Contents of this file

Text S1: site classification methodology

This study uses the K-means clustering method to partition the DA sites and VE sites. The procedure is as follows. Firstly, k initial cluster centers are randomly chosen, then each observation site is assigned to the cluster that is closest to its nearest cluster center. After that, the center of each cluster is recalculated, and these two steps are repeated until the cluster centers no longer change or the maximum number of iterations is reached. The number of clusters is determined using the elbow method, which involves plotting the relationship between the number of clusters and the sum of squared errors (SSE) of the clustering results (Fig. S1a). The appropriate number of clusters is identified at the inflection point where the rate of decrease in SSE slows down (14 clusters are identified in this study). The final site division result is shown in Fig. S1b.

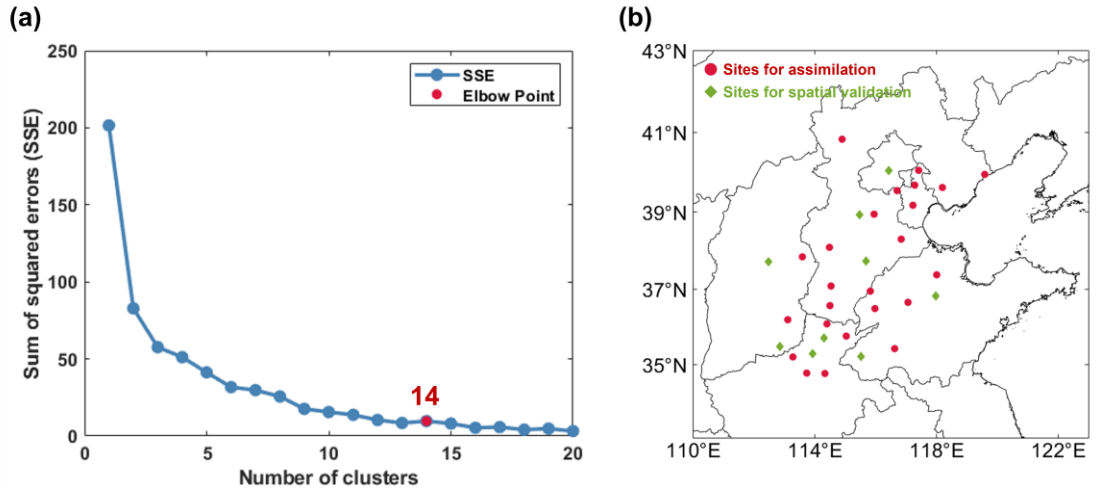

Figure S1: The relationship between the number of clusters and the sum of squared errors (a), the site division result (b), the red sites for assimilation, and the green sites for spatial independent validation.

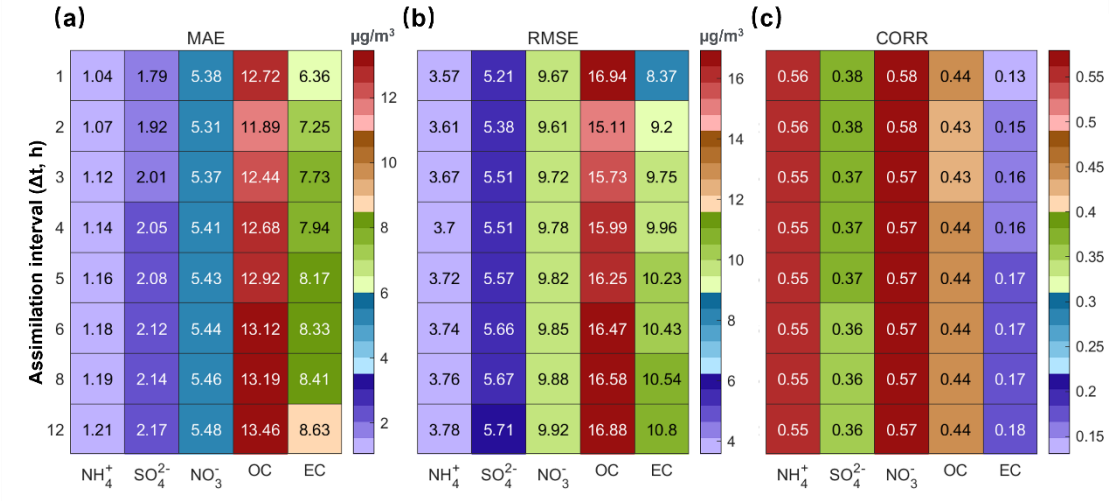

Figure S2: Assessment of assimilation interval dependency based on mean absolute error (MAE) (a), root

mean square error (RMSE) (b), correlation coefficient (CORR) (c) at forecast step.

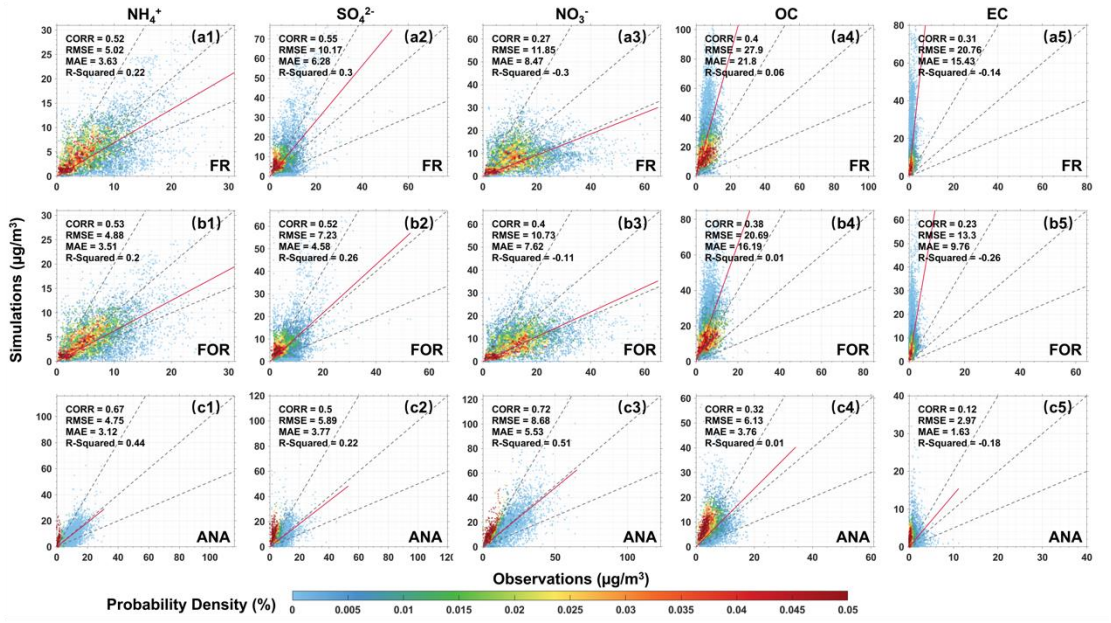

Figure S3: Scatterplots of the VE-site simulations versus the VE-site observations with probability density for the free-running field (FR, a1-a5), forecast field (FOR, b1-b5), and analysis field (ANA, c1-c5). The dotted gray lines represent the 2:1, 1:1, and 1:2 lines, and the solid red line represents the fitting regression line.

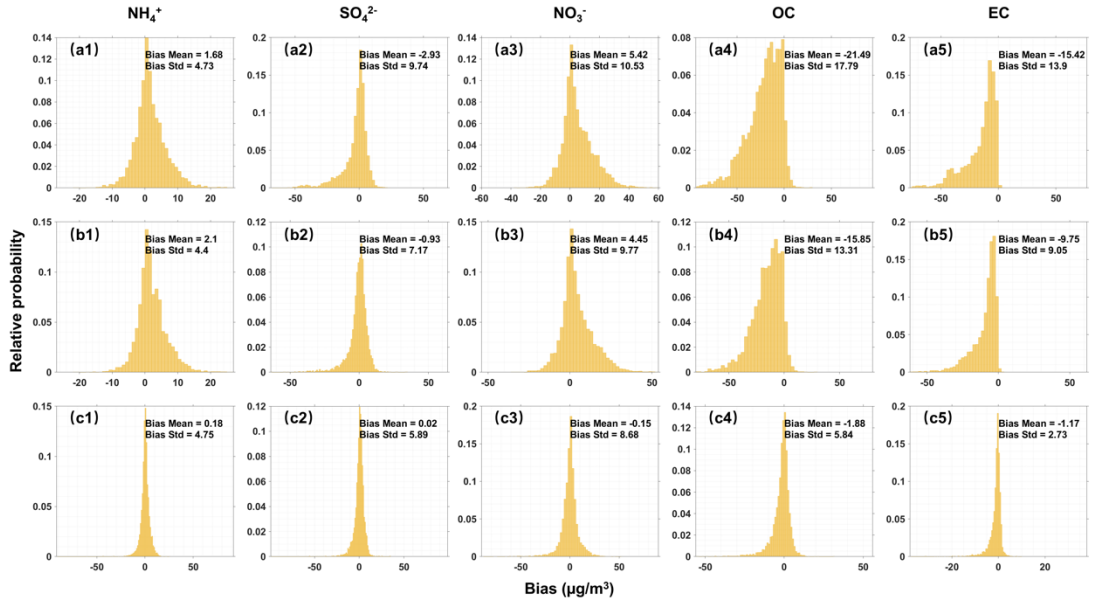

Figure S4: Probability distributions of bias between VE-site observations and VE-site simulations for the free-running field (FR, a1-a5), forecast field (FOR, b1-b5), and analysis field (ANA, c1-c5).

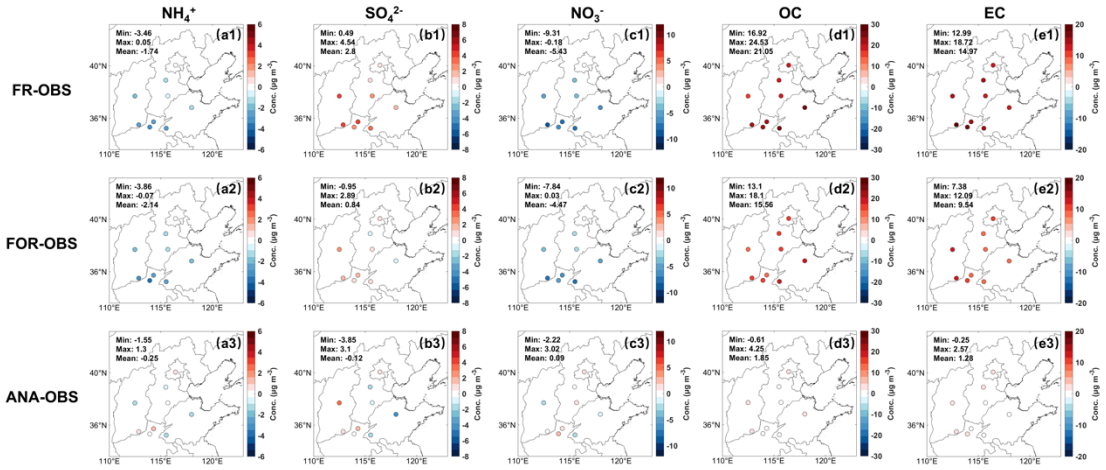

Figure S5: Spatial distribution of VE-site bias for five $PM_{2.5}$ chemical components from observation (OBS) for the free-running field (FR, a1-e1), forecast field (FOR, a2-e2) and analysis field (ANA, a3-e3).

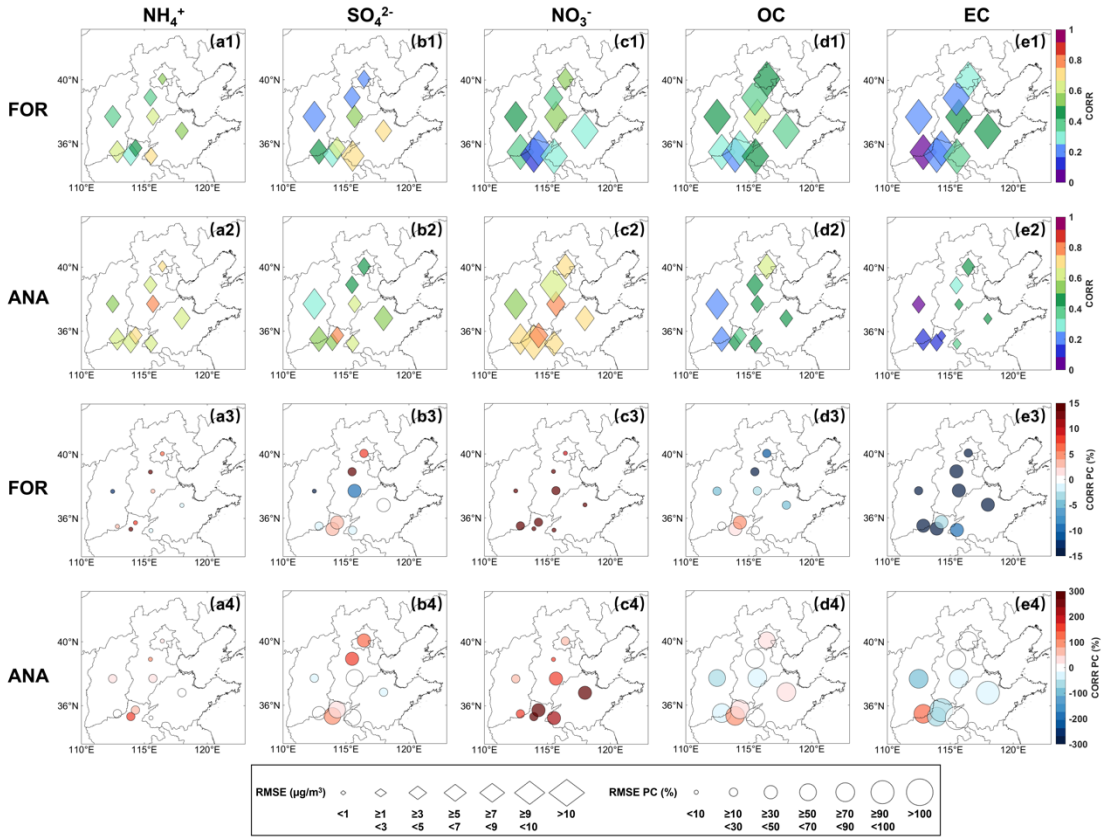

Figure S6: Spatial distribution of VE-site statistical indicators for five $PM_{2.5}$ chemical components. (a1-e1) represents the values of RMSE and CORR for the forecast field (FOR), (a2-e2) same as (a1-e1) but for analysis field (ANA), (a3-e3) represents the improvement of RMSE and CORR for the forecast field (FOR), (a4-e4) same as (a3-e3) but for analysis field (ANA). The size represents the value of RMSE in (a1-e2) and the improvement percentage compared to non-assimilation in (a3-e4), respectively.

Table S1: The physicochemical process parameterization schemes in NAQPMS.

Process	Scheme	Feature	Reference
Advection process	High precision positive	Fixed-grid size, mass	(Walcek and Aleksic,

	definite mass conserving difference scheme	conserving, stability, computational accuracy	1998)
Eddy diffusion	Gradient transport theory	High precision, local K theory scheme	(Byun and Dennis, 1995)
Dry deposition	Resistance model	Consideration of gravity deposition for aerosol	(Wesely, 1989; Slinn and Slinn, 1980)
Wet scavenging	$W_{ash} = W_a C_i$	Calculation of gas and particulate matter	\
Gas-phase chemistry	Carbon Bond Mechanism version Z (CBM-Z)	67 species and 176 reactions, lumped structure, carbon number conservation, computational efficiency	(Zaveri and Peters, 1999)
Aqueous chemistry	Regional Acid Deposition Model version 2 (RADM2)	63 species and 156 reactions, high vertical coverage range	(Chang et al., 1987; Stockwell et al., 1990)
Aerosol physicochemical process	Thermodynamic equilibrium model (ISORROPIA), Secondary Organic Aerosol Production (SOAP) model	Calculation of inorganic aerosol fractions and secondary organic aerosol, respectively	(Nenes et al., 1998; Strader et al., 1999)
Heterogeneous reactions	The first-order rate constant (k_{het})	28 reactions, simulation of the mixing of aerosols with anthropogenic gases	(Li et al., 2012)
Other processes	Radiative transfer model, online emission calculation	Simulation of the mixing of aerosols with gases, calculation of dimethyl sulfide, sea salt, and dust	(Li et al., 2011; Li et al., 2018; Lana et al., 2011; Athanasopoulou et al., 2008; Wang et al., 2000)

Table.S2. Parameterization schemes in WRF.

Physical processes	Parameterization schemes	Reference
Cloud microphysics process	Lin scheme	(Lin et al., 1983)
Longwave radiation	Rapid Radiative Transfer Model (RRTM)	(Mlawer et al., 1997)
Shortwave radiation	Goddard scheme	(Chou and Suarez, 1999)
Planetary boundary layer	Yonsei University (YSU) scheme	(Hong et al., 2006)

process		
Cumulus convection process	Kain-Fritsch scheme	(Kain and Fritsch, 1990)
Land surface processes	Noah Land Surface Model (LSM)	(Niu et al., 2011)
Surface layer scheme	MM5 similarity theory scheme	(Hari Prasad et al., 2016)

Table S3: The statistical indicators (CORR and RMSE) in the free-running field (FR), forecast field (FOR), and analysis field (ANA) at three representative DA sites and three representative VE sites.

Stations		FR		FOR		ANA	
		CORR	RMSE ($\mu g/m^3$)	CORR	RMSE ($\mu g/m^3$)	CORR	RMSE ($\mu g/m^3$)
DA	Tianjin	0.66	52.26	0.67	38.78	0.94	22.98
	Langfang	0.61	50.73	0.63	35.15	0.94	14.08
	Jiaozuo	0.48	67.86	0.52	41.19	0.91	11.45
VE	Beijing	0.48	58.53	0.46	49.10	0.75	19.57
	Zibo	0.66	73.87	0.64	51.61	0.75	19.80
	Xinxiang	0.49	66.21	0.51	41.18	0.71	22.10

Reference

Athanasopoulou, E., Tombrou, M., Pandis, S., and Russell, A.: The role of sea-salt emissions and heterogeneous chemistry in the air quality of polluted coastal areas, Atmospheric Chemistry Physics, 8, 5755-5769, 2008.

Byun, D. W. and Dennis, R.: Design artifacts in eulerian air quality models: Evaluation of the effects of layer thickness and vertical profile correction on surface ozone concentrations, Atmospheric Environment, 29, 105-126, https://doi.org/10.1016/1352-2310(94)00225-A, 1995.

Chang, J., Brost, R., Isaksen, I., Madronich, S., Middleton, P., Stockwell, W., and Walcek, C.: A three-dimensional Eulerian acid deposition model: Physical concepts and formulation, Journal of Geophysical Research: Atmospheres, 92, 14681-14700, 1987.

Chou, M.-D. and Suarez, M. J.: A solar radiation parameterization for atmospheric studies, (No. NASA/TM-1999-104606/VOL15). Greenbelt, MD: NASA Goddard Space Flight Center, 1999.

Hari Prasad, K. B. R. R., Venkata Srinivas, C., Venkateswara Naidu, C., Baskaran, R., and Venkatraman, B.: Assessment of surface layer parameterizations in ARW using micro-meteorological observations from a tropical station, Meteorological Applications, 23, 191-208, 10.1002/met.1545, 2016.

Hong, S.-Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment processes, Monthly weather review, 134, 2318-2341, 2006.

Kain, J. S. and Fritsch, J. M.: A One-Dimensional Entraining/Detraining Plume Model and Its Application in Convective Parameterization Journal of Atmospheric Sciences, 47, 2784-2802, https://doi.org/10.1175/1520-0469(1990)047<2784:AODEPM>2.0.CO;2, 1990.

Lana, A., Bell, T., Simó, R., Vallina, S., Ballabrera-Poy, J., Kettle, A., Dachs, J., Bopp, L., Saltzman, E., and Stefels, J.: An updated climatology of surface dimethlysulfide concentrations and emission fluxes in the global ocean, Global biogeochemical cycles, 25, 2011.

Li, J., Wang, Z., Zhuang, G., Luo, G., Sun, Y., and Wang, Q.: Mixing of Asian mineral dust with anthropogenic pollutants over East Asia: a model case study of a super-duststorm in March 2010, Atmospheric Chemistry and Physics, 12, 7591-7607, 2012.

Li, J., Chen, X., Wang, Z., Du, H., Yang, W., Sun, Y., Hu, B., Li, J., Wang, W., and Wang, T.: Radiative and heterogeneous chemical effects of aerosols on ozone and inorganic aerosols over East Asia, Science of the Total Environment, 622, 1327-1342, 2018.

Li, J., Wang, Z., Wang, X., Yamaji, K., Takigawa, M., Kanaya, Y., Pochanart, P., Liu, Y., Irie, H., and Hu, B.: Impacts of aerosols on summertime tropospheric photolysis frequencies and photochemistry over Central Eastern China, Atmospheric Environment, 45, 1817-1829, 2011.

Lin, Y.-L., Farley, R. D., and Orville, H. D.: Bulk Parameterization of the Snow Field in a Cloud Model Journal of Applied Meteorology and Climatology, 22, 1065-1092, https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2, 1983.

Mlawer, E. J., Taubman, S. J., Brown, P. D., Iacono, M. J., and Clough, S. A.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave, Journal of Geophysical Research: Atmospheres, 102, 16663-16682, 10.1029/97jd00237, 1997.

Nenes, A., Pandis, S. N., and Pilinis, C.: ISORROPIA: A new thermodynamic equilibrium model for multiphase multicomponent inorganic aerosols, Aquatic geochemistry, 4, 123-152, 1998.

Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., and Rosero, E. J. J. o. G. R. A.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, 116, 2011.

Slinn, S. A. and Slinn, W. G. N.: Predictions for particle deposition on natural waters, Atmospheric Environment (1967), 14, 1013-1016, https://doi.org/10.1016/0004-6981(80)90032-3, 1980.

Stockwell, W. R., Middleton, P., Chang, J. S., and Tang, X.: The second generation regional acid deposition model chemical mechanism for regional air quality modeling, Journal of Geophysical Research: Atmospheres, 95, 16343-16367, 1990.

Strader, R., Lurmann, F., and Pandis, S. N.: Evaluation of secondary organic aerosol formation in winter, Atmospheric Environment, 33, 4849-4863, 1999.

Walcek, C. J. and Aleksic, N. M.: A simple but accurate mass conservative, peak-preserving, mixing ratio bounded advection algorithm with FORTRAN code, Atmospheric Environment, 32, 3863-3880, https://doi.org/10.1016/S1352-2310(98)00099-5, 1998.

Wang, Z., Ueda, H., and Huang, M.: A deflation module for use in modeling long-range transport of yellow sand over East Asia, Journal of Geophysical Research: Atmospheres, 105, 26947-26959, 2000.

Wesely, M. L.: Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models, Atmospheric Environment (1967), 23, 1293-1304, https://doi.org/10.1016/0004-6981(89)90153-4, 1989.

Zaveri, R. A. and Peters, L. K.: A new lumped structure photochemical mechanism for large-scale

applications, Journal of Geophysical Research: Atmospheres, 104, 30387-30415, 1999.