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]

[1] *Department of Earth and Environmental Science, 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*

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 wad developed. The 'Reactive Nitrogen species simulation using Deep neural network' (RND) has been implemented in Python. It was trained, validated, and tested with HONO measurement data obtained in Seoul during the warm months from 2016 to 2019.

A k-fold cross validation and test results confirmed the performance of RND v1.0 with an Index Of Agreement (IOA) of 0.79 ~ 0.89 and a Mean Absolute Error (MAE) of 0.21 ~ 0.31 ppbv. The RNDV1.0 adequately represents the main characteristics of HONO and thus, RND v1.0 is proposed as a supplementary model for calculating the HONO mixing ratio in a high-NO_x 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 and intensity of surface O_3 peaks 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 $\leq$ 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 NOx 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)), and 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, heterogeneous reaction mechanism at surface is major concern in recently 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), 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

92 such as Convolution NN (CNN) or Long-Short Term Memory (LSTM) because it works well
93 for discrete spatiotemporal data. In general, the performance of DNN is similar to or better than
94 other ML methods for small number of data as well as large data set (Baek and Jung, 2021;Dang
95 et al., 2021;Sumathi and Pugalendhi, 2021).

96 When the DNN method is applied to atmospheric chemical constituents, it requires
97 large amount of data for training and thus, the size of measurement data becomes a limiting
98 factor for trace species such as HONO, which are not routinely measured such as O_3 or $PM_{2.5}$.
99 In this regard, the daily average HONO mixing ratio was attempted to be estimated using
100 ensemble ML models with satellite measurements (Cui and Wang, 2021). In comparison, the
101 hourly HONO mixing ratio was calculated using a simple NN architecture with measured
102 variables, which were thought to be closely linked with HONO formation (Gil et al., 2021). The
103 accuracy of the hourly HONO estimated from input variables such as aerosol surface areas and
104 mixed layer height was better than the daily HONO estimate.

105 In this study, we aimed to construct a user-friendly 'Reactive Nitrogen species
106 simulation using DNN' (RND) model and estimate HONO mixing ratio using routinely
107 measured atmospheric variables in a highly polluted urban area. Finally, the model results will
108 be incorporated into operational photochemical models for air quality forecasting and improve
109 their performance. Since this is the first attempt to calculate HONO mixing ratios using a first
110 version of RND model (RNDv1.0), we describe the entire modeling process and evaluate the
111 model results by comparing them with the measurements.

112

113 **2. Model description**

114

115 The development of RNDv1.0 model follows the systematic steps similar to a general
116 machine learning model construction workflow, including collecting data, preprocessing data,
117 building the DNN, training and validating the model, and testing the performance of the model
118 (Figure 1). The RNDv1.0 was written in Python and necessary libraries to build and operate
119 RNDv1.0 are listed in Table 1. The dataset used to train-test-validation can be downloaded from
120 Gil et al., 2021.

121

2.1. Collection of measurement data for model construction

123

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 using a Quantum Cascade – Tunable Infrared Laser Differential Absorption Spectrometer (QC-TILDAS) system in Seoul during May–June 2016, June 2018, and April-June 2019 (Lee et al., 2011;Gil et al., 2021). 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). 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 in 2018 ($37.59°N$, $127.03°E$), and at the site near the campus in 2019 ($37.59°N$, $127.08°E$) (NIER, 2020) (Figure S1). Of the three sites, the Korea University campus and Olympic Park have served as measurement sites representing the air quality of Seoul. In fact, it has been known that O_3 and $PM_{2.5}$ levels are strongly influenced by the synoptic circulation throughout the Korean peninsula (Peterson et al., 2019;Jordan et al., 2020). In addition, 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 10th and 90th percentile mixing ratios of HONO, NO_2, and O_3 are 0.3 ppbv and 1.9 ppbv, 10.7 ppbv and 48.2 ppbv, and 12.0 ppbv and 80.9 ppbv, respectively for the entire experiment periods.

147

2.2. Data preprocessing

149

150   In the next step, the observation data set was prepared for RNDv1.0 model construction.

151 As input variables, hourly measurements of chemical and meteorological parameters are used,

152 including the mixing ratios of O_3, NO_2, CO, and SO_2, along with temperature (T), relative

153 humidity (RH), wind speed (WS), wind direction (WD), and solar zenith angle (SZA) to

154 estimate the target species, HONO, as the output. Wind direction in degrees were converted to

155 a cosine value for continuity. As a last step in data processing, missing values were filtered out

156 from the input dataset. Finally, 50.7 % of all available measurement data (1636) were used to

157 construct the RNDv1.0 in this study.

158   Since the measurements of these nine variables vary over a wide range in different units,

159 they were normalized to avoid bias during the calculations. Among the widely used

160 normalization methods, '*min-max scaling*' method was adopted and input variables were

161 normalized against the minimum and maximum values in this study (Eq. 1):

162

163 $$x_{sca} = \frac{x_{raw} - F_2(X)}{F_1(X)}, \tag{1}$$

164

165   where x_{raw} is raw data of input variable (X), x_{sca} is scaled data of X, F_1 and F_2 are scale

166 factors of X, and are given for each input variable used in Table 2.

167

168 **2.3. Neural network architecture and hyperparameters**

169

170   At this stage, the network is built to calculate HONO using those input variables. The

171 RNDv1.0 is composed of five hidden layers (Figure 2), which employed an exponential linear

172 unit (ELU) as an activation function (Eq. 2).

173

174 $$ELU : \phi(x) = \begin{cases} e^x - 1 \ (x < 0) \\ x \ (x \geq 0) \end{cases}. \tag{2}$$

175

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. Train, validation, and test

The RNDv1.0 model was trained, validated, and tested with HONO measurements obtained during May ~ June in 2016 and 2019, in June 2018, and in April 2019, respectively (Figure 3). The number of data used for train, validation, and test were 1122, 381, and 133, respectively.

With the hyperparameters specified in previous section, the performance of the model was firstly validated using the k-fold cross-validation method, which is especially useful when the size of dataset is small (Bengio and Grandvalet, 2003). In the k-fold cross-validation 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) and Mean Absolute Error (MAE) expressed by the following equation (Eq. 3, Eq. 4):

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

$$MAE = \frac{\sum_{i=1}^{n}|O_i - P_i|}{n}, \tag{4}$$

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. The overall accuracy of

As IOA and MAE vary according to the number of nodes, they were 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

204   the averaged IOA and MAE were 0.89 ± 0.01 (mean ± standard deviation) and 0.31 ± 0.02 ppbv,

205   respectively (Figure 4). The high level of IOA and low MAE demonstrates that the performance

206   of RNDv1.0 model is adequate, and it is capable of simulating the ambient HONO mixing ratio

207   using the routinely measured chemical and meteorological parameters. In particular, MAE was

208   commensurate with the detection limit of HONO measurement.

209   After the network validation, HONO mixing ratio was calculated for May ~ June in 2016

210   and 2019, and the model results were compared with the measured values (Figure 5). The

211   average mixing ratios of measured and calculated HONO was 0.94 ppbv and 0.89 ppbv in 2016,

212   and 1.02 ppbv and 0.96 ppbv in 2019, respectively. The MAE and IOA of the measurement and

213   calculation were 0.27 ppbv and 0.90 in 2016, and 0.29 ppbv and 0.91 in 2019, respectively,

214   demonstrating the ability of the RNDv1.0 to simulate ambient HONO levels. In both cases,

215   however, the model slightly underestimated the highest and lowest HONO mixing ratios, which

216   is mainly due to the limited number of data used for training, but also related to the intrinsic

217   nature of DNN. The model calculation well captured the diurnal variation of ambient HONO

218   with a slight underestimation (Figure 6). In addition, the correlation between $HONO_{mod}$ and

219    $HONO_{obs}$ was better in 2019 (MAE = 0.06 ppbv) than in 2016 (MAE = 0.08 ppbv). Since the

220   MAE of the two cases was far below the detection limit of HONO measurements (~ 0.1 ppbv),

221   the RNDv1.0 is considered adequate to simulate HONO in urban areas.

222   Finally, the RND model was validated and tested against the measurement data obtained in

223   June 2018 and April 2019. The calculated HONO mixing ratios are compared with those

224   measured in Figure 7, and their MAE and IOA are listed in Table 3. The two sets of model

225   performance test showed that the model reasonably traced what was observed. As the validation

226   result of RND, the MAE and IOA of the calculated and measured in June 2018 are comparable

227   to those of 2016~2019 result. However, the MAE and IOA of the April 2019 measurements

228   were relatively poor compared to the validation results. Especially, the MAE of the April 2019

229   is about twice as high as those of validation.

230   In these two test periods, HONO levels were lower than those observed on validation days

231   (Figure 5), and the model tended to overestimate high HONO concentrations. The large

232   discrepancy in April 2019 is probably due to seasonality: the difference in meteorological and

233   chemical regime of the atmosphere. For example, the monthly average temperature, relative

234   humidity, and NO_2 mixing ratio of Seoul in 2019 were 12.1 ℃, 50.9 %, and 29 ppbv in April

235 2019 and 22.5 ℃, 60.6 %, and 21 ppbv in June 2019 (https://cleanair.seoul.go.kr;

236 https://weather.go.kr). Note that the RNDv1.0 model was trained with the 9 variables measured

237 in early summer (Table 2). Therefore, the more measurement data spanning a full year for

238 training, the more accurate the model estimates will be.

239

240 **2.5. Influence of input variables to HONO concentration**

241

242 Additionally, a simple bootstrapping test was conducted by setting each variable to zero

243 with keeping other variables (Kleinert et al., 2021). Then, the importance of each input variable

244 to HONO concentration was evaluated using MAE and root mean square error (RMSE). Of

245 nine input variables, NO_2 was found to have the most significant influence on HONO

246 concentration, followed by RH, temperature, and solar zenith angle (Table S2). The result of

247 bootstrap test is in good agreement with those from our previous study (Gil et al., 2021), where

248 more detailed information such as aerosol surface area and mixing layer height were

249 incorporated into the model and highlighted the role of precursor gases and heterogeneous

250 conversion in HONO formation. Therefore, these results demonstrate that the RND model

251 constructed using routinely observed variables, reasonably traced the level of HONO in urban

252 atmosphere.

253

254 **3. Operation and application of RNDv1.0**

255

256 The RNDv1.0 package is provided as an operational model, .h5 files that can be opened in

257 Python. To run the RNDv1.0, the measurement data for nine input variables are required and

258 need to be properly prepared as described in Section 2.2. A sample of preprocessed input dataset

259 is provided as a .csv file (Dataset_for_model.csv). Once the input data is ready, open the RND

260 model with input data files using the code provided in the example (Figure S3). Then, RND

261 v1.0 calculates and presents the HONO results as scaled values (x_{sca}), which will be finally

262 converted to HONO mixing ratio (ppbv) by the two scale factors in Table 2 (Eq. 5):

263

264 $$HONO \ (ppbv) = HONO_{sca} \times F_1(HONO) + F_2(HONO). \qquad (5)$$

265

266   The result of the RNDv1.0, HONO, can be applied to an urban photochemical cycle
267 simulation. It is already known that the photolysis of HONO is a major source of OH radicals
268 in the early morning when the OH level is low, and this OH affects daytime O_3 formation
269 through photochemical reactions with VOCs and NO_x, which are primarily emitted during
270 morning rush hour in urban areas. Therefore, the OH produced from HONO expedites
271 photochemical reactions, promoting O_3 formation. However, the HONO formation mechanism
272 is still poorly understood, and concentrations are not correctly simulated in conventional
273 photochemical models; therefore, the absence of HONO causes great uncertainty in O_3
274 prediction (Figure 8).

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

291

292   **4. Summary and implications**

293

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 trace gases 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 measured using the same instrument. The validation and test results demonstrate that RND adequately captures the characteristic variation of HONO and confirms the efficacy of RND v1.0.

RNDv1.0 was constructed using measurements made in a high NO_x environment during early summer (May–June). It is noteworthy that in this period, the HONO mixing ratio was raised above 3 ppbv with the highest O_3 levels under stagnant conditions. If RND 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 RND can serve as a supplementary tool for conventional photochemical models.

5. Acknowledgements

This study was financially supported by the National Research Foundation of Republic of Korea (2020R1A2C3014592).

6. Code availability

322

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

325

326 **7. Author contributions**

327

328 JG and ML designed the manuscript and developed the model code. JK, GL, and JA
329 provided the measurement data and validated the model. All the authors contributed to the
330 manuscript.

331

332 **8. Competing interests**

333

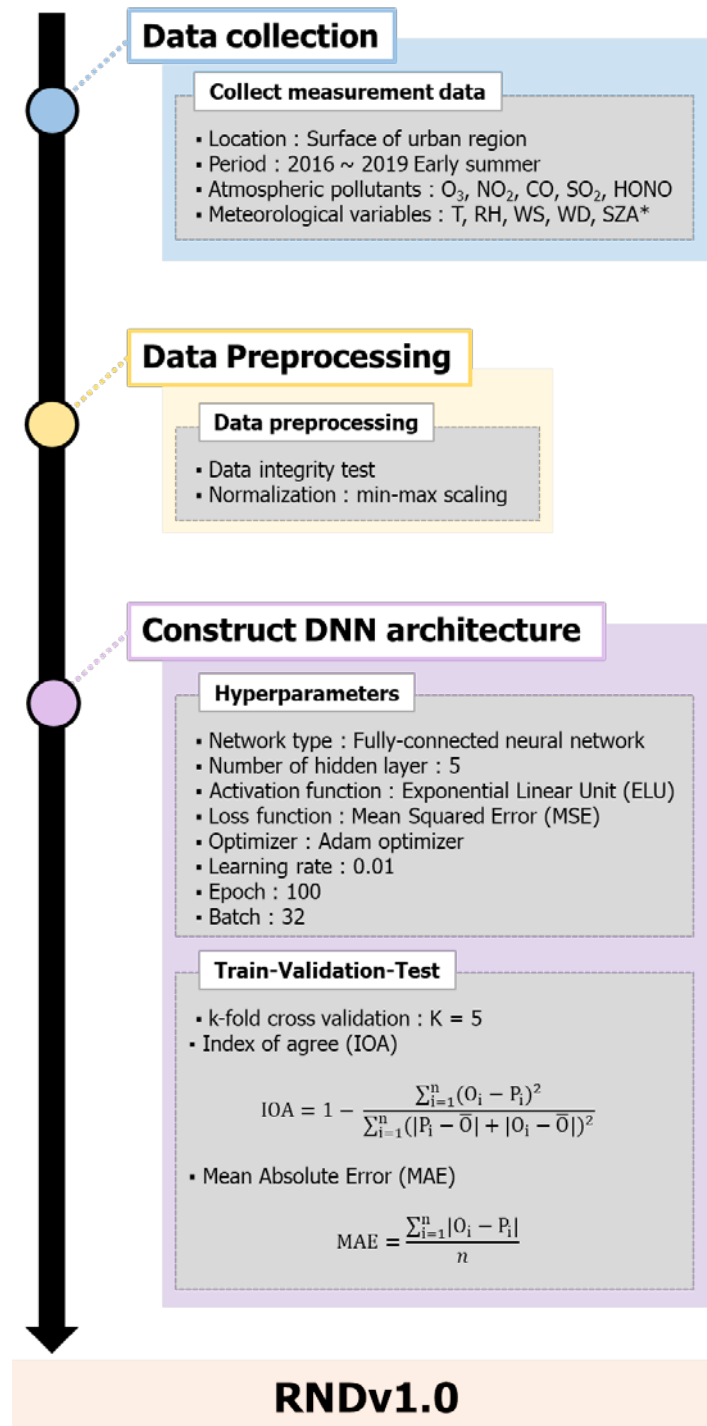
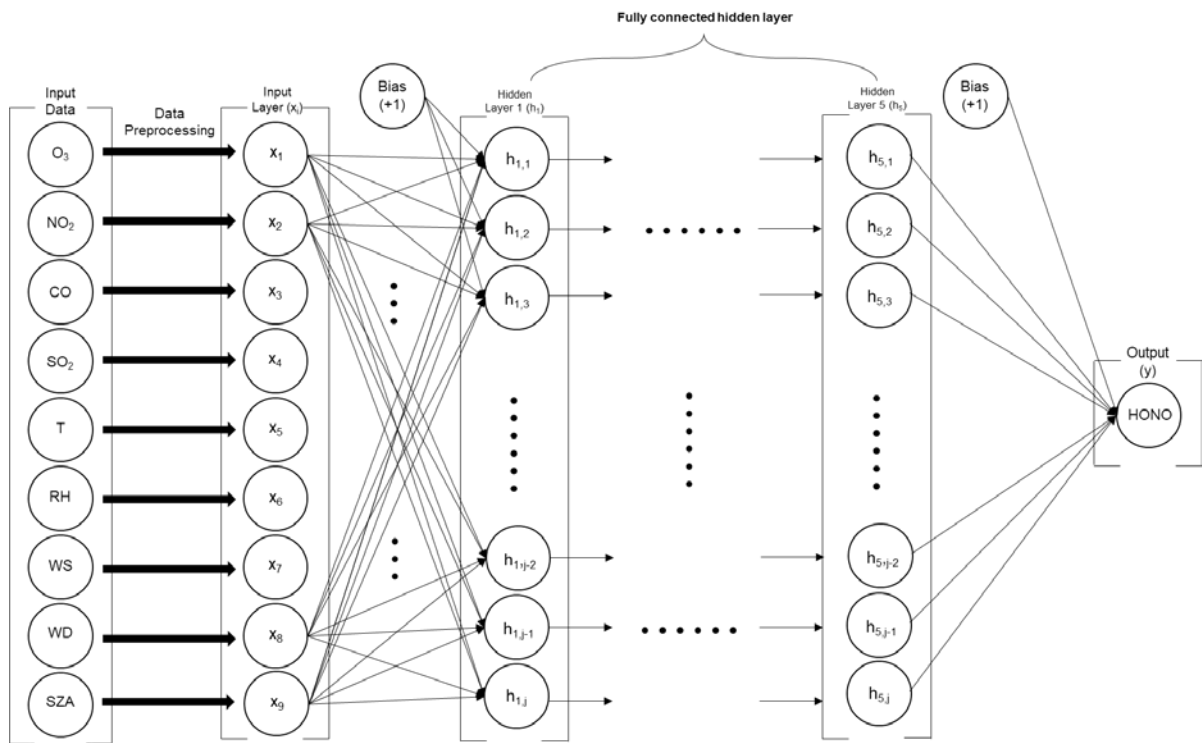
334 The authors declare that they have no conflict of interest.

335

336

337 **Figures and Tables**

338

339

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

341

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

343

344

345

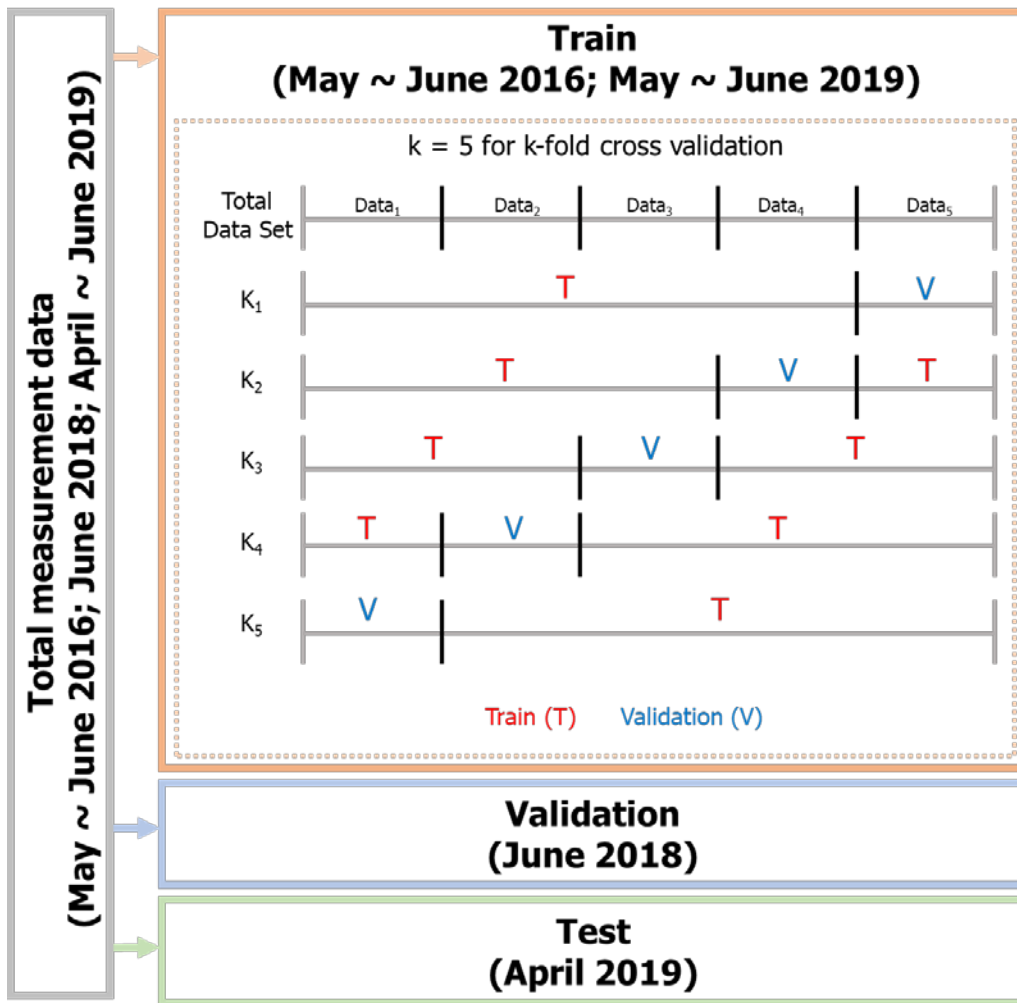
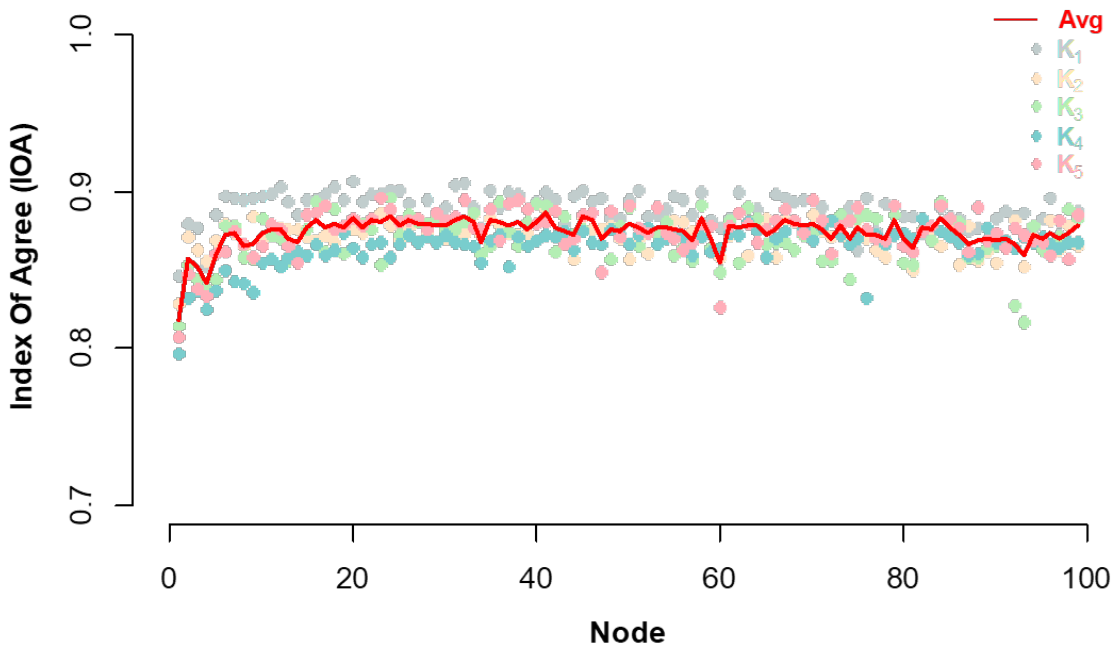



Figure 3. Design of training, validation, and test to build RNDv1.0 using measurement data. **The k-fold cross validation were performed using randomly divided five subsets of training data set.**

349

350

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

353

354

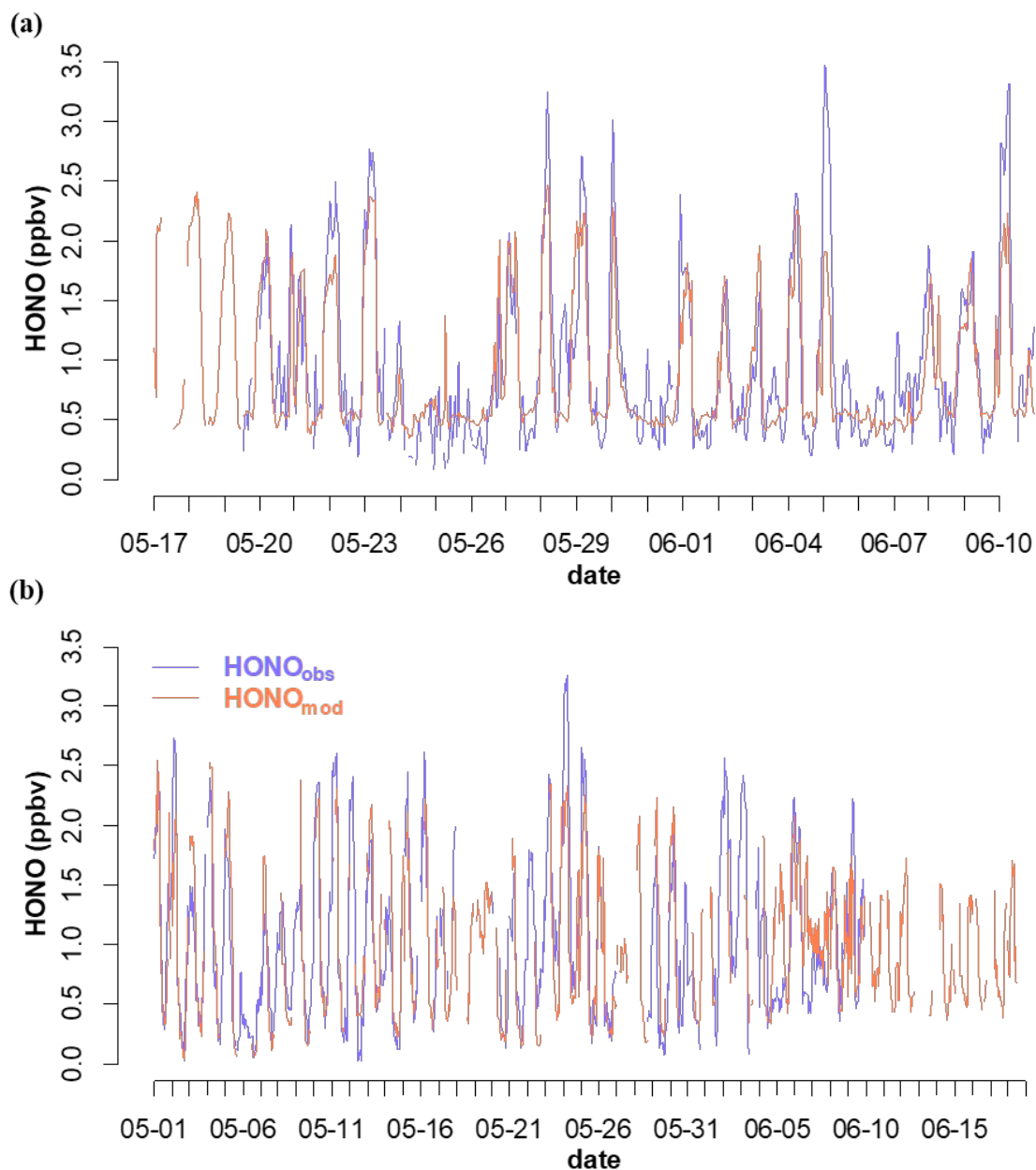



Figure 5. Comparison between the measured (HONO$_{obs}$) and calculated (HONO$_{mod}$) HONO mixing ratios in Seoul during May~June in (a) 2016 and (b) 2019. The blue and red lines indicate the measured and calculated HONO mixing ratio, respectively.

358

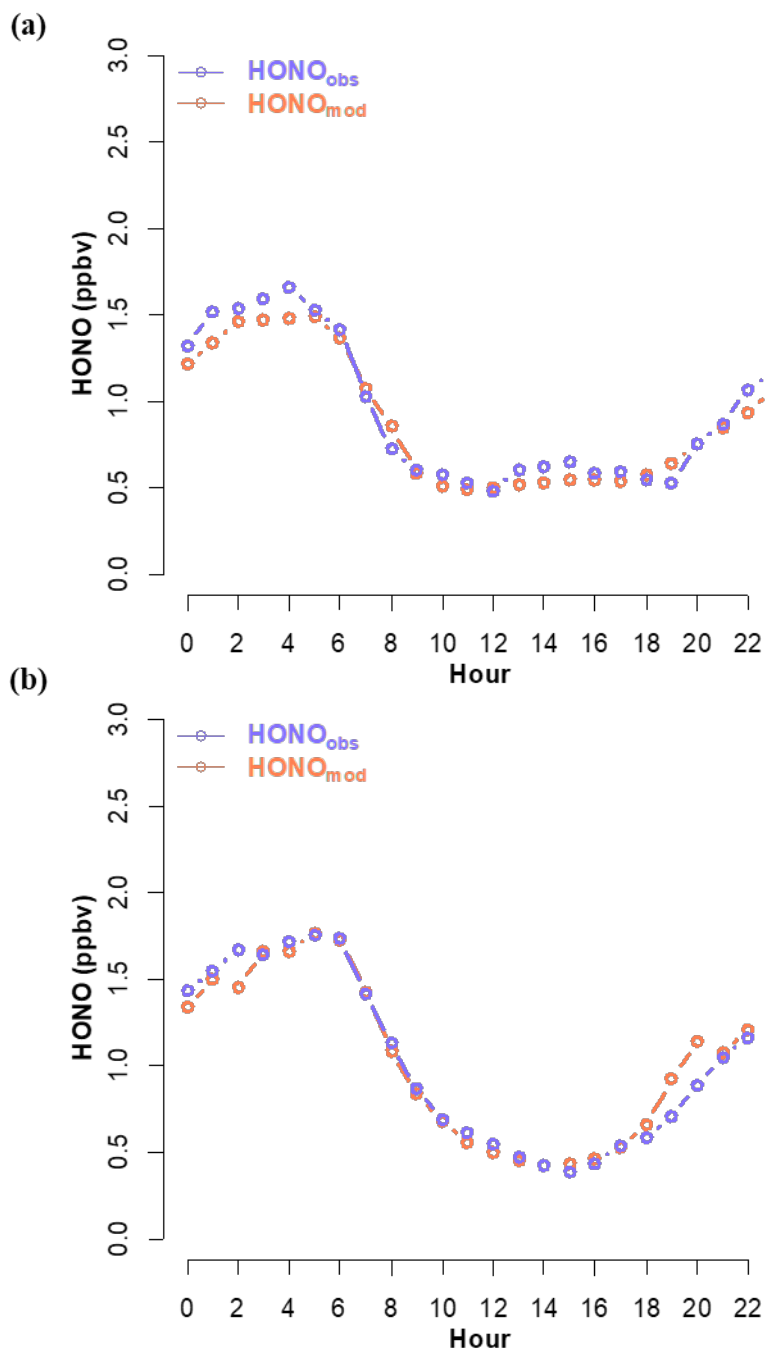



Figure 6. Average diurnal variations of the measured (HONO_obs) and the calculated (HONO_mod) HONO mixing ratios in Seoul during May ~ June in (a) 2016 and (b) 2019.

363

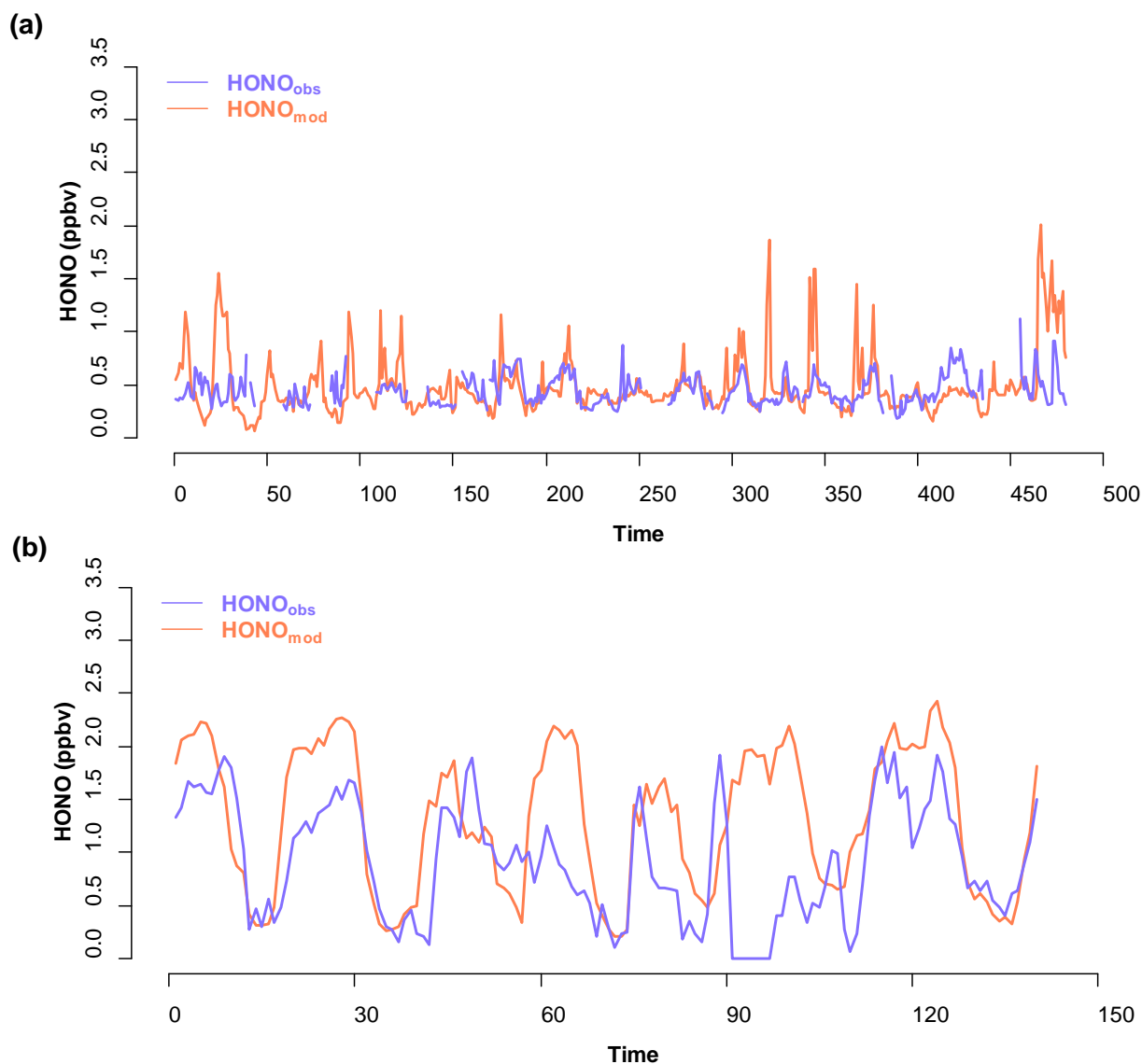



Figure 7. Comparison between the measured (HONO_obs) and calculated (HONO_mod) HONO
mixing ratios in Seoul during (a) June 2018 and (b) April 2019. The blue and red lines indicate
the measured and calculated HONO mixing ratio, respectively. The x axis indicates the hour
from the beginning of the experiment, which is (a) 00:00 on 1st June 2018 and (b) 00:00 on 12th
April 2019.

369

370

371

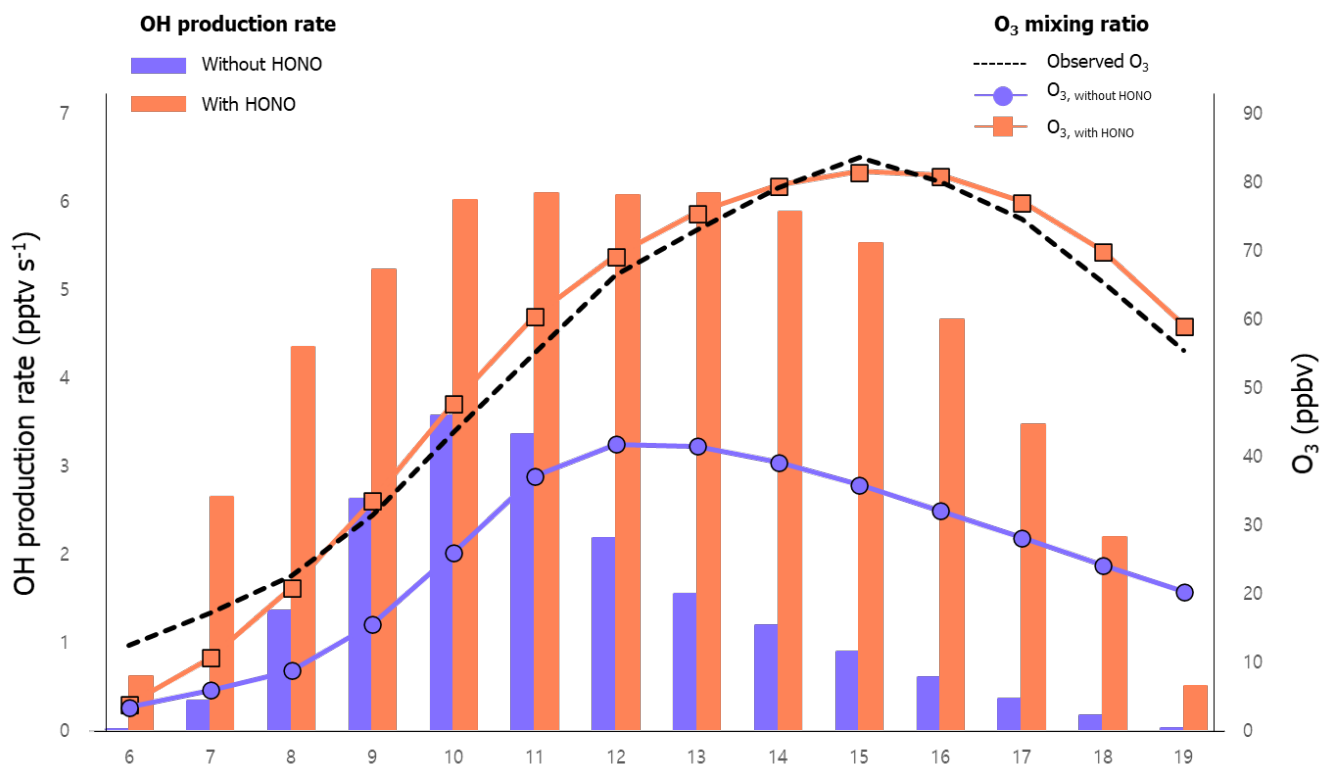



Figure 8. For June 2016, diurnal variations of O_3 (line) and OH production rate (bar) calculated from the F0AM photochemical model with (orange) and without (blue) HONO estimated from the RNDv1.0 model. The measured O_3 is compared with the calculated.

375

376   **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*

377   *GPU denotes graphic processing unit

378 **Table 2**. Input variables of the RNDv1.0 model and their ranges (10th and 90th percentile)

379 observed in Seoul during May ~ June in 2016 and 2019.

	10th~90th 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 (°)	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 (°)	99.4	359.565	0.235
Output Variables				
HONO	0.3 ~ 2.0 (ppbv)	81.1%	3.447	0.013

380 * Maximum – Minimum

381 ** Minimum value

382

383   **Table 3.** The result of validation and test of RNDv1.0 model using measurement data.

Measurement data	Validation		Test	
	MAE (ppbv)	IOA	MAE (ppbv)	IOA
May 2016*	0.26	0.93		
June 2016*	0.29	0.86		
June 2018	0.21	0.79		
April 2019			0.56	0.65
May 2019*	0.26	0.93		
June 2019*	0.36	0.76		

384   *Re-using the data already used for training

385

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.

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.

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.

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.

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.

571 Sun, Y., Wang, L., Wang, Y., Quan, L., and Zirui, L.: In situ measurements of SO2, NOx, NOy,
572 and O3 in Beijing, China during August 2008, Science of the Total Environment, 409, 933-940,
573 2011.
574 Tie, X., Geng, F., Guenther, A., Cao, J., Greenberg, J., Zhang, R., Apel, E., Li, G., Weinheimer,
575 A., and Chen, J.: Megacity impacts on regional ozone formation: observations and WRF-Chem
576 modeling for the MIRAGE-Shanghai field campaign, Atmospheric Chemistry and Physics, 13,
577 5655-5669, 2013.
578 VandenBoer, T., Markovic, M., Sanders, J., Ren, X., Pusede, S., Browne, E., Cohen, R., Zhang,
579 L., Thomas, J., and Brune, W. H.: Evidence for a nitrous acid (HONO) reservoir at the ground
580 surface in Bakersfield, CA, during CalNex 2010, Journal of Geophysical Research:
581 Atmospheres, 119, 9093-9106, 2014.
582 Varotsos, K., Giannakopoulos, C., and Tombrou, M.: Assessment of the Impacts of climate
583 change on european ozone levels, Water, Air, & Soil Pollution, 224, 1-13, 2013.
584 Wang, H., Chen, X., Lu, K., Hu, R., Li, Z., Wang, H., Ma, X., Yang, X., Chen, S., and Dong,
585 H.: NO3 and N2O5 chemistry at a suburban site during the EXPLORE-YRD campaign in 2018,
586 Atmospheric Environment, 224, 117180, 2020.
587 Wang, T., Qin, Z., and Zhu, M.: An ELU network with total variation for image denoising,
588 International Conference on Neural Information Processing, 2017a, 227-237.
589 Wang, X., Wang, H., Xue, L., Wang, T., Wang, L., Gu, R., Wang, W., Tham, Y. J., Wang, Z.,
590 and Yang, L.: Observations of N2O5 and ClNO2 at a polluted urban surface site in North China:
591 High N2O5 uptake coefficients and low ClNO2 product yields, Atmospheric environment, 156,
592 125-134, 2017b.
593 Wolfe, G. M., Marvin, M. R., Roberts, S. J., Travis, K. R., and Liao, J.: The framework for 0-
594 D atmospheric modeling (F0AM) v3. 1, Geoscientific Model Development, 9, 3309, 2016.
595 Xu, Z., Liu, Y., Nie, W., Sun, P., Chi, X., and Ding, A.: Evaluating the measurement interference
596 of wet rotating-denuder–ion chromatography in measuring atmospheric HONO in a highly
597 polluted area, Atmospheric Measurement Techniques, 12, 6737-6748, 2019.
598 Xue, C., Ye, C., Ma, Z., Liu, P., Zhang, Y., Zhang, C., Tang, K., Zhang, W., Zhao, X., and Wang,
599 Y.: Development of stripping coil-ion chromatograph method and intercomparison with CEAS
600 and LOPAP to measure atmospheric HONO, Science of The Total Environment, 646, 187-195,
601 2019.
602 Ye, C., Zhou, X., Pu, D., Stutz, J., Festa, J., Spolaor, M., Tsai, C., Cantrell, C., Mauldin, R. L.,
603 and Campos, T.: Rapid cycling of reactive nitrogen in the marine boundary layer, Nature, 532,
604 489-491, 2016.

605