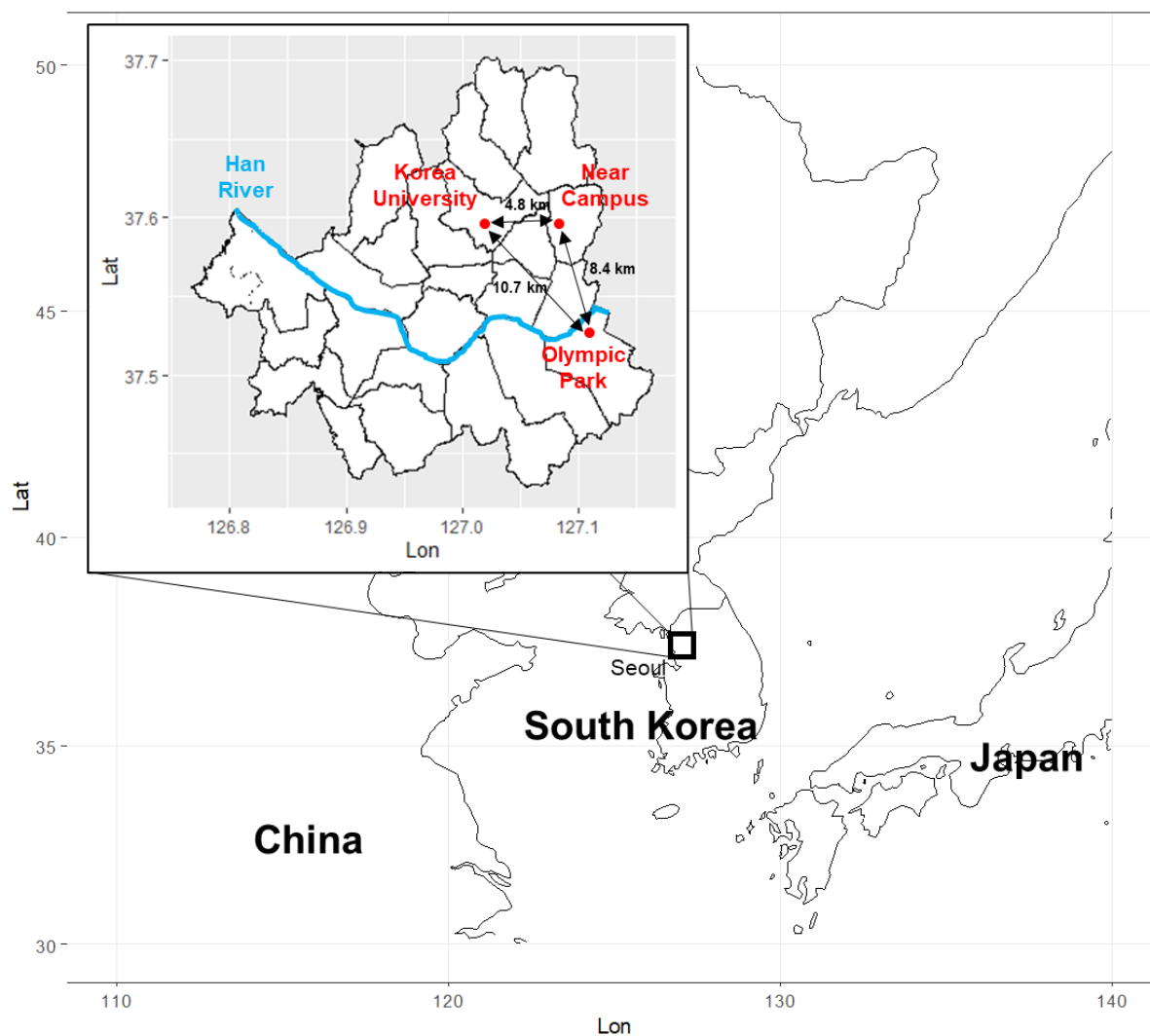Geoscientific
Model Development

Supplement of

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

Junsu Gil et al.

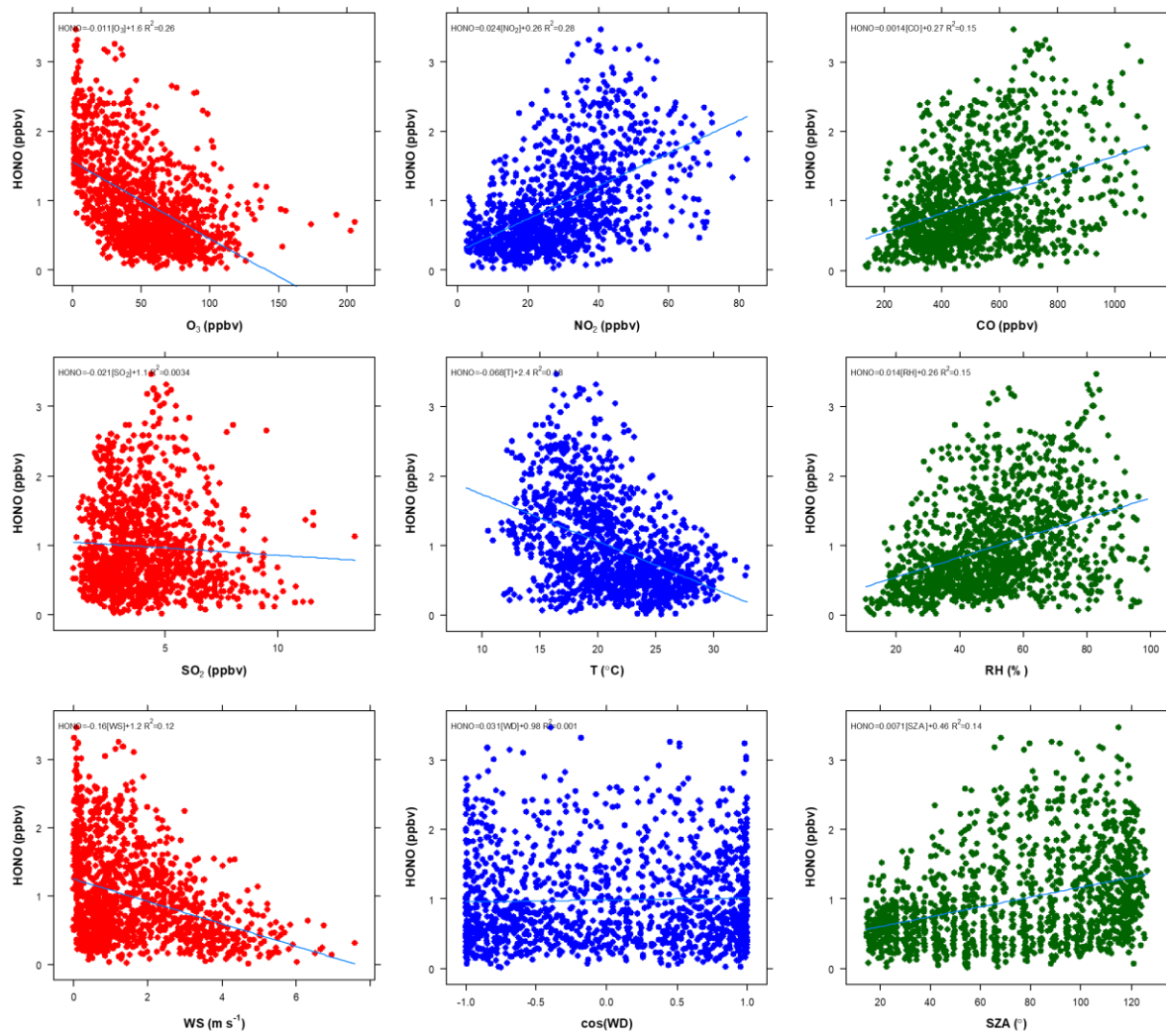*Correspondence to:* Meehye Lee (meehye@korea.ac.kr)

Supplementary

Figure S1. The map shows the locations where HONO measurements were made in Seoul from 2016 to 2021.

Figure S2. The correlation between HONO and other input variables.

```
1    #Set the pathway of files and model
2    cd C:\Users\atmos\Desktop\
3
4    #import libraries
5    import pandas as pd
6    import numpy as np
7    import tensorflow as tf
8    import keras
9
10   #Read preprocessed input data file and makes the shpae of data be proper to model input
11   raw_data = pd.read_csv('Dataset_for_model.csv', encoding = 'cp949')
12   data_X = raw_data[["O3_sc","NO2_sc","CO_sc","SO2_sc","SZA_sc","T_sc","RH_sc","WS_sc","WD_sc"]]
13   final_val_X = data_X.dropna(axis=0)
14   final_val_X1 = final_val_X.to_numpy().astype('float64')
15
16   #Load model
17   model = keras.models.load_model('RND_v1.0.h5')
18
19   #Calculate HONO
20   final_val_Y1 = model.predict(final_val_X1)
```

Figure S3. Total code for simulation of HONO using the RND model
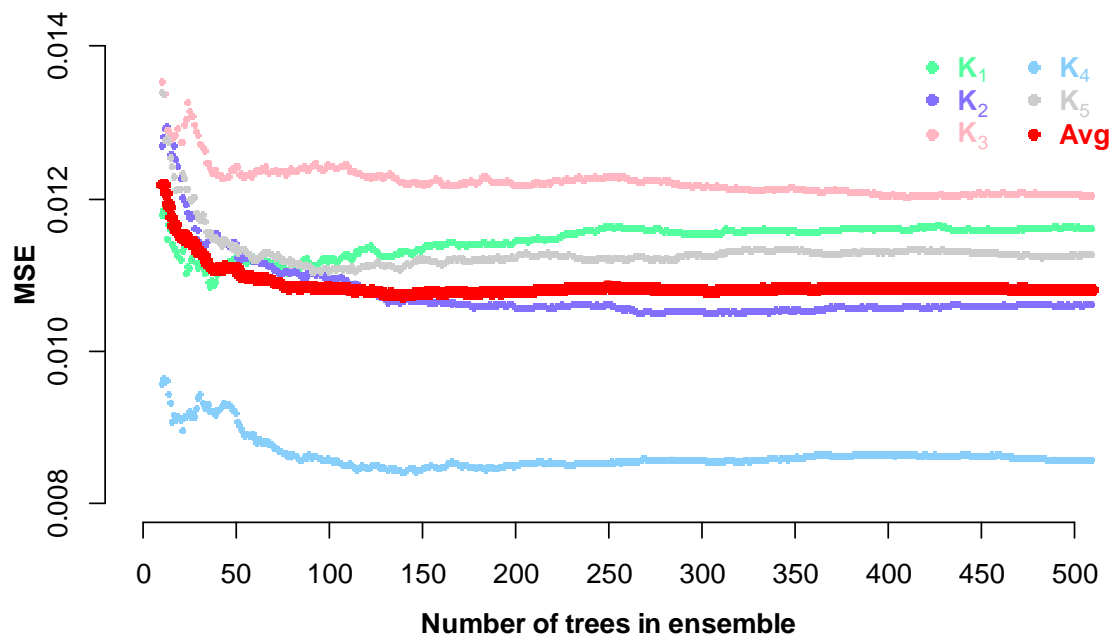
Figure S4. K-fold cross validation for Random Forest (RF) model by changing the number of trees in ensemble.
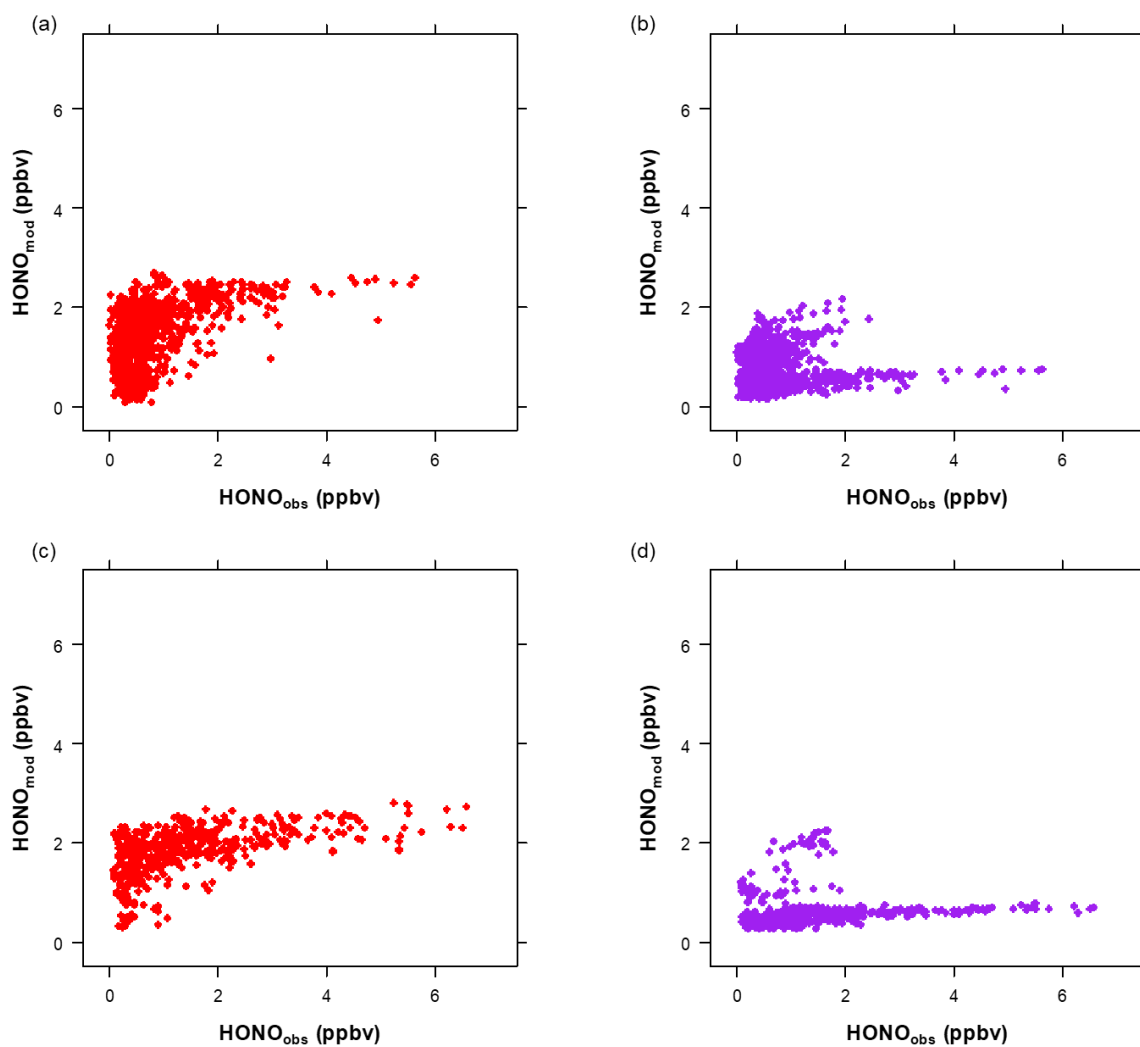
Figure S5. Relationship between measured HONO (HONO$_{obs}$) and modeled HONO (HONO$_{mod}$) using RNDv1.0 (red) and a Random Forest model (purple) for the test dataset. (a) and (b) present data in which all input variables are within the range of the train dataset, and (c) and (d) are the others that do not meet the criteria.

Table S1. The range of input variables (hourly measurements) used in this study.

Minimum	O$_3$ (ppbv)	NO$_2$ (ppbv)	CO (ppbv)	SO$_2$ (ppbv)	SZA (°)	T (°C)	RH (%)	WS (m s^{-1})	HONO (ppbv)
Train-Validation	0.8	2.4	137.3	1.0	14.195	8.6	10.6	0.010	0.01
Test	1.0	1.7	165.6	0.1	14.900	-2.2	9.7	0.270	0.10
Maximum	O$_3$ (ppbv)	NO$_2$ (ppbv)	CO (ppbv)	SO$_2$ (ppbv)	SZA (°)	T (°C)	RH (%)	WS (m s^{-1})	HONO (ppbv)
Train-Validation	205.6	82.3	1112.5	13.4	126.512	32.9	99.1	7.586	3.5
Test	132.0	88.5	1500.0	10.4	163.600	32.5	100.0	8.179	6.6

Table S2. The performance of RNDv1.0 and a Random Forest (RF) model on the test dataset that is divided into 'in' where all input parameters fall within the range of the train dataset and 'out' that do not meet the criteria.

	RNDv1.0$_{in}$	RNDv1.0$_{out}$	RF$_{in}$	RF$_{out}$
IOA	0.71	0.82	0.28	0.73
r	0.55	0.52	-0.02	0.10
MAE	0.64	0.63	0.55	0.66
RMSE	0.86	0.96	0.87	1.24