The authors have addressed all of my major comments from the original submission. The revised manuscript shows significant improvements. I believe that GPU-accelerated computing will be important in the future if the data transfer bottleneck between the host and device can be resolved. Future generations of accelerated devices and CPUs are expected to be more tightly integrated.

Response: We sincerely thank the referee for the careful re-evaluation and positive assessment of the revised manuscript. We also appreciate the constructive comments, which have helped us improve the presentation of the model evaluation results. Our point-by-point responses are provided below.

1) **Figures 6 and 7 have not been referenced or discussed in the main text.**

Response: Thanks for pointing this out. We have now explicitly referenced and discussed Figures 6 and 7 in Sect. 4.2.1. Visually, the spatial distributions and concentration magnitudes calculated by the Fortran and HIP-Opt2 versions are nearly identical. Further comparison shows that the differences in HONO concentrations are within ± 0.0001 ppbV, while the differences for the other species are within ± 0.001 ppbV for gases or $\pm 0.001\ \mu g \cdot m^{-3}$ for aerosols. Cao et al. (2024) reported that the differences between the HIP and Fortran versions of CAMx were within ± 0.01 ppbV for O_3 and NO_2 and within $\pm 0.005\ \mu g \cdot m^{-3}$ for PSO_4. By comparison, in the present study, the differences in O_3 and NO_2 are within ± 0.001 ppbV, while the difference in ASO_4 is within $\pm 0.001\ \mu g \cdot m^{-3}$, all of which are smaller than the corresponding differences reported by Cao et al. (2024). For each species, the grid-cell differences were further averaged. The resulting mean differences are 5.1×10^{-9}, 3.6×10^{-7}, 1.7×10^{-6}, and 3.2×10^{-4} ppbV for HONO, SO_2, NO_2, and O_3, respectively, and 1.1×10^{-6}, 1.8×10^{-5}, 4.0×10^{-6}, and $1.0 \times 10^{-6}\ \mu g \cdot m^{-3}$ for BC, $PM_{2.5}$, ASO_4, and ANH_4, respectively. Taken together, the smaller difference ranges and near zero domain mean differences confirm the close numerical agreement between the Fortran and HIP-Opt2 versions. We have added this part in **lines 490-503**, which are as follows:

Figures 6 and 7 compare the gas and aerosol concentrations calculated by the original Fortran and HIP-Opt2 versions after the 24 h integration. Visually, the two versions produce nearly identical spatial distributions and concentration magnitudes. The differences in HONO concentrations are within ± 0.0001 *ppbV, while the differences for the other species are within* ± 0.001 *ppbV for gases or* $\pm 0.001\ \mu g \cdot m^{-3}$ *for aerosols. Cao et al. (2024) reported that the differences between the HIP and Fortran versions of CAMx were within* ± 0.01 *ppbV for* O_3 *and* NO_2 *and within* $\pm 0.005\ \mu g \cdot m^{-3}$ *for* PSO_4*. By comparison, in the present study, the differences in* O_3 *and* NO_2 *are within* ± 0.001 *ppbV, while the difference in* ASO_4 *is within* $\pm 0.001\ \mu g \cdot m^{-3}$*, all of which are smaller than the corresponding differences reported by Cao et al. (2024). For each species, the grid-cell differences were further averaged. The resulting mean differences are* 5.1×10^{-9}*,* 3.6×10^{-7}*,* 1.7×10^{-6}*, and* 3.2×10^{-4} *ppbV for HONO,* SO_2*,* NO_2*, and* O_3*, respectively, and* 1.1×10^{-6}*,* 1.8×10^{-5}*,* 4.0×10^{-6}*, and* $1.0 \times 10^{-6}\ \mu g \cdot m^{-3}$ *for BC,* $PM_{2.5}$*,* ASO_4*, and* ANH_4*, respectively. Taken together, the smaller difference ranges and near zero domain mean differences confirm the close numerical agreement between the Fortran and HIP-Opt2 versions.*

2) **Figure 6, the caption should be expanded to describe the left, middle, and right panels. Specify the units of color bar (ppbV, ugm-3, %). Reduce the color bar scale to see the difference between the two models. Include O3 in your evaluation because it is an important air pollutant.**

Response: Thanks for this helpful suggestion. We have redrawn Figures 6 and 7 to provide a clearer comparison between the two model versions. In both figures, the left, middle, and right columns show the Fortran results, HIP-Opt2 results, and differences between the two versions, respectively. The units are now marked in the upper-right corner of every panel. To make the small differences more distinguishable, we reduced the color-bar ranges of the difference panels from the original ± 0.1 to ± 0.0001 for HONO and ± 0.001 for the other species, with units of ppbV for gases and $\mu g \cdot m^{-3}$ for aerosols. In addition, NH_3 has been replaced by O_3 in Figure 6, and the corresponding O_3 difference has been included. The captions of Figures 6 and 7 have also been expanded to describe the left, middle, and right panels and their units. The updated Figures 6 and 7 are provided below.

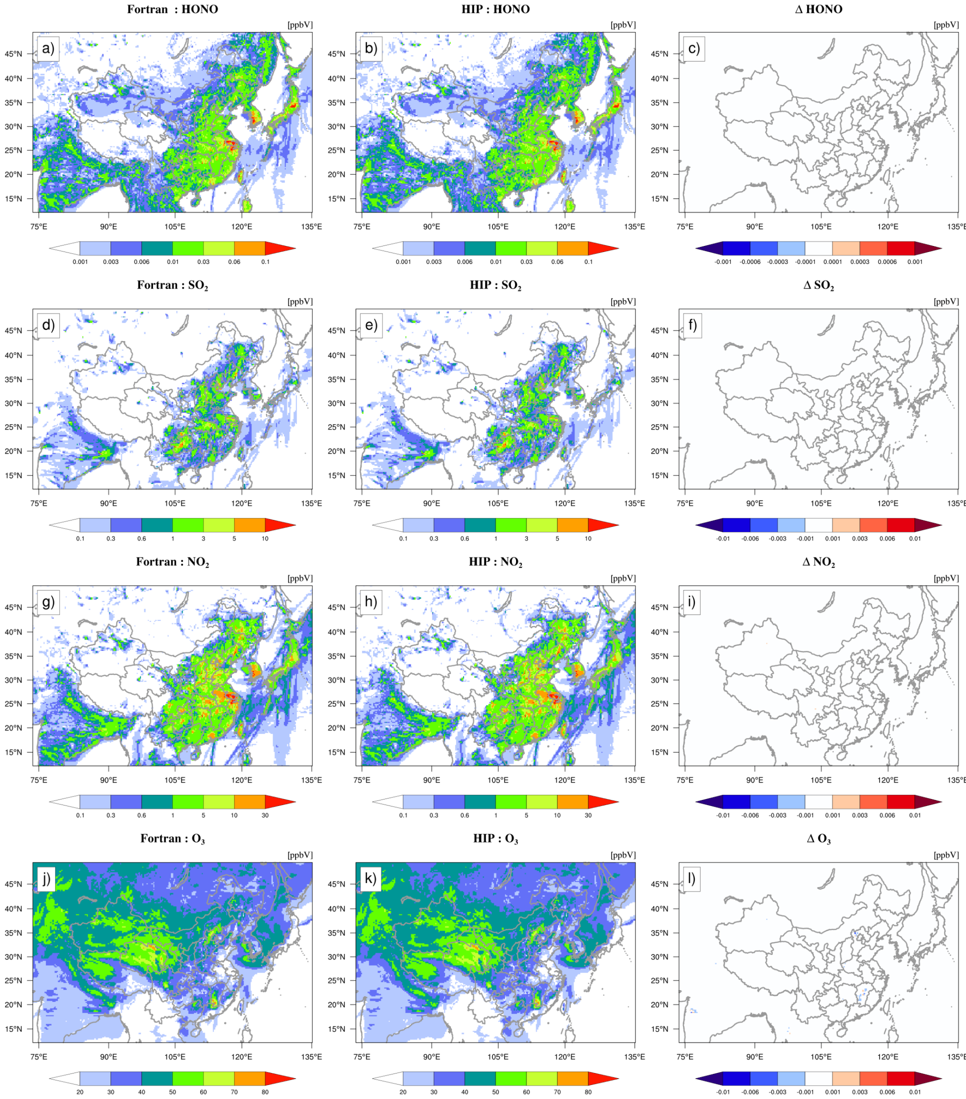

Figure 6. *Spatial distributions of HONO, SO_2, NO_2, and O_3 concentrations calculated by the original Fortran and HIP-Opt2 versions after the 24 h integration. Panels (a), (d), (g), and (j) show the Fortran results; panels (b), (e), (h), and (k) show the HIP-Opt2 results; and panels (c), (f), (i), and (l) show the differences between the two versions. The units of the concentrations and differences are ppbV.*

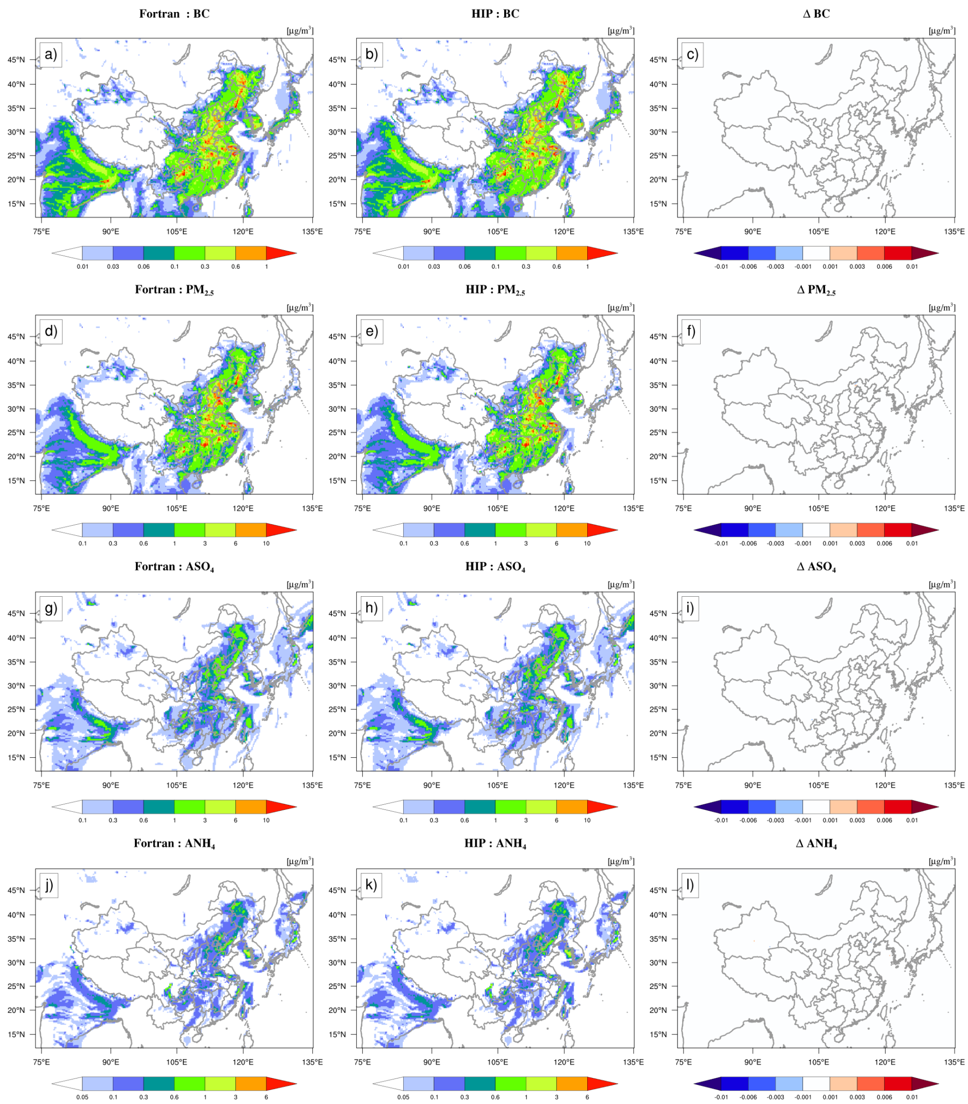

Figure 7. *Spatial distributions of BC, $PM_{2.5}$, ASO_4, and ANH4 concentrations calculated by the original Fortran and HIP-Opt2 versions after the 24 h integration. Panels (a), (d), (g), and (j) show the Fortran results; panels (b), (e), (h), and (k) show the HIP-Opt2 results; and panels (c), (f), (i), and (l) show the differences between the two versions. The units of the concentrations and differences are* $\mu g \cdot m^{-3}$.

3) Tables 5 and 6, include the normalized mean bias, as it is an important metric. Since each species has different concentration range. For example, the mean concentration of O3 is about 30 ppbV, while HONO is around 1 ppbV.

Response: Thanks for this important suggestion. We agree that normalized mean bias (NMB) is useful for comparing species with substantially different concentration ranges. We have therefore added NMB to Tables 5 and 6 and incorporated its interpretation into the corresponding discussions. For Table 5, NMB represents the overall signed difference of the HIP-Opt2 results relative to the Fortran results. The NMB values range from $-1.0 \times 10^{-4}\%$ for HONO to $2.9 \times 10^{-3}\%$ for ASO_4, indicating negligible systematic bias between the two model versions. For Table 6, NMB represents the normalized difference between the simulated and observed concentrations. Positive and negative NMB values indicate model overestimation and underestimation, respectively. We recalculated RMSE, r, and NMB using the same city-level simulated and observed time series and updated both Table 6 and the accompanying discussion. The contents of the above two parts have been revised in **lines 518-529** and **lines 593-608**, which are as follows:

Lines 518-529:

To further evaluate the suitability of HIP-Opt2 for scientific applications, we followed Wang et al. (2021) and Cao et al. (2024) and calculated the root mean square error (RMSE), standard deviation (std), RMSE/std ratio, and normalized mean bias (NMB). RMSE quantifies the differences between HIP-Opt2 and the original Fortran version, while std represents the spatial variability of each species in the Fortran results. The RMSE/std ratio therefore measures the magnitude of the computational differences relative to the spatial variability of the corresponding species. NMB measures the overall signed difference of the HIP-Opt2 results relative to the Fortran results. As shown in Table 5, the RMSE/std ratios range from $6.9 \times 10^{-4}\%$ *for* NO_2 *to* $9.2 \times 10^{-2}\%$ *for* ANH_4. *The NMB values range from* $-1.0 \times 10^{-4}\%$ *for HONO to* $2.9 \times 10^{-3}\%$ *for* ASO_4. *The small RMSE/std ratios and near-zero NMB values demonstrate that the computational differences introduced by the HIP implementation are negligible relative to the spatial variability and concentration magnitudes of the evaluated species.*

Table 5. *Root mean square error (RMSE) between the Fortran and HIP-Opt2 versions, standard deviation (std) of the Fortran results, RMSE/std ratio, and normalized mean bias (NMB) of HIP-Opt2 relative to the Fortran version.*

	RMSE	***std***	***RMSE/std (%)***	***NMB (%)***

HONO (ppbV)	1.1×10^{-7}	*0.01*	1.1×10^{-3}	-1.0×10^{-4}
SO_2 (ppbV)	1.2×10^{-5}	*0.5*	2.4×10^{-3}	1.9×10^{-4}
NO_2 (ppbV)	1.8×10^{-5}	*2.7*	6.9×10^{-4}	1.9×10^{-4}
O_3 (ppbV)	3.6×10^{-4}	*14.2*	2.5×10^{-3}	9.6×10^{-4}
BC ($\mu g \cdot m^{-3}$)	4.5×10^{-6}	*0.1*	3.6×10^{-3}	2.0×10^{-3}
$PM_{2.5}$ ($\mu g \cdot m^{-3}$)	4.3×10^{-4}	*2.1*	2.2×10^{-2}	2.0×10^{-3}
ASO_4 ($\mu g \cdot m^{-3}$)	2.1×10^{-5}	*0.3*	7.9×10^{-3}	2.9×10^{-3}
ANH_4 ($\mu g \cdot m^{-3}$)	9.6×10^{-5}	*0.1*	9.2×10^{-2}	2.4×10^{-3}

Lines 593-608:

Figures 8 and 9, together with Table 6, present the time-series comparisons and statistical evaluation of the simulated and observed $PM_{2.5}$ and O_3 concentrations. The formulas used to calculate the statistical metrics are provided in the Supplement. Positive and negative NMB values indicate model overestimation and underestimation, respectively. For daily $PM_{2.5}$, the EPICC-Model reproduced the temporal variations reasonably well in most cities. Fuzhou had an RMSE of 3.9 $\mu g \cdot m^{-3}$, an r of 0.87, and an NMB of −14.6%. Across the eight cities, the mean RMSE, r, and city-level NMB were 9.4 $\mu g \cdot m^{-3}$, 0.70, and 19.7%, respectively. The positive NMB values in Beijing and Hefei, 89.1% and 74.0%, indicate overestimation, whereas the negative NMB values in Guangzhou and Fuzhou, -28.1% and -14.6%, indicate underestimation. For MDA8 O_3, the mean RMSE, r, and city-level NMB were 12.1 ppbV, 0.76, and −17.7%, respectively, indicating an overall tendency toward underestimation. The largest negative NMB values occurred in Taiyuan and Hangzhou, at -41.5% and -34.3%, respectively, while Qingdao showed a small positive bias of 4.7%. Qingdao and Guangzhou had r values of 0.91 and 0.87 and RMSE values of 12.2 and 7.8 ppbV, respectively. These statistical results indicate that the EPICC-Model V1.6.0, integrated with GPU-HADVPPM4HIP V1.0, reasonably reproduces the spatiotemporal characteristics of $PM_{2.5}$ and O_3 concentrations on China' s domestic heterogeneous computing clusters.

Table 6. *Statistical evaluation of the $PM_{2.5}$ and O_3 simulations of the EPICC-Model in eight Chinese cities. NMB denotes normalized mean bias. The average row represents*

the arithmetic mean of the city-level metrics.

	$PM_{2.5}$			***O_3***		
	RMSE ($\mu g \cdot m^{-3}$)	*r*	*NMB (%)*	*RMSE (ppbV)*	*r*	*NMB (%)*
Beijing	*17.3*	*0.84*	*89.1*	*13.2*	*0.77*	*-7.0*
Taiyuan	*11.0*	*0.54*	*18.6*	*21.6*	*0.78*	*-41.5*
Hangzhou	*7.4*	*0.50*	*-3.7*	*14.8*	*0.80*	*-34.3*
Hefei	*12.4*	*0.55*	*74.0*	*14.2*	*0.70*	*-15.3*
Fuzhou	*3.9*	*0.87*	*-14.6*	*7.5*	*0.77*	*-17.0*
Qingdao	*11.0*	*0.90*	*20.6*	*12.2*	*0.91*	*4.7*
Guangzhou	*7.2*	*0.65*	*-28.1*	*7.8*	*0.87*	*-14.9*
Kunming	*5.3*	*0.76*	*2.0*	*6.0*	*0.46*	*-16.4*
Average	*9.4*	*0.70*	*19.7*	*12.1*	*0.76*	*-17.7*

4) Provide the definition of GPU-like accelerator.

Response: Thanks for requesting clarification of this terminology. In this study, a GPU-like accelerator refers to a domestically developed general-purpose accelerator in China that performs the same massively parallel acceleration role as a conventional GPU and supports GPU-style programming through the HIP framework. Its architecture and programming model are similar to those of AMD GPUs, including the single-instruction multiple-thread (SIMT) execution model and the grid-block-thread parallel hierarchy. The GPU-like accelerator used in this study has 3840 computing units and 16 GB of device memory, as summarized in Table 1. The term "GPU-like accelerator" follows Cao et al. (2024), who used HIP to port and accelerate the PPM advection solver of the CAMx air quality model on the same type of domestically developed accelerator in China. The broader applicability of HIP to geoscientific models was also demonstrated by Wang et al. (2021), who developed and applied the HIP version of the LICOM3 ocean model on large-scale AMD GPU systems. Accordingly, we have clarified the definition and application context of GPU-like accelerators in Sect. 2.2. We have added this part in **lines 232-248**, which are as follows:

All performance testing of the EPICC-Model V1.6.0 and the heterogeneous adaptation and optimization of its advection module were conducted at the Earth System Numerical Simulation Facility (EarthLab; Chai et al., 2021). Jointly *developed by the Institute of Atmospheric Physics, Chinese Academy of Sciences, and collaborating institutions, EarthLab is designed for Earth system modeling and high-resolution regional environmental simulations and employs a CPU-GPU heterogeneous architecture. Within this platform, the term "GPU-like accelerator" refers to a domestically developed general-purpose accelerator in China that performs the same massively parallel acceleration role as a conventional GPU and supports GPU-style programming through the HIP framework. The GPU-like accelerator used in this study has 3,840 computing units and 16 GB of device memory. As summarized in Table 1, each GPU-like computing node contains two domestically developed CPU processors and two GPU-like accelerators connected through PCIe 4.0 buses, and both the CPUs and GPU-like accelerators are first-generation devices. Cao et al. (2024) previously used HIP to accelerate the PPM advection solver of the CAMx air quality model on the same type of domestic GPU-like accelerator, while Wang et al. (2021) demonstrated the large-scale application of the HIP-based LICOM3 ocean model. The software environment includes the Intel OneAPI 2021.3.0 toolkit for compiling the CPU code and the dtk-23.04.1 toolkit for compiling the GPU-like code.*

References:

Cao, K., Wu, Q., Wang, L., Guo, H., Wang, N., Cheng, H., Tang, X., Li, D., Liu, L., Li, D., Wu, H., and Wang, L.: GPU-HADVPPM4HIP V1.0: using the heterogeneous-compute interface for portability (HIP) to speed up the piecewise parabolic method in the CAMx (v6.10) air quality model on China's domestic GPU-like accelerator, Geosci. Model Dev., 17, 6887–6901, https://doi.org/10.5194/gmd-17-6887-2024, 2024.

Wang, P., Jiang, J., Lin, P., Ding, M., Wei, J., Zhang, F., Zhao, L., Li, Y., Yu, Z., Zheng, W., Yu, Y., Chi, X., and Liu, H.: The GPU version of LASG/IAP Climate System Ocean Model version 3 (LICOM3) under the heterogeneous-compute interface for portability (HIP) framework and its large-scale application, Geosci. Model

Dev., 14, 2781–2799, https://doi.org/10.5194/gmd-14-2781-2021, 2021.