Zhu, R.: Autumn crop yield prediction using data-driven approaches:-support vector machines, random forest, and deep neural network methods, Canadian Journal of Remote Sensing, 47, 162-181, 2021.21.

Cui, L., and Wang, S.: Mapping the daily nitrous acid (HONO) concentrations across China during 2006-2017 through ensemble machine-learning algorithm, Science of The Total Environment, 147325, 2021.

Gil, J., Kim, J., Lee, M., Lee, G., Ahn, J., Lee, D. S., Jung, J., Cho, S., Whitehill, A., Szykman, J., and Lee, J.: Characteristics of HONO and its impact on O3 formation in the Seoul Metropolitan Area during the Korea-US Air Quality study, Atmospheric Environment, 2021, https://doi.org/10.1016/j.atmosenv.2020.118182., 2021.

Kleinert, F., Leufen, L. H., and Schultz, M. G.: IntelliO3-ts v1. 0: a neural network approach to predict near-surface ozone concentrations in Germany, Geoscientific Model Development, 14, 1-25, 2021.