1 Text S1. Fine-tuning experiments

In this article, we adopted the fine-tuning method that is commonly used in ensemble-based atmospheric data assimilation to create a series of random emission scenarios (Jin et al., 2023). In order to fully represent the variability of emission species, the emission coefficient matrix for each emission variables was created independently. We first constructed a spatial correlation matrix L to represent the correlation between different spatial positions of emission grids.

$$S_{i,j} = \frac{D_{i,j}}{L_{\text{thres}}} \tag{1}$$

$$L_{i,j} = \begin{cases} 1 - \frac{5}{3} \times S_{i,j}^2 + \frac{5}{8} \times S_{i,j}^3 + \frac{1}{2} \times S_{i,j}^4 - \frac{1}{4} \times S_{i,j}^5, & S_{i,j} < 1 \\ -\frac{2}{3} \times S_{i,j}^{-1} + 4 - 5 \times S_{i,j} + \frac{5}{3} \times S_{i,j}^2 + \frac{5}{8} \times S_{i,j}^3 - \frac{1}{2} \times S_{i,j}^4 + \frac{1}{12} \times S_{i,j}^5, & 1 \leq S_{i,j} < 2 \\ 0, & S_{i,j} \geq 2 \end{cases} \tag{2}$$

where S refers to local support designed to guarantee non-zero values in local region (Gaspari and Cohn, 1999; Pang et al., 2023); L_{thres} is the correlation length threshold with the value of 800m; $D_{i,j}$ represents the spatial distance between two grid cells i and j. With the distance $D_{i,j}$ closes to the threshold L_{thres}, the correlation between two grid cell i and j is declining and reduces to zero later (Pang et al., 2023), indicating that the variability of emission coefficients in adjacent grids would be relatively smooth.

Additionally, we used 11 constants, namely 0.05, 0.2, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 1.8, and 1.95, as the mean value, and constructed regional spatial correlation matrix L as the standard deviation, to fit a series of multivariate normal distributions (MND) with different distributing states. The emission coefficient matrices were generated through independently repeated sampling of the corresponding MND. For instance, in the tun0.6 scenario depicted in Fig **??**, coefficient matrix for each emission variable was created from multiple sampling of the MND with 0.6 as the mean.

$$x_{i,j} \sim N(c, L) \tag{3}$$

where $x_{i,j}$ refers to the emission coefficient of each grid obtained by random sampling from the NMD; c represents the selected constant.

It should be noted that when the average value for fitting the distribution is relatively small, the final emission factors may contain negative values. Thus, all the abnormal values are uniformly set to a small value of 0.0001.

Table S1. The R^2, RMSE, and MAE values of TGEOS model for twelve indicators of $PM_{2.5}$ and O_3 within the test set. RMSE and MAE are expressed in $\mu g/m_3$.

Metrics	PM2.5_mean	PM2.5_max	PM2.5_min	PM2.5_mid	PM2.5_p25	PM2.5_p75	O3_mean	O3_max	O3_min	O3_mid	O3_p25	O3_p75
R2	0.985	0.983	0.958	0.982	0.980	0.984	0.992	0.989	0.975	0.990	0.987	0.991
RSME	2.972	4.810	4.261	3.278	3.241	3.447	1.506	4.483	0.872	1.544	1.219	2.066
MAE	1.700	2.874	2.549	1.949	1.928	2.055	0.758	2.328	0.433	0.792	0.602	1.085

Table S2. The R^2 values of the Multi-Layer Perceptron (MLP), Random Forest (RF), and TGEOS models for twelve indicators of $PM_{2.5}$ and O_3 within the test set.

Model	PM2.5_mean	PM2.5_max	PM2.5_min	PM2.5_mid	PM2.5_p25	PM2.5_p75	O3_mean	O3_max	O3_min	O3_mid	O3_p25	O3_p75
RF	0.936	0.932	0.878	0.934	0.925	0.938	0.822	0.842	0.810	0.816	0.810	0.824
MLP	0.983	0.974	0.944	0.978	0.974	0.981	0.989	0.988	0.941	0.983	0.979	0.986
TGEOS	0.985	0.983	0.958	0.982	0.980	0.984	0.992	0.989	0.975	0.990	0.987	0.991

Table S3. The MAE values of the Random Forest (RF), Multi-Layer Perceptron (MLP), and TGEOS models on twelve indicators of $PM_{2.5}$ and O_3 within the test set. MAE are expressed in $\mu g/m_3$.

Model	PM2.5_mean	PM2.5_max	PM2.5_min	PM2.5_mid	PM2.5_p25	PM2.5_p75	O3_mean	O3_max	O3_min	O3_mid	O3_p25	O3_p75
RF	3.739	5.686	4.355	3.781	3.652	4.165	3.585	9.040	1.010	3.176	2.177	4.570
MLP	2.002	3.554	3.279	2.349	2.359	2.511	0.935	2.740	0.697	1.080	0.856	1.419
TGEOS	1.700	2.874	2.549	1.949	1.928	2.055	0.758	2.328	0.433	0.792	0.602	1.085

References

Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, Quarterly Journal of the Royal Meteorological Society, 125, 723–757, 1999.

Jin, J., Fang, L., Li, B., Liao, H., Wang, Y., Han, W., Li, K., Pang, M., Wu, X., and Lin, H. X.: 4DEnVar-based inversion system for ammonia emission estimation in China through assimilating IASI ammonia retrievals, Environmental Research Letters, 18, 034 005, 2023.

Pang, M., Jin, J., Segers, A., Jiang, H., Fang, L., Lin, H. X., and Liao, H.: Dust storm forecasting through coupling LOTOS-EUROS with localized ensemble Kalman filter, Atmospheric Environment, 306, 119 831, 2023.

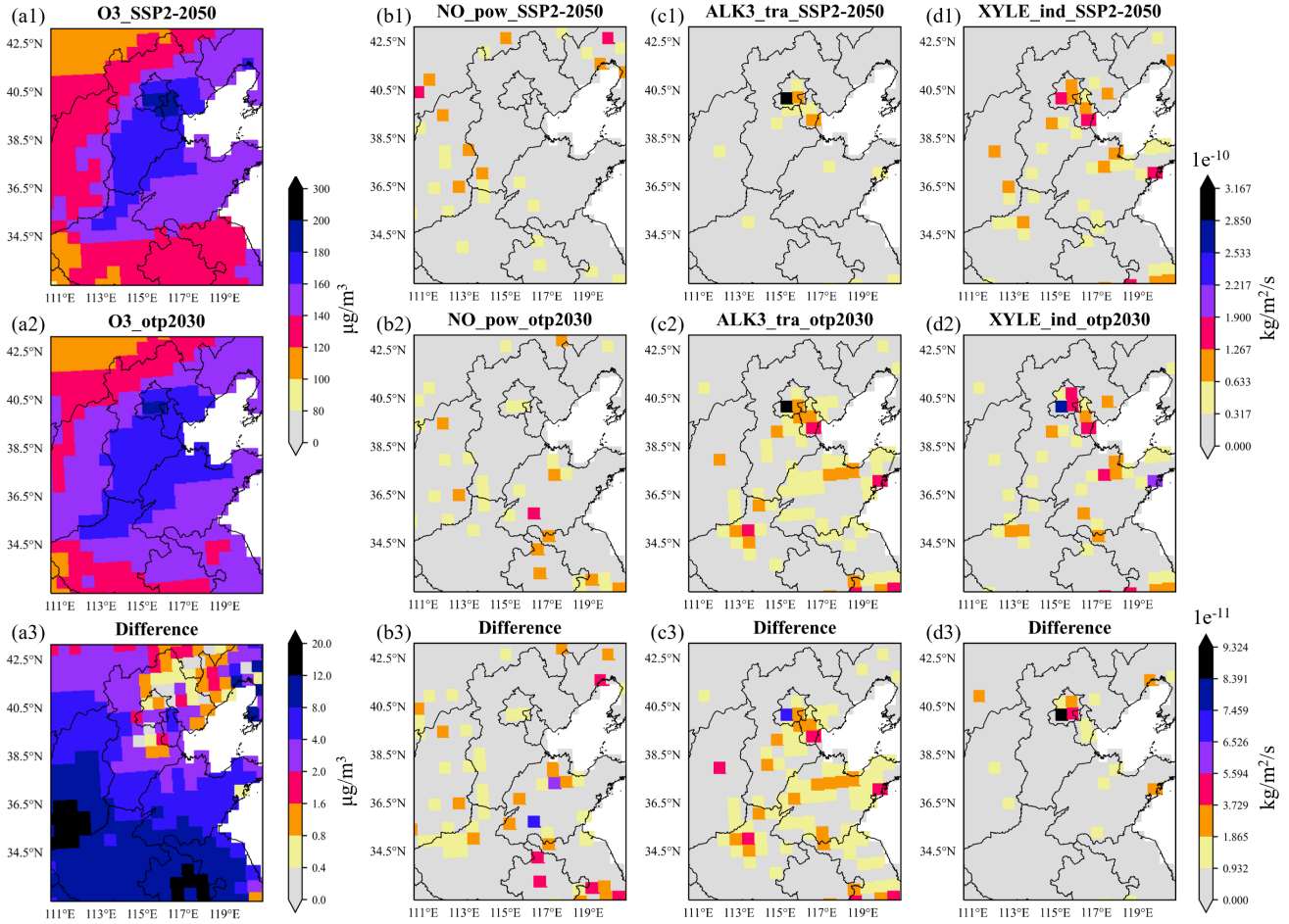

Figure S1. Spatial distributions of mean O_3 concentration and three emission variables in SSP2_2050 (a) and otp2030 (b) scenarios in July, along with the quantified absolute differences between two scenarios (c).

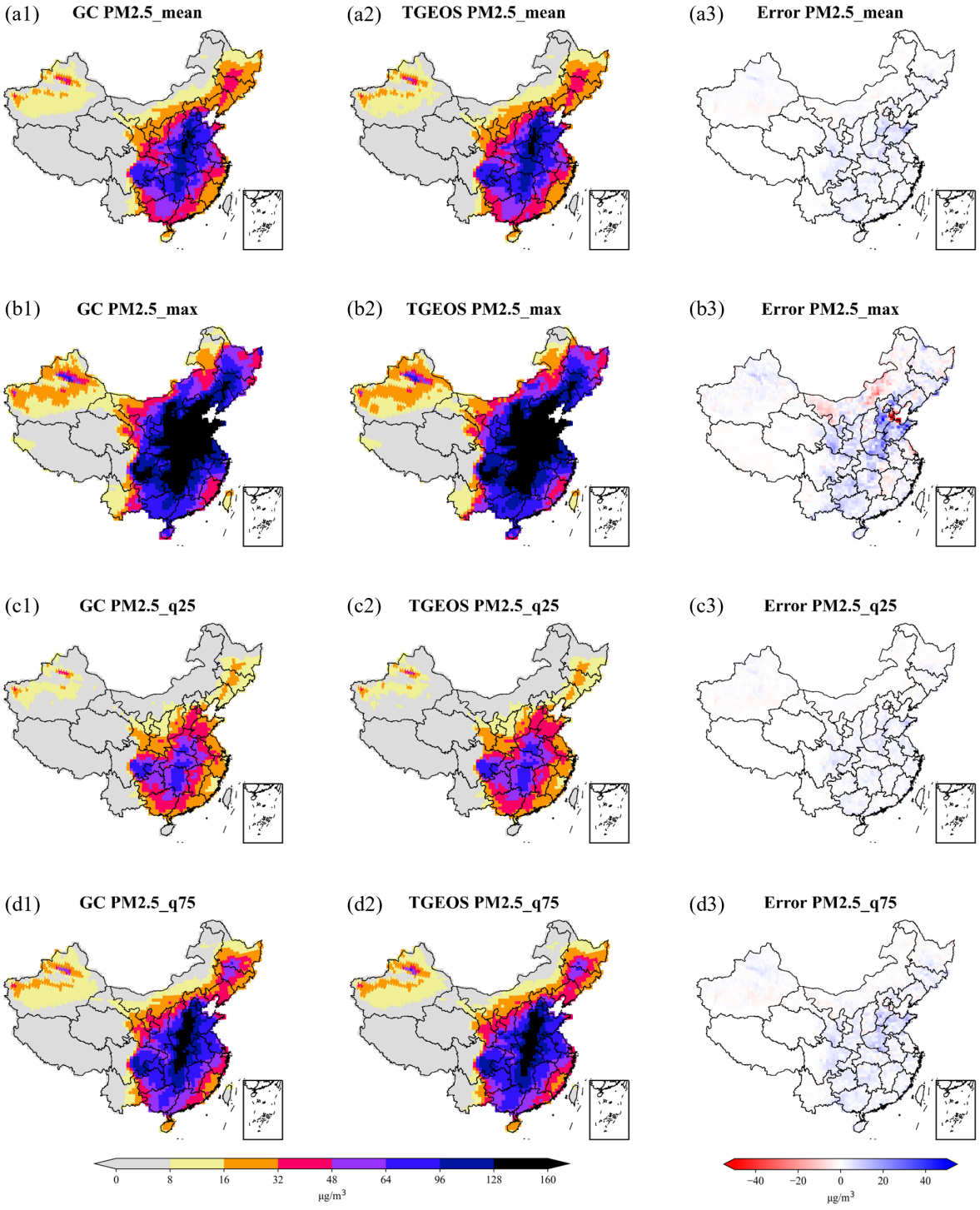

Figure S2. Spatial comparison of GEOS-Chem simulated and TGEOS estimated four statistical indicators of $PM_{2.5}$ concentrations in January under SSP3_2040 scenario (high emissions) and the corresponding error maps. (a1) to (d1) represent GC simulation while (a2) to (d2) represent TGEOS prediction, and (a3) to (d3) represent the error between this two for each indicator, including mean, maximum, 25-quantile, and 75-quantile.

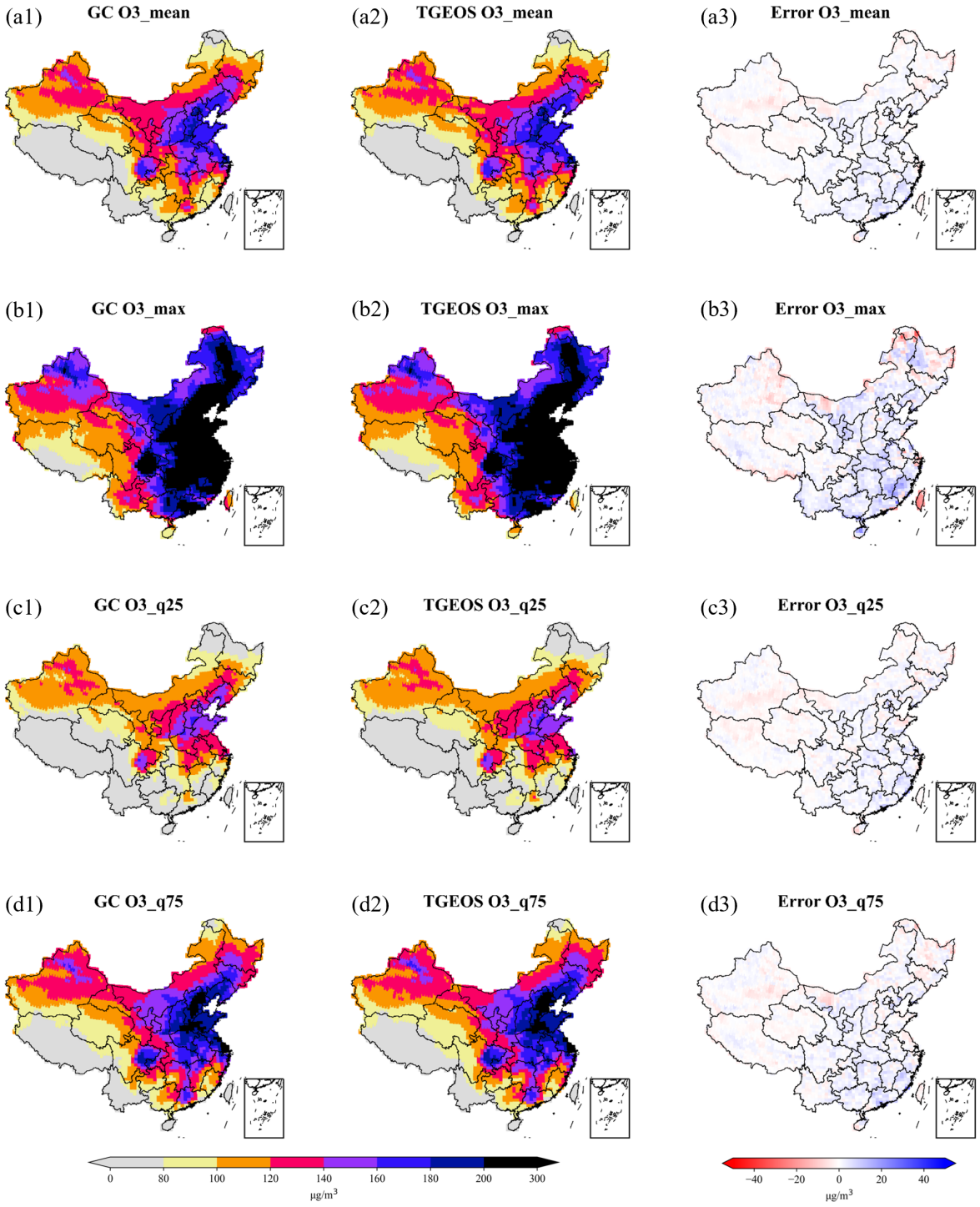

Figure S3. Same as Figure S2 but for O_3.

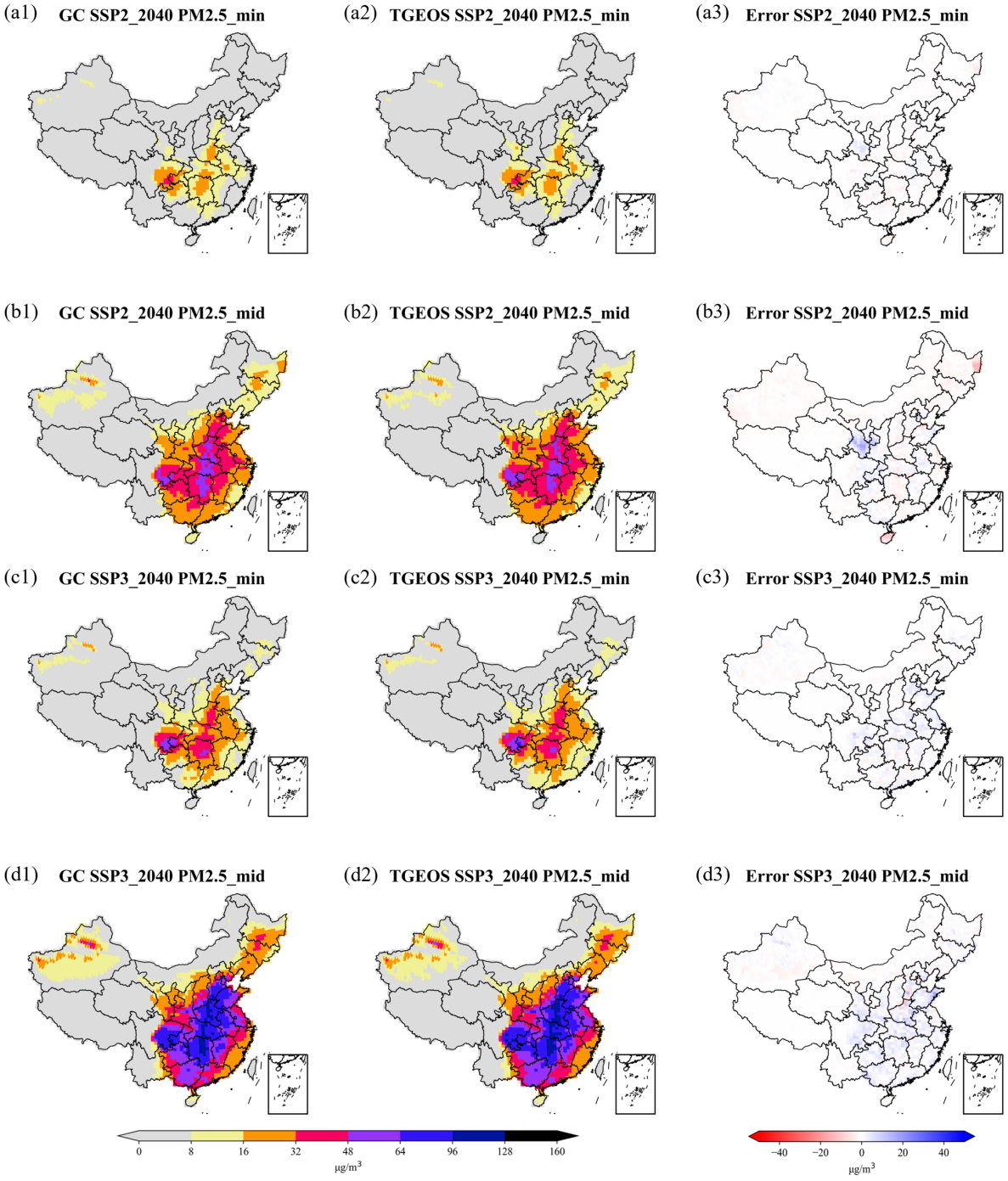

Figure S4. Spatial comparison of GC and TGEOS estimated minium and median values of $PM_{2.5}$ concentrations in January for SSP2_2040 (a1 to b3) and SSP3_2040 (c1 to d3) scenarios.

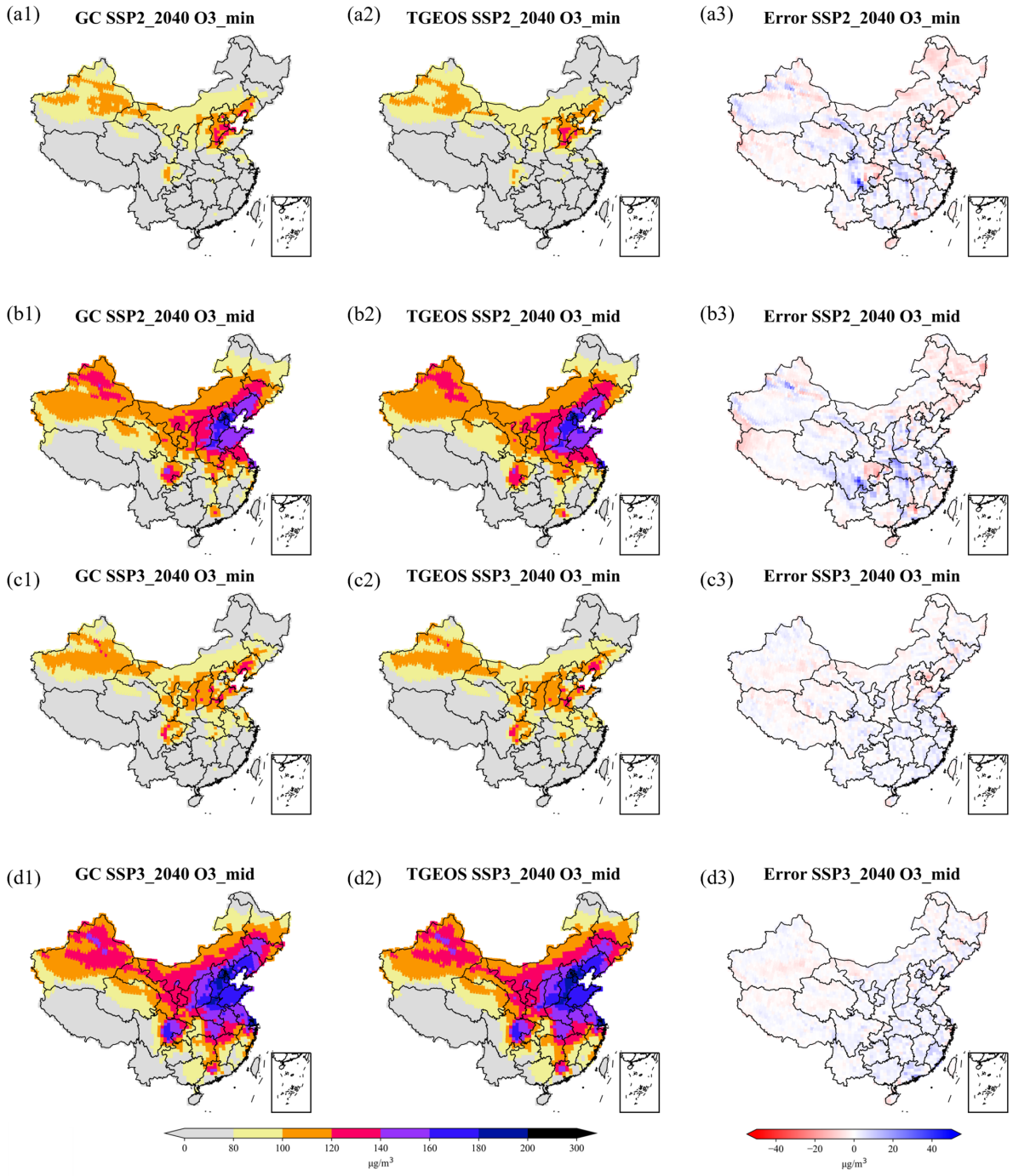

Figure S5. Same as Figure S4 but for O_3 in July.

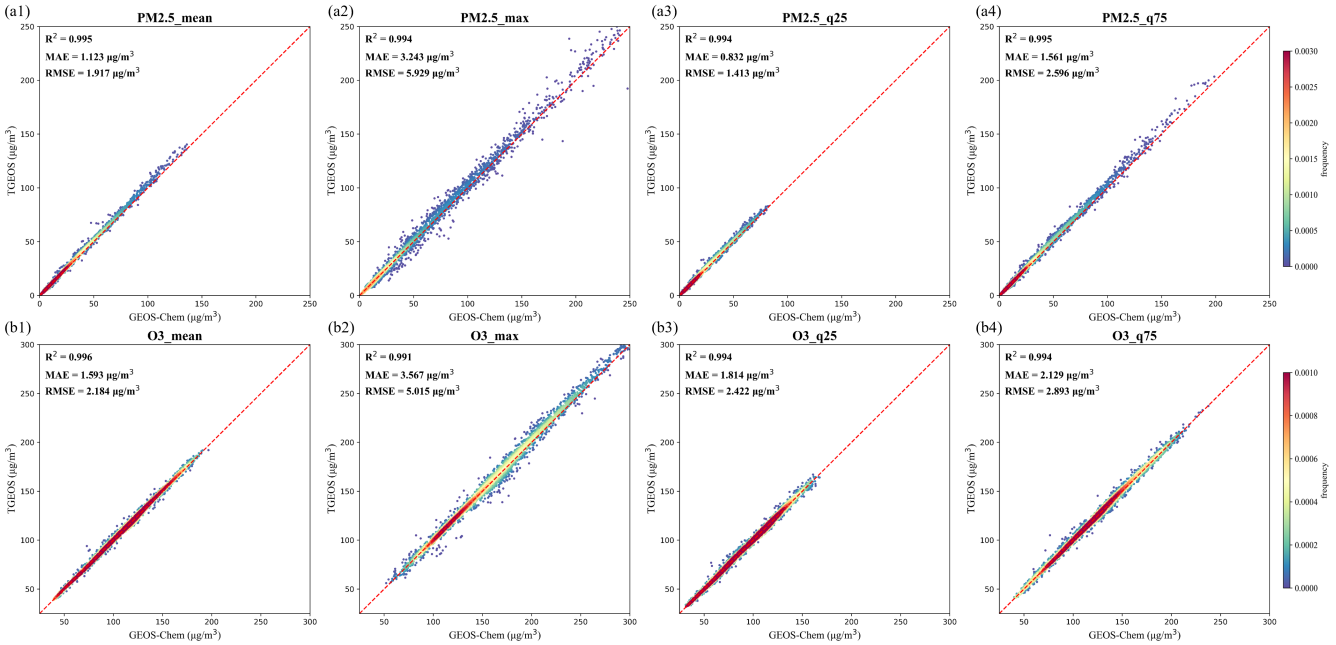

Figure S6. Density scatter plots between GEOS-Chem simulations and TGEOS predictions for eight indicators of $PM_{2.5}$ and O_3 concentrations in SSP3_2040 scenario. (a1 to a4) denotes the mean, maximum, 25th percentile, and 75th percentile of January $PM_{2.5}$ concentration; (b1 to b4) denotes the corresponding statistics for July O_3 concentration.

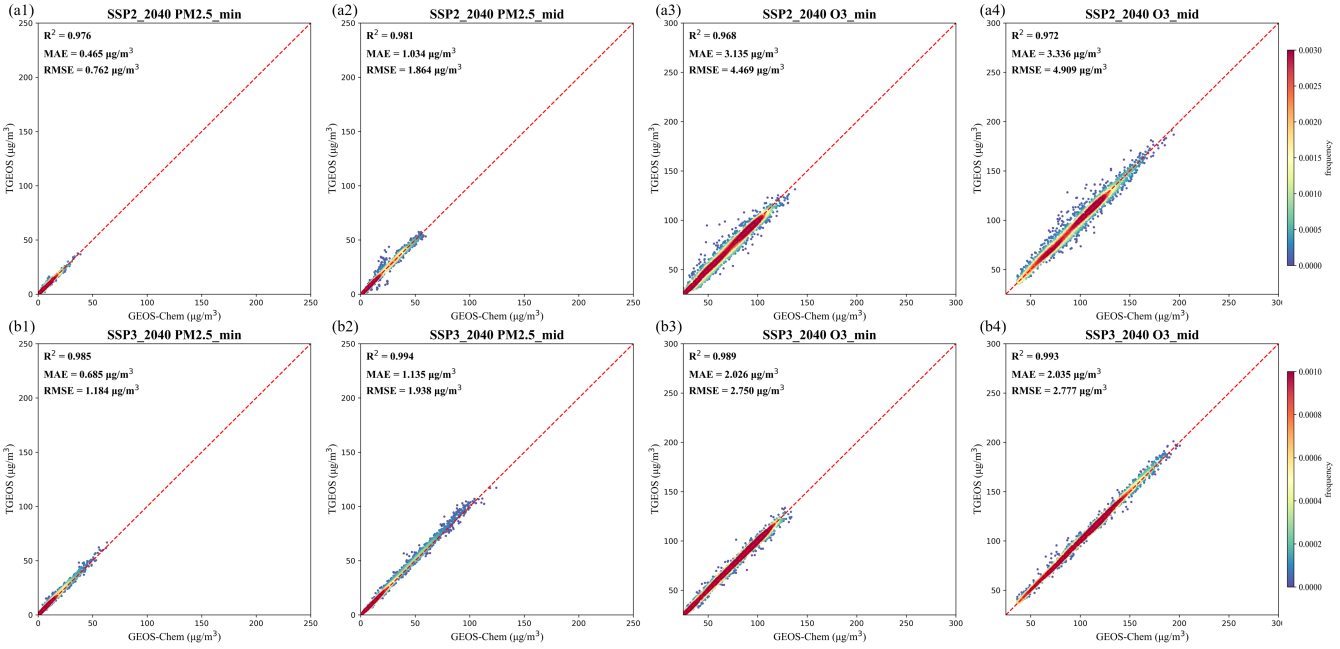

Figure S7. Density scatter plots for median and minium values of $PM_{2.5}$ and O_3 concentrations in SSP2_2040 (a) and SSP3_2040 (b) scenarios.

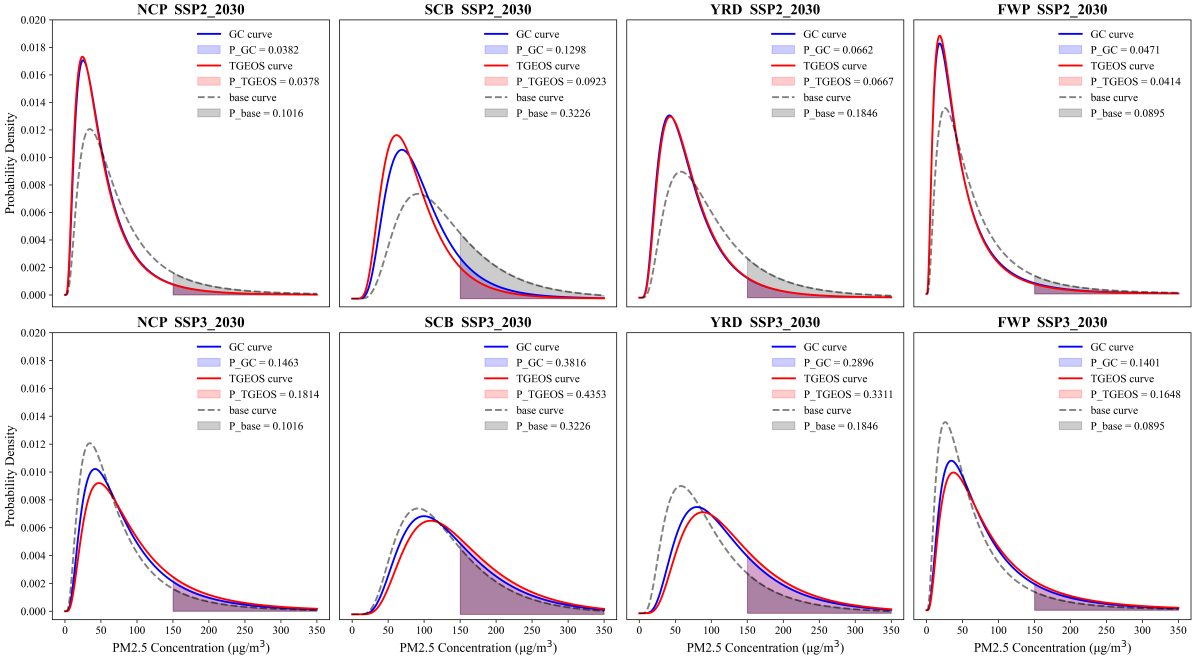

Figure S8. Probability distribution curves of winter $PM_{2.5}$ fitted by GC and TGEOS estimates under SSP2_2030 and SSP3_2030 scenarios in NCP (a), SCB (b), YRD (c), and FWP (d).

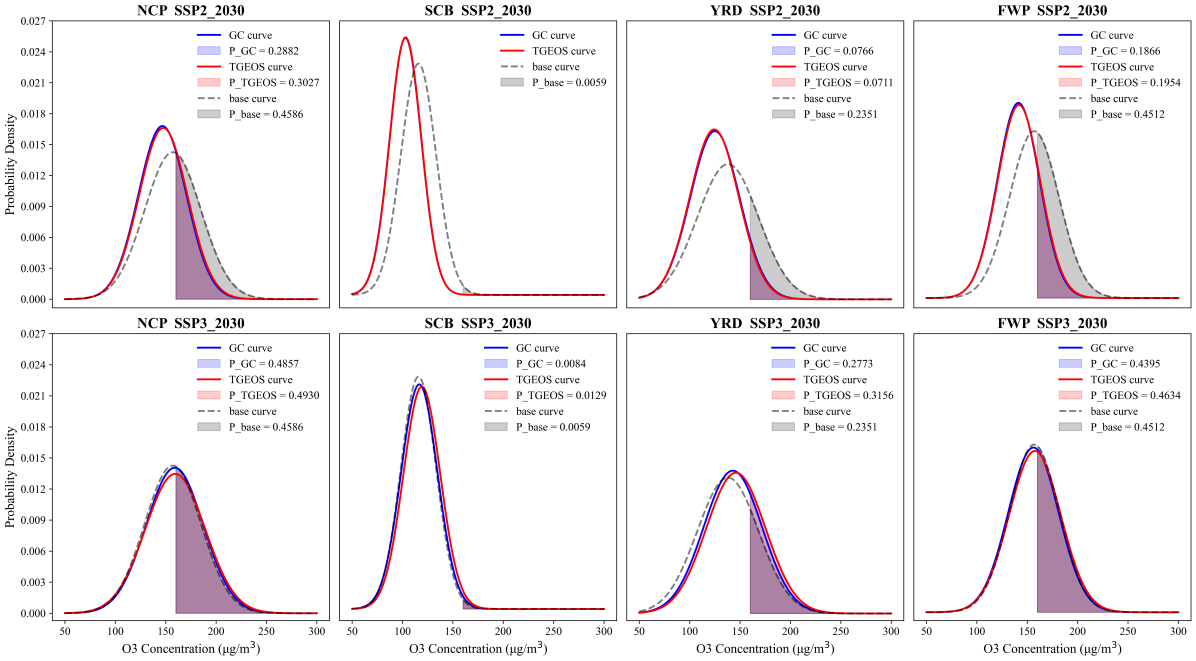

Figure S9. Same as Figure S8 but for summer O_3.

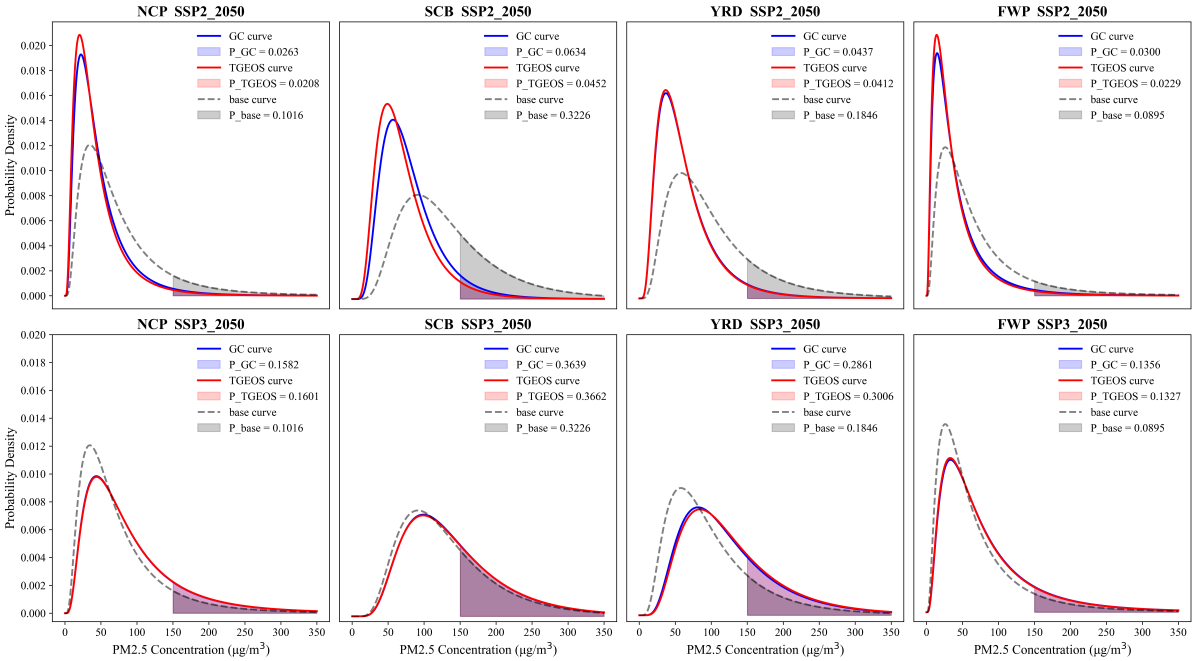

Figure S10. Same as Figure S8 but for SSP2_2050 and SSP3_2050 scenarios.

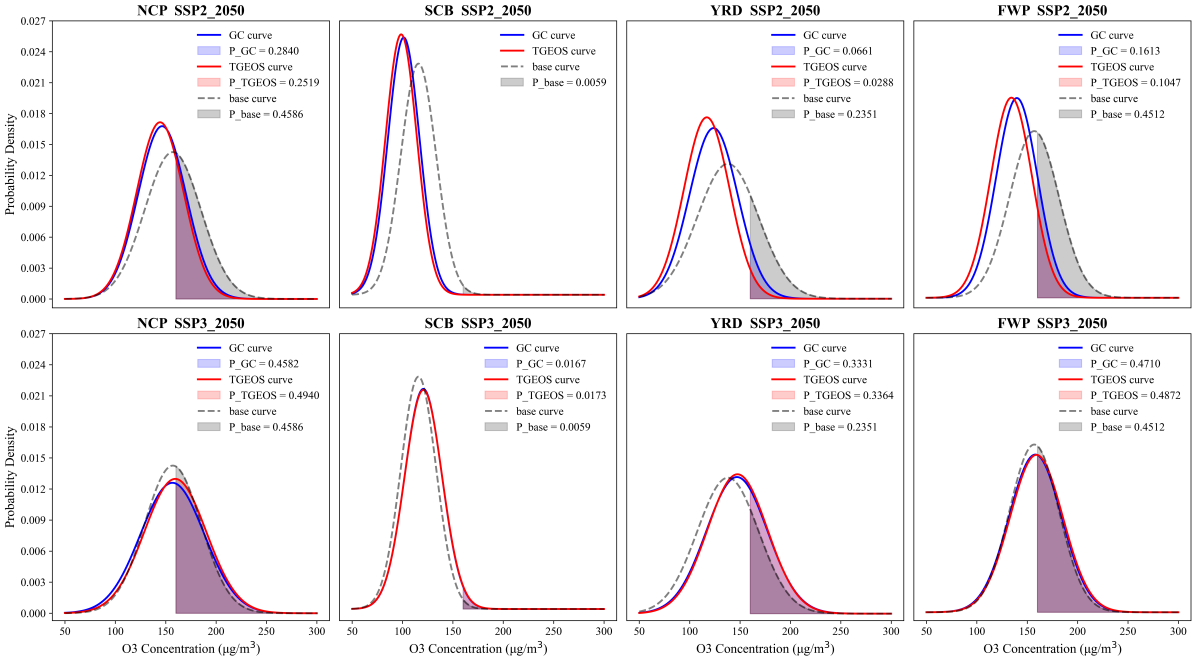

Figure S11. Same as Figure S8 but for SSP2_2050 and SSP3_2050 scenarios and summer O_3.