Response to Referee #1 Comments

We sincerely thank the referee for the time and effort devoted to reading our manuscript and for providing thoughtful and constructive comments. We have carefully considered all of the points raised and have revised the manuscript accordingly. Below, we list the reviewer's original comments (reproduced in *italic text*), followed by our responses in blue text. All line numbers and figure numbers cited in our replies refer to the revised manuscript.

Referee #1

Overview:

The paper presents a description and evaluation of the aerosol simulation performance of the coupled climate-aerosol CMA-CPSv4 climate modelling system for a 20-year period (2001-2020). The meteorology is nudged to ERA5 and CMIP forcing and emissions data were used. The model simulation is evaluated with AOD (Aeronet) and surface PM25 observations (East Asia networks) and compared against the MERRA-2 reanalysis and the CMIP MME. The authors find overall a good simulation of global aerosols with some issues related regional desert dust and sulphate over the oceans.

General remarks

The paper is a solid evaluation of 20-year aerosol simulation with chemistry-climate model. The study's conclusions for climate applications are limited since CMA-CPSv4 is nudged to ERA5 analysis. It is also unclear how prognostic aerosols are used in both the radiation scheme (direct effects) and cloud microphysics (indirect effects). While CMA-CPSv4 appears to function similarly to the CTM, the title implies that aerosol feedback is included. The authors should clarify which aspects of aerosol-climate coupling are examined and evaluated.

Response: We thank you for these important and constructive comments. We realize that the objective and scope of this study were not sufficiently clear in Section "1 Introduction" part of the original manuscript. CMA-CPSv4 is the latest generation of the climate prediction system developed by CMA. The present work focuses on a comprehensive evaluation of the aerosol module in CMA-CPSv4, and it represents only the first part of the comprehensive evaluations for CMA-CPSv4. We have revised the relevant text in the Introduction (Lines 78–80) to clarify this point. If permitted by the GMD editorial office, we would also like to revise the title to **"Global fully coupled climate-aerosol CMA-CPSv4: part I. aerosol simulation performance"** to more accurately reflect the scope of this paper.

As the reviewer pointed out, the evaluation framework adopted in this work is broadly consistent with that commonly used in chemical transport model (CTM) studies. Accordingly, the primary purpose of the present work is to assess whether CMA-CPSv4 can realistically reproduce the spatiotemporal variability of aerosols under realistic large-scale meteorological forcing. Aerosol–climate coupling and aerosol feedback processes, including aerosol–radiation interactions (ARI) and aerosol–cloud interactions (ACI), and their respective impacts on climate prediction will be systematically investigated and evaluated in a separate follow-up paper due to the paper's length limit.

The paper does not provide evidence that CMA-CPSv4 represents an improvement over v3. If the inclusion of aerosol weather and climate feedback is considered a significant innovation in v4, the authors should demonstrate this by illustrating how these feedbacks enhance specific aspects of the climate simulation. Despite nudging to ERA5, differences in cloud and precipitation simulations are expected and should be addressed. If v3 simulation data are unavailable, it would be appropriate to compare v4 simulations with and without aerosol interactions.

Response: We thank you for this valuable suggestion. A systematic comparison between CMA-CPSv4 and CMA-CPSv3 will be presented in our follow-up paper. Our preliminary analyses indicate that CMA-CPSv4, which includes aerosol feedback processes, generally exhibits improved prediction skill for near-surface temperature and total precipitation compared with CMA-CPSv3. As an illustrative example, we provide supplementary comparisons in the following Figs. R1 and R2, showing the prediction skill of near-surface temperature and total precipitation for March in CMA-CPSv3 and CMA-CPSv4.

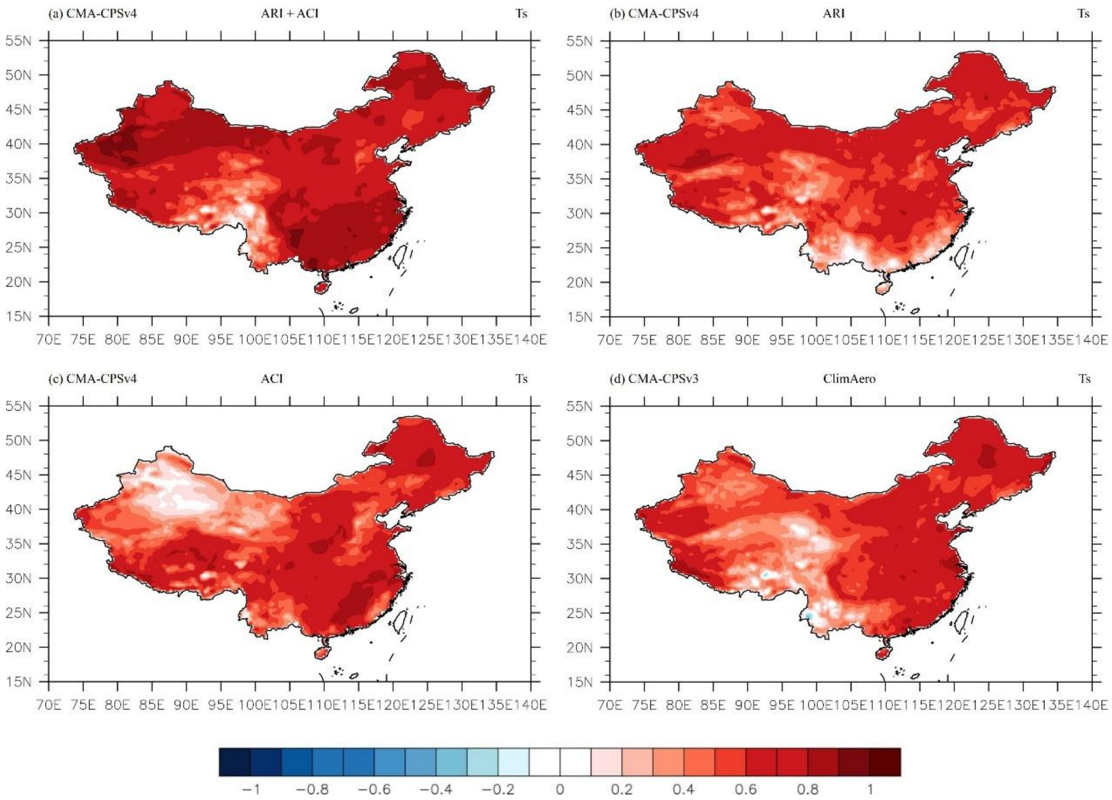

Figure R1. Temporal correlation coefficients (TCC) of March surface air temperature forecasts over China during 2008–2022, verified against CRA40 (CMA's global atmospheric reanalysis; http://data.cma.cn/analysis/cra40). Forecasts are initialized on 28 February of each year and produced by (a) CMA-CPSv4 with fully interactive prognostic aerosols including both aerosol–radiation interactions (ARI) and aerosol–cloud interactions (ACI), (b) CMA-CPSv4 with ARI only, (c) CMA-CPSv4 with ACI only, and (d) CMA-CPSv3 with climatological aerosols (ClimAero).

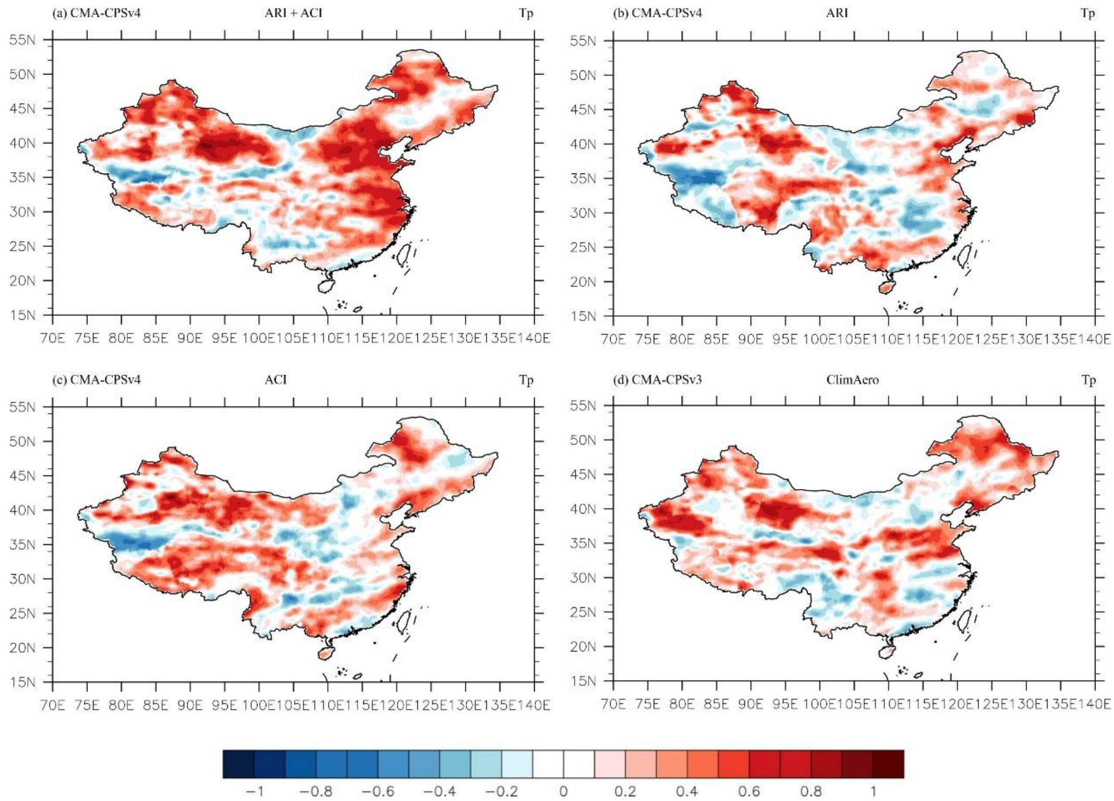

Figure R2. Temporal correlation coefficients (TCC) of March total precipitation forecasts over China during 2008–2022, verified against CRA40 (CMA's global atmospheric reanalysis; http://data.cma.cn/analysis/cra40). Forecasts are initialized on 28 February of each year and produced by (a) CMA-CPSv4 with fully interactive prognostic aerosols including both aerosol–radiation interactions (ARI) and aerosol–cloud interactions (ACI), (b) CMA-CPSv4 with ARI only, (c) CMA-CPSv4 with ACI only, and (d) CMA-CPSv3 with climatological aerosols (ClimAero).

Regarding the issue of cloud and precipitation simulations in the present manuscript, differences between simulations and observations were not analyzed. We think that cloud and precipitation simulations under nudging toward ERA5 reanalysis fields have certain limitations. In CMA-CPSv4, nudging is applied to large-scale prognostic variables, including horizontal winds, temperature, and specific humidity, toward 6-hourly ERA5 reanalysis fields. As a result, the large-scale circulation is constrained by ERA5, which may lead to stronger upward motion and enhanced aerosol transport to higher altitudes, as shown in Fig. 3. At the same time, the nudging approach may suppress small-scale circulation variability and related precipitation processes (as shown in Fig. R3), which is also a limitation of this method. This suppression may weaken aerosol wet scavenging and consequently contribute to higher aerosol concentrations in the upper troposphere.

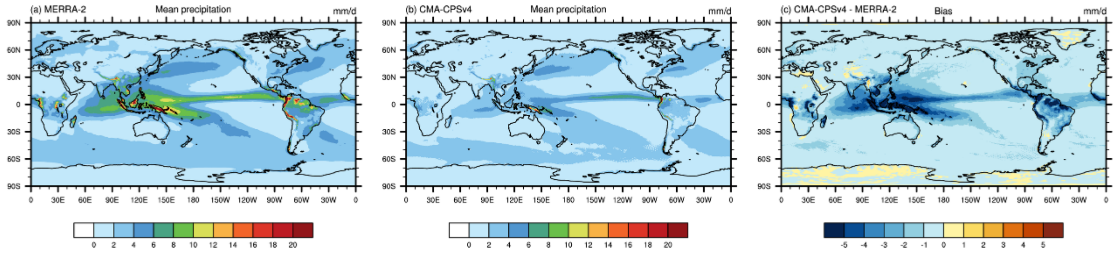

Figure R3. Global distributions of annual mean total precipitation (units: mm d^{-1}) from MERRA-2 (left), CMA-CPSv4 (middle), and their bias (right) over the 2001–2020 period.

Utilizing MERRA-2 and the CMIP multi-model ensemble as reference is problematic due to well-documented limitations in these data sets. Specifically, for the scatter plots shown in Figure 7 (PM observations) and Figure 9 (Aeronet AOD), it is advisable to always display observations on the x-axis and model outputs on the y-axis. Additionally, Aeronet analyses should include relevant CMIP data for comprehensive comparison.

Response: The referee mentioned that utilizing MERRA-2 and the CMIP multi-model ensemble as reference datasets is problematic. We think that MERRA-2 are widely used reference datasets in aerosol evaluation studies (e.g., Buchard et al., 2017; Gui et al., 2026; Randles et al., 2017; Su et al., 2022; Turnock et al., 2020; Zhou et al., 2023). MERRA-2 assimilates meteorological observations and aerosol-related satellite retrievals, providing an observationally constrained estimate of aerosol distributions. The CMIP6 MME represents the ensemble-mean behavior of ESMs and is also frequently used in model evaluation (Li et al., 2021; Li et al., 2024; Mortier et al., 2020; Turnock et al., 2020). We think that CMIP6 MME also provide useful references to evaluating aerosol simulations in ESMs.

Following your suggestions, we have revised Section 3.6 and added the CMIP6 multi-model ensemble mean (MME) AOD results to Figures 8, 9 and 10 to provide a more comprehensive comparison.

Buchard, V., Randles, C. A., Silva, A. M. da, Darmenov, A., Colarco, P. R., Govindaraju, R., Ferrare, R., Hair, J., Beyersdorf, A. J., Ziemba, L. D., and Yu, H.: The MERRA-2 Aerosol Reanalysis, 1980 Onward. Part II: Evaluation and Case Studies, https://doi.org/10.1175/JCLI-D-16-0613.1, 2017.

Gui, K., Zhang, X., Che, H., Li, L., Zheng, Y., An, L., Miao, Y., Zhao, H., Dubovik, O., Holben, B., Wang, J., Gupta, P., Lind, E. S., Toledano, C., Wang, H., Wang, Z., Wang, Y., Huang, X., Dai, K., Xia, X., Xu, X., and Zhang, X.: Advancing operational global aerosol forecasting with machine learning, Nature, 651, 658–665, https://doi.org/10.1038/s41586-026-10234-y, 2026.

Li, F., Song, X., Harrison, S. P., Marlon, J. R., Lin, Z., Leung, L. R., Schwinger, J., Marécal, V., Wang, S., Ward, D. S., Dong, X., Lee, H., Nieradzik, L., Rabin, S. S., and Séférian, R.: Evaluation of global fire simulations in CMIP6 Earth system models, Geoscientific Model Development, 17, 8751–8771, https://doi.org/10.5194/gmd-17-8751-2024, 2024.

Li, X., Liu, Y., Wang, M., Jiang, Y., and Dong, X.: Assessment of the Coupled Model Intercomparison Project phase 6 (CMIP6) Model performance in simulating the spatial-temporal variation of aerosol optical depth over Eastern Central China, Atmospheric Research, 261, 105747, https://doi.org/10.1016/j.atmosres.2021.105747, 2021.

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, Atmospheric Chemistry and Physics, 20, 13355–13378, https://doi.org/10.5194/acp-20-13355-2020, 2020.

Randles C. A., Silva A. M., Buchard V., Colarco P. R., Darmenov A., Govindaraju R., Smirnov A., Holben B., Ferrare R., Hair J., Shinozuka Y., and Flynn C. J.: The MERRA-2 Aerosol Reanalysis, 1980 Onward. Part I: System Description and Data Assimilation Evaluation, https://doi.org/10.1175/JCLI-D-16-0609.1, 2017.

Su, X., Wu, T., Zhang, J., Zhang, Y., Jin, J., Zhou, Q., Zhang, F., Liu, Y., Zhou, Y., Zhang, L., Turnock, S. T., and Furtado, K.: Present-Day PM2.5 over Asia: Simulation and Uncertainty in CMIP6 ESMs, J. Meteorol. Res., 36, 429–449, https://doi.org/10.1007/s13351-022-1202-7, 2022.

Turnock, S. T., Allen, R. J., Andrews, M., Bauer, S. E., Deushi, M., Emmons, L., Good, P., Horowitz, L., John, J. G., Michou, M., Nabat, P., Naik, V., Neubauer, D., O'Connor, F. M., Olivié, D., Oshima, N., Schulz, M., Sellar, A., Shim, S., Takemura, T., Tilmes, S., Tsigaridis, K., Wu, T., and Zhang, J.: Historical and future changes in air pollutants from CMIP6 models, Atmos. Chem. Phys., 20, 14547–14579, https://doi.org/10.5194/acp-20-14547-2020, 2020.

Zhou, Y., Wu, T., Zhang, J., Zhang, Y., Zhang, Y., Zhao, H., Jie, W., Zhou, Y., and Furtado, K.: Variation of Dust in Northern China and Its Reproduction in BCC-ESM1 since 1980, J. Meteorol. Res., 37, 617–631, https://doi.org/10.1007/s13351-023-2195-6, 2023.

While the case study of two dust events in May 2011 is interesting, it raises the question of the representativeness of these events for the 20-year study period. A more scientifically valid analysis would have involved evaluating the model's ability to forecast the frequency, timing, and intensity of dust events across China or other regions worldwide.

Response: We have revised "Section 3.7 Case study of severe dust events in China" to "Section 3.7 Case analysis of pollution events in China". Following your suggestion, two new figures (Figs. 11 and 12 in the revised manuscript) have been added to Section 3.7: (1) the spatial distribution of $PM_{2.5}$ pollution-event frequency across China, and (2) the site-based daily evolution of $PM_{2.5}$ concentrations in Beijing during December for 2014–2019. Daily surface $PM_{2.5}$ observations from the China National Environmental Monitoring Center (CNEMC) for 2014–2019 were used to evaluate the frequency and temporal evolution of pollution events. The performance of CMA-CPSv4 was assessed and compared with MERRA-2. The results show that both CMA-CPSv4 and MERRA-2 reasonably reproduce the temporal evolution and peak timing of pollution events, although both systematically underestimate event intensity relative to observations. These corresponding revisions have been incorporated into Section 3.7 of the revised manuscript.

The frequent use of subjective terms like "good agreement" or "reasonable agreement" is a weakness; quantitative assessments should be used instead.

Response: We thank you for this helpful suggestion. Following this suggestion, we have carefully revised the relevant descriptions throughout the manuscript and, wherever possible, replaced subjective statements with quantitative assessments supported by statistical metrics.