Point-to-Point Responses to Reviewer's Comments

We would like to thank for reviewer's thoughtful comments on our manuscript.

Lyu et al. put forth the 'FastCTM' model which seems to be a reduced complexity model that discretizes changes in concentrations for 10 air pollution species. Though interesting, the presentation of methods, results, and context of the study needs to be heavily refined before being accepted. The details of the study are currently not sufficient as they stand.

Introduction:

"The air pollutant and species concentrations can be then calculated by solving these complicated equations with numeric methods (Byun and Schere, 2006), which is often time-consuming and requires intense computational resources." --> This thought is not very well flushed out. A single reference from 2006 does not detail at all what makes these models computationally expensive.

Response: This sentence is revised with two more related references are added, as follows.

The air pollutant concentrations can be then calculated by solving these complicated equations with numeric methods (Byun and Schere, 2006), which is often time-consuming (Leal et al., 2017) and require intense computational resources such as high-performance computing (Efstathiou et al., 2024).

Leal, A. M., Kulik, D. A., Smith, W. R., and Saar, M. O.: An overview of computational methods for chemical equilibrium and kinetic calculations for geochemical and reactive transport modeling, Pure and Applied Chemistry, 89, 597-643, 2017.

Efstathiou, C. I., Adams, E., Coats, C. J., Zelt, R., Reed, M., McGee, J., Foley, K. M., Sidi, F. I., Wong, D. C., and Fine, S.: Enabling high-performance cloud computing for the Community Multiscale Air Quality Model (CMAQ) version 5.3. 3: performance evaluation and benefits for the user community, Geoscientific Model Development, 17, 7001-7027, 2024.

"Quantifying the contributions of individual processes would provide fundamental explanations for a model's predictions, and therefore is also useful in identifying potential sources of error in the model formulation or its inputs (Liu et al., 2010)." --> I find this introduction quite poor. The authors provide minimal examples of emulating entire CTMs but give no examples of using ML to emulate and replace CTM model components which there are many for chemistry, photolysis, deposition, etc. This needs much greater discussion as it shows a lack of awareness by the authors of what currently exists, below of which are only several examples:

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., and Chalikov, D. V.: New Approach to Calculation of Atmospheric Model Physics: Accurate and Fast Neural Network Emulation of Longwave Radiation in a Climate Model, Monthly Weather Review, 133, 1370–1383, https://doi.org/10.1175/MWR2923.1, 2005.

Kelp, M. M., Jacob, D. J., Lin, H., and Sulprizio, M. P.: An Online-Learned Neural Network Chemical Solver for Stable LongTerm Global Simulations of Atmospheric Chemistry, Journal of Advances in

Modeling Earth Systems, 14, e2021MS002926, https://doi.org/10.1029/2021MS002926, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021MS002926, 2022.

Xia, Z., Zhao, C., Du, Q., Yang, Z., Zhang, M., and Qiao, L.: Advancing Photochemistry Simulation in WRF-Chem V4.0: Artificial Intelligence PhotoChemistry (AIPC) Scheme with Multi-Head Self-Attention Algorithm, https://www.authorea.com/users/816476/articles/1217166-advancing-photochemistry-simulation-in-wrf-chem-v4-0-artificial-intelligence-photochemistry-aipc-scheme-with-multi-head-self-attention-algorithm, 2024.

Zhong, X., Ma, Z., Yao, Y., Xu, L., Wu, Y., and Wang, Z.: WRF–ML v1.0: a bridge between WRF v4.3 and machine learning parameterizations and its application to atmospheric radiative transfer, Geoscientific Model Development, 16, 199–209, https://doi.org/10.5194/gmd16-199-2023, publisher: Copernicus GmbH, 2023.

Silva, S. J., Heald, C. L., Ravela, S., Mammarella, I., and Munger, J. W.: A Deep Learning Parameterization for Ozone Dry Deposition Velocities, Geophysical Research Letters, 46, 983–989, https://doi.org/10.1029/2018GL081049, tex.copyright: ©2018. American Geophysical Union. All Rights Reserved., 2019.

Response: Given the suggestive comments from the reviewer, we have added an independent paragraph in the introduction, to analyze related studies and progress, as follows,

Quantifying individual processes would provide fundamental explanations for a model's predictions, and therefore is also useful in identifying potential sources of error in the model formulation or its inputs (Liu et al., 2010). With this motivation, there are studies dedicated to developing models to learn one specific atmospheric process, i.e. chemical and deposition, in the CTM model. Kelp et al. (2022) developed a neural network chemical solver for stable long-term global simulations of atmospheric chemistry, learned from the GEOS-Chem model. Xia et al. (2024) simulated 74 chemical species and 229 reactions following the SAPRC-99 mechanism with an artificial intelligence photochemistry (AIPC) scheme to achieve around 8-time speed-up. Sturm and Wexler (2020) developed a mass- and energy-conserving framework for using machine learning to speed computations with a successful application in a photochemistry example. For the deposition process, Silva et al. (2019) proposed a deep learning parameterization for ozone dry deposition velocities with accurate predictions in independent new date sets, revealing the potential of neural network in encoding complex spatio-temporal processes. Liu et al. (2025) proposed a Neural Network Emulator, named ChemNNE, for fast chemical concentration modelling, which achieved good performance in accuracy and efficiency. Even though these successful applications using deep learning methods to simulate individual atmospheric chemical and physical processes, there is a missing gap in coupling these NN operator replacements together as a complete deep learning based CTM.

"process analysis" --> I don't know what this means

Response: Revised to "*internal chemical and physical process analysis*".

" Interpretations of the FastCTM are also widely vowed to improve deep learning model applications

in earth system science and climate studies. " --> Not sure how you can claim this given no evidence, more aspirational than substantive

Response: It's revised as follows, "*Interpretations of deep learning network are also widely vowed to improve their applications in earth system science and climate studies.*".

"The FastCTM is currently configured to simulate hourly concentrations of 10 pollutant variables, including and major species of PM2.5 (SO4 2−, NO3 −, NH4+, organic matters and other inorganic components, coarse part in PM10, CO, NO2, SO2 and O3." --> Not sure how many atmospheric chemists and climate scientists want a CTM with only ten species. Needs much more motivation. Even small chemical mechanisms in operational use have around ~70 species.

Response: We sincerely appreciate the reviewer's comment, as it raises an important point regarding the limited number of pollutant species in the FastCTM model. FastCTM is designed to address real-time air quality forecasting, where operational usage is critical. The 10 species were selected based on their direct relevance to regulatory standards (e.g., $PM_{2.5}$, PM_{10}, O_3, NO_2, SO_2, CO) and their dominance in driving health and environmental impacts in urban and industrial regions (e.g., China, where $PM_{2.5}$ components like SO_4^{2-}, NO_3^-, and NH_4^+ account for most of the fine aerosol mass). By prioritizing these species, FastCTM balances accuracy with computational speed, making it suitable for rapid decision-making in policy and emergency response scenarios. While traditional CTMs (e.g., CMAQ) include ~70 species for comprehensive chemical analysis, operational forecasts often focus on criteria pollutants and key $PM_{2.5}$ components due to their regulatory importance. FastCTM replicates the outputs most frequently used in air quality management, ensuring compatibility with existing regulatory frameworks. Besides, FastCTM's performance was validated against both CMAQ simulations and ground observations (Sect. 3.1–3.2). Results show high agreement for all 10 species (R^2 = 0.7–0.9), confirming that the selected variables adequately represent key atmospheric processes. We acknowledge that FastCTM may benefit from expanded mechanisms with detailed gas-phase chemistry or aerosol microphysics. FastCTM's design supports incremental integration of additional species (e.g., via user-defined modules) without overhauling the core framework. Future versions will explore adding VOCs and secondary organics to address broader research needs.

We clarified the motivation for the 10-species configuration in the Introduction and Section 2.1 and Section 4, emphasizing regulatory and operational priorities driving species selection and plans for modular expansion in future work, as follows,

The 10 species were selected based on their direct relevance to regulatory standards (e.g., $PM_{2.5}$, PM_{10}, O_3, NO_2, SO_2, CO) and their dominance in driving health and environmental impacts in urban and industrial regions.

FastCTM's design supports incremental integration of additional species (e.g., via user-defined modules) without overhauling the core framework. Future versions will explore adding VOCs and secondary organics to address broader research needs.

Methods:

"CMAQ structures" --> I don't know what structures means here

-Is this predicting only surface level concentrations?

-I would not really call this model a CTM, this feels more like a reduced order model. There are potentially hundreds of chemical species that lead to the formation of PM2.5, O3, etc. And yet you do not mention the chemical mechanism at all in the WRF-CMAQ model. This work is basically mapping emissions to concentrations in a fairly naive way.

Response: To rule out the possibilities of FastCTM as a simple model mapping emissions to concentrations, we tested the land use regression (LUR) framework with machine learning models of random forest, XGBoost, and also a linear regression model. The input data for these LUR models include emissions, meteorological forecasts from WRF, and geophysical covariates, the same as those used in FastCTM. The LUR model carries out direct mapping from emission and weather data to 10 pollutants. Results have exhibited LUR's poor performance in predicting air pollutant concentrations. Related studies are included in Section 3.1, as follows.

To validate FastCTM model, three land use regression (LUR) models were constructed, namely the linear regression model, the random forest model (with the number of trees set at 500), and the XGBoost model (with the booster specified as gbtree). These LUR models were developed using the same input meteorological data, emission, and geophysical variables. When compared with the FastCTM model, the performance of the LUR models was found to be significantly inferior (as demonstrated in Figure S10 – S12 in the SI). This outcome is, in fact, anticipated when we consider the complex nature of air quality dynamics. Air quality is not a static entity, but it varies both spatially and temporally. For instance, the transport of air pollution is a highly dynamic process that hinges on wind fields and air pollution concentrations in a reciprocal manner. The wind direction and speed dictate the trajectory along which pollutants travel, while the existing pollutant concentrations in different regions influence the overall dispersion and mixing patterns. LUR models, which predominantly rely on local input data (Wong et al., 2021; Cheng et al., 2021), struggle to capture these intricate, non-local interactions. They lack the capacity to account for the far-reaching effects such as wind-driven pollutant transport and the consequential changes in air quality over larger geographical areas.

The supplementary Figures exhibit the performances of three machine learning models as follows.

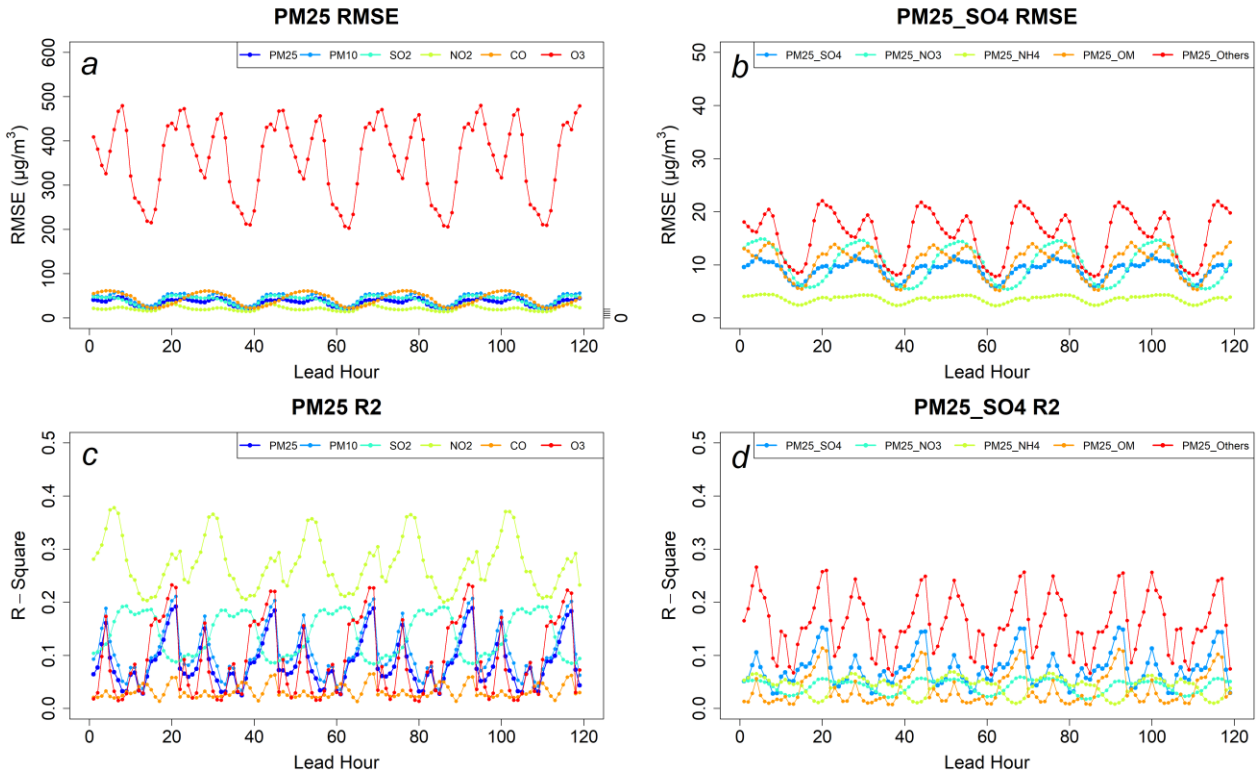

Figure S1: The evaluation performances of linear regression forecasts against CMAQ forecasts in 2023. Panel (a) and (b) respectively show RMSE values of criteria pollutants and the $PM_{2.5}$ components of. Panel (c) and (d) respectively show R^2 values. It should be noted that RMSE value of CO corresponds to the right axis in panel (a).

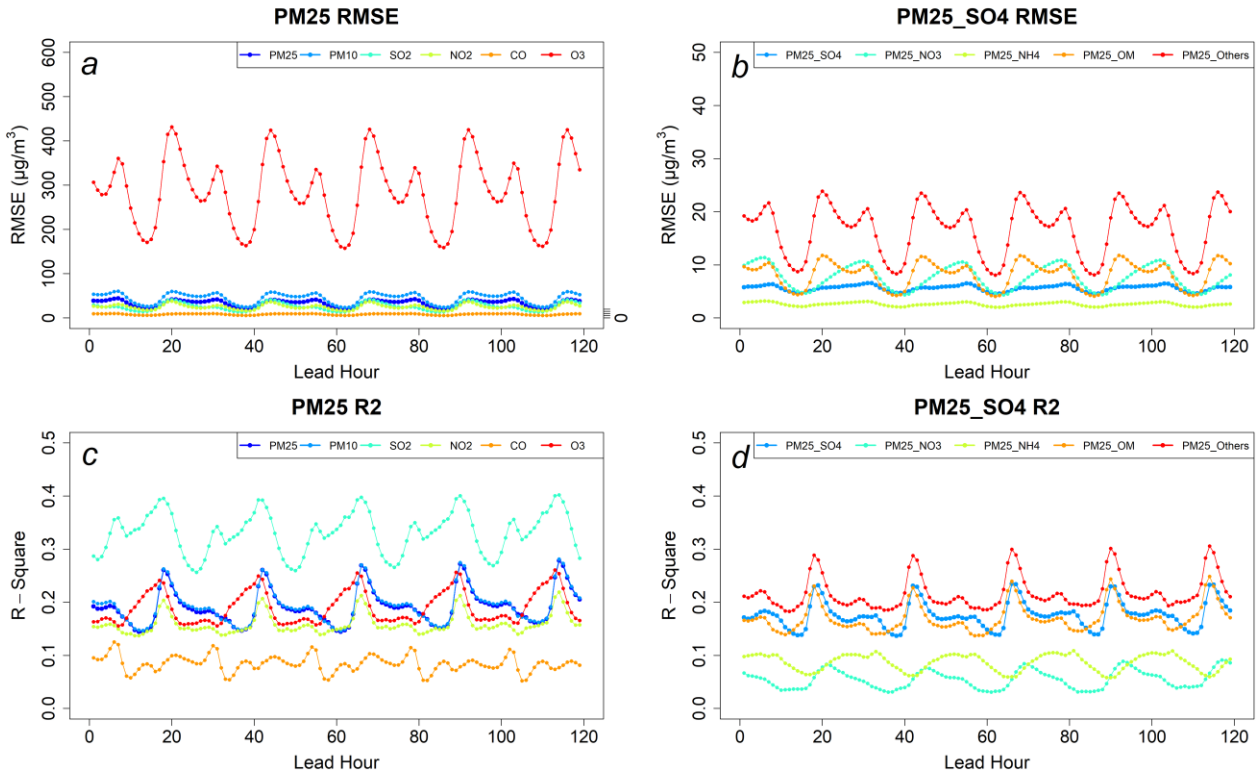

Figure S2: The evaluation performances of random forest forecasts against CMAQ forecasts in 2023. Panel (a) and (b) respectively show RMSE values of criteria pollutants and the $PM_{2.5}$ components of. Panel (c) and (d)

respectively show R^2 values. It should be noted that RMSE value of CO corresponds to the right axis in panel (a).

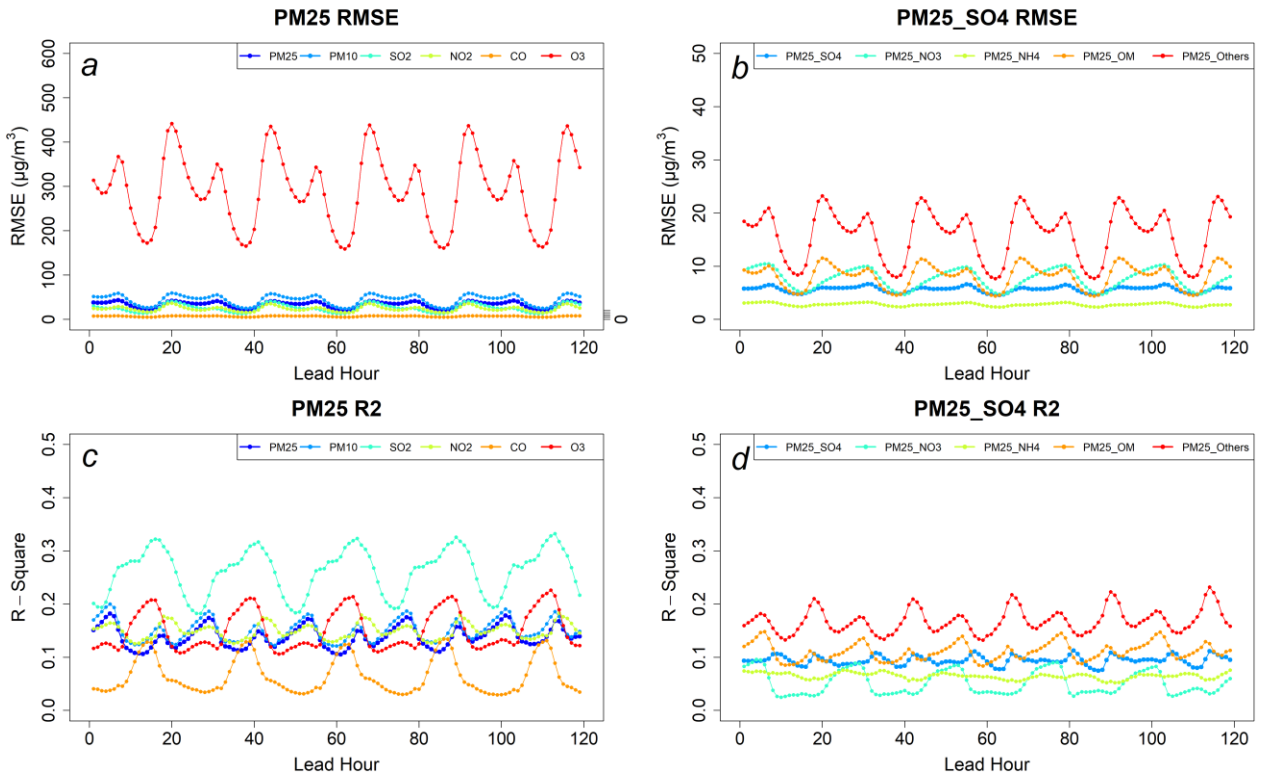

Figure S3: The evaluation performances of XGBoost forecasts against CMAQ forecasts in 2023. Panel (a) and (b) respectively show RMSE values of criteria pollutants and the $PM_{2.5}$ components of. Panel (c) and (d) respectively show R^2 values. It should be noted that RMSE value of CO corresponds to the right axis in panel (a).

" A detailed description of CMAQ principles is available elsewhere (Byun and Schere, 2006) " --> I find this lazy. This paper is 20 years old and I do not know what you would like the reader to find in it.

Response: The reference provided a detailed description of the theory, model framework, and numerical methods. We added another late review study to reflect more recent developments of the chemical transport model CMAQ.

Appel, K. W., Napelenok, S. L., Foley, K. M., Pye, H. O., Hogrefe, C., Luecken, D. J., Bash, J. O., Roselle, S. J., Pleim, J. E., and Foroutan, H.: Description and evaluation of the Community Multiscale Air Quality (CMAQ) modeling system version 5.1, Geoscientific model development, 10, 1703-1732, 2017.

"Chemical Reaction Module" --> This just sounds like a first order approximation using idealized rate constants. There is a very rich and long history of using ODE solvers to get the solution to complex chemical mechanisms. There really is not enough discussion with this module (or really any of the

preceding module sections). You are highly simplifying each of these processes without an underlying discussion of why you are doing so. There already exist data-driven and reduced complexity modeling systems that accomplish similar air quality regulation goals (e.g., InMAP, EASIUR, APEEP).

Response: The simplification of chemical kinetics in FastCTM is motivated by balancing data availability with physical interpretability. While traditional CTMs (e.g., CMAQ) use detailed ODE solvers for hundreds of species and reactions, FastCTM focuses on key variables and pathways for air quality dynamics, such as secondary inorganic aerosol formation (SO_4^{2-}, NO_3^-, NH_4^+) and ozone photochemistry.

Comparing to other reduced-form modelling systems, models like InMAP and EASIUR focus on annual-average exposure, while FastCTM provides hourly-resolved simulations critical for real-time management. Unlike reduced-form models that aggregate source impacts, FastCTM quantifies hourly contributions from individual processes (transport, chemistry, emissions) via its modular design. Furthermore, FastCTM explicitly couples meteorology (PBLH, T, RH) with chemistry, whereas InMAP/APEEP assumes static meteorology, limiting their utility in capturing diurnal or synoptic-scale variations. Therefore, FastCTM is more like a learnable CTM model in the neural network form with some simplifications in input variables and space domain (3D to 2D) by embedding physical principles (e.g., mass conservation in transport, Arrhenius-like rate dependencies in chemistry). Besides, the modular architecture of FastCTM allows incremental addition of species/reactions (e.g., VOC oxidation pathways) without retraining the entire model. Related discussion have been added in the section,

Reduced-form models like InMAP (Tessum et al., 2017) and EASIUR (Gentry et al., 2023) focus on annual-average exposure, while FastCTM provides hourly-resolved simulations for real-time management. FastCTM quantifies hourly contributions from individual processes (transport, chemistry, emissions) via its modular design, rather than aggregating source impacts (e.g., EASIUR's source-receptor matrices) in reduced-form models. Furthermore, FastCTM explicitly couples meteorology (PBLH, T, RH) with chemistry, whereas InMAP/APEEP (Muller and Mendelsohn, 2006) assumes static meteorology, limiting their utility in capturing diurnal or synoptic-scale variations.

And also in the Section 4,

Besides, FastCTM may also benefit from expanded mechanisms with detailed gas-phase chemistry or aerosol microphysics. FastCTM's design supports incremental integration of additional species (e.g., via user-defined modules) without overhauling the core framework. Future versions will explore adding VOCs and secondary organics to address broader research needs.

Also for the diffusion module and deposition module, related discussions have been added in the corresponding sections as follows,

Diffusion involves the physical and chemical processes that disperse pollutants in the atmosphere. It's influenced by meteorological conditions, i.e. atmospheric stability and humidity, and surface features, i.e. land terrains and vegetation (Jiang et al., 2021).

Air pollutant deposition refers to the process by which atmospheric pollutants are transferred to Earth's surfaces (land, water, vegetation) or removed from the air. This phenomenon plays a critical role in environmental pollution dynamics and ecosystem impacts. The deposition was closely

influenced by meteorological conditions and surface characteristics (Janhäll, 2015). For example, high wind disperses pollutants, while turbulence enhances dry deposition. Forests and crops act as sinks due to large surface areas for adsorption.

-I don't explicitly understand how this is a machine learning model. You describe a sequence-to-sequence modeling framework reminiscent of an LSTM, but no mention of memory or hyper parameters in general. The inclusion of these equations may seem more like a symbolic regression kind of ML framework, but the details are sparse and lack substance. Are all the modules trained jointly so that error influences each other? Chemistry is constantly affected by other modules (and vice versa) yet these interaction terms can't be considered during training at all. That is, how does error propagate from one time step to the next in training? Is the underlying WRF-CMAQ simulations two-way coupled such that weather influences chemistry and chemistry feedbacks via aerosol effects to influence the weather? Not enough details in the underlying simulations or the joint training of modules. There are examples of this kind of offline training here:

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.

Yang, X., Guo, L., Zheng, Z., Riemer, N., and Tessum, C. W.: Atmospheric chemistry surrogate modeling with sparse identification of nonlinear dynamics, https://doi.org/10.48550/arXiv.2401.06108, 2024.

Liu, Z.-S., Clusius, P., and Boy, M.: Neural Network Emulator for Atmospheric Chemical ODE, https://doi.org/10.48550/arXiv.2408.01829, 2024.

Response: (1) The five modules in FastCTM are defined in the form of operator, where operator parameters are estimated, rather than in the form of pure predictor mapping concentrations from one hour to the next. For example, in the diffusion module, FastCTM learns to encode diffusion coefficient K from meteorological conditions before performing upwind finite difference procedure to solve the diffusion process $\nabla(K\nabla C_i)$. It's also the same for processes such as reaction, advection, and deposition. Therefore, it is impossible for one process to represent all atmospheric processes simultaneously. The independent contribution of each process is depicted in Figure 12 of section 3.3. Each process exhibited its patterns of contribution to hourly air pollutant concentration changes, constrained by the form of the operator in the processes. The related description is added in Section 2.3 Model Training,

Even though five modules are defined in FastCTM, individual processes are not trained separately. The model was trained as a whole with hour-to-hour air pollutant concentrations, while each process could learn its parameters under the constrains of its dedicated formulation. Specifically, FastCTM was tuned to minimize the loss function $\mathcal{L}$, which was determined to be L2 loss (Bühlmann and Yu, 2003) of the regularized mean squared error (MSE) as shown in Eq. 15.

(2) The configuration of the parent model was added in Section 2.1

WRF-CMAQ simulations are not two-way coupled so that weather and chemistry do not have feedback to influence each other.

"The main objective of our study is to build and validate a principles-guided neural network based FastCTM that could simulate spatial-temporal fields of hourly concentrations of major air pollutant species like a traditional CTM. Besides, the FastCTM could model individual contributions from each of the atmospheric processes of transport, diffusion, deposition, reaction and emission. " --> this should be stated earlier. The term "principles-guided" is vague, and I don't really consider this 'like a traditional CTM'. You discretize the potential processes that affect air quality outputs, but this is more like a traditional reduced complexity model approach. I think a deeper review into the literature would help the authors situate their work in this established landscape.

Response: These two sentences are moved to the last paragraph of the Introduction section, to make it clearer for the general purpose of the study. Unlike the traditional reduced-from models, FastCTM is time-resolved with a 60 seconds step to simulate the evolution of air pollutants. It generates hourly air quality simulations based on hourly meteorological conditions and emissions. The simulations have good correlations with its parent numerical CTM model CMAQ. FastCTM also exhibited reasonable responses to emission changes, also in close agreement with that of CMAQ. Besides, internal atmospheric processes could also be checked to reflect specific contributions from each process. We have added a related literature review on the application of neural networks in simulating atmospheric processes in the introduction, as follows.

Quantifying individual processes would provide fundamental explanations for a model's predictions, and therefore is also useful in identifying potential sources of error in the model formulation or its inputs (Liu et al., 2010). With this motivation, there are studies dedicated to developing models to learn one specific atmospheric process, i.e. chemical and deposition, in the CTM model. Kelp et al. (2022) developed a neural network chemical solver for stable long-term global simulations of atmospheric chemistry, learned from the GEOS-Chem model. Xia et al. (2024) simulated 74 chemical species and 229 reactions following the SAPRC-99 mechanism with an artificial intelligence photochemistry (AIPC) scheme to achieve around 8-time speed-up. Sturm and Wexler (2020) developed a mass- and energy-conserving framework for using machine learning to speed computations with a successful application in a photochemistry example. For the deposition process, Silva et al. (2019) proposed a deep learning parameterization for ozone dry deposition velocities with accurate predictions in independent new date sets, revealing the potential of neural network in encoding complex spatio-temporal latent processes. Liu et al. (2025) proposed a Neural Network Emulator, named ChemNNE, for fast chemical concentration modelling, which achieved good performance in accuracy and efficiency. Even though these successful applications using deep learning methods to simulate individual atmospheric chemical and physical processes, there is an missing gap in coupling these NN operator replacements together as a complete deep learning based CTM.

Also in Section 2.3.5 Chemical Reaction Module, the comparison of FastCTM with reduced-form models is discussed as follows,

Reduced-form models like InMAP (Tessum et al., 2017) and EASIUR (Gentry et al., 2023) focus on annual-average exposure, while FastCTM provides hourly-resolved simulations for real-time management. FastCTM quantifies hourly contributions from individual processes (transport, chemistry, emissions) via its modular design, rather than aggregating source impacts (e.g., EASIUR's

source-receptor matrices) in reduced-form models. Furthermore, FastCTM explicitly couples meteorology (PBLH, T, RH) with chemistry, whereas InMAP/APEEP (Muller and Mendelsohn, 2006) assumes static meteorology, limiting their utility in capturing diurnal or synoptic-scale variations.

"Furthermore, CMAQ and FastCTM forecasts were both evaluated by hourly observations from national monitoring sites (as shown in Figure S5 in the supplementary material) for six criteria pollutants (PM2.5, PM10, SO2, NO2, CO, and O3)." --> What is the point of this if CMAQ is your ground truth?

Response: We agree with the reviewer's point that it does not make much sense to compare FastCTM to station observations. Related comparisons are removed from the manuscript.

Results:

"Besides, since the FastCTM is a 2-D model only considering atmospheric processes within the boundary layer, lower consistency with the CMAQ model during daytime could be due to more active vertical turbulence which is not fully represented." --> Isn't the point of having this processed-based emulation the ability to attribute errors to processes? This sounds hand-wavy and does not explain the variability very well

Response: We sincerely appreciate the reviewer's insightful feedback regarding the attribution of errors in FastCTM's daytime performance. The comment highlights a critical aspect of our process-based emulation framework and motivates a deeper exploration of error sources. FastCTM can simulate the contributions from each process. We do not have the CMAQ process analysis in the test period. Therefore, it is not possible to attribute FastCTM's simulation errors to specific processes, by comparing the process data of CMAQ. CMAQ uses a non-local closure scheme for vertical diffusion, explicitly resolving turbulent mixing across layers. FastCTM's 2D framework parameterizes this via horizontal diffusivity and PBLH, which cannot capture vertical advection or entrainment. Besides, we considered a 2-D model in FastCTM, which means process analysis could be different from that of CMAQ in its definition natures. We are going to apply FastCTM in 3D dimensions in the later version. We added further explanation in this section as follows,

Besides, since the FastCTM is a 2-D model only considering atmospheric processes within the boundary layer, lower consistency with the CMAQ model during the daytime could be due to more active vertical turbulence. Studies show that strong vertical mixing of air pollutants to the height above PBLH have been found (Li et al., 2017; Tang et al., 2016), which could not be fully represented in FastCTM.

"It is important to note that the relatively low R2 values observed for NH4+ can be attributed to the fact that it is the sole cation included in the FastCTM model without a corresponding acid-base balance, which may affect the model's predictive accuracy." --> I don't see how this is the reason. WRF-CMAQ has many base pairings that can neutralize NH4+ that are not represented here. I don't recall

conservation of mass as a constraint in your chemical module. Furthermore, how do you know that NH4+ does not precipitate out as it is very hydrophilic.

Response: We appreciate the reviewer's critical assessment of the NH_4+ prediction performance and agree that our initial explanation simplified the issue. FastCTM's chemical module (Eq. 11–12) approximates NH_4+ dynamics using a data-driven approach trained on CMAQ outputs. While CMAQ explicitly resolves NH_4+ formation via reactions with HNO3 ($NH_3 + HNO_3 \rightarrow NH_4NO_3$) and H_2SO_4 ($2NH_3 + H_2SO_4 \rightarrow (NH_4)_2SO_4$), FastCTM does not explicitly encode these pathways. Instead, the neural network learns relationships between NH_4+ and precursor emissions (NH_3, NOx, SO_2) and meteorological variables (e.g., temperature, humidity). This simplification omits acid-base equilibria and aerosol thermodynamics, which are critical for partitioning NH4+ between gas and particle phases. The reviewer correctly notes that FastCTM's chemical module does not enforce mass conservation. While CMAQ rigorously tracks nitrogen and sulfur species across gas, aerosol, and aqueous phases, FastCTM's neural network predicts NH_4+ concentrations directly from emissions and meteorology without explicit mass-balance constraints. This can lead to unphysical predictions, especially when precursor emissions (e.g., NH_3) are over/underestimated or when thermodynamic conditions (e.g., high humidity) favor aerosol formation. The low R^2 for NH_4+ primarily reflects FastCTM's simplified chemical mechanism, which lacks explicit acid-base pairing and aerosol thermodynamics. We have revised the text accordingly in Section 3.1, as follows,

While CMAQ explicitly resolves NH_4^+ formation reactions, FastCTM does not explicitly encode these pathways. Instead, the neural network implicitly learns relationships between NH_4^+ and precursor emissions (NH3, NOx, SO2) and meteorological variables (e.g., temperature, humidity). This simplification omits acid-base equilibria and aerosol thermodynamics, which are critical for partitioning NH_4^+ between gas and particle phases. The low R^2 for NH_4^+ primarily reflects FastCTM's simplified chemical mechanism in this part, which could be improved by adding related species in the simulation.

-I actually believe it is quite concerning that the RMSEs vary diurnally. You should also plot the WRF-CMAQ and FastCTM time series against each other. A diurnal error actually may suggest that you are not correctly learning the atmospheric dynamics of the system well. You may be predicting an average concentration across all time and that's why you see a diurnal error profile.

"FastCTM forecasts using zero values as input air quality data were almost the same as that using ordinary input in the long leading hours, indicating that FastCTM simulations in long leading hours are not affected by initial conditions, just like deterministic numeric CTMs (such as CMAQ)" --> This is hard to conclude, you need to plot actual concentration time series instead of RMSEs. It seems like the error is always the same, this could mean the FastCTM always predicts the same values given the time of day. More results need to be presented.

Response: As reviewer kindly pointed, FastCTM possibly has taken average pollutant concentration from five-year training data in 2018-2022. In order to confirm that FastCTM was able to predict air quality based on given meteorological conditions and emissions, daily average FastCTM simulation in the fifth leading day (leading hours 96-119) in the test year of 2023 is compared with daily average CMAQ simulations in 2023 and in the training years of 2018-2022. Results revealed that FastCTM

forecasts are generally in good correlation with CMAQ forecasts in 2023, rather than that in 2018-2023. It means FastCTM has learned the evolution rules of air pollutant concentrations, instead of just giving average air pollutant concentration according to time of the year. Related results have been added in the manuscript in section 3.1.1, as follows.

Annually, the daily air quality typically exhibits similar fluctuations to those in other years, which can be primarily attributed to the cyclical nature of meteorological conditions and pollutant emission patterns. The FastCTM model was trained using a comprehensive dataset spanning five years, from 2018 to 2022. In light of this, it was crucial to rule out the possibility that the model was merely reproducing historical averages during the test year of 2023. To this end, the daily national average concentrations of $PM_{2.5}$ and O_3 in 2023, as predicted by FastCTM, were meticulously compared with those simulated by CMAQ in the same test year, as well as with the CMAQ forecasts from the training years of 2018-2022. As illustrated in Figure 5, it becomes evident that the predictions made by FastCTM in 2023 align more closely with the actual CMAQ forecasts for that year, rather than with the forecasts generated from the training data of 2018-2022. This finding not only validates the adaptive learning capabilities of the FastCTM model but also indicates that the model is not resorting to a simplistic approach of taking the average concentration from the previous five years based on the time of day. Instead, it is likely incorporating real-time meteorological feedback, adjusting for any shifts in emission patterns, and leveraging its learned relationships to provide more accurate and contemporaneous predictions.

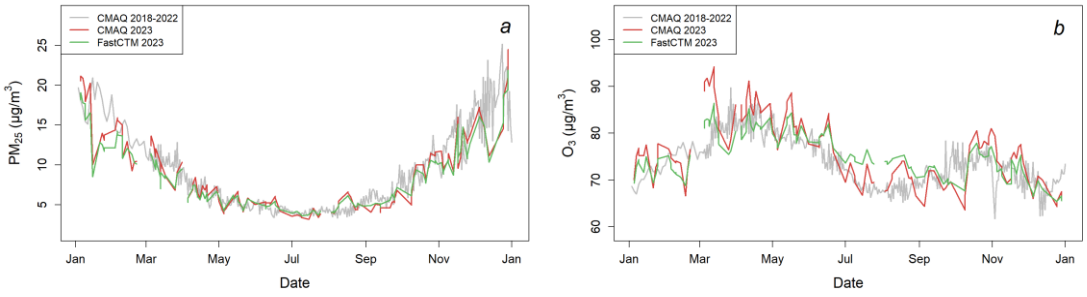

Figure 5: The timeseries of FastCTM forecasts against CMAQ forecasts.

Figure 3 is unwieldy. There are 60 mulitplots and not well labeled on the figure. Here you should show spatial differences in terms of both absolute and relative error. Seems like FastCTM does not capture the highest concentration values, which is concerning given that is the largest impact on health and climate. Hard to have any substantive discussion of results without any quantitative measure regarding Figure 3.

Response: As the reviewer suggested, we revised the manuscript in the section as follows,

The spatial distributions of the mean absolute error (MAE) and the normalized mean absolute error (NMAE) are presented in Figure 3. For the six criteria pollutants, the MAE values are higher in polluted areas. This could be attributed to the complex and dynamic nature of pollutant interactions in such regions. In polluted environments, there are often multiple sources of emissions, complex chemical reactions, and variable meteorological conditions that can lead to greater discrepancies between the model-predicted and actual pollutant concentrations. Conversely, the NMAE values

exhibit an opposite trend, being lower in polluted areas. In these regions, the NMAE values are typically around 0.2, in contrast to the relatively higher values of approximately 1 in cleaner areas. The NMAE is a normalized metric that takes into account the magnitude of the actual pollutant concentrations. A lower NMAE in areas with high pollution levels suggests that the FastCTM model is effectively capturing the overall magnitude and trends of pollutant concentrations relative to the reference CMAQ model.

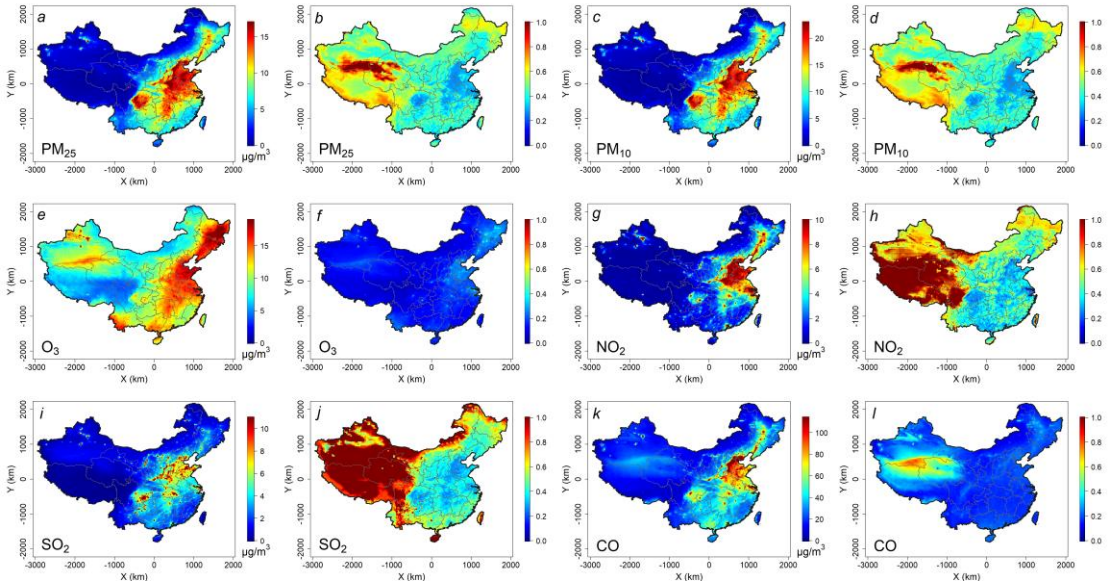

Figure 4: Spatial distribution of mean absolute error (panels a, c, e, g, i, and k) and normalized mean absolute error for the six criteria pollutants (panels b, d, f, h, j and l) of FastCTM comparing to CMAQ in 2023.

Section 3.1.2. Again, I don't see why this comparison makes sense. You do not incorporate any station data, so why would you make comparisons against it? The WRF-CMAQ model is the ground truth here.

Response: We agree with the reviewer's point that related comparisons between FastCTM to station observations are removed from the manuscript.

Sections 3.2: These don't have much meaning if we do not understand how the FastCTM model behaves in relation to the parent model

Response: We agree with the reviewer's comment that more analysis are needed to verify FastCTM's capabilities to capture the impact of changes in emissions, especially compared to the parent model of WRF-CMAQ. We added a comparison between FastCTM and CMAQ under 11 emission scenarios in the winter month of January 2019 and in the summer month of July 2019. The results signified that the FastCTM simulations manifested a good agreement with those of CMAQ, which was manifested in two principal aspects. Firstly, the FastCTM model forecasted positive responses to augmented emissions in the no-control (NCtrl) scenario and negative responses in the other emission-controlled scenarios just like CMAQ. This implies that when emissions were unrestricted and increased, as in the

NCtrl scenario, FastCTM could capture the increasing trend as that of CMAQ. In scenarios where emissions were reduced, they both predicted a decline. Secondly, in scenarios characterized by more substantial emission reductions, the FastCTM model simulated a more pronounced decrease in air pollutant concentrations. This is of particular significance as it shows the model's sensitivity to the magnitude of emission inputs. It suggests that the FastCTM model is not only capable of discerning changes in emission scenarios but can also reflect the degree of impact on air quality, thereby reinforcing its reliability and utility in simulating air quality dynamics in agreement with CMAQ. Related results in the manuscript are shown as follows.

The sensitivities of FastCTM simulations to emission interventions were contrasted with those of CMAQ. Specifically, CMAQ was employed to simulate 11 emission scenarios over the two-month periods of January and July 2019 in Southwest China (Huang et al., 2022). The alterations in emissions relative to the base case are presented in Table 1. Among these scenarios, 10 involved reduced emissions of major species, with only the no-control scenario exhibiting increased emissions. Utilizing the identical emissions and meteorological data, FastCTM also conducted simulations, which were then compared to those of CMAQ. For the 11 scenarios in question, the changes in air pollutant concentrations relative to the base case at the locations of 139 national air quality monitoring stations (Figure S14 in the SI) were extracted and compared in the winter month of January 2019 (Figure 9) and in summer month of July 2019 (Figure 10). The results indicated that, overall, the FastCTM simulations were in good agreement with those of CMAQ reflected in two aspects. First, FastCTM predicted positive responses to increased emissions in the nocontrol (NCtrl) scenario and negative responses to other emission-controlled scenarios just as CMAQ. Second, FastCTM simulated larger air pollutant concentration decrease in those scenarios with higher emission reductions. Specifically, in January 2019, with the exception of NO_2, FastCTM responded to emission changes with an interquartile range (IQR, 25% - 75% percentile) similar to that of CMAQ (Figure 9). For NO_2, in the same emission reduction scenarios, FastCTM simulated lower NO_2 values. In the summer month of July 2019, as depicted in Figure 10, all the criteria pollutants except CO demonstrated a comparable degree of response to emission reductions. The comparison suggests that the FastCTM model is not only capable of discerning changes in emission scenarios but can also reflect the degree of impact on air quality, thereby reinforcing its reliability and utility in simulating air quality dynamics in tandem with CMAQ. It should be noted that in both months, FastCTM exhibited slightly larger median values, suggesting its greater sensitivity to emission interventions.

***Table 1**. The emission change details of emission scenarios*

Scenario	***abbreviation***	***Sector***	***NO_x***	***VOCs***	***SO_2***	***CO***	***$PM_{2.5}$***	***PMC***
nocontrol	***NCtrl***	*Industrial*	*30%*	*30%*	*30%*	*30%*	*30%*	*30%*
		Traffic	*20%*	*20%*	*20%*	*20%*	*20%*	*20%*
medianX	***MedX***	*Industrial*	*-36%*	*-35%*	*-48%*	*-23%*	*-9%*	*-9%*
		Traffic	*-40%*	*-10%*	*0*	*-26%*	*-10%*	*-10%*
medianY	***MedY***	*Industrial*	*-26%*	*-20%*	*-38%*	*-13%*	*-4%*	*-4%*

		Traffic	*-30%*	*0%*	*0*	*-16%*	*-5%*	*-5%*
medianZ	***MedZ***	*Industrial*	*-36%*	*-10%*	*-48%*	*-23%*	*-9%*	*-9%*
		Traffic	*-40%*	*0%*	*0*	*-26%*	*-10%*	*-10%*
median-3	***Med-3***	*Industrial*	*-10%*	*-10%*	*-18%*	*0*	*0*	*0*
		Traffic	*-10%*	*0%*	*0*	*0*	*0*	*0*
median-2	***Med-2***	*Industrial*	*-16%*	*-20%*	*-28%*	*-3%*	*0*	*0*
		Traffic	*-20%*	*0%*	*0*	*-6%*	*0*	*0*
median-1	***Med-1***	*Industrial*	*-26%*	*-35%*	*-38%*	*-13%*	*-4%*	*-4%*
		Traffic	*-30%*	*-10%*	*0*	*-16%*	*-5%*	*-5%*
median0	***Med0***	*Industrial*	*-36%*	*-50%*	*-48%*	*-23%*	*-9%*	*-9%*
		Traffic	*-40%*	*-20%*	*0*	*-26%*	*-10%*	*-10%*
median+1	***Med+1***	*Industrial*	*-46%*	*-65%*	*-58%*	*-33%*	*-19%*	*-19%*
		Traffic	*-50%*	*-30%*	*0*	*-36%*	*-20%*	*-20%*
median2030	***Med30***	*Industrial*	*-55%*	*-70%*	*-80%*	*-40%*	*-40%*	*-40%*
		Traffic	*-60%*	*-40%*	*0*	*-40%*	*-40%*	*-40%*
median2035	***Med35***	*Industrial*	*-80%*	*-80%*	*-90%*	*-60%*	*-50%*	*-50%*
		Traffic	*-80%*	*-60%*	*0*	*-60%*	*-50%*	*-50%*

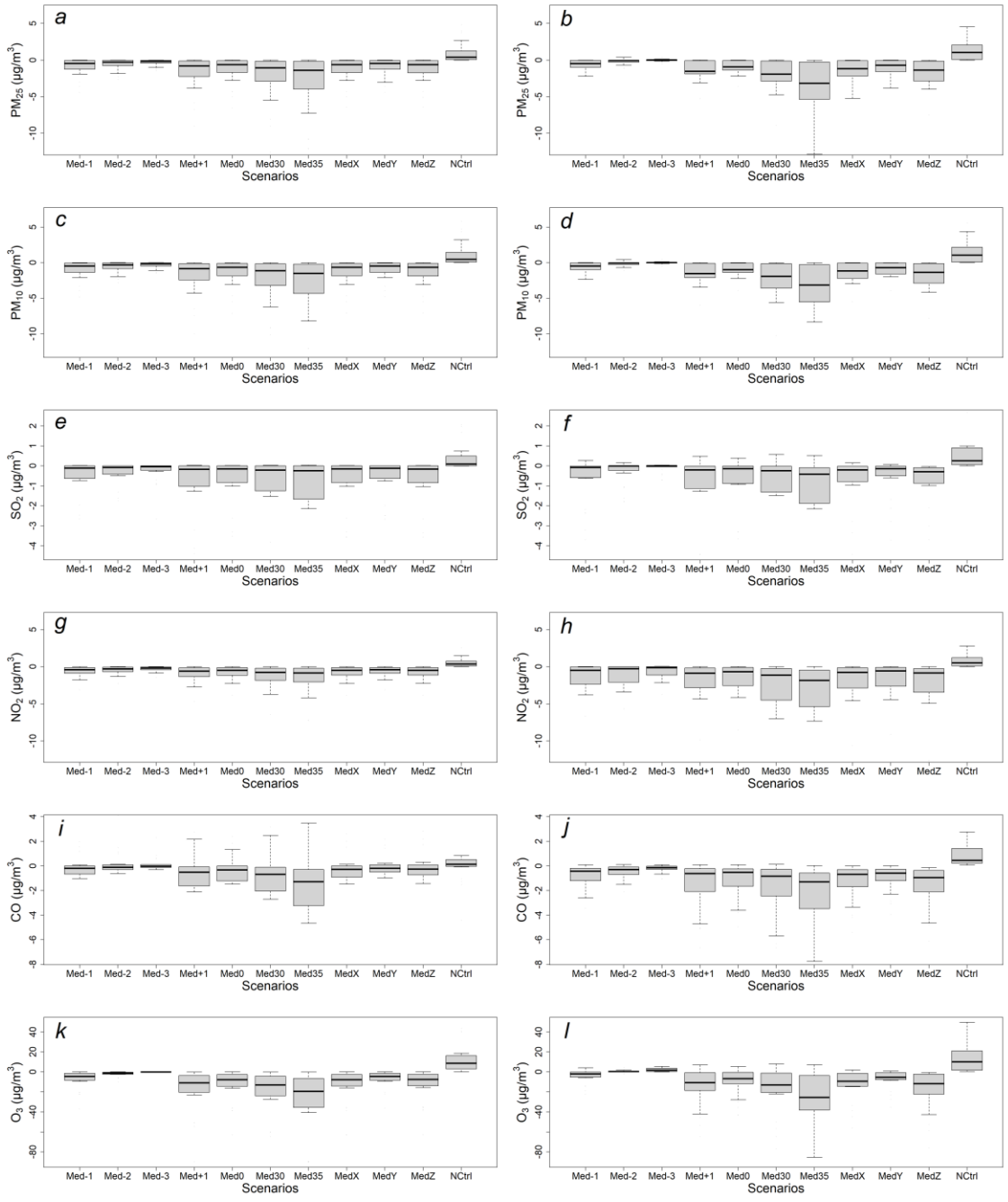

Figure 9: Air pollutant concentration changes in terms of base case simulated by CMAQ (subplots of a, c, e, g, i and k in first column) and by FastCTM (subplots of b, d, f, h, j and l in second column) in January 2019.

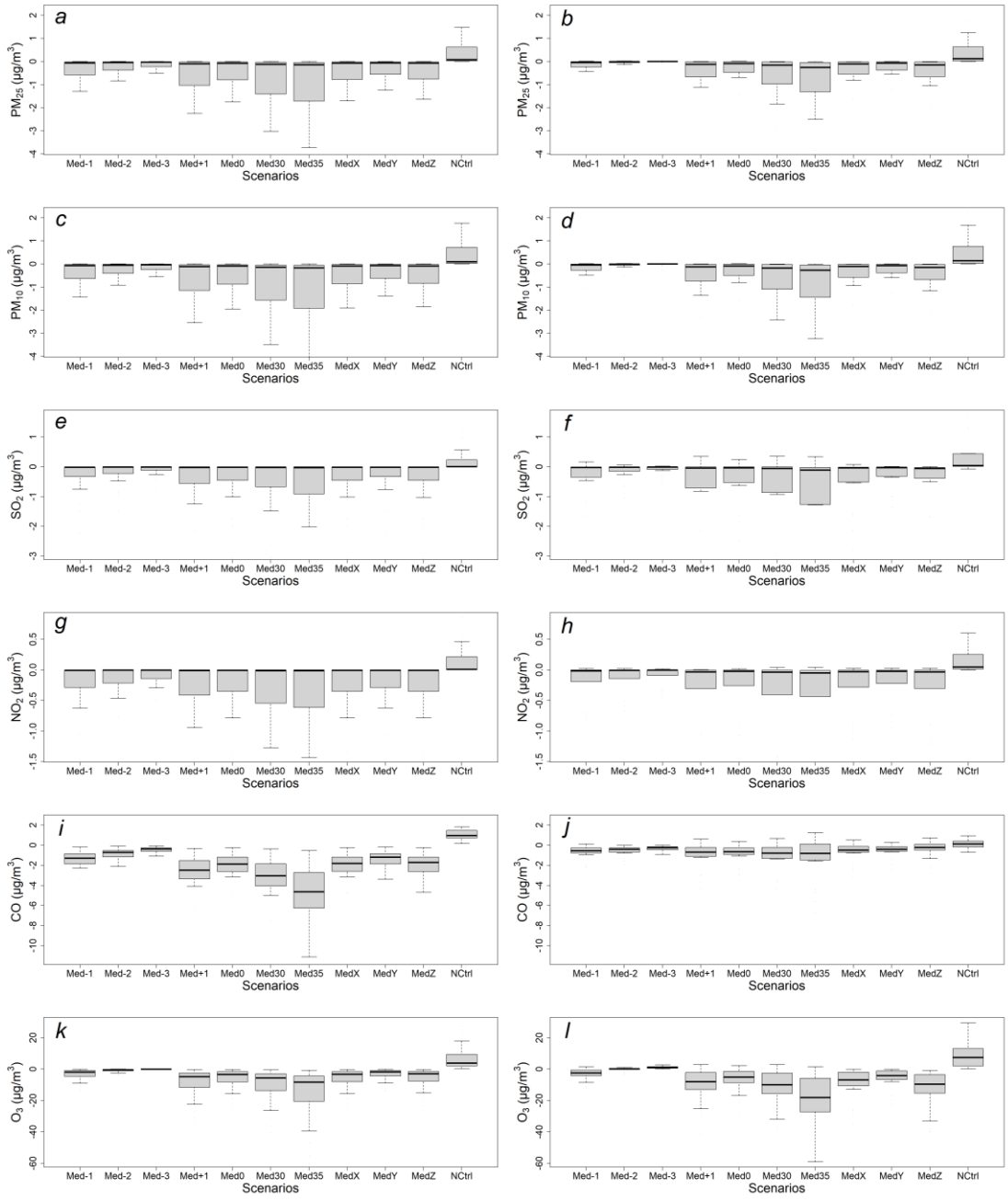

Figure 10: Air pollutant concentration changes in terms of base case simulated by CMAQ (subplots of a, c, e, g, i, and k in the first column) and by FastCTM (subplots of b, d, f, h, j, and l in the second column) in July 2019.

Figure 8: These color bars are difficult to discern changes in concentrations. Does adding d through h yield panels a or b? Again, individual contribution doesn't matter if we don't know how the model actually behaves.

Response: Adding panel d through h yield panel c. To validate the process analysis by FastCTM, its simulation results are compared to those by CMAQ. We added related results and discussion in Section 3.3 as follows,

In this study, we further selected the data recorded at 23:00 on October 13, 2024, to compare the

impacts of the five major atmospheric physical and chemical processes, as simulated by FastCTM and CMAQ, on $PM_{2.5}$ concentration changes (Figure 12). Emissions, advection processes, and diffusion processes demonstrated a relatively high degree of consistency between the two models. Regarding the simulation of chemical reactions, while the spatial distribution of high-value areas in the FastCTM results was comparable to that of CMAQ, the simulated values in FastCTM were notably higher. Correspondingly, FastCTM overestimated the contribution of the deposition process. This overestimation counterbalanced the impact of the higher chemical reaction values. The difference in the simulated deposition contributions between the two models could be due to differences in how they represent these influencing factors. Overall, the simulation results of the process contributions by FastCTM and its parent model CMAQ were relatively consistent. This consistency indicates that, despite some differences in the magnitude of certain process simulations, FastCTM is capable of capturing the essential features of atmospheric processes related to $PM_{2.5}$ concentration changes, similar to CMAQ. Such consistency provides confidence in the reliability of FastCTM for simulating and understanding the complex interplay of atmospheric processes of $PM_{2.5}$.

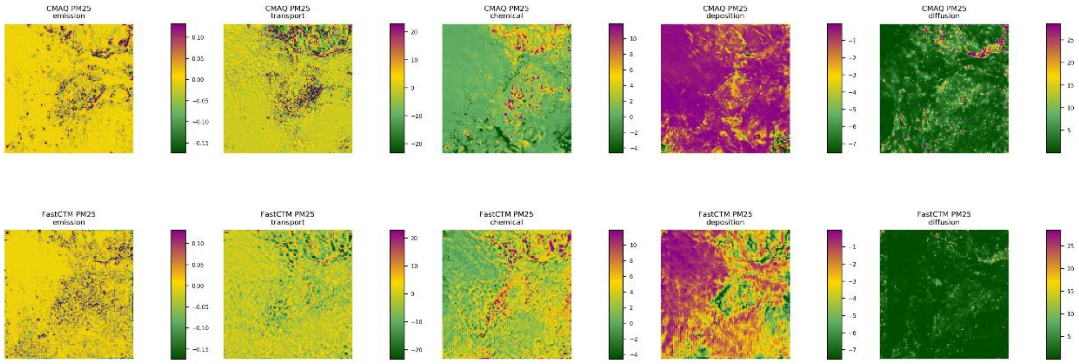

Figure 5: An example of contributions from five major atmospheric processes to $PM_{2.5}$ changes ($\mu g/m^3$) by CMAQ (first row) and FastCTM (second row) at 23:00 on October 13, 2024.