Supporting Information

Evaluating the Impact of Aerosol Treatments in TROPOMI NO_2 Retrievals on Top-Down NO_x Emission Inversions over Eastern China

Wenhui Dong[1], Yi Wang[1,*], Yanyu Wang[2,*], Jiajun Xiong[1], Jintai Lin[3], Mengying Wang[3], Minghui Tao[1], Lunche Wang[1]

[1]Hubei Key Laboratory of Regional Ecology and Environmental Change, School of Geography and Information Engineering, China University of Geosciences, Wuhan, 430074, China

[2]State Environmental Protection Key Laboratory of Formation and Prevention of Urban Air Pollution Complex, Shanghai Academy of Environmental Sciences, Shanghai, 200233, China

[3]Laboratory for Climate and Ocean–Atmosphere Studies, Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing, 100871, China

[*]Corresponding author: Yi Wang (wangyi34@cug.edu.cn); Yanyu Wang (wangyy@saes.sh.cn)

Text S1. Validation of the observation operator

The primary purpose of the adjoint model is to provide accurate gradients of the cost function with respect to the control variables. Any error in the observation operator can be amplified through the gradient calculation, potentially affecting the convergence and physical consistency of the inversion. In this study, gradients computed by the adjoint model are verified using the finite difference method. Specifically, horizontal transport is disabled while all other model settings are held fixed. Small perturbation is applied to selected control variables, and the resulting changes in the cost function are then computed. Numerical approximation of the corresponding gradient is obtained from finite difference (Eq. S1). The finite difference gradients are subsequently compared with the gradients directly output by the adjoint model without any perturbation, in order to assess their consistency in sign, magnitude, and spatial distribution.

$$\frac{\partial J(\ln \boldsymbol{\sigma})}{\partial \ln \boldsymbol{\sigma}} \approx \frac{J(\ln(\boldsymbol{\sigma}+0.01))-J(\ln(\boldsymbol{\sigma}-0.01))}{\ln(\boldsymbol{\sigma}+0.01)-\ln(\boldsymbol{\sigma}-0.01)} \quad \text{(S1)}$$

This comparison is conducted for simulations over the fixed China nested domain at 0.25°×0.3125° resolution (NESTED_CH) and for simulations over a user-defined small domain at the same resolution (NESTED_SD). Fig S1. (a) and (b) show that both NESTED_CH and NESTED_SD exhibit extremely high linear correlations with the adjoint gradient and the finite difference gradient. The correlation coefficient (R) of the two experiments reaches 0.987 and 0.982, respectively, and the fitting slopes are both 1.01. Although the sample size (N) decreased from 36,225 to 5049 due to the reduction of the calculation domain, the statistical indicators did not show significant degradation.

From the spatial distribution, the local gradient features captured by the finite difference method, Figure S1 (c, d) are perfectly reproduced in the accompanying model results, Figure S1 (e, f). This proves the rationality of the custom nested region code and observation operator, meeting the accuracy requirements of the subsequent assimilation inversion.

Text S2. Recalculation of TROPOMI NO_2 Columns Using GEOS-Chem Vertical Profiles

Satellite NO_2 retrievals first derive a slant column density (SCD) from the measured radiance spectra and subsequently convert the SCD into a vertical column density (VCD) using an air mass factor (AMF). The AMF describes the sensitivity of the satellite measurement to NO_2 at different altitudes and therefore depends on the assumed a priori vertical distribution of NO_2. Because different retrieval algorithms and atmospheric models generally adopt or produce different NO_2 vertical profiles, the resulting AMFs can differ substantially even for the same atmospheric column.

This dependence on the a priori profile needs to be explicitly considered when comparing satellite-retrieved and model-simulated NO_2 columns. In particular, directly comparing the TROPOMI tropospheric VCD with the GEOS-Chem tropospheric VCD would cause differences in their vertical profiles to be embedded in the apparent satellite–model discrepancy. Consequently, part of the difference would reflect the inconsistent vertical-profile assumptions used in the satellite retrieval rather than an actual difference in the atmospheric NO_2 column. To ensure a consistent comparison, the vertical sensitivity of TROPOMI was therefore mapped onto the GEOS-Chem

vertical grid, and the TROPOMI NO_2 column was recalculated using the GEOS-Chem NO_2 vertical profile. This procedure minimizes the influence of a priori profile differences and allows the satellite and model columns to be compared on a consistent vertical basis. Following Qu et al. (2019), we calculated the scattering weights (SW) and the mid-level atmospheric pressure (SWP) for model grid (i,j) at vertical layers (l):

$$SW_{(i,j,l)} = AMF_{(i,j)} \times AK_{(i,j,l)} \quad \text{(S2)}$$

$$SWP_{(i,j,l)} = \frac{1}{2}\left(a1_{(l)} + a2_{(l)}\right) + \frac{1}{2}\left(b1_{(l)} + b2_{(l)}\right) \times P_{\text{surf}_{(i,j)}} \quad \text{(S3)}$$

determining the layer's center pressure in a hybrid sigma-pressure coordinate system, $a1$ and $a2$ represent pressure-dependent coefficients of the upper and lower boundary on the respective layers; $b1$ and $b2$ represent the dimensionless sigma coefficients; P_{surf} is surface pressure provided by satellite product.

Based on the pressure (P_{GC}) at layer k of the model, SW is then mapped to GEOS-Chem vertical profile using linear interpolation, yielding scattering weight aligned with the model profile (SW_{GC}):

$$\text{SW}_{\text{GC}_{(i,j,k)}} = SW_{(i,j,l)} + (SW_{(i,j,l-1)} - SW_{(i,j,l)}) \times \frac{P_{\text{GC}_{(i,j,k)}} - SWP_{(i,j,l)}}{SWP_{(i,j,l-1)} - SWP_{(i,j,l)}} \quad \text{(S4)}$$

Subsequently, the model NO_2 tropospheric VCD (VCD_{GC}) is converted to the corresponding tropospheric SCD (SCD_{GC}) using the model AMF (AMF_{GC}), enabling direct comparison with the satellite-observed tropospheric SCD.

$$\text{AMF}_{\text{GC}} = \frac{\sum_k c_{(i,j,k)} \times \Delta P_{(i,j,k)} \times \text{SW}_{\text{GC}_{(i,j,k)}}}{\sum_k c_{(i,j,k)} \times \Delta P_{(i,j,k)}} \quad \text{(S5)}$$

$$SCD_{\text{GC}} = \text{AMF}_{\text{GC}} \times VCD_{\text{GC}} \quad \text{(S6)}$$

where $c_{(i,j,k)}$ represents the NO_2 concentration in the k layer of GEOS-Chem model.

ΔP is the pressure difference between the top layer and the bottom layer.

For comparison with simulated vertical columns, the satellite tropospheric NO_2 VCD is recalculated using $\mathrm{AMF_{GC}}$:

$$\mathrm{VCD_{sat}^{GC}} = \mathrm{VCD_{sat}}\frac{\mathrm{AMF_{sat}}}{\mathrm{AMF_{GC}}} = \frac{\mathrm{SCD_{sat}}}{\mathrm{AMF_{GC}}} \quad \text{(S7)}$$

Here, $\mathrm{VCD_{sat}}$ and $\mathrm{AMF_{sat}}$ are the original satellite tropospheric VCD and AMF, respectively, and $\mathrm{SCD_{sat}} = \mathrm{VCD_{sat}} \times \mathrm{AMF_{sat}}$ is the corresponding tropospheric slant column. The superscript GC denotes a satellite VCD recalculated using the GEOS-Chem profile, rather than a simulated VCD. Together, Eqs. (S2), (S4), and (S5) recover the retrieval scattering weights, interpolate them onto the model pressure levels, and calculate an AMF weighted by the GEOS-Chem NO_2 profile. Equation (S7) then replaces the original AMF while retaining the satellite slant column (Douros et al., 2023).

Text S3. Spatial Autocorrelation Analysis Methods

To quantify the overall degree of spatial clustering in surface NO_2 concentrations across the study domain, we employed Global Moran's I statistic (Moran, 1950). The Global Moran's I statistic is defined as:

$$I = \frac{N}{W} \cdot \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\omega_{ij}(x_i-\bar{x})(x_j-\bar{x})}{\sum_{i=1}^{N}(x_i-\bar{x})^2} \quad \text{(S8)}$$

where: N is the total number of spatial units, x_i and x_j are the values of the variable at locations i and j, $\bar{x}$ is the mean of the variable, w_{ij} is the row-standardized K nearest neighbor spatial weight between locations i and j, defined as the sum of all spatial weights. where $\mathcal{N}_K(i)$ denotes the set of the K nearest grid cells

to location i, determined from the distances between grid-cell centers, and w_{ii}=0.

$$w_{ij} = \begin{cases} 1/K, & j \in \mathcal{N}_K(i), i \neq j, \\ 0, & otherwise, \end{cases} \quad \text{(S9)}$$

$\mathcal{N}_K(i)$ denotes the set of the K nearest grid cells to location i, determined from the distances between grid-cell centers, and w_{ii}=0.

The formulation evaluates the overall spatial autocorrelation by comparing the similarity between neighboring observations relative to the global variance. This index serves as a global measure of spatial autocorrelation, evaluating whether the spatial distribution of a variable is clustered, dispersed, or random. The value of Global Moran's I ranges from -1 to 1. A positive value (I > 0) indicates positive spatial autocorrelation, implying that locations with similar attribute values (either high or low) tend to cluster together spatially. In the context of this study, a high positive I value signifies that high-pollution areas are adjacent to other high-pollution areas, or clean areas are adjacent to clean areas. Conversely, a negative value (I < 0) indicates negative spatial autocorrelation, suggesting a dispersed pattern where high values are typically surrounded by low values, and vice versa. A value near zero suggests a random spatial distribution with no significant autocorrelation. The statistical significance of Global Moran's I is assessed using a permutation test to generate a pseudo-significance level (p-value).

While Global Moran's I provides a single summary statistic for the entire region, it does not identify the specific locations of spatial clusters or outliers. To address this, we utilized the Local Indicators of Spatial Association (LISA), specifically the Local Moran's I, to decompose the global statistic into the contribution of each individual

observation unit. The Local Moran's I for each spatial unit i is defined as:

$$I_i = \frac{(x_i - \bar{x})}{S^2} \sum_{j=1}^{N} \omega_{ij}(x_j - \bar{x}) \quad \text{(S10)}$$

where:

$$S^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2 \quad \text{(S11)}$$

I_i represents the local contribution of location i to the overall spatial autocorrelation, other symbols have the same meanings as in the global statistic.

The LISA analysis allows for the detection of significant local spatial patterns and classifies the spatial association at each location into four distinct cluster types based on the relationship between the value at a specific location and the weighted average of its neighbors. The four categories of local spatial association are defined as follows: (1) High-High (HH) clusters, often referred to as hotspots, represent locations with high NO_2 concentrations surrounded by neighbors with similarly high concentrations. These areas typically correspond to regional pollution centers driven by intensive anthropogenic emissions. (2) Low-Low (LL) clusters, or coldspots, represent locations with low NO_2 concentrations surrounded by low-value neighbors, indicating regionally extensive clean background areas. (3) High-Low (HL) outliers indicate locations with high concentrations surrounded by low-concentration neighbors. These represent isolated pollution sources or local anomalies where the concentration is significantly higher than the regional background. (4) Low-High (LH) outliers indicate locations with low concentrations surrounded by high-concentration neighbors, representing isolated clean spots within a polluted region.

Text S4. Getis-Ord G Statistic

While traditional spatial autocorrelation metrics (e.g., Global Moran's I) evaluate the general degree of spatial dependence, they do not distinguish between the clustering of high values (hotspots) and low values (coldspots). The Getis-Ord G statistic is specifically designed to isolate and measure the extent to which high or low values are clustered across the entire study domain. The Global Getis-Ord G is calculated based on a specified spatial distance or contiguity matrix. The formula is defined as follows:

$$G = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\omega_{ij}x_i x_j}{\sum_{i=1}^{N}\sum_{j=1}^{N}x_i x_j} \quad \text{(S12)}$$

Where, x_i and x_j represent the values at spatial grid cells i and j, respectively. N is the total number of valid spatial grid cells within the study region. ω_{ij} is the element of the spatial weight matrix defining the spatial proximity or interaction between grid i and j. In this study, a K-Nearest Neighbors spatial weight matrix was utilized to represent the effective spatial influence range of atmospheric advection and diffusion.

To evaluate the statistical significance of the computed Getis-Ord G statistic, it is standardized into a Z score. The Z score tests the null hypothesis of Complete Spatial Randomness (CSR), which assumes that the spatial distribution of NO_2 concentrations is completely random and devoid of any underlying clustering structure. The standardized Z score for the Global Getis-Ord G is computed as:

$$Z = \frac{G-\mu_G}{\sigma_G} \quad \text{(S13)}$$

Where, μ_G is the mathematical expectation of the Getis-Ord G under the null hypothesis of CSR. σ_G is the standard deviation of the Getis-Ord G under the null hypothesis.

Text S5. Mann–Whitney U test

The Mann–Whitney U test (also known as the Wilcoxon rank-sum test) is a nonparametric statistical method used to assess whether two independent samples originate from the same distribution. Unlike parametric tests such as the *t*-test, it does not assume normality of the data and is therefore particularly suitable for skewed or non-Gaussian datasets.

The test is based on the relative ranking of all observations from the two groups combined. Suppose two independent samples have sizes n_1 and n_2. After pooling and ranking all observations, the test statistic U for one sample is calculated as:

$$U = R_1 - \frac{n_1(n_1+1)}{2} \tag{S14}$$

where R_1 is the sum of the ranks assigned to observations in the first sample. An equivalent statistic can be computed for the second sample, and the smaller of the two is typically used for significance testing.

Under the null hypothesis that the two samples come from the same distribution, the mean and variance of U are given by:

$$\mu_U = \frac{n_1 n_2}{2} \tag{S15}$$

$$\sigma_U^2 = \frac{n_1 n_2 (n_1+n_2+1)}{12} \tag{S16}$$

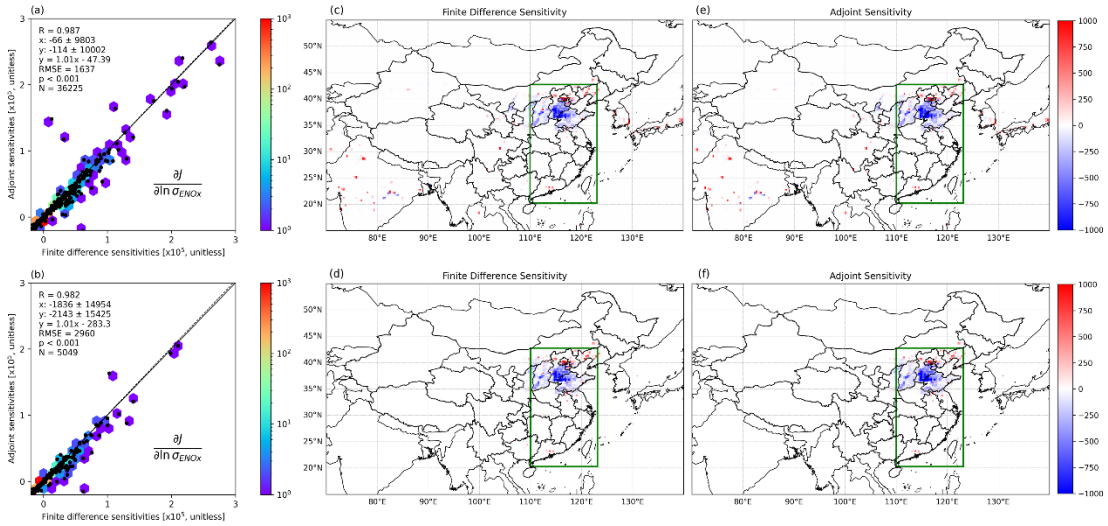
For sufficiently large sample sizes, U can be approximated by a normal distribution, allowing the computation of a standardized test statistic:

$$Z = \frac{U-\mu_U}{\sigma_U} \tag{S17}$$

p-value is then derived from Z to evaluate statistical significance. In this study, regions where the test fails to reject the null hypothesis at the 95% confidence level (i.e.,

195 $p > 0.05$) are considered statistically insignificant and are indicated by the gray overlay.

196



197

198 **Figure S1.** Finite-difference validation of adjoint gradients for the observation operator.

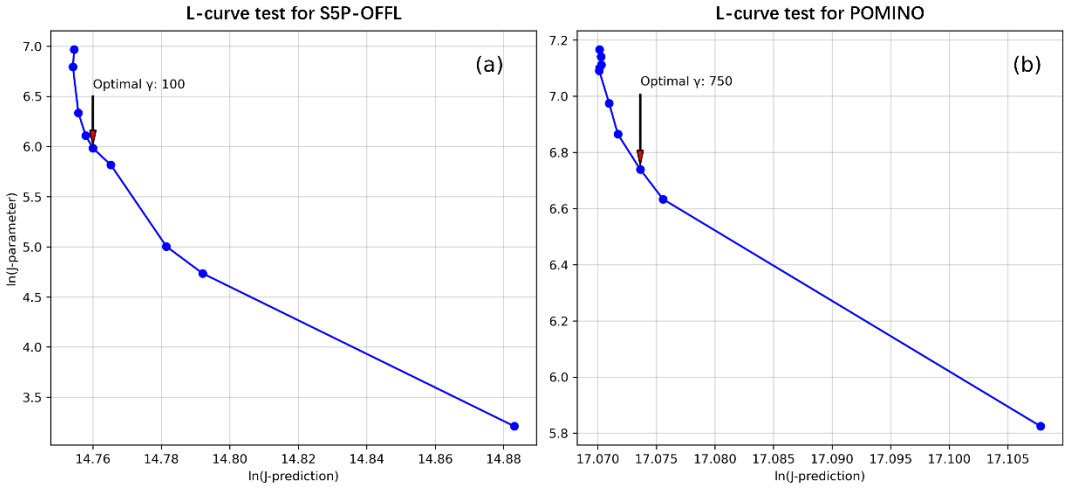
199 Panels (a, b) show scatter plots of adjoint versus finite-difference gradients for the

200 default China nested domain (NESTED_CH) and the user-defined nested domain

201 (NESTED_SD). Panels (c, d) display finite-difference gradients, and panels (e, f) show

202 gradients computed directly by the adjoint model.

203



204

205 **Figure S2.** L-curve test for S5P-OFFL and POMINO NO_2 simulation. Dependence of

206 the cost function of observation and prior on the regularization parameter γ. Selected

207 γ values in the test are: 1, 20, 50, 80, 100, 150, 500, 750, 5000.

208

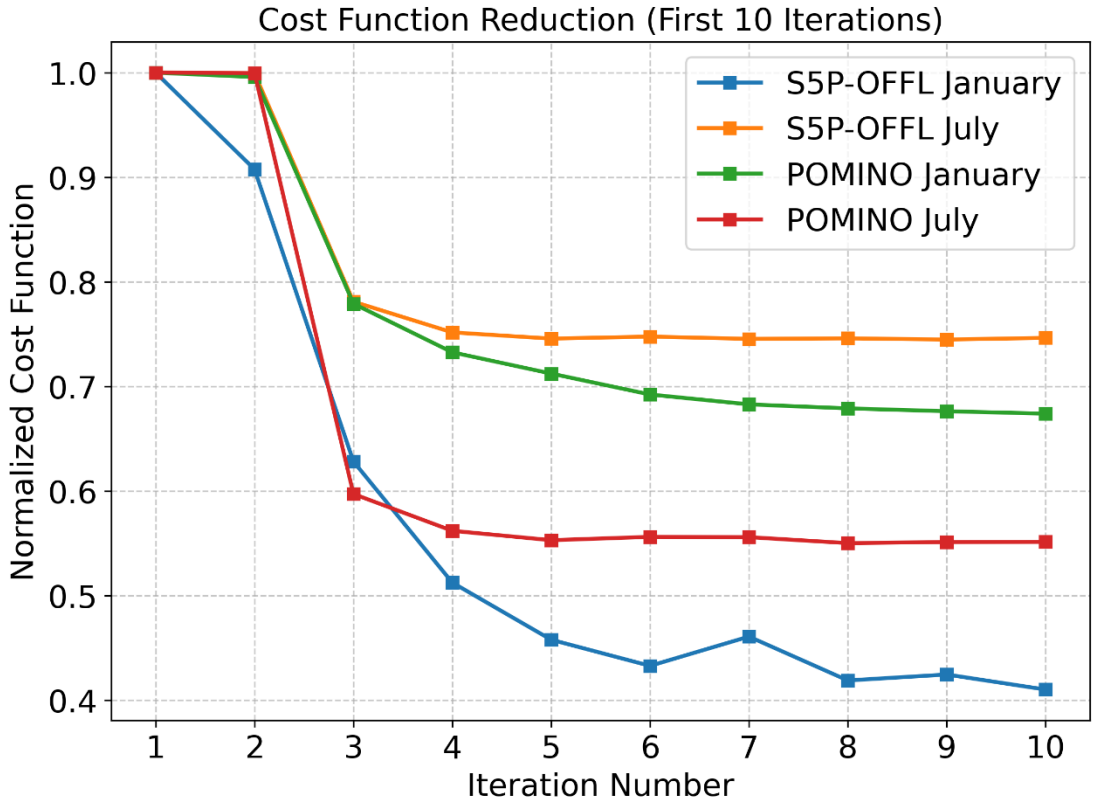


Figure S3. Trend of the cost function, including S5P-OFFL-01 (blue), S5P-OFFL-07 (orange), POMINO-01 (green), and POMINO-07 (red). As the iteration steps increase, the value of the normalized cost function gradually decreases. All simulations can achieve convergence within 10 steps.

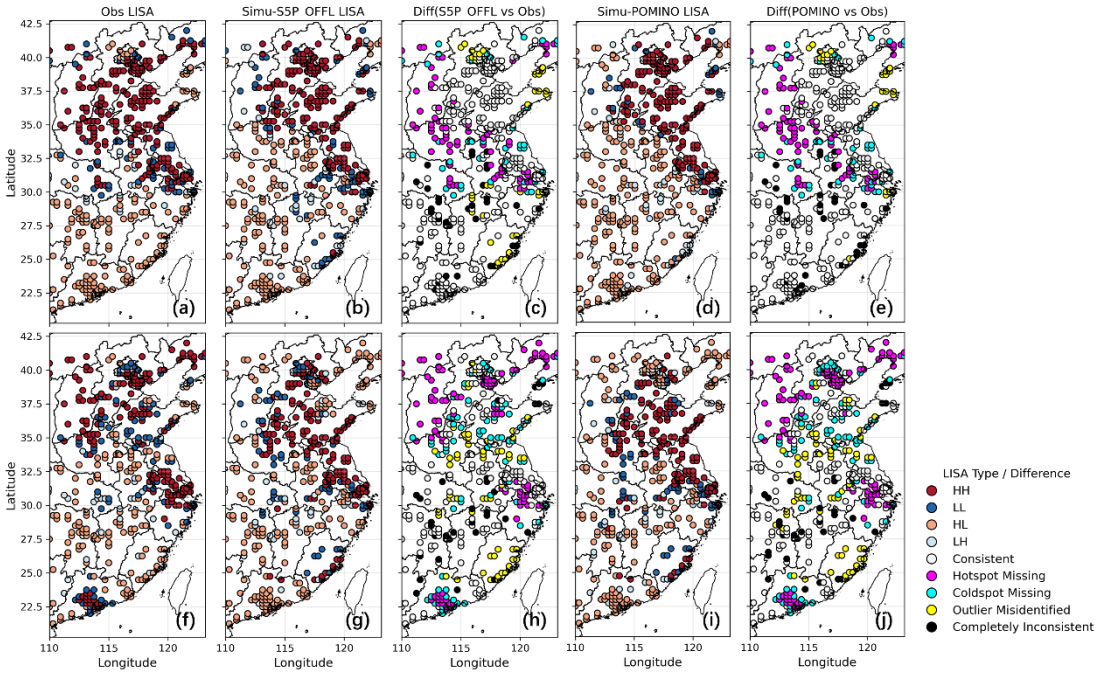


Figure S4. Spatial distribution of Local Indicators of Spatial Association (LISA) clusters and consistency comparisons between simulations and in-situ observations for surface NO_2 in January (top row, a–e) and July (bottom row, f–j) 2023. Columns from left to right display: (a, f) LISA clusters for in-situ observations; (b, g) LISA clusters for S5P-OFFL-driven posterior simulations; (c, h) consistency classification between S5P-OFFL posterior and observations; (d, i) LISA clusters for POMINO-driven posterior simulations; and (e, j) consistency classification between POMINO posterior and observations. HH (High-High) indicates spatial clusters of high values (pollution hotspots); LL (Low-Low) indicates clusters of low values (clean background); HL (High-Low) and LH (Low-High) represent spatial outliers where the value at a location is significantly different from its neighbors. Consistent implies the simulation matches the observed LISA type; Hotspot Missing indicates observed HH clusters failed to be captured by the simulation; Coldspot Missing indicates observed LL clusters missed by the simulation; Outlier Misidentified refers to mismatches involving spatial outliers

(HL/LH); and Completely Inconsistent represents a fundamental contradiction (e.g., HH vs. LL) between observed and simulated patterns.

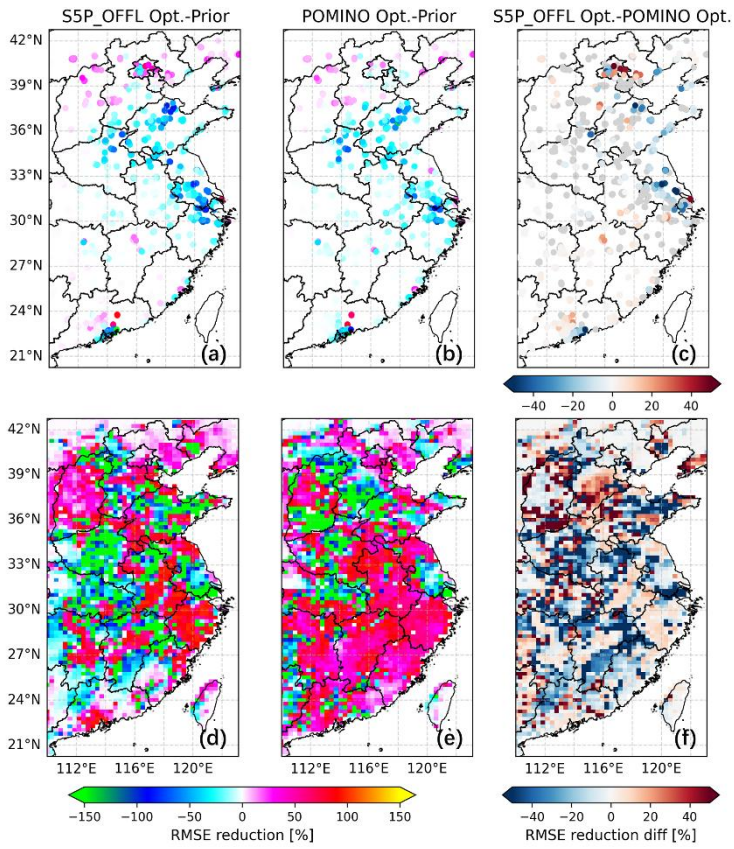


Figure S5. Regional comparison of root-mean-square error (RMSE) reduction for measurement of tropospheric NO_2 concentration against assimilated simulations over East China on January 3, 2023. RMSE reduction (%) resulting from: (a) the S5P-OFFL-driven surface NO_2 concentration relative to the model (surface NO_2) prior; (b) the POMINO-driven surface NO_2 concentration relative to the model (surface NO_2) prior; (d) the S5P-OFFL-driven NO_2 VCD data relative to the model prior of VCD NO_2; (e) the POMINO-driven NO_2 VCD data relative to the model prior of VCD NO_2. Difference in RMSE reduction between: (c) S5P-OFFL-driven surface NO_2 concentration and POMINO-driven surface NO_2 concentration; (f) S5P-OFFL-driven NO_2 VCD data and POMINO-driven NO_2 VCD data (S5P-OFFL-driven minus POMINO-driven) (%). In panels (c, f), a gray overlay indicates regions where the

differences are statistically insignificant at the 95% confidence level using the Mann-Whitney U test.

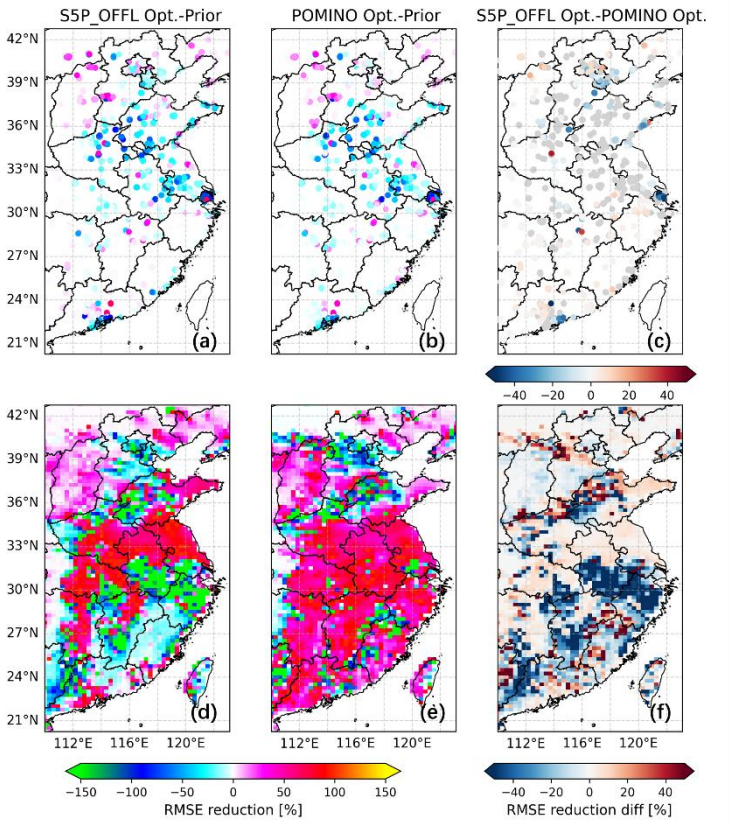


Figure S6. Similar to Figure S5, but for January 29, 2023.

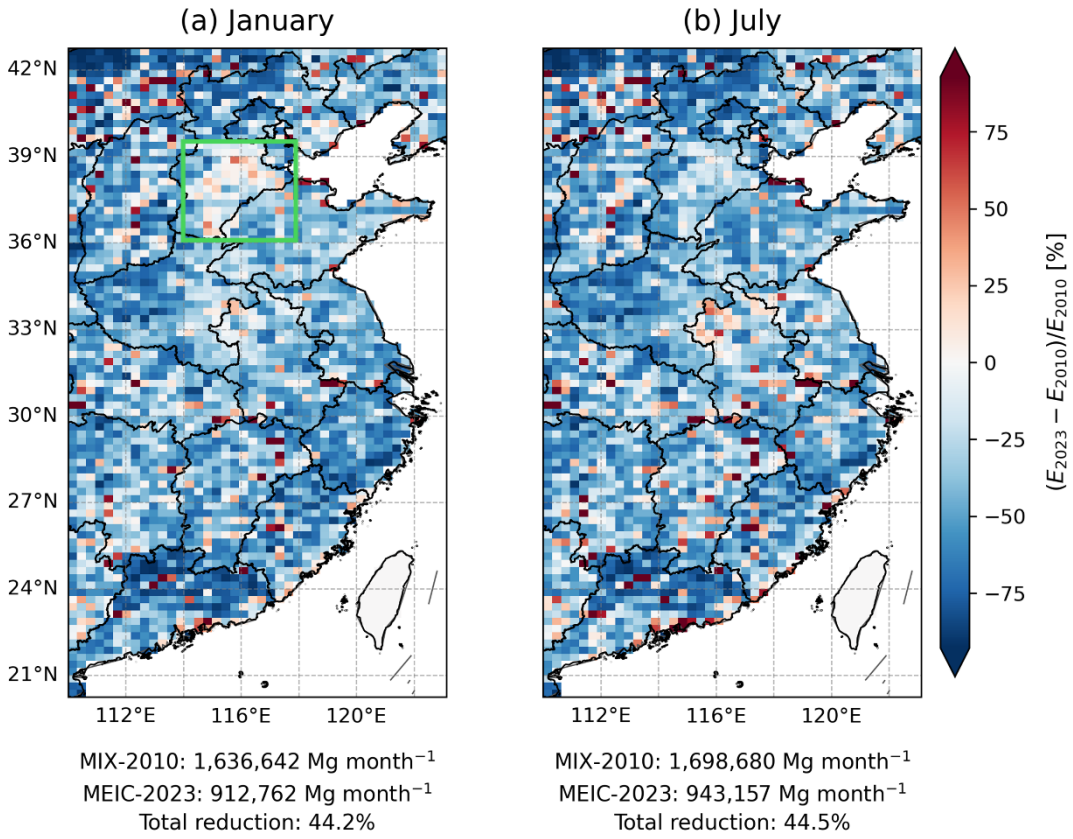


Figure S7. Spatial differences in anthropogenic NO_x emissions between the MIX-2010 and MEIC-2023 inventories for (a) January and (b) July over the inversion domain. The green rectangular box marks the southern part of Hebei.

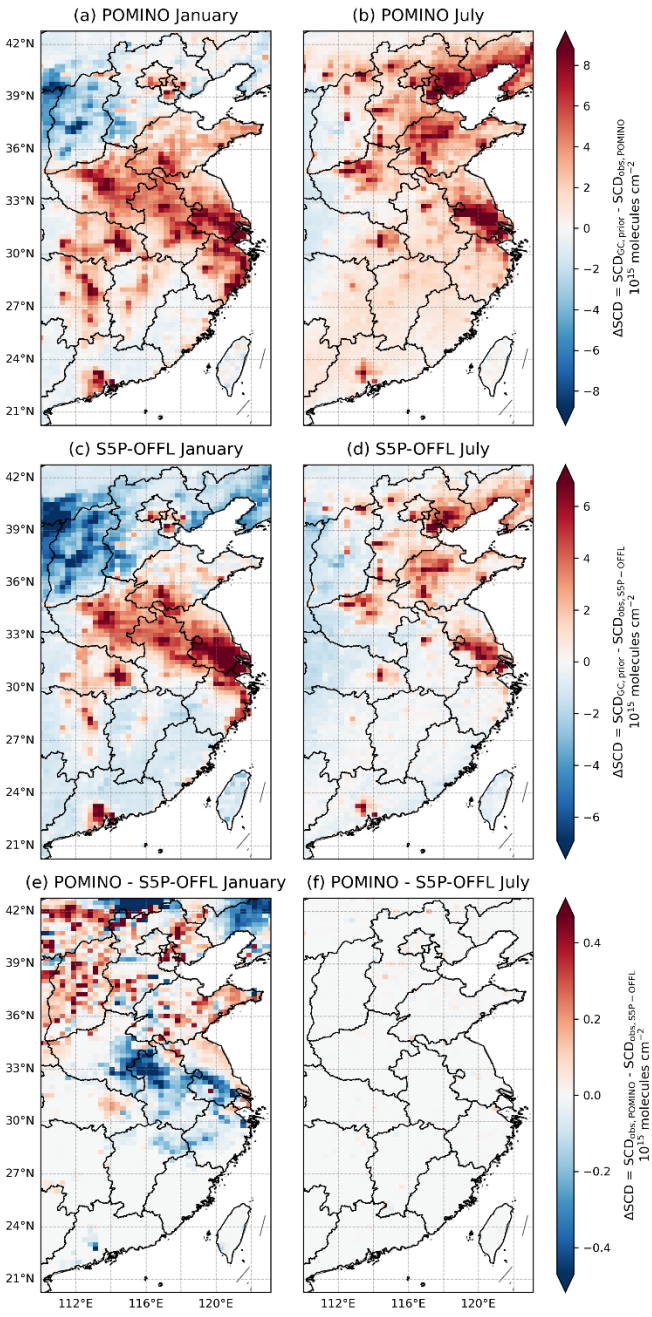


Figure S8. Spatial distributions of the mean NO_2 slant column density (SCD) differences over the inversion domain. For panels (a–d), ΔSCD is calculated as the prior GEOS-Chem SCD minus the observed SCD: (a) POMINO in January, (b) POMINO in July, (c) S5P-OFFL in January, and (d) S5P-OFFL in July. Panels (e) and (f) show the observed SCD differences between POMINO and S5P-OFFL (POMINO minus S5P-OFFL) in January and July, respectively.

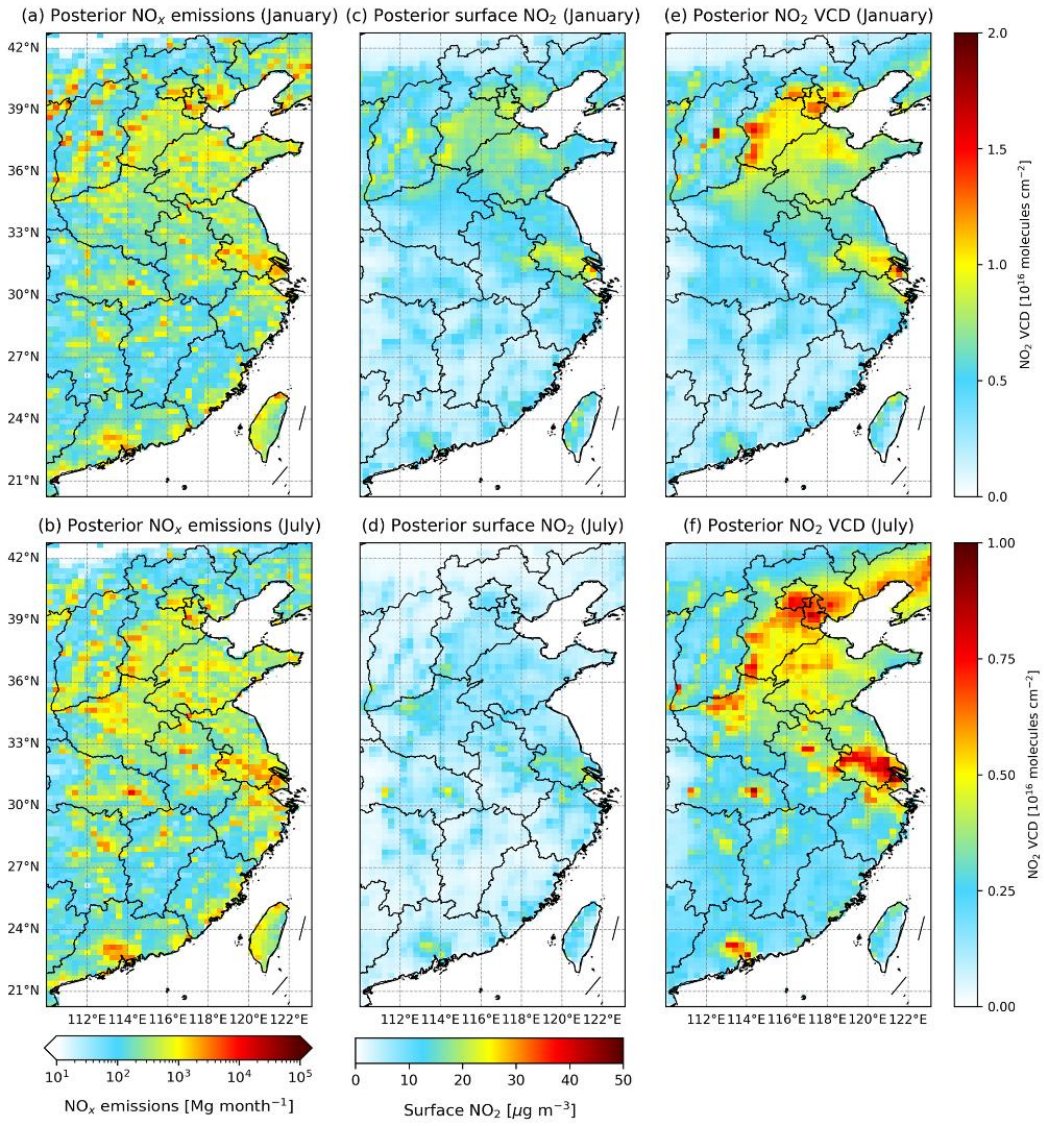


Figure S9. Posterior fields from the common-weight POMINO sensitivity experiment over Eastern China in January and July 2023. The POMINO observations and product-specific observation operator were retained, while the S5P-OFFL observation-error field and regularization parameter (γ) were applied. Panels show posterior anthropogenic NO_x emissions in (a) January and (b) July (Mg/month), posterior surface NO_2 concentrations in (c) January and (d) July ($\mu g/m^3$), and posterior tropospheric NO_2 vertical column densities in (e) January and (f) July (10^{16} molecules/cm^2).

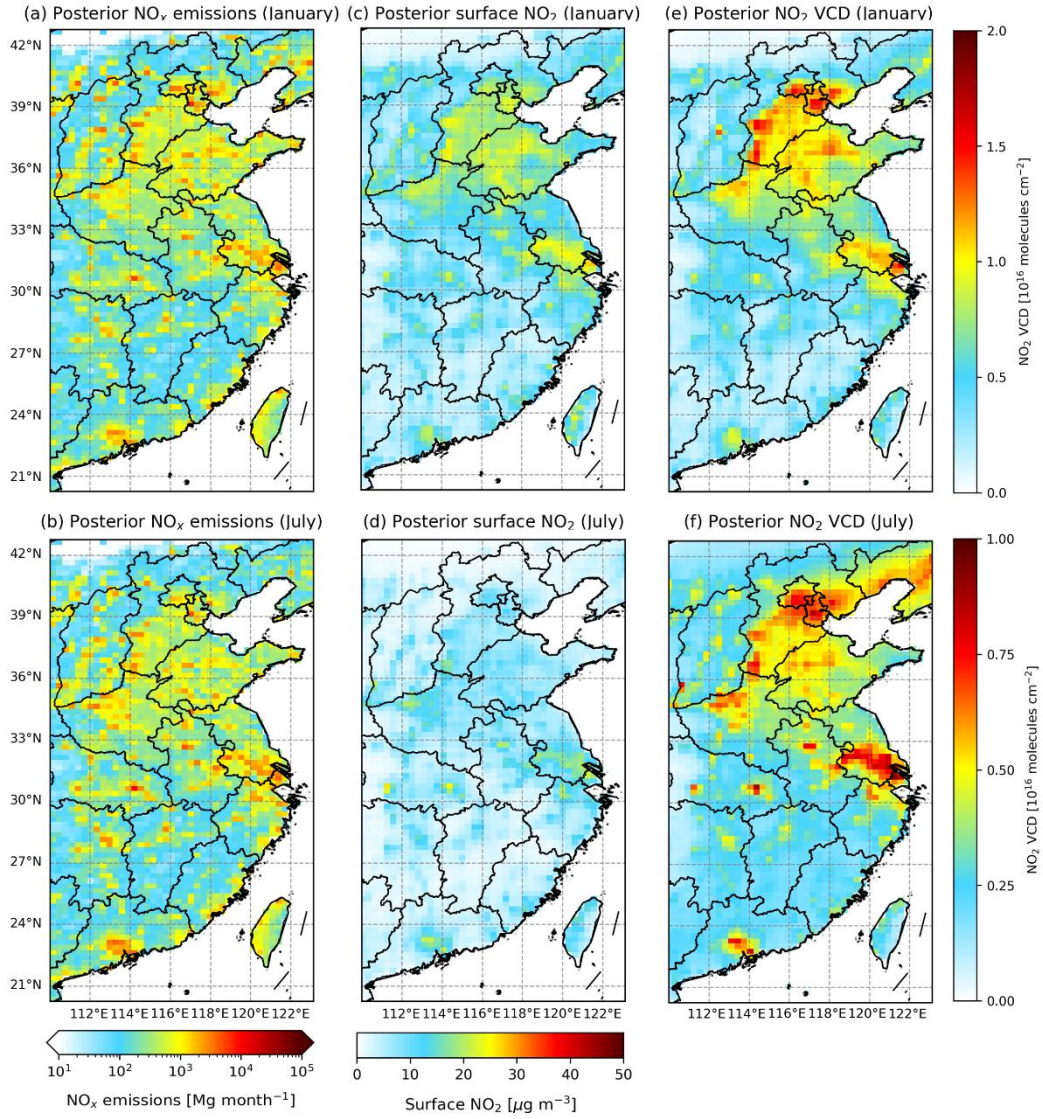
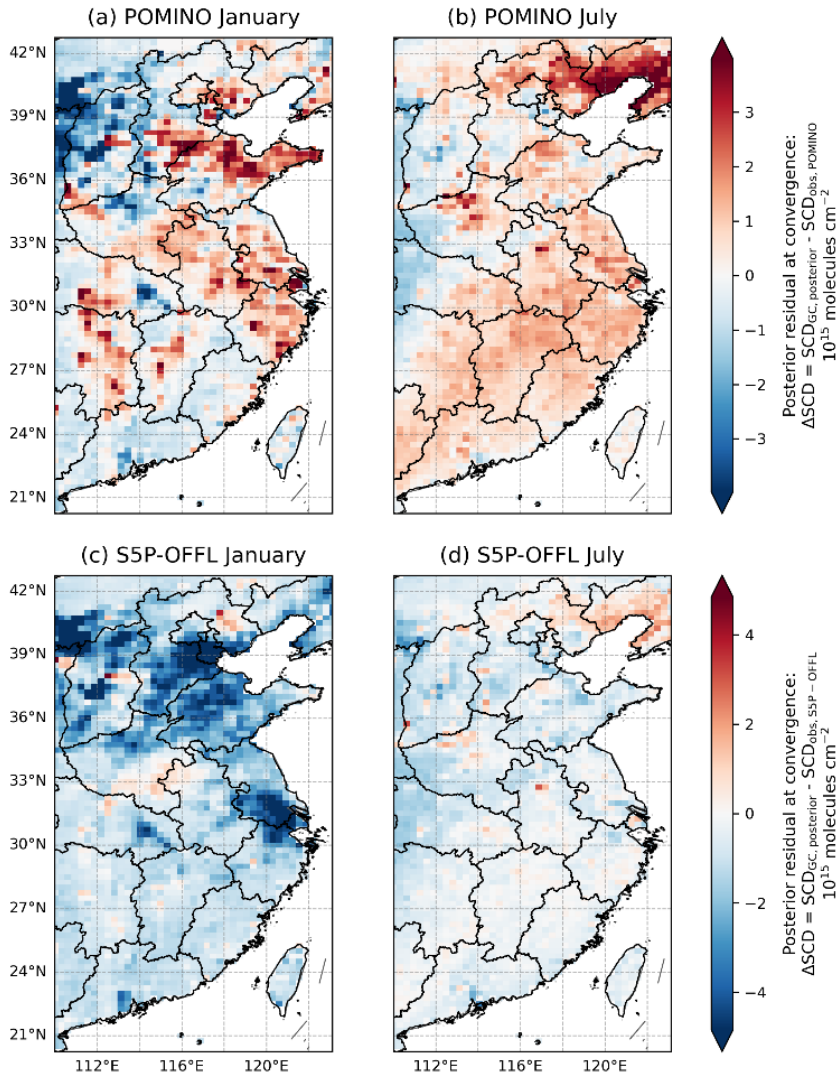


Figure S10. It is the same as Figure S9, except that gamma is scaled by the ratio of the cost functions. Specifically, gamma for January is reduced by a factor of 25, and gamma for July is reduced by a factor of 3.



277

278 **Figure S11.** Spatial distributions of the mean posterior NO_2 ΔSCD over the inversion

279 domain after convergence. ΔSCD is calculated as the posterior GEOS-Chem SCD

280 minus the observed SCD: (a) POMINO in January, (b) POMINO in July, (c) S5P-OFFL

281 in January, and (d) S5P-OFFL in July. Positive values indicate that the posterior

282 simulation exceeds the satellite observation, whereas negative values indicate that the

283 posterior simulation is lower than the observation.

284

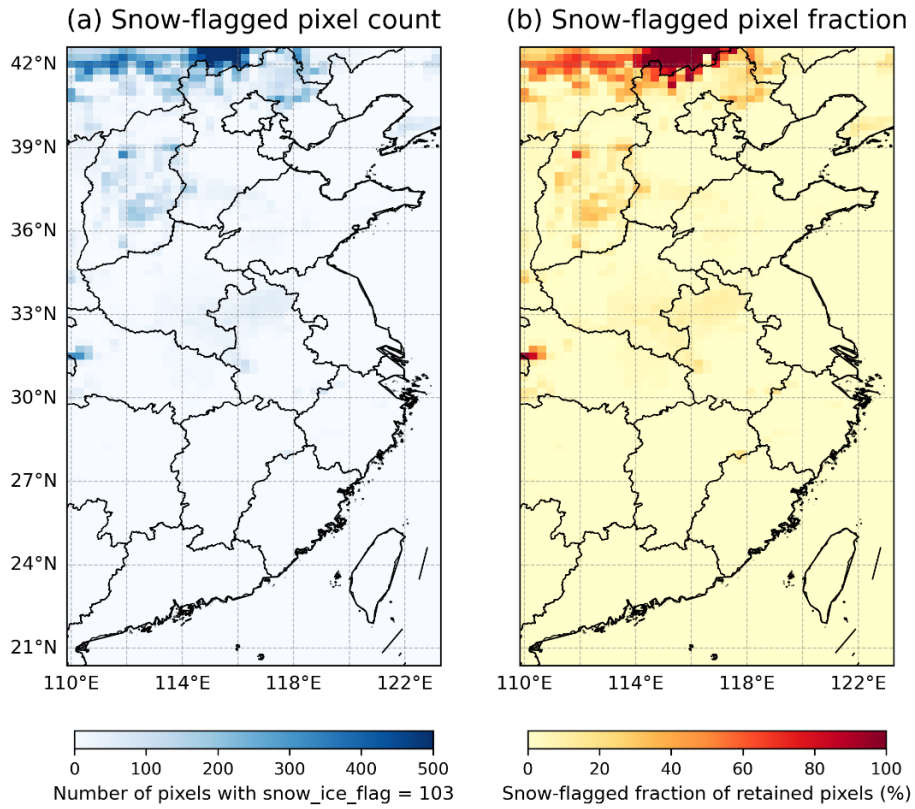


Figure S12. Spatial distributions of snow-flagged pixels in January 2023 over the inversion domain. (a) Number of retained valid pixels with snow_ice_flag = 103. High values may reflect either frequent snow-covered conditions or a larger number of valid satellite observations within a grid cell. (b) Fraction of snow-flagged pixels among all retained valid pixels used in the monthly average, calculated as $F_{\mathrm{snow}} = N_{\mathrm{snow}} / N_{\mathrm{valid}} \times 100\%$.

Douros, J., Eskes, H., van Geffen, J., Boersma, K. F., Compernolle, S., Pinardi, G., Blechschmidt, A. M., Peuch, V. H., Colette, A., & Veefkind, P. (2023). Comparing Sentinel-5P TROPOMI NO2 column observations with the CAMS regional air quality ensemble. *Geosci. Model Dev.*, *16*(2), 509-534. https://doi.org/10.5194/gmd-16-509-2023

Moran, P. A. P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, *37*(1/2), 17-23. https://doi.org/10.2307/2332142

Qu, Z., Henze, D. K., Theys, N., Wang, J., & Wang, W. (2019). Hybrid Mass Balance/4D-Var Joint Inversion of NOx and SO2 Emissions in East Asia. *J Geophys Res Atmos*, *124*(14), 8203-8224. https://doi.org/10.1029/2018JD030240