

Supporting Materials

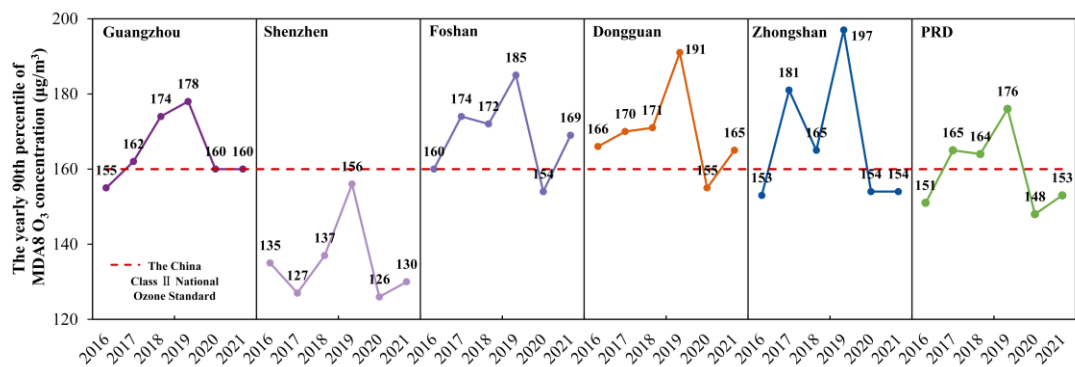
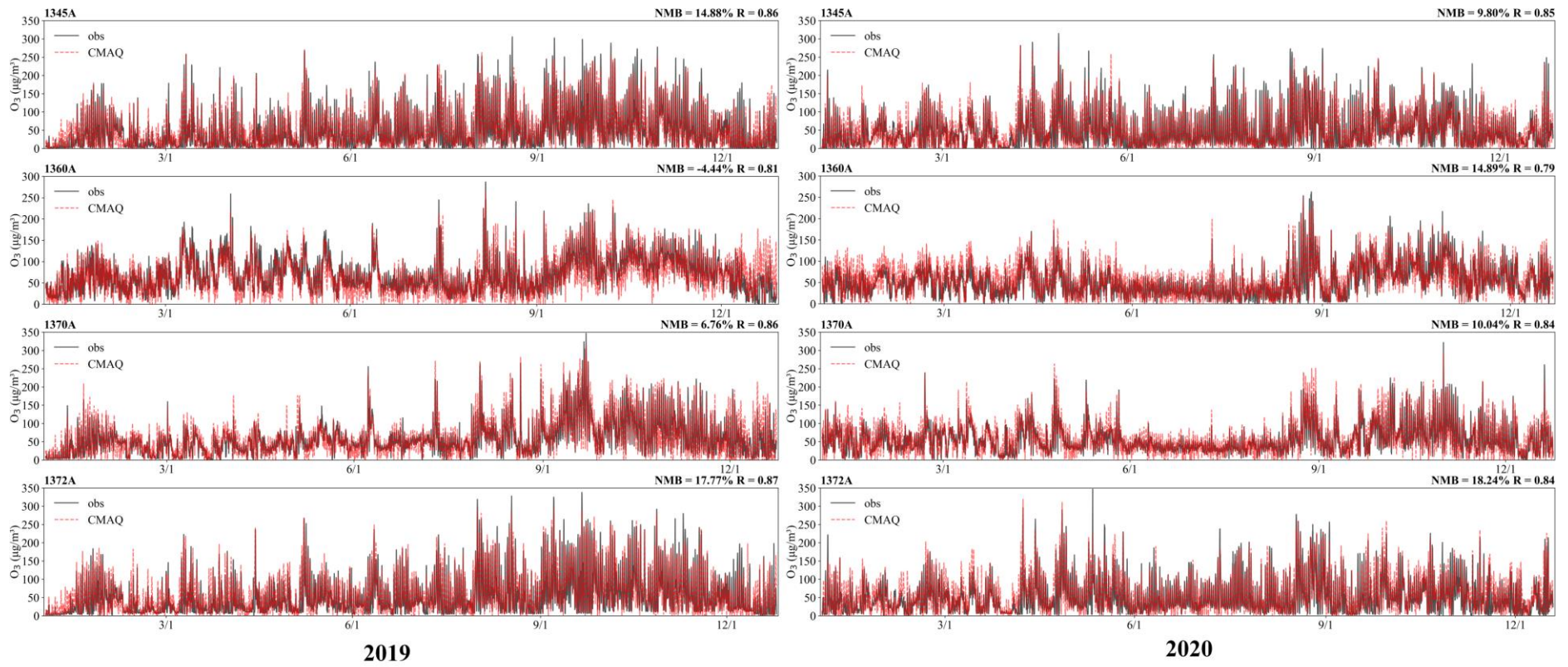
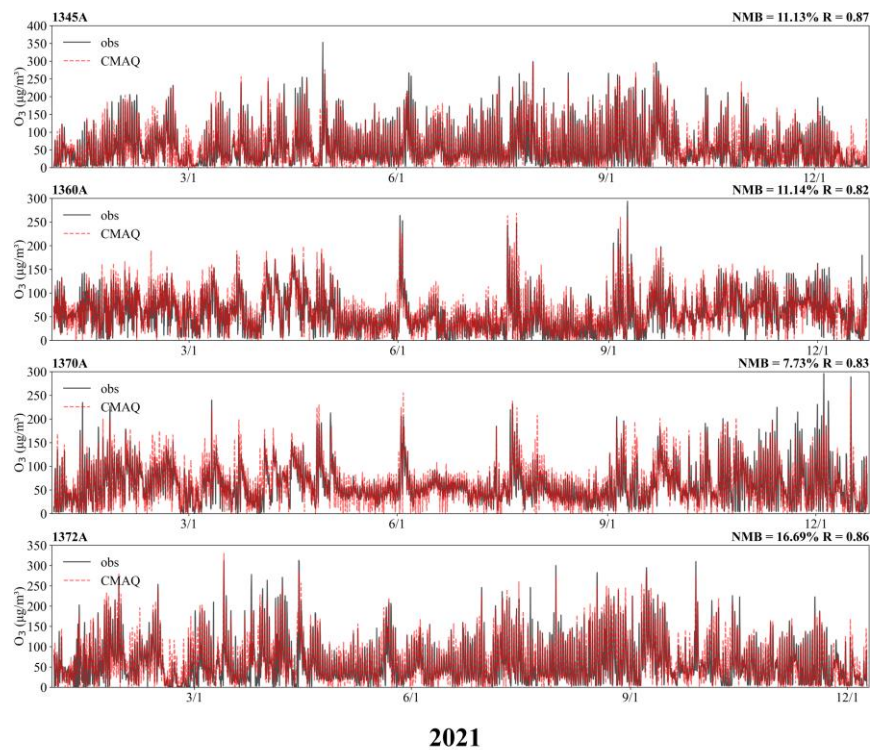


Fig. S1 The yearly 90th percentile of the maximum daily 8-h average (MDA8-90%) O₃ concentration in core PRD cities and PRD from 2016 to 2021 (GDEEP, 2023).



(a)



(b)

Fig. S2 Comparison of O₃ hourly series observed (obs) and simulated by WRF-CAMQ (CMAQ) at representative sites in (a) 2019, 2020 and (b) 2021 (1345A: Guangya High School in Guangzhou (GZ); 1360A: Saltpan site in Shenzhen (SZ); 1370A: site of Doumen in Zhuhai (ZH); 1372A: Huazai Vocational High School in Foshan (FS). Among them, 1345A, 1372A are inland sites, 1360A and 1370A are coastal sites).

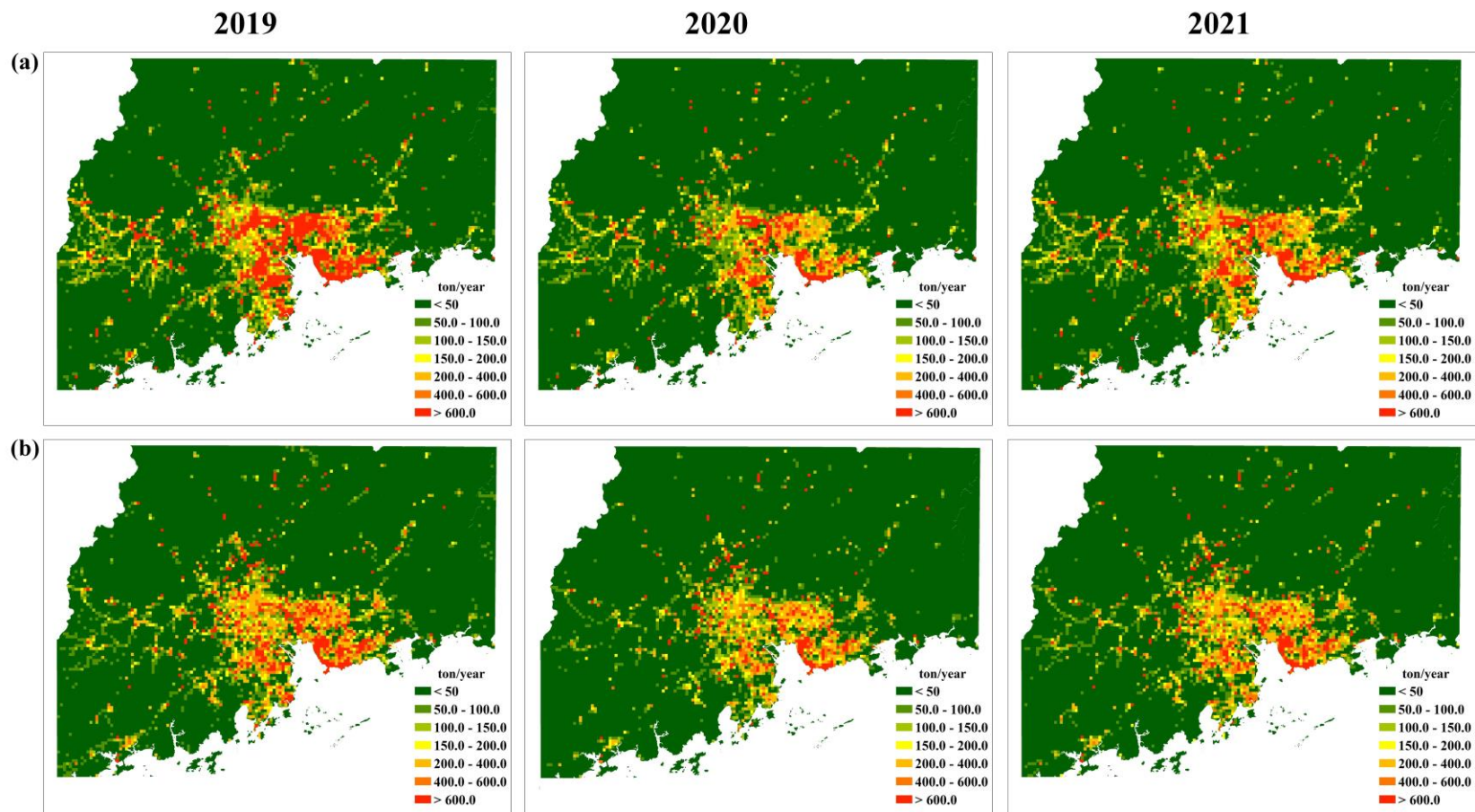


Fig. S3 Spatial distribution of total (a) NO_x and (b) anthropogenic VOC (AVOC) emissions in 2019, 2020 and 2021.

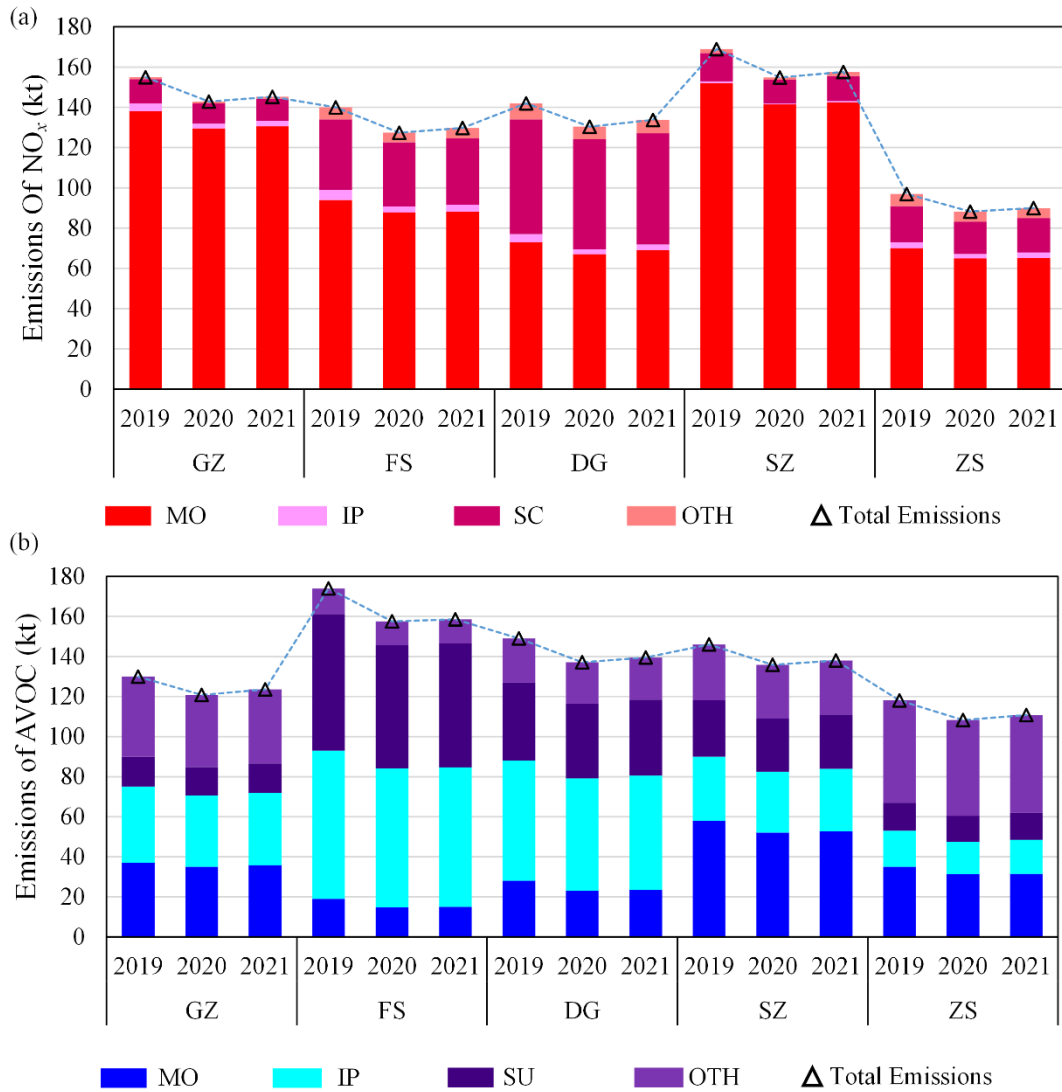


Fig. S4 Emissions of (a) NO_x and (b) AVOC from different sectors in each city of the D3 domain from 2019 to 2021.

GZ - Guangzhou; FS - Foshan; ZS - Zhongshan; DG - Dongguan; SZ - Shenzhen. MO: mobile; IP: industrial process; SU: solvent utilization; SC: stationary combustion; OTH of NO_x: other sources of NO_x (biomass burning and storage); OTH of AVOC: other sources of VOC (storage, stationary combustion, biomass burning, waste disposal).

	Case	GZ_NO _x	GZ_SO ₂	GZ_NH ₃	GZ_VOC	...	Oth_NO _x	Oth_SO ₂	Oth_NH ₃	Oth_VOC	
Baseline	1	1	1	1	1	...	1	1	1	1	Training Dataset
19 randomized controlled scenarios	2	0.6567	0.6419	1.3615	0.2665	...	0.6567	0.6419	1.3615	0.2665	
	3	1.1701	1.0344	1.0469	0.7264	...	1.1701	1.0344	1.0469	0.7264	
	4	0.8641	1.3689	0.2075	0.8403	...	0.8641	1.3689	0.2075	0.8403	
	5	0.3174	0.3161	1.4164	1.1947	...	0.3174	0.3161	1.4164	1.1947	
	6	1.0035	0.3850	0.3050	0.0351	...	1.0035	0.3850	0.3050	0.0351	
	7	1.2207	1.3483	0.8323	0.6373	...	1.2207	1.3483	0.8323	0.6373	
	8	0.0320	1.1819	0.8108	0.5965	...	0.0320	1.1819	0.8108	0.5965	
	9	0.6862	1.4308	1.1885	1.2062	...	0.6862	1.4308	1.1885	1.2062	
	10	0.2091	0.0293	0.3673	1.1546	...	0.2091	0.0293	0.3673	1.1546	
	11	0.9028	0.1938	1.2256	0.9879	...	0.9028	0.1938	1.2256	0.9879	
	12	0.0071	0.1522	1.3171	0.2891	...	0.0071	0.1522	1.3171	0.2891	
	13	1.4901	0.7964	1.0663	1.4324	...	1.4901	0.7964	1.0663	1.4324	
	14	0.5098	1.4804	1.4150	1.3414	...	0.5098	1.4804	1.4150	1.3414	
	15	1.3750	0.2971	0.5615	0.1926	...	1.3750	0.2971	0.5615	0.1926	
	16	0.3564	0.6427	0.0847	1.4959	...	0.3564	0.6427	0.0847	1.4959	
	17	1.4301	0.7920	1.0452	1.0024	...	1.4301	0.7920	1.0452	1.0024	
	18	0.5187	0.1754	0.5036	1.3776	...	0.5187	0.1754	0.5036	1.3776	
	19	0.1927	1.4964	0.6820	0.3866	...	0.1927	1.4964	0.6820	0.3866	
	20	1.1052	0.6031	0.1134	0.0160	...	1.1052	0.6031	0.1134	0.0160	
Clean	21	0	0	0	0	...	0	0	0	0	Construction of DeepRSM-PRD
	22	0	0	0	0	...	1	1	1	1	
	23	1	1	1	1	...	1	1	1	1	
	24	1	1	1	1	...	1	1	1	1	
	25	1	1	1	1	...	1	1	1	1	
	26	1	1	1	1	...	1	1	1	1	
	27	1	1	1	1	...	0	0	0	0	

Fig. S5 Design of the control matrix.

The training dataset encompasses baseline, clean (full-control), and random-control scenarios. The “Construction of DeepRSM-PRD” involves creating a distinct full-control scenario for each subregion. (GZ: Guangzhou, FS: Foshan, DG: Dongguan, SZ: Shenzhen, ZS: Zhongshan, Oth: other regions in the D3 domain).

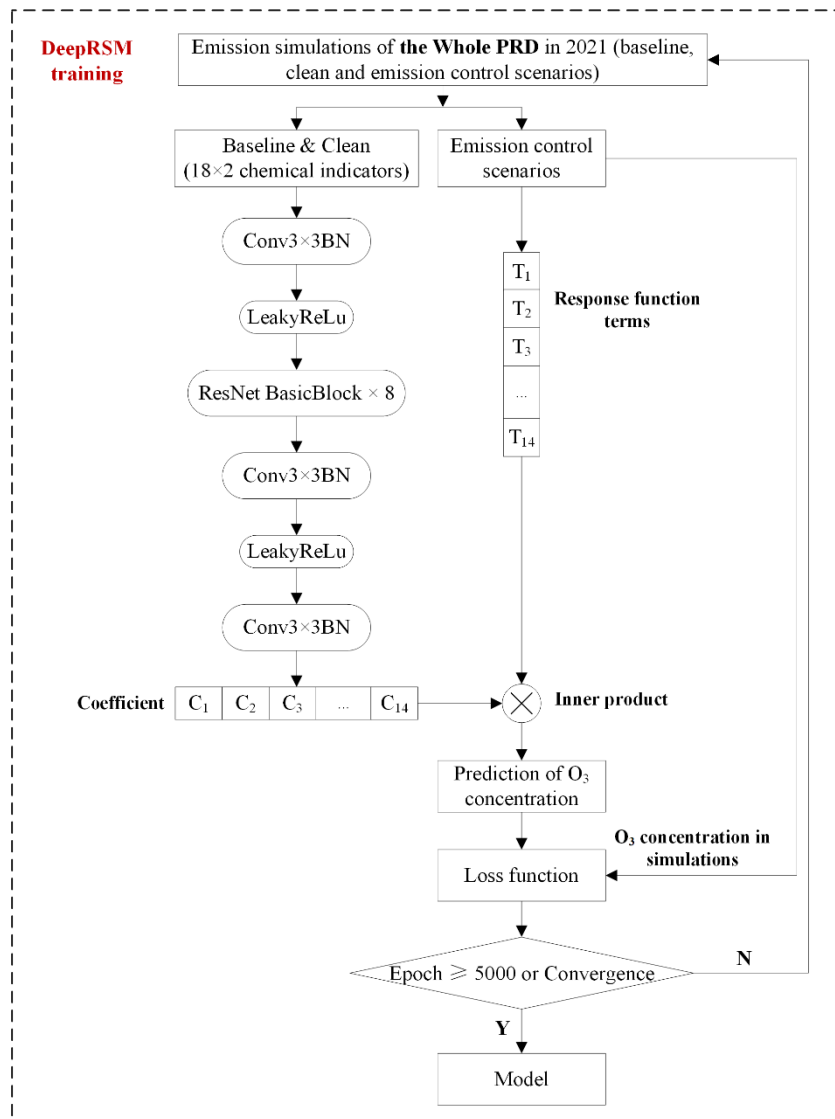


Fig. S6 The training process of DeepRSM-PRD.

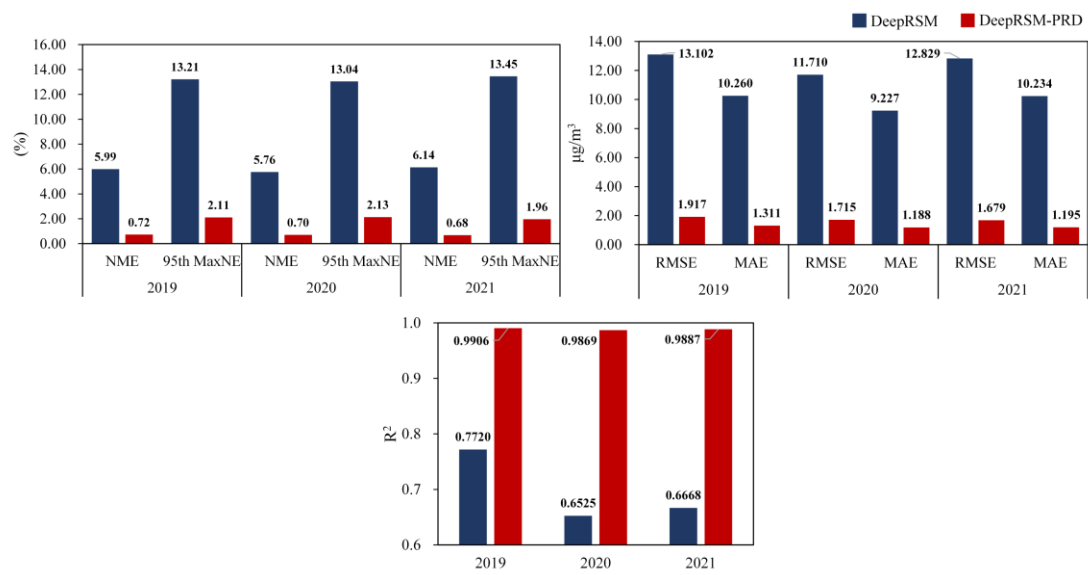


Fig. S7 Comparison of the performance between DeepRSM-PRD and DeepRSM in OOS validation.

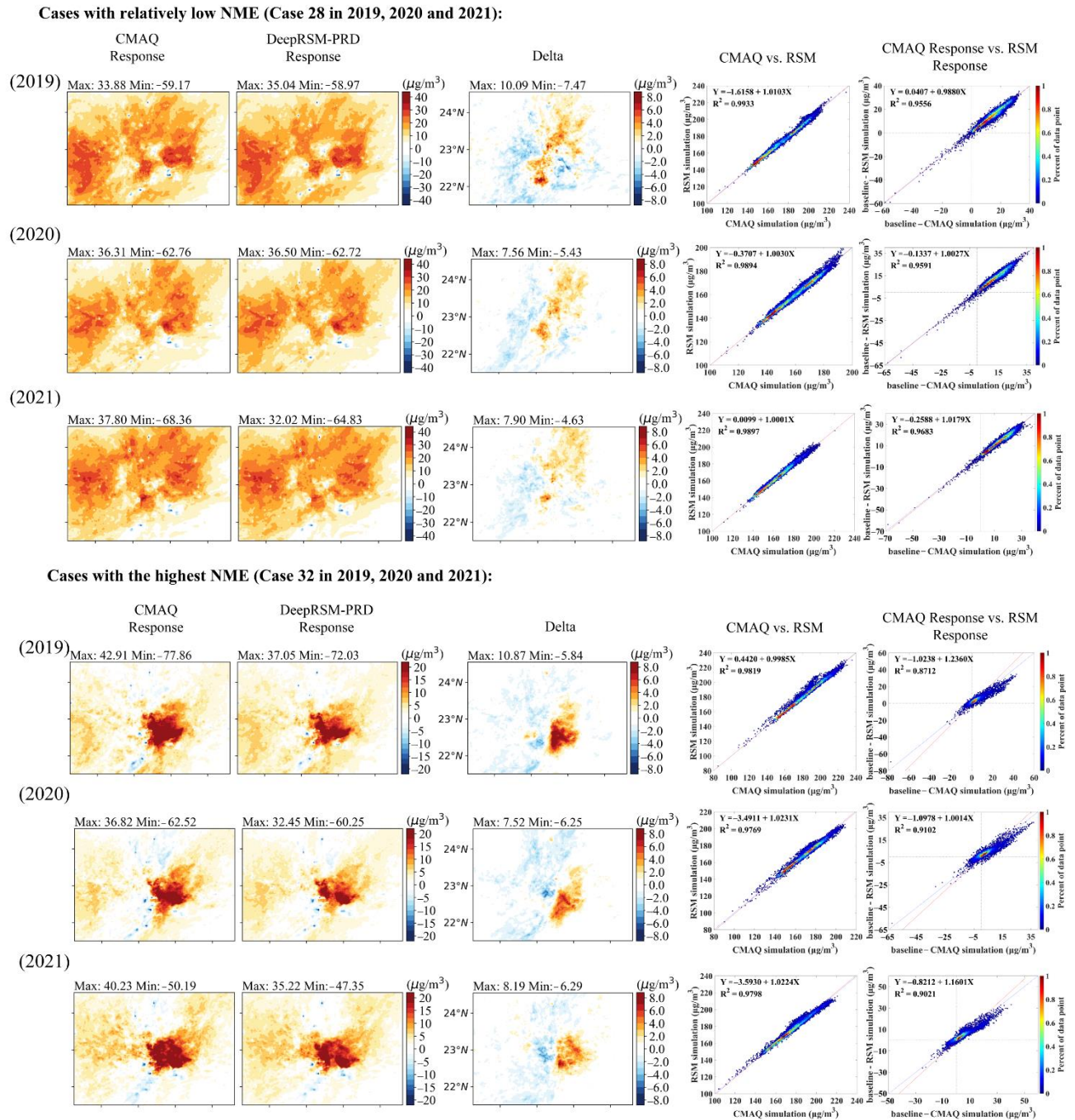


Fig. S8 Spatial distribution of O_3 response (MDA8-90% O_3) and density scatter plot in OOS validation.

CMAQ Response: difference in concentration between control and baseline scenarios (baseline – CMAQ simulation); DeepRSM Response: difference in concentration between control and baseline scenarios (baseline – DeepRSM simulation); Delta: difference in concentration between CMAQ response and DeepRSM response.

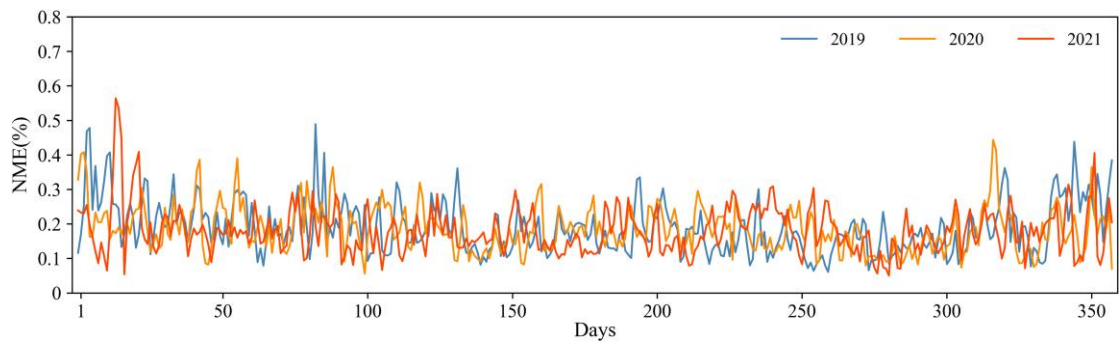


Fig. S9 Line plot of daily NME differences between 2019, 2020 and 2021 in OOS validation (MDA8 O₃).

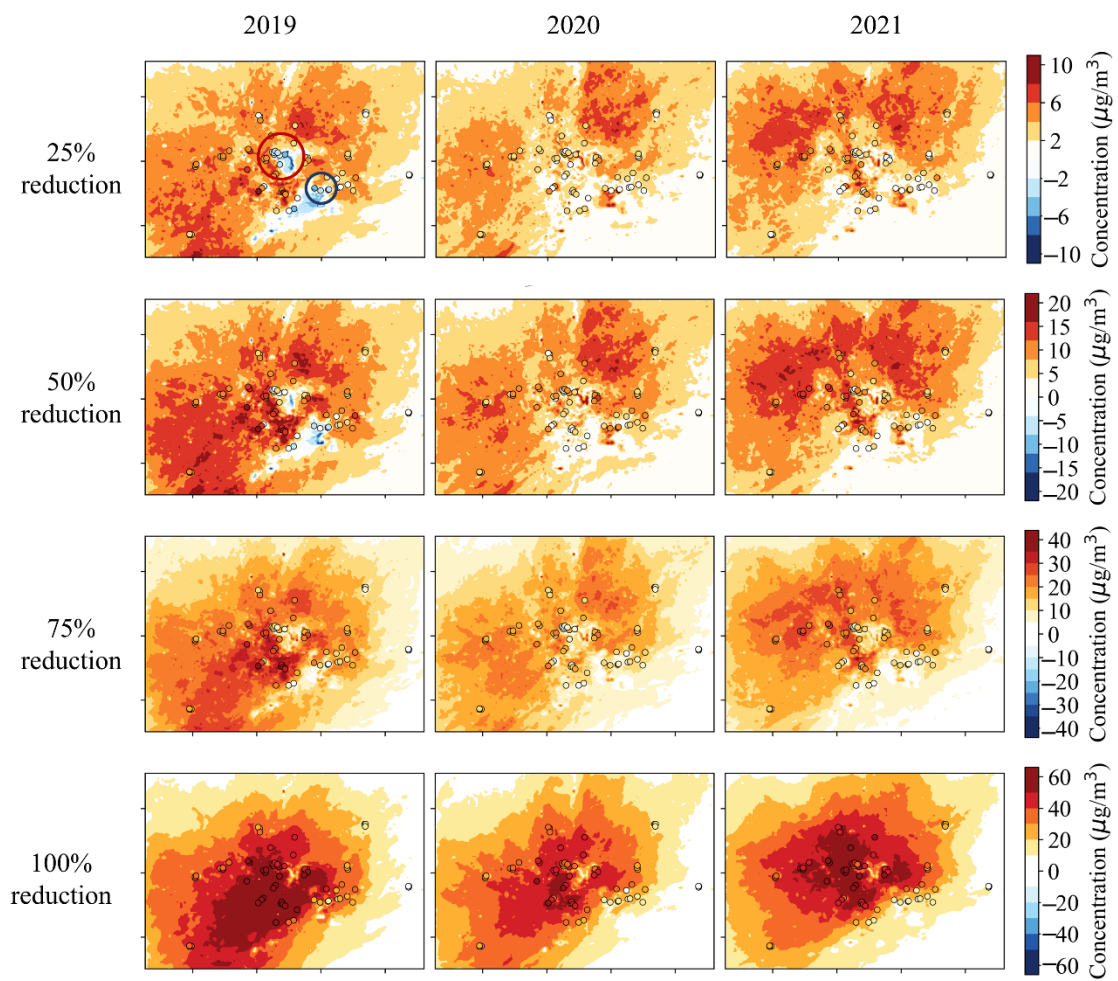


Fig. S10 Spatial of O₃ response concentration for NO_x reduction only (Note: The small circles on the figure represent national air quality monitoring sites).

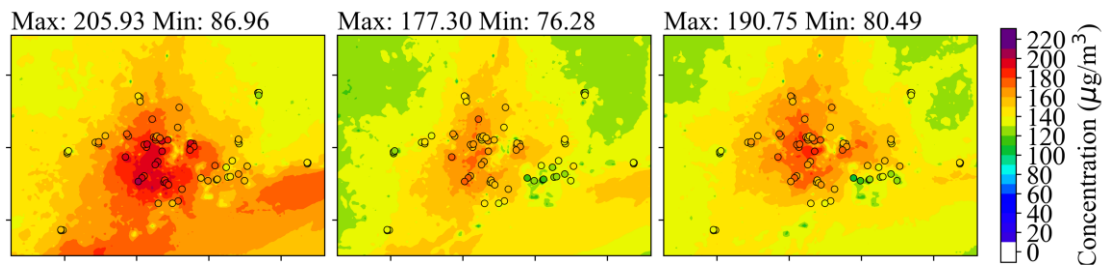


Fig. S11 The spatial distribution of MDA8-90% O₃ concentration from 2019 to 2021 (Note: The small circles on the figure represent national air quality monitoring sites).

Table S1 WRF model configuration.

Parameters	WRF version 3.9.1
Vertical layers	34
Nested grid	Three nested
Land-surface physics scheme	Noah
PBL physics scheme	Yenisei University
Cumulus	Kain-Fritsch cumulus cloud parameterization
Longwave radiation	Rapid radiative transfer model
Shortwave radiation	Goddard
Microphysics	Purdue Lin

Table S2 CMAQ model configuration.

Parameters	CMAQ version 5.3
Vertical layers	14
Nested grid	Three nested
Horizontal advection	Yamartino global mass conserving scheme
Horizontal diffusion	Explicit
Vertical advection	Yamartino global mass conserving scheme
Vertical diffusion	ACM2
Gas-phase chemistry	CB06
Gas-phase chemistry algorithm	EBI
Aerosol chemistry	AERO6
Dry deposition	M3DRY Accumulation and coarse mode particles completely
Wet deposition	absorbed in cloud water; Nuclei mode slowly scavenged; Henry's law equilibrium for gases

Table S3 Validation of hourly O₃ concentrations for the WRF-CMAQ simulations at representative sites (two inland and two coastal sites) ^a.

Year	Site	R	MAE (µg/m ³)	NMB (%)
2019	1345A	0.86	20.71	14.87
	1360A	0.81	18.90	-4.44
	1370A	0.86	19.12	6.76
	1372A	0.87	21.42	17.77
2020	1345A	0.85	19.04	9.80
	1360A	0.79	19.36	14.89
	1370A	0.84	18.27	10.04
	1372A	0.84	20.56	18.24
2021	1345A	0.87	20.28	11.13
	1360A	0.82	18.61	11.14
	1370A	0.83	17.94	7.73
	1372A	0.86	21.51	16.69

^a 1345A: Guangya High School in Guangzhou (GZ); 1360A: Saltpan site in Shenzhen (SZ); 1370A: site of Doumen in Zhuhai (ZH); 1372A: Huazai Vocational High School in Foshan (FS). Among them, 1345A, 1372A are inland sites, 1360A and 1370A are coastal sites.

Table S4 Total emissions of NO_x and AVOC in each area of the D3 domain in 2019, 2020 and 2021 (unit: kt/yr) ^a.

Year	Pollutant	GZ	FS	DG	SZ	ZS	Oth
2019	NO _x	155	139	142	169	97	607
	AVOC	130	174	149	146	118	579
2020	NO _x	143	127	130	155	88	566
	AVOC	121	157	137	136	108	555
2021	NO _x	145	130	134	158	90	579
	AVOC	124	158	139	138	111	561

^a GZ: Guangzhou; FS: Foshan; DG: Dongguan; SZ: Shenzhen; ZS: Zhongshan; Oth: Other regions in the D3 domain.

Table S5 Main hyper-parameters of DeepRSM method.

Hyper-parameters	Value
Batch size	24
Learning rate	0.001
Epoch	5000
Optimizer	Adam ($\beta_1=0.9$, $\beta_2=0.999$)
Kernel size	3×3
Activation function	Leaky ReLu

Table S6 Design of the 5 randomly selected control scenarios for OOS validation.

Case Number	Precursors	GZ	SZ	FS	DG	ZS	Oth
28	NO _x	0.7356	1.2463	1.2297	0.6827	1.1656	0.5588
	SO ₂	0.3977	0.6344	0.9806	1.1287	0.848	0.5039
	NH ₃	0.7128	0.8322	0.3698	1.4454	1.4296	0.1211
	VOC	0.6726	1.0852	1.4714	0.6249	0.6337	0.5420
29	NO _x	0.0071	1.3881	0.4101	0.7504	0.4617	1.4014
	SO ₂	0.1522	0.5035	0.6191	0.0254	1.1156	0.999
	NH ₃	1.3171	0.9106	0.1773	0.0157	1.1994	0.9977
	VOC	0.2891	1.1333	0.7317	1.4996	0.8257	1.3976
30	NO _x	0.7203	0.5693	0.6587	0.5927	1.0627	0.5539
	SO ₂	1.0389	1.1355	0.7297	0.3924	0.7206	0.3671
	NH ₃	0.6826	0.5742	1.2498	0.5117	0.277	1.3221
	VOC	0.8156	0.7426	0.4936	0.7289	0.5734	1.1081
31	NO _x	0.6862	0.4270	0.8726	1.3959	0.7276	0.4616
	SO ₂	1.4308	1.2105	1.4421	1.3557	0.2544	0.3497
	NH ₃	1.1885	0.759	0.8464	0.8859	0.5619	1.1988
	VOC	1.2062	0.374	0.2968	1.229	0.0081	1.3872
32	NO _x	1.4901	0.3181	1.276	0.9255	0.1061	0.6629
	SO ₂	0.7964	0.1248	0.0361	0.9731	0.5698	1.0517
	NH ₃	1.0663	0.4613	1.2859	0.985	0.7920	0.7779
	VOC	1.4324	0.2615	1.0166	0.0588	1.3707	1.2616

Table S7 The mean results of evaluation metrics in OOS validation (MDA8-90% O₃) in 2019, 2020, and 2021 ^a.

Year	RMSE ($\mu\text{g}/\text{m}^3$)	MAE ($\mu\text{g}/\text{m}^3$)	NME (%)	95th MaxNE (%)	R^2
2019	1.917 (1.530-2.455)	1.311 (1.091-1.487)	0.72 (0.61-0.81)	2.11 (1.68-2.51)	0.991 (0.982-0.993)
2020	1.715 (1.357-2.157)	1.188 (0.963-1.411)	0.70 (0.58-0.83)	2.13 (1.71-2.66)	0.987 (0.977-0.991)
2021	1.679 (1.380-2.030)	1.195 (0.984-1.363)	0.68 (0.59-0.77)	1.96 (1.74-2.37)	0.989 (0.980-0.994)

^a Values in parentheses are the minimum and maximum of the error between these OOS cases.

Table S8 Paired sample t-test results about the performance of DeepRSM-PRD in different years.

Serial Number	Paired Samples	Mean Difference	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		t	Degree of freedom	p-value
					Lower	Upper			
Pair 1	(2019,2020)	0.00463	0.08840	0.00467	-0.00456	0.01382	0.992	357	0.322
Pair 2	(2019,2021)	0.00988	0.10183	0.00538	-0.00070	0.02047	1.837	357	0.067
Pair 3	(2020,2021)	0.00525	0.08819	0.00466	-0.00392	0.01442	1.126	357	0.261

Text S1 Data Fusion technique.

The data fusion technique enhances the baseline results of the WRF-CMAQ simulation by integrating monitoring data, thereby ensuring greater consistency with the actual spatial distribution of concentration. This method employs specific algorithms such as VNA, eVNA, and Downscaler, with more details in the paper (Li et al., 2019).

Text S2 Validation of WRF-CMAQ Simulations.

The WRF-CMAQ modeling results were compared with monitoring data, specifically the O₃ hourly series values in each year from air quality monitoring stations (i.e., 1345A: Guangya High School in Guangzhou (GZ); 1360A: Saltpan site in Shenzhen (SZ); 1370A: site of Doumen in Zhuhai (ZH); 1372A: Huazai Vocational High School in Foshan (FS). Among them, 1345A, 1372A are inland sites, 1360A and 1370A are coastal sites). The assessment metrics, including R (correlation coefficient), MAE (Mean Absolute Error), and NMB (Normalized Mean Bias), which are detailed in Text S7. In 2019, the results showed R ranging from 0.81 to 0.87, MAE ranging from 18.90 to 21.42 $\mu\text{g}/\text{m}^3$, and NMB spanning from -4.44% to 17.77% . For 2020, R was concentrated in the range of 0.79 to 0.85, MAE varied from 18.27 to 20.56 $\mu\text{g}/\text{m}^3$, and NMB fell within the range of 9.80% to 18.24% . Further, in 2021, R ranged from 0.82 to 0.87, MAE ranged from 17.94 to 21.51 $\mu\text{g}/\text{m}^3$, and NMB fell within the range of 7.73% to 16.69% . Generally, these metric results align with recommended values from previous studies, meeting the requirement of $R > 0.5$ and NMB falling within ($\pm 15\%$) (Emery et al., 2017).

Table S3 provides a detailed summary of the evaluation results for representative sites and Fig. S2 illustrates the fitting trend.

Text S3 The polynomial function for O₃ response to precursor emission changes.

$$\Delta C_{O_3} = \sum_{i=1}^n K_i \times (E_{NO_x})^{a_i} \times (E_{VOCs})^{b_i} \times (E_{SO_2})^{c_i} \times (E_{NH_3})^{d_i} \quad (1)$$

Where ΔC_{O_3} represents the response concentration of O₃ relative to the baseline due to emission changes. n denotes the number of terms of the equation and K_i refers to the coefficients of the polynomial term i (in the pf-RSM, a minimum of 20 CTM scenarios is required for fitting). E_{NO_x} , E_{VOCs} , E_{SO_2} and E_{NH_3} correspond to emission ratios, while a_i , b_i , c_i and d_i respectively indicate the non-negative integer exponents of E_{NO_x} , E_{VOCs} , E_{SO_2} and E_{NH_3} .

Text S4 DeepRSM-PRD across the non-training times available.

To comprehensively assess the impact of temporal variations on this model, we created DeepRSM-PRD using daily maximum 8-h average (MDA8) O₃ concentrations for each day in 2019, 2020, and 2021.

Paired sample t-tests were utilized to analyze the significance of the impact of temporal scale changes on this trained model (i.e., the temporal scale generalization performance). This method helps determine if significant differences exist between results under different conditions (Dabbour et al., 2021; Zhou et al., 2021; Zhang et al., 2022a). For DeepRSM-PRD, the model trained on 2021 data was tested for its performance in non-training years (2019, 2020). These tests were conducted based on the daily NME for DeepRSM-PRD in 2019, 2020, and 2021 using IBM SPSS Statistics 26. Specifically, paired comparisons were made between 2019 and 2020 (Pair 1), 2019 and 2021 (Pair 2), and 2020 and 2021 (Pair 3). The significance threshold was set at $\alpha = 0.05$.

The trend of daily NME for different years is illustrated in Fig. S9, which depicts noticeable fluctuations in daily errors across 2021 and the preceding two years, with similar trends observed among them. The results of paired sample t-tests in Table S8 reveal the range of differences within a 95% confidence interval. It is noticeable that the p-values for all pairwise year comparisons exceed 0.05. This indicates that the differences in model performance error levels between 2019, 2020, and 2021 are not statistically significant. Thus, within the 95% confidence interval, the impact of temporal scale changes on the model is insignificant.

This underscores the model across the non-training times available, attributed to the information-capturing capabilities of deep learning neural networks (Cabaneros et al., 2019; Janarthanan et al., 2021; Zang et al., 2021). The extensive training data, which includes environmental concentrations of chemical indicators, provides rich O₃ feature information about the PRD, enhancing the model's ability to fit the non-linear relationships among O₃ precursors in the region (Mao et al., 2022b). Additionally, previous research suggests that key factors influencing O₃ concentrations, such as emissions and meteorology, are effectively incorporated into the neural network through variations in chemical indicator concentrations under baseline and clean conditions (Xing et al., 2020). Importantly, the non-linear relationships of O₃ with precursor species in DeepRSM are quantified through specific indicator species in the CTM, independent of time scale (Womack et al., 2019; Xing et al., 2020). These features enhance the performance of DeepRSM-PRD, providing it with cross-temporal transfer capability.

Text S5 The peak ratio (PR) of O₃.

$$PR = 1 + E_{NO_x} \left| \frac{\partial \Delta Conc O_3}{\partial \Delta E_{NO_x}} \right|_{=0} \quad (2)$$

Where $\frac{\partial \Delta Conc O_3}{\partial E_{NO_x}}$ is the first derivation of the O₃ response to the NO_x emission rate.

Text S6 Baseline Simulations.

The fusion result of the baseline scenario with monitoring data is depicted in Fig. S11, presenting the spatial distribution of MDA8-90% O₃ concentration in the PRD from 2019 to 2021. High O₃ concentration areas consistently appeared in the core region of the PRD during these years, with severe pollution notably prevalent in 2019. Specifically, the mean concentration in 2019 was higher than in the subsequent years, a trend aligned with government-reported data on MDA8-90% O₃ concentration (Fig. S1). This trend likely resulted from the combined effects of emissions and meteorological factors. From 2019 to 2021, pollutant control measures gradually reduced NO_x and VOC emissions (Li et al., 2022b). The COVID-19 outbreak in early 2020 led to a virtual halt in human activities. NO_x and VOC emissions in the PRD decreased by 57% and 19% during this period, respectively, causing a significant fluctuation in O₃ precursor emissions in 2020 compared to the previous year (Pei et al., 2022). Additionally, persistent and intensified anomalies in the western Pacific subtropical high-pressure system were noted in 2019 and 2021 (Hu et al., 2023a; Liu et al., 2023). These unfavorable meteorological conditions and changes in anthropogenic emissions collectively influenced the fluctuation trend in O₃ concentrations from 2019 to 2021.

Text S7 Formulas of evaluation metrics.

Correlation coefficient (R):

$$R = \frac{\sum_{i=1}^N (S_i - \bar{S}) (O_i - \bar{O})}{\sqrt{\sum_{i=1}^N (S_i - \bar{S})^2} \sqrt{\sum_{i=1}^N (O_i - \bar{O})^2}}$$

Normalized mean bias (NMB):

$$NMB = \frac{\sum_{i=1}^N (S_i - O_i)}{\sum_{i=1}^N O_i}$$

Where N denotes the number of hours in a year; S_i and O_i are WRF-CMAQ simulated and observed values of hour i , respectively; \bar{S} is mean value of all WRF-CMAQ simulated values; \bar{O} is mean value of all observed values.

Coefficient of determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^N (M_i - O_i)^2}{\sum_{i=1}^N (O_i - \frac{1}{N} \sum_{i=1}^N O_i)^2}$$

Mean absolute error (MAE):

$$MAE = \frac{1}{N} \sum_{i=1}^N |M_i - O_i|$$

Root mean squared error (RMSE):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (M_i - O_i)^2}$$

Normalized mean error (NME):

$$NME = \frac{\sum_{i=1}^N |M_i - O_i|}{\sum_{i=1}^N O_i}$$

Max normalized error (maxNE):

$$maxNE = \max \left(\frac{|M_i - O_i|}{O_i} \right)$$

Where N denotes the number of cases multiplied by the number of grid cells; M_i and O_i are the DeepRSM-simulated and CMAQ-simulated.

References

- Cabaneros S M, Calautit J K, Hughes B R (2019). A review of artificial neural network models for ambient air pollution prediction. *Environmental Modelling & Software*, 119: 285-304
- Dabbour L, Abdelhafez E, Hamdan M (2021). Effect of climatology parameters on air pollution during COVID-19 pandemic in Jordan. *Environmental Research*, 202: 111742
- Emery C, Liu Z, Russell A G, Odman M T, Yarwood G, Kumar N (2017). Recommendations on statistics and benchmarks to assess photochemical model performance. *Journal Of The Air & Waste Management Association*, 67(5): 582-598
- Gdeep (2023). 2022 report on the state of Guangdong provincial environment
- Hu W, Liu R, Chen Z, Ouyang S, Hu T, Wang Y, Cui Z, Jiang B, Chen D, Liu S C (2023). Processes conducive to high ozone formation in Pearl River Delta in the presence of Pacific tropical cyclones. *Atmospheric Environment*, 307: 119859
- Janarthanan R, Partheeban P, Somasundaram K, Navin Elamparithi P (2021). A deep learning approach for prediction of air quality index in a metropolitan city. *Sustainable Cities and Society*, 67: 102720
- Li J, Zhu Y, Kelly J T, Jang C J, Wang S, Hanna A, Xing J, Lin C J, Long S, Yu L (2019). Health benefit assessment of PM_{2.5} reduction in Pearl River Delta region of China using a model-monitor data fusion approach. *Journal of Environmental Management*, 233: 489-498
- Li Y, Wang L, Chang S, Yang Z, Luo Y, Liao C (2022). An Integrated Air Quality Improvement Path of Energy-Environment Policies in the Guangdong-Hong Kong-Macao Greater Bay Area. *Atmosphere*, 13(11): 1841
- Liu C, He C, Wang Y, He G, Liu N, Miao S, Wang H, Lu X, Fan S (2023). Characteristics and mechanism of a persistent ozone pollution event in Pearl River Delta induced by typhoon and subtropical high. *Atmospheric Environment*, 310: 119964
- Mao W, Jiao L, Wang W (2022). Long time series ozone prediction in China: A novel dynamic spatiotemporal deep learning approach. *Building and Environment*, 218
- Pei C, Yang W, Zhang Y, Song W, Xiao S, Wang J, Zhang J, Zhang T, Chen D, Wang Y, Chen Y, Wang X (2022). Decrease in ambient volatile organic compounds during the COVID-19 lockdown period in the Pearl River Delta region, south China. *Science of The Total Environment*, 823: 153720
- Womack C C, McDuffie E E, Edwards P M, Bares R, De Gouw J A, Docherty K S, Dubé W P, Fibiger D L, Franchin A, Gilman J B, Goldberger L, Lee B H, Lin J C, Long R, Middlebrook A M, Millet D B, Moravek A, Murphy J G, Quinn P K, Riedel T P, Roberts J M, Thornton J A, Valin L C, Veres P R, Whitehill A R, Wild R J, Warneke C, Yuan B, Baasandorj M, Brown S S (2019). An Odd Oxygen Framework for Wintertime Ammonium Nitrate Aerosol Pollution in Urban Areas: NO_x and VOC Control as Mitigation Strategies. *Geophysical Research Letters*, 46(9): 4971-4979
- Xing J, Zheng S, Ding D, Kelly J T, Wang S, Li S, Qin T, Ma M, Dong Z, Jang C, Zhu Y, Zheng H, Ren L, Liu T Y, Hao J (2020). Deep Learning for Prediction of the Air Quality Response to Emission Changes. *Environmental Science & Technology*, 54(14): 8589-8600
- Zang Z, Guo Y, Jiang Y, Zuo C, Li D, Shi W, Yan X (2021). Tree-based ensemble deep learning model for spatiotemporal surface ozone (O₃) prediction and interpretation. *International Journal of Applied Earth Observation and Geoinformation*, 103
- Zhang L, Li X, Chen H, Wu Z, Hu M, Yao M (2022). Haze Air Pollution Health Impacts of Breath-Borne VOCs. *Environmental Science & Technology*, 56(12): 8541-8551
- Zhou L, Zhang J, Lu T, Bao M, Deng X, Hu X (2021). Pollution patterns and their meteorological analysis all over China. *Atmospheric Environment*, 246