Responses to Editor and Reviewers' Comments

Responses to Editor

Many thanks for addressing the reviewers' comments and revising your manuscript.

The response of the reviewers is diverse. While reviewer #1 indicates technical corrections that I would like to ask you to consider carefully, reviewer #2 still has substantial comments concerning the approach (see reviewer comments).

I am inclined to accept this as an scientific debate that should be documented.

However, this requires a deeper discussion of these points. Please consider the reviewers comments carefully. I propose to include this in the discussion section, e.g. as individual subsections.

For example the reviewers concerns of

- Accuracy: In fact, literature precedent suggests that they are less accurate and when coupled together will have a greater, unexplained uncertainty.
- Choice of NN: if the deposition component of a CTM is not a computational bottleneck and you replace it with a NN solver that is less accurate, then what is the utility of that?
- Error identification: How do you move beyond a 'black box' model?

Response: Thank you for your guidance to deepen the discussion of key debates raised by reviewers. We have revised the Discussion section by adding three dedicated subsections (4.1-4.3) to explicitly address the core concerns: model accuracy/uncertainty, the rationale for neural network (NN) component selection, and interpretability beyond "black-box" limitations. These sections integrate perspectives from the reviewers and our responses, ensuring the scientific debate is thoroughly documented. The revised Discussion section are shown as follows,

"***4 Discussions***

4.1 Model Accuracy and Uncertainty

One debatable concern is the accuracy of neural network (NN)-based components in integrated chemical transport models (CTMs) and the potential for amplified uncertainty when coupling multiple NN modules. Literature precedent suggests that individual NN emulators may exhibit lower accuracy compared to traditional physical parameterizations, but their integration could introduce unexplained uncertainties. This is a valid consideration that aligns with broader discussions in Earth system modeling about the trade-offs between computational efficiency and physical fidelity (Irrgang et al., 2021).

In FastCTM, we address this by adopting a principle-informed modular design where each module (transport, chemistry, deposition, etc.) is constrained by governing physical/chemical equations (e.g., Eqs. 3-14). This distinguishes it from unconstrained "black-box" NN models, as each process is guided by known atmospheric dynamics. For example, the transport module explicitly enforces

mass conservation via upwind schemes (Eqs. 5-7), and the chemical reaction module links reaction rates to meteorological conditions (Eq. 12) based on kinetic theory. Our evaluation shows that FastCTM maintains high consistency with CMAQ across 119-hour forecasts (Section 3.1), with R^2 values exceeding 0.8 for most pollutants, indicating that physical constraints effectively mitigate accuracy losses.

However, we acknowledge that uncertainty can accumulate when coupling modules, particularly for species involved in complex multi-process interactions due to limited chemical constraints in our current training datasets(e.g., NH_4^+, Section 3.1). This is partly due to simplifications in FastCTM's chemical mechanism, which omits some aerosol thermodynamics included in CMAQ. Future work will reduce such uncertainties by incorporating additional species (e.g., VOCs) and refining process formulations by adding CMAQ's integrated process rate (IPR) data for supervised training of individual modules.

4.2 Choosing Neural Network Components over Traditional Parameterizations

One question might arise about the utility of replacing non-bottleneck CTM components (e.g., deposition) with NN solvers, given the argument that traditional parameterizations may already be accurate and fast. This highlights a critical design choice in FastCTM: balancing computational efficiency with fidelity to the parent model (CMAQ).

It is important to note that even non-bottleneck components in traditional CTMs can benefit from NN acceleration in integrated simulations. For example, CMAQ's deposition module, while not a primary computational burden, relies on parameterizations based on similarity theory and limited flux measurements (Janhäll, 2015), which may oversimplify complex surface-atmosphere interactions (e.g., vegetation-specific uptake). NN-based parameterizations have shown promise in improving such processes. Silva et al. (2019), for instance, developed a deep learning model for ozone dry deposition that outperformed traditional schemes in independent validation. In FastCTM, the deposition module (Eq. 14) leverages NN to capture nonlinear relationships between meteorology (e.g., wind speed, land cover) and deposition rates, while retaining compatibility with CMAQ's output.

Moreover, FastCTM's modular architecture allows flexible integration of traditional parameterizations as an option. For example, users could replace the NN-based deposition module with CMAQ's original parameterization if higher fidelity to that specific process is prioritized. This hybrid approach addresses concerns about unnecessary replacement of robust components while retaining the overall speed advantage of NN for bottleneck processes (e.g., chemical reactions, which dominate CTM runtime; Xia et al., 2024).

4.3 Beyond "Black Boxes": Interpretability and Error Identification

A central goal of FastCTM is to advance beyond opaque deep learning models by enabling process-level interpretability, addressing concerns about error attribution. Traditional "black-box" NN models obscure how individual processes contribute to predictions, hindering error analysis. In contrast, FastCTM's modular design quantifies hourly contributions from transport, diffusion,

emissions, chemistry, and deposition separately (Section 3.3), allowing targeted identification of error sources. For example, in the January 2023 pollution episode (Figure 10), transport was found to dominate $PM_{2.5}$ concentration changes, while deposition acted as a secondary sink This process-level attribution aligns well with CMAQ's process analysis (Figure 11), ensuring that uncertainties are traced to specific physical processes rather than being attributed to arbitrary model behavior.

We anticipate that incorporating abundant CMAQ's integrated process rate (IPR) data for supervised training of individual modules will further refine the FastCTM's process level predictions. However, a comprehensive process-oriented error analysis that would further enhancing interpretability, for instance isolating and quantifying whether transport or chemistry drives urban-rural accuracy discrepancies, requires long-term process simulations and systematic perturbations plus observational datasets (e.g., tracer experiments) to validate specific processes predictions from both CMAQ and FastCTM.

4.4 Limitations and Future Directions

FastCTM's current limitations include simplified vertical dynamics (2D boundary layer representation) and incomplete chemical mechanisms, which affect performance during vigorous daytime mixing (Section 3.1). A future extension to a 3D framework will improve representation of vertical transport and in-cloud chemistry. Additionally, while FastCTM efficiently reproduces CMAQ simulations, it does not claim superiority over traditional CTMs across all scenarios; rather, it serves as a complementary tool for applications requiring rapid simulations (e.g., ensemble forecasting, emission scenario screening).

By addressing these limitations and engaging with ongoing debates about NN integration in atmospheric modeling, FastCTM aims to bridge the gap between computational efficiency and physical rigor, providing a flexible framework for air quality research and management."

Responses to Reviewer #1

I am grateful to the authors for their responses to my requests. The figures are much improved in particular, and I appreciate the additional statistical analysis. I believe that the manuscript is acceptable for publication pending copy-editing (there remain some typographical errors, e.g. line 406: "which could be caused by the reason that increased VOC"). I otherwise have no further comment.

Response: We appreciate your confirmation of the manuscript's readiness pending copy-editing. We have carefully revised the text to correct typographical errors. Thank you for your meticulous feedback.

Responses to Reviewer #2

The reviewers' replies to several of my comments were lackluster and did not address the main text

at all, instead opting to opine as a comment directly to me. Any reviewer comments I leave should be addressed in the text explicitly. The authors responded to my second comment about motivating this work in a poor manner. Each component or operator in a CTM has its own physics associated with it. Although the authors train individual NN operators with physics/chemistry constraints, this does not mean they are any more accurate than a traditional CTM. In fact, literature precedent suggests that they are less accurate and when coupled together will have a greater, unexplained uncertainty. For example, if the deposition component of a CTM is not a computational bottleneck and you replace it with a NN solver that is less accurate, then what is the utility of that? Having a "complete" deep learning CTM is ill-posed in a situation when specific CTM model components are accurate and fast. Further, as evidenced by the copious comments by Reviewer #1 concerning the presentation of results in a robust/quantitiatve manner, just because you incorporate physical constraints into the NN operators does not mean that they are more accurate/stable.

"However, implementing such a comprehensive process-oriented error analysis would require extensive retrospective simulations with systematic perturbation of individual modules, along with detailed validation against process-specific observational data (e.g., tracer studies, chamber experiments). This would demand substantial computational resources and time that extend beyond the scope of the current study."

--> This response is also lacking. The authors champion FastCTM as a tool that is able to circumvent the computational cost of a CTM, but then running a simple error analysis would "demand substantial computational resources". There is no need to run tracer studies or use chamber experiments but a simple error analysis visualized similarly to Figure 10 would be helpful. Otherwise, is this not the same as being a 'black box' model, which you claim to move beyond?

Response: We sincerely appreciate your detailed comments, which have helped strengthen the manuscript. We apologize for any previous oversights in addressing your concerns and have revised the main text extensively to incorporate our responses explicitly, particularly in the new subsections of the Discussion.

1. Accuracy and uncertainty of coupled NN modules

As highlighted in Section 4.1, we acknowledge that coupling NN modules may introduce uncertainties, as noted in the literature. To mitigate this, FastCTM adopts a principle-informed design where each module is constrained by physical/chemical equations (e.g., mass conservation in transport, kinetic theory in reactions). Our evaluation shows high consistency with CMAQ ($R^2 >$ 0.8 for most pollutants), but we also explicitly discuss limitations (e.g., lower R^2 for NH_4^+ due to simplified aerosol thermodynamics) and plans to reduce uncertainties using CMAQ's integrated process rate (IPR) data for supervised module training in the future.

2. Utility of NN for non-bottleneck components

Section 4.2 now addresses the utility of applying NNs to non-bottleneck components, clarifying our rationale on three fronts. First, we explain that even efficient traditional parameterizations have

known limitations (e.g., in deposition) where NNs can offer significant improvements. Second, we highlight that FastCTM is a flexible, hybrid framework that allows users to retain original CMAQ modules if they prioritize fidelity for a specific process. Finally, we emphasize that accelerating any component contributes to a crucial reduction in overall runtime, which is vital for computationally demanding applications like ensemble forecasting.

3. Beyond "black-box" models: interpretability and error analysis

Section 4.3 demonstrates that FastCTM's modular design enables process-level attribution (e.g., Figure 10 shows transport dominating $PM_{2.5}$ changes in a pollution episode). While comprehensive process-oriented error analysis (e.g., isolating urban-rural discrepancies) is beyond the current scope, we outline plans to use observational datasets (e.g., tracer studies) for such analyses, enhancing transparency.

4. Integration of responses into the main text

All key points from our responses are now incorporated into the revised manuscript, particularly in the Discussion subsections. We have ensured that all the concerns raised (e.g. uncertainty accumulation, "black-box" risks) are explicitly addressed in the text, not just in replies to reviewers. The revised texts in the Discussion section are shown as follows,

"***4 Discussions***

4.1 Model Accuracy and Uncertainty

One debatable concern is the accuracy of neural network (NN)-based components in integrated chemical transport models (CTMs) and the potential for amplified uncertainty when coupling multiple NN modules. Literature precedent suggests that individual NN emulators may exhibit lower accuracy compared to traditional physical parameterizations, but their integration could introduce unexplained uncertainties. This is a valid consideration that aligns with broader discussions in Earth system modeling about the trade-offs between computational efficiency and physical fidelity (Irrgang et al., 2021).

In FastCTM, we address this by adopting a principle-informed modular design where each module (transport, chemistry, deposition, etc.) is constrained by governing physical/chemical equations (e.g., Eqs. 3-14). This distinguishes it from unconstrained "black-box" NN models, as each process is guided by known atmospheric dynamics. For example, the transport module explicitly enforces mass conservation via upwind schemes (Eqs. 5-7), and the chemical reaction module links reaction rates to meteorological conditions (Eq. 12) based on kinetic theory. Our evaluation shows that FastCTM maintains high consistency with CMAQ across 119-hour forecasts (Section 3.1), with R^2 values exceeding 0.8 for most pollutants, indicating that physical constraints effectively mitigate accuracy losses.

However, we acknowledge that uncertainty can accumulate when coupling modules, particularly for species involved in complex multi-process interactions due to limited chemical constraints in our current training datasets(e.g., NH_4^+, Section 3.1). This is partly due to simplifications in FastCTM's chemical mechanism, which omits some aerosol thermodynamics included in CMAQ.

Future work will reduce such uncertainties by incorporating additional species (e.g., VOCs) and refining process formulations by adding CMAQ's integrated process rate (IPR) data for supervised training of individual modules.

4.2 Choosing Neural Network Components over Traditional Parameterizations

One question might arise about the utility of replacing non-bottleneck CTM components (e.g., deposition) with NN solvers, given the argument that traditional parameterizations may already be accurate and fast. This highlights a critical design choice in FastCTM: balancing computational efficiency with fidelity to the parent model (CMAQ).

It is important to note that even non-bottleneck components in traditional CTMs can benefit from NN acceleration in integrated simulations. For example, CMAQ's deposition module, while not a primary computational burden, relies on parameterizations based on similarity theory and limited flux measurements (Janhäll, 2015), which may oversimplify complex surface-atmosphere interactions (e.g., vegetation-specific uptake). NN-based parameterizations have shown promise in improving such processes. Silva et al. (2019), for instance, developed a deep learning model for ozone dry deposition that outperformed traditional schemes in independent validation. In FastCTM, the deposition module (Eq. 14) leverages NN to capture nonlinear relationships between meteorology (e.g., wind speed, land cover) and deposition rates, while retaining compatibility with CMAQ's output.

Moreover, FastCTM's modular architecture allows flexible integration of traditional parameterizations as an option. For example, users could replace the NN-based deposition module with CMAQ's original parameterization if higher fidelity to that specific process is prioritized. This hybrid approach addresses concerns about unnecessary replacement of robust components while retaining the overall speed advantage of NN for bottleneck processes (e.g., chemical reactions, which dominate CTM runtime; Xia et al., 2024).

4.3 Beyond "Black Boxes": Interpretability and Error Identification

A central goal of FastCTM is to advance beyond opaque deep learning models by enabling process-level interpretability, addressing concerns about error attribution. Traditional "black-box" NN models obscure how individual processes contribute to predictions, hindering error analysis. In contrast, FastCTM's modular design quantifies hourly contributions from transport, diffusion, emissions, chemistry, and deposition separately (Section 3.3), allowing targeted identification of error sources. For example, in the January 2023 pollution episode (Figure 10), transport was found to dominate $PM_{2.5}$ concentration changes, while deposition acted as a secondary sink This process-level attribution aligns well with CMAQ's process analysis (Figure 11), ensuring that uncertainties are traced to specific physical processes rather than being attributed to arbitrary model behavior.

We anticipate that incorporating abundant CMAQ's integrated process rate (IPR) data for supervised training of individual modules will further refine the FastCTM's process level predictions. However, a comprehensive process-oriented error analysis that would further enhancing interpretability, for instance isolating and quantifying whether transport or chemistry drives urban-

rural accuracy discrepancies, requires long-term process simulations and systematic perturbations plus observational datasets (e.g., tracer experiments) to validate specific processes predictions from both CMAQ and FastCTM.

4.4 Limitations and Future Directions

FastCTM's current limitations include simplified vertical dynamics (2D boundary layer representation) and incomplete chemical mechanisms, which affect performance during vigorous daytime mixing (Section 3.1). A future extension to a 3D framework will improve representation of vertical transport and in-cloud chemistry. Additionally, while FastCTM efficiently reproduces CMAQ simulations, it does not claim superiority over traditional CTMs across all scenarios; rather, it serves as a complementary tool for applications requiring rapid simulations (e.g., ensemble forecasting, emission scenario screening).

By addressing these limitations and engaging with ongoing debates about NN integration in atmospheric modeling, FastCTM aims to bridge the gap between computational efficiency and physical rigor, providing a flexible framework for air quality research and management."

Thank you again for your insightful feedback, which has significantly improved the rigor and clarity of our work.