Supplement of Atmos. Meas. Tech., 16, 1279–1294, 2023
https://doi.org/10.5194/amt-16-1279-2023-supplement

Atmospheric Measurement Techniques
Open Access

Supplement of

Typhoon-associated air quality over the Guangdong–Hong Kong–Macao Greater Bay Area, China: machine-learning-based prediction and assessment

Yilin Chen et al.

Correspondence to: Yuanjian Yang (yyj1985@nuist.edu.cn)

1. Section S1: The data preprocessing procedure

First, the typhoon, ERA5, and air quality data are matched in time (Beijing Time, BJT), and then in space. The ERA5 dataset is gridded, with a spatial resolution of 0.5° × 0.5°. Linear interpolation is adopted to interpolate the meteorological variables to the 36 monitoring stations. There are also missing values in the air quality data, due to monitoring equipment malfunctions and communication network failures. These missing values are eliminated. Ultimately, the sample size after preprocessing for AQI, $PM_{2.5}$, PM_{10}, SO_2, NO_2 and O_3 in TY (NTY) days are 55141(66789), 63894(78961), 61639(74698), 65170(79369), 64972(79122) and 64837(78818), respectively. According to previous studies (Chow et al. 2018; Deng et al. 2019), only typhoons that are located within a certain extent (10°–35°N, 100°–130°E) will cause the favorable meteorological conditions necessary for air pollution in the GBA. Therefore, in this study, only those typhoons that enter this extent are introduced into the model. Meanwhile, given the year when CNEMC started to release their air quality data publicly (i.e., 2014), the study period used here is from June 2014 to December 2020.

2. Section S2: Hyperparameters tuning

There are 2 hyperparameters adjusted in the present study: the number of decision trees (*n-estimators*) and the maximum depth of the decision tree (*max-depth*). If these two hyperparameters are too low, the model will not capture the characteristics of the data very well, and its performance on both the training set and testing set is not ideal. This phenomenon is called underfitting. If the two hyperparameters are too high, overfitting will occur, i.e., the model over-captures the data feature of the training set, resulting in poor generalization ability of the model. In this case, the model performs well in the training set, but the performance of the test set is very poor. To find out the best *n-estimators* and *max-depth*, a tuning method that combines random search and grid search is adopted. Grid search is a kind of exhausted tuning method, which iterates over all specified hyperparameter combinations (parameter space) and calculates the score of all parameter combinations to find the optimal one, it has high accuracy in finding the best hyperparameters, but also spends a long time. Random search is to randomly select some points in the specified parameter space to determine the best hyperparameter combination among these selected, its advantage is to save the program running time, but the accuracy is reduced compared to grid search. Taking their strengths and weaknesses into account, the two methods are combined. First a random search is performed in large parameter space, roughly find the location of the best parameter combination, and then use a smaller step grid search at this position to obtain the best model parameter combination.

3. Tables S1 to S10 and captions

Table S1. Evaluation metrics of the model prediction of the case Danas in Guangzhou.

	AQI	$PM_{2.5}$	PM_{10}	SO_2	NO_2	O_3
MAE	10.078	6.841μg/m^3	11.879μg/m^3	6.029μg/m^3	7.636μg/m^3	26.085μg/m^3
RMSE	12.657	8.237μg/m^3	14.042μg/m^3	6.547μg/m^3	8.646μg/m^3	32.675μg/m^3
Bias	-2.337	-2.941μg/m^3	-0.731μg/m^3	-6.029μg/m^3	-2.129μg/m^3	-11.427μg/m^3
R	*0.692	*0.626	*0.569	*0.611	*0.802	*0.900
SD_O	14.274	8.441μg/m^3	14.772μg/m^3	2.159μg/m^3	12.900μg/m^3	65.342μg/m^3
SD_P	12.812	7.290μg/m^3	9.529μg/m^3	2.606μg/m^3	9.708μg/m^3	52.706μg/m^3
IA	0.826	0.775	0.705	0.488	0.873	0.933

Note: the correlation coefficient marked with "*" is significant with a significance level of 0.05.

Table S2. Evaluation metrics of the model prediction of the case Danas in Shenzhen.

	AQI	$PM_{2.5}$	PM_{10}	SO_2	NO_2	O_3
MAE	8.518	6.287μg/m^3	11.709μg/m^3	3.810μg/m^3	6.259μg/m^3	16.461μg/m^3
RMSE	12.704	9.299μg/m^3	15.592μg/m^3	4.097μg/m^3	7.502μg/m^3	26.950μg/m^3
Bias	-0.096	1.608μg/m^3	0.417μg/m^3	-3.810μg/m^3	-1.404μg/m^3	3.400μg/m^3
R	*0.813	*0.722	*0.656	*0.537	*0.810	*0.951
SD_O	16.793	11.120μg/m^3	17.174μg/m^3	1.045μg/m^3	11.251μg/m^3	50.982μg/m^3
SD_P	11.846	7.364μg/m^3	11.218μg/m^3	1.459μg/m^3	9.669μg/m^3	35.882μg/m^3
IA	0.853	0.793	0.746	0.386	0.884	0.931

Note: the correlation coefficient marked with "*" is significant with a significance level of 0.05.

Table S3. Evaluation metrics of the model prediction of the case Danas in Zhuhai.

	AQI	$PM_{2.5}$	PM_{10}	SO_2	NO_2	O_3
MAE	10.928	6.781μg/m^3	9.388μg/m^3	3.623μg/m^3	5.407μg/m^3	18.245μg/m^3
RMSE	15.805	8.836μg/m^3	14.914μg/m^3	4.121μg/m^3	6.143μg/m^3	28.338μg/m^3
Bias	-1.395	0.257μg/m^3	-1.605μg/m^3	-3.532μg/m^3	-3.510μg/m^3	5.614μg/m^3
R	*0.786	*0.726	*0.563	*0.701	*0.685	*0.909
SD_O	17.708	10.105μg/m^3	12.808μg/m^3	1.219μg/m^3	5.561μg/m^3	40.205μg/m^3
SD_P	14.796	10.279μg/m^3	15.641μg/m^3	2.480μg/m^3	4.899μg/m^3	28.394μg/m^3
IA	0.847	0.848	0.757	0.470	0.738	0.896

Note: the correlation coefficient marked with "*" is significant with a significance level of 0.05.

Table S4. Evaluation metrics of the model prediction of the case Danas in Foshan.

	AQI	$PM_{2.5}$	PM_{10}	SO_2	NO_2	O_3
MAE	9.076	5.285μg/m^3	9.026μg/m^3	6.512μg/m^3	5.707μg/m^3	20.849μg/m^3
RMSE	11.909	6.778μg/m^3	11.282μg/m^3	7.229μg/m^3	7.664μg/m^3	26.611μg/m^3
Bias	-7.040	-4.139μg/m^3	-6.785μg/m^3	-5.625μg/m^3	-0.501μg/m^3	-14.166μg/m^3
R	*0.802	*0.806	*0.711	*0.590	*0.727	*0.926
SD_O	12.311	7.577μg/m^3	9.859μg/m^3	3.848μg/m^3	10.091μg/m^3	54.042μg/m^3
SD_P	13.341	7.714μg/m^3	10.240μg/m^3	3.890μg/m^3	7.605μg/m^3	52.41μg/m^3
IA	0.851	0.851	0.779	0.603	0.824	0.948

Note: the correlation coefficient marked with "*" is significant with a significance level of 0.05.

Table S5. Evaluation metrics of the model prediction of the case Danas in Zhaoqin.

	AQI	$PM_{2.5}$	PM_{10}	SO_2	NO_2	O_3
MAE	8.686	7.516μg/m^3	11.837μg/m^3	6.708μg/m^3	5.939μg/m^3	21.788μg/m^3
RMSE	11.615	9.674μg/m^3	14.281μg/m^3	7.037μg/m^3	7.074μg/m^3	26.161μg/m^3
Bias	-5.839	-4.282μg/m^3	-6.016μg/m^3	-4.863μg/m^3	-0.203μg/m^3	-15.739μg/m^3
R	*0.570	*0.489	*0.528	-0.065	*0.731	*0.904
SD_O	9.368	7.500μg/m^3	11.947μg/m^3	2.968μg/m^3	9.132μg/m^3	42.858μg/m^3
SD_P	9.454	6.795μg/m^3	9.797μg/m^3	1.875μg/m^3	5.702μg/m^3	45.133μg/m^3
IA	0.702	0.643	0.667	0.324	0.806	0.924

Note: the correlation coefficient marked with "*" is significant with a significance level of 0.05.

Table S6. Evaluation metrics of the model prediction of the case Danas in Jiangmen.

	AQI	$PM_{2.5}$	PM_{10}	SO_2	NO_2	O_3
MAE	12.485	8.475μg/m^3	16.111μg/m^3	6.087μg/m^3	7.161μg/m^3	19.059μg/m^3
RMSE	15.500	10.158μg/m^3	19.183μg/m^3	7.037μg/m^3	10.020μg/m^3	24.934μg/m^3
Bias	-3.203	-3.315μg/m^3	-1.537μg/m^3	-6.087μg/m^3	1.322μg/m^3	-11.936μg/m^3
R	*0.583	*0.593	0.287	0.248	0.324	*0.905
SD_O	15.511	10.011μg/m^3	15.837μg/m^3	1.137μg/m^3	9.305μg/m^3	40.216μg/m^3
SD_P	15.109	9.304μg/m^3	12.866μg/m^3	3.018μg/m^3	3.214μg/m^3	44.330μg/m^3
IA	0.773	0.760	0.568	0.255	0.470	0.937

Note: the correlation coefficient marked with "*" is significant with a significance level of 0.05.

Table S7. Evaluation metrics of the model prediction of the case Danas in Huizhou.

	AQI	$PM_{2.5}$	PM_{10}	SO_2	NO_2	O_3
MAE	11.576	5.991μg/m^3	14.129μg/m^3	3.453μg/m^3	7.033μg/m^3	14.471μg/m^3
RMSE	12.565	6.961μg/m^3	15.528μg/m^3	3.987μg/m^3	9.379μg/m^3	19.372μg/m^3
Bias	4.569	0.896μg/m^3	6.249μg/m^3	-3.430μg/m^3	2.511μg/m^3	0.291μg/m^3
R	*0.768	*0.759	*0.760	0.416	*0.830	*0.941
SD_O	14.200	9.453μg/m^3	18.886μg/m^3	1.528μg/m^3	14.065μg/m^3	45.979μg/m^3
SD_P	10.838	6.880μg/m^3	13.448μg/m^3	1.614μg/m^3	8.663μg/m^3	38.424μg/m^3
IA	0.827	0.843	0.813	0.454	0.855	0.960

Note: the correlation coefficient marked with "*" is significant with a significance level of 0.05.

Table S8. Evaluation metrics of the model prediction of the case Danas in Dongguan.

	AQI	$PM_{2.5}$	PM_{10}	SO_2	NO_2	O_3
MAE	13.188	8.983μg/m^3	14.345μg/m^3	4.490μg/m^3	13.661μg/m^3	26.806μg/m^3
RMSE	17.469	10.935μg/m^3	19.310μg/m^3	5.645μg/m^3	16.595μg/m^3	35.262μg/m^3
Bias	-3.925	-3.940μg/m^3	-4.593μg/m^3	-4.293μg/m^3	3.044μg/m^3	-5.662μg/m^3
R	*0.694	*0.645	*0.501	*0.550	*0.547	*0.863
SD_O	18.737	11.298μg/m^3	17.765μg/m^3	2.535μg/m^3	18.365μg/m^3	60.518μg/m^3
SD_P	14.038	7.730μg/m^3	10.959μg/m^3	3.146μg/m^3	7.361μg/m^3	56.954μg/m^3
IA	0.800	0.737	0.629	0.546	0.627	0.922

Note: the correlation coefficient marked with "*" is significant with a significance level of 0.05.

Table S9. Evaluation metrics of the model prediction of the case Danas in Zhongshan.

	AQI	$PM_{2.5}$	PM_{10}	SO_2	NO_2	O_3
MAE	11.927	5.502μg/m^3	10.218μg/m^3	4.055μg/m^3	5.640μg/m^3	27.207μg/m^3
RMSE	15.588	6.966μg/m^3	13.369μg/m^3	5.222μg/m^3	6.982μg/m^3	33.673μg/m^3
Bias	-3.383	-1.840μg/m^3	-3.893μg/m^3	-4.055μg/m^3	-3.380μg/m^3	-2.186μg/m^3
R	*0.848	*0.824	*0.745	*0.713	*0.732	*0.850
SD_O	20.816	10.453μg/m^3	16.695μg/m^3	1.658μg/m^3	7.547μg/m^3	49.516μg/m^3
SD_P	17.041	10.071μg/m^3	15.441μg/m^3	3.433μg/m^3	6.747μg/m^3	44.730μg/m^3
IA	0.881	0.900	0.848	0.515	0.805	0.912

Note: the correlation coefficient marked with "*" is significant with a significance level of 0.05.

Table S10. Evaluation metrics of the model prediction of the case Danas in Hong Kong.

	$PM_{2.5}$	PM_{10}	SO_2	NO_2	O_3
MAE	6.080μg/m^3	10.217μg/m^3	3.687μg/m^3	6.481μg/m^3	23.921μg/m^3
RMSE	8.575μg/m^3	13.425μg/m^3	4.058μg/m^3	7.700μg/m^3	35.786μg/m^3
Bias	3.168μg/m^3	3.409μg/m^3	-3.563μg/m^3	-3.437μg/m^3	13.108μg/m^3
R	*0.767	*0.698	*0.550	*0.799	*0.923
SD_O	11.080μg/m^3	16.317μg/m^3	1.621μg/m^3	9.571μg/m^3	47.446μg/m^3
SD_P	7.664μg/m^3	11.709μg/m^3	1.836μg/m^3	10.203μg/m^3	23.685μg/m^3
IA	0.814	0.789	0.480	0.864	0.839

Note: the correlation coefficient marked with "*" is significant with a significance level of 0.05.

4. Figures S1 to S8 and captions

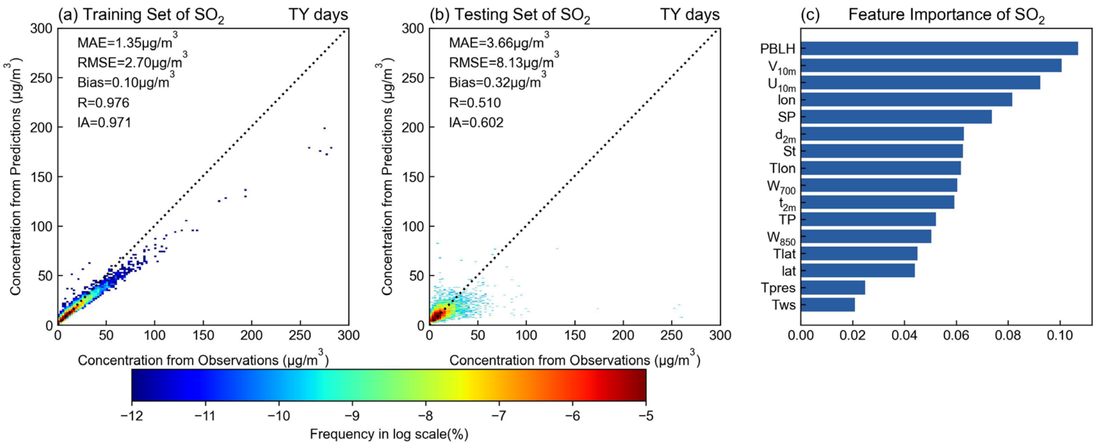

Figure S1. Comparison of the predicted and observed concentration of SO_2 in TY days: (a) training set; (b) testing set; (c) feature importance.

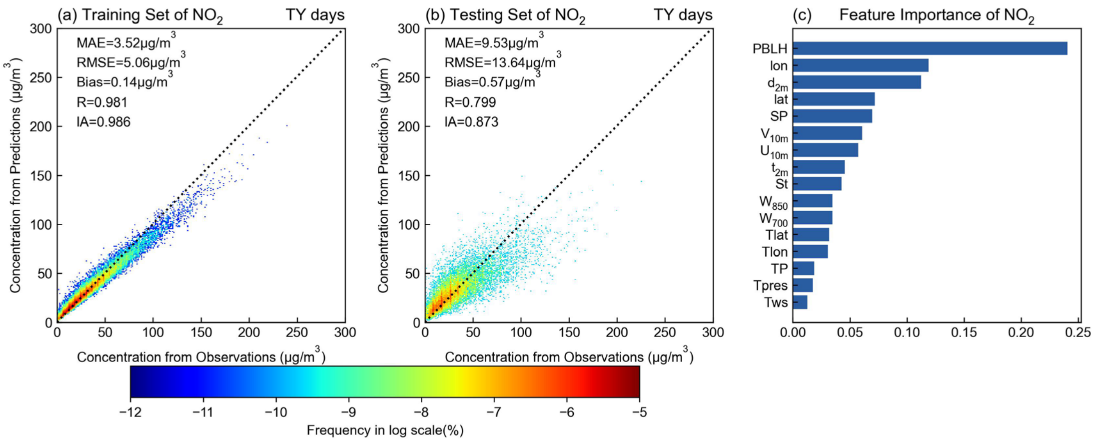

Figure S2. Comparison of the predicted and observed concentration of NO_2 in TY days: (a) training set; (b) testing set; (c) feature importance.

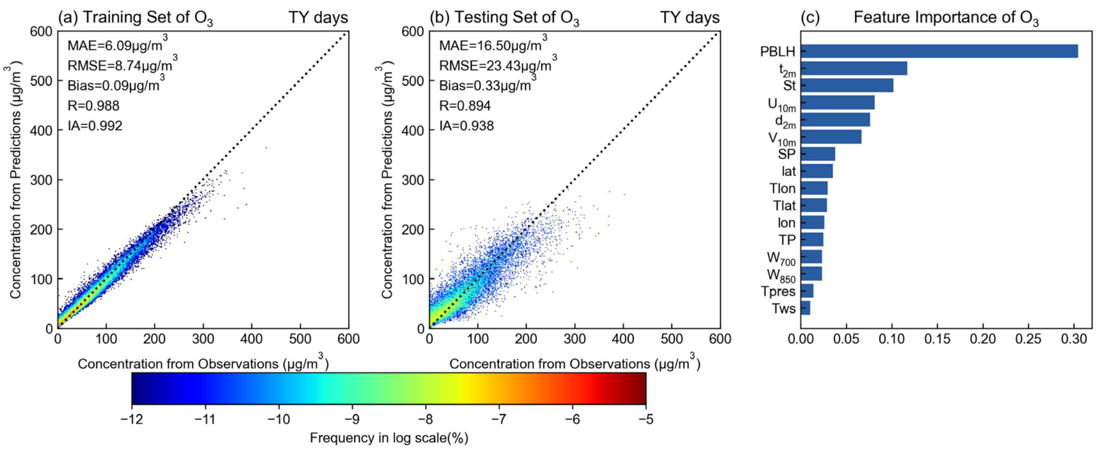

Figure S3. Comparison of the predicted and observed concentration of O_3 in TY days: (a) training set; (b) testing set; (c) feature importance.

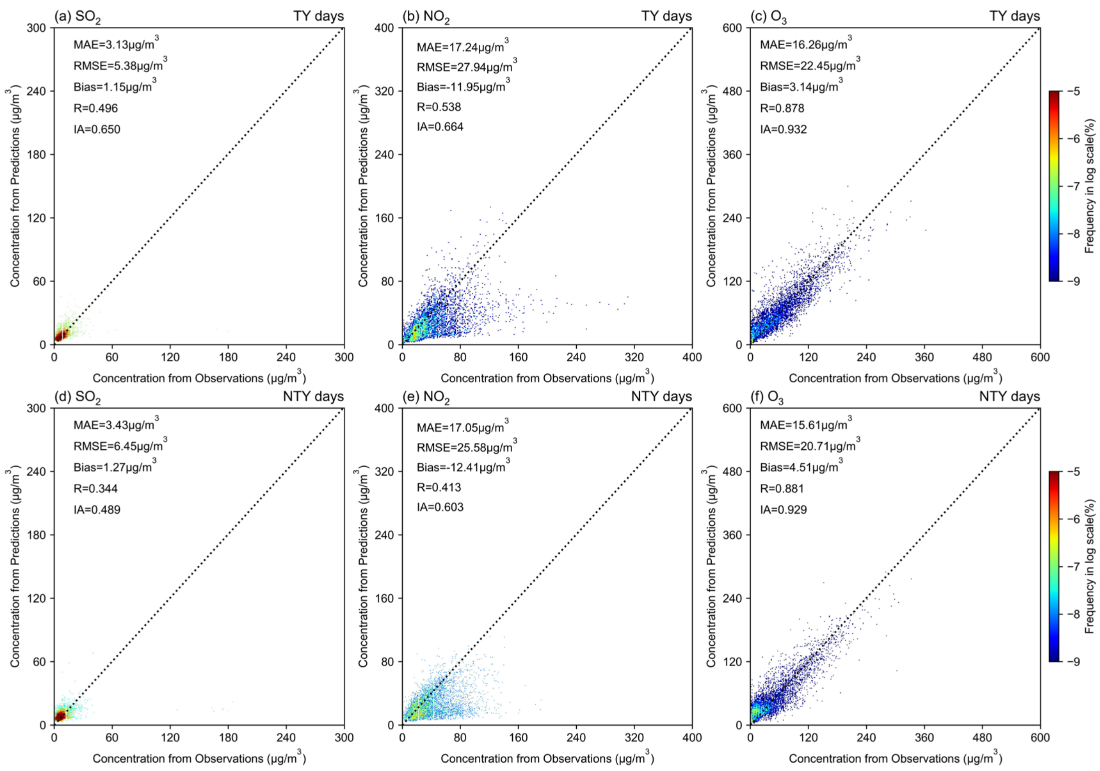

Figure S4. The result of SO_2, NO_2 and O_3 of 3 additional monitoring stations predicted by the RF model. (a) SO_2 of TY days; (b) NO_2 of TY days; (c) O_3 of TY days; (d) SO_2 of NTY days; (e) NO_2 of NTY days; (f) O_3 of NTY days.

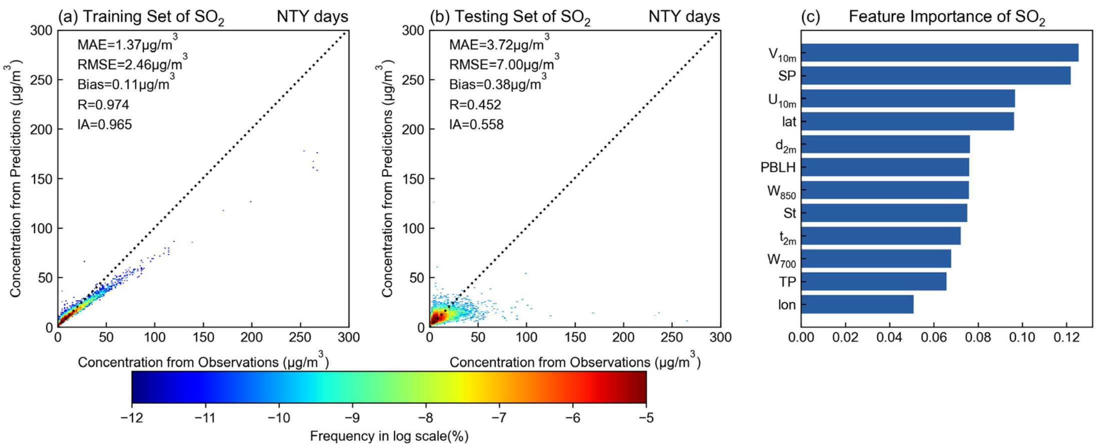

Figure S5. Comparison of the predicted and observed concentration of SO_2 in NTY days: (a) training set; (b) testing set; (c) feature importance.

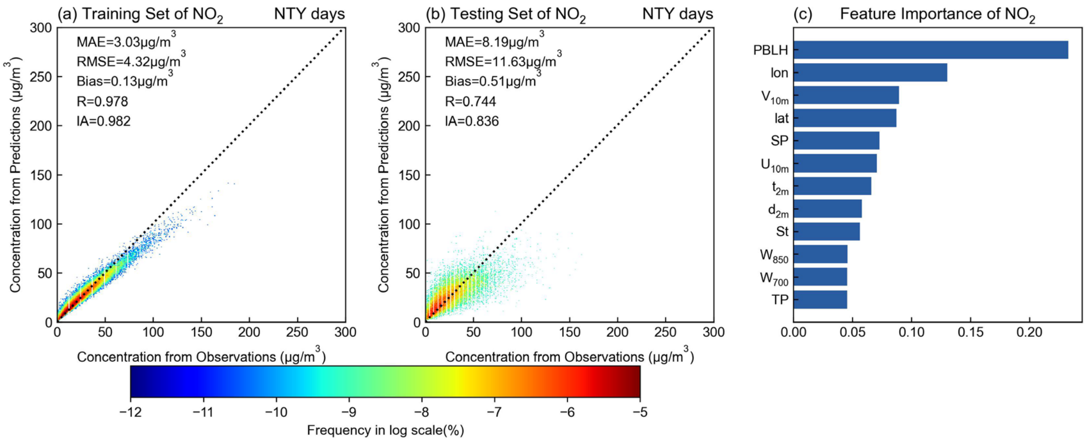

Figure S6. Comparison of the predicted and observed concentration of NO_2 in NTY days: (a) training set; (b) testing set; (c) feature importance.

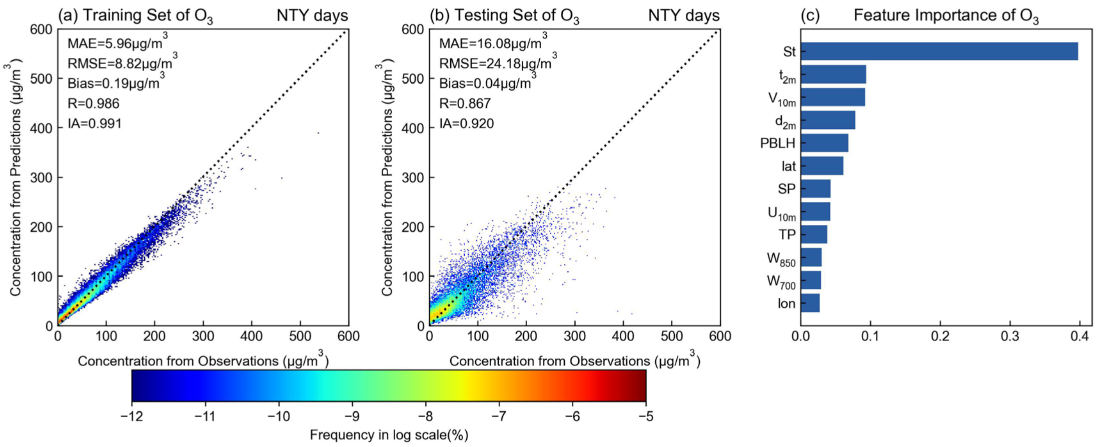

Figure S7. Comparison of the predicted and observed concentration of O_3 in NTY days: (a) training set; (b) testing set; (c) feature importance.

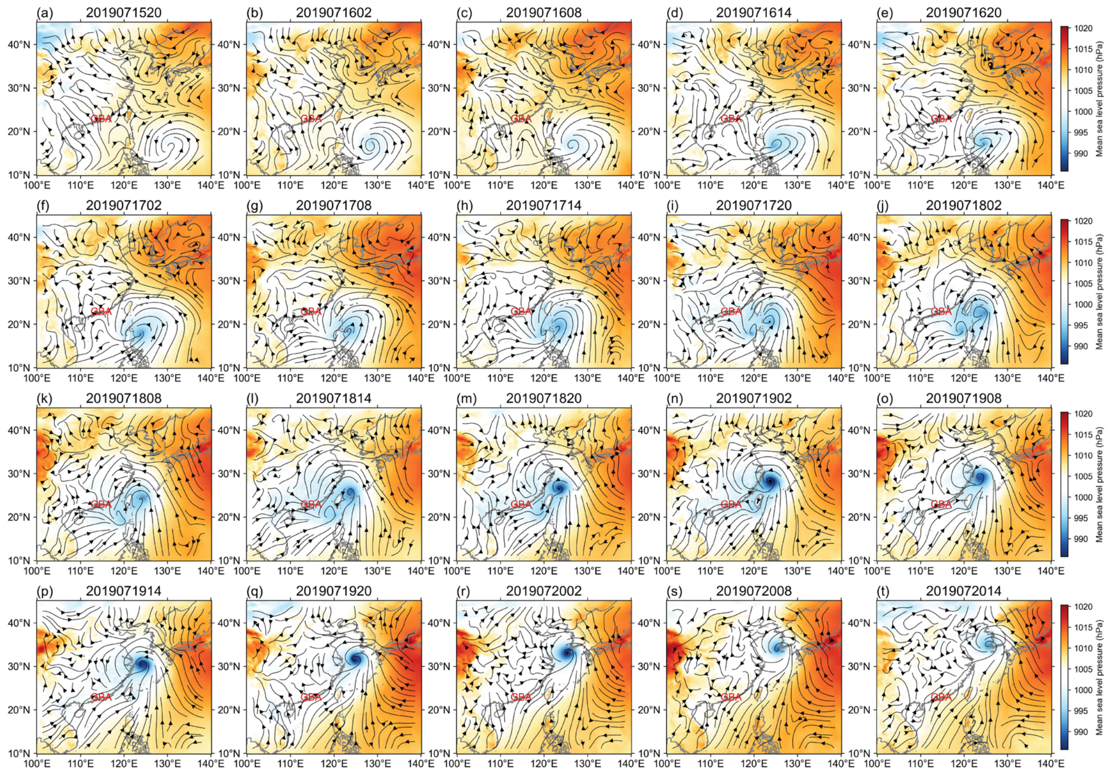

Figure S8. The synoptic chart of typhoon Danas: the streamline indicates surface wind, the shading indicates sea-level pressure.

5. Reference

Chow, E. C. H., Li, R. C. Y., and Zhou, W.: Influence of Tropical Cyclones on Hong Kong Air Quality, Adv. Atmos. Sci., 35, 1177–1188, https://doi.org/10.1007/s00376-018-7225-4, 2018.

Deng, T., Wang, T., Wang, S., Zou, Y., Yin, C., Li, F., Liu, L., Wang, N., Song, L., Wu, C., and Wu, D.: Impact of typhoon periphery on high ozone and high aerosol pollution in the Pearl River Delta region, Sci. Total Environ., 668, 617–630, https://doi.org/10.1016/j.scitotenv.2019.02.450, 2019.