General Comments

The authors present a hardware-accelerated implementation of the EPICC air quality model's horizontal advection scheme, demonstrating impressive computational performance improvements achieved through clever and careful source code translation and refactoring, although these improvements are offset by data transfer overhead. The authors couple the accelerated advection module back into the full model and demonstrate that this heterogeneous model implementation retains code integrity to produce reasonable simulation results. Moreover, the authors present a thorough performance analysis. Taken together, this is a comprehensive and impressive effort that may offer an impetus for the air quality modeling community to collectively modernize model codebases towards enabling the kind of ultra-high-resolution simulations that are necessary to characterize urban- and neighborhood-scale pollution, the scales at which communities experience harmful exposures. My comments and brief and minor.

Response: We sincerely thank the referee for the positive and encouraging assessment of our work. We also appreciate the referee's constructive comments, which have helped us improve the clarity and technical completeness of the manuscript. Our point-by-point responses are provided below.

Specific comments

Line 130: Please include, along with brief term definitions, the species continuity equation explicitly.

Response: Thank you for this helpful suggestion. We agree that explicitly presenting the species continuity equation and defining its terms will improve the clarity and completeness of the model description. We will add the species continuity equation and definitions of the relevant terms in Sect. 2.1 in the revised manuscript, as follows:

The model framework is fundamentally based on the following species continuity equation and is used to simulate the complex physical and chemical processes of pollutants in the atmosphere.

$$\frac{\partial C_i}{\partial t} + u\frac{\partial C_i}{\partial x} + v\frac{\partial C_i}{\partial y} + \omega\frac{\partial C_i}{\partial z} = \frac{\partial}{\partial x}\left(K_x\frac{\partial C_i}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_y\frac{\partial C_i}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_z\frac{\partial C_i}{\partial z}\right) + S + P - R_d - W_{ash}$$

where C_i is the concentration of pollutant species i, u, v, and ω are the wind-velocity components in the x, y, and z directions, respectively. K_x, K_y, and K_z are the turbulent diffusion coefficients in *the longitudinal, latitudinal, and vertical directions, respectively. S represents the emission, P represents chemical transformation, R_d represents dry deposition, and W_{ash} represents wet scavenging.*

Line 217, 599-600: Considering the substantial CPU-GPU data transfer overhead, could a cluster deploying EPICC with GPU-HADVPPM4HIP V1.0 on a cluster with PCIe 5.0 buses, offering double the bandwidth of PCIe 4.0, expect to see additional efficiency improvements? Please include a brief discussion of hardware outlook. Regardless, I commend the authors' focus on software optimization rather than relying purely on hardware capabilities, so that the model is maximally performant across diverse hardware and not only the latest generation.

Response: Thank you for this valuable and forward-looking comment. We fully agree with the referee that upgrading the interconnect from PCIe 4.0 to PCIe 5.0 would improve CPU-GPU communication efficiency. CPU-GPU communication is one of the key factors limiting model performance on heterogeneous clusters. In this study, we reconstructed the loop and array structures of the advection module, reducing the number of CPU-GPU data transfers from approximately 9.8 million to only 10 per time step. In future work, we will further investigate chunked transfers using pinned memory, shared-memory and common-field caching, asynchronous communication, and unified memory to reduce data-transfer overhead and overlap CPU-GPU communication with GPU computation.

In addition to these software-level optimizations, advances in hardware interconnects provide a complementary means of improving CPU-GPU communication efficiency. PCI Express (PCIe) is a general-purpose, high-speed interconnect widely used for data exchange between host CPUs and accelerator devices. Its nominal per-lane raw data rate has increased from 16 GT/s with PCIe 4.0 to 32 GT/s with PCIe 5.0 and 64 GT/s with PCIe 6.0, while the PCIe 7.0 specification further increases it to 128 GT/s (Lender, 2019; Yanes, 2022; PCI-SIG, 2025). The EarthLab heterogeneous cluster used in this study is currently equipped with PCIe 4.0 buses; therefore, we cannot directly evaluate the performance improvement offered by PCIe 5.0. If access to a heterogeneous computing platform equipped with PCIe 5.0 or a later-generation interconnect becomes available, we will systematically evaluate the performance of the EPICC-Model coupled with GPU-HADVPPM4HIP V1.0 on that platform. By combining software optimization with advances in hardware interconnects, we aim to further reduce CPU-GPU communication overhead and improve model performance across different generations of heterogeneous computing platforms. We have added a discussion of PCIe development in **lines 771–786**, as follows::

First, future work will further improve CPU-GPU communication efficiency through chunked transfers using pinned memory, shared-memory and common-field caching, asynchronous communication, and unified memory. These strategies will reduce repeated data transfers and increase the overlap between CPU-GPU communication and GPU computation. In addition to these software-level optimizations, advances in hardware interconnects provide a complementary means of improving CPU-GPU communication efficiency. PCI Express (PCIe) is a general-purpose, high-speed interconnect widely used for data exchange between host CPUs and accelerator devices. Its nominal per-lane raw data rate has increased from 16 GT/s with PCIe 4.0 to 32 GT/s with PCIe 5.0 and 64 GT/s with PCIe 6.0, while the PCIe 7.0 specification further increases it to 128 GT/s (Lender, 2019; Yanes, 2022; PCI-SIG, 2025). The EarthLab heterogeneous cluster used in this study is currently equipped with PCIe 4.0 buses; therefore, the performance improvement offered by PCIe 5.0 cannot be directly evaluated in the present study. If access to a heterogeneous computing platform equipped with PCIe 5.0 or a later-generation interconnect becomes available, we will evaluate GPU-HADVPPM4HIP V1.0 on that platform and further improve CPU-GPU communication efficiency through hardware-software co-optimization.

Reference:

Lender, K.: PCI-SIG in 2019: A year in review, PCI-SIG, https://pcisig.com/blog/pci-sig%C2%AE-2019-year-review (last access: 6 August 2026), 2019.

Yanes, A.: PCIe 6.0 specification released to members: Double the bandwidth for next-generation applications, PCI-SIG, https://pcisig.com/blog/pcie%C2%AE-60-specification-released-members-double-bandwidth-next-generation-applications (last access: 6 August 2026), 2022.

PCI-SIG: The PCIe 7.0 specification, Version 1.0 is now available to members, PCI-SIG, https://pcisig.com/specifications/pcie-70-specification-version-03-now-available-members (last access: 6 August 2026), 2025.

Technical corrections

Line 45: “Air pollution, a source of…” may be better worded as “Air pollution, including…”

Response: Thanks for this correction. We have revised the sentence accordingly at the original **line 45**, which are as follows:

Air pollution, including fine particulate matter in both urban and rural areas, is associated with an elevated risk of strokes, heart diseases, lung cancer, and acute and chronic respiratory diseases (Atkinson et al., 2010; Kim et al., 2015; Liu et al., 2016; Milton and White, 2020).

Line 83: Please define “HPL score”.

Response: Thank you for pointing this out. We will define HPL and briefly explained what the benchmark measures at the **lines 83-86**. The sentence has been revised as follows:

In December 2024, the 64th TOP500 list ranked the El Capitan system first, with a performance of 1.742 EFLOP/s in the High Performance Linpack (HPL) benchmark, which measures the sustained double-precision floating-point performance achieved when solving a dense system of linear equations (TOP500, 2024).

Line 92: Reword “enabling automated generates…” to “enabling automated generation of…”.

Response: Thank you for identifying this grammatical error. We have revised the sentence at the original **line 95** as follows:

Alvanos and Christoudias (2017) developed a software package for the global atmospheric chemistry model ECHAM/MESSy Atmospheric Chemistry (EMAC), enabling automated generation of CUDA kernels for the numerical integration of atmospheric chemical kinetics using the Kinetic PreProcessor (KPP; Damian et al., 2002).

Line 108: “Air quality models-the advection module…” may be scan better with a comma instead of a hyphen.

Response: Thank you for this suggestion. We have replaced the hyphen with a comma at **lines 111-112**, which are as follows:

Regarding the advection module, another computational hotspot in air quality models, Cao et al.

(2023, 2024) implemented GPU acceleration for the CAMx advection module using CUDA and HIP technologies, respectively.

Line 127: “Air pollution complex in China” maybe is meant to be “complex air pollution in China”?

Response: Thanks for this correction. We have changed “air pollution complex in China” to “complex air pollution in China”, which are as follows:

The emission and atmospheric Processes Integrated and Coupled Community Model version 1.6.0 (EPICC-Model V1.6.0; EPICC-Model Working Group, 2025; Wang et al., 2025) is an air quality modeling system specifically designed for complex air pollution in China (Zhu et al., 2023).

Line 133: Change “developed using Fortran program language…” to include “the: between “using” and “Fortran”.

Response: Thanks for this correction. We have revised the sentence at **line 144**, which are as follows:

The EPICC-Model V1.6.0 adopts a modular architecture developed using the Fortran programming language, a high-performance computing language designed for scientific applications.

Figure 1: Wet scavenging typo for RADM; aqueous-phase chemistry seems to reference the vertical diffusion schemes.

Response: Thank you for pointing out these errors, and we apologize for the plotting and spelling mistakes. We have corrected the corresponding entries in Figure 1 as follows:

Wet scavenging: *Regional Acid Deposition Model (RADM; Chang et al., 1987).*

Aqueous-phase chemistry: *Regional Acid Deposition Model (RADM; Chang et al., 1987).*

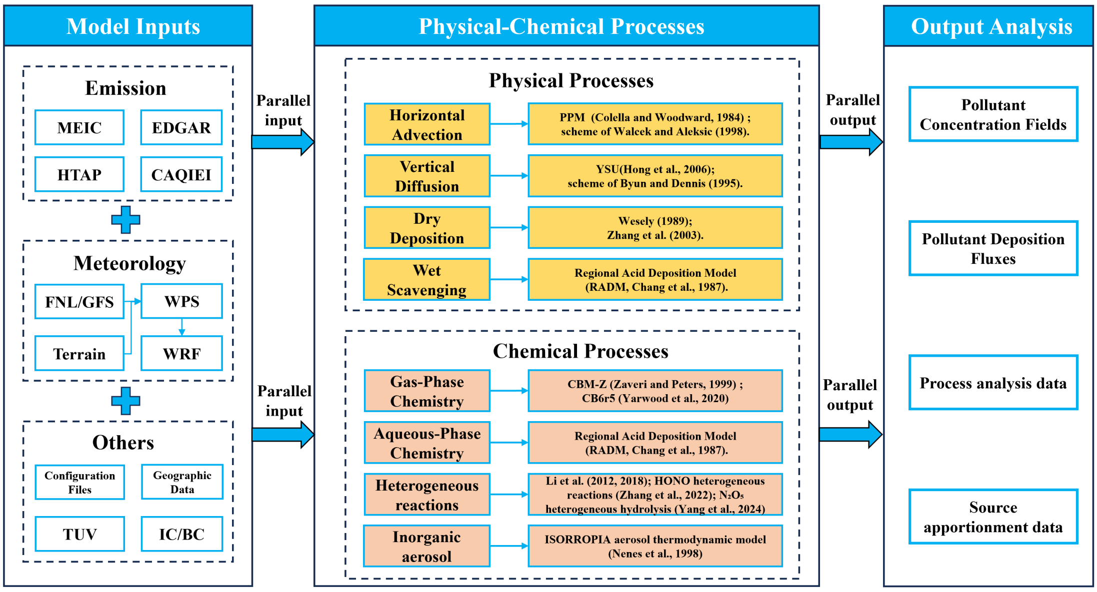

Figure 1. The computational framework and workflow of the EPICC-Model V1.6.0. In the section of physical-chemical processes, yellow represents the physical module, and orange indicates the chemical module.

Line 216: Change “two China’ s domestic CPU…” to “two of China’ s domestic CPU…”

Response: Thank you for this correction. We have revised the sentence at the original **line 228** as follows:

Each GPU-like node contains two of China’s domestic CPU processors and two GPU-like accelerators (Cao et al., 2024) interconnected via PCIe 4.0 buses.