Simulation Model of Reactive Nitrogen Species in an Urban Atmosphere using a Deep Neural Network: RNDv1.0

Junsu Gil[1], Meehye Lee[1*], Jeonghwan Kim[2], Gangwoong Lee[2], Joonyoung Ahn[3], Cheol-hee Kim[4]

[1] *Department of Earth and Environmental Sciences, Korea University, Seoul, South Korea*

[2] *Department of Environmental Science, Hankuk University of Foreign Studies, Yongin, South Korea*

[3] *Air Quality Forecasting Center, Climate and Air Quality Research Department, National Institute of Environmental Research (NIER), Incheon, South Korea*

[4] *Department of Atmospheric Science, Pusan National University, Busan, South Korea*

* *Corresponding author*: Meehye Lee (meehye@korea.ac.kr)

Abstract

Nitrous acid (HONO), one of the reactive nitrogen oxides (NO_y), plays an important role in the formation of ozone (O_3) and fine aerosols ($PM_{2.5}$) in the urban atmosphere. In this study, a new simulation approach to calculate HONO mixing ratios using a deep neural technique based on measured variables was developed. The 'Reactive Nitrogen species simulation using Deep neural network (RND)' has been implemented in Python. The first version of RND (RNDv1.0) was trained, validated, and tested with HONO measurement data obtained in Seoul during the several months from 2016 to 2021.

RNDv1.0 was constructed utilizing k-fold cross validation and evaluated with an Index Of Agreement (IOA), correlation coefficient (r), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). The RNDv1.0 adequately represents the main characteristics of the measured HONO and thus, RND v1.0 is proposed as a supplementary model for calculating the HONO mixing ratio in a polluted urban environment.

1. Introduction

Surface ozone (O_3) pollution has been reported to be worsen over continental areas (Arnell et al., 2019;Monks et al., 2015;Varotsos et al., 2013;IPCC, 2014). In particular, a warmer climate is expected to increase surface O_3 concentrations and peak levels in polluted regions, depending on its precursor levels (IPCC 2021). As one of the short-lived climate pollutants (SLCPs), O_3 also interacts with the global temperature via positive feedback (Shindell et al., 2013;Myhre et al., 2017;Stevenson et al., 2013). Therefore, it is imperative to accurately predict the mixing ratios and variations of surface O_3. While operational models such as community multiscale air quality (CMAQ) have been used widely for this purpose, uncertainties still arise from poorly understood chemical mechanisms involving reactive nitrogen oxides (NOy) and volatile organic compounds (VOCs), and lack of their measurements (Mallet and Sportisse, 2006;Canty et al., 2015;Akimoto et al., 2019;Shareef et al., 2019;Cheng et al., 2022).

In the urban atmosphere, NO_y typically includes NO_x ($NO + NO_2$), HONO, HNO_3, organic nitrates (e.g., PAN), NO_3, N_2O_3, and particulate NO_3^-. These species are produced and recycled through photochemical reactions until they are removed through wet or dry deposition (Liebmann et al., 2018;Brown et al., 2017;Wang et al., 2020;Li et al., 2020). NO_y play an important role in critical environmental issues concerning the Earth's atmosphere, spanning from local air pollution to global climate change (Sun et al., 2011;Ge et al., 2019). The oxidation of NO to NO_2, and finally to HNO_3, is the backbone of the chemical mechanism producing ozone (O_3) and $PM_{2.5}$ (particulate matter of size $\leqslant$ 2.5 µm), and it determines the oxidization capacity of the atmosphere. Recently, as O_3 has increased along with a decrease in NO_x emission over many regions including East Asia, interest in the heterogeneous reaction of reactive nitrogen oxides, which is yet to be understood, has been newly raised (Brown et al., 2017;Stadtler et al., 2018). Currently, the lack of measurement of individual NO_y species hindered a comprehensive understanding of the heterogeneous reactions (Anderson et al., 2014;Wang et al., 2017b;Chen et al., 2018b;Akimoto and Tanimoto, 2021;Stadtler et al., 2018).

In particular, there are growing number of evidence for heterogeneous formation of HONO in relation to high $PM_{2.5}$ and O_3 occurrence in urban areas (e.g., (Li et al., 2021b). As an OH reservoir, HONO will expedite the photochemical reactions involving VOCs and NO_x in the early morning, leading to O_3 and fine aerosol formation. Nonetheless, its formation mechanism has not been elucidated clearly enough to be constrained in conventional photochemical models. In addition to the reaction of NO with OH (Bloss et al., 2021), various pathways of HONO formation have been suggested from laboratory experiments, field measurements, and model simulations: direct emissions from vehicles (e.g., (Li et al., 2021a) and soil (e.g.,(Bao et al., 2022), photolysis of particulate nitrate (e.g., (Gen et al., 2022), heterogeneous conversion of NO_2 on various aerosol surfaces (e.g., (Jia et al., 2020), ground surface (e.g.,(Meng et al., 2022), and microlayers of sea surface (e.g., (Gu et al., 2022). Among these, the heterogeneous reaction mechanism on the surface is of major interest in the recent HONO study.

HONO has been measured mostly during intensive campaigns in urban areas using various techniques such as a LOng Path Absorption Photometer (LOPAP) (Kleffmann et al., 2006;Xue et al., 2019), Chemical Ionization Mass Spectrometry (CIMS) (Levy et al., 2014;Roberts et al., 2010), Ion Chromatography (IC) (VandenBoer et al., 2014;Gil et al., 2020;Ye et al., 2016;Xu et al., 2019), Monitor for AeRosols and Gases in ambient Ari (MARGA) (Xu et al., 2019), and Quantum Cascade - Tunable Infrared Laser Differential Absorption Spectrometry (QC-TILDAS) (Lee et al., 2011;Gil et al., 2021). Of these methods, QC-TILDAS has served as a reference for intercomparison of measurement data from different techniques due to high time resolution and stability (Pinto et al., 2014). These studies reported the maximum HONO of several ppb levels at nighttime. In comparison, the model captured at most 67~90 % of the observed HONO in megacities such as Beijing (Tie et al., 2013;Liu et al., 2019).

In recent years, Machine Learning (ML) method has been adopted in the atmospheric science for pattern classification (e.g. New Particle Formation event) and forecasting and spatiotemporal modelling of O_3 and $PM_{2.5}$ (Arcomano et al., 2021;Shahriar et al., 2020;Krishnamurthy et al., 2021;Cui and Wang, 2021;Joutsensaari et al., 2018;Chen et al., 2018a;Kang et al., 2021). 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 O_3 or $PM_{2.5}$. 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 deeply involved in the formation of HONO (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.

The aim of this study is to develop a user-friendly 'Reactive Nitrogen species simulation using DNN' model (RNDv1.0) that estimates HONO mixing ratios from real-time measurements of criteria pollutants and meteorological parameters and is ultimately to be incorporated into operational models that forecast urban air quality. Since this study is the first attempt to calculate the HONO mixing ratio using RNDv1.0, the entire construction process is described in detail, and the performance is evaluated by comparing the results with simulations using a commonly used model and observations over several years.

2. Model description

The development of RNDv1.0 model follows the systematic steps similar to a general machine learning model construction workflow, including collecting data, preprocessing data, building the DNN, training and validating the model, and testing the performance of the model (Figure 1). The RNDv1.0 was written in Python and necessary libraries to build and operate

RNDv1.0 are listed in Table 1. The dataset used to train-test-validation can be downloaded from Gil et al., 2021.

2.1. Collection of measurement data for model construction

As the first step constructing the RNDv1.0, measurement data were obtained including HONO, reactive gases, and meteorological parameters. It is noteworthy that the HONO measurement data is for model construction and is not required to run the RND model. The HONO mixing ratio was measured in Seoul using a QC-TILDAS system during May–June 2016, June 2018, and April–June 2019 (Lee et al., 2011;Gil et al., 2021) and a MARGA system during May–June and October–November 2021 (Gil, 2022). When testing and evaluating atmospheric HONO measurement methods, QC-TILDAS has been chosen as the reference method for comparing ambient HONO mixing ratios measured using several different techniques owing to its advantages of low detection limits (~ 0.1 ppbv) and high temporal resolution (Pinto et al., 2014). More details on measurements can be found elsewhere (Gil et al., 2021;Gil, 2022).

HONO was measured at Olympic Park (37.52°N, 127.12°E) during the Korea-United States Air Quality (KORUS-AQ) study in 2016 (Kim et al., 2020;Gil et al., 2021), at the campus of Korea University (37.59°N, 127.03°E) in 2018 and 2021, and at the site near the campus (37.59°N, 127.08°E) in 2019 (NIER, 2020) (Figure S1). Although measurements were made at three sites, O_3 and $PM_{2.5}$ levels have been known to be greatly influenced by the synoptic circulation throughout the Korean peninsula (Peterson et al., 2019;Jordan et al., 2020), and the Korea University campus and Olympic Park have served as measurement sites representing the air quality of Seoul. In addition to HONO, trace gases including O_3, NO_2, CO, and SO_2 and meteorological parameters including temperature (T), relative humidity (RH), wind speed (WS) and direction (WD) were measured. Note that HONO was not significantly correlated with any of these variables (Figure S2). The measurement statistics are presented in Table 2 and Table S1. Briefly summarizing, the 10^{th} and 90^{th} percentile mixing ratios of HONO, NO_2, and O_3 are 0.3 ppbv and 2.0 ppbv, 10.0 ppbv and 47.0 ppbv, and 8.0 ppbv and 75.0 ppbv, respectively for the entire experiment periods.

2.2. Data preprocessing

In the next step, the observation data set was prepared for RNDv1.0 model construction. As input variables, hourly measurements of chemical and meteorological parameters are used, including the mixing ratios of O_3, NO_2, CO, and SO_2, 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. Wind direction in degrees were converted to a cosine value for continuity. As a last step in data processing, missing values were filtered out from the input dataset. Finally, 54.2 % of all available measurement data (2847) were used to construct and evaluate the RNDv1.0 in this study.

Since the measurements of these nine variables vary over a wide range in different units, they were normalized to avoid bias during the calculations. Among the widely used normalization methods, '*min-max scaling*' method was adopted, and input variables were normalized against the minimum and maximum values in this study (Eq. 1):

$$x_{sca} = \frac{x_{raw} - F_2(X)}{F_1(X)}, \qquad \text{(Eq. 1)}$$

where x_{raw} is raw data, x_{sca} is scaled value, and F_1 and F_2 are scale factors of input variable (X), which are listed in Table 2.

2.3. Neural network architecture and hyperparameters

At this stage, the network is built to calculate HONO using those input variables. The RNDv1.0 is composed of five hidden layers (Figure 2), which employed an exponential linear unit (ELU) as an activation function (Eq. 2).

$$\mathrm{ELU}:\ \phi(x) = \begin{cases} e^{x} - 1\ (x < 0) \\ x\ (x \geq 0) \end{cases}. \qquad \text{(Eq. 2)}$$

In a DNN, an activation function creates a nonlinear relationship between an input variable and an output variable. When constructing a DNN model, an ELU has the advantage of a fast-training process and better performance in handling negative values than other activation functions (Wang et al., 2017a;Ding et al., 2018). In addition, the mean squared error and Adam optimizer were applied as loss function and optimize function, respectively. The learning rate, epoch, and batch were set to 0.01, 100, and 32, respectively.

2.4. Model training and k-fold cross validation

The RNDv1.0 model was trained-and-validated and tested with HONO measurements obtained during May ~ June in 2016 and June in 2018, April ~ June 2019, and May ~ June and October ~ November in 2021, respectively (Figure 3). The number of data used for train-validation and test were 1122 and 1725, respectively.

With the hyperparameters specified in previous section, the performance of the model was firstly validated using the K-Fold Cross-Validation (KFCV) method, which is especially useful when the size of dataset is small (Bengio and Grandvalet, 2003). In the KFCV method (Figure 3), the entire data is randomly divided into k subsets, of which k-1 sets were used for training and the rest one was used for validation. k was set to 5 in this study. The accuracy was determined by Index Of Agreement (IOA) expressed by the following equation (Eq. 3):

$$\mathrm{IOA} = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(|P_i - \bar{O}| + |O_i - \bar{O}|)^2}, \qquad \text{(Eq. 3)}$$

where O_i, P_i, $\bar{O}$, and n are the observed value, predicted value, average of the observed values, and number of nodes, respectively.

As IOA vary according to the number of nodes, it was calculated for the measured ($HONO_{obs}$) and calculated ($HONO_{mod}$) mixing ratios by varying the number of nodes from 0 to 100 in each hidden layer. The best performance was found with 41 nodes, with which the averaged IOA was 0.89 ± 0.01 (Figure 4). The high level of IOA demonstrates that the performance of RNDv1.0 model is adequate, and it is capable of simulating the ambient HONO mixing ratio using the routinely measured criteria pollutants and meteorological parameters.

The performance of RNDv1.0 was compared with that of other models, including Community Multi-scale Air Quality Model (CMAQv5.3.1, Appel et al., 2021), Random Forest (RF), and 1-layer Artificial Neural Network (ANN, Gil et al., 2021) using 2016 measurement data. A RF model was constructed using KFCV method and the same input parameters as RNDv1.0 (Figure S4). Their performance was evaluated by Mean Absolute Error (MAE), Root mean Square Deviation (RMSE), and Pearson correlation coefficient (r) (Eq. 4 - 6).

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}|O_i - P_i|}{n}, \qquad \text{(Eq. 4)}$$

$$\mathrm{RMSE} = \sqrt[2]{\frac{\sum_{i=1}^{n}(O_i - P_i)^2}{n}}, \qquad \text{(Eq. 5)}$$

$$r = \frac{cov(O,P)}{\sigma_O\,\sigma_P}, \qquad \text{(Eq. 6)}$$

where σ and cov implies standard deviation and covariance, respectively.

The measured HONO mixing ratios correlated well with those calculated except for the CMAQ (Figure 5), which not only severely underestimated the measured HONO, but also failed to represent the diurnal variation (Figure 6). The statistical information about the performance of the four models is given in Table 3. The mean HONO mixing ratio measured and calculated from CMAQ, RF, ANN, and RNDv1.0 was 0.94 ppbv, 0.09 ppbv, 0.95 ppbv, 0.88 ppbv, and 0.89 ppbv, respectively. Of the four models, RF showed the best performance, followed by RND. ANN has advantage of being able to calculate HONO more accurately than RND using more input variables, but it results in a lower data capture rate (41.5 %) compared to RND (97.7 %) or RF (85.3 %).

2.6. Model test

The RNDv1.0 and RF models were tested using June in 2018, April in 2019, and May ~ June and October ~ November in 2021 (Figure 3). Of the test dataset, the early winter (October ~ November) data is particularly valuable for demonstrating the applicability of the RNDv1.0 because they were produced in different weather conditions from those of the train dataset. Note that the RF performance was the best among the four models in train-validation process (Figure 5). Interestingly, the performance of RF was much worse than RNDv1.0 in test (Figure 7). The IOA and correlation coefficient of RF were extremely low (0.29 and -0.02, respectively), which are similar to or worse than those for CMAQv5.3.1 (Table 3).

The performance of RNDv1.0 was slightly lessened, but it well tracing the HONO mixing ratio. When the data in which at least one input parameters do not fall within the range of the train dataset is excluded from the test dataset, there is no significant difference in the performance of RNDv1.0 between the two that meet same atmospheric conditions or do not meet the criteria (Figure S5 and Table S2). And this test dataset includes severe haze pollution events when the daily average $PM_{2.5}$ concentration was raised up to 120 μg m^{-3}, and the HONO mixing ratio also increased to 4 ppbv or more in Seoul. Except for these extremes, RNDv1.0 traces well the variation of HONO mixing ratio. These test results, therefore, are convincing evidence for the applicability of the RNDv1.0 to the estimation of HONO levels in the urban atmosphere.

2.6. Bootstrap test and feature importance

A simple bootstrapping test was conducted for RNDv1.0 and RF to evaluate the relative importance of the input variable to HONO concentration. In this analysis, each variable was set to zero and MAE was calculated as an evaluation metrics (Kleinert et al., 2021). Of nine input variables, NO_2 was found to have the most significant influence on HONO concentration, followed by RH, temperature in RNDv1.0 (Table 5). The highest MAE of 0.59 ppbv can be considered as the maximum uncertainty of RNDv1.0 due to the input variable. The bootstrap test result is in good agreement with those of our previous study (Gil et al., 2021), where more variables such as aerosol surface area and mixing layer height were incorporated into the model, highlighting the crucial role of precursor gases and heterogeneous conversion in HONO formation.

In contrast, O_3 was the most important variable for RF. This is likely due to the distinct inverse relationship between O_3 and HONO in the diurnal patterns and O_3 variations over a wide range. In conjunction with the evaluation of test presented in the previous section, the results of feature importance for the two models demonstrates the ability of the deep learning model to simulate the HONO mixing ratio more adequately in polluted urban areas compared to the general machine learning model. Thus, it is reasonable to argue that the RNDv1.0 constructed using routinely measured criteria pollutants and meteorological parameters can sufficiently capture the HONO variability in the urban atmosphere.

3. Operation and application of RNDv1.0

The RNDv1.0 package is provided as an operational model, .h5 files that can be opened in Python. To run the RNDv1.0, the measurement data for nine input variables are required and need to be properly prepared as described in Section 2.2. A sample of preprocessed input dataset is provided as a .csv file (Dataset_for_model.csv). Once the input data is ready, open the RNDv1.0 with input data files using the code provided in the example (Figure S3). Then, RND v1.0 calculates and presents the HONO results as scaled values (x_{sca}), which will be finally converted to HONO mixing ratio (ppbv) by the two scale factors in Table 2 (Eq. 5):

$$\mathrm{HONO\ (ppbv)} = \mathrm{HONO_{sca}} \times \mathrm{F_1(HONO)} + \mathrm{F_2(HONO)}. \quad (5)$$

The HONO calculated by Eq. 5 can be applied to an urban photochemical cycle simulation. It is already known that the photolysis of HONO is a major source of OH radicals in the early morning when the OH level is low, and this OH affects daytime O_3 formation through photochemical reactions with VOCs and NO_x, which are primarily emitted during morning rush hour in urban areas. In addition, the OH produced from HONO promotes the photochemical oxidation of SO_2 and VOCs, leading to aerosol formation. However, the HONO formation mechanism is still poorly understood, hindering O_3 and fine aerosols as well as HONO from being correctly simulated in conventional photochemical models.

The 0-Dimension Atmospheric Modelling (F0AM) utilizing the MCM v3.3.1 chemical reaction mechanisms (Wolfe et al., 2016), can be used to simulate the diurnal variation of O_3 with the measurements of several reactive gases (NO, NO_2, CO, HCHO, VOCs, and HONO). Detailed information about F0AM can be found in (https://sites.google.com/site/wolfegm/models) and in previous works published elsewhere (Wolfe et al., 2016; Gil et al., 2020). When the F0AM model is run without HONO, it is not able to reproduce the concentration and diel cycle of the observed O_3 (Figure 8). In comparison, the model simulates the O_3 well within 2 ppbv when adding HONO, which is the product of RND v1.0. This is mainly due to the missing OH produced by HONO photolysis in the early morning. Its production rate is estimated to be 0.57 pptv s^{-1}, contributing approximately 2.28 pptv to OH budget during 06:00 ~ 11:00 (LST) (Gil et al., 2021). Given that OH is mainly produced from the photolysis of O_3 under high sun, the early morning source of OH will expedite the photochemical cycle involving NO_x and VOCs, promoting O_3 and secondary aerosol formation. Since the presence of HONO in the photochemical model allows for accurate estimation of OH radicals, the incorporation of RNDv1.0 into conventional models will improve their overall performance.

4. Summary and implications

In this study, we developed the RND model to calculate the mixing ratio of NO_Y in an urban atmosphere using a DNN along with measurement data. The target species of RNDv1.0 is HONO, and its mixing ratio is calculated using criteria pollutants including O_3, NO_2, CO, and SO_2, and meteorological variables including T, RH, WS, and WD, along with the SZA. These variables are routinely measured through monitoring networks. The RNDv1.0 was trained and validated using the HONO measurements obtained in Seoul by adopting a k-fold cross validation method and tested with other HONO datasets. The test results demonstrate that RNDv1.0 adequately captures the characteristic variation of HONO and confirms the efficacy of RND v1.0.

RNDv1.0 was constructed using the measurement made in a high NO_x environment. It is noteworthy that during the measurement period, the HONO mixing ratio was raised above 7 ppbv with the highest O_3 levels under stagnant conditions. If RNDv1.0 is applied to areas under significant influence of outflows, the model possibly overestimates or underestimate the level of HONO without detailed information such as nanoparticles. In the previous study, the formation of HONO was shown to be intimately related with surface areas of submicron particles (Gil et al., 2021). Nevertheless, the HONO concentration produced from RNDv1.0 with routine measurements provides the benefit of relatively inexpensive test for measurement quality control, location selection, and supports the data used for traditional chemistry model based on the current knowledge of the urban photochemical cycle. Therefore, it is reasonable to argue that RNDv1.0 can serve as a supplementary tool for conventional forecasting models. As attempts are currently being made to estimate ground HONO from satellite observations (Clarisse et al., 2011;Theys et al., 2020;Armante et al., 2021), RNDv1.0 will also be useful for validating satellite-derived HONO by supplementing measurement data.

5. Acknowledgements

This study was supported by the National Research Foundation of Republic of Korea (2020R1A2C3014592) and Korea Institute of Science and Technology (KIST2E31650-22-P019).

6. Code availability

The RND model codes (.h5 files) with preprocessed sample data can be downloaded from (Gil, 2021).

7. Author contributions

JG and ML designed the manuscript and developed the model code. JK, GL, and JA provided HONO measurements and CK provided CMAQ model data. All the authors contributed to the manuscript.

8. Competing interests

The authors declare that they have no conflict of interest.

Figures and Tables

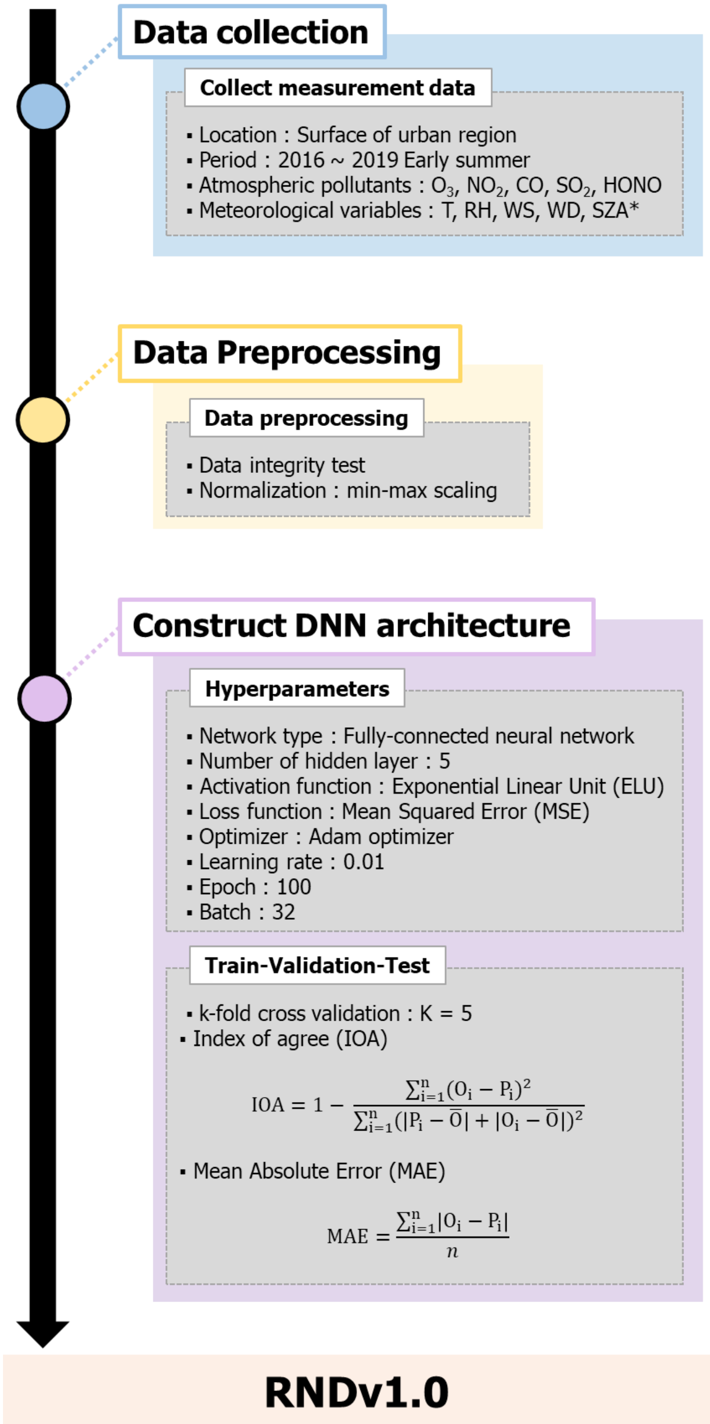

Figure 1. The main processes for configuring the RNDv1.0 (*: calculated values)

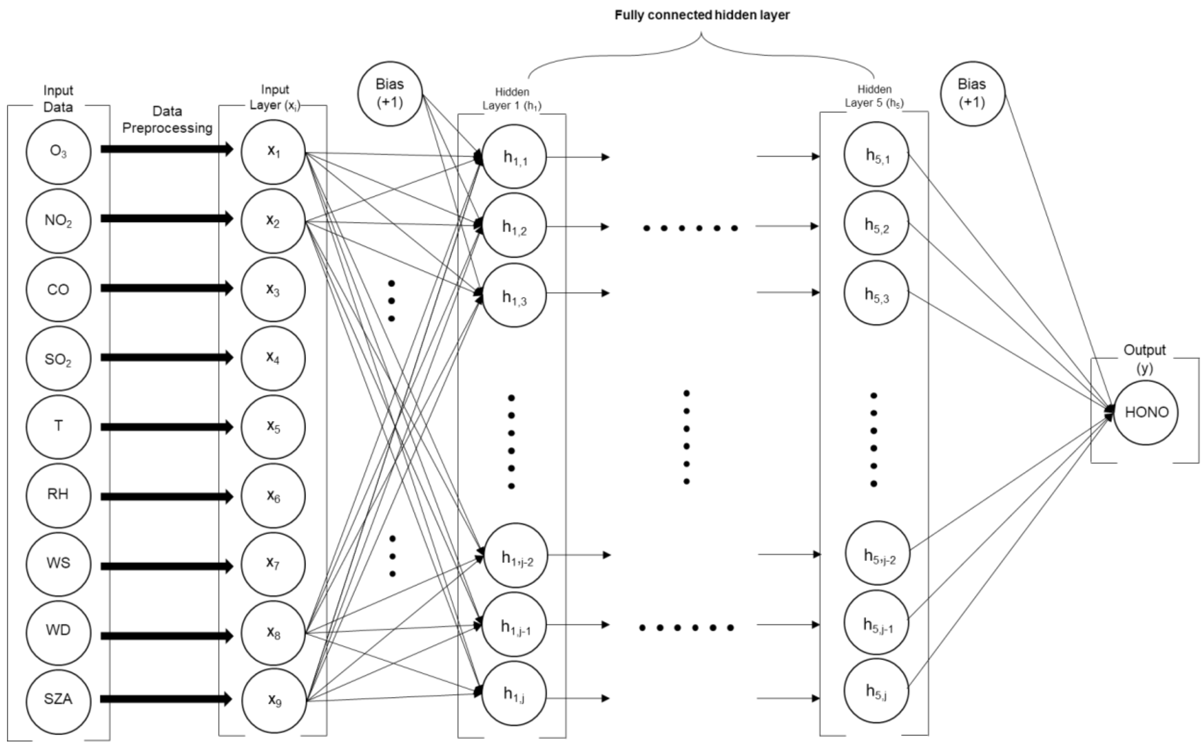

Figure 2. The structure of deep neural network built for RND v1.0.

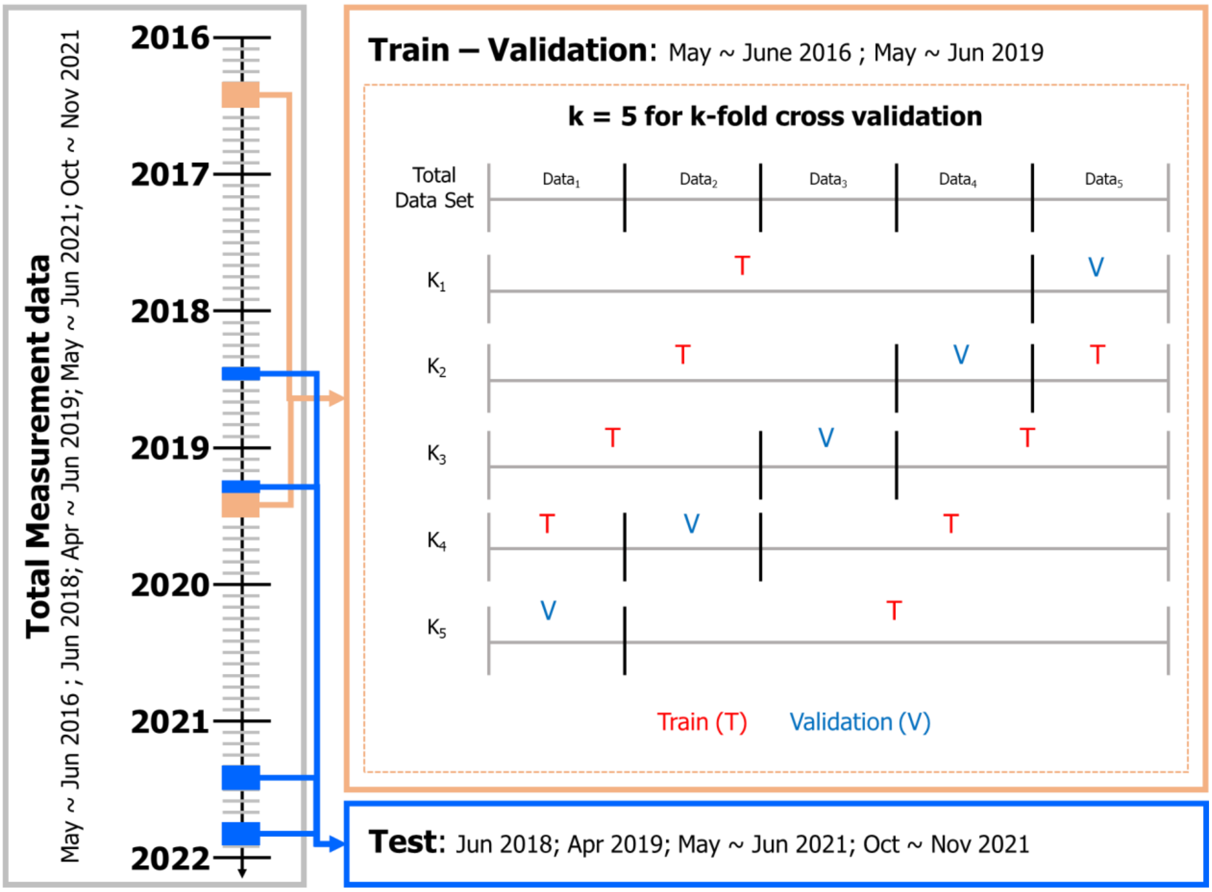

Figure 3. Design of training, validation, and test to build RNDv1.0 using measurement data. The k-fold cross validation was performed using randomly divided five subsets of training data set.

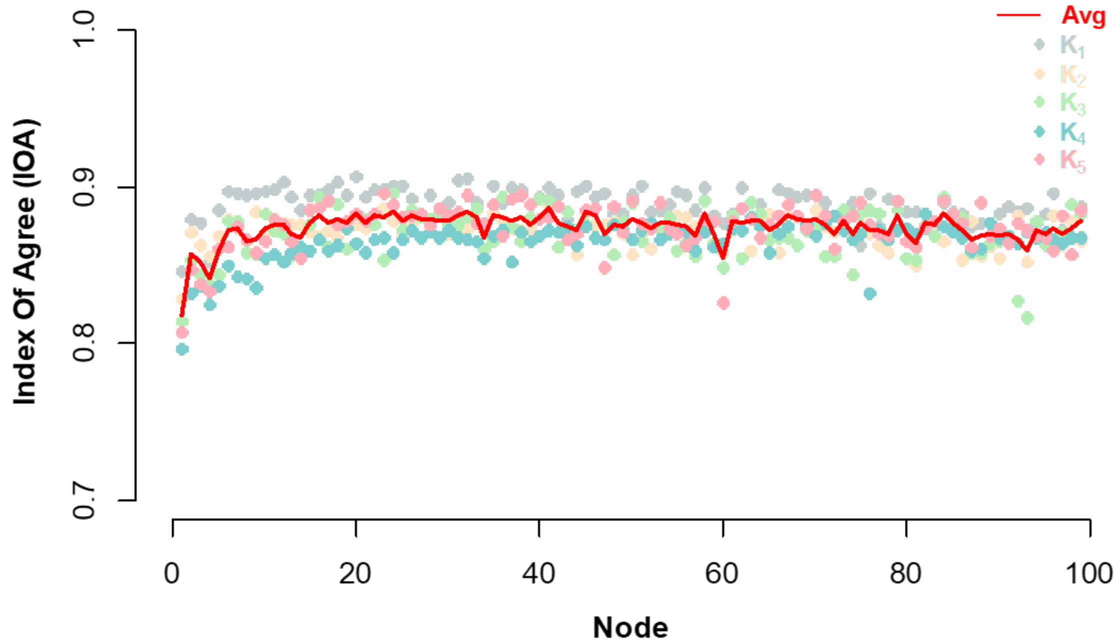

Figure 4. Index Of Agreement (IOA) for k-fold cross validation. Solid circle and red line represent IOA for each validation (k=5) and the average of 5 validation sets at each node number

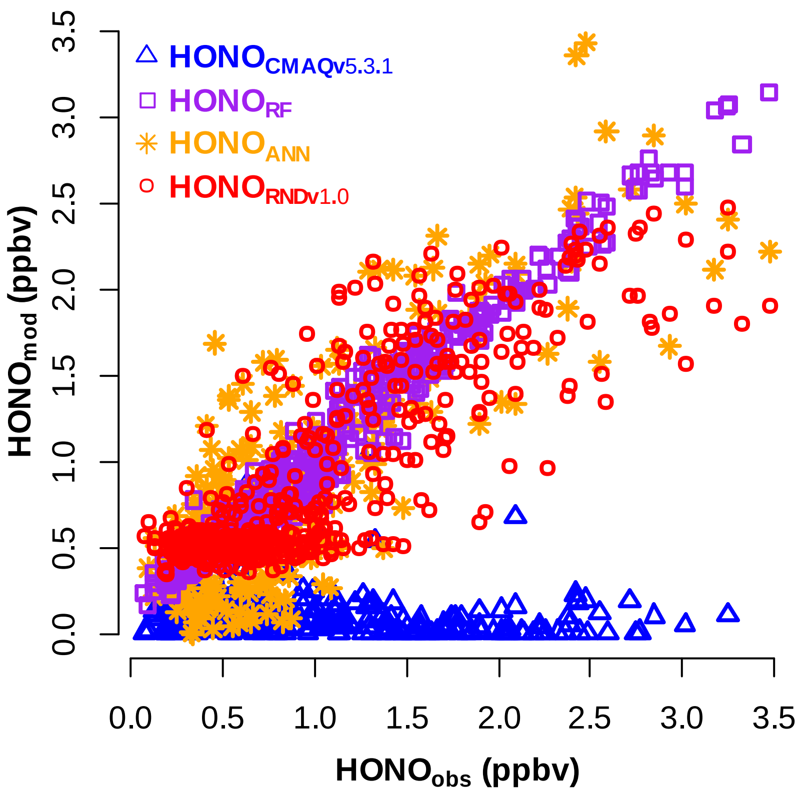

Figure 5. Comparison between measured HONO (HONO$_{obs}$) and calculated HONO (HONO$_{mod}$) using CMAQv5.3.1 (blue triangle), RF (purple square), ANN (orange star), and RNDv1.0 (red circle) during the KORUS-AQ campaign (may-June 2016)

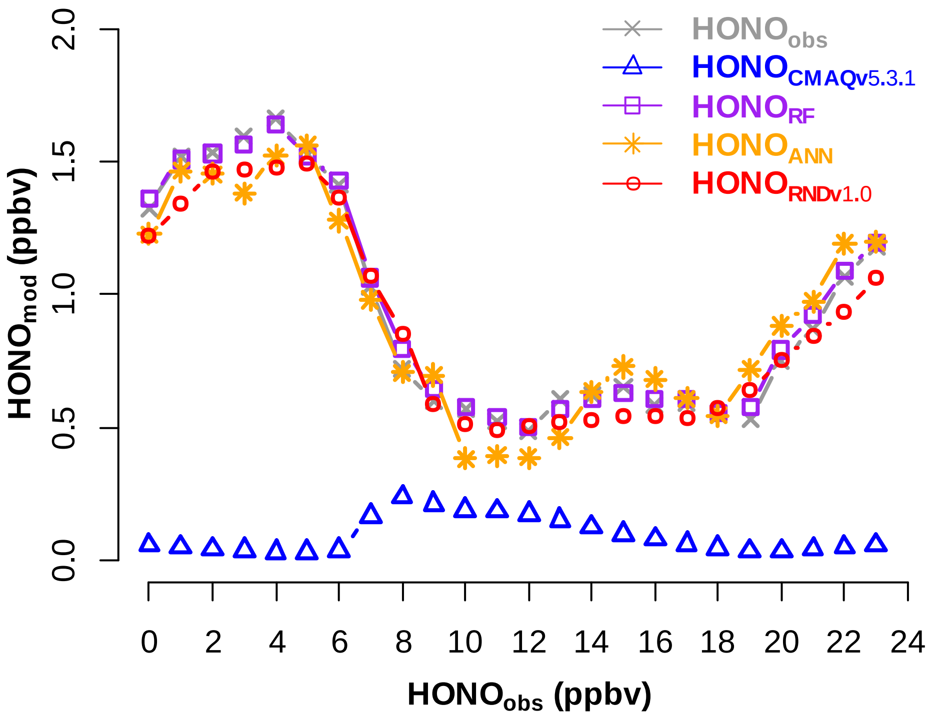

Figure 6. Average diurnal variation of measured HONO ($HONO_{obs}$) and calculated HONO ($HONO_{mod}$) using CMAQv5.3.1 (blue triangle), RF (purple square), ANN (orange star), and RNDv1.0 (red circle) during the KORUS-AQ campaign (may-June 2016)

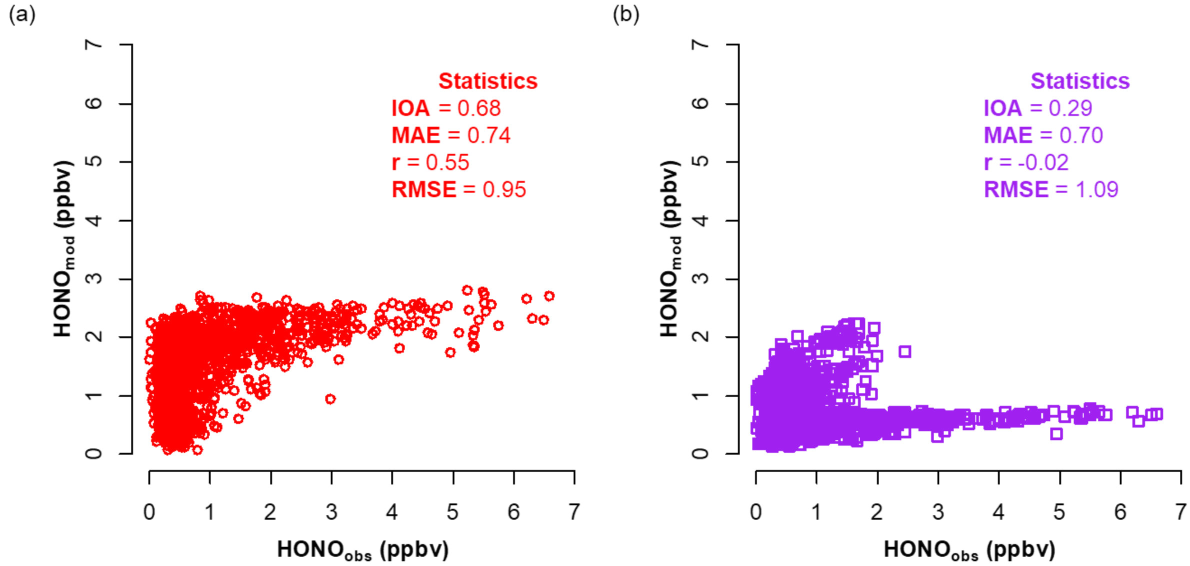

Figure 7. Relationship between measured HONO ($HONO_{obs}$) and modeled HONO ($HONO_{mod}$) using (a) RNDv1.0 and (b) a Random Forest model for the test dataset.

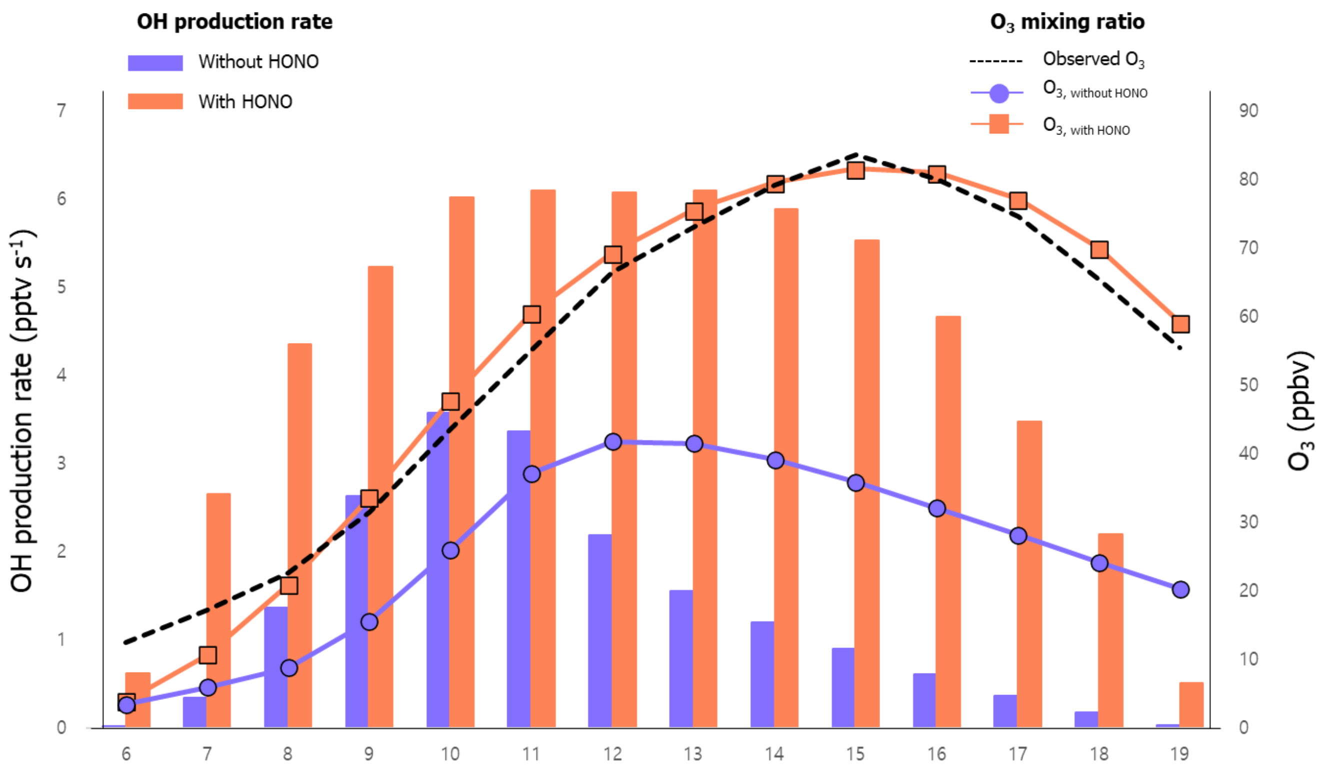

Figure 8. For June 2016, diurnal variations of O_3 (line) and OH production rate (bar) calculated from the F0AM photochemical model with (orange) and without (blue) HONO estimated from the RNDv1.0 model. The measured O_3 is compared with the calculated.

Table 1. Resources for constructing RND model.

	Version	Remark
Python	v3.8.3	
CUDA	v10.1	*If using GPU
CuDNN	v7.6.5	*If using GPU
Tensorflow	v2.3.0	*Python library*
Keras	v2.4.3	*Python library*
Pandas	v1.0.5	*Python library*
Numpy	v1.18.5	*Python library*

*GPU denotes graphic processing unit

Table 2. Input variables and their concentrations (10^{th} ~ 90^{th} percentile), coverage, and scale factors for RNDv1.0 model. Measurements were made in Seoul during May ~ June in 2016 and 2019.

	10^{th}~90^{th} percentile (unit)	Coverage (%)	Scale Factor1 (F_1)*	Scale Factor 2 (F_2)**
Input Variables				
O_3	12.1 ~ 90.4 (ppbv)	95.5	204.738	0.842
NO_2	11.0 ~ 48.6 (ppbv)	80.6	79.925	2.375
CO	252 ~ 743 (ppbv)	95.1	975.248	137.253
SO_2	1.9 ~ 6.4 (ppbv)	95.6	12.479	0.958
Solar Zenith Angle	22.7 ~ 118.4 (º)	100.0	112.317	14.195
Temperature	15.9 ~ 26.7 (°C)	99.4	24.240	8.610
Relative Humidity	29.2 ~ 79.1 (%)	99.4	88.545	10.555
Wind Speed	0.2 ~ 3.7 (m/s)	99.4	7.581	0.005
Wind Direction	45.4 ~ 287.5 (º)	99.4	359.565	0.235
Output Variables				
HONO	0.3 ~ 2.0 (ppbv)	81.1%	3.447	0.013

* Maximum – Minimum
** Minimum value

Table 3. The performance of chemical transport model (CMAQv5.3.1) and machine learning (ML) models including Random Forest (RF), Artificial Neural Network (ANN), and RNDv1.0 on measurement data from 2016 KORUS-AQ campaign that were used for training.

	CMAQv5.3.1	RF	ANN	RNDv1.0
IOA	0.44	0.99	0.86	0.9
r	-0.07	0.99	0.81	0.84
MAE	0.82	0.1	0.38	0.27
RMSE	1.06	0.12	0.41	0.37

Table 4. The result of bootstrap test of measurement data used to train the RF and RNDv1.0 model. The greater the MAE, the greater the influence of variable.

Variable	RF		RNDv1.0	
	MAE	Feature Importance	MAE	Feature Importance
-	0.10	-	0.28	-
O_3	0.57	1	0.29	8
NO_2	0.24	4	0.59	1
CO	0.19	7	0.37	5
SO_2	0.17	8	0.34	6
Solar zenith Angle (SZA)	0.25	2	0.41	4
Temperature (T)	0.21	5	0.52	2
Relative humidity (RH)	0.25	3	0.52	2
Wind speed (WS)	0.20	6	0.34	6
Wind direction (WD)	0.13	9	0.29	8

Reference

Akimoto, H., Nagashima, T., Li, J., Fu, J. S., Ji, D., Tan, J., and Wang, Z.: Comparison of surface ozone simulation among selected regional models in MICS-Asia III–effects of chemistry and vertical transport for the causes of difference, Atmospheric Chemistry and Physics, 19, 603-615, 2019.

Akimoto, H., and Tanimoto, H.: Review of Comprehensive Measurements of Speciated NOy and its Chemistry: Need for Quantifying the Role of Heterogeneous Processes of HNO3 and HONO, Aerosol and Air Quality Research, 21, 200395, 2021.

Anderson, D. C., Loughner, C. P., Diskin, G., Weinheimer, A., Canty, T. P., Salawitch, R. J., Worden, H. M., Fried, A., Mikoviny, T., and Wisthaler, A.: Measured and modeled CO and NOy in DISCOVER-AQ: An evaluation of emissions and chemistry over the eastern US, Atmospheric Environment, 96, 78-87, 2014.

Appel, K. W., Bash, J. O., Fahey, K. M., Foley, K. M., Gilliam, R. C., Hogrefe, C., Hutzell, W. T., Kang, D., Mathur, R., and Murphy, B. N.: The Community Multiscale Air Quality (CMAQ) model versions 5.3 and 5.3. 1: system updates and evaluation, Geoscientific Model Development, 14, 2867-2897, 2021.

Arcomano, T., Szunyogh, I., Wikner, A., Pathak, J., Hunt, B. R., and Ott, E.: A Hybrid Approach to Atmospheric Modeling that Combines Machine Learning with a Physics-Based Numerical Model, Journal of Advances in Modeling Earth Systems, e2021MS002712, 2021.

Armante, R., Perrin, A., Kwabia Tchana, F., and Manceron, L.: The ν4 bands at 11 μm: linelists for the Trans-and Cis-conformer forms of nitrous acid (HONO) in the 2019 version of the GEISA database, Molecular Physics, e1951860, 2021.

Arnell, N. W., Lowe, J. A., Challinor, A. J., and Osborn, T. J.: Global and regional impacts of climate change at different levels of global temperature increase, Climatic Change, 155, 377-391, 2019.

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, 2021.

Bao, F., Cheng, Y., Kuhn, U., Li, G., Wang, W., Kratz, A. M., Weber, J., Weber, B., Pöschl, U., and Su, H.: Key Role of Equilibrium HONO Concentration over Soil in Quantifying Soil–Atmosphere HONO Fluxes, Environmental science & technology, 2022.

Bengio, Y., and Grandvalet, Y.: No unbiased estimator of the variance of K-fold cross-validation, Citeseer, 2003.

Bloss, W. J., Kramer, L., Crilley, L. R., Vu, T., Harrison, R. M., Shi, Z., Lee, J. D., Squires, F. A., Whalley, L. K., and Slater, E.: Insights into air pollution chemistry and sulphate formation from nitrous acid (HONO) measurements during haze events in Beijing, Faraday Discussions, 226, 223-238, 2021.

Brown, S. S., An, H., Lee, M., Park, J.-H., Lee, S.-D., Fibiger, D. L., McDuffie, E. E., Dubé, W. P., Wagner, N. L., and Min, K.-E.: Cavity enhanced spectroscopy for measurement of nitrogen oxides in the Anthropocene: results from the Seoul tower during MAPS 2015, Faraday discussions, 200, 529-557, 2017.

Canty, T., Hembeck, L., Vinciguerra, T., Anderson, D., Goldberg, D., Carpenter, S., Allen, D., Loughner, C., Salawitch, R., and Dickerson, R.: Ozone and NO x chemistry in the eastern US: evaluation of CMAQ/CB05 with satellite (OMI) data, Atmospheric Chemistry and Physics, 15, 10965-10982, 2015.

Chen, G., Li, S., Knibbs, L. D., Hamm, N. A., Cao, W., Li, T., Guo, J., Ren, H., Abramson, M. J., and Guo, Y.: A machine learning method to estimate PM2. 5 concentrations across China with remote sensing, meteorological and land use information, Science of the Total Environment, 636, 52-60, 2018a.

Chen, Y., Wolke, R., Ran, L., Birmili, W., Spindler, G., Schröder, W., Su, H., Cheng, Y., Tegen, I., and Wiedensohler, A.: A parameterization of the heterogeneous hydrolysis of N2O5 for mass-based aerosol models: improvement of particulate nitrate prediction, Atmos. Chem. Phys, 18, 673-689, 2018b.

Cheng, P., Pour-Biazar, A., White, A. T., and McNider, R. T.: Improvement of summertime surface ozone prediction by assimilating Geostationary Operational Environmental Satellite cloud observations, Atmospheric Environment, 268, 118751, 2022.

Clarisse, L., R'Honi, Y., Coheur, P. F., Hurtmans, D., and Clerbaux, C.: Thermal infrared nadir observations of 24 atmospheric gases, Geophysical Research Letters, 38, 2011.

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.

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.

Ding, B., Qian, H., and Zhou, J.: Activation functions and their characteristics in deep neural networks, 2018 Chinese control and decision conference (CCDC), 2018, 1836-1841.

Ge, B., Xu, X., Ma, Z., Pan, X., Wang, Z., Lin, W., Ouyang, B., Xu, D., Lee, J., and Zheng, M.: Role of Ammonia on the Feedback Between AWC and Inorganic Aerosol Formation During Heavy Pollution in the North China Plain, Earth and Space Science, 6, 1675-1693, 2019.

Gen, M., Liang, Z., Zhang, R., Mabato, B. R. G., and Chan, C. K.: Particulate nitrate photolysis in the atmosphere, Environmental Science: Atmospheres, 2022.

Gil, J., Son, J., Kang, S., Park, J., Lee, M., Jeon, E., and Shim, M.: HONO measurement in Seoul during Summer 2018 and its Impact on Photochemistry, Journal of Korean Society for Atmospheric Environment, 36, 579-588, 10.5572/KOSAE.2020.36.5.579, 2020.

Gil, J.: RNDv1.0 and example, https://doi.org/10.5281/zenodo.5540180, in, Zenodo, 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.

Gil, J.: Formation pathways of HONO and its impact on O3 and fine aerosol: based on measurement and modelling study, Doctoral Thesis, 262, 2022.

Gu, R., Wang, W., Peng, X., Xia, M., Zhao, M., Zhang, Y., Liu, Y., Shen, H., Xue, L., and Wang, T.: Nitrous acid in the polluted coastal atmosphere of the South China Sea: Ship emissions, budgets, and impacts, Science of The Total Environment, 153692, 2022.

IPCC: Summary for policymakers. In:Climate Change 2014: Impacts, Adaption, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Field, C.B., V.R. Barros, D.J. Dokken, K.J. Mach, M.D. Mastrandrea, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L.White (eds.)], Cambridge, United Kingdom and New York, NY, USA, 1-32, 2014.

Jia, C., Tong, S., Zhang, W., Zhang, X., Li, W., Wang, Z., Wang, L., Liu, Z., Hu, B., and Zhao, P.: Pollution characteristics and potential sources of nitrous acid (HONO) in early autumn 2018 of Beijing, Science of The Total Environment, 735, 139317, 2020.

Jordan, C., Crawford, J. H., Beyersdorf, A. J., Eck, T. F., Halliday, H. S., Nault, B. A., Chang, L.-S., Park, J., Park, R., Lee, G., Kim, H., Ahn, J.-y., Cho, S., Shin, H. J., Lee, J. H., Jung, J., Kim, D.-S., Lee, M., Lee, T., Whitehill, A., Szykman, J., Schueneman, M. K., Campuzano-Jost, P., Jimenez, J. L., DiGangi, J. P., Diskin, G. S., Anderson, B. E., Moore, R. H., Ziemba, L. D., Fenn, M. A., Hair, J. W., Kuehn, R. E., Holz, R. E., Chen, G., Travis, K., Shook, M., Peterson, D. A., Lamb, K. D., and Schwarz, J. P.: Investigation of factors controlling PM2.5 variability across the South Korean Peninsula during KORUS-AQ, Elementa: Science of the Anthropocene, in review, 2020.

Joutsensaari, J., Ozon, M., Nieminen, T., Mikkonen, S., Lähivaara, T., Decesari, S., Facchini, M. C., Laaksonen, A., and Lehtinen, K. E.: Identification of new particle formation events with deep learning, Atmospheric Chemistry and Physics, 18, 9597-9615, 2018.

Kang, Y., Choi, H., Im, J., Park, S., Shin, M., Song, C.-K., and Kim, S.: Estimation of surface-level NO2 and O3 concentrations using TROPOMI data and machine learning over East Asia, Environmental Pollution, 288, 117711, 2021.

Kim, H., Gil, J., Lee, M., Jung, J., Whitehill, A., Szykman, J., Lee, G., Kim, D., Cho, S., Ahn, J., Hong, J., and Park, M.: Overview and characteristics of air quality in the Seoul Metropolitan Area during the KORUS-AQ campaign, Elementa: Science of the Anthropocene, in review, 2020.

Kleffmann, J., Lörzer, J., Wiesen, P., Kern, C., Trick, S., Volkamer, R., Rodenas, M., and Wirtz, K.: Intercomparison of the DOAS and LOPAP techniques for the detection of nitrous acid (HONO), Atmospheric Environment, 40, 3640-3652, 2006.

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.

Krishnamurthy, R., Newsom, R. K., Berg, L. K., Xiao, H., Ma, P.-L., and Turner, D. D.: On the estimation of boundary layer heights: a machine learning approach, Atmospheric Measurement Techniques, 14, 4403-4424, 2021.

Lee, B. H., Wood, E. C., Zahniser, M. S., McManus, J. B., Nelson, D. D., Herndon, S. C., Santoni, G., Wofsy, S. C., and Munger, J. W.: Simultaneous measurements of atmospheric HONO and NO 2 via absorption spectroscopy using tunable mid-infrared continuous-wave quantum cascade lasers, Applied Physics B, 102, 417-423, 2011.

Levy, M., Zhang, R., Zheng, J., Zhang, A. L., Xu, W., Gomez-Hernandez, M., Wang, Y., and Olaguer, E.: Measurements of nitrous acid (HONO) using ion drift-chemical ionization mass spectrometry during the 2009 SHARP field campaign, Atmospheric Environment, 94, 231-240, 2014.

Li, S., Song, W., Zhan, H., Zhang, Y., Zhang, X., Li, W., Tong, S., Pei, C., Wang, Y., and Chen, Y.: Contribution of Vehicle Emission and NO2 Surface Conversion to Nitrous Acid (HONO) in Urban Environments: Implications from Tests in a Tunnel, Environmental Science & Technology, 55, 15616-15624, 2021a.

Li, Y., Wang, X., Wu, Z., Li, L., Wang, C., Li, H., Zhang, X., Zhang, Y., Li, J., and Gao, R.: Atmospheric nitrous acid (HONO) in an alternate process of haze pollution and ozone pollution in urban Beijing in summertime: Variations, sources and contribution to atmospheric photochemistry, Atmospheric Research, 260, 105689, 2021b.

Li, Z., Xie, P., Hu, R., Wang, D., Jin, H., Chen, H., Lin, C., and Liu, W.: Observations of N2O5 and NO3 at a suburban environment in Yangtze river delta in China: Estimating heterogeneous N2O5 uptake coefficients, Journal of Environmental Sciences, 2020.

Liebmann, J., Karu, E., Sobanski, N., Schuladen, J., Ehn, M., Schallhart, S., Quéléver, L., Hellen, H., Hakola, H., and Hoffmann, T.: Direct measurement of NO3 radical reactivity in a boreal forest, Atmospheric Chemistry and Physics, 2018.

Liu, Y., Lu, K., Li, X., Dong, H., Tan, Z., Wang, H., Zou, Q., Wu, Y., Zeng, L., and Hu, M.: A comprehensive model test of the HONO sources constrained to field measurements at rural North China Plain, Environmental science & technology, 53, 3517-3525, 2019.

Mallet, V., and Sportisse, B.: Uncertainty in a chemistry-transport model due to physical parameterizations and numerical approximations: An ensemble approach applied to ozone modeling, Journal of Geophysical Research: Atmospheres, 111, 2006.

Meng, F., Qin, M., Fang, W., Duan, J., Tang, K., Zhang, H., Shao, D., Liao, Z., Feng, Y., and Huang, Y.: Measurement of HONO flux using the aerodynamic gradient method over an agricultural field in the Huaihe River Basin, China, Journal of Environmental Sciences, 2022.

Monks, P. S., Archibald, A., Colette, A., Cooper, O., Coyle, M., Derwent, R., Fowler, D., Granier, C., Law, K. S., and Mills, G.: Tropospheric ozone and its precursors from the urban to the global scale from air quality to short-lived climate forcer, Atmospheric Chemistry and Physics, 15, 8889-8973, 2015.

Myhre, G., Aas, W., Cherian, R., Collins, W., Faluvegi, G., Flanner, M., Forster, P., Hodnebrog, Ø., Klimont, Z., and Lund, M. T.: Multi-model simulations of aerosol and ozone radiative forcing due to anthropogenic emission changes during the period 1990–2015, Atmospheric Chemistry and Physics, 17, 2709-2720, 2017.

Peterson, D. A., Hyer, E. J., Han, S.-O., Crawford, J. H., Park, R. J., Holz, R., Kuehn, R. E., Eloranta, E., Knote, C., and Jordan, C. E.: Meteorology influencing springtime air quality, pollution transport, and visibility in Korea, Elem Sci Anth, 7, 2019.

Pinto, J., Dibb, J., Lee, B., Rappenglück, B., Wood, E., Levy, M., Zhang, R. Y., Lefer, B., Ren, X. R., and Stutz, J.: Intercomparison of field measurements of nitrous acid (HONO) during the SHARP campaign, Journal of Geophysical Research: Atmospheres, 119, 5583-5601, 2014.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., and Carvalhais, N.: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195-204, 2019.

Roberts, J. M., Veres, P., Warneke, C., Neuman, J., Washenfelder, R., Brown, S., Baasandorj, M., Burkholder, J., Burling, I., and Johnson, T. J.: Measurement of HONO, HNCO, and other inorganic acids by negative-ion proton-transfer chemical-ionization mass spectrometry (NI-PT-CIMS): Application to biomass burning emissions, Atmospheric Measurement Techniques, 3, 981, 2010.

Schultz, M., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L., Mozaffari, A., and Stadtler, S.: Can deep learning beat numerical weather prediction?, Philosophical Transactions of the Royal Society A, 379, 20200097, 2021.

Shahriar, S. A., Kayes, I., Hasan, K., Salam, M. A., and Chowdhury, S.: Applicability of machine learning in modeling of atmospheric particle pollution in Bangladesh, Air Quality, Atmosphere & Health, 13, 1247-1256, 2020.

Shareef, M. M., Husain, T., and Alharbi, B.: Studying the Effect of Different Gas-Phase Chemical Kinetic Mechanisms on the Formation of Oxidants, Nitrogen Compounds and Ozone in Arid Regions, Journal of Environmental Protection, 10, 1006-1031, 2019.

Shindell, D. T., Lamarque, J.-F., Schulz, M., Flanner, M., Jiao, C., Chin, M., Young, P., Lee, Y. H., Rotstayn, L., and Mahowald, N.: Radiative forcing in the ACCMIP historical and future climate simulations, Atmospheric Chemistry and Physics, 13, 2939-2974, 2013.

Stadtler, S., Simpson, D., Schröder, S., Taraborrelli, D., Bott, A., and Schultz, M.: Ozone impacts of gas–aerosol uptake in global chemistry transport models, Atmospheric chemistry and physics, 18, 3147-3171, 2018.

Stevenson, D., Young, P., Naik, V., Lamarque, J.-F., Shindell, D. T., Voulgarakis, A., Skeie, R. B., Dalsoren, S. B., Myhre, G., and Berntsen, T. K.: Tropospheric ozone changes, radiative forcing and attribution to emissions in the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), Atmospheric Chemistry and Physics, 13, 3063-3085, 2013.

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.

Sun, Y., Wang, L., Wang, Y., Quan, L., and Zirui, L.: In situ measurements of SO2, NOx, NOy, and O3 in Beijing, China during August 2008, Science of the Total Environment, 409, 933-940, 2011.

Theys, N., Volkamer, R., Müller, J.-F., Zarzana, K. J., Kille, N., Clarisse, L., De Smedt, I., Lerot, C., Finkenzeller, H., and Hendrick, F.: Global nitrous acid emissions and levels of regional oxidants enhanced by wildfires, Nature geoscience, 13, 681-686, 2020.

Tie, X., Geng, F., Guenther, A., Cao, J., Greenberg, J., Zhang, R., Apel, E., Li, G., Weinheimer, A., and Chen, J.: Megacity impacts on regional ozone formation: observations and WRF-Chem modeling for the MIRAGE-Shanghai field campaign, Atmospheric Chemistry and Physics, 13, 5655-5669, 2013.

VandenBoer, T., Markovic, M., Sanders, J., Ren, X., Pusede, S., Browne, E., Cohen, R., Zhang, L., Thomas, J., and Brune, W. H.: Evidence for a nitrous acid (HONO) reservoir at the ground surface in Bakersfield, CA, during CalNex 2010, Journal of Geophysical Research: Atmospheres, 119, 9093-9106, 2014.

Varotsos, K., Giannakopoulos, C., and Tombrou, M.: Assessment of the Impacts of climate change on european ozone levels, Water, Air, & Soil Pollution, 224, 1-13, 2013.

Wang, H., Chen, X., Lu, K., Hu, R., Li, Z., Wang, H., Ma, X., Yang, X., Chen, S., and Dong, H.: NO3 and N2O5 chemistry at a suburban site during the EXPLORE-YRD campaign in 2018, Atmospheric Environment, 224, 117180, 2020.

Wang, T., Qin, Z., and Zhu, M.: An ELU network with total variation for image denoising, International Conference on Neural Information Processing, 2017a, 227-237.

Wang, X., Wang, H., Xue, L., Wang, T., Wang, L., Gu, R., Wang, W., Tham, Y. J., Wang, Z., and Yang, L.: Observations of N2O5 and ClNO2 at a polluted urban surface site in North China: High N2O5 uptake coefficients and low ClNO2 product yields, Atmospheric environment, 156, 125-134, 2017b.

Wolfe, G. M., Marvin, M. R., Roberts, S. J., Travis, K. R., and Liao, J.: The framework for 0-D atmospheric modeling (F0AM) v3. 1, Geoscientific Model Development, 9, 3309, 2016.

Xu, Z., Liu, Y., Nie, W., Sun, P., Chi, X., and Ding, A.: Evaluating the measurement interference of wet rotating-denuder–ion chromatography in measuring atmospheric HONO in a highly polluted area, Atmospheric Measurement Techniques, 12, 6737-6748, 2019.

Xue, C., Ye, C., Ma, Z., Liu, P., Zhang, Y., Zhang, C., Tang, K., Zhang, W., Zhao, X., and Wang, Y.: Development of stripping coil-ion chromatograph method and intercomparison with CEAS and LOPAP to measure atmospheric HONO, Science of The Total Environment, 646, 187-195, 2019.

Ye, C., Zhou, X., Pu, D., Stutz, J., Festa, J., Spolaor, M., Tsai, C., Cantrell, C., Mauldin, R. L., and Campos, T.: Rapid cycling of reactive nitrogen in the marine boundary layer, Nature, 532, 489-491, 2016.