Authors' responses to the reviewers

Dear Editor and Reviewers,

Thank you so much for conducting such a careful and timely review of our manuscript (Manuscript Number: EGUSPHERE-2025-4441), entitled "Evaluating the EPICC-Model for Regional Air Quality Simulation: A Comparative Study with CAMx and CMAQ", and for recommending major revisions before it can be considered for publication in ***Geoscientific Model Development***.

We sincerely appreciate the thoughtful evaluations and constructive feedback provided by the anonymous reviewers, which have been very helpful in improving the quality of our manuscript. We have made every effort to address all the concerns raised and have revised the manuscript accordingly.

A **Revision List** and **detailed Point-By-Point Responses** are provided in the following sections. We have carefully evaluated and thoroughly investigated the references suggested by the reviewers. In addition, we have consulted a broader range of literature and revised parts of the manuscript where the original wording lacked clarity, in order to enhance the scientific rigor and overall readability of the paper. All major revisions in the manuscript are marked in blue to facilitate the review process. The line numbers provided refer to the revised manuscript with tracked changes.

Thank you once again for your time and effort in handling our manuscript.

Yours sincerely,

Qizhong Wu
Beijing Normal University
2026/03/20

Responses to Reviewer

Comment 1:

A) The paper stresses systematic consistency in the model comparison. But according to Table 1 that is not entirely accurate. The EPICC-Model uses 20 vertical layers, while CAMx and CMAQ use 14. The authors stress the better layer structure in EPICC-Model as"demonstrating superior vertical resolution compared to the 14 layers coordinates used in other two models" especially "for nearsurface turbulent mixing and nocturnal stable layer structures." CAMx and CMAQ should use the same 20 layer structure as EPICC-Model to improve model-to-model consistency and remove a key source of uncertainty in the comparison of model results. If a single WRF run was used to drive all 3 models (run with 30 layers as shown in Table 2), use of consistent layer structures is possible. While CAMx allows for sub-setting WRF layers to fewer CAMx layers, the CMAQ MCIP preprocessor does not allow for this and requires the use of the full 30 layer WRF structure. So it is not clear how the authors transmitted 30-layer WRF meteorological fields to a subset of 14 CMAQ layers, unless they ran a separate configuration of WRF with 14 layers or developed their own WRF-to-CMAQ meteorological interface. Additionally, the authors state that both CMAQ and CAMx use "default" boundary conditions to specify concentrations on the 45-km outer grid. These values and their sources should be stated in the paper with a rationale justifying their selection. Both CMAQ and CAMx can utilize the same MOZART global model output data to derive boundary conditions, as was done for EPICC-Model. While I understand that boundary conditions do not appear to have appreciable effects on model results overall, use of consistent boundary conditions is preferable.

B) A comparison of EPICC-Model and CAMx in Table 1, and the descriptions of these models in the text, indicate that most EPIC-Model algorithms are similar or identical to CAMx algorithms. The main difference is in the chemical treatments. CAMx was run with the CB05 photochemistry mechanism, which is an odd choice given that it represents 20-year old science. CAMx v7.00 used in this study includes CB6r4, and CAMx v7.10 released just 6 months later in 2020 includes CB6r5 (the same CAMx mechanism that was adopted by EPICC-Model). This is a critical issue and CAMx should be rerun with CB6 chemistry. If the MEIC VOC emissions were uniformly speciated to CB6 compounds for all 3 models, then that speciation would lead to missing VOC mass in CB05 due to newer CB6 species that are unrecognized by CB05. That alone would lead to different photochemical responses. A full explanation is needed on how MEIC emissions were temporally, spatially, and chemically processed for each of the 3 models.

C) Additionally, several of the simple descriptions for CAMx processes in Table 1 are inadequate or inaccurate. The PPM vertical advection solver was not yet available in CAMx v7.00 or v7.10, so that entry in Table 1 is either incorrect, or the authors used CAMx v7.20 or later. It is inaccurate to state that CAMx employs the YSU vertical diffusion scheme, rather it employs a first-order K-theory approach (apparently like EPICC-Model) where the vertical diffusivities are diagnostically derived from WRF meteorological fields based on the YSU approach. CAMx v7.00 does not use the original SOAP scheme of Strader et al. (1999), but an updated SOAP2 scheme unique to CAMx (refer to the CAMx v7 user's guide). CAMx uses the full-science TUV as a pre-processor to generate a photolysis rate input file, then internally applies a streamlined

TUV version to adjust those for effects from clouds and aerosols. This identical TUV process was adopted in EPICC-Model (according to Wang et al., 2025: Development and evaluation of photolysis and gas-phase reaction scheme in EPICCmodel: Impacts on tropospheric ozone simulation, Atmos. Environ., https://doi.org/10.1016/j.atmosenv.2025.121373).

Response to comment 1:

We thank the reviewer for the constructive comments, which have helped us further clarify model configurations and improve the rigor of the manuscript.

A) Regarding the different vertical layer structures used by the three models, the primary reason is that EPICC-Model adopts a terrain-following height coordinate, whereas CAMx and CMAQ employ terrain-following sigma-pressure coordinates. Due to differences in vertical coordinate definitions and internal numerical implementations, it is not feasible to maintain a completely identical vertical layer structure among the three models. With respect to the transmission of WRF meteorological fields, this study used WRF v3 as the meteorological driver. This version of WRF adopts a terrain-following sigma-pressure coordinate system. The meteorological fields were processed using MCIP v4, which supports vertical layer reduction of WRF outputs. It should be noted that since WRF v4 introduced a hybrid vertical coordinate system, vertical layer reduction is generally not recommended, and newer versions of MCIP no longer support this operation. Regarding boundary conditions, the term "default" in the manuscript refers to the built-in profile-type boundary condition schemes in both models. These schemes construct vertical concentration profiles based on average background concentrations and generate lateral boundary concentrations through interpolation. The corresponding configuration parameters can be found in the model setup files.

B) Concerning the choice of chemical mechanism, although CB6 represents a more recent update compared to CB05, there is currently no consistent conclusion that CB6 necessarily performs better than CB05 over China. CB05 has been widely applied in long-term regional air quality simulations in China and has undergone extensive evaluation, demonstrating good stability and applicability. Therefore, CB05 was retained in this study. For the MEIC emission inventory, temporal allocation factors were first applied to distribute emissions at monthly, daily, and hourly scales, followed by spatial allocation using a first-order conservative interpolation method. Chemical speciation was conducted with reference to the SPECIATE database (https://www.epa.gov/air-emissions-modeling/speciate), and species mapping was performed separately according to the chemical mechanism used in each model to ensure mass conservation within each mechanism framework. Therefore, no VOC mass was lost due to incompatibility between CB6 species and CB05.

C) Regarding the vertical advection scheme, after re-examination we confirm that CAMx v7.0 was used in this study, and the PPM option can be enabled in this version. For vertical diffusion, CAMx applies a first-order K-theory framework to calculate vertical diffusion fluxes, and the vertical diffusivity coefficients are diagnostically derived from WRF meteorological fields following the YSU boundary layer parameterization approach; therefore, the YSU-related description has been retained. With respect to secondary organic aerosol treatment, the reviewer is correct that the updated SOAP2.2 scheme was used rather than the original SOAP scheme. This has been revised in both the main text and Table 1. For photolysis processes, CAMx first uses the full-science TUV model as a preprocessor to generate photolysis rate input files, and then applies a streamlined TUV module within the model to dynamically adjust photolysis rates according to cloud and aerosol conditions. EPICC-Model adopts the same approach. To avoid ambiguity, both the full-science TUV and streamlined TUV components have been explicitly listed in Table 1.

Comment 2:

Abstract, Line 41: I didn't see specific references or comparisons of EPICC-Model performance "against international benchmarks" in the body of the paper. Do you mean "against international models"?

Response to comment 2:

Thank the reviewer for pointing this out. The phrase "international benchmarks" in the original text was indeed not precise. We agree that the more accurate phrasing should be "against international models".

Comment 3:

Line 56: I suggest adding a phrase like "due to a non-linear response to precursor NOx reductions" or similar to add clarity about the cause of increasing urban ozone.

Response to comment 3:

Thank the reviewer for the valuable suggestion. We fully agree that adding the causal explanation "due to a non-linear response to precursor NOx reductions" when describing the rise in ozone levels makes the statement more rigorous and complete. We have incorporated this phrase as a parenthetical explanation into the original text. The revised sentence is as follows:
(Lines 55-58) Despite $PM_{2.5}$ reductions, O_3 levels have risen steadily due to a non-linear response to precursor NOx reductions, with warm-season maximum daily 8-hour average (MDA8) O_3 increasing by 1.2 ± 1.3 ppb yr^{-1} in major urban clusters during 2013-2022 (Wang et al., 2024).

Comment 4:

Numerous locations: Certain words are used throughout the paper that carry vague meaning, most frequently the word "stable". For example, Lines 199-200, "Simulations of 10 m wind speed were generally stable", and again on Line 212, "Surface pressure simulations showed strong stability." It is unclear what the word stable is meant to convey. I have not seen other uses of "stable" in the context of model performance *evaluation, so I cannot suggest an alternative word. Other occurrences:*

Lines 338: "EPICC-Model ... reflecting a more balanced and stable performance overall."

388: "CAMx is stable (IOA=0.78) and suitable for multi-seasonal averages."

426: "Comprehensive evaluation indicates that the EPICC-Model shows stable seasonal MDA8 O3 performance"

437: "Correlation coefficients exceed 0.75 for all models, while RMSE and STD remain relatively low, indicating higher stability and accuracy in MDA8 O3 simulations compared to PM2.5."

540: "BC exhibits the most compact R distribution, with low inter-city variability, reflecting stable and consistent simulation by all models."

545: "SO_4^{2-} and NH_4^+ exhibit moderate RMSE, indicating relatively stable simulations"

579: "The EPICC-Model shows stable performance in forecasting good to moderate pollution levels"

645: "incorporating pollutant transport pathways and the stability of meteorological fields."

Response to comment 4:

We thank the reviewer for highlighting the ambiguous use of the word "stable" in the manuscript. We have revised the relevant expressions into more precise academic terminology based on the specific context. The revised sentences are as follows:

(Lines 201-203) Simulations of 10 m wind speed showed consistent performance, with R exceeding 0.7 and RMSE below 8 m/s in most regions.

(Lines 214-215) Surface pressure simulations exhibited high consistency, with R values exceeding 0.7 across the country.

(Lines 340-342) Notably, the EPICC-Model not only accurately captured peak concentrations but also achieved the lowest annual NMB, reflecting a more robust performance overall.

(Lines 392-394) CAMx shows consistent performance (IOA=0.78) and is suitable for multi-seasonal averages but underestimates high-humidity regions and responds weakly to severe pollution.

(Lines 434-436) Comprehensive evaluation indicates that the EPICC-Model shows consistent seasonal MDA8 O_3 performance with an annual IOA of 0.85, capturing O_3 distribution across China.

(Lines 450-452) Correlation coefficients exceed 0.75 for all models, while RMSE and STD remain relatively low, indicating lower variability and higher accuracy in MDA8 O_3 simulations compared to $PM_{2.5}$.

(Lines 557-559) BC exhibits the most compact R distribution, with low inter-city variability, reflecting robust and consistent simulation by all models.

(Lines 561-564) The EPICC-Model performs well in this aspect. SO_4^{2-} and NH_4^+ exhibit moderate RMSE, indicating relatively low variability simulations, while BC shows the lowest RMSE and most compact distribution, suggesting high accuracy.

(Lines 596-600) The EPICC-Model shows reliable performance in forecasting good to moderate pollution levels, indicating strong capability in simulating the initial formation and accumulation of secondary pollutants under relatively low to moderate

precursor concentrations.
(Lines 664-668) To further evaluate the capability of the EPICC-Model, CAMx, and CMAQ in simulating trans-regional and long-duration pollution episodes, this study selects typical persistent $PM_{2.5}$ and O_3 events in 2021, and conducts a dedicated analysis incorporating pollutant transport pathways and the spatiotemporal consistency of meteorological fields.

Comment 5:

Line 214: A surface pressure MB of 20 hPa (20 mb!) is extraordinarily large and highly concerning. Perhaps this is a typographical error and the authors meant 2 hPa (2 mb)?

Response to comment 5:

Thank the reviewer for raising this point. The statement "MB < 20 hPa" is correct as written. It indicates that the maximum regional mean bias between simulations and observations in parts of Southwest and Central China was less than 20 hPa.
For context, the typical surface pressure is approximately 1013 hPa, so a 20 hPa bias represents a relative error of about 2%. In regional climate modeling over complex terrain like southwestern China, biases of this magnitude are common and considered acceptable. Similar magnitudes of regional mean bias (10-20 hPa) in surface pressure are frequently reported in comparable model evaluation studies. These biases primarily stem from the challenges in simulating boundary-layer processes and local circulations over complex topography, rather than from typographical or unit errors.

Comment 6:

Lines 229-235: The authors appropriately argue about the applicability of different emission inventories representing different years. On Line 230, "emission inventories do not exactly match the simulation year", I suggest the removal of "exactly" as it's clear they do not match. I agree with this statement: "Under the assumption of no substantial changes in regional climate and socioeconomic activities, such inventories can reasonably represent the emission patterns of the study period." However, I suggest adding an explicit statement to the effect of "No projections of the individual emission inventory data to the 2021 simulation year were performed" to add clarity that the emission inventories were used without modification or alignment.

Response to comment 6:

We thank the reviewer for this meticulous and important suggestion. We fully agree that removing "exactly" makes the statement more direct, and adding a clarification that no temporal adjustment was performed significantly enhances the transparency of our methodology. We have incorporated both suggestions in the revised manuscript: 1) the word "exactly" has been deleted, and 2) an explicit statement, "It is important to clarify that no projection or temporal adjustment of the individual emission inventory data to the 2021 simulation year was performed." has been added at the end of the paragraph.

Comment 7:

Lines 291-292: "In contrast, the EPICC-Model shows greater adaptability..." This statement is not clear, although I think I understand the intent. Perhaps a better word like "variability" or "spatial/temporal response" would be better than "adaptability" to reflect the authors' meaning.

Response to comment 7:

We thank the reviewer for this helpful suggestion. We agree that the term "adaptability" may be ambiguous in the context of model performance evaluation. Our intention was to describe the model's more distinct seasonal response under varying background concentration conditions. In the revised manuscript, we have replaced "adaptability" with "more pronounced seasonal response" to better reflect the intended meaning. The revised sentences are as follows:
(Lines 295-296) In contrast, the EPICC-Model exhibits a more pronounced seasonal response.

Comment 8:

Lines 313-317: "This discrepancy mainly stems from differences in the treatment of photolysis and chemical mechanisms among the models" and "In addition, the EPICC-Model employed the CB6r5 chemical mechanism, which offered more comprehensive representation of BVOC oxidation (especially isoprene) compared to CB6r3 (used in CMAQ) and CB05 (used in CAMx), thereby increasing O3 formation potential." Certainly true – see my comments above about running CAMx with CB05 rather than CB6r5. Yet the photolysis treatments in EPICC-Model and CAMx are apparently identical so the reference to different photolysis should be removed.

Response to comment 8:

We thank the reviewer for this important comment. We agree that the photolysis treatments in EPICC-Model and CAMx are essentially identical, and therefore differences in photolysis schemes should not be cited as a primary source of discrepancy among the models. In the revised manuscript, we have removed the reference to general differences in photolysis treatment and clarified that the discrepancy mainly arises from differences in chemical mechanism design, including the implementation of heterogeneous HONO formation and nitrate photolysis pathways in EPICC-Model. The revised sentences are as follows:
(Lines 317-319) This discrepancy mainly stems from differences in chemical mechanisms design among the models.

Comment 9:

Lines 403-405: "CAMx generally underestimated O3 nationwide, likely due to limitations of the CB05 mechanism and its coarse $14\sigma_p$ vertical resolution." See my comments above. The selection of CB05 and 14 vertical layers are the authors' choice, not fixed features of CAMx or CMAQ.

Response to comment 9:

We thank the reviewer for this important clarification. We agree that the use of the CB05 chemical mechanism and the 14-layer vertical configuration reflects modeling choices made in this study rather than inherent limitations of CAMx. Our original wording was not sufficiently precise and may have unintentionally implied otherwise. The revised sentences are as follows:
(Lines 408-413) CAMx generally underestimated O_3 nationwide, which may be associated with the use of the CB05 chemical mechanism and the 14σ_p vertical configuration adopted in this study. (Ren et al., 2022; Tang et al., 2011).

Comment 10:

Lines 411-412: "Given the use of identical meteorological fields". This is untrue, see my comments about Table 1 above. Meteorology is not identical using different layer structures and boundary layer treatments.

Response to comment 10:

We thank the reviewer for this clarification. We agree that the phrase "identical meteorological fields" was not sufficiently precise and may have caused confusion. Our intention was to indicate that all three air quality models were driven by meteorological fields generated from the same WRFv3 simulation with 30 vertical layers, rather than implying that the meteorological representations within each model were strictly identical. The revised sentences are as follows:
(Lines 418-421) Given that all three models were driven by meteorological fields derived from the same WRFv3 simulation, regional O_3 levels are primarily controlled by photochemical production rather than constrained by boundary inputs (Li et al., 2019a).

Comment 11:

Figure 5 and associated text: Please include a description of Taylor diagrams and how to interpret them. What does standard deviation (STD) refer to, the modeled variability or model-measurement variability, and what are the units? Why is STD=1 the best? The diagrams appear to be missing value labels for RMSE.

Response to comment 11:

We thank the reviewer for this helpful suggestion. We agree that additional explanation of the Taylor diagrams and their interpretation would improve clarity. The Taylor diagram provides a concise statistical summary of model performance by simultaneously displaying the correlation coefficient (R), normalized standard deviation (STD), and centered root mean square error (RMSE) between simulations and observations (Taylor, 2001). In this study, STD represents the standard deviation of model simulations normalized by the observed standard deviation; therefore, it is dimensionless. A normalized STD of 1 indicates that the model reproduces the same variability magnitude as the observations. Values greater than 1 indicate overestimation of variability, while values less than 1 indicate underestimation. The radial distance from the origin represents normalized STD, the angular position corresponds to the correlation coefficient, and the dashed arcs denote contours of centered RMSE relative to observations. The reference point (REF) represents observations with R = 1 and STD = 1. The revised sentences are as follows:

(Lines 443-448) The Taylor diagram simultaneously displays the correlation coefficient (R), normalized standard deviation (STD), and centered root mean square error (RMSE). In the diagram, the radial distance represents STD, the angular coordinate indicates R, and dashed contours denote RMSE relative to observations. The "REF" point (R = 1, STD = 1) represents perfect agreement with observations.

Comment 12:

Line 446: "CMAQ demonstrating greater sensitivity to pollution peaks". The word sensitivity seems inappropriate here, perhaps better to say "greater ability to replicate pollution peaks". Also, Line 448 "indicating enhanced error sensitivity", perhaps simply remove the word "sensitivity" because including it makes the meaning unclear.

Response to comment 12:

We thank the reviewer for this helpful suggestion. We agree that the term "sensitivity" may not accurately convey our intended meaning in this context. Our intention was to describe CMAQ's ability to reproduce pollution peaks rather than implying parameter sensitivity. Accordingly, we have revised "CMAQ demonstrating greater sensitivity to pollution peaks" to "CMAQ demonstrating a greater ability to replicate pollution peaks." We have also removed the term "error sensitivity" and rephrased the sentence to improve clarity. The manuscript has been revised accordingly. The revised sentences are as follows:

(Lines 457-460) Seasonally, $PM_{2.5}$ simulation accuracy exhibits marked variations: during winter and spring high-concentration periods, all three models show comparable performance, with CMAQ demonstrating greater ability to replicate pollution peaks (STD≈1, closest to observations).

(Lines 460-462) Model performance declines significantly under summer's low-background conditions (R<0.55, elevated RMSE/STD), indicating enhanced error during high-concentration periods.

Comment 13:

Lines 478-481: "The insufficient formation of SO_4^{2-} not only limits the production of ammonium sulphate [$(NH_4)_2SO_4$] but also reduces the consumption potential of NH_3 (Gao et al., 2018), thereby suppressing the simulated concentration of ammonium (NH_4^+)." Certainly true. This also frees up more NH_3 to bond with HNO_3, driving up nitrate levels – one possible reason for nitrate over predictions.

Response to comment 13:

We thank the reviewer for this insightful comment. We agree that reduced sulfate formation may increase the availability of free NH_3, which can subsequently react with HNO_3 to form ammonium nitrate and potentially contribute to nitrate overestimation.

This is an important mechanistic linkage that strengthens the interpretation of our results. The revised paragraph is as follows:
(Lines 489-498) Meanwhile, all three models consistently underestimated sulfate (SO_4^{2-}), a bias consistent with existing research findings (Shao et al., 2019), which may stem from underestimating SO_2 oxidation rates or uncertainties in intra-cloud aqueous-phase chemical mechanisms. The insufficient formation of SO_4^{2-} not only limits the production of ammonium sulphate [$(NH_4)_2SO_4$] but also reduces the consumption potential of NH_3 (Gao et al., 2018), thereby suppressing the simulated concentration of ammonium (NH_4^+). In addition, the reduced formation of ammonium sulfate may leave more free NH_3 available to react with HNO_3, potentially enhancing the formation of ammonium nitrate (NH_4NO_3) and contributing to the overestimation of nitrate in some regions.

Comment 14:

Lines 576-579: "CAMx performes best in predicting AQI (84%), PM2.5 (81%), and MDA8 O3 (89%) levels, primarily attributed to its detailed representation of gas-phase precursor reaction chains under low concentration conditions." The word "performes" is misspelled. Attempting to attribute CAMx performance in this manner is conjecture and unclear, especially given the use of CB05. I suggest removing such conjecture if concrete reasoning backed by appropriate evidence (such as process analysis) and/or a reference cannot be provided.

Response to comment 14:

We thank the reviewer for this careful comment. We have corrected the spelling error ("performes" to "performs"). We agree that our previous attribution of CAMx performance to its gas-phase reaction representation was speculative and not sufficiently supported by process-level analysis or specific references. To avoid conjecture, we have removed this interpretation and revised the text to focus on objective performance statistics. The revised sentences are as follows:
(Lines 593-596) In terms of overall forecasting accuracy, CAMx performs best in predicting AQI (84%), $PM_{2.5}$ (81%), and MDA8 O_3 (89%) levels.

Comment 15:

Lines 580-581: "indicating strong capability in simulating regional pollution transformation and the early evolution stages of pollution processes." What is meant by "transformation" and "early evolution stages" here? Please explain or simply revise this sentence to be more specific.

Response to comment 15:

We thank the reviewer for pointing out the lack of specificity in our wording. We agree that the terms "transformation" and "early evolution stages" were overly general and could be unclear. In the revised manuscript, we have replaced these expressions with more specific descriptions, clarifying that the EPICC-Model demonstrates strong capability in simulating the initial formation and accumulation of secondary pollutants

under relatively low to moderate precursor concentrations. The text has been revised accordingly to improve clarity and precision. The revised sentences are as follows:
(Lines 596-600) The EPICC-Model shows reliable performance in forecasting good to moderate pollution levels, indicating strong capability in simulating the initial formation and accumulation of secondary pollutants under relatively low to moderate precursor concentrations.

Comment 16:

Lines 589-592: "This advantage is attributed to its comprehensive aerosol-radiation-cloud feedback mechanisms, such as the AERO6 module, which provide detailed representation of pollutant accumulation and regional transport processes." CMAQ AERO6 is stated, but Table 1 indicates that CMAQ was run with AERO7. Like my comment on Lines 576-579, I suggest removing such conjecture about the attribution of CMAQ performance without concrete evidence (process analysis results).

Response to comment 16:

We thank the reviewer for this careful observation. We agree that attributing CMAQ's performance advantage to specific aerosol-radiation-cloud feedback mechanisms without supporting process analysis is speculative. To avoid unsupported conjecture, we have removed the attribution statement and revised the text to focus on objective performance differences.

Comment 17:

Lines 621-622: "CAMx achieves the lowest false alarm rate (4.9%), demonstrating stronger robustness." The word robustness seems to be inappropriate in this context (like "stable" examples above). More simply, the false alarm rate is low because CAMx underpredicts $PM_{2.5}$ levels and does not predict high PM events.

Response to comment 17:

We thank the reviewer for pointing this out. We agree that the term "robustness" was not appropriate in this context. We have revised the sentence to clarify that the lower false alarm rate is likely associated with CAMx's tendency to underpredict $PM_{2.5}$ concentrations and forecast fewer high pollution events. We also retained the explanation regarding its relatively low hit rate, which reflects a conservative forecasting tendency. The revised sentences are as follows:
(Lines 639-643) CAMx achieves the lowest false alarm rate (4.9%), likely because it tends to underpredict $PM_{2.5}$ levels and forecast fewer high pollution events. However, its hit rate is also relatively low (45.0%), reflecting a more conservative forecasting tendency.

Comment 18:

Lines 623-625: "The EPICC-Model strikes a balance between hit rate (55.0%) and false alarm rate (7.6%), suggesting an optimized trade-off between accuracy and reliability." The trade-off is true, but it cannot be considered purposely "optimized". I *suggest removing the word "optimized" as it is misleading.*

Response to comment 18:

We thank the reviewer for this helpful comment. We agree that the term "optimized" may be misleading, as no explicit optimization procedure was performed in this study. Our intention was to describe the statistical balance between hit rate and false alarm rate rather than implying purposeful optimization. We have removed the word "optimized" and revised the sentence to describe the trade-off more objectively. The manuscript has been updated accordingly. The revised sentences are as follows:
(Lines 643-646) The EPICC-Model strikes a balance between hit rate (55.0%) and false alarm rate (7.6%), suggesting a trade-off between detection capability and false alarm occurrence.

Comment 19:

Lines 630-631: "This reflects a conservative strategy that prioritizes avoiding false positives under extreme pollution conditions." This statement is probably lost in translation. The models do not present a strategy of any kind, especially that avoids certain performance issues – they are objective methods based on the science algorithms they employ. Perhaps a more accurate statement would be, "This reflects an inability of these models to simulate extreme pollution conditions".

Response to comment 19:

We thank the reviewer for this helpful clarification. We agree that the previous wording was inappropriate, as the models do not employ any "strategy" or intentionally prioritize specific performance outcomes. Our original expression was imprecise. The revised sentences are as follows:
(Lines 651-653) This reflects limitations of the models in simulating extreme pollution conditions.

Comment 20:

Lines 632-633: "the EPICC-Model achieves the highest hit rate (45.6%), significantly outperforming CAMx (43.0%) and CMAQ (38.4%)." Be careful using words like "significantly" in a qualitative and subjective sense, as readers will have different ideas on what "significant" means to them. I suggest removing that word, especially since a difference of 2-7% seems small and likely not statistically significant (unless you report a statistical hypothesis test with a set p-value).

Response to comment 20:

We thank the reviewer for this important comment. We agree that the term "significantly" may imply statistical significance, which was not formally tested in this study. Our intention was only to describe numerical differences rather than to claim statistical significance. The revised sentences are as follows:
(Lines 653-656) For O_3 pollution events (MDA8 O_3 > 160 μg m^{-3}), the EPICC-Model achieves the highest hit rate (45.6%), slightly higher than CAMx (43.0%) and CMAQ

(38.4%), while also maintaining a low false alarm rate (3.8%).

Comment 21:

Lines 634-636: "This suggests advantages in its representation of O3 precursor transport, photochemical mechanisms, and boundary layer feedback processes." Again, like my comment on Lines 576-579, I suggest removing such conjecture about the attribution of EPICC-Model performance without concrete evidence (process analysis results), especially given small differences in the hit rate statistics and many similarities in model treatments (photochemical mechanisms, boundary layer dynamics) and source of inputs.

Response to comment 21:

We thank the reviewer for the comment. We have revised the sentence to remove the speculative attribution of the EPICC-Model's performance. The text now objectively reports the model hit rate without inferring specific causes.

Comment 22:

Lines 639-640: "CAMx exhibits relatively low DROC across all thresholds, consistent with its conservative simulation style." Like my comments for Lines 630-631, CAMx is not developed to possess a "style" of any kind. I suggest removing the last part of the sentence starting at "consistent".

Response to comment 22:

We thank the reviewer for the comment. We have removed the speculative phrase referring to CAMx's "conservative simulation style" and now report only the observed DROC values. The revised sentences are as follows:
(Lines 661-662) CAMx exhibits relatively low DROC across all thresholds.

Comment 23:

Lines 657-658: "The EPICC-Model showed superior performance in reproducing spatial gradients and temporal patterns." The word "superior" is too strong. Given results in Figure 9 it's difficult to assess this visually. I suggest replacing "superior" with "better" or "improved".

Response to comment 23:

We thank the reviewer for the comment. We have replaced "superior" with "better" to provide a more objective description of the EPICC-Model's performance. The revised sentences are as follows:
(Lines 679-682) The EPICC-Model showed better performance in reproducing spatial gradients and temporal patterns, particularly in the centre of the North China Plain and Yangtze River Middle-Reach corridor.

Comment 24:

Lines 661-662: "likely due to chemical and physical parameterizations enhancing *pollutant production and mix". I suggest removing this conjecture for similar reasons provided above.*

Response to comment 24:

We thank the reviewer for the comment. We have removed the speculative attribution regarding CMAQ's overestimation and now report only the observed concentration differences. The revised sentences are as follows:
(Lines 682-685) CAMx reproduced the northeastward transport path but underestimated pollution intensity, failing to capture the strong core in Henan during the peak. CMAQ systematically overestimated concentrations.

Comment 25:

Lines 674-676: "In terms of magnitude, CAMx underestimated and CMAQ overestimated concentrations, while the EPICC-Model produced comparatively robust results." Except for Zhengzhou, all 3 models look to track very similarly. Perhaps make the quoted statement specific to Zhengzhou and then say all models otherwise performed similarly.

Response to comment 25:

We thank the reviewer for the comment. We have revised the sentence to make the statement specific to Zhengzhou and added that all three models otherwise tracked concentrations similarly in the other cities. The revised sentences are as follows:
(Lines 697-700) In terms of magnitude, CAMx underestimated and CMAQ overestimated concentrations in Zhengzhou, while the EPICC-Model produced comparatively robust results. In the other three cities, all models tracked concentrations similarly.

Comment 26:

Lines 694-696: "In contrast, CAMx and CMAQ systematically underestimated pollution intensity, particularly during the peak on June 6 and the initial clearing stage on June 10." From Figure 11, the contoured predictions match the color of the measured dots better in lower ozone areas on June 6 and 10, especially in central China. It looks to me that EMICC-Model tends to overpredict ozone in those areas.

Response to comment 26:

We thank the reviewer for the comment. We have revised the description to clarify that CAMx and CMAQ tended to underestimate ozone mainly in high-concentration areas, while predictions in lower-concentration regions were generally consistent with observations. We also noted that the EPICC-Model slightly overpredicted ozone in some low-concentration areas. The revised text avoids absolute statements and provides a more nuanced, observation-based comparison. The revised sentences are as follows:
(Lines 718-727) CAMx and CMAQ tended to underestimate pollution intensity in high-concentration areas, particularly during the peak on June 6 and the initial clearing stage on June 10, and did not fully reproduce the strong pollution exceeding 160 μg m^{-3} across

Henan Province. In lower ozone areas, especially in central China, their predictions were generally consistent with the observed values, while the EPICC-Model slightly overpredicted ozone in some of these regions.

Comment 27:

Lines 697-699: "This discrepancy may result from insufficient temporal resolution of emission inventories or limited photochemical sensitivity to precursors." Is this possible? I thought all 3 models were given the same emissions. Photochemical sensitivity should not be substantially different between the various CB chemistry mechanisms among the models. Again, I suggest removing such conjectures without solid evidence from process analyses. Figure 12 does not show much difference between the models except in Guangzhou where CMAQ seemed to perform best and the other 2 models under predicted substantially.

Response to comment 27:

We thank the reviewer for pointing this out. You are correct that all three models were driven by the same emission inventories, and the photochemical sensitivity among the CB chemistry mechanisms is not expected to differ substantially. Therefore, we agree that the previous statement suggesting possible causes ("insufficient temporal resolution of emission inventories or limited photochemical sensitivity") is speculative. We have removed this sentence from the revised manuscript to avoid unwarranted conjecture.

Comment 28:

Lines 769-770: "The EPICC-Model performs exceptionally well in identifying PM2.5 light to moderate pollution levels and MDA8 O3 general pollution." I suggest removing the subjective word "exceptionally" without a numerical example to provide evidence for it. See also my comments above for Lines 632-633 with reference to the small 2-7% differences in hit rate.

Response to comment 28:

We thank the reviewer for the comment. We agree that the term "exceptionally" is subjective and may be misleading without quantitative support. In the revised manuscript, we have removed this term and now describe the EPICC-Model's performance based on numerical results. Specifically, the EPICC-Model achieved a hit rate of 45.6% for MDA8 O_3 > 160 μg m^{-3}, which is higher than CAMx (43.0%) and CMAQ (38.4%), and it shows relatively balanced false alarm rates across all pollution types. This factual description conveys the model's performance without subjective wording. The revised sentences are as follows:
(Lines 797-808) The EPICC-Model shows higher hit rates in identifying $PM_{2.5}$ light to moderate pollution levels and MDA8 O_3 general pollution, achieving 45.6% for MDA8 O_3 > 160 μg m^{-3}, compared with 43.0% for CAMx and 38.4% for CMAQ. The forecast accuracy of all three models for single extreme pollution events was less than 2%, indicating that the response capability of existing models to sudden pollution events

still needs further improvement. In addition, the EPICC-Model demonstrated relatively balanced false alarm rates across all pollution types, suggesting a simulation capability comparable to other mainstream models and particular effectiveness in capturing light to moderate pollution and O_3 pollution.

Technical Corrections:

Be sure to define all acronyms at the time of their first occurrence. There are many undefined acronyms throughout the text.
Line 25: Remove the unneeded word "the" in the phrase "the EPICC-Model".
Line 26 (and elsewhere): What is meant by "unified" in this context? Either use a more appropriate adjective or remove entirely.
Line 28: The meaning of "high spatial consistency" is not clear in this limited context, perhaps simply replace with "high spatial consistency with measurements" if that is what is meant.
Line 34: Perhaps a better word for "identifying" is "forecasting".
Line 36: Suggest removing "compound" because it is unclear without additional context.
Line 65: Replace "ENVIRON Corporation" with "Ramboll" as the former company no longer exists.
Line 129: Replace "(Anon, 2020)" with "(Ramboll, 2020)" and update in the reference list.
Line 132: Replace "(Anon, 2021)" with "(EPA, 2021)" and update in the reference list.
Line 348: Replace "错误!未找到引用源"with English. I think it is a reference to Table 3.
Line 558-560: "While the EPICC-Model demonstrates relatively lower consistency in species simulation, yet maintains more stable performance in terms of RMSE and NMB." I suggest removing either the word "While" or "yet" to make the syntax correct.

Response to Technical Corrections:

We thank the reviewer for the detailed technical comments. All suggested corrections have been carefully addressed in the revised manuscript. Specifically, all acronyms are now defined at first occurrence, unnecessary words and unclear terms such as "the" and "unified" have been removed, wording has been revised for clarity (e.g., "identifying" → "forecasting", "high spatial consistency" → "high spatial consistency with measurements"), outdated company names and references have been updated (e.g., ENVIRON → Ramboll; Anon → Ramboll/EPA), the incorrect placeholder in Line 348 has been replaced with the correct English Table 3 reference, and the syntax issue in Lines 558–560 has been corrected. These revisions have been implemented throughout the manuscript.