

Point-to-Point Responses to Reviewer's Comments

Comments from Reviewer 1

The authors describe a new, neural network-based reduced-order model of atmospheric chemistry and transport (FastCTM) which has been trained using an extensive dataset of output from CMAQ. FastCTM uses a novel and interesting approach, building physics-informed networks for five separate operators. The authors show that FastCTM is able to reproduce the general patterns of concentrations calculated by CMAQ for an out-of-training-data year (2023), and that the sensitivities of FastCTM to key meteorological variables or nation-wide changes in emissions mostly follow expected patterns. If FastCTM can be shown to be reliable in policy-relevant contexts then it could be a very useful tool.

This approach to modelling is interesting, and this methodological advance has the potential to significantly accelerate air quality scenario analysis. A CTM which can respect key physical constraints (e.g. mass conservation) while also accurately reproducing the effect of different perturbations to emissions and meteorological fields would have great value. However, the manuscript as written does not quite live up to this promise. Along with some minor concerns, the key challenge is that the authors do not show evidence that this new model can fulfil the roles of a CTM and produce accurate results for one of the most common use cases (i.e. understanding the effects of different perturbations). I explain this concern in more detail below, and until this concern is addressed I do not believe the manuscript should be accepted for publication by GMD.

Major comments

The most significant concern relates to the validation/evaluation of the model. The authors appear to have trained the five physical operators based on several years of output from the CMAQ chemistry transport model. While I have some questions regarding the training process, I will take it as read for the moment that the training was done in such a way as to avoid overfitting. However, the verification of the model rests on its ability to predict, from the 2018-2022 data, the performance in 2023. This approach is inadequate for two reasons. First, the authors do not compare the performance of the model to simpler approaches with the same data such as generalized additive models, gradient boosting, or linear regression with land use (see e.g. Wong et al., 2021 and Cheng et al., 2021). Without such a comparison to evaluate how such models would have performed in predicting 2023, it is difficult to say what the magnitude of FastCTM's advance is. This is exacerbated by the relatively shallow quantitative assessment in section 3.1.1. RMSE and R² values are provided, but it is not clear how these were calculated; given that these are calculated as a function of time, are these calculated based on the difference in each of the 158,742 grid cells between CMAQ and FastCTM? A deeper analysis which investigates how model

performance varies between (e.g.) rural and urban areas, coastal and inland areas, winter and summer, and so on would provide a much more robust test of the model's ability to predict the effect of changing meteorology. This could be informed by (e.g.) taking the difference of FastCTM for 2023 against CMAQ for 2023, and comparing that to the difference between CMAQ for 2023 and the average of CMAQ from 2018 to 2022. This would at least demonstrate whether FastCTM provides more explanatory power for the mean atmospheric state than taking the average concentration from the previous five years.

Response: Thanks. We agree with review's point that current analysis is not adequately to demonstrate that FastCTM has actual capabilities to simulate air quality changes by learning and representing physical and chemical processes. As reviewer suggested, more tests, comparisons and analysis are performed.

(1) For comparing with simpler machine learning models, we tested three models of Linear Regression, Random Forest and XGBoost with the same train and test dataset. The results are added in the manuscript as follows,

In addition, 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 gbtrees). These LUR models were developed using the same input meteorological, emission, and geophysical covariates. When compared with the FastCTM model, the performance of the LUR models was found to be significantly inferior (as demonstrated in Figure S6 - S8 in the Supplementary Information). 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.

With the supplementary Figures demonstrating performances of three machine learning models displaying in the following part.

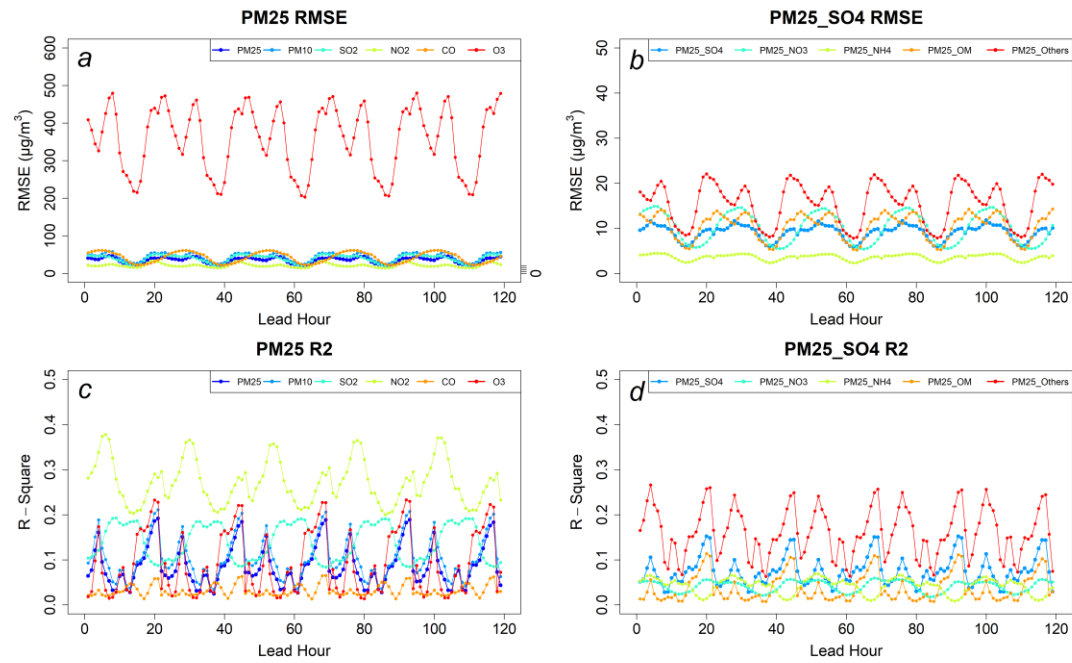


Figure S1: The evaluation performances of linear regression forecasts against CMAQ forecasts in 2023. Panel (a) and (b) respectively show RMSE values of criteria pollutants and the PM_{2.5} components of. Panel (c) and (d) respectively show R^2 values. It should be noted that RMSE value of CO corresponds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

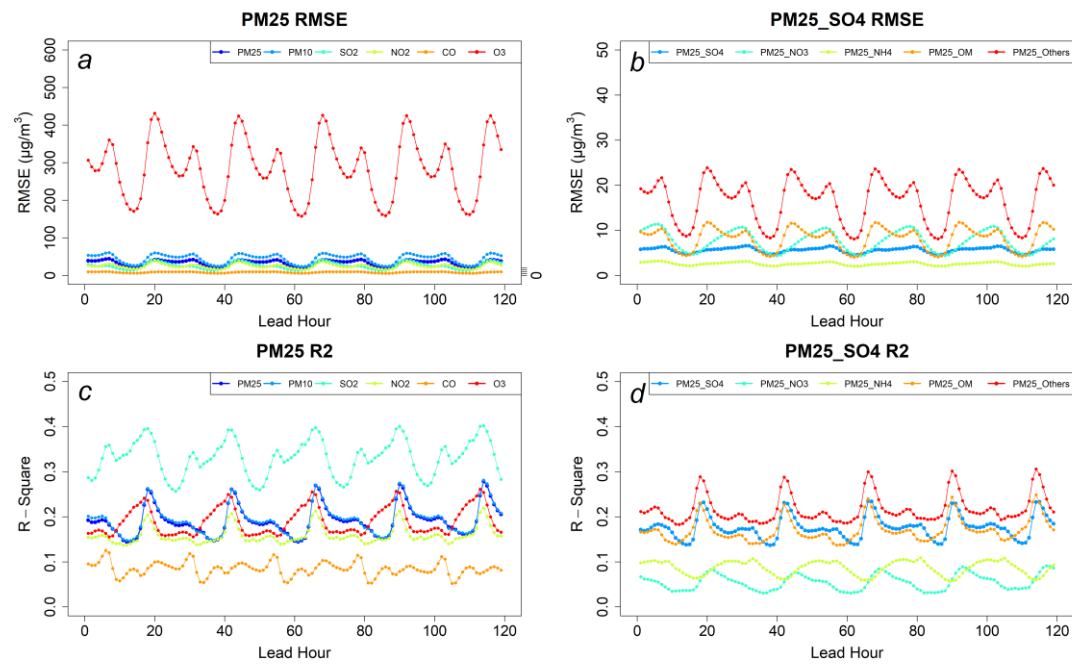


Figure S2: The evaluation performances of random forest forecasts against CMAQ forecasts in 2023. Panel (a) and (b) respectively show RMSE values of criteria pollutants and the PM_{2.5} components of. Panel (c) and (d) respectively show R^2 values. It should be noted that RMSE value of CO corresponds to the right axis in panel (a).

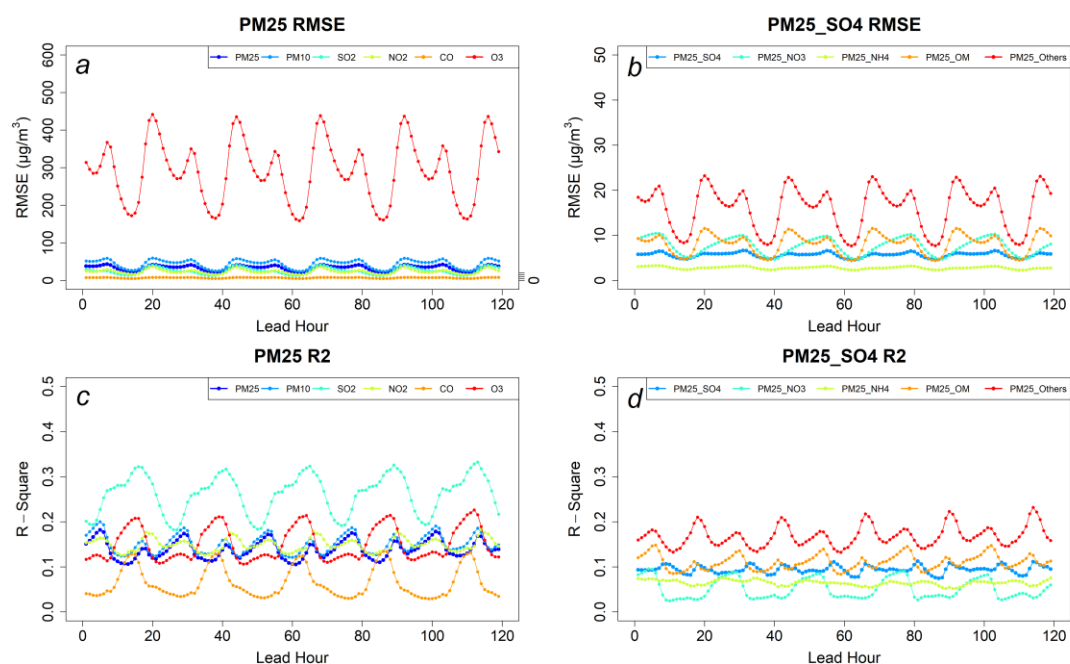


Figure S3: The evaluation performances of XGboost forecasts against CMAQ forecasts in 2023. Panel (a) and (b) respectively show RMSE values of criteria pollutants and the PM_{2.5} components of. Panel (c) and (d) respectively show R² values. It should be noted that RMSE value of CO corresponds to the right axis in panel (a).

(2) As for the calculation process for the metrics of RMSE and R², they are elaborated in the section 2.5 Model Evaluation, as follows.

The metrics of root mean square error (RMSE) and coefficient of determination (R²) were calculated daily in each of 119 leading hours on the difference in each of the 158,742 grid cells between CMAQ and FastCTM. Therefore, 119 static values for each metric of R² and RMSE were obtained on each day of the independent test year of 2023. The statistic values on each day are then averaged for the same leading hour for comparison.

(3) As reviewer suggested, evaluations of FastCTM comparing to CMAQ in rural/urban, inland/coastal areas, and cold/warm seasons are further performed. Generally, they have similar performances in comparative areas or seasons. But FastCTM exhibited lower correlations in urban areas and coastal areas. In urban area, emission sources and chemical processes are more complex than that in rural area, making it harder for FastCTM to simulate due to its 2D setting and fewer chemicals considered than CMAQ. It is also same true for FastCTM's performance in coastal area, where meteorological conditions are more varied in time. Related discussions and results are added in the section 3.1.1. and in the supplementary material, which as follows.

Defining the warm season as the months from April to September and the winter and cold season as the remaining months, the FastCTM model exhibited comparable performances. As shown in Figure 3 (with detailed information in Figure S9 in the SI), the coefficient of determination R² values for the six criteria pollutants were 0.82, 0.8, 0.8, 0.82, 0.91, and 0.7 in the warm season, and 0.8, 0.79, 0.78, 0.83, 0.88, and 0.68 in

the cold season, respectively. To assess the performance variations of FastCTM across different spatial locations, comparative evaluations were carried out in urban and rural areas as well as in inland and coastal regions. Generally, FastCTM demonstrated slightly higher accuracies in rural areas compared to urban areas (as presented in Figure S10 in the SI). This outcome is reasonable given the more intricate emission and chemical processes prevalent in urban settings (Guo et al., 2014). Similarly, FastCTM exhibited comparable performances in inland areas to those in coastal areas, with the exception of $PM_{2.5}$ and PM_{10} (Figure S11 in the SI).

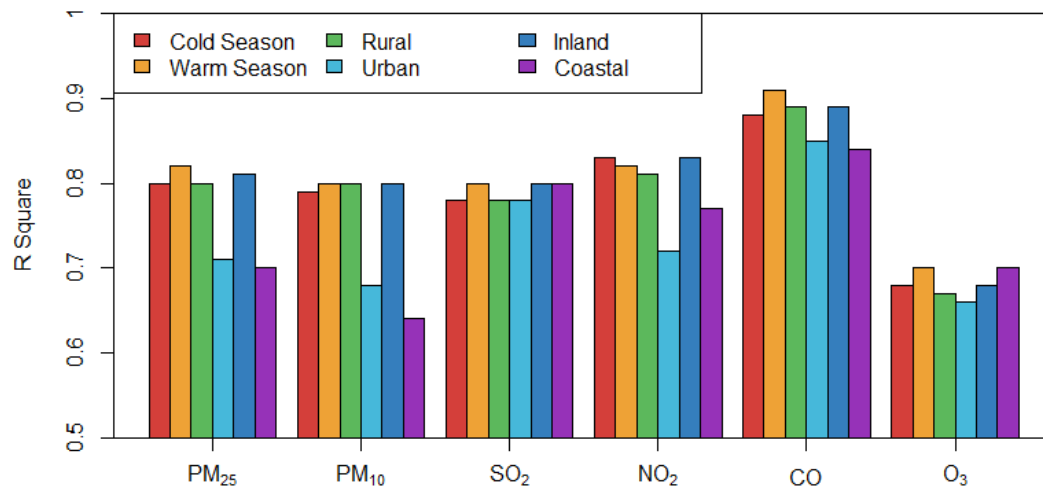


Figure 4: The evaluation R^2 values of FastCTM forecasts in warm/cold seasons, rural/urban areas and coastal/inland areas.

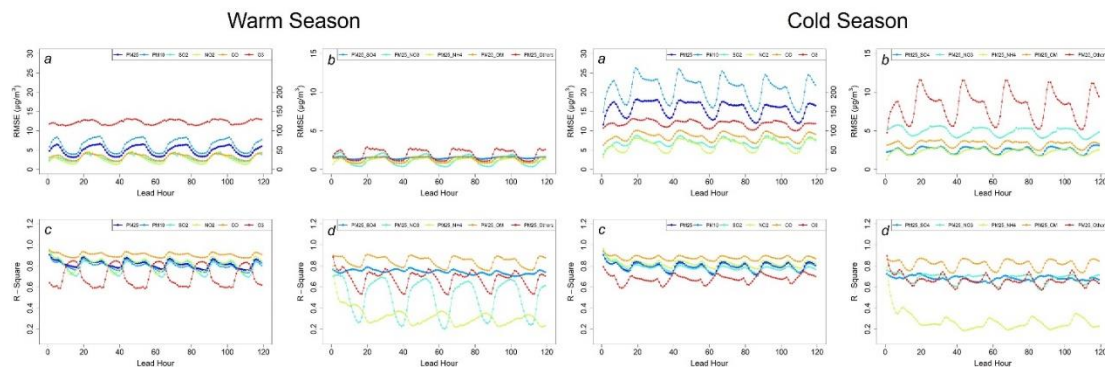


Figure S5: The evaluation performances of random forest forecasts against CMAQ forecasts in warm season of 2023. Panel (a) and (b) respectively show RMSE values of criteria pollutants and the $PM_{2.5}$ components of. Panel (c) and (d) respectively show R^2 values. It should be noted

that RMSE value of CO corresponds to the right axis in panel (a).

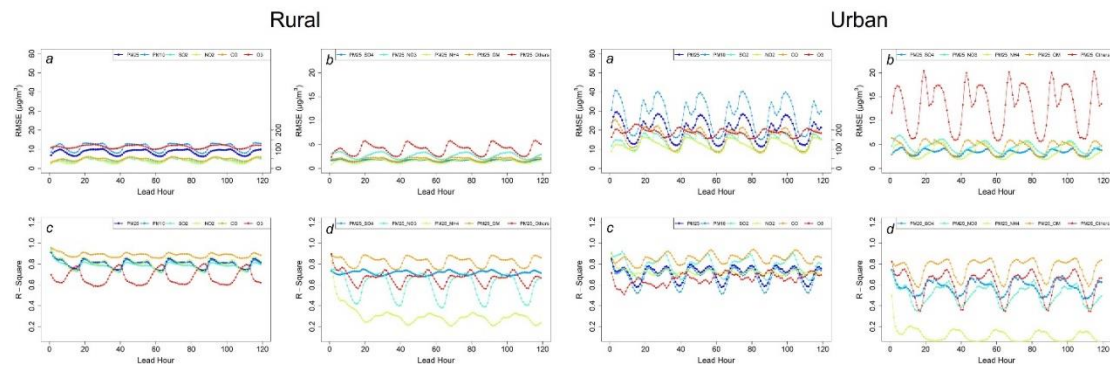


Figure S6: The evaluation performances of XGBoost forecasts against CMAQ forecasts in rural and urban areas in 2023. Panel (a) and (b) respectively show RMSE values of criteria pollutants and the PM_{2.5} components of. Panel (c) and (d) respectively show R² values. It should be noted that RMSE value of CO corresponds to the right axis in panel (a).

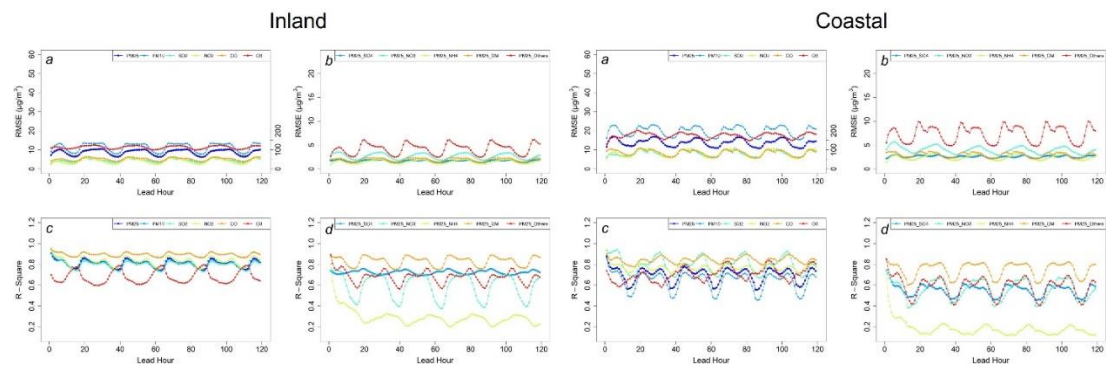


Figure S7: The evaluation performances of FastCTM forecasts against CMAQ forecasts in inland and coastal areas in 2023. Panel (a) and (b) respectively show RMSE values of criteria pollutants and the PM_{2.5} components of. Panel (c) and (d) respectively show R² values. It should be noted that RMSE value of CO corresponds to the right axis in panel (a).

(4) 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 generally in good correlation with CMAQ forecasts in 2023, rather than that in 2018-2023. It means FastCTM has learn 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 the 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 4, 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.

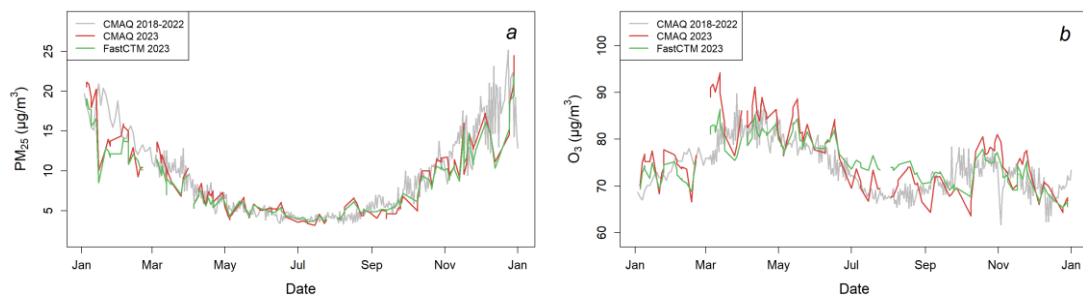


Figure 5: The evaluation performances of FastCTM forecasts against CMAQ forecasts in 2023.

Second, and perhaps more importantly, the function of FastCTM is to reproduce the results of high-fidelity CTMs at a fraction of the computational cost – specifically to support air pollution simulations, sensitivity analysis, and internal process analysis (abstract lines 30-32). The comparison to 2023 only tells us that FastCTM can reproduce the general pattern of air pollution in 2023, but does not tell us whether FastCTM will accurately predict the effect of interventions. The sensitivity tests in section 3.2 have no basis for comparison, and are in any case so broad (representing nationwide changes in temperature, PBL height, or emissions) that they are a limited test of the CTM’s capabilities. At the very minimum, an evaluation is needed which shows that FastCTM’s trends actually match the underlying trends in WRF-CMAQ; this should be straightforward for the emissions cases. Since the goal of FastCTM is to reduce computational costs, it is critical that FastCTM be shown to be faithful to its parent model for realistic applications such as projecting the impact of a change in emissions. Going further and comparing sensitivities for local or single-sector emissions changes would provide even more powerful proof, and I strongly recommend that the authors consider such a comparison.

Without these kinds of quantitative comparisons I can only judge the model’s success based on data such as Figure 3, where I am concerned because the patterns do not – speaking qualitatively – appear to match that well between CMAQ and FastCTM. I am particularly concerned that the model may be mostly reproducing emissions maps and

historical scalings, rather than accurately representing chemistry and transport (especially given that transport is 2-D only). A more critical, quantitative analysis of the models strengths and weaknesses would be necessary before I would recommend its use in a scientific or regulatory context.

Response: We agree with reviewer's comment that more analysis needed to verify FastCTM's capabilities to project the impact of a change in emissions. Since the emissions are same for each year from 2018-2023, it is not possible to test FastCTM's trends to that of WRF-CMAQ. Instead, we added comparison between FastCTM and CMAQ in 11 emission intervention scenarios in the winter month of January 2019 and in summer month of July 2019. The results signified that, on the whole, the FastCTM simulations manifested a high level of concordance with those of CMAQ, which was manifested in two principal aspects. Firstly, akin to CMAQ, the FastCTM model forecasted positive responses to augmented emissions in the no-control (NCtrl) scenario and negative responses in the other emission-controlled scenarios. This implies that when emissions were unrestricted and increased, as in the NCtrl scenario, both models detected a corresponding upward trend in pollutant levels. Conversely, in scenarios where emissions were curbed, 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 showcases the model's sensitivity to the magnitude of emission interventions. 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 tandem with CMAQ. Related results in the manuscript were 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. 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 NO₂, FastCTM responded to emission changes with an interquartile range (IQR, 25% - 75% percentile) similar to that of CMAQ (Figure 9). For NO₂, in the same emission reduction scenarios, FastCTM

simulated lower NO₂ 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

<i>Scenario</i>	<i>abbreviation</i>	<i>Sector</i>	<i>NO_x</i>	<i>VOC_s</i>	<i>SO₂</i>	<i>CO</i>	<i>PM_{2.5}</i>	<i>PMC</i>
<i>nocontrol</i>	<i>NCtrl</i>	<i>Industrial</i>	30%	30%	30%	30%	30%	30%
		<i>Traffic</i>	20%	20%	20%	20%	20%	20%
<i>medianX</i>	<i>MedX</i>	<i>Industrial</i>	-36%	-35%	-48%	-23%	-9%	-9%
		<i>Traffic</i>	-40%	-10%	0	-26%	-10%	-10%
<i>medianY</i>	<i>MedY</i>	<i>Industrial</i>	-26%	-20%	-38%	-13%	-4%	-4%
		<i>Traffic</i>	-30%	0%	0	-16%	-5%	-5%
<i>medianZ</i>	<i>MedZ</i>	<i>Industrial</i>	-36%	-10%	-48%	-23%	-9%	-9%
		<i>Traffic</i>	-40%	0%	0	-26%	-10%	-10%
<i>median-3</i>	<i>Med-3</i>	<i>Industrial</i>	-10%	-10%	-18%	0	0	0
		<i>Traffic</i>	-10%	0%	0	0	0	0
<i>median-2</i>	<i>Med-2</i>	<i>Industrial</i>	-16%	-20%	-28%	-3%	0	0
		<i>Traffic</i>	-20%	0%	0	-6%	0	0
<i>median-1</i>	<i>Med-1</i>	<i>Industrial</i>	-26%	-35%	-38%	-13%	-4%	-4%
		<i>Traffic</i>	-30%	-10%	0	-16%	-5%	-5%
<i>median0</i>	<i>Med0</i>	<i>Industrial</i>	-36%	-50%	-48%	-23%	-9%	-9%
		<i>Traffic</i>	-40%	-20%	0	-26%	-10%	-10%
<i>median+1</i>	<i>Med+1</i>	<i>Industrial</i>	-46%	-65%	-58%	-33%	-19%	-19%
		<i>Traffic</i>	-50%	-30%	0	-36%	-20%	-20%
<i>median2030</i>	<i>Med30</i>	<i>Industrial</i>	-55%	-70%	-80%	-40%	-40%	-40%
		<i>Traffic</i>	-60%	-40%	0	-40%	-40%	-40%

<i>median20</i>		<i>Industrial</i>	-80%	-80%	-90%	-60%	-50%	-50%
<i>35</i>	<i>Med35</i>	<i>Traffic</i>	-80%	-60%	0	-60%	-50%	-50%

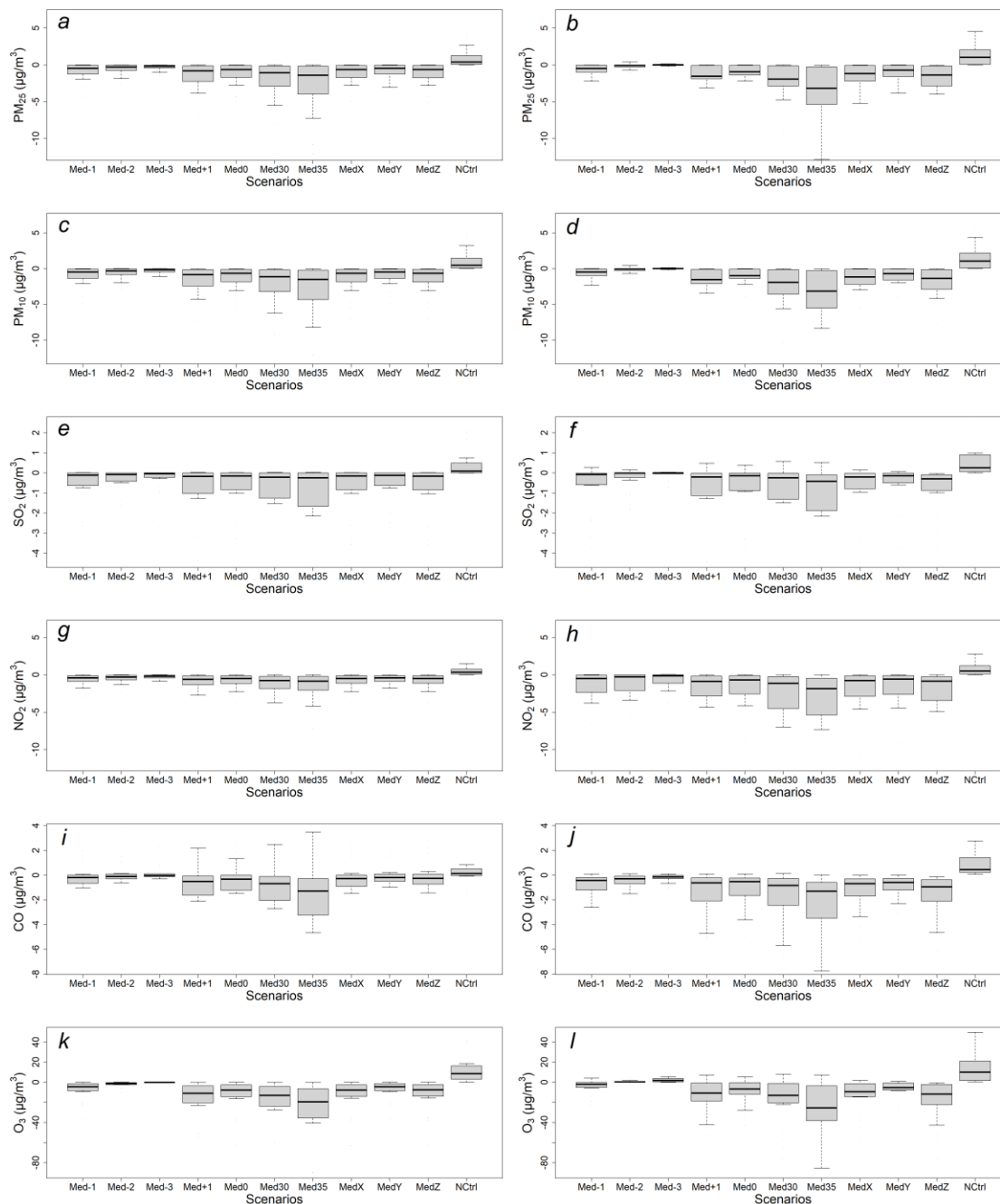


Figure 9: Air pollutant concentration changes in terms of base case simulated by CMAQ (subplots of a, c, e, g, i and k in first column) and by FastCTM (subplots of b, d, f, h, j and l in second column) in January 2019.

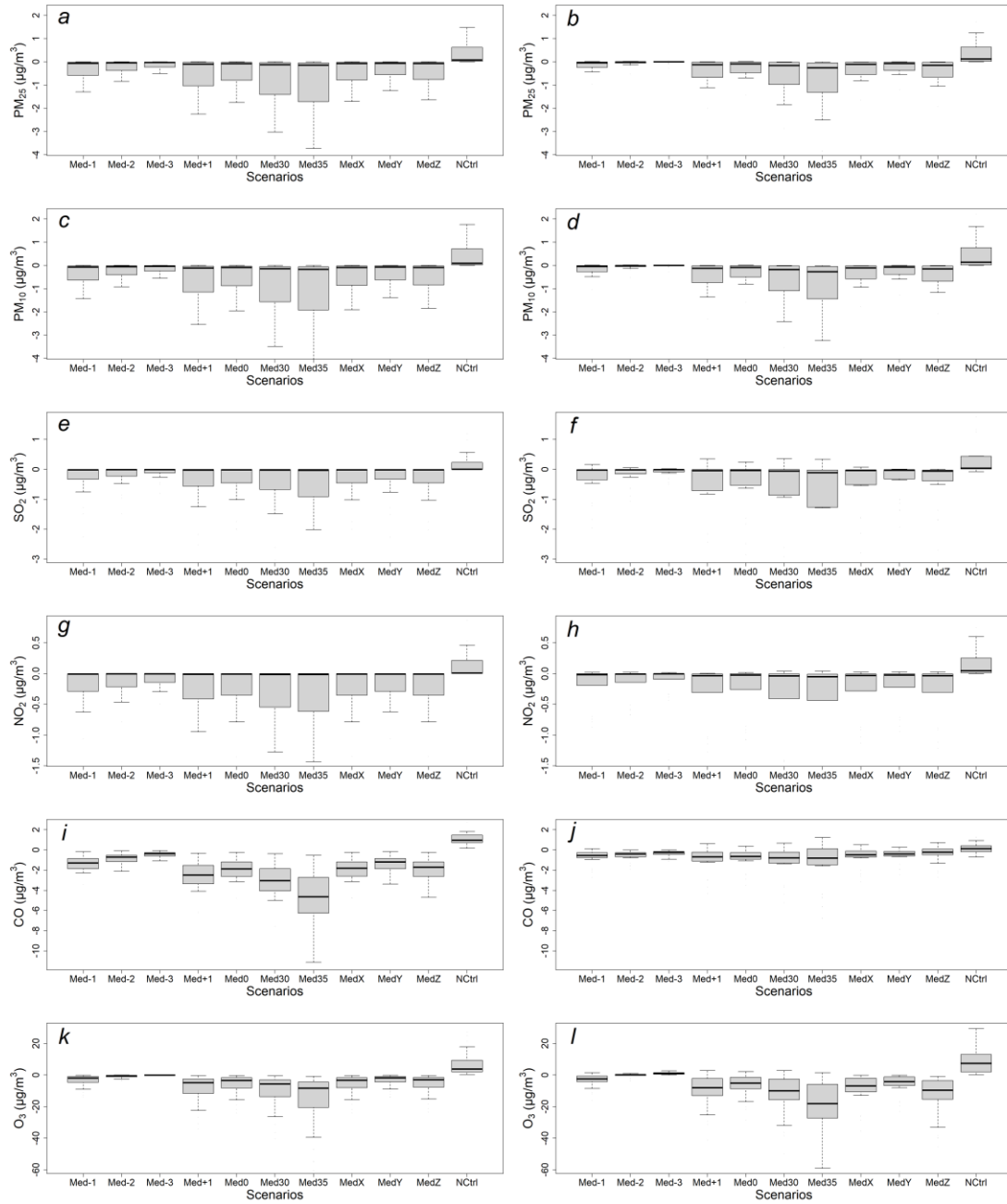


Figure 10: Air pollutant concentration changes in terms of base case simulated by CMAQ (subplots of a, c, e, g, i and k in first column) and by FastCTM (subplots of b, d, f, h, j and l in second column) in July 2019.

Minor comments

The description of the five operators does not quite seem to verify that physical constraints are being satisfied but this may simply be a misinterpretation on my part. For example, can you confirm that the method you used to generate the convolution kernels (Eq 5-7) for transport ensures mass conservation? This seems to depend on C_i being in units of molec/cm³ rather than ppbv, but the units of C_i are not clearly specified.

Response: We added description for the model framework in the Section 2.3. The unit for all pollutants is $\mu\text{g}/\text{m}^3$. We used a upwind-scheme to simulate diffusion and advection processes. For the scheme, masses are conserved. However, FastCTM as a whole is not mass conserved, because it also consists other neural network modules such as reaction and deposition. These deep learning modules are learned to minimize the loss function of mean squared error. Revised model description and figures are shown as follows.

Instead, we use 1-hour initial pollutant concentration ($J=1$) to simulate 24-hour air quality pollutants ($K=24$), to ensure FastCTM is dedicated to learn air quality changes between neighboring two hours as shown in Figure 1a. In other words, at time $t = 0$, FastCTM predicted K -hour air pollutant concentrations of $C_{t=0}, C_{t=1}, \dots, C_{t=K-1}$, given the input air pollutant concentration in previous hour $C_{t=-1}$ and corresponding meteorological data and emissions at time $t = 0, 1, \dots, K-1$. The unit of concentrations is $\mu\text{g}/\text{m}^3$ for all pollutants.

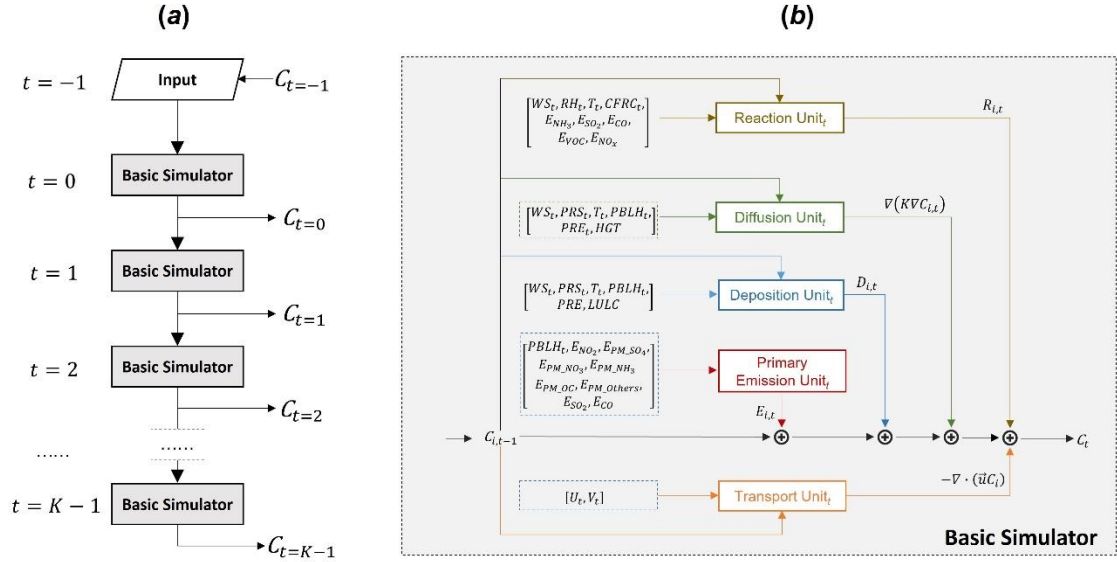


Figure 8: (a) General model workflow, and (b) the basic simulator module structure at the time step t of deep learning simulation model FastCTM designed according to Eq.1. Arrows and boxes with different colours represent calculation modules of different atmospheric physical and chemical processes.

A related concern is that surface layer winds are used and treated conservatively, which neglects the fact of rapid vertical mixing. Can the authors provide evidence that the surface winds (which would be expected to be slower than the mean wind speed in the boundary layer) are accurately predicting pollutant motion? It seems that any model which is designed to predict transport using only the horizontal near-surface winds will underestimate overall transport. Should the model not be using the PBL-averaged horizontal winds in Eq. 4 instead of the surface winds at 10-meter height (lines 83-84)?

Response: We appreciate the reviewer's critical observation regarding the use of

surface-layer winds (10 m) in FastCTM's transport module and the potential underestimation of pollutant transport. Regarding the suggestion of using PBL - averaged horizontal winds, we understand the potential benefits of such an approach. However, implementing PBL - averaged winds would require additional data processing and model complexity. In our case, we aimed to strike a balance between model accuracy and computational efficiency. The use of 10 - meter surface winds allows for a more straightforward implementation, which is beneficial for the model's scalability and real - time application. Moreover, the evaluation of the FastCTM against CMAQ simulations and observational data, as presented in our manuscript, indicates that the model performs reasonably well in predicting pollutant concentrations. The model achieved high agreements with CMAQ in terms of RMSE and R^2 values for multiple pollutants, suggesting that despite using surface winds, it is able to capture the essential aspects of pollutant transport and transformation. Besides, in Section 3.2.1, responses to increased wind speed are evaluated, it would be reasonable to assume decreased air pollutant concentration predicted by FastCTM using larger PBL-averaged wind speed. For example, with a 10% increase of wind speed would lead to about 2% decrease of air pollutant concentrations except for O_3 . Therefore, it would make much difference with PBL-averaged wind speed. Future revision of FastCTM in 3D dimensions would consider the vertically varied wind fields to better simulate wind-related processes.

It would be helpful to get more detail on how components such as the diffusion encoder were trained. Currently the manuscript states that 5 years of data (2018 – 2022 inclusive) were used in training, but not how the five different models were trained using that data. A naïve assessment would assume that all five sub-models were trained based simply on hour-to-hour pollutant concentrations, but that would suggest that the models were each trying to represent all atmospheric processes simultaneously.

Response: 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 diffusion process $\nabla(K\nabla C_i)$. It's also same for processes such as reaction, advection, and deposition. Therefore, it is not possible for one process to represent all atmospheric processes simultaneously. The independent contribution of each process depicted in Figure 12 of section 3.3. Clearly, each process exhibited their patterns of contribution to hourly air pollutant concentrations changes, constrained by the form of the operator in the processes. Related description was added in the 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 constraints 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.

Line 172 says that the reaction encoder in Equation 12 “has the same structure as that of reaction and deposition encoder models (Eq. 10)”. This is recursive, but also Eq. 10 refers to the diffusion module?

Response: This error was revised as follows in corresponding section,

Therefore, the reaction rate constant k is simulated using a spatial encoder function Encoder as shown in Eq. 12, which has the same structure as that of diffusion encoder modules (Eq. 10).

On line 194, “We did not use the fixed area as that in the previous studies (Xing et al., 2022)” – can you elaborate? It was not clear to me what this meant.

Response: Revised as follows,

The random sampling tactics would help model learn inherent physical and chemical principles model rather than just statistical spatiotemporal autocorrelations using data in constant spatial area (Xing et al., 2022). Besides, the spatio-temporal random samples contain varied emissions which would improve FastCTM adaption to changing emission levels.

The y-axis labels on Figure 5 say “Percentage”, but from context it appears these must really be the factor difference from the baseline (as all cross at 1.0).

Response: Revised.

Finally, there are numerous minor grammatical errors (e.g. L14: simulations and managements; L67: interpretations of the FastCTM are also widely vowed; L70: including and major; and so on). This is not important for judgment of the paper’s appropriateness for publication, but I recommend the authors take another look at the paper to correct such minor issues.

Response: We have read through the manuscript and made through revision on the writing of the manuscript.

Citations

Wong et al., “Using a land use regression model with machine learning to estimate

ground level PM_{2.5}." Environ. Pollut., 2021.

Cheng et al. "Influence of weather and air pollution on concentration change of PM_{2.5} using a generalized additive model and gradient boosting machine." Atmospheric environment, 2021.

Comments from Reviewer 2

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 and species 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 requires intense computational resources such as high-performance computing (Efsthathiou 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.

Efsthathiou, 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](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 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 the motivation, there are studies dedicated to develop model 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 ~8 time speed-up. Sturm and Wexler (2020) developed a mass- and energy-conserving framework for

using machine learning to speed computations with an 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 data 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 an 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: We are sorry that this is a written error. 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 PM_{2.5} (SO₄²⁻, 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 utility are critical. The 10 species were selected based on their direct relevance to regulatory standards (e.g., PM_{2.5}, PM₁₀, O₃, NO₂, SO₂, CO) and their dominance in driving health and environmental impacts in urban and industrial regions (e.g., China, where PM_{2.5} components like SO₄²⁻, NO₃⁻, and NH₄⁺ account for most part of 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 rigorously validated against both CMAQ simulations and ground observations (Sect. 3.1–3.2). Results show high agreement for all 10 species ($R^2 = 0.7\text{--}0.9$), 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 will clarify 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.

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 PM_{2.5}, 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 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, same to that 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.

In addition, 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 gbtrees). These LUR models were developed using the same input meteorological, emission, and geophysical covariates. When compared with the FastCTM model, the performance of the LUR models was found to be significantly inferior (as demonstrated in Figure S6 - S8 in the Supplementary Information). 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.

With the supplementary Figures exhibiting performances of three machine learning models as follows.

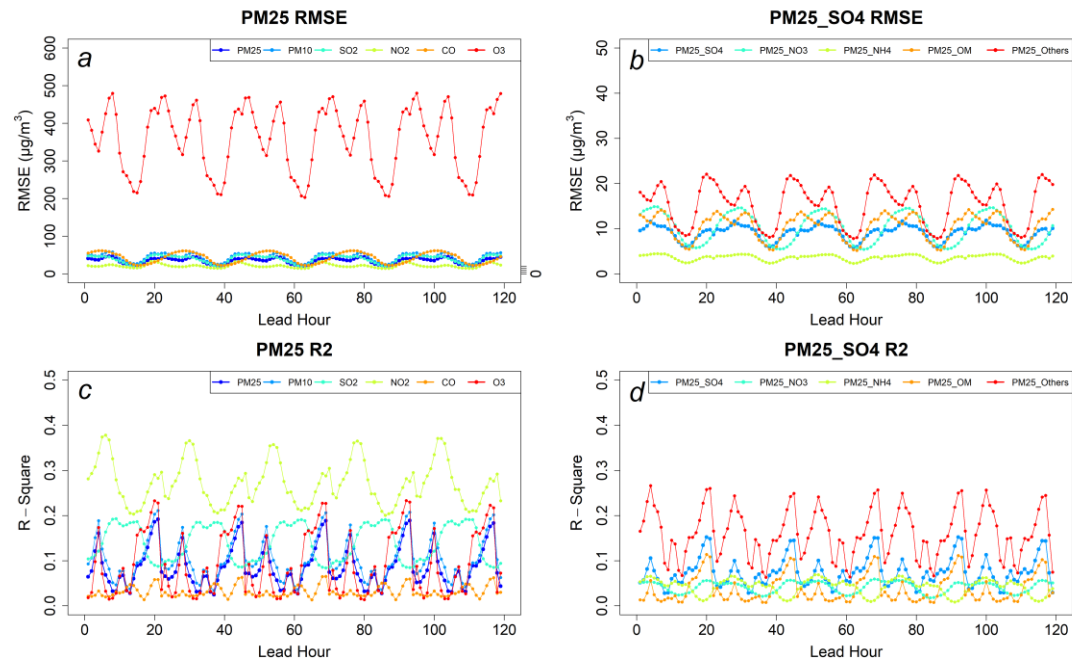


Figure S9: The evaluation performances of linear regression forecasts against CMAQ forecasts in 2023. Panel (a) and (b) respectively show RMSE values of criteria pollutants and the $\text{PM}_{2.5}$ components of. Panel (c) and (d) respectively show R^2 values. It should be noted that RMSE value of CO corresponds to the right axis in panel (a).

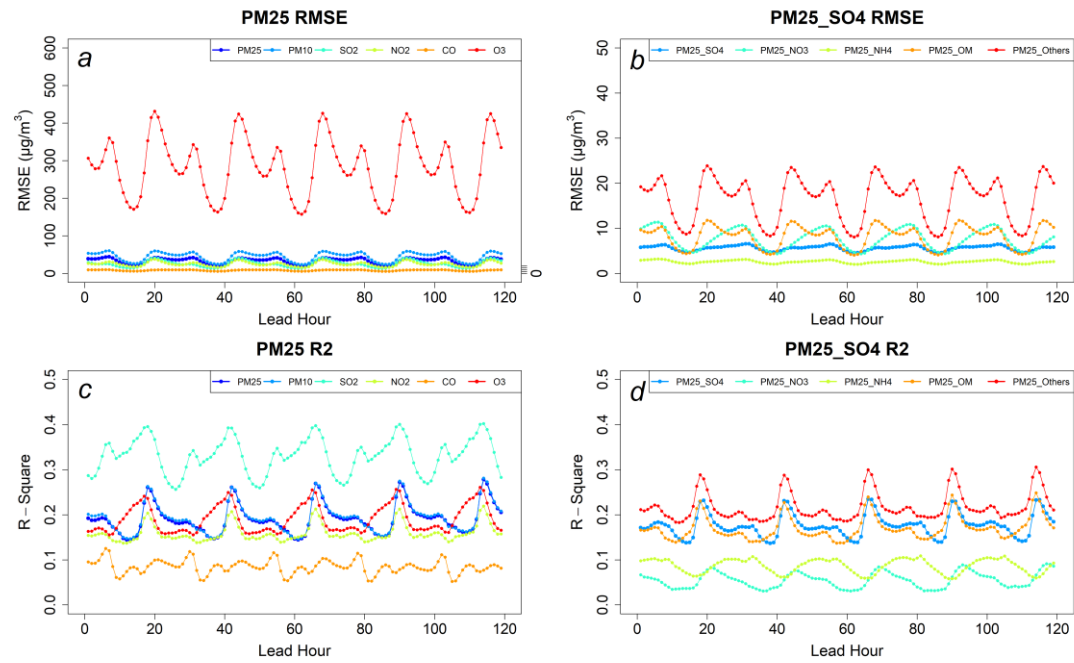


Figure S10: The evaluation performances of random forest forecasts against CMAQ forecasts in 2023. Panel (a) and (b) respectively show RMSE values of criteria pollutants and the $\text{PM}_{2.5}$ components of. Panel (c) and (d) respectively show R^2 values. It should be noted that RMSE value of CO corresponds to the right axis in panel (a).

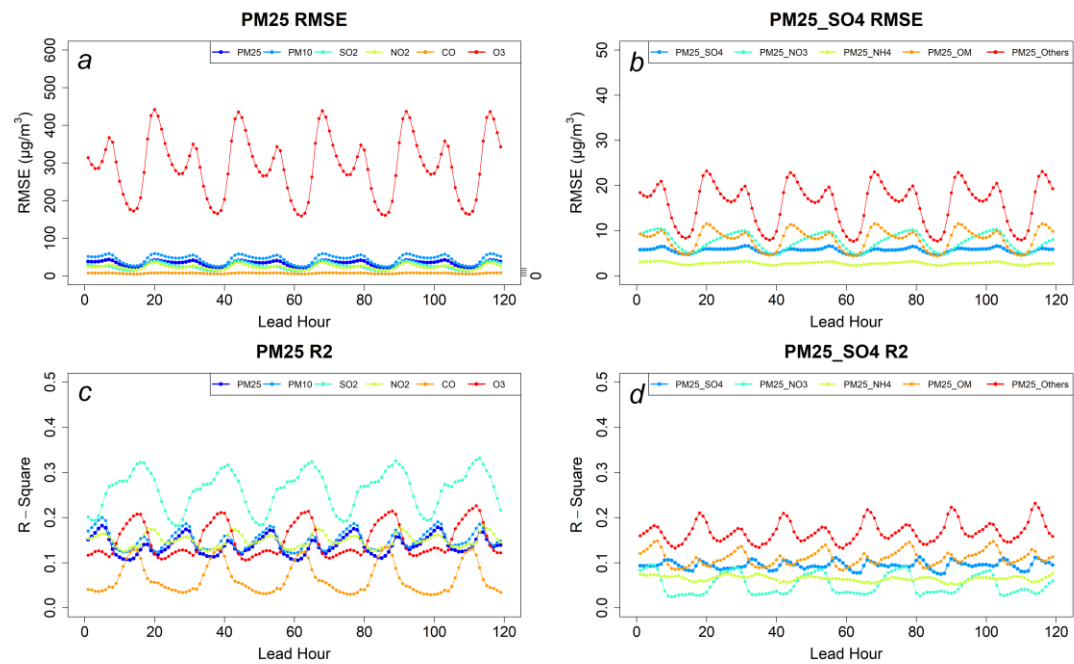


Figure S11: The evaluation performances of XGboost forecasts against CMAQ forecasts in 2023. Panel (a) and (b) respectively show RMSE values of criteria pollutants and the $\text{PM}_{2.5}$ components of. Panel (c) and (d) respectively show R^2 values. It should be noted that RMSE value of CO corresponds 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 the detailed description about the theory, model framework and numerical methods. We added one another lately review studies to reflect recent developments of 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 assume 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) within a neural framework. 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 critical 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) assume 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 diffusion process $\nabla(K\nabla C_i)$. It's also same for processes such as reaction, advection, and deposition. Therefore, it is not possible for one process to represent all atmospheric processes simultaneously. The independent contribution of each process depicted in Figure 12 of section 3.3. Clearly, each process exhibited their patterns of contribution to hourly air pollutant concentrations changes, constrained by the form of the operator in the processes. Related description was added in the 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 constraints 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 parent model was added in the Section 2.1

WRF-CMAQ simulations are not two-way coupled so that weather and chemistry and chemistry do not have feedbacks 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-form models, FastCTM is time-resolved with a 60 seconds step to simulate evolution of air pollutants. It generates hourly air quality simulations given hourly meteorological conditions and emissions. The simulations have good correlations with its parent numerical CTM model CMAQ. FastCTM also exhibited reasonable responses to emission interventions, also in close agreement to that of CMAQ. Besides, internal atmospheric processes could also be checked to reflect specific contributions from each process. We have added related literature review on the application of neural network 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 the motivation, there are studies dedicated to develop model 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 ~8 time speed-up. Sturm and Wexler (2020) developed a mass- and energy-conserving framework for using machine learning to speed computations with an 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 data 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 an complete deep learning based CTM.

Also in the Section 2.3.5 Chemical Reaction Module, comparison of FastCTM to reduced-form models are 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 critical 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) assume 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 (PM_{2.5}, 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 process-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 could simulate the contributions from each processes. 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 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 processes analysis could be different to that of CMAQ in its definition natures. 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 daytime could be due to more active vertical turbulence. 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 not 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 NH₄⁺ prediction

performance and agree that our initial explanation oversimplified 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 HNO_3 ($\text{NH}_3 + \text{HNO}_3 \rightarrow \text{NH}_4\text{NO}_3$) and H_2SO_4 ($2\text{NH}_3 + \text{H}_2\text{SO}_4 \rightarrow (\text{NH}_4)_2\text{SO}_4$), FastCTM does not explicitly encode these pathways. Instead, the neural network implicitly learns relationships between NH_4^+ and precursor emissions (NH_3 , NO_x , SO_2) 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 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 agree with the reviewer that attributing the error to NH_4^+ being a "sole cation" was incomplete and have revised the text accordingly in the 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 (NH_3 , NO_x , SO_2) 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 generally in good correlation with CMAQ forecasts in 2023, rather than that in 2018-2023. It means FastCTM has learn 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 the 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 4, 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.

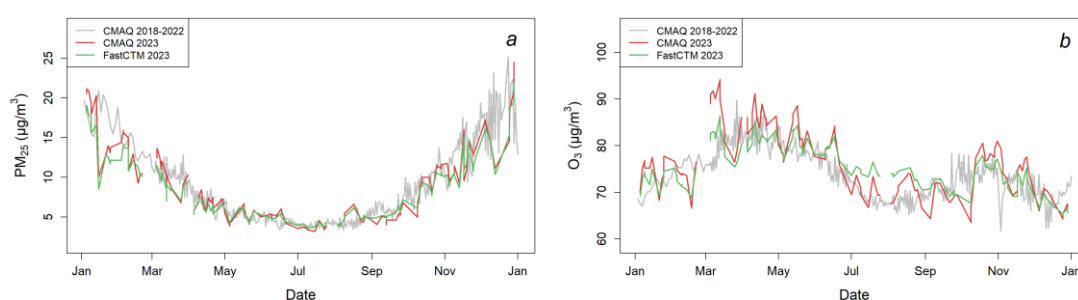


Figure 5: The evaluation performances of FastCTM forecasts against CMAQ forecasts in 2023.

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 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 all six pollutants under consideration, it is a notable finding that the MAE values tend to be higher in polluted areas. This can 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 typically hover 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.*

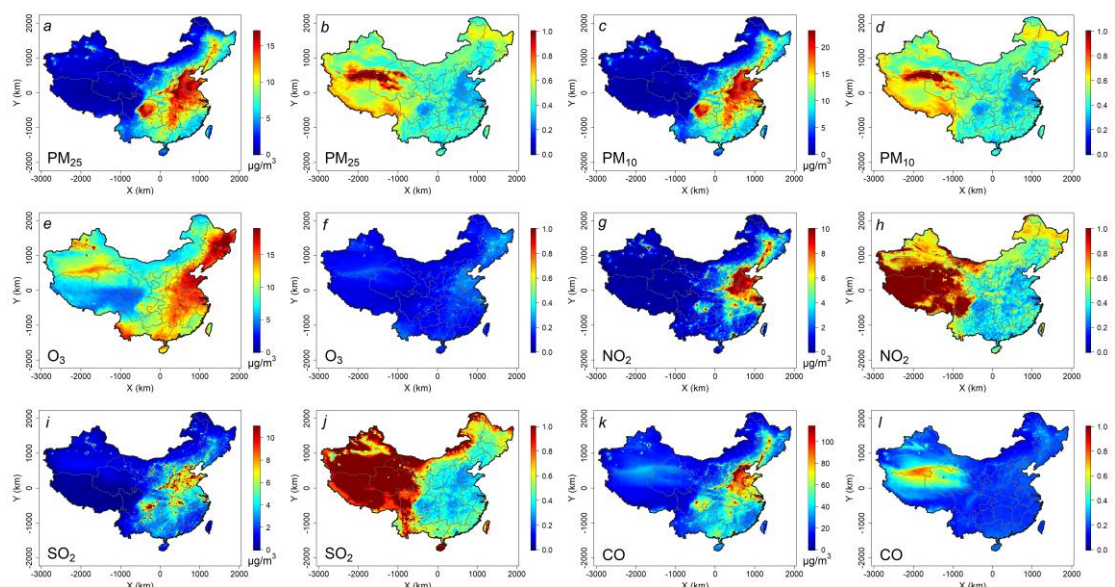


Figure 12: Spatial distribution of mean absolute error (panel a, c, e, g, i, and k) and normalized mean absolute error for six pollutants (panel b, d, f, h, j and l) of FastCTM comparing to CMAQ in 2023.

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 reviewer's comment that more analysis needed to verify FastCTM's capabilities to project the impact of a change in emissions, especially comparing to the parent model of WRF-CMAQ. We added comparison between FastCTM and CMAQ in simulating 11 emission intervention scenarios in the winter month of January 2019 and in summer month of July 2019. The results signified that, on the whole, the FastCTM simulations manifested a high level of concordance with those of CMAQ, which was manifested in two principal aspects. Firstly, akin to CMAQ, the FastCTM model forecasted positive responses to augmented emissions in the no-control (NCtrl) scenario and negative responses in the other emission-controlled scenarios. This implies that when emissions were unrestricted and increased, as in the NCtrl scenario, both models detected a corresponding upward trend in pollutant levels. Conversely, in scenarios where emissions were curbed, 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 showcases the model's sensitivity to the magnitude of emission interventions. 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 tandem with CMAQ. Related results in the manuscript were 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. 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 NO₂, FastCTM responded to emission changes with an interquartile range (IQR, 25% - 75% percentile) similar to that of CMAQ (Figure 9). For NO₂, in the same emission reduction scenarios, FastCTM

simulated lower NO₂ 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 2. The emission change details of emission scenarios

<i>Scenario</i>	<i>abbreviation</i>	<i>Sector</i>	<i>NO_x</i>	<i>VOCs</i>	<i>SO₂</i>	<i>CO</i>	<i>PM_{2.5}</i>	<i>PMC</i>
<i>nocontrol</i>	<i>NCtrl</i>	<i>Industrial</i>	30%	30%	30%	30%	30%	30%
		<i>Traffic</i>	20%	20%	20%	20%	20%	20%
<i>medianX</i>	<i>MedX</i>	<i>Industrial</i>	-36%	-35%	-48%	-23%	-9%	-9%
		<i>Traffic</i>	-40%	-10%	0	-26%	-10%	-10%
<i>medianY</i>	<i>MedY</i>	<i>Industrial</i>	-26%	-20%	-38%	-13%	-4%	-4%
		<i>Traffic</i>	-30%	0%	0	-16%	-5%	-5%
<i>medianZ</i>	<i>MedZ</i>	<i>Industrial</i>	-36%	-10%	-48%	-23%	-9%	-9%
		<i>Traffic</i>	-40%	0%	0	-26%	-10%	-10%
<i>median-3</i>	<i>Med-3</i>	<i>Industrial</i>	-10%	-10%	-18%	0	0	0
		<i>Traffic</i>	-10%	0%	0	0	0	0
<i>median-2</i>	<i>Med-2</i>	<i>Industrial</i>	-16%	-20%	-28%	-3%	0	0
		<i>Traffic</i>	-20%	0%	0	-6%	0	0
<i>median-1</i>	<i>Med-1</i>	<i>Industrial</i>	-26%	-35%	-38%	-13%	-4%	-4%
		<i>Traffic</i>	-30%	-10%	0	-16%	-5%	-5%
<i>median0</i>	<i>Med0</i>	<i>Industrial</i>	-36%	-50%	-48%	-23%	-9%	-9%
		<i>Traffic</i>	-40%	-20%	0	-26%	-10%	-10%
<i>median+1</i>	<i>Med+1</i>	<i>Industrial</i>	-46%	-65%	-58%	-33%	-19%	-19%
		<i>Traffic</i>	-50%	-30%	0	-36%	-20%	-20%
<i>median2030</i>	<i>Med30</i>	<i>Industrial</i>	-55%	-70%	-80%	-40%	-40%	-40%
		<i>Traffic</i>	-60%	-40%	0	-40%	-40%	-40%
	<i>Med35</i>	<i>Industrial</i>	-80%	-80%	-90%	-60%	-50%	-50%

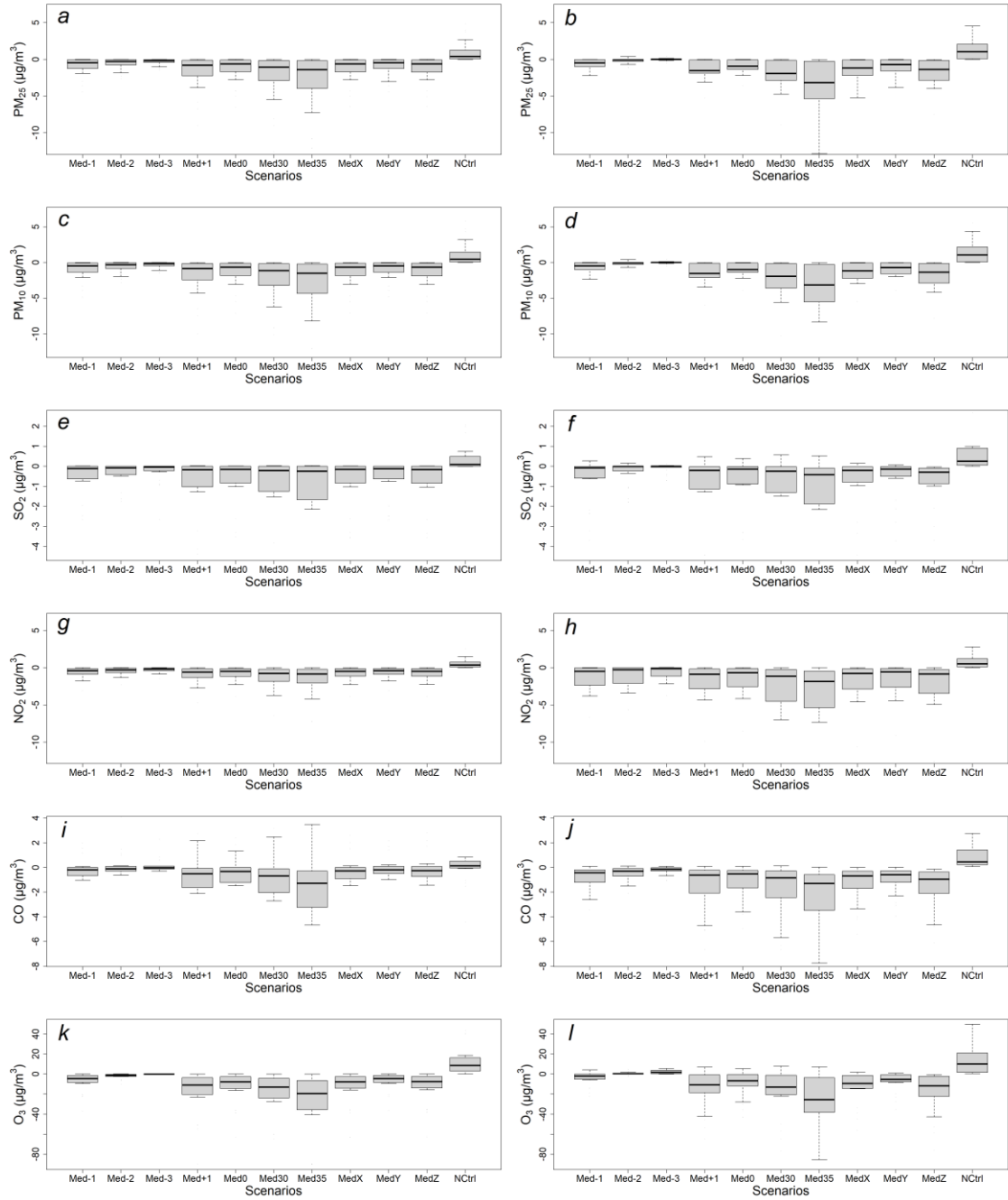


Figure 9: Air pollutant concentration changes in terms of base case simulated by CMAQ (subplots of a, c, e, g, i and k in first column) and by FastCTM (subplots of b, d, f, h, j and l in second column) in January 2019.

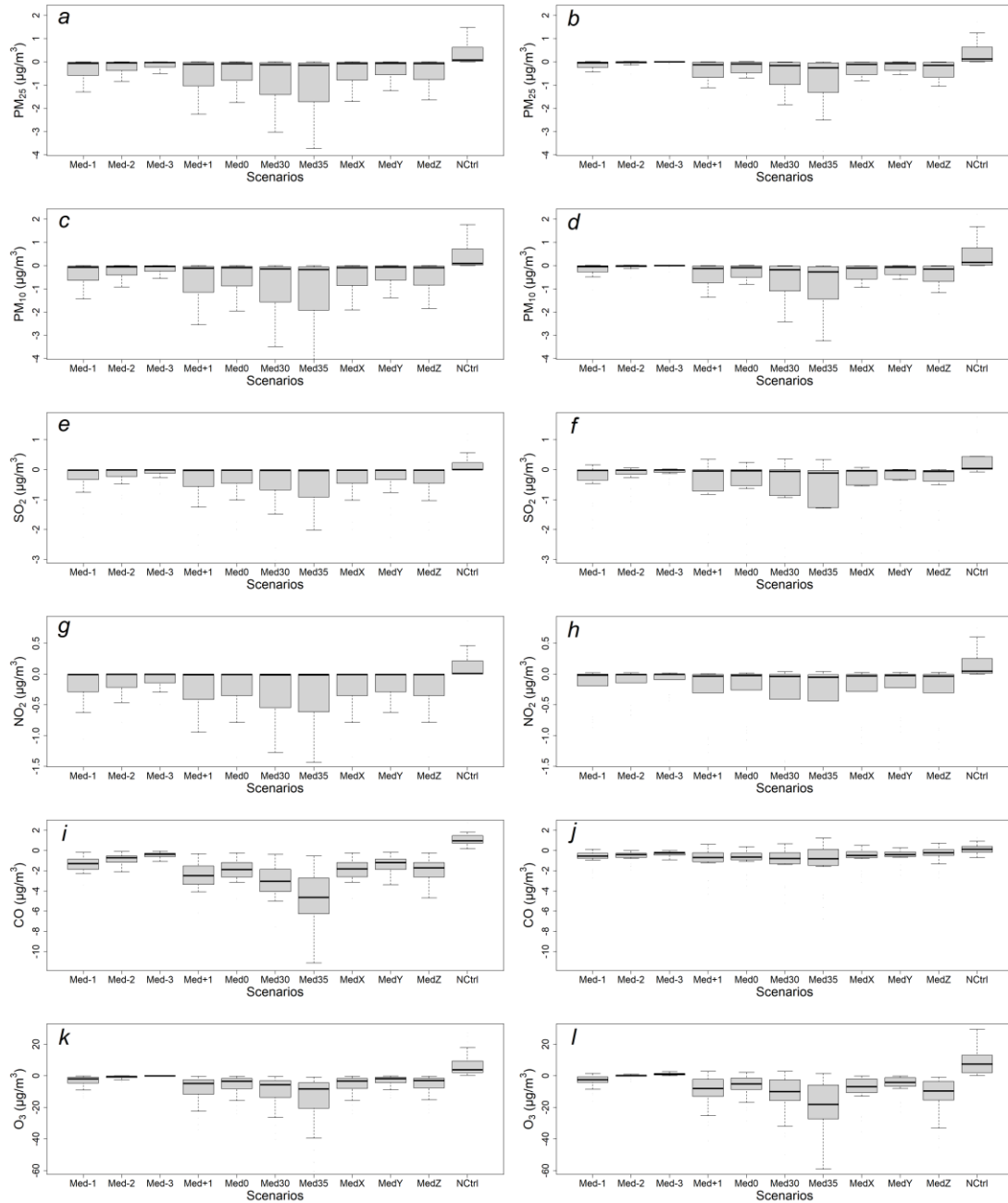


Figure 10: Air pollutant concentration changes in terms of base case simulated by CMAQ (subplots of a, c, e, g, i and k in first column) and by FastCTM (subplots of b, d, f, h, j and l in second column) in July 2019.

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 that 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). Specifically, the simulation outcomes of atmospheric 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 that govern PM_{2.5} levels.

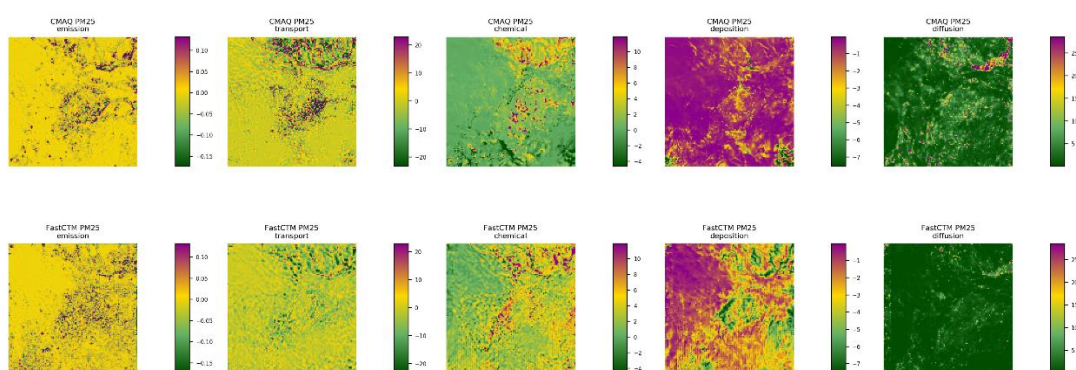


Figure 13: An example of contributions from five major atmospheric processes to PM_{2.5} changes ($\mu\text{g}/\text{m}^3$) by CMAQ (first row) and FastCTM (second row) in 23:00 on October 13, 2024.