Editor

General comments:

- ***Public justification (visible to the public if the article is accepted and published):***
  One reviewer further raised some critical comments for the revised manuscript. In particular, the reviewer was not satisfied with the presentation of the manuscript. Further improvement in the presentation of the manuscript is needed.
- ***Notification to the authors:***
  Coloured or marked text in *.pdf supplement file is not allowed. Please provide a clean version of the *.pdf supplement file (with black text) with the next revision. I would recommend you combine the coloured version of the supplement and the coloured version of the manuscript into one Author's tracked changes version of the files. This way you will be able to save data about changes both in the manuscript and in the supplement.

Dear Editors:

We sincerely appreciate your editorial effort and the constructive comments provided by the reviewer. We have carefully addressed all the concerns raised, particularly those regarding the presentation of the manuscript, to further improve its clarity and readability.

Additionally, as requested, we have provided a clean version of the supplementary file (with black text only) and merged the tracked changes from both the manuscript and the supplementary materials into a single Author's tracked changes file for clearer documentation of revisions.

Thank you for your time and valuable guidance. We hope the revised manuscript now meets the journal's standards.

Yours sincerely,

Zihan Xia and co-authors

Reviewer #1

Specific Comments:

- ***1. The term 'plug-and-play' is not scientific, rephrase or remove it as it sounds superfluous.***

Response: Thank you for your valuable feedback on the use of the term "plug-and-play" in our manuscript. We agree that the phrasing may come across as informal and have revised it to better align with scientific terminology. In the revised manuscript, we have replaced "plug-and-play nature" with "modular design" to more accurately convey the model's adaptability and ease of implementation within broader climate frameworks. The updated text in the Abstract (Lines 27-29) now reads:

"AIMACI's generalization capability and its modular design suggest potential for future coupling into global climate models, which are expected to enhance the precision and efficiency of aerosol simulations in climate modeling that neglects or simplifies ACI processes."

- ***2. "After that, the MOSAIC scheme is employed to address aerosol chemistry and interactions. The detailed processes included in aerosol chemistry and interactions are chemical reactions and phase equilibrium, gas-particle partitioning, particle size growth, coagulation, and nucleation, as explained in my response to your previous review comment No. 3."***

  -->This is still vague. your review comments are directed at me but should be directed towards the text. You write that "Aerosol chemistry and interactions (ACI) involve a range of highly nonlinear processes, including chemical reactions and phase equilibrium, gas-particle partitioning, particle size growth, coagulation, and nucleation ...", so aerosol chemistry is the entire WRF-Chem chemistry model? Are you just emulating and replacing MOSAIC? You need to explicitly state what components you are replacing instead of referring to general physical/chemical processes like coagulation.

Response: Thank you for highlighting the need for clarity regarding the scope of AIMACI within the WRF-Chem. We apologize for any ambiguity in our previous manuscript version.

As shown in Figure 1, in the WRF-Chem chemistry model, the Aerosol Chemistry and Interactions (ACI) is one of several components (e.g., emissions, photolysis, photochemistry, cloud chemistry, etc.). The full ACI consists of multiple subprocesses, including chemical reactions, phase equilibrium, gas-particle partitioning, particle size growth, coagulation, and nucleation. The MOSAIC scheme within WRF-Chem can be used to simulate the full ACI. In our study, the chosen MOSAIC scheme features four discrete size bins and primarily focuses on processes related to inorganic aerosols. It also considers the impact of marine biogenic sources of dimethyl sulfide on atmospheric aerosols and some aqueous reactions. However, secondary organic aerosols and complex heterogeneous chemical processes (e.g., oxidation of dissolved S(IV) by H_2O_2, O_3, NO_2 and O_2 catalyzed

by transition metal ions (TMI) in aerosol water, (Ruan et al., 2022)) are not included in the chosen MOSAIC scheme in this study. These aspects will be considered in future development. The AIMACI scheme developed by us is used to achieve end-to-end simulation of the full ACI within a 3D atmospheric numerical model, replacing the chosen MOSAIC numerical scheme to improve computational efficiency.

To emphasize this point, we have revised the relevant text and added a pseudo-code diagram, as follows:

- Section 1 (Lines 37-45)

Full aerosol chemistry and interactions (ACI) consists of multiple highly nonlinear subprocesses, including chemical reactions and phase equilibrium, gas-particle partitioning, particle size growth, coagulation, and nucleation, which have a significant impact on the concentration of atmospheric aerosols (Zaveri et al., 2008). Numerical models stand as indispensable analytical tools, pivotal for comprehending the aforementioned phenomena, and are instrumental in air quality management and the formulation of mitigation strategies for climate change. However, simulating full ACI processes within these models pose a significant computational challenge (Carmichael et al., 1999; Ebel et al., 2006). As quantified in Figure S1, using the Model for Simulating Aerosol Interactions and Chemistry (MOSAIC) scheme with four bins for full ACI simulation in the Weather Research and Forecasting with Chemistry (WRF-Chem) model accounts for 31.4% of the total computational time of the chemistry module.

- Section 2.1 (Lines 115-129)

The conventional numerical scheme adopted in this study for full ACI simulation is the MOSAIC scheme (Zaveri et al., 2008), coupled with the CBM-Z (Carbon Bond Mechanism Version Z) photochemistry scheme (Zaveri and Peters, 1999). The MOSAIC scheme stands out for its innovative approach to address the long-standing issues in solving the dynamic partitioning of semivolatile inorganic gases (HNO_3, HCl, and NH_3) to size-distributed atmospheric aerosol particles. It has been validated against a benchmark model version utilizing a rigorous solver for the integration of stiff differential equations, demonstrating both computational efficiency and high fidelity (Zaveri et al., 2008). The MOSAIC scheme used in this study features four discrete size bins (0.039~0.156, 0.156~0.625, 0.625~2.5, and 2.5~10.0 μm in diameter) and treats all the major aerosol species important at urban, regional, and global scales, including sulfate (SO_4^{2-}), nitrate (NO_3^-), chloride (Cl^-), carbonate (CO_3^{2-}), ammonium ($NH4^+$), sodium (Na^+), calcium (Ca^{2+}), black carbon (BC), organic carbon (OC), other inorganic mass (OIN), mineral dust, methanesulfonic acid (MSA) and liquid water content of aerosol (Water). It also considers the impact of marine biogenic sources of dimethyl sulfide on atmospheric aerosols and some aqueous processes. The chemical reactions among various species are detailed in Zaveri et al. (2008). However, secondary organic aerosols (SOA) and complex heterogeneous chemical processes (e.g., oxidation of dissolved S(IV) by H_2O_2, O_3, NO_2 and O_2 catalyzed by transition metal ions (TMI) in aerosol water, (Ruan et al., 2022)) are not included in the chosen MOSAIC scheme in this study. These aspects will be considered in future development.

- Section 2.2.1 (Lines 139-149)

In this study, for the first time, we attempt to use an AI algorithm to achieve end-to-end simulation of the full ACI within a 3D atmospheric numerical model, replacing the chosen MOSAIC numerical scheme to improve computational efficiency. Given the complexity of multiple subprocesses involved in full ACI, there is a clear need for AI algorithms with superior representational capacity for nonlinear systems. The Multi-Head Self-Attention (MHSA) mechanism, a pivotal component of state-of-the-art transformer architectures, has demonstrated exceptional performance across diverse domains such as Natural Language Processing (Vaswani et al., 2017), Computer Vision (Liu et al., 2021), and Weather Forecasting (Bi et al., 2023). Additionally, in the development of an Artificial Intelligence PhotoChemistry scheme, Xia et al. (2024) have highlighted that the MHSA algorithm excels in capturing the intricate chemical relationships among different species through calculating attention weights. It offers not only high accuracy and computational efficiency but is also less susceptible to the increase in the number of chemical species. Building upon these advancements, this study leverages the MHSA algorithm to develop the Artificial Intelligence Model for Aerosol Chemistry and Interactions (AIMACI).

- Section 2.2.1 (Lines 150-174)

Figure 1 illustrates the development strategy of the AIMACI scheme and its integration into our hybrid atmospheric model with physics and AI schemes (physics-AI hybrid model). The corresponding pseudocode schematic diagram for replacing the traditional numerical scheme with the AIMACI scheme in WRF-Chem is provided in Figure S2. The AI model architecture shown in Figure 1 is designed with three main components, each serving a distinct function in the simulation process: (1) Input Embedding Layer: This initial layer receives meteorological variables and chemical species as input features. The input embedding layer is designed as a fully connected layer, which maps the input data into a higher-dimensional space where interdependencies between variables can be more effectively captured. (2) Integrator: As the core of the AI model, it is composed of 2 identical blocks, each of which contains two sub-layers: an attention layer and a feed-forward layer. We apply residual connections around each of these two sub-layers, followed by layer normalization. This integrator is responsible for learning the complex and high nonlinear processes of ACI within the data and integrate them over time. (3) Output Representation Layer: Following the integrator, it also implemented as a fully connected layer. This layer translates the processed information from the integrator into chemical concentrations, providing the output targets for the simulation. Furthermore, the AI model is complemented by pre-processing and post-processing steps, such as min-max normalization, to constitute the AIMACI scheme. The trained AIMACI scheme was packaged into a static (or dynamic) library and then coupled into WRF-Chem, utilizing TorchScript and Libtorch tools officially provided by PyTorch. Compared to coupling approaches relying on third-party libraries such as CFFI (C Foreign Function Interface for Python), our method demonstrates two distinct technical advantages: (1) Enhanced Computational Efficiency: TorchScript and LibTorch, as core components of the PyTorch ecosystem, are specifically optimized for AI model deployment in C++ environments. This optimization reduces computational overhead compared to generic third-party libraries. (2) Streamlined Implementation: PyTorch's official documentation provides standardized workflows for serializing trained AI models into static or dynamic libraries via

TorchScript and LibTorch. In contrast, third-party libraries lack native support for AI model deployment, necessitating manual reimplementation of low-level interfaces. Therefore, our coupling approach minimizes alterations to the original codebase and offers a lightweight, adaptable, and modular solution. It is capable of encapsulating a wide range of complex AI algorithms and coupling them with diverse atmospheric and climate models.

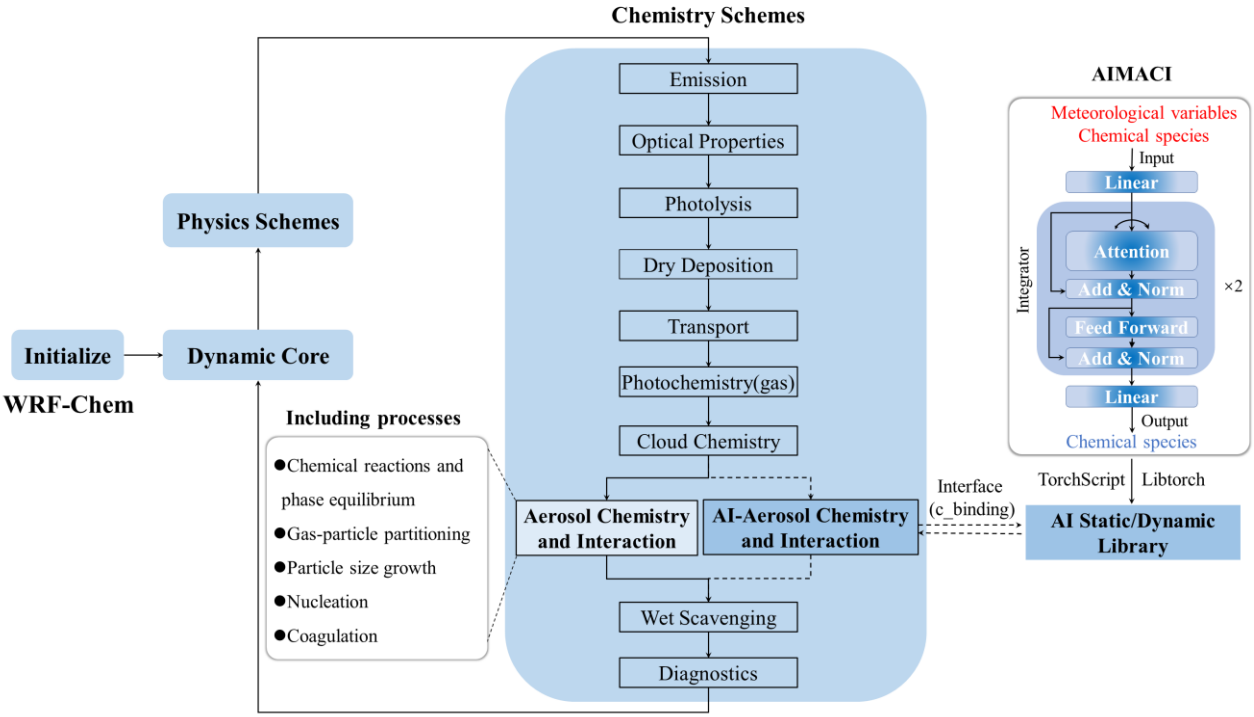

Figure 1: The Artificial Intelligence Model for Aerosol Chemistry and Interactions (AIMACI) in the Weather Research and Forecasting with Chemistry (WRF-Chem). The trained AIMACI is packaged into a static/dynamic library using TorchScript and Libtorch, and can be called by WRF-Chem through an interface to replace the numerical scheme for simulating aerosol chemistry and interactions, while the remaining processes maintain the original numerical scheme.

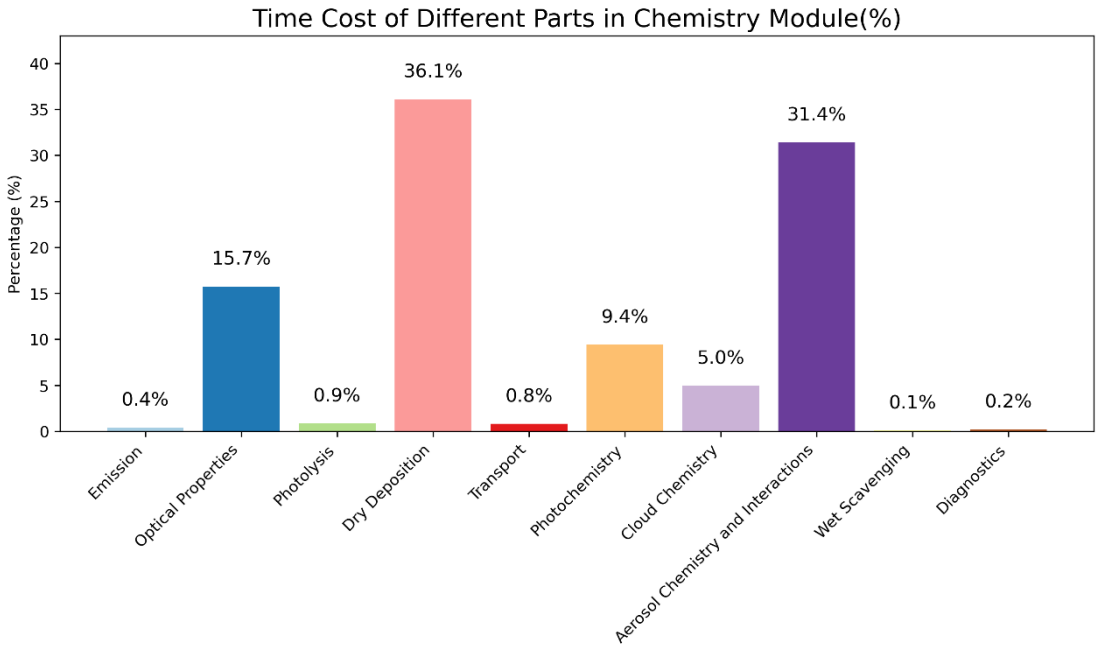

Figure S1: Proportion of computational time for different parts of the chemistry module in this study. The photochemistry is modeled using the CBM-Z (Carbon Bond Mechanism Version Z) scheme, which does not account for gas species related to complex SOA, while aerosol chemistry and interactions are simulated using the Model for Simulating Aerosol Interactions and Chemistry (MOSAIC) scheme, which primarily focuses on inorganic aerosol species and utilizes a 4-bin configuration.

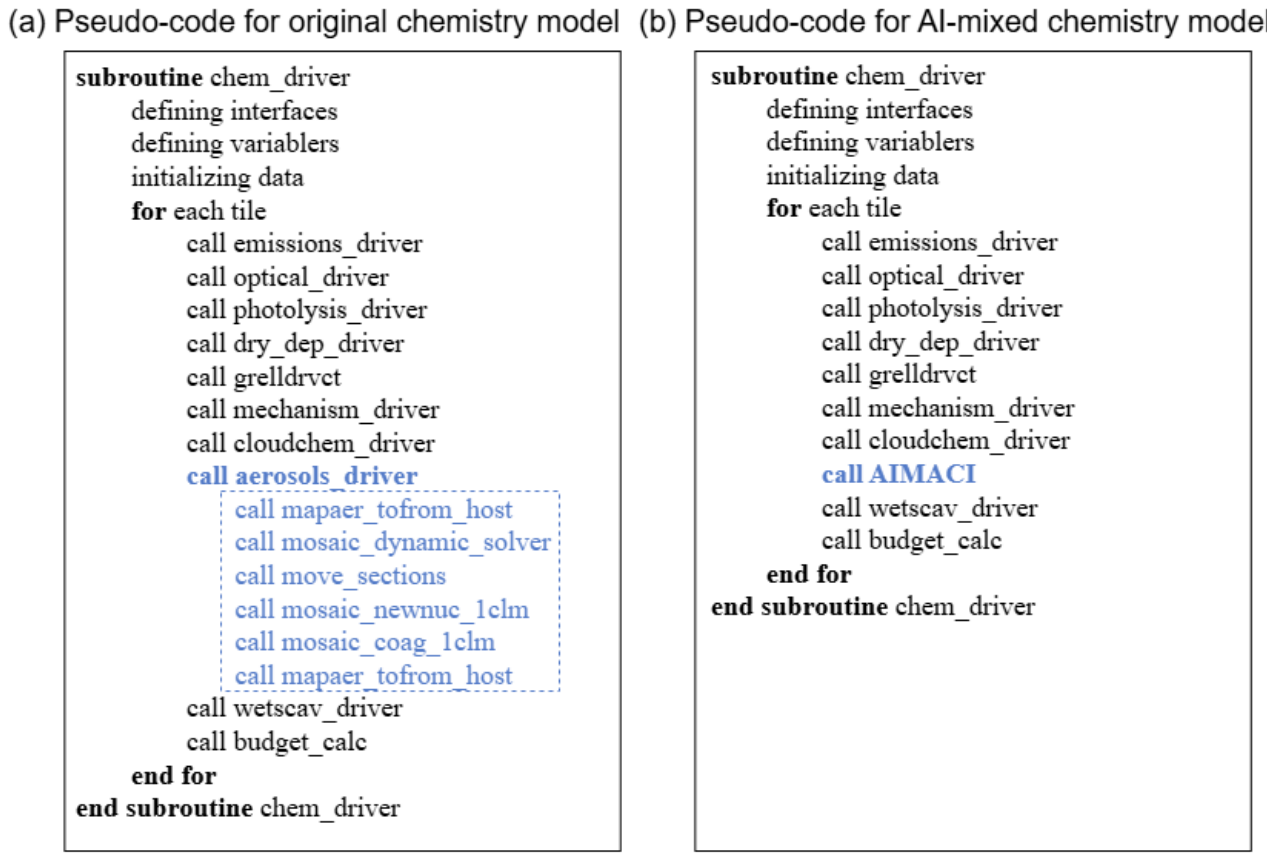

Figure S2: Schematic diagram of the specific components replaced by the AIMACI scheme in WRF-Chem. (a) The pseudo code for original chemistry model in WRF-Chem. The part surrounded by the blue dotted box is the key subroutine inside aerosols_driver. (b) The pseudo code for AI-mixed chemistry model in WRF-Chem. The AIMACI scheme enables end-to-end simulation of full aerosol chemistry and interactions processes, including chemical reactions, phase equilibrium, gas-particle partitioning, particle size growth, coagulation, and nucleation. Note that in our study, the chosen MOSAIC scheme features four discrete size bins and primarily focuses on processes related to inorganic aerosols. It also considers the impact of marine biogenic sources of dimethyl sulfide on atmospheric aerosols and some aqueous reactions. However, secondary organic aerosols and complex heterogeneous chemical processes are not included in the chosen MOSAIC scheme. These aspects will be considered for future development.

- ***3. "As the first step, this study focuses on inorganic aerosols, because the chemistry of organic aerosols (i.e., secondary organic aerosols) still has large uncertainties and lacks a convincing numerical scheme for AI scheme to emulate" --> There are many SOA numerical schemes that exist. I do not know how you can say there is a lack of a convincing scheme.***

Response: Thank you for raising this important point and sorry for the inappropriate statement. We acknowledge that there are existing numerical schemes for secondary organic aerosols (SOA) that have been widely implemented in atmospheric models. Our original phrasing regarding the "lack of a convincing scheme" was imprecise and has been revised to better reflect the motivation for prioritizing inorganic aerosols in this initial study. In the revised manuscript (Section 1, Lines 97-99), the text now states:

"As a first step, this study primarily focuses on inorganic aerosols because they constitute significant amounts of secondary aerosols globally and serve as the important driver of aerosol radiative forcing and cloud condensation nuclei activity in current climate models. Given that the production of secondary organic aerosols (SOA) involves significantly more complex chemical pathways, encompassing a wider array of precursor species and heterogeneous reaction mechanisms, the current AIMACI does not include them. Their incorporation will be considered in the future development of AIMACI."

- ***4. " its utilization in ACI simulation represents a new horizon. The algorithm's ability to globally attend to input variables and conduct parallel computations across multiple heads is pivotal in tackling the challenges posed by the curse of dimensionality, capturing complex interdependencies, and significantly enhancing computational efficiency."***

  Again, tone down this language. This is a scientific paper, not an infomercial. The results are promising but it's more of a demo than a useable model, only trained on a tiny amount of useable data.

Response: Thank you for your valuable feedback. We have removed this description and revised the relevant text to reflect a more measured and scientific tone. The revised text now reads as follows (Section 2.2.1, Lines 139-149):

"In this study, for the first time, we attempt to use an AI algorithm to achieve end-to-end simulation of the full ACI within a 3D atmospheric numerical model, replacing the chosen MOSAIC numerical scheme to improve computational efficiency. Given the complexity of multiple subprocesses involved in full ACI, there is a clear need for AI algorithms with superior representational capacity for nonlinear systems. The Multi-Head Self-Attention (MHSA) mechanism, a pivotal component of state-of-the-art transformer architectures, has demonstrated exceptional performance across diverse domains such as Natural Language Processing (Vaswani et al., 2017), Computer Vision (Liu et al., 2021), and Weather Forecasting (Bi et al., 2023). Additionally, in the development of an Artificial Intelligence PhotoChemistry scheme, Xia et al. (2024) have highlighted that the MHSA algorithm excels in capturing the intricate chemical relationships among different species through calculating attention weights. It offers not only high accuracy and computational efficiency but is also less susceptible to the increase in the number of chemical species. Building upon these advancements, this study leverages the MHSA algorithm to develop the Artificial Intelligence Model for Aerosol Chemistry and Interactions (AIMACI)."

- ***5. "After training, the AIMACI scheme was flexibly coupled into WRF-Chem"***

  -->Not sure what flexibly means here, vague

Response: Thank you for your comments. What we intended to convey is that the modular design and lightweight interface of the AIMACI scheme make it easy to couple with WRF-Chem. Considering that we have introduced the advantages of our coupling method over other third-party libraries in detail, we delete "flexibly" here and revised the relevant text, as follows (Section 2.2.1, Lines 163-173):

"The trained AIMACI scheme was packaged into a static (or dynamic) library and then coupled into WRF-Chem, utilizing TorchScript and Libtorch tools officially provided by PyTorch. Compared to coupling approaches relying on third-party libraries such as CFFI (C Foreign Function Interface for Python), our method demonstrates two distinct technical advantages: (1) Enhanced Computational Efficiency: TorchScript and LibTorch, as core components of the PyTorch ecosystem, are specifically optimized for AI model deployment in C++ environments. This optimization reduces computational overhead compared to generic third-party libraries. (2) Streamlined Implementation: PyTorch's official documentation provides standardized workflows for serializing trained AI models into static or dynamic libraries via TorchScript and LibTorch. In contrast, third-party libraries lack native support for AI model deployment, necessitating manual reimplementation of low-level interfaces. Therefore, our coupling approach minimizes alterations to the original codebase and offers a lightweight, adaptable, and modular solution. It is capable of encapsulating a wide range of complex AI algorithms and coupling them with atmospheric and climate models."

- ***6. "Secondly, in terms of compatibility, LibTorch supports cross-platform deployment, offering more flexibility than third-party libraries. "***

  --> This does not sound scientific, sounds like a commercial. Cross-platform is vague in the context of WRF-Chem and scientific computing

Response: Thank you for pointing out this. We have deleted this description and adjusted the relevant text, as follows (Section 2.2.1, Lines 165-173):

"Compared to coupling approaches relying on third-party libraries such as CFFI (C Foreign Function Interface for Python), our method demonstrates two distinct technical advantages: (1) Enhanced Computational Efficiency: TorchScript and LibTorch, as core components of the PyTorch ecosystem, are specifically optimized for AI model deployment in C++ environments. This optimization reduces computational overhead compared to generic third-party libraries. (2) Streamlined Implementation: PyTorch's official documentation provides standardized workflows for serializing trained AI models into static or dynamic libraries via TorchScript and LibTorch. In contrast, third-party libraries lack native support for AI model deployment, necessitating manual reimplementation of low-level interfaces. Therefore, our coupling approach minimizes alterations to the original codebase and offers a lightweight, adaptable, and modular solution. It is capable of encapsulating a

wide range of complex AI algorithms and coupling them with atmospheric and climate models."

- ***7. "The average NMB for these species is 3.02%, reflecting only a slight deviation from the numerical scheme's outcomes and highlighting the AIMACI scheme's impressive accuracy in simulating ACI. However, as shown in Table 3, some species still exhibit relatively poorer statistical indicators compared to others, such as carbonates. "***

  --> The manuscript still has way too many value judgments throughout using the terms 'impressive' and 'promising'. You say something is impressive and then the next line you say some parts do poorly. Please take out these instances, they are not scientific and detract from the quality of the paper. Some elements are indeed impressive, but at the end of the day, you are using a relatively simple transformer model trained on 2 weeks of data which does not feel particularly 'impressive' from the ML-weather and ML-NLP communities. Please temper this kind of unscientific language

Response: We sincerely appreciate your rigorous critique regarding the use of subjective language in our manuscript. We fully agree that value judgments like "impressive" and "promising" detract from scientific objectivity. We have systematically revised the text to eliminate such terminology and reframe the discussion around quantitative metrics, as follows (Section 3.1, Lines 265-269):

"Quantitative evaluation across 61 output targets demonstrates close alignment between the simulations using the AIMACI scheme and the MOSAIC scheme (hereinafter referred to as the numerical scheme), as evidenced by an average R^2 of 0.98 and an average NMB of 3.02%. While 55 output targets, including major inorganic aerosols such as sulfate, nitrate, and ammonium, achieve R^2 values $\geqslant$0.95, there are still a few species exhibit relatively poorer statistical metrics, particularly carbonates (Table 3)."

- ***8. "From the absolute error figures, it is observed that for each particle size, AIMACI tends to underestimate the higher concentration regions and overestimate lower values, particularly those near zero."***

  -->Same findings in Kelp et al. (2022) which uses much simpler ML methods. Why do you mention the loss function (RMSE)? What loss can you change it to for better results?

Response: Thanks for your insightful comment. The systematic bias pattern observed in our results indeed aligns with the findings of Kelp et al. (2022), who reported similar behavior in MSE-optimized emulators. The tendency of the AI model to underestimate high values and overestimate low values can be related to the mathematical formulation of the RMSE (or MSE) loss function. Specifically, the quadratic term in RMSE (or MSE) penalizes large errors more heavily, which biases the model towards predicting the mean of the target distribution to minimize the overall loss (Gneiting, T., 2011).

While alternative loss functions, such as Huber loss, could potentially mitigate this bias, RMSE (or MSE) remains widely used in autoregressive simulation studies. One of the key rationales stems from its inherent capability to effectively prevent

large errors, which is critical for ensuring the stability of long-term simulations. To alleviate above bias pattern, we plan to explore data transformation techniques to reduce the skewness of the training data distribution. This approach may help alleviate the bias without compromising the stability of the model.

We have revised the relevant text to clarify this point, as follows (Section 3.2.1, Lines 318-337):

"Figure 3 illustrates the spatial distribution of sulfate column concentrations across different size bins at the end of the 10-day continuous simulation (i.e., 2019-03-30 00:00 UTC), accompanied by the temporal evolution of RMSE during the simulation period. The results reveal that the high-value areas of sulfate column concentrations for different particle sizes exhibit a hook-like structure, The results reveal that the high-value areas of sulfate column concentrations for different particle sizes exhibit a hook-like structure, extending northeastward from the Yangtze River Economic Belt to the northeastern regions of China. The distinct patterns may be attributed to the complex interplay of meteorological conditions, emission sources, and atmospheric transport processes. The sulfate column concentrations are predominantly concentrated within the 0.156 to 0.625 μm (size bin 2), with relatively lower column concentrations in the 2.5 to 10 μm (size bin 4), which is consistent with the findings in Figure 2. AIMACI reproduces these spatial distribution patterns with strong fidelity, as evidenced by R^2 exceeding 0.88 across all size bins. However, systematic bias pattern is observed that for each particle size, AIMACI tends to underestimate the higher concentration regions and overestimate lower values. This aligns with the findings of Kelp et al. (2022), who report similar bias pattern in MSE-optimized emulators for photochemistry simulation. We attribute this bias pattern to two main factors: (1) the imbalance in the training dataset, where high-value samples are underrepresented compared to low-value samples, leading to insufficient learning of high-value instances by the AI model; and (2) the use of the RMSE (or MSE) loss function, whose quadratic term penalizes large errors more heavily, thereby biasing the model towards predicting the mean of the target distribution to minimize the overall loss (Gneiting, T., 2011). While alternative loss functions (e.g., Huber loss) may reduce this bias, RMSE (or MSE) remains widely used in autoregressive simulation studies. One of the key rationales stems from its inherent capability to effectively prevent large errors, which is critical for ensuring the stability of long-term simulations. To alleviate above bias pattern, we plan to explore data transformation techniques to reduce the skewness of the training data distribution in future work. This approach may help alleviate the bias without compromising the stability of the AIMACI."

Reference

Gneiting, T.: Making and Evaluating Point Forecasts, Journal of the American Statistical Association, 106, 746–762, https://doi.org/10.1198/jasa.2011.r10138, 2011.

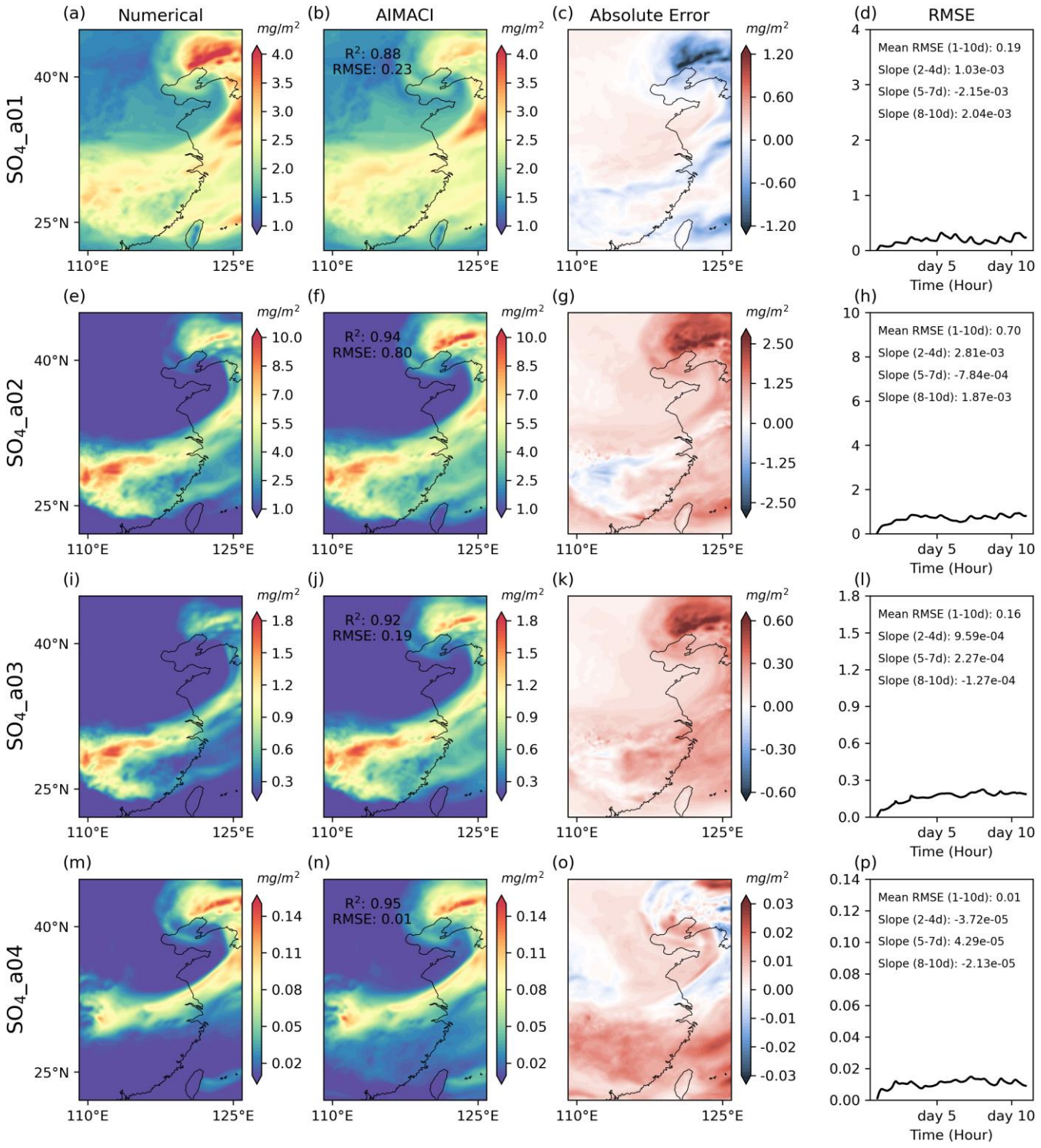

Figure 3: Sulfate column concentration simulations across different size bins. The first and second column depict the spatial distribution at the 10-day continuous simulation's end (2019-03-30 00:00 UTC), as simulated online by both the numerical scheme and the AIMACI scheme. The third column is the absolute error between them. The fourth column shows the temporal evolution of the hourly RMSE over the 10-day period. The mean RMSE (unit: mg /m^2) for all days and the slope (unit: mg m^{-2} h^{-1}) for different simulation stages (2-4day, 5-7day, 8-10day) are given inset.

- ***9. "Additionally, the results shown are from the last time step of a 10-day continuous simulation, and the simulation errors could be influenced not only by the biases of a single simulation instance but also by potential inaccuracies in the inputs at that time step"***

  --> Vague reason, what do inaccuracies in the inputs at that time step even mean? The ground truth is the WRF-Chem simulation

Response: Thank you for your comment. In Figure 3, we compare the results of the last time step from a 10-day continuous simulation using the AIMACI scheme and the MOSAIC scheme. While both AIMACI and MOSAIC schemes operate within the same WRF-Chem framework with identical meteorological/emission inputs initially, there are slight deviations in the concentrations of different chemical species simulated by the two schemes during the continuous simulation. Therefore, there are certain differences in the inputs of the two schemes in the last time step of the simulation. This is what we originally intended to convey by "potential inaccuracies in the inputs at that time step." However, we agree that this explanation is vague and may not be necessary, as the ground truth for comparison is indeed the WRF-Chem simulation itself. Therefore, we have removed this statement from the manuscript to improve clarity and focus on the primary findings.

- ***10. "This complex error variation may be related to the online simulation approach, as the aerosol concentrations simulated by the AIMACI scheme are subject to other processes in the numerical model such as dry deposition, wet scavenging."***

  --> You cannot say this without showing this. It comes off as deflecting the deficiencies of the ML model.

Response: Thank you for pointing this out. We have reorganized the language in this section to acknowledge the limitations of the AIMACI scheme and analyze the possible reasons for the complex error variation, as shown below (Section 3.2.1, Lines 338-354):

"Additionally, RMSE temporal evolution reveals that during the entire simulation period, it does not exhibit rapid growth but maintains oscillations within a constrained range. The analysis of the RMSE trend slopes across different simulation stages for four size bins demonstrates non-uniform error progression patterns (e.g., in Figure 3l, the slope for the 2–4 day period is 0.000959 mg m^{-2} h^{-1}, while the slope for the 8–10 day period is -0.000127 mg m^{-2} h^{-1}), where even within identical simulation phases, distinct size bins of the same species manifest divergent error trends (e.g., during the 8–10 day period, the slopes for different size bins are 0.00204 mg m^{-2} h^{-1}, 0.00187 mg m^{-2} h^{-1}, -0.000127 mg m^{-2} h^{-1}, and -0.0000213 mg m^{-2} h^{-1}, respectively). The emergence of this complex error variation is related to the dual influence governing each grid point errors in online continuous simulations: (1) The inherent limitations of the AIMACI scheme, which achieves accurate simulations when encountering well-learned input feature combinations, while exhibiting degraded performance under insufficiently trained input patterns; and (2) The compound error propagation mechanisms during continuous simulations, where input biases of species concentrations at each timestep are affected by three factors: local error accumulation from preceding steps, error

propagation through transport processes in numerical models from neighboring grids, and perturbations induced by other processes in numerical models (e.g., dry deposition and wet scavenging). Consequently, input biases exhibit nonlinear variability, with AIMACI's simulation accuracy being inversely correlated to input error magnitudes. In operational implementations of physics-AI hybrid models for online simulations, the influences of above two factors are often interrelated rather than independent, thereby amplifying the complexity of error variation. Introducing an error-correcting operators in the continuous simulation process may potentially enhance the stability of long-term simulations."

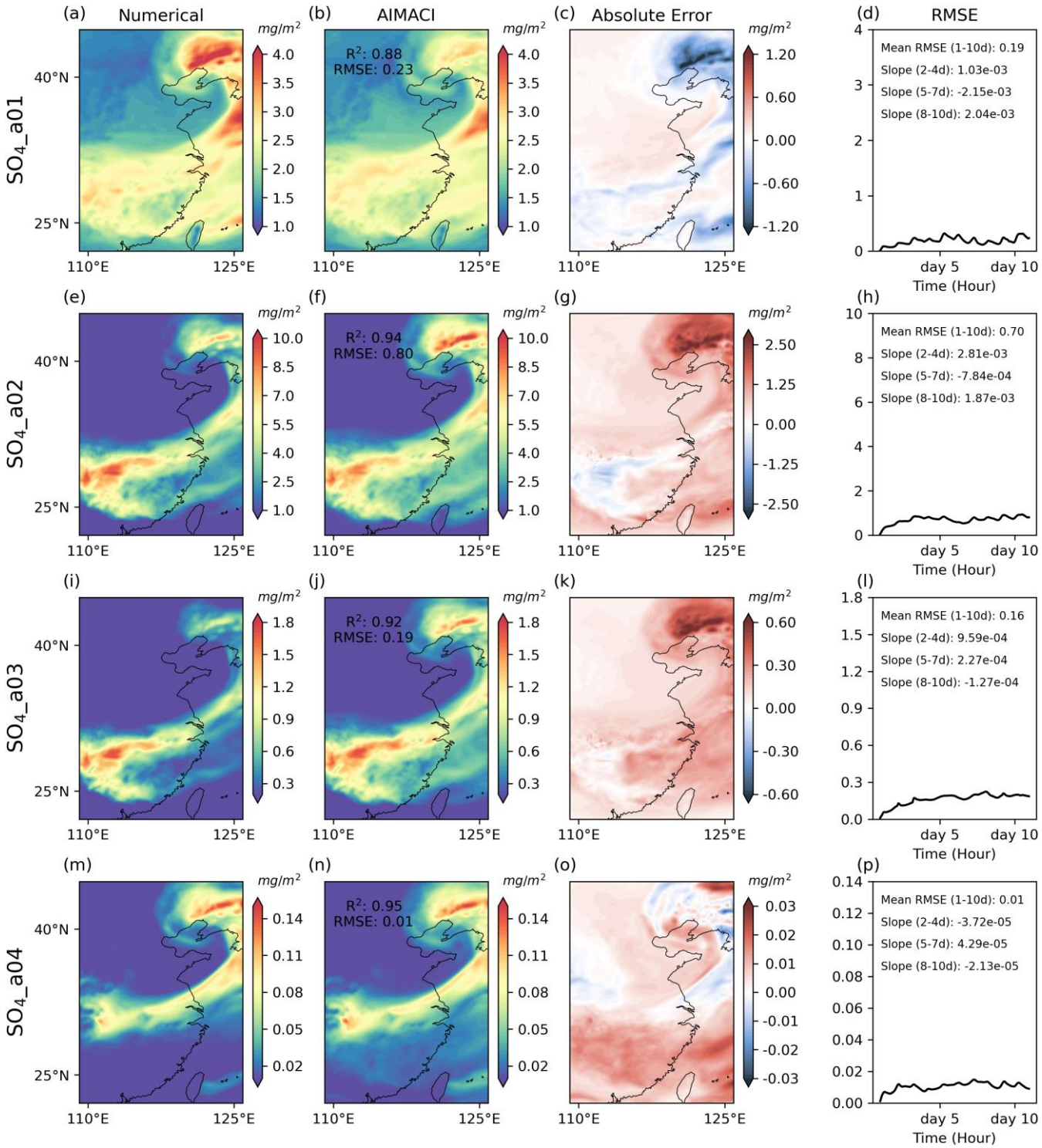

Figure 3: Sulfate column concentration simulations across different size bins. The first and second column depict the spatial distribution at the 10-day continuous simulation's end (2019-03-30 00:00 UTC), as simulated online by both the numerical

scheme and the AIMACI scheme. The third column is the absolute error between them. The fourth column shows the temporal evolution of the hourly RMSE over the 10-day period. The mean RMSE (unit: mg /m^2) for all days and the slope (unit: mg m^{-2} h^{-1}) for different simulation stages (2-4day, 5-7day, 8-10day) are given inset.

- ***11. Figure 4: These errors seem quite large, what are the effects of overestimating nitrate in this scheme? Are there more chemical reactions as a result? What are the chemical sinks of nitrate and MOSAIC and would this cause an error increase in other species?***

Response: Thank you for your insightful comment. You are correct that, if we only look at the relative error graph for nitrate (Figure 4h), we can see that the relative error at high altitudes is indeed large. We acknowledge that the original manuscript did not sufficiently explain this phenomenon. The observed behaviour arises because species concentrations at higher altitudes are typically below 1.0 μg/kg, and the relative error calculation divides the absolute error by these low concentration values. Consequently, even small absolute errors at high altitudes can result in significantly large relative errors. This effect is particularly pronounced for nitrate, which exhibits concentrations at higher altitudes that are one to two orders of magnitude lower than sulfate, often approaching zero. When considering the absolute error distribution (Figure 4g), the errors at higher altitudes are, in fact, relatively small.

However, we recognize that the AIMACI scheme does exhibit a tendency to overestimate nitrate concentrations. This overestimation can introduce biases in the input concentrations at each timestep during continuous simulations, potentially affecting the simulated concentrations of other species. When the overestimation is minor, its impact on other species is limited. In the atmosphere, the primary sinks of nitrate include: (1) Wet Deposition: Nitrate is removed from the atmosphere through precipitation (rain, snow); (2) Dry Deposition: Nitrate aerosols are deposited onto surfaces (e.g., vegetation, soil); (3) Photolysis and Re-release: Under specific conditions, nitrate aerosols may decompose and re-release NO_x. From the perspective of chemical sinks, underestimation of the photolysis process of nitrate could lead to overestimated nitrate concentrations. Additionally, from the perspective of chemical production, overestimation of the reaction between nitric acid and ammonia could also result in overestimated nitrate concentrations. However, in this study, our AIMACI scheme achieves end-to-end simulation of full aerosol chemistry and interactions. Therefore, it can't directly distinguish which process is causing the overestimation. Further investigation will be needed to identify the specific causes.

We have revised the relevant description of the manuscript to provide a clearer explanation of the significant relative errors observed for nitrate at high altitudes, as follows (Section 3.2.1, Lines 363-369):

"The absolute error exhibits a vertical gradient, with larger errors near the surface and smaller errors at higher altitudes. Conversely, the relative error distribution shows an inverse pattern, with lower relative errors near the surface and higher relative errors at elevated altitudes. This behavior arises because species concentrations at higher altitudes are typically below 1.0 μg/kg, and the relative error calculation divides the absolute error by these low concentration values. Consequently, even

small absolute perturbations at high altitudes can result in significantly large relative errors. This effect is particularly pronounced for nitrate, which exhibits concentrations at higher altitudes that are one to two orders of magnitude lower than sulfate, often approaching zero."

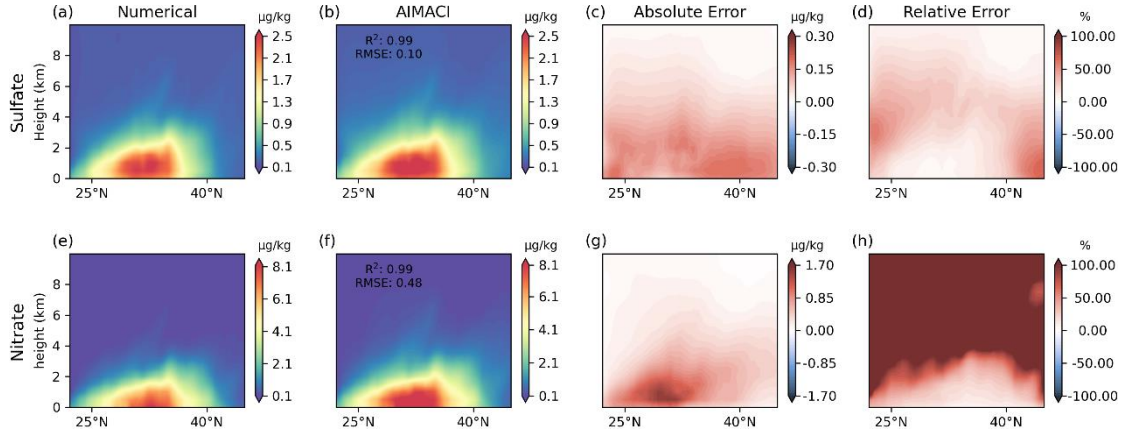

Figure 4: Zonal mean total concentrations (summed across 4 size bins) of sulfate and nitrate between 109.1 ℈ and 125.9 ℈, as simulated online by both the numerical scheme and the AIMACI scheme. Results are averages over the entire 10-day simulation period.

- ***12. The analysis with seasonal weather is nice, thank you for this type of discussion***

Response: Thank you for your positive feedback on the seasonal weather analysis.

- ***13. Figure 3. Slope misspelled***

Response: Thank you for carefully reviewing the manuscript and identifying this typographical error. The misspelled label "Slpoe" in Figure 3 has been corrected to "Slope" in the revised manuscript.

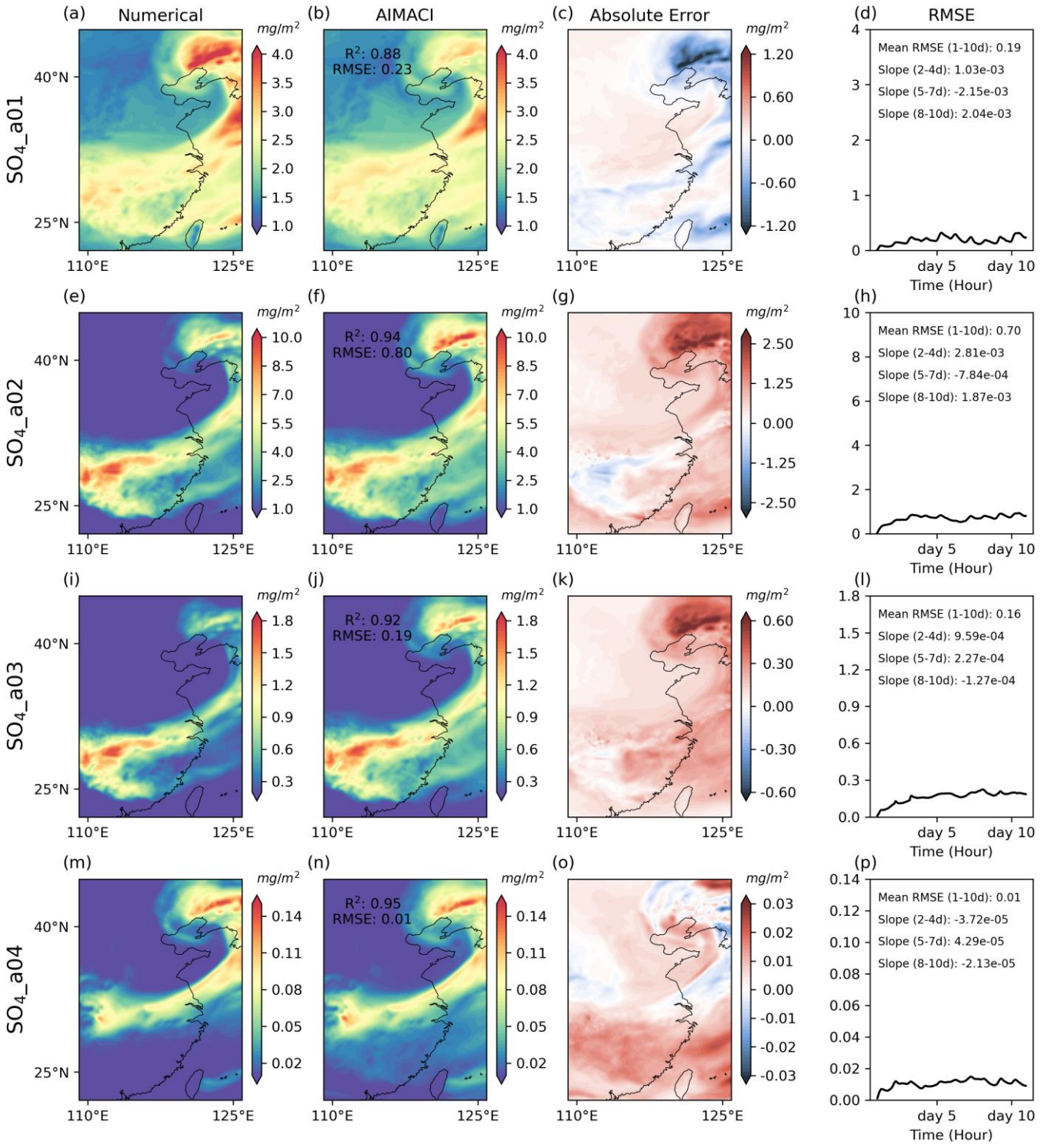

Figure 3: Sulfate column concentration simulations across different size bins. The first and second column depict the spatial distribution at the 10-day continuous simulation's end (2019-03-30 00:00 UTC), as simulated online by both the numerical scheme and the AIMACI scheme. The third column is the absolute error between them. The fourth column shows the temporal evolution of the hourly RMSE over the 10-day period. The mean RMSE (unit: mg /m^2) for all days and the slope (unit: mg m^{-2} h^{-1}) for different simulation stages (2-4day, 5-7day, 8-10day) are given inset.

- ***14. "In the development of our AIMACI scheme, we faced a bifurcation of choices: whether to input all features of a single grid point and predict for that grid point individually, followed by iterating through all grid points, or to input all features for the entire 3D grid space simultaneously and predict for the entire 3D space at once. Most current AI large models such as Pangu (Bi et al., 2023) and Fengwu (Chen et al., 2023), opt for the latter approach, which inevitably requires the use of convolutional networks... "***

  --> I am not sure why you do this, or your justification. Random forests and ANNs/LSTMs can work like this as you are basically creating a tabular data problem. Unless you are inserting some correlation between grids/points, I do not see the strength of using this MHSA. Maybe you can say that this approach is computationally simpler, but we know we can use FlashAttention or other techniques to alleviate such a bottleneck. It is true that ViTs perform very well, but without a counterfactual involved, I do not see the inherent benefit of using this ML model architecture unless you have tried others and failed.

Response: Thank you for your comment. We acknowledge that the approach of creating training data based on individual grid points is algorithm-agnostic, meaning that other algorithms such as random forests, ANNs/LSTMs can also perform grid-based predictions.

Our intention in the original manuscript was to highlight the benefits of choosing to use individual grid points as training samples over using the entire 3D grid space as a single sample when developing an AI scheme coupled with a numerical model: (1) Larger Training Dataset: By using individual grid points, we can generate a much larger volume of training data since there are numerous grid points within a single 3D grid space at any given moment. (2) Flexibility in Simulation Domain Size: When the AI scheme is trained on individual grid points samples, it can be applied to simulate any size domains by iterating calculations for each grid point. In contrast, if the AI scheme is trained using the entire 3D grid space as a single sample, the shape of the input array to the AI scheme would be fixed to the size of the training sample. This would restrict the flexibility in setting the size of simulation domain such as the number and resolution of grids. (3) Reduced Computational Load: Opting for individual grid points generally requires less computation. If the entire 3D grid space is used as a single sample, convolutional networks are often necessary, which can be computationally intensive.

However, considering that current related studies predominantly use grid-based training data and that this discussion may appear somewhat abrupt in the context of the section, we have removed this statement from the revised manuscript to maintain coherence and focus.

- ***15. "Despite these pronounced fluctuations"***

  --> You did not address the concerns of the other reviewer regarding this point. You offer a good explanation but that should be put directly in the text. I do not need any direct comments from the authors addressed to me, everything should be in the text.

Response: Thank you for your comment and sorry for not making the text clearer in the original manuscript. In the revised manuscript, we ensured that all responses were reflected in the text. In response to this comment, we have carefully reviewed our previous response and realized that it did not adequately address the concerns regarding the increase in fluctuation frequency of aerosol species concentrations after five days. The changes in local aerosol concentration are influenced by multiple factors, and the increased frequency of fluctuations may be related to the following reasons: (1) Variations in meteorological conditions, such as changes in wind speed and direction, can significantly influence the dispersion and transport of aerosols. Increased wind speed or frequent shifts in wind direction may lead to more pronounced concentration fluctuations. Additionally, changes in humidity can affect the wet scavenging and hygroscopic growth of aerosols. Precipitation events, in particular, can enhance the wet scavenging effect, leading to a significant reduction in aerosol concentrations. (2) Variations in anthropogenic emissions, such as differences in human activities during weekdays and weekends, can also contribute to changes in aerosol concentrations.

Given that the focus of this study is to evaluate the performance of the AIMACI scheme rather than to analyse the mechanisms behind the fluctuations in aerosol concentrations, we have not conducted a more detailed analysis to identify the specific causes. We have adjusted the relevant text to include a simple analysis of the generation of concentration fluctuations as follows (Lines 377-385):

"The occurrence of these pronounced fluctuations may be related to the following factors: (1) Variations in meteorological conditions, such as changes in wind speed and direction, can significantly influence the dispersion and transport of aerosols. Increased wind speed or frequent shifts in wind direction may lead to more pronounced concentration fluctuations. Additionally, changes in humidity can affect the wet scavenging and hygroscopic growth of aerosols. Precipitation events, in particular, can enhance the wet scavenging effect, leading to a significant reduction in aerosol concentrations. (2) Variations in anthropogenic emissions, such as differences in human activities during weekdays and weekends, can also contribute to changes in aerosol concentrations. A more detailed analysis would be required to identify the specific causes. Despite these pronounced fluctuations, the AIMACI scheme effectively reproduces these features without introducing systematic bias, achieving R^2 values larger than 0.97."