Responses to Reviewer 1's comments

This paper explores the performance of 14 CMIP6 models in simulating the spatial distribution, temporal variations, and components of $PM_{2.5}$ concentrations in China by comparing the models' historical run from 2000–2014 with satellite-based total $PM_{2.5}$ concentrations and ground-based $PM_{2.5}$ components data derived from the literature. It is found that $PM_{2.5}$ concentrations are generally underestimated, especially in eastern China. The concentrations of five individual components (OC, BC, sulfate, nitrate, and ammonium) are also largely underestimated. The potential causes of model biases and climate impacts on aerosol radiative forcing are also discussed. Overall, the paper is well-written, offering a thorough analysis and discussion that enhances our current understanding of the capabilities of the latest Earth system models. I have a few minor suggestions for the authors to consider.

Reply: We thank a lot the Reviewer #1 for the comments. We have studied the comments carefully and tried to incorporate as many suggested changes as possible, which have greatly helped us in improving the manuscript. Our responses to the comments and suggestions are as follows. The original comments are in green while our replies are in black.

1. Since the satellite-based $PM_{2.5}$ is one of the primary datasets used to validate CMIP6 output, it would be informative to include details about the dataset in section 2.2. This could involve specifying the aerosol species simulated by the GEOS-Chem model, discussing the accuracy of this dataset in comparison with $PM_{2.5}$ ground observations (if available), and providing information on the ground observations in China used in the dataset.

Reply: As suggested, we have added the description in Lines 175-182 and cited it here:

"The GEOS-Chem aerosol simulations include primary and secondary carbonaceous aerosols, sulfate, nitrate, ammonium, mineral dust, and sea salt. The dataset provides the annual average $PM_{2.5}$ concentrations during the period 2000–2014 with a high spatial resolution of 0.01° × 0.01° (~1 × 1 km^2). The adjusted satellite-derived $PM_{2.5}$ concentrations over Asia are compared with surface $PM_{2.5}$ observations collected from the Global Burden of Disease (GBD) collaborators during the period 2008–2013 ($Mean_{satellite}$ = 61.5 μg m^{-3} versus $Mean_{obs}$ = 59.1 μg m^{-3}) (van Donkelaar et al., 2016) and from the China National Environmental Monitoring Center (CNEMC) during the period 2015–2019 ($Mean_{satellite}$ = 45.9 μg m^{-3} versus $Mean_{obs}$ = 43.4 μg m^{-3}) (van Donkelaar et al., 2021)."

2. It may be helpful to explain why dust and sea salt concentrations are excluded from the $PM_{2.5}$ component analysis. For instance, are the ground observations available for these components?

Reply: Thanks for pointing it out. As suggested, we have added the explanation in Lines 148-150 and cited it here:

"In evaluating $PM_{2.5}$ components (Sect. 4), the evaluation of dust and sea salt concentrations is excluded due to the lack of available ground-based observations. We compare OC, BC, sulfate, nitrate, and ammonium simulations with the observed data available for these components."

3. In Section 3.2, is the positive trend in $PM_{2.5}$ concentrations in eastern China consistent with

findings from previous studies? Do the emissions also show a positive trend in the same period?

Reply: The positive trend in $PM_{2.5}$ concentrations over eastern China is consistent with findings from previous studies. The aerosol optical depth (AOD) retrieved from the MODerate resolution Imaging Spectroradiometer (MODIS) over eastern China increased from 2000 to 2013 and decreased from 2015 to 2018 (de Leeuw et al., 2022). Geng et al. (2021) developed an air pollutant database named Tracking Air Pollution in China (TAP, http://tapdata.org.cn/) using information from monitor-, satellite-, and simulation-based sources. The TAP data also captures the $PM_{2.5}$ concentrations increasing rapidly before 2006 and dropping sharply after 2013.

CMIP6 emissions over eastern China also have a positive trend in the same period (Fig. R1). We have added the explanation in Lines 244-246 and cited it here:

"The positive trend of satellite data over the eastern regions is consistent with findings from previous studies of AOD and $PM_{2.5}$ (de Leeuw et al., 2022; Geng et al., 2021), as caused mainly by emission changes (Hoesly et al., 2018; Wang et al., 2022)."

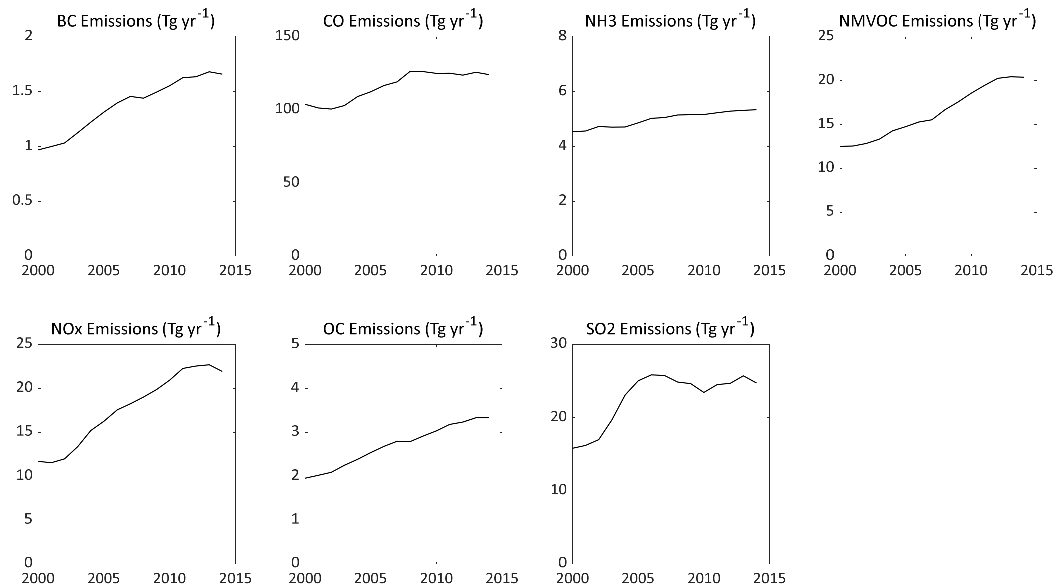

Figure R1. Trends of CMIP6 emissions for BC, CO, NH_3, NMVOC, NO_x, OC, and SO_2 over eastern China from 2000 to 2014.

4. In line 85, I'm not sure Sockol and Griswold (2017) examined $PM_{2.5}$ concentrations.

Reply: Thanks for pointing it out. We have modified the expression in Lines 83-86 and cited it here:

"Several studies have evaluated total $PM_{2.5}$ simulations of CMIP models over China, using AOD data from satellite retrievals (Sockol and Small Griswold, 2017; Michou et al., 2020) and ground-based aerosol networks (Mortier et al., 2020)."

5. In line 214, none of the correlations in Fig. 2 are greater than 0.9.

Reply: The correlation coefficient mentioned in Line 214 is the spatial correlation between simulations and satellite-based data over the eastern regions. We wanted to show that four models reproducing the spatial pattern over the eastern regions well with correlation coefficients greater than 0.9.

To make it clearer, we have added the spatial correlation coefficient values over the eastern and western regions into Table S2, respectively. We have also modified the expression in Lines 226-228 and cited it here:

"Nevertheless, the spatial pattern over the eastern regions is well simulated by four models (GFDL-ESM4, GISS-E2-1-OMA, MIROC-ES2L, and MPI-ESM-1-2-HAM) (R > 0.9, as shown in Table S2) with the maximum center over North China correctly reproduced."

Table S2. The specific values of a_1 and a_2 from Eq. 1. The average, trend, and spatial correlation coefficients of $PM_{2.5}$ concentrations over the eastern regions and western regions during 2000−2014.

	Model	a_1	a_2	Eastern regions Average ($\mu g\ m^{-3}$)	Eastern regions Trend ($\mu g\ m^{-3}\ yr^{-1}$)[a]	Eastern regions Spatial Corr.[b]	Western regions Average ($\mu g\ m^{-3}$)	Western regions Trend ($\mu g\ m^{-3}\ yr^{-1}$)	Western regions Spatial Corr.
Satellite-based				39.0	0.72	1	22.7	0.06*	1
Total $PM_{2.5}$ from Direct ESM output	GFDL-ESM4			37.7	1.14	0.92	22.1	0.28	0.66
	GISS-E2-1-OMA			24.4	0.69	0.91	10.9	0.13	0.79
	MIROC-ES2L			20.3	0.49	0.90	8.9	0.13	0.59
	MPI-ESM-1-2-HAM			36.6	0.93	0.91	22.5	0.20*	0.36
	MRI-ESM2-0			30.4	0.57	0.83	24.5	0.24	0.71
	NorESM2-LM			22.1	0.32	0.87	35.5	0.03*	0.49
	NorESM2-MM			23.6	0.40	0.90	43.1	–0.10*	0.53
Total $PM_{2.5}$ from Eq. 1	BCC-ESM1			19.5	0.40	0.87	10.2	0.15	0.62
	CESM2-WACCM	0.25	0.1	24.0	0.73	0.92	10.1	0.22	0.67
	CNRM-ESM2-1	0.02	0.25	18.9	0.42	0.90	5.5	0.11	0.51
	EC-Earth3-AerChem	0.25	0.1	21.4	0.56	0.91	8.3	0.18	0.53
	GISS-E2-1-MATRIX	0.25	0.1	17.0	0.43	0.92	6.4	0.10	0.67

HadGEM3-GC31-LL	0.27	0.35	26.5	0.80	0.89	7.9	0.18	0.50
UKESM1-0-LL	0.27	0.35	26.5	0.71	0.89	8.1	0.18	0.52

[a] Trends are estimated using the Theil-Sen Median method (Theil, 1950; Sen, 1968). Significant changes are identified using the non-parametric Mann-Kendall test (Kendall, 1938). * represents non-significant monotonous change at p = 0.05. [b] Spatial correlation coefficients between simulations and satellite-based data over the eastern and western regions are calculated. The spatial correlation coefficients of 14 models are at the 0.05 significance level.

6. In Fig. 2, what does NMB stand for? Normalized mean bias?

Reply: Thanks for pointing it out. NMB stands for normalized mean bias (i.e., NMB = ($Mean_{simulation}$ / $Mean_{observation}$ − 1) × 100%). $Mean_{simulation}$ and $Mean_{observation}$ are the spatial average of simulated and observed concentrations, respectively. We have added the definition of NMB in Lines 269-273 and Lines 764-765, and cited it here:

Lines 269-273: "The national average of the 14-model mean (6.5 μg m^{-3}, normalized mean bias (NMB) = –59.0%), which are spatially coincidently sampled with the ground-based observations (i.e., model values are obtained from grid cells with available observations), severely underestimates the observations, especially over parts of North China with the bias reaching –40 μg m^{-3} (Fig. 5 b)."

Lines 766-767: "R stands for spatial correlation, and NMB stands for normalized mean bias."

7. In Fig. 4, consider marking the area where correlations are statistically significant.

Reply: As suggested, we have marked the regions where the correlation coefficient is at the 0.05 significance level.

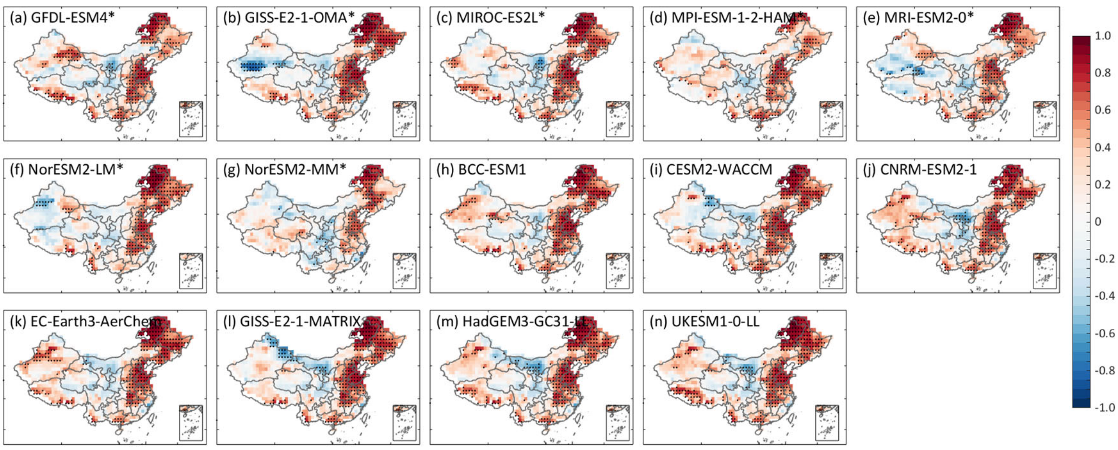

Figure 4. Spatial distribution of correlation coefficients between modeled and satellite-based data for interannual variations of annual mean total $PM_{2.5}$ concentrations during 2000–2014. Black dots indicate a significance level of 0.05.

8. In Fig. 6, what do the dotted black lines denote?

Reply: The dotted black lines denote the spatial correlation coefficient value of 0.5. We plotted the dotted black lines as references to compare the correlations of 14 models more clearly. We have added the description of dotted black lines behind Fig. 6 (Line 784) and cited it here:

“The black dotted lines denote the spatial correlation coefficient value of 0.5.”

Reference

de Leeuw, G., Fan, C., Li, Z., Dong, J., Li, Y., Ou, Y., and Zhu, S.: Spatiotemporal variation and provincial scale differences of the AOD across China during 2000–2021, Atmos. Pollut. Res., 13, 101359, https://doi.org/10.1016/j.apr.2022.101359, 2022.

Geng, G., Xiao, Q., Liu, S., Liu, X., Cheng, J., Zheng, Y., Xue, T., Tong, D., Zheng, B., Peng, Y., Huang, X., He, K., and Zhang, Q.: Tracking Air Pollution in China: Near Real-Time PM2.5 Retrievals from Multisource Data Fusion, Environ. Sci. Technol., 55, 12106-12115, https://doi.org/10.1021/acs.est.1c01863, 2021.

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., Kurokawa, J. I., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q.: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS), Geosci. Model Dev., 11, 369-408, https://doi.org/10.5194/gmd-11-369-2018, 2018.

Kendall, M. G.: A new measure of rank correlation, Biometrika, 30, 81-93, https://doi.org/10.1093/biomet/30.1-2.81, 1938.

Michou, M., Nabat, P., Saint-Martin, D., Bock, J., Decharme, B., Mallet, M., Roehrig, R., Séférian, R., Sénési, S., and Voldoire, A.: Present-Day and Historical Aerosol and Ozone Characteristics in CNRM CMIP6 Simulations, J. Adv. Model. Earth Syst., 12, e2019MS001816, https://doi.org/10.1029/2019MS001816, 2020.

Mortier, A., Gliß, J., Schulz, M., Aas, W., Andrews, E., Bian, H., Chin, M., Ginoux, P., Hand, J., Holben, B., Zhang, H., Kipling, Z., Kirkevåg, A., Laj, P., Lurton, T., Myhre, G., Neubauer, D., Olivié, D., von Salzen, K., Skeie, R. B., Takemura, T., and Tilmes, S.: Evaluation of climate model aerosol trends with ground-based observations over the last 2 decades – an AeroCom and CMIP6 analysis, Atmos. Chem. Phys., 20, 13355-13378, https://doi.org/10.5194/acp-20-13355-2020, 2020.

Sen, P. K.: Estimates of the Regression Coefficient Based on Kendall's Tau, J. Am. Stat. Assoc., 63, 1379-1389, https://doi.org/10.1080/01621459.1968.10480934, 1968.

Sockol, A. and Small Griswold, J. D.: Intercomparison between CMIP5 model and MODIS satellite-retrieved data of aerosol optical depth, cloud fraction, and cloud-aerosol interactions, Earth Space Sci., 4, 485-505, https://doi.org/10.1002/2017EA000288, 2017.

Theil, H.: A Rank-Invariant Method of Linear and Polynomial Regression Analysis, Proc. R. Neth. Acad. Sci., 386–392, van Donkelaar, A., Martin, R. V., Brauer, M., Hsu, N. C., Kahn, R. A., Levy, R. C., Lyapustin, A., Sayer, A. M., and Winker, D. M.: Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors, Environ. Sci. Technol., 50, 3762-3772, https://doi.org/10.1021/acs.est.5b05833, 2016.

van Donkelaar, A., Hammer, M. S., Bindle, L., Brauer, M., Brook, J. R., Garay, M. J., Hsu, N. C., Kalashnikova, O. V., Kahn, R. A., Lee, C., Levy, R. C., Lyapustin, A., Sayer, A. M., and Martin, R. V.: Monthly Global Estimates of Fine Particulate Matter and Their Uncertainty, Environ. Sci. Technol., 55, 15287-15300, https://doi.org/10.1021/acs.est.1c05309, 2021.

Wang, C., Wang, Z., Lei, Y., Zhang, H., Che, H., and Zhang, X.: Differences in East Asian summer monsoon responses to Asian aerosol forcing under different emission inventories, Advances in Climate Change Research, 13, 309-322, https://doi.org/10.1016/j.accre.2022.02.008, 2022.