https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

Toward a Learnable Artificial Intelligence Model for Aerosol Chemistry and Interactions (AIMACI) based on the Multi-Head Self-Attention Algorithm

Zihan Xia [1], Chun Zhao [1,2,3*], Zining Yang [1], Qiuyan Du [1], Jiawang Feng [1], Chen Jin [1], Jun Shi [4], Hong An [2,4]

[1]Deep Space Exploration Laboratory/School of Earth and Space Sciences/CMA-USTC Laboratory of Fengyun Remote Sensing/State Key Laboratory of Fire Science/Institute of Advanced Interdisciplinary Research on High-Performance Computing Systems and Software, University of Science and Technology of China, Hefei, China
[2]Laoshan Laboratory, Qingdao, China
[3]CAS Center for Excellence in Comparative Planetology, University of Science and Technology of China, Hefei, China.
[4]School of Computer Science and Technology, University of Science and Technology of China, Hefei, 230026, China

Correspondence to: Chun Zhao (chunzhao@ustc.edu.cn)

Abstract. Simulating aerosol chemistry and interactions (ACI) is crucial in climate and atmospheric model, yet conventional numerical schemes are computationally intensive due to stiff differential equations and iterative methods involved. While artificial intelligence (AI) have demonstrated the potential in accelerating photochemistry simulations, it has not been applied for simulating the full ACI processes which encompass not only chemical reactions but also other processes such as nucleation and coagulation. To bridge this gap, we develop a novel Artificial Intelligence Model for Aerosol Chemistry and Interactions (AIMACI), focusing initially on inorganic aerosols. Trained based on conventional numerical scheme, it has been validated both offline and online coupled into three dimensional numerical atmospheric model. Results demonstrate that AIMACI are not only comparable to those with the conventional numerical scheme in spatial distributions, temporal variations, and evolution of particle size distribution of 8 aerosol species including water content in aerosols, but also exhibits robust generalization ability, reliably simulating one month under different environmental conditions across four seasons despite being trained on limited data from merely 16 days. Remarkably, it exhibits a ~5× speedup with a single CPU and ~277× speedup with a single GPU compared to conventional numerical scheme. While global long-term simulations have not yet been implemented, AIMACI's robust generalization capability, coupled with our easily plug-and-play solution, paves the way for its coupling into global climate models for further testing in near future. This advancement promises to enhance the precision and efficiency of atmospheric aerosol simulations in climate modeling.

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

1 Introduction

Atmospheric aerosols, which consist of a suspension of solid and liquid particles in the air, exert a profound influence on Earth's climate system and air quality (Charlson et al., 1992). Their multifaceted impacts are evident in their capacity to alter the Earth's radiation balance through the scattering and absorption of solar and longwave radiation, as well as in their role as cloud condensation nuclei that influence the formation and characteristics of clouds (Li et al., 2022; Twomey, 1974). The presence of atmospheric aerosols extends its reach to environmental well-being, with implications that span visibility, human health, and the integrity of ecological ecosystems (Arfin et al., 2023; Pöschl, 2005). 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, coupling aerosol chemistry and interactions into these models poses a significant computational challenge (Carmichael et al., 1999; Ebel et al., 2006). This is primarily due to the requirement to solve a complex set of stiff nonlinear differential equations governing aerosol processes, coupled with the use of implicit integration schemes to ensure numerical stability (Sandu et al., 1997a, b). Furthermore, to accommodate the diverse methodologies for describing the evolution of particle size distribution (PSD), some aerosol processes may require repeated calculations (Wang et al., 2022). For example, when employing a discrete model, the coagulation collision frequency functions need to be computed for each discrete size (Zhang et al., 2020). Consequently, the computational burden is significantly amplified. This computational intensity often creates a dilemma, as it competes with other priorities in numerical modeling, such as enhancing spatial resolution (Gu et al., 2022), recognized as helpful for minimizing uncertainties in numerical models. Numerous numerical models opt for simplified or even deactivate aerosol chemistry and interactions scheme during long-term simulations, particularly in high-resolution atmospheric and climate models, introducing considerable uncertainties into the simulation results (Lee et al., 2016; Zhang et al., 2020). Consequently, there is a pressing need to achieve rapid, accurate, and stable simulation of atmospheric aerosol chemistry and interactions within numerical models.

Over the past few decades, extensive research efforts have been dedicated to striking a tradeoff between the accuracy and computational efficiency in simulating aerosol chemistry and interactions. Researchers have primarily approached this challenge from two distinct perspectives: one is the exploration of various methodologies for describe the evolution of PSD. For instance, in the discrete model, the PSD is divided into discrete sizes, with calculations performed for each individual size. This approach yields the most precise results but also demands the highest computational resources (Landgrebe and Pratsinis, 1990; Zhao et al., 2013b). The moment model, which tracks the lower-order moments of an unknown aerosol distribution, is particularly well-suited for scenarios where the size distribution is lognormal (Pratsinis, 1988). Concurrently, researchers have been engaged in employing diverse methodologies to solve the system of stiff differential equations. For example, the Multicomponent Taylor Expansion Method (MTEM) has been developed to compute activity coefficients in aqueous atmospheric aerosols (Zaveri et al., 2005). This method offers an efficient non-iterative solution for systems rich in

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

sulfate aerosols. The Adaptive Step Time-split Euler Method (ASTEM) leverages several key characteristics of the atmospheric gas-particle partitioning, systematically reducing stiffness while preserving the integrity of the numerical solution (Zaveri et al., 2008). Despite these advancements significantly improving the computational efficiency of simulating aerosol chemistry and interactions, current progress remains far from sufficient.

An alternative approach is to utilize artificial intelligence (AI) schemes to replace conventional numerical schemes in atmospheric and climate models, which could potentially bring about a transformative impact. Recent studies by Liu et al. (2021) developed an AI scheme based on a Residual Neural Network (ResNet) algorithm for simulating atmospheric photochemistry, achieving a nearly 10.6× increase in computational efficiency. However, they adopted a hybrid approach, combining the numerical scheme for radicals and oxidants with the AI scheme for volatile organic compounds (VOCs). Kelp et al. (2022) employed an online training strategy to refine an AI scheme for a simplified Super-Fast chemistry scheme (12 species) in atmospheric models, achieving stable simulations over a year with a nearly 5× speedup. Yet, this method necessitated a complex preparation process, involving training 48 separate AI model for each chemical species across four seasons. Sharma et al. (2023) developed a physics-informed AI approach to study isoprene epoxydiols in acidic aqueous aerosols over the Amazon rainforest, halving computational costs but requiring training separate AI model for each size bin. Xia et al. (2024) have taken a step further by developing an AIPC scheme leveraging the Multi-Head Self-Attention algorithm to simulate a full complex atmospheric photochemistry with a unified AI model. When coupled with three-dimensional (3D) numerical models, their approach not only reliably simulates the continuous spatiotemporal evolution of 74 chemical species over 15 consecutive days but also achieved a nearly 8× speedup.

While the studies discussed above have highlighted the impressive performance of AI algorithms in capturing highly nonlinear relationships between different chemical species and reproducing complex spatiotemporal distributions, to date, no AI-based scheme exists for simulating the aerosol chemistry and interactions in 3D numerical models. Unlike photochemistry which only involves chemical reactions between species, the full aerosol chemistry and interactions encompasses numerous other intricate processes such as nucleation, coagulation, thermodynamics. Furthermore, since the PSD of an aerosol significantly influences aerosol behavior, an accurate depiction of the evolution of PSD is as critical as the precise simulation of concentration of aerosol species. These factors collectively present a heightened challenge for the development of an AI scheme capable of simulating the full atmospheric aerosol chemistry and interactions. The feasibility of establishing such an AI scheme for aerosols remains an open question.

To bridge this gap, in this study, we have developed a novel Artificial Intelligence Model for Aerosol Chemistry and Interactions, termed AIMACI, which is based on the Multi-Head Self-Attention algorithm and has been online coupled with a 3D numerical atmospheric model. 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, which certainly deserves further investigation in future. The results demonstrate that, compared to conventional numerical schemes, the AIMACI scheme: (1) accurately reproduces the spatiotemporal distributions and the evolution of PSD of different aerosol species, (2) has robust generalization ability, reliably simulating one month under

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

different environmental conditions across four seasons despite training on a limited dataset of just 16 days, and (3) achieves significantly computational speed-ups of approximately 5× on a single CPU and 277× on a single GPU. These advancements hold great promise for the future of climate modeling, enabling fast, accurate, and stable simulations of aerosol chemistry and interactions, thereby reducing uncertainties stemming from simplified representations of these processes. The structure of this paper is organized as follows: Section 2 provides a detailed description of the Weather Research and Forecasting with Chemistry (WRF-Chem) model and the establishment of the AIMACI scheme. Section 3 discusses the results, and Section 4 presents the conclusion, outlining the implications of our findings for the field.

2 Methods

2.1 WRF-Chem Model and MOSAIC Scheme

In this study, we utilize the updated version of WRF-Chem developed by the University of Science and Technology of China (USTC) for conducting all simulations. This USTC version of WRF-Chem boasts additional functionalities compared to the publicly released version, including the capability to diagnose radiative forcing of aerosol species, land-surface-coupled biogenic volatile organic compound emissions, and aerosol-snow interactions (Du et al., 2020; Hu et al., 2019; Zhang et al., 2021; Zhao et al., 2013a, b, 2014, 2016).

The conventional numerical scheme of aerosol chemistry and interactions adopted in this study for comparison is the MOSAIC (Model for Simulating Aerosol Interactions and Chemistry) (Zaveri et al., 2008) scheme within WRF-Chem, 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 (HNO3, HCl, and NH3) 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 treats all the major aerosol species important at urban, regional, and global scales, including sulfate (SO42-), nitrate (NO3-), chloride (Cl-), carbonate (CO32-), ammonium (NH4+), sodium (Na+), calcium (Ca2+), black carbon (BC), organic carbon (OC), other inorganic mass (OIN), mineral dust, methanesulfonic acid (MSA) and liquid water content of aerosol (Water). It employs a sectional approach, dividing the PSD into 4 discrete size bins in this study. The first, second, third, and fourth size bins are set to be 0.039~0.156, 0.156~0.625, 0.625~2.5, and 2.5~10.0 μm in diameter, respectively. In addition, it further considers the impact of marine biogenic sources of dimethyl sulfide on atmospheric aerosols and aqueous aerosol processes.

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

2.2 Learnable AIMACI Scheme

2.2.1 Scheme Construction

Previous attempts into the substitution of conventional numerical schemes with AI schemes have predominantly utilized simple AI algorithms, such as Random Forest Regression (Keller and Evans, 2019). This preference stems from the complex challenge of coupling sophisticated AI algorithms, often written in Python, with numerical models coded in Fortran. While some studies have explored the use of advanced AI algorithms, such as Fully Connected Neural Networks (Sharma et al., 2023) and Residual Neural Networks (Kelp et al., 2018, 2020, 2022; Liu, 2021; Wang et al., 2022), these have occasionally encountered difficulties when dealing with high-dimensional input variables and have demonstrated limitations in accurately simulating highly nonlinear systems (Xia et al., 2024).

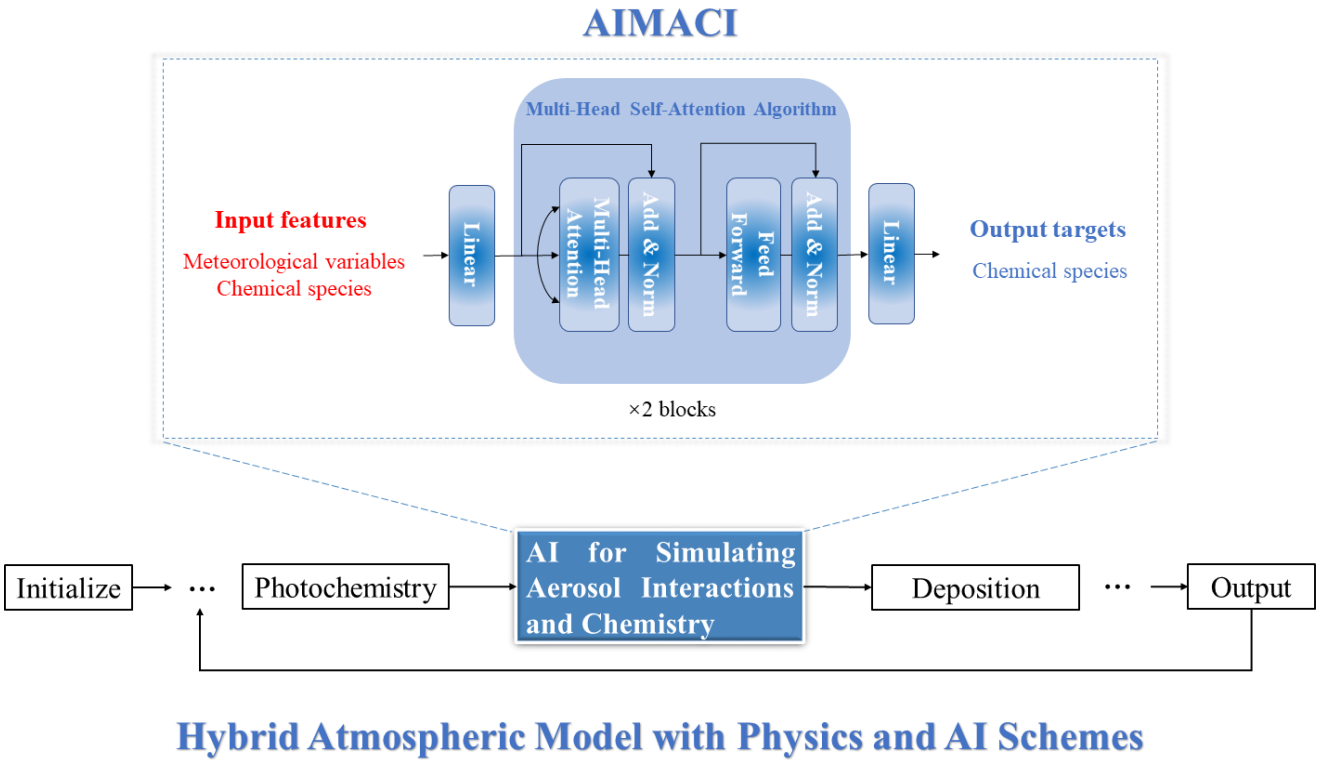
In this study, with an AI algorithm, we, first time, attempt to emulate a sophisticated scheme of aerosol chemistry and interactions (i.e., the MOSAIC scheme), which encompasses a range of highly nonlinear processes such as chemical reactions, nucleation, coagulation, thermodynamics and phase equilibrium, gas-particle partitioning, and particle transfer (Zaveri et al., 2008). Given the complexity of these interactions, there is a clear need for AI algorithms with superior representational capacity for nonlinear systems. Xia et al. (2024) have highlighted that the Multi-Head Self-Attention (MHSA) algorithm excels in capturing the intricate chemical relationships among different species. It offers not only high simulation accuracy and computational efficiency but is also less susceptible to the increase in the number of chemical species.

Therefore, here we innovate by pioneering a novel Artificial Intelligence Model for Aerosol Chemistry and Interactions (AIMACI), leveraging the MHSA algorithm. The MHSA algorithm, serving as the foundational architecture of state-of-the-art transformer models, has been successfully deployed in domains such as Natural Language Processing and Computer Vision, achieving significant advancements (e.g. Vaswani et al., 2017; Liu et al., 2021; Bi et al., 2023). It is distinguished by its ability to globally attend to input variables and conduct parallel calculations across multiple heads. This attribute empowers it to adeptly navigate the curse of dimensionality and capture overarching dependencies, while effectively enhancing computational efficiency.

Figure 1 provides a schematic representation of the AI model architecture utilized within the AIMACI scheme, showcasing the integration of MHSA algorithm in our hybrid atmospheric model with physics and AI schemes (physics-AI hybrid model). The AI model architecture is intricately designed with three principal 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 raw 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: a multi-head 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 aerosol

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

chemistry and interactions 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 comprehensive AIMACI scheme.



160

Figure 1: The model architecture employed in the Artificial Intelligence Model for Aerosol Chemistry and Interactions (AIMACI). In the hybrid atmospheric model with physics and AI schemes, we only use AI schemes for simulating aerosol interactions and chemistry, and the rest maintain the original numerical schemes.

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

2.2.2 Training and Testing Procedure

To generate the training, validation, and test datasets, we conducted the WRF-Chem simulations over East China, spanning the period from 2019-03-01 00:00 UTC to 2019-03-19 23:00 UTC. The simulation result was segmented as follows: the initial 16 days from 2019-03-02 00:00 UTC, were designated as the training set, the penultimate day served as the validation set, and the final day constituted the test set. The simulation was configured with a 0.2 ° horizontal resolution, covering 140 × 105 grid cells within the geographical bounds of 107.1 °E to 127.9 °E and 19.7 °N to 47.5 °N, and featured 49 vertical layers extending up to 50 hPa. A dynamic time step of 2 minutes and a chemical time step of 1 hour were employed. Concentrations of aerosol and gas species pertinent to gas-particle partitioning were recorded hourly, along with key meteorological variables influencing chemistry: temperature, pressure, air density, and water vapor mixing ratio. Each training sample included 65 input features (4 meteorological variables, 5 gas species, and 14 aerosol species with 4 size bins) and 37 output targets (5 gas species and 8 aerosol species with 4 size bins). It should be noted that in the AIMACI scheme's simulation of aerosol processes, the concentrations of other inorganic mass (OIN), mineral dust, black carbon (BC), organic carbon (OC), calcium (Ca^{2+}), and carbonate (CO_3^{2-}) are not altered. Consequently, these aerosol species are not output variables. However, they play a significant role in affecting the acidity or alkalinity of the atmospheric environment, which in turn influences the formation of aerosols. Therefore, these species are necessary as input variables to ensure that the model accurately reflects the conditions affecting aerosol production. A comprehensive list of variables used for training the AIMACI scheme is presented in Table 1.

After training, the AIMACI scheme was flexibly coupled into WRF-Chem, utilizing TorchScript and Libtorch tools officially provided by PyTorch. This coupling approach encapsulates the AIMACI scheme within a static library, minimizing alterations to the original codebase while offering a lightweight, adaptable, and easily plug-and-play solution. It is not only efficient but also versatile, capable of encapsulating a wide range of complex AI algorithms and coupling them with diverse atmospheric and climate models.

Furthermore, we conducted three sets of additional experiments to comprehensively evaluate the performance of the AIMACI scheme: (1) predictions were made on the test dataset prior to coupling the scheme with WRF-Chem; (2) a 10-day continuous simulation was performed from a period outside the training phase following the coupling of the AIMACI scheme into WRF-Chem; (3) a month-long continuous simulation was carried out under different environmental conditions across all four seasons. Through the analysis of spatial distributions, temporal series, and the evolution of PSD, we aimed to assess the AIMACI scheme's potential on climate research applications.

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

Table 1 Input and output variables of Artificial Intelligence Model for Aerosol Chemistry and Interactions (AIMACI)

Type	Input variables	Output variables
Meteorological variables	temperature	-
	air density	-
	pressure	-
	water vapor mixing ratio	-
Gas species	H_2SO_4	H_2SO_4
	HNO_3	HNO_3
	NH3	NH_3
	HCL	HCL
	MSA	MSA
Aerosol species	SO_4^{2-} [Size:1-4]	SO_4^{2-} [Size:1-4]
	NO_3^- [Size:1-4]	NO_3^- [Size:1-4]
	NH_4^+ [Size:1-4]	NH_4^+ [Size:1-4]
	Na^+ [Size:1-4]	Na^+ [Size:1-4]
	Cl^- [Size:1-4]	Cl^- [Size:1-4]
	MSA[Size:1-4]	MSA[Size:1-4]
	Water[Size:1-4]	Water [Size:1-4]
	Num [Size:1-4]	Num [Size:1-4]
	OIN [Size:1-4]	-
	DUST [Size:1-4]	-
	OC [Size:1-4]	-
	BC [Size:1-4]	-
	Ca^{2+} [Size:1-4]	-
	CO_3^{2-} [Size:1-4]	-

2.3 Evaluation metric

In this research, a comprehensive evaluation of the AIMACI scheme's effectiveness was conducted utilizing three recognized statistical measures. For every species examined, the calculation of the Pearson correlation coefficient (R2), the root mean square error (RMSE) and the normalized mean bias (NMB) was performed.

$$\uparrow (R^2) = \frac{\left(\sum_{i=1}^{N_{lat}} \sum_{j=1}^{N_{lon}} L(i)\left(c_{i,j} - \bar{c}\right)\left(\hat{c}_{i,j} - \bar{\hat{c}}\right)\right)^2}{\sum_{i=1}^{N_{lat}} \sum_{j=1}^{N_{lon}} L(i)\left(c_{i,j} - \bar{c}\right)^2 \times \sum_{i=1}^{N_{lat}} \sum_{j=1}^{N_{lon}} L(i)\left(\hat{c}_{i,j} - \bar{\hat{c}}\right)^2} \quad (1)$$

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

$$\downarrow RMSE = \sqrt{\frac{\sum_{i=1}^{N_{lat}} \sum_{j=1}^{N_{lon}} L(i)\left(\hat{c}_{i,j} - c_{i,j}\right)^2}{N_{lat} \times N_{lon}}} \quad (2)$$

$$\downarrow NMB = \frac{\sum_{i=1}^{N_{lat}} \sum_{j=1}^{N_{lon}} L(i)(\hat{c}_{i,j} - c_{i,j})}{\sum_{i=1}^{N_{lat}} \sum_{j=1}^{N_{lon}} L(i) c_{i,j}} \quad (3)$$

In above, $L(i) = N_{lat} \times \frac{\cos \varphi_i}{\sum_{i=1}^{N_{lat}} \cos \varphi_i}$ is the weight at latitude φ_i , $\hat{C}$ denotes the concentration simulated by the AIMACI scheme, C denotes the concentration simulated by the MOSAIC scheme, ↑ denotes higher values are better, ↓ denotes lower values are better.

2.4 Computational configuration

A primary incentive for coupling AI schemes into atmospheric and climate models is the pursuit of substantial computational acceleration. However, such acceleration is not inherently guaranteed, as demonstrated by Keller et al. (2019). Consequently, it is imperative to meticulously compare the temporal expenditure of AI schemes against those of traditional numerical schemes.

In this study, we undertook a comparative analysis of the computational time required by the numerical scheme and the AIMACI scheme for simulating aerosol chemistry and interactions in 720,300 discrete grid cells, which roughly corresponds to a global simulation at 2.5° × 2.5° horizontal resolution with 72 vertical layers. To ensure a holistic and unbiased assessment of the speedup achieved, we measured the computational time by averaging the duration of 24 consecutive daily simulations. Both schemes were tested utilizing a single CPU core and additionally evaluated the AIMACI scheme with a GPU-accelerated scenario using a single GPU. The computational hardware employed in our tests consisted of an Intel Xeon Scalable 8358 CPU and an NVIDIA A100-80G GPU.

3 Results

3.1 Offline Single-step Simulations with the AIMACI Scheme

Before coupling the AIMACI scheme with the 3D numerical model WRF-Chem for continuous simulation, we first evaluated its performance on a test dataset that was separate from the training data. The test dataset, as detailed in previous sections, comprises a series of 3D spatial outcomes taken at 24-hourly intervals on March 19, 2019. It provides representative samples that span a wide range of meteorological conditions and species concentrations. The evaluation on this dataset provides insight into the AIMACI scheme's performance in various atmospheric conditions. Table 2 presents the statistical metrics for all simulated species, offering a comprehensive assessment of the scheme's simulation capabilities.

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

The results are promising with an average R^2 of 0.99 for all 37 evaluated species. This high degree of correlation indicates a strong consistency between the simulations with the AIMACI scheme and the MOSAIC scheme (hereinafter referred to as numerical scheme). The average NMB for these species is 0.95%, reflecting only a slight deviation from the numerical scheme's outcomes and highlighting the AIMACI scheme's impressive accuracy in simulating aerosol chemistry and interactions. Atmospheric aerosols significantly impact the climate system through direct radiative forcing, such as scattering and absorption of solar radiation, and indirect radiative forcing, such as serving as cloud condensation nuclei (CCN) (Bellouin et al., 2020). As depicted in Table 2, major aerosol species that play key roles in these processes, including sulfates, nitrates, ammonium, sea salt (e.g., sodium and chloride) (Lohmann and Feichter, 2005), are all accurately simulated by the AIMACI scheme. This is crucial for the future coupling the AIMACI scheme into climate models for precise climate simulations.

Table 2 Statistical metrics on the test dataset of Artificial Intelligence Model for Aerosol Chemistry and Interactions (AIMACI) (The RMSE of different species has different unit: aerosol (μg/kg), num (kg-1), gas (ppmv)).

Number	Variable	R2	RMSE	NMB(%)
1	H_2SO_4	0.97	2.99E-07	-2.10
2	HNO_3	1.00	3.61E-05	0.19
3	NH_3	1.00	4.84E-05	3.49
4	HCL	1.00	9.68E-06	-0.41
5	MSA	0.82	1.03E-09	0.12
6	SO_4_a01	1.00	1.14E-02	0.16
7	NO_3_a01	1.00	4.90E-02	-0.43
8	NH_4_a01	1.00	1.44E-02	-0.45
9	Na_a01	1.00	6.10E-06	0.32
10	Cl_a01	0.99	2.52E-03	-1.42
11	MSA_a01	1.00	1.78E-05	5.70
12	Water_a01	1.00	5.60E-01	0.46
13	Num_a01	1.00	4.45E+07	-0.10
14	SO_4_a02	1.00	4.16E-02	0.38
15	NO_3_a02	1.00	7.04E-02	0.69
16	NH_4_a02	1.00	2.31E-02	0.51
17	Na_a02	1.00	1.65E-04	0.10
18	Cl_a02	1.00	4.63E-03	-1.25
19	MSA_a02	1.00	1.51E-05	3.05

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

20	Water_a02	0.99	1.32E+00	1.58
21	Num_a02	1.00	9.20E+06	0.61
22	SO_4_a03	0.94	2.90E-02	0.80
23	NO_3_a03	1.00	2.68E-02	0.89
24	NH_4_a03	1.00	5.37E-03	0.48
25	Na_a03	1.00	2.01E-03	2.44
26	Cl_a03	1.00	4.98E-03	1.29
27	MSA_a03	1.00	2.94E-06	2.85
28	Water_a03	0.99	4.11E-01	2.03
29	Num_a03	1.00	1.05E+05	0.45
30	SO_4_a04	1.00	7.25E-04	0.27
31	NO_3_a04	0.99	5.11E-02	0.75
32	NH_4_a04	0.98	5.62E-03	-3.21
33	Na_a04	1.00	9.19E-03	4.13
34	Cl_a04	1.00	1.13E-02	3.28
35	MSA_a04	0.79	2.00E-05	7.93
36	Water_a04	0.99	1.35E+00	-3.32
37	Num_a04	1.00	5.89E+03	2.86

To illustrate the simulated results of the AIMACI scheme, we selected four key aerosol species, which include chemical species, liquid water content in aerosol, and number concentration of aerosols. Figure 2 presents the data density and distribution for column concentration of the four key aerosol species. Sulfate, derived primarily from fossil fuel combustion emissions, plays a crucial role in acid rain, aerosol formation, and aerosol-cloud interactions (Calvert et al., 1985; Fuzzi et al., 2015; Penkett et al., 1979). Nitrate, formed through the oxidation of nitrogen oxides, is a major inorganic component of aerosols influencing air quality and ecosystem health (Parrish et al., 2012; Saiz-Lopez et al., 2017). Liquid water content within aerosols is pivotal for understanding the hygroscopic nature of particles and their ability to act as cloud condensation nuclei, thereby influencing cloud formation and precipitation processes (Hodas et al., 2014; Liu et al., 2019; Nguyen et al., 2016; Wu et al., 2018). The number concentration of aerosols is a critical metric for assessing the overall aerosol loading and its direct impact on visibility, radiation balance, and climate feedback mechanisms (Spracklen et al., 2010). These four species collectively provide a comprehensive view of the aerosol's multifaceted role in atmospheric processes. The results from the numerical scheme simulations indicate that sulfate, nitrate, and liquid water content of aerosol exhibit higher column concentrations within 0.156 to 0.625 μm (size bin 2), whereas the number concentration is notably larger within 0.039 to 0.156 μm (size bin 1). This indicates that, despite the greater number of smaller particles in size bin 1, their overall contribution to the total mass is less significant due to their lower individual mass compared to the larger particles in size bin

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

2. The AIMACI scheme effectively captures these nuanced aerosol characteristics, as corroborated by the exemplary R^2 values of 1.0 depicted in Figure 2, which underscore the scheme's fidelity in modeling aerosol behavior across various particle sizes.

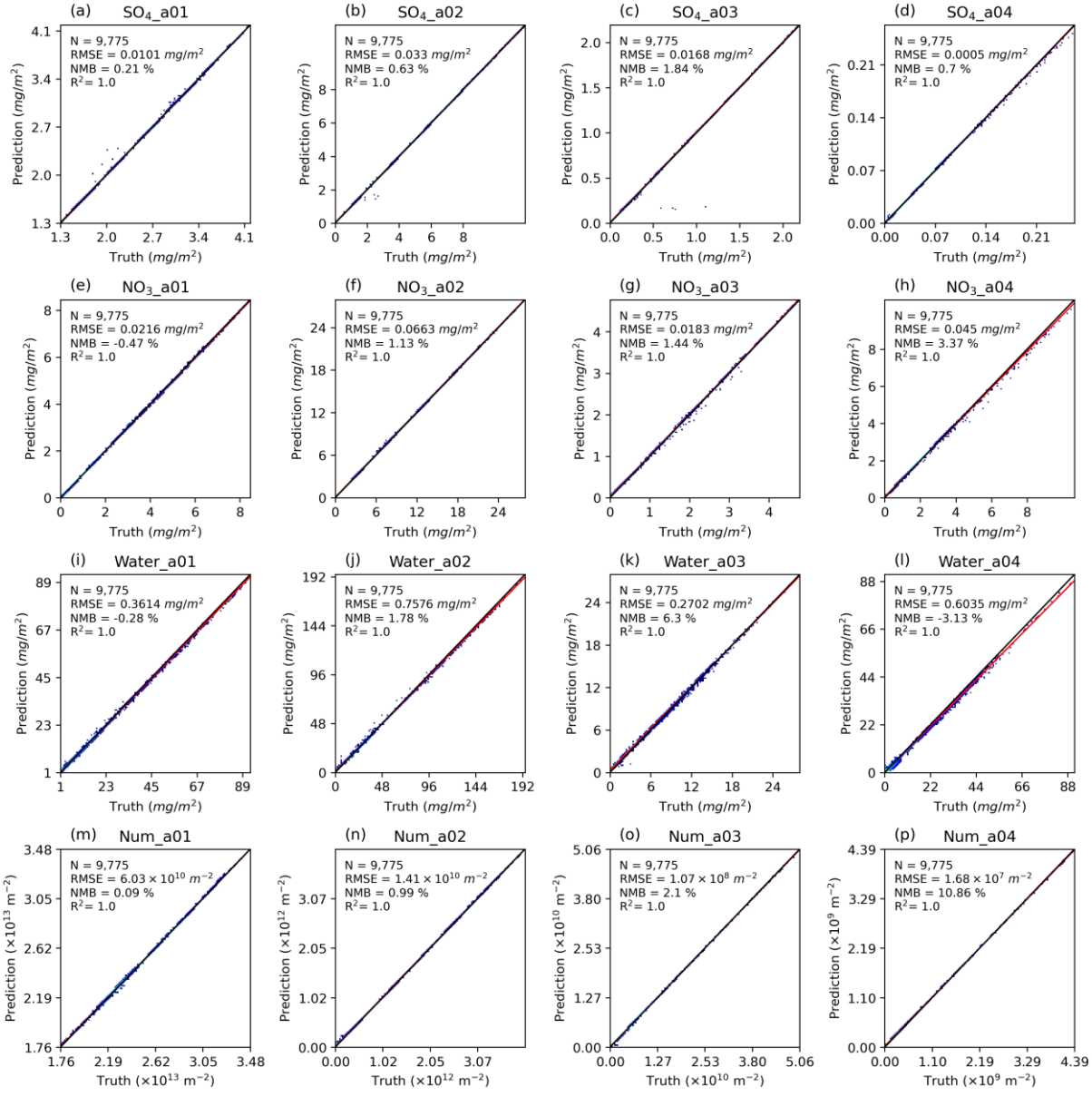


Figure 2: The density plot of column concentration of four key aerosol species (sulfate (SO42-), nitrate (NO3-), liquid water content of aerosol (Water), and number concentration of aerosol (Num)) simulated by the AIMACI scheme on the test dataset. The results are calculated by covering the region spanning from 109.1 °E to 125.9 °E and from 22.1 °N to 44.9 °N, with the data being averaged across the time dimension. The black line is the identical line (y=x), and the red line is fitted line.

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

3.2 Offline Single-step Simulations with the AIMACI Scheme

Unlike offline single-step simulations, coupling the AIMACI scheme into 3D numerical models to form a physics-AI hybrid model for continuous simulation entails interactions and feedback with numerous other processes. Consequently, a thorough evaluation of the AIMACI scheme's online simulation performance is essential. We focus on its performance across three critical dimensions: (1) Stable and Accurate Simulation Capability: The AIMACI scheme should accurately reproduce the spatiotemporal and size distribution of various aerosol species without rapid accumulation of errors during the simulation process. (2) Robust Generalization Ability: The AIMACI scheme should be applicable to scenarios beyond the training data, such as different seasons, demonstrating its robustness in a variety of environmental conditions. (3) High Computational Efficiency: Compared to the conventional numerical scheme, the AIMACI scheme should offer enhanced computational efficiency, which is vital for high-resolution, long-term simulations. Although these requirements are often challenging to satisfy simultaneously, achieving these benchmarks is crucial for leveraging the full potential of the AIMACI scheme in advancing our understanding of aerosol interactions and their impact on climate change.

3.2.1 Stable and Accurate Simulation Capability

Coupling AI schemes into numerical models for stable and accurate simulations across multiple time steps has long been a formidable challenge. While the simulation errors for individual species at each time step may be minimal, they can accumulate over multiple time steps, and may even spread to other species and physical-chemical processes, leading to chaotic simulation outcomes at the end. Typically, simulations with sophisticated aerosol processes at high-resolution, such as those in WRF-Chem, are limited to a few weeks due to computational costs. In this study, we conducted a 10-day continuous simulation from 2019-03-20 00:00 UTC to 2019-03-30 00:00 UTC to evaluate the performance of the AIMACI scheme in a coupled mode.

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). The figure also tracks the temporal evolution of the RMSE throughout the simulation period. The results reveal that the high-value areas of sulfate column concentrations for different particle sizes exhibit a hook-like structure, stretching from the Yangtze River Economic Belt northeastward 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.

A notable aspect of the AIMACI scheme is its grid-based training and prediction methodology, which contrasts with existing AI large models such as Pangu (Bi et al., 2023) and Fengwu (Chen et al., 2023) that operate on entire fields. This approach offers a significant reduction in computational costs by avoiding the necessity for convolutional networks. Furthermore, the grid-based AI scheme is versatile, capable of being applied to simulations of regions of any size, without

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

constraints imposed by the size of the training area. However, this approach also presents a challenge in accurately simulating spatial distributions, given the potential for error propagation from neighboring grid points due to physical processes like transport. In Figure 3, the AIMACI scheme has successfully captured and reproduced the intricate spatial patterns of sulfate column concentrations across different particle sizes with R2 values all exceeding 0.88, even after a prolonged 10-day simulation. This achievement underscores the AIMACI scheme's exceptional stability and accuracy. Despite localized instances of underestimation or overestimation, the time series of RMSE for each size bin exhibit a stable trend with low values throughout the entire simulation period, suggesting that these discrepancies do not lead to runaway error growth. This sustained performance further substantiates the AIMACI scheme's reliability, positioning it as a robust tool for extended atmospheric and climate simulations.

Figure 4 presents a comparison of the zonal average total concentrations (summed across all size bins) of sulfate and nitrate, simulated by both the numerical scheme and the AIMACI scheme, with results averaged over the entire 10-day simulation period. Observations from Figure 4a and 4c indicate that high concentration zones for both sulfate and nitrate are predominantly situated between 25 °N and 40 °N, coinciding with the latitude range of the Yangtze River Economic Belt. This distribution pattern is likely influenced by the significant anthropogenic emissions in this area. Through turbulent and convective transport processes, sulfate and nitrate from lower altitudes are transported to higher altitudes, with concentrations gradually diminishing with increasing altitude. In Figures 4b and 4d, the AIMACI scheme exhibits a notable alignment with the outcomes from the numerical scheme, as evidenced by the R^2 values, which are exceptionally high at 0.99 for both sulfate and nitrate. The Root Mean Square Error (RMSE) values are 0.10 μg/kg for sulfate and 0.48 μg/kg for nitrate, suggesting that the discrepancies are minimal, further supporting the AIMACI scheme's accuracy. Although slightly overestimations of sulfate and nitrate concentrations are observed in the outcomes from the AIMACI scheme, these minor discrepancies do not detract from the overall consistency between the two schemes. The high R^2 and low RMSE values underscore the AIMACI scheme's remarkable capability to accurately reproduce the distribution patterns of aerosols, affirming its reliability and effectiveness in simulating atmospheric aerosol chemistry and interactions.

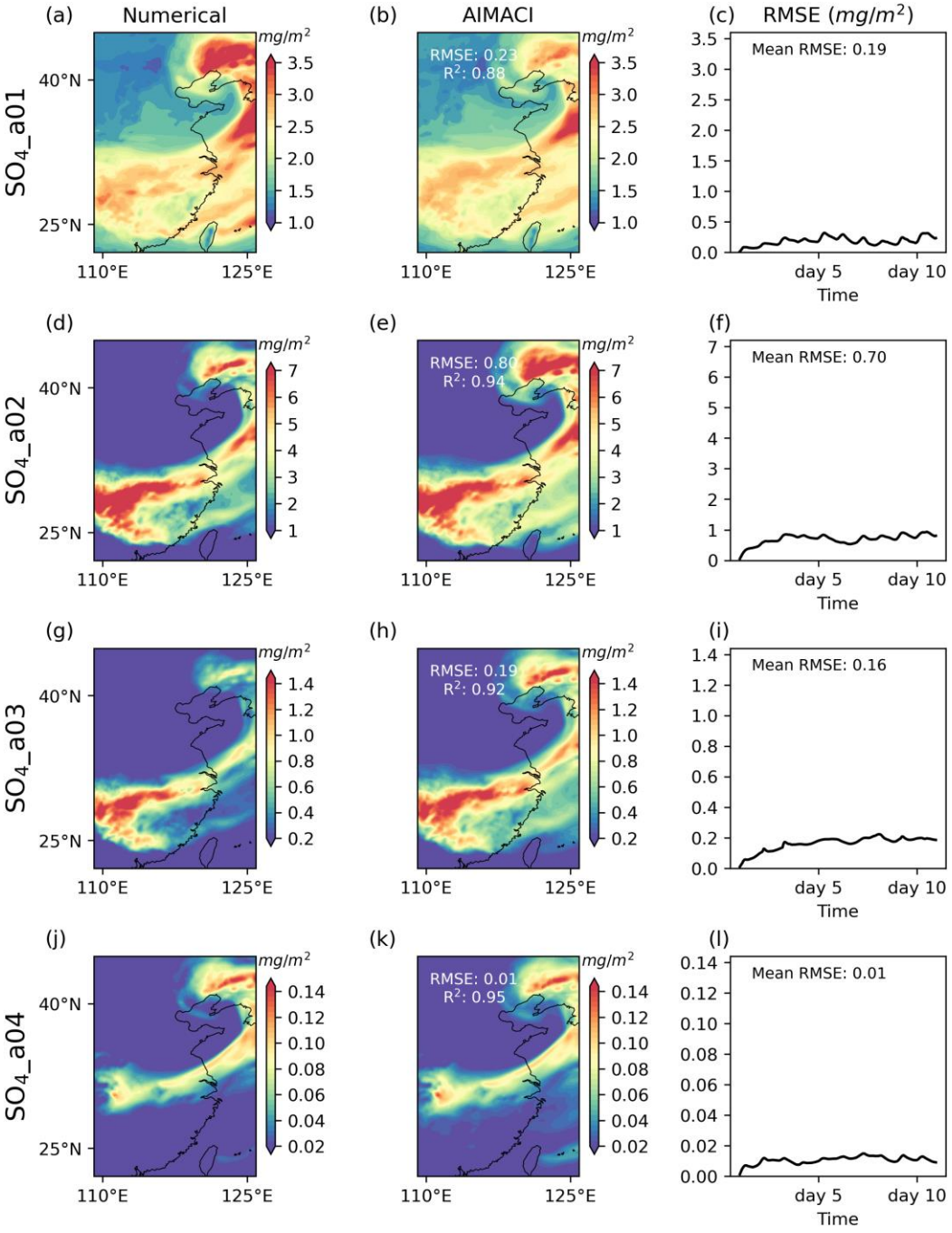
Figure 5 illustrates the ability of the AIMACI scheme to reproduce temporal variations of surface total concentrations of four key aerosol species. These results represent the calculated averages for the Yangtze River Delta region, a crucial urban agglomeration in China, spanning the coordinates 119.1 °E to 121.9 °E and 30.1 °N to 31.9 °N. Throughout the simulation period, sulfate concentrations primarily fluctuate within the range of 0 to 6 μg/kg, while nitrate concentrations exhibit a broader variability, predominantly ranging from 0 to 20 μg/kg. Notably, all four key aerosol species experience several instances of abrupt concentration spikes and declines. For instance, between the 11th and 30th hour of the simulation, the liquid water content of aerosol experiences a dramatic increase from 32.86 μg/kg to 263.47 μg/kg, followed by a sharp decrease to 28.72 μg/kg. Despite these pronounced fluctuations, the AIMACI scheme adeptly reproduces these features without introducing systematic bias, achieving R^2 values larger than 0.97.

As discussed above, the AIMACI scheme 's proficiency in simulating the spatiotemporal distribution and variation trend of different aerosol species is well-established. However, accurately reproducing the evolution of the aerosol PSD is equally

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

vital, given the significant role particle size plays in dictating the interactions of aerosols with clouds and radiation`, which are pivotal for atmospheric processes. Figure 6 presents the PSD and frequency distribution for the surface concentrations of four key aerosol species simulated by the AIMACI scheme. The frequency distributions of sulfate and nitrate surface concentrations exhibit a relatively uniform pattern, whereas the liquid water content and number concentration of aerosols display extreme values, leading to pronounced skewness in their distributions. The AIMACI scheme accurately captures these distributions, although it tends to overestimate minimal concentration values that approach zero across different particle sizes, a minor deviation that could be addressed in future refinements. Notably, there are significant differences in the PSD among the aerosol species. Sulfate and nitrate concentrations peak within the 0.156 to 0.625 μm (size bin 2), whereas the liquid water content of aerosols is most concentrated in the 2.5 to 10.0 μm (size bin 4) and the number concentration of aerosols is predominantly found in the 0.039 to 0.156 μm (size bin 1). These findings highlight the complexity of accurately modeling PSD and the AIMACI scheme's commendable performance in reproducing these intricate patterns.

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024



345 **Figure 3: Sulfate column concentration simulations across different size bins. The left and middle panels depict the spatial distribution at the 10-day continuous simulation's end (2019-03-30 00:00 UTC), as simulated by the numerical scheme and AIMACI scheme, respectively. The right panel shows the temporal evolution of the hourly RMSE over the 10-day period.**

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

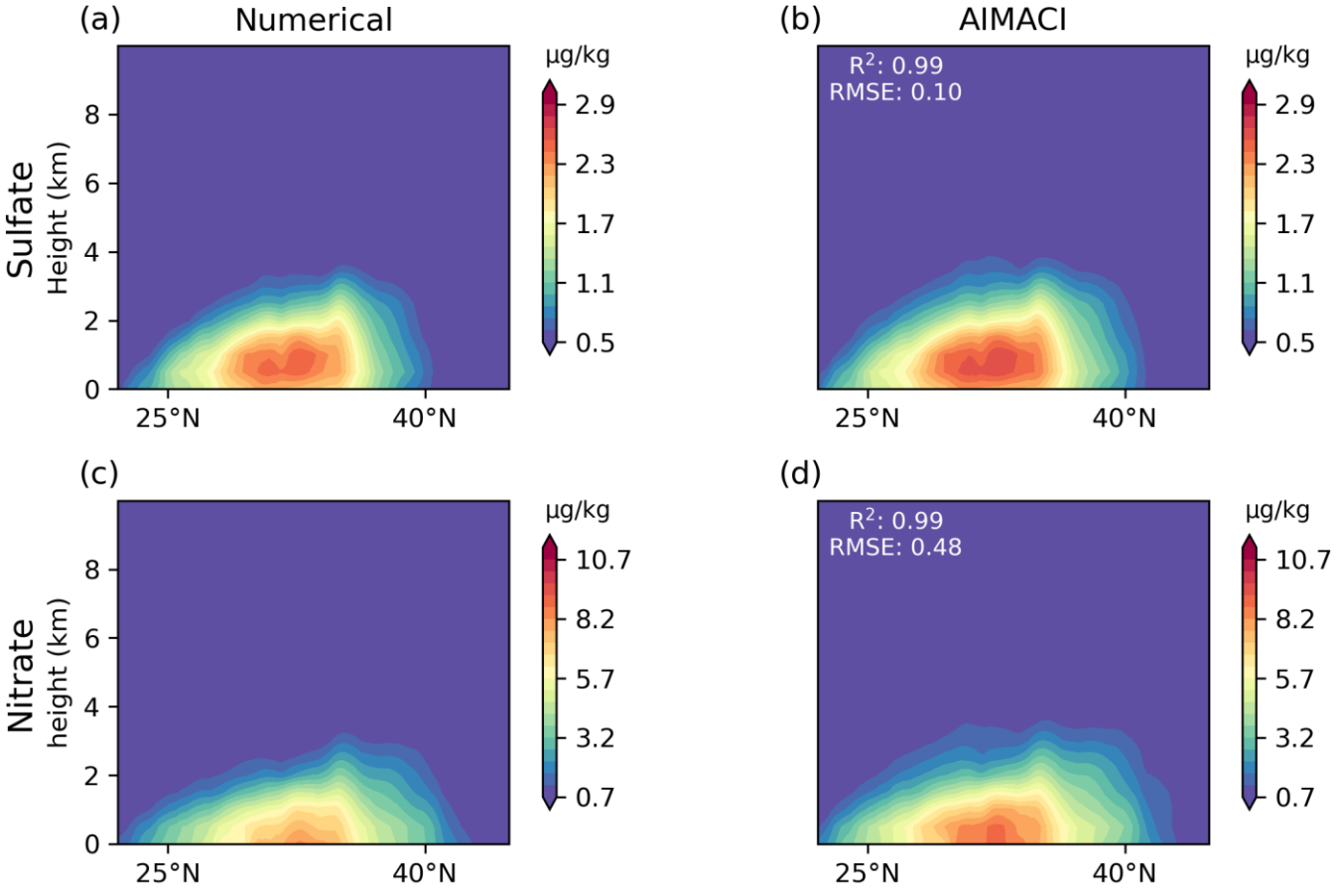


Figure 4: Zonal mean total concentrations (summed across 4 size bins) of sulfate and nitrate between 109.1 °E and 125.9 °E, as simulated by the numerical scheme and AIMACI scheme. Results are averages over the entire 10-day simulation period.

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

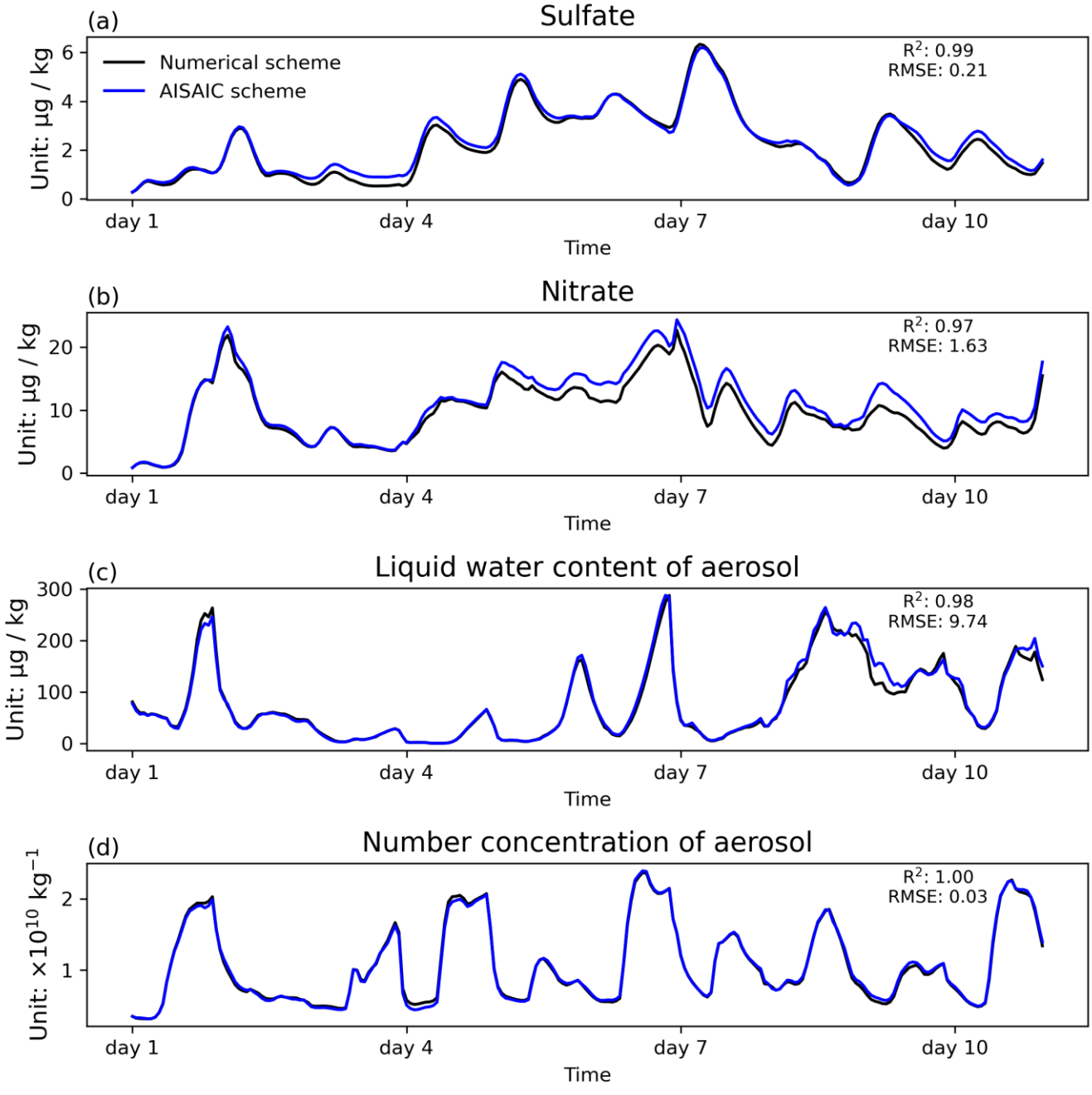


Figure 5: Time series of surface total concentrations (summed across 4 size bins) of four key aerosol species (sulfate, nitrate, liquid water content of aerosol, and number concentration of aerosol), as simulated by the numerical scheme and the AIMACI scheme. Results represent the calculated averages for the Yangtze River Delta region (119.1 °E~121.9 °E, 30.1 °N~31.9 °N).

355

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

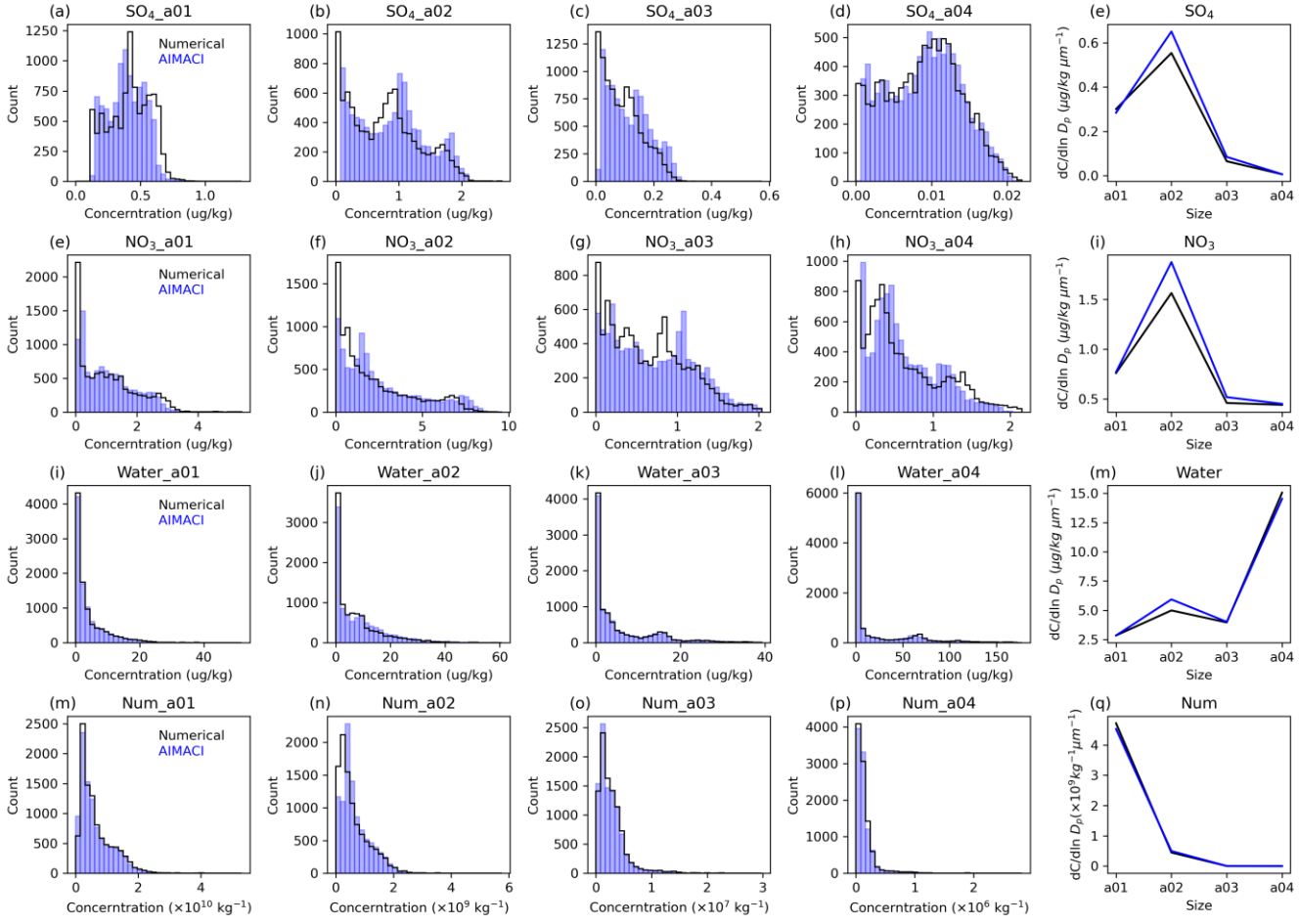


Figure 6: Frequency distributions of different particle sizes for the surface concentrations of four key aerosol species (sulfate (SO4), nitrate (NO3), liquid water content of aerosol (Water), and number concentration of aerosol (Num)), as simulated by the numerical scheme and the AIMACI scheme. The last column showcases the particle size distributions of these key aerosol species surface concentrations. The results are calculated by covering the region spanning from 109.1 °E to 125.9 °E and from 22.1 °N to 44.9 °N, with the data being averaged over the entire simulation period.

3.2.2 Robust Generalization Ability

In the preceding section, we have demonstrated the AIMACI scheme's remarkable success in simulating the 3D spatiotemporal distributions and PSD of various aerosol species concentrations. Building upon these findings, this section delves into an exploration of the AIMACI scheme's generalization ability under diverse environmental conditions, a critical aspect for its future integration into climate models to mitigate uncertainties stemming from oversimplified or absent aerosol processes. To evaluate this, we conducted a series of supplementary experiments. These experiments involved a comparative analysis of one-month simulations for each of the four seasons—spring, summer, autumn, and winter—between the

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

numerical scheme and the AIMACI scheme. This comprehensive evaluation ensures that the AIMACI scheme's performance is not limited to specific conditions but is consistently accurate across a range of environmental scenarios, thereby bolstering its applicability and reliability in climate modeling endeavors.

Figure 7 illustrates the simulation with AIMACI scheme of monthly average surface total concentrations of nitrate for different environmental conditions across four seasons. The results reveal distinct seasonal variations in nitrate surface concentrations, with higher values observed in January and lower in July. This seasonal contrast can be primarily ascribed to the heightened anthropogenic emissions associated with winter heating activities, which contribute to a surge in nitrate aerosols. Conversely, the summer months are often characterized by meteorological phenomena such as the East Asian monsoon, marked by increased precipitation and wind velocities. These conditions enhance the atmospheric dispersion and removal of aerosol, thereby resulting in reduced nitrate concentrations at the surface. The spatial distribution of nitrate concentrations, as depicted in Figure 7, is notably concentrated between the latitudes of 25 ℃N and 40 ℃N, a pattern that remains consistent across months in different seasons. However, the specific distribution of high-concentration zones and the concentrations over the sea areas exhibit discernible differences. The AIMACI scheme, despite being trained on data from only 16 days in March, demonstrates a remarkable ability to reproduce these distribution characteristics across different environmental conditions. This is corroborated by the R^2 values, which larger than 0.93, indicating a strong agreement between the AIMACI scheme and the numerical scheme. However, the AIMACI scheme's performance in July is somewhat less accurate compared to other months, with some overestimations in certain areas. Based on our analysis, deviations in the AISAIC simulation results are frequently associated with the occurrence of typhoon events. Our training dataset, which comprises only partial data from March, may not adequately encompass the unique meteorological conditions associated with extreme events such as typhoons. Consequently, these omissions may contribute to the observed discrepancies in the simulations. To address this limitation and enhance the AIMACI scheme's accuracy, future iterations could incorporate a more diverse training dataset that encompasses various environmental conditions, including extreme events such as typhoons. By expanding the training dataset, the AIMACI scheme can be further refined to provide more reliable simulations, especially during months with unique meteorological phenomena.

Figure 8 presents the time series of surface total concentrations of nitrate across different environmental conditions for the four seasons, as simulated by the AIMACI scheme for the Yangtze River Delta region. The analysis reveals that January exhibit more pronounced fluctuations in nitrate surface concentrations, with peaks surpassing 30 μg/kg, while other months display more rapid variations, characterized by a higher frequency of concentration peaks and troughs. Reproducing the hourly concentration changes poses a greater challenge than simulating multi-day average concentrations, as the averaging process can offset positive and negative errors, which is not the case for time series data. Nonetheless, the AIMACI scheme maintains high consistency with the numerical scheme, demonstrating its robust generalization capabilities. Despite occasional significant deviations during July simulation period, such as the notable overestimation on July 20th, these anomalies do not lead to a rapid error escalation or the disruption of subsequent simulation results. Instead, as the simulation progresses, the errors are gradually mitigated, bringing the AIMACI scheme's results back in close alignment with the

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

numerical scheme. This phenomenon is closely related to the interactions between aerosols and other processes within the physics-AI hybrid model, which differ significantly from offline simulation scenarios.

405

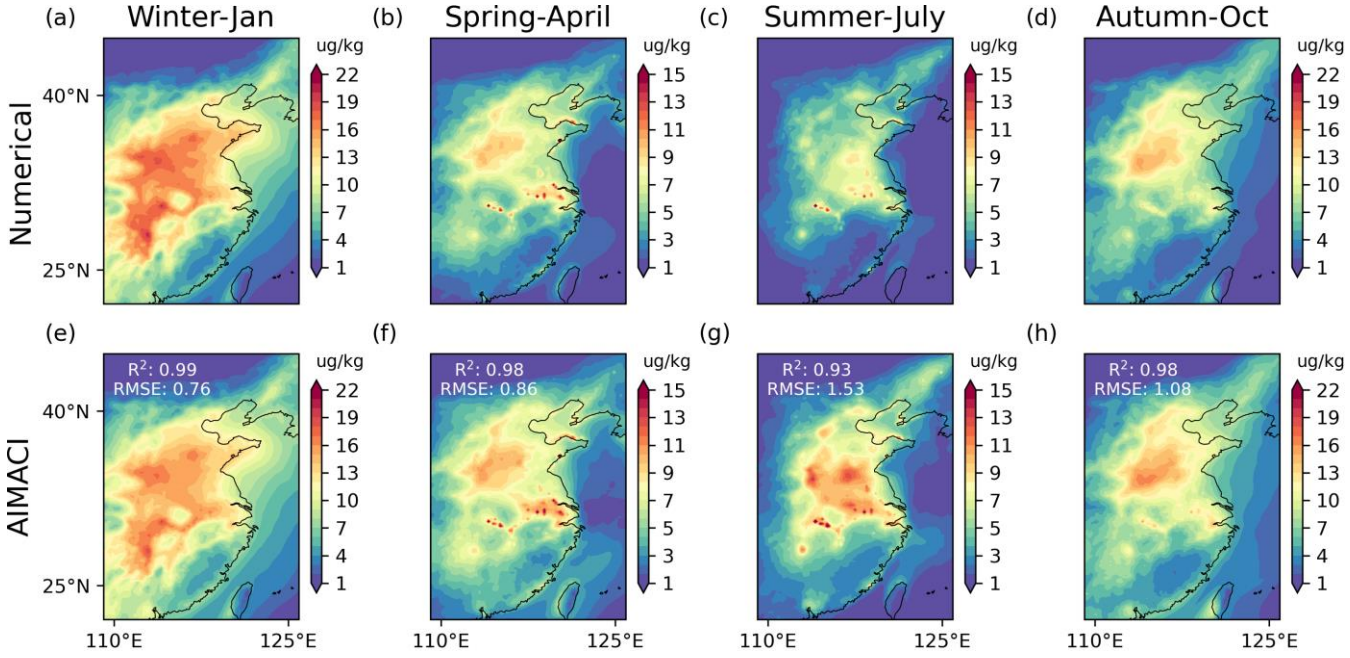


Figure 7: Monthly average surface total concentrations (summed across 4 size bins) of nitrate for different environmental conditions across seasons, as simulated by the numerical scheme and the AIMACI scheme.

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

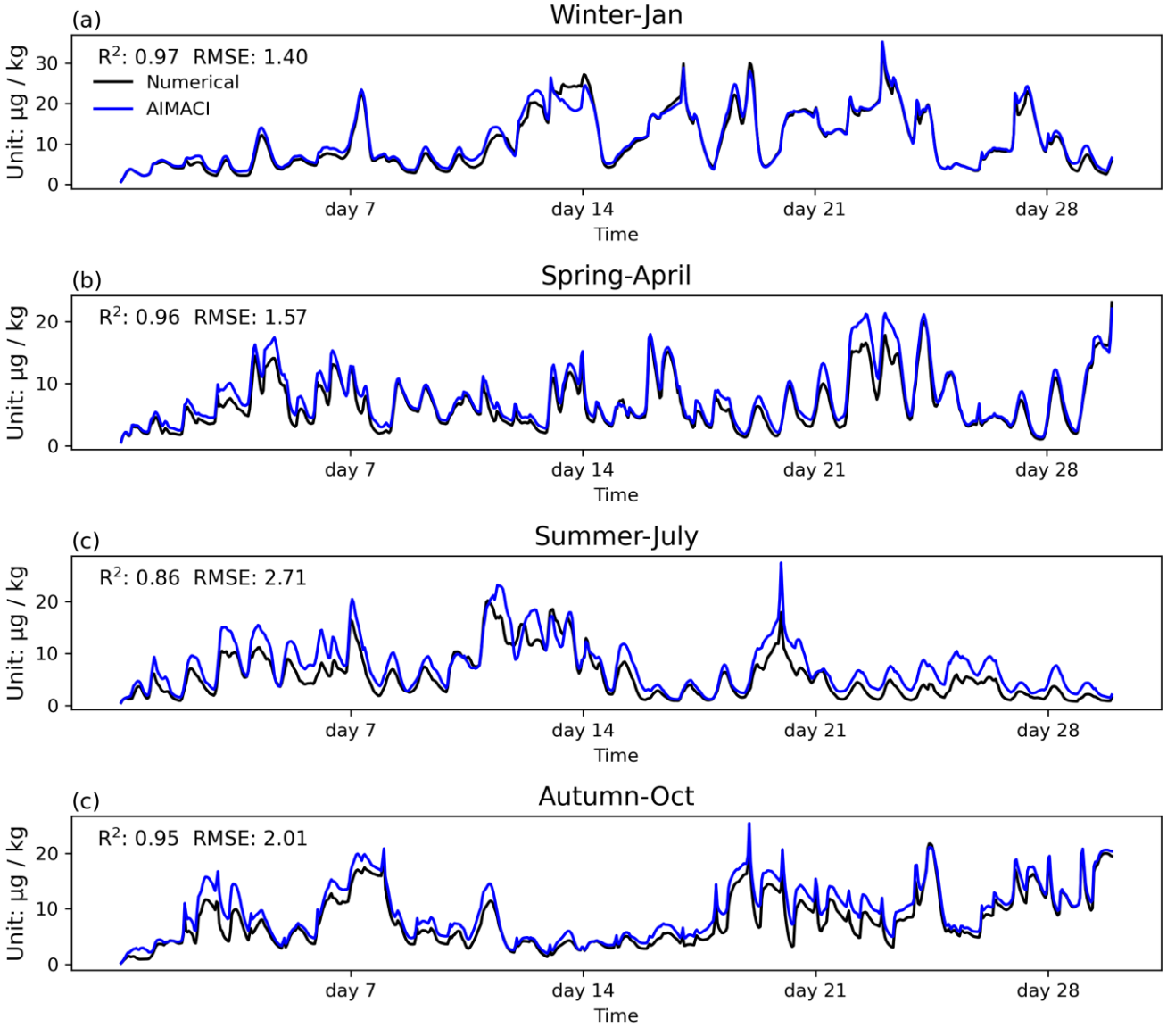


Figure 8: Time series of surface total concentrations (summed across 4 size bins) of nitrate for different environmental conditions across seasons, as simulated by the numerical scheme and the AIMACI scheme. Results represent the calculated averages for the Yangtze River Delta region (119.1 °E~121.9 °E, 30.1 °N~31.9 °N).

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

3.2.3 High Computational Efficiency

A primary motivation for the development of the AIMACI scheme is the potential for increased computational efficiency offered by AI schemes compared to conventional numerical schemes. However, past research has indicated that such computational efficiency gains are not always guaranteed (Keller and Evans, 2019), necessitating a direct comparison of the computational speeds of the AIMACI scheme and the numerical scheme. Given that the WRF-Chem, written in Fortran, is not conducive to GPU acceleration, we conducted offline tests of the AIMACI scheme's computational speed on a GPU and compared it with the numerical scheme on a CPU, where the AIMACI scheme was coupled into the WRF-Chem.

Figure 9 demonstrates that when utilizing a single CPU core, the AIMACI scheme achieves a computational speedup of approximately 5×, with a time cost of 48.51 seconds compared to the numerical scheme's 229.74 seconds. This advancement is further amplified when employing a single GPU, the AIMACI scheme completes the computation in a mere 0.83 seconds, which is approximately 277 × faster than the numerical scheme running on a single CPU core. Although we have not yet tested the online simulation speed of the physics-AI hybrid model on a GPU, it is reasonable to anticipate that future implementation of heterogeneous computing platforms, integrating both CPUs and GPUs, will yield significant enhancements in computational efficiency.

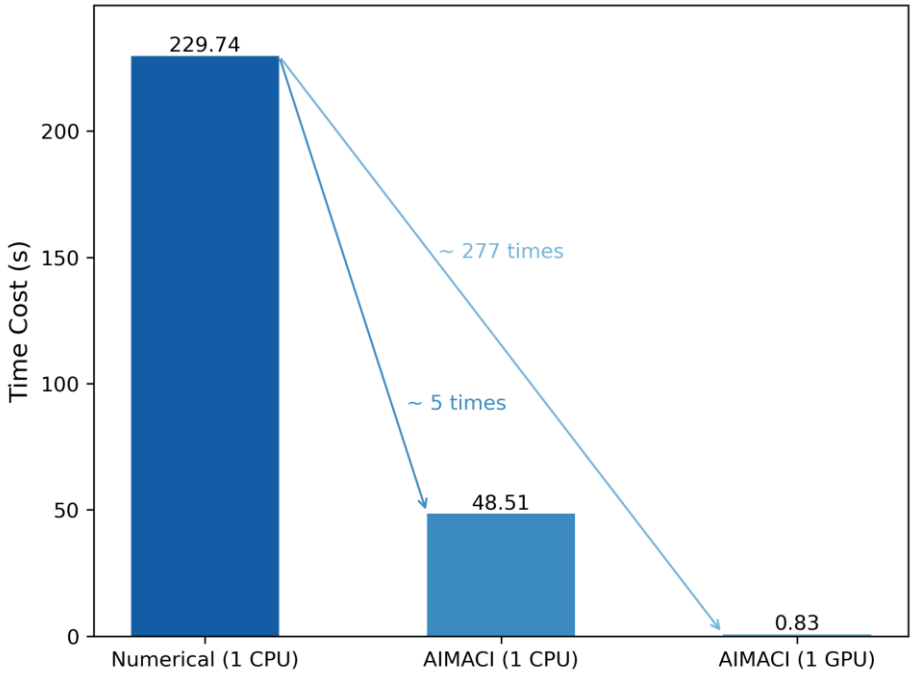


Figure 9: Comparison of computational speeds between the numerical scheme and the AIMACI scheme under different computational configurations. The time cost for the GPU is measured in a mode where the AIMACI scheme is not yet coupled to the model, while the time cost for the CPU is measured in a mode with the AIMACI scheme coupled into the model. The calculations are based on simulating the concentrations of 37 chemical species across 720,300 grid cells.

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

4 Conclusions

This study develops and evaluates a novel Artificial Intelligence Model for Aerosol Chemistry and Interactions, termed AIMACI, with a special focus on addressing the long-standing challenge of significant computational burden associated with enabling numerical scheme for simulating atmospheric aerosol chemistry and interactions in atmospheric models. The differential equations governing aerosol chemistry are notably stiff, coupled with a stringent time integration scheme required for numerical stability, resulting in limited breakthroughs in simulation speed with available numerical techniques. While previous studies have explored AI schemes as alternatives for conventional numerical schemes for simulating photochemical processes in atmospheric models, the use of AI schemes for simulating aerosol chemistry and interactions has not been studied. Moreover, atmospheric aerosol chemistry and interaction processes encompass not only chemical reactions but also a suite of highly nonlinear processes, including nucleation, coagulation, thermodynamics and phase equilibrium, gas-particle partitioning, and particle transfer. This study, therefore, aim to addresses the critical question of whether an AI scheme can effectively supplant the entire numerical scheme for these processes, achieving both high fidelity in simulation accuracy and a marked increase in computational efficiency.

To this end, the AIMACI scheme was established based on the state-of-the-art Multi-Head Self-Attention algorithm, renowned for its powerful nonlinear representation capabilities. This algorithm can efficiently capture complex reaction relationships between different chemical species, remaining robust even as the number of simulated species increases. In an offline mode, where the AIMACI scheme was not yet integrated with a 3D numerical model, it demonstrated remarkable statistical metrics on a test dataset, all 37 evaluated species exhibiting an average R^2 of 0.99, and an average Normalized Mean Bias (NMB) of 0.95%. This high degree of consistency between the numerical and AIMACI schemes lays a solid foundation for further online continuous simulations.

To facilitate the coupling of the Python-written AIMACI scheme with Fortran-based numerical models, we utilized PyTorch's TorchScript and LibTorch tools to encapsulate the AIMACI scheme into a static library callable by the numerical model. This approach entails minimal changes to the existing numerical model's codebase and offers a highly flexible and easily plug-and-play solution for coupling AI algorithms of diverse complexities with a range of numerical models.

Employing hybrid model with physics and AI schemes, we implemented additional experiments to evaluate the online simulation performance of the AIMACI scheme. The 10-day continuous simulation results indicate that the AIMACI scheme not only accurately captures the spatiotemporal distribution of various aerosol species but also effectively reproduces their size distributions, maintaining stability throughout the simulation period without rapid error growth. Furthermore, the AIMACI scheme exhibits robust generalization capabilities, applicable across various environmental conditions in all four seasons for month-long continuous simulations, despite being trained on data from only 16 days in March. The simulation results for nitrate's monthly average surface total concentrations and hourly time series illustrate a high degree of consistency with the numerical scheme. However, the AIMACI scheme's performance in July was less accurate compared to other

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

months, potentially due to the prevalence of typhoons in summer, which differ from the environmental conditions represented in the training data. This indicates that there is potential for further refinement to enhance the simulation accuracy of the AIMACI scheme, such as by incorporating a more diverse training dataset that includes extreme events.

In terms of computational speed, the AIMACI scheme is approximately 5 times faster than the conventional numerical scheme when predicting 720,300 grid points with 37 chemical species using a single CPU core. This speedup increases significantly to about 277 times faster when utilizing a GPU. Future simulations on heterogeneous platforms, integrating both CPUs and GPUs, are expected to further improve this speedup ratio. This anticipated enhancement will enable higher spatial resolutions, extended simulation durations, and substantially reduced computation times for chemical transport models.

An important outcome of this work is the first-time successful application of an AI scheme to replace entire numerical scheme for aerosol chemistry and interactions within the numerical model, achieving fast, accurate, and stable simulations. The high fidelity in reproducing the complex spatiotemporal distributions and PSD of aerosol species including water content in aerosols, coupled with a significant acceleration, highlights the potential of the AIMACI scheme in advancing climate modeling and atmospheric science. As the first step, we concentrate on inorganic aerosols, which are a fundamental aspect of atmospheric chemistry. While organic aerosols also play a crucial role, the inherent chemical complexity and the absence of a convincing numerical scheme for AI scheme to emulate have led us to defer their inclusion. However, the exploration of AI scheme for organic aerosols is firmly on our agenda for future studies, with the aim of integrating them into our comprehensive AIMACI scheme. While our application was constrained to have no global long-term simulations, the convenience and flexibility of the method we employed in this study—coupling the AIMACI scheme with the numerical model via a static library—as well as the robust generalization ability demonstrated by the AIMACI scheme, provide substantial promise for successfully conducting global long-term simulations in the near future. This not only helps to reduce uncertainties in global models due to simplified or absent detailed aerosol chemistry and interactions but also advances our comprehension of aerosol behavior, vital for climate studies and policy-making. By harnessing the power of artificial intelligence in conjunction with sophisticated atmospheric models, we aspire to achieve faster, more accurate, and reliable predictions of future climate scenarios.

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

Code availability

The hybrid version of WRF-Chem with both physics and AI schemes is publicly accessible. The source code for this hybrid model, along with the parameter file for the Multi-Head Self-Attention model used in the Artificial Intelligence Model for Aerosol chemistry and Interactions (AIMACI) is available under an open license at https://zenodo.org/records/13736859.

Author contributions

Zihan Xia and Chun Zhao designed the experiments, conducted and analyzed the simulations. All authors contributed to the discussion and final version of the paper.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

This research was supported by the National Key Research and Development Program of China (No. 2022YFC3700701), the Strategic Priority Research Program of Chinese Academy of Sciences (XDB0500303), National Natural Science Foundation of China (41775146), the USTC Research Funds of the Double First-Class Initiative (YD2080002007, KY2080000114), the Science and Technology Innovation Project of Laoshan Laboratory (LSKJ202300305), the National Key Scientific and Technological Infrastructure project "Earth System Numerical Simulation Facility" (EarthLab), and CMA-USTC Laboratory of Fengyun Remote Sensing. The study used the computing resources from the Supercomputing Center of the University of Science and Technology of China (USTC) and the Qingdao Supercomputing and Big Data Center.

References

Arfin, T., Pillai, A. M., Mathew, N., Tirpude, A., Bang, R., and Mondal, P.: An overview of atmospheric aerosol and their effects on human health, Environ Sci Pollut Res, 30, 125347–125369, https://doi.org/10.1007/s11356-023-29652-w, 2023.

Bellouin, N., Quaas, J., Gryspeerdt, E., Kinne, S., Stier, P., Watson-Parris, D., Boucher, O., Carslaw, K. S., Christensen, M., Daniau, A.-L., Dufresne, J.-L., Feingold, G., Fiedler, S., Forster, P., Gettelman, A., Haywood, J. M., Lohmann, U., Malavelle, F., Mauritsen, T., McCoy, D. T., Myhre, G., Mülmenstädt, J., Neubauer, D., Possner, A., Rugenstein, M., Sato,

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

Y., Schulz, M., Schwartz, S. E., Sourdeval, O., Storelvmo, T., Toll, V., Winker, D., and Stevens, B.: Bounding Global Aerosol Radiative Forcing of Climate Change, Reviews of Geophysics, 58, e2019RG000660, https://doi.org/10.1029/2019RG000660, 2020.

Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., and Tian, Q.: Accurate medium-range global weather forecasting with 3D neural networks, Nature, 619, 533–538, https://doi.org/10.1038/s41586-023-06185-3, 2023.

Calvert, J. G., Lazrus, A., Kok, G. L., Heikes, B. G., Walega, J. G., Lind, J., and Cantrell, C. A.: Chemical mechanisms of acid generation in the troposphere, Nature, 317, 27–35, https://doi.org/10.1038/317027a0, 1985.

Carmichael, G. R., Sandu, A., Song, C. H., He, S., Phandis, M. J., Daescu, D., Damian-Iordache, V., and Potra, F. A.: Computational Challenges of Modeling Interactions Between Aerosol and Gas Phase Processes in Large Scale Air Pollution Models, in: Large Scale Computations in Air Pollution Modelling, edited by: Zlatev, Z., Brandt, J., Builtjes, P. J. H., Carmichael, G., Dimov, I., Dongarra, J., van Dop, H., Georgiev, K., Hass, H., and Jose, R. S., Springer Netherlands, Dordrecht, 99–136, https://doi.org/10.1007/978-94-011-4570-1_10, 1999.

Charlson, R. J., Schwartz, S. E., Hales, J. M., Cess, R. D., Coakley, J. A., Hansen, J. E., and Hofmann, D. J.: Climate Forcing by Anthropogenic Aerosols, Science, 255, 423–430, https://doi.org/10.1126/science.255.5043.423, 1992.

Chen, K., Han, T., Gong, J., Bai, L., Ling, F., Luo, J.-J., Chen, X., Ma, L., Zhang, T., Su, R., Ci, Y., Li, B., Yang, X., and Ouyang, W.: FengWu: Pushing the Skillful Global Medium-range Weather Forecast beyond 10 Days Lead, https://doi.org/10.48550/arXiv.2304.02948, 6 April 2023.

Du, Q., Zhao, C., Zhang, M., Dong, X., Chen, Y., Liu, Z., Hu, Z., Zhang, Q., Li, Y., Yuan, R., and Miao, S.: Modeling diurnal variation of surface PM2.5 concentrations over East China with WRF-Chem: impacts from boundary-layer mixing and anthropogenic emission, Atmospheric Chemistry and Physics, 20, 2839–2863, https://doi.org/10.5194/acp-20-2839-2020, 2020.

Ebel, A., Memmesheimer, M., Friese, E., Jakobs, H. J., Feldmann, H., Kessler, C., and Piekorz, G.: Long-Term Atmospheric Aerosol Simulations — Computational and Theoretical Challenges, in: Large-Scale Scientific Computing, Berlin, Heidelberg, 490–497, https://doi.org/10.1007/11666806_56, 2006.

Fuzzi, S., Baltensperger, U., Carslaw, K., Decesari, S., Denier van der Gon, H., Facchini, M. C., Fowler, D., Koren, I., Langford, B., Lohmann, U., Nemitz, E., Pandis, S., Riipinen, I., Rudich, Y., Schaap, M., Slowik, J. G., Spracklen, D. V., Vignati, E., Wild, M., Williams, M., and Gilardoni, S.: Particulate matter, air quality and climate: lessons learned and future needs, Atmospheric Chemistry and Physics, 15, 8217–8299, https://doi.org/10.5194/acp-15-8217-2015, 2015.

Gu, J., Feng, J., Hao, X., Fang, T., Zhao, C., An, H., Chen, J., Xu, M., Han, W., Yang, C., Li, F., and Chen, D.: Establishing a non-hydrostatic global atmospheric modeling system at 3-km horizontal resolution with aerosol feedbacks on the Sunway supercomputer of China, Science Bulletin, 67, 1170–1181, https://doi.org/10.1016/j.scib.2022.03.009, 2022.

Hodas, N., Sullivan, A. P., Skog, K., Keutsch, F. N., Collett, J. L. Jr., Decesari, S., Facchini, M. C., Carlton, A. G., Laaksonen, A., and Turpin, B. J.: Aerosol Liquid Water Driven by Anthropogenic Nitrate: Implications for Lifetimes of

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

Water-Soluble Organic Gases and Potential for Secondary Organic Aerosol Formation, Environ. Sci. Technol., 48, 11127–11136, https://doi.org/10.1021/es5025096, 2014.

Hu, Z., Huang, J., Zhao, C., Bi, J., Jin, Q., Qian, Y., Leung, L. R., Feng, T., Chen, S., and Ma, J.: Modeling the contributions of Northern Hemisphere dust sources to dust outflow from East Asia, Atmospheric Environment, 202, 234–243, https://doi.org/10.1016/j.atmosenv.2019.01.022, 2019.

Keller, C. A. and Evans, M. J.: Application of random forest regression to the calculation of gas-phase chemistry within the GEOS-Chem chemistry model v10, Geoscientific Model Development, 12, 1209–1225, https://doi.org/10.5194/gmd-12-1209-2019, 2019.

Kelp, M. M., Tessum, C. W., and Marshall, J. D.: Orders-of-magnitude speedup in atmospheric chemistry modeling through neural network-based emulation, https://doi.org/10.48550/arXiv.1808.03874, 11 August 2018.

Kelp, M. M., Jacob, D. J., Kutz, J. N., Marshall, J. D., and Tessum, C. W.: Toward Stable, General Machine-Learned Models of the Atmospheric Chemical System, Journal of Geophysical Research: Atmospheres, 125, e2020JD032759, https://doi.org/10.1029/2020JD032759, 2020.

Kelp, M. M., Jacob, D. J., Lin, H., and Sulprizio, M. P.: An Online-Learned Neural Network Chemical Solver for Stable Long-Term Global Simulations of Atmospheric Chemistry, Journal of Advances in Modeling Earth Systems, 14, e2021MS002926, https://doi.org/10.1029/2021MS002926, 2022.

Landgrebe, J. D. and Pratsinis, S. E.: A discrete-sectional model for particulate production by gas-phase chemical reaction and aerosol coagulation in the free-molecular regime, Journal of Colloid and Interface Science, 139, 63–86, https://doi.org/10.1016/0021-9797(90)90445-T, 1990.

Lee, L. A., Reddington, Kenneth, and Carslaw, S.: On the relationship between aerosol model uncertainty and radiative forcing uncertainty, Proceedings of the National Academy of Sciences, 113, 5820–5827, https://doi.org/10.1073/pnas.1507050113, 2016.

Li, J., Carlson, B. E., Yung, Y. L., Lv, D., Hansen, J., Penner, J. E., Liao, H., Ramaswamy, V., Kahn, R. A., Zhang, P., Dubovik, O., Ding, A., Lacis, A. A., Zhang, L., and Dong, Y.: Scattering and absorbing aerosols in the climate system, Nat Rev Earth Environ, 3, 363–379, https://doi.org/10.1038/s43017-022-00296-7, 2022.

Liu: Emulation of an atmospheric gas-phase chemistry solver through deep learning: Case study of Chinese Mainland, Atmospheric Pollution Research, 12, 101079, https://doi.org/10.1016/j.apr.2021.101079, 2021.

Liu, Y., Wu, Z., Huang, X., Shen, H., Bai, Y., Qiao, K., Meng, X., Hu, W., Tang, M., and He, L.: Aerosol Phase State and Its Link to Chemical Composition and Liquid Water Content in a Subtropical Coastal Megacity, Environ. Sci. Technol., 53, 5027–5033, https://doi.org/10.1021/acs.est.9b01196, 2019.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B.: Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows, Proceedings of the IEEE/CVF International Conference on Computer Vision, 10012–10022, 2021.

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

Lohmann, U. and Feichter, J.: Global indirect aerosol effects: a review, Atmospheric Chemistry and Physics, 5, 715–737, https://doi.org/10.5194/acp-5-715-2005, 2005.

Nguyen, T. K. V., Zhang, Q., Jimenez, J. L., Pike, M., and Carlton, A. G.: Liquid Water: Ubiquitous Contributor to Aerosol Mass, Environ. Sci. Technol. Lett., 3, 257–263, https://doi.org/10.1021/acs.estlett.6b00167, 2016.

Parrish, D. D., Law, K. S., Staehelin, J., Derwent, R., Cooper, O. R., Tanimoto, H., Volz-Thomas, A., Gilge, S., Scheel, H.-E., Steinbacher, M., and Chan, E.: Long-term changes in lower tropospheric baseline ozone concentrations at northern mid-latitudes, Atmospheric Chemistry and Physics, 12, 11485–11504, https://doi.org/10.5194/acp-12-11485-2012, 2012.

Penkett, S. A., Jones, B. M. R., Brich, K. A., and Eggleton, A. E. J.: The importance of atmospheric ozone and hydrogen peroxide in oxidising sulphur dioxide in cloud and rainwater, Atmospheric Environment (1967), 13, 123–137, https://doi.org/10.1016/0004-6981(79)90251-8, 1979.

Pöschl, U.: Atmospheric Aerosols: Composition, Transformation, Climate and Health Effects, Angewandte Chemie International Edition, 44, 7520–7540, https://doi.org/10.1002/anie.200501122, 2005.

Pratsinis, S. E.: Simultaneous nucleation, condensation, and coagulation in aerosol reactors, Journal of Colloid and Interface Science, 124, 416–427, https://doi.org/10.1016/0021-9797(88)90180-4, 1988.

Saiz-Lopez, A., Borge, R., Notario, A., Adame, J. A., Paz, D. de la, Querol, X., Artiñano, B., Gómez-Moreno, F. J., and Cuevas, C. A.: Unexpected increase in the oxidation capacity of the urban atmosphere of Madrid, Spain, Sci Rep, 7, 45956, https://doi.org/10.1038/srep45956, 2017.

Sandu, A., Verwer, J. G., Blom, J. G., Spee, E. J., Carmichael, G. R., and Potra, F. A.: Benchmarking stiff ode solvers for atmospheric chemistry problems II: Rosenbrock solvers, Atmospheric Environment, 31, 3459–3472, https://doi.org/10.1016/S1352-2310(97)83212-8, 1997a.

Sandu, A., Verwer, J. G., Van Loon, M., Carmichael, G. R., Potra, F. A., Dabdub, D., and Seinfeld, J. H.: Benchmarking stiff ode solvers for atmospheric chemistry problems-I. implicit vs explicit, Atmospheric Environment, 31, 3151–3166, https://doi.org/10.1016/S1352-2310(97)00059-9, 1997b.

Sharma, H., Shrivastava, M., and Singh, B.: Physics informed deep neural network embedded in a chemical transport model for the Amazon rainforest, npj Clim Atmos Sci, 6, 1–10, https://doi.org/10.1038/s41612-023-00353-y, 2023.

Spracklen, D. V., Carslaw, K. S., Merikanto, J., Mann, G. W., Reddington, C. L., Pickering, S., Ogren, J. A., Andrews, E., Baltensperger, U., Weingartner, E., Boy, M., Kulmala, M., Laakso, L., Lihavainen, H., Kivekäs, N., Komppula, M., Mihalopoulos, N., Kouvarakis, G., Jennings, S. G., O'Dowd, C., Birmili, W., Wiedensohler, A., Weller, R., Gras, J., Laj, P., Sellegri, K., Bonn, B., Krejci, R., Laaksonen, A., Hamed, A., Minikin, A., Harrison, R. M., Talbot, R., and Sun, J.: Explaining global surface aerosol number concentrations in terms of primary emissions and particle formation, Atmospheric Chemistry and Physics, 10, 4775–4793, https://doi.org/10.5194/acp-10-4775-2010, 2010.

Twomey, S.: Pollution and the planetary albedo, Atmospheric Environment (1967), 8, 1251–1256, https://doi.org/10.1016/0004-6981(74)90004-3, 1974.

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. ukasz, and Polosukhin, I.: Attention is All you Need, in: Advances in Neural Information Processing Systems, 2017.

Wang, J. L., Curtis, J. H., Riemer, N., and West, M.: Learning Coagulation Processes With Combinatorial Neural Networks, Journal of Advances in Modeling Earth Systems, 14, e2022MS003252, https://doi.org/10.1029/2022MS003252, 2022.

Wang, Z., Li, J., Wu, L., Zhu, M., Zhang, Y., Ye, Z., and Wang, Z.: Deep learning-based gas-phase chemical kinetics kernel emulator: Application in a global air quality simulation case, Front. Environ. Sci., 10, https://doi.org/10.3389/fenvs.2022.955980, 2022.

Wu, Z., Wang, Y., Tan, T., Zhu, Y., Li, M., Shang, D., Wang, H., Lu, K., Guo, S., Zeng, L., and Zhang, Y.: Aerosol Liquid Water Driven by Anthropogenic Inorganic Salts: Implying Its Key Role in Haze Formation over the North China Plain, Environ. Sci. Technol. Lett., 5, 160–166, https://doi.org/10.1021/acs.estlett.8b00021, 2018.

Xia, Z., Zhao, C., Du, Q., Yang, Z., Zhang, M., Qiao, L.: Advancing Photochemistry Simulation in WRF-Chem V4.0: Artificial Intelligence PhotoChemistry (AIPC) Scheme with Multi-Head Self-Attention Algorithm, Journal of Advances in Modeling Earth Systems, submitted, 2024.

Zaveri, R. A. and Peters, L. K.: A new lumped structure photochemical mechanism for large-scale applications, Journal of Geophysical Research: Atmospheres, 104, 30387–30415, https://doi.org/10.1029/1999JD900876, 1999.

Zaveri, R. A., Easter, R. C., and Wexler, A. S.: A new method for multicomponent activity coefficients of electrolytes in aqueous atmospheric aerosols, Journal of Geophysical Research: Atmospheres, 110, https://doi.org/10.1029/2004JD004681, 2005.

Zaveri, R. A., Easter, R. C., Fast, J. D., and Peters, L. K.: Model for Simulating Aerosol Interactions and Chemistry (MOSAIC), Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2007JD008782, 2008.

Zhang, H., Sharma, G., Dhawan, S., Dhanraj, D., Li, Z., and Biswas, P.: Comparison of discrete, discrete-sectional, modal and moment models for aerosol dynamics simulations, Aerosol Science and Technology, 54, 739–760, https://doi.org/10.1080/02786826.2020.1723787, 2020.

Zhang, M., Zhao, C., Yang, Y., Du, Q., Shen, Y., Lin, S., Gu, D., Su, W., and Liu, C.: Modeling sensitivities of BVOCs to different versions of MEGAN emission schemes in WRF-Chem (v3.6) and its impacts over eastern China, Geoscientific Model Development, 14, 6155–6175, https://doi.org/10.5194/gmd-14-6155-2021, 2021.

Zhao, C., Ruby Leung, L., Easter, R., Hand, J., and Avise, J.: Characterization of speciated aerosol direct radiative forcing over California, Journal of Geophysical Research: Atmospheres, 118, 2372–2388, https://doi.org/10.1029/2012JD018364, 2013a.

Zhao, C., Chen, S., Leung, L. R., Qian, Y., Kok, J. F., Zaveri, R. A., and Huang, J.: Uncertainty in modeling dust mass balance and radiative forcing from size parameterization, Atmospheric Chemistry and Physics, 13, 10733–10753, https://doi.org/10.5194/acp-13-10733-2013, 2013b.

Zhao, C., Hu, Z., Qian, Y., Ruby Leung, L., Huang, J., Huang, M., Jin, J., Flanner, M. G., Zhang, R., Wang, H., Yan, H., Lu, Z., and Streets, D. G.: Simulating black carbon and dust and their radiative forcing in seasonal snow: a case study over North

https://doi.org/10.5194/egusphere-2024-2860
Preprint. Discussion started: 14 October 2024

China with field campaign measurements, Atmospheric Chemistry and Physics, 14, 11475–11491, https://doi.org/10.5194/acp-14-11475-2014, 2014.

Zhao, C., Huang, M., Fast, J. D., Berg, L. K., Qian, Y., Guenther, A., Gu, D., Shrivastava, M., Liu, Y., Walters, S., Pfister, G., Jin, J., Shilling, J. E., and Warneke, C.: Sensitivity of biogenic volatile organic compounds to land surface parameterizations and vegetation distributions in California, Geoscientific Model Development, 9, 1959–1976, https://doi.org/10.5194/gmd-9-1959-2016, 2016.