Responses to Reviewer 2

This manuscript provides a comparison of simulated regional PM2.5 and MDA8O3 fields generated by the Emission and Atmospheric Processes Integrated and Coupled Community Model (EPICC-Model) with results from the CAMx and CMAQ modeling systems. Operational model evaluation statistics for all three models, benchmarked against a set of surface observations, are presented to facilitate an assessment of their performance across average spatial and temporal scales. Comparison of EPICC performance results alongside those from established and widely used models such as CAMx and CMAQ is a primary motivation for this study. As expected, statistical metrics averaged over broad spatial (domain-wide) and temporal (seasonal) scales reveal comparable performance among the models, which are fundamentally based on similar process representations.

Establishing and documenting this baseline for EPICC performance relative to other models is valuable for prospective users – thus, a manuscript of this nature will be of interest to the journal readership. Equally critical for the adoption and advancement of a modeling system is the explicit demonstration of its unique strengths and a clear articulation of how its formulation and structural features support its intended applications – an aspect of the manuscript that could be improved further. This could be accomplished by enhancing the clarity of diagnostic evaluation for the EPICC model, thereby highlighting the influence of distinct attributes such as process representation, numerical schemes, and the integration with meteorology and emissions, on the simulation of PM2.5 and O3, relative to CAMx and CMAQ. The following comments and suggestions are offered that may help improve the uniqueness and usefulness of the manuscript.

Response to comment 1:

We thank the reviewer for the careful and constructive review of our manuscript, and the point-to-point responses is present as follows.

Comment 1:

Throughout the manuscript the EPICC model is referred to as "independently developed", but it is not apparent to me what that implies? Was the code development independent and if so, independent of what? Is the science in the model not based on other experiences and knowledge accumulated over decades of international research? Looking at the EPICC code, the overall structure of the model bears resemblance to other contemporary systems including CAMx and CMAQ, share similar names for routines, and often also borrows process routines and numerical approaches. Though the individual routines acknowledge derivative work and include some history of the code development, given the similarities between many aspects of the EPICC model structure and code with other existing open-source models, some clarity on what "independently developed" is intended to convey would be useful for the readers.

Response to comment 1:

We thank the reviewer for this important comment. We agree that the original phrase "independently developed" was ambiguous and could be interpreted as implying that the EPICC-Model was developed independently of established modeling systems and scientific knowledge. This was not our intended meaning. Here, "independent" refers specifically to the development of the modular model framework, plug-in architecture, and system integration, rather than to the independent origin of all process routines and numerical algorithms. We have therefore revised the description throughout the manuscript and explicitly identified the contributions of CAMx and CMAQ. The revised text distinguishes the routines inherited or adapted from these models from the framework development and chemistry integration conducted by the EPICC team. The revised text is as follows:

(Lines 129-136) *The EPICC-Model is a three-dimensional tropospheric CTM developed in China with an independent modular framework (EPICC-Model Working Group, 2026). Several major process-level routines and numerical algorithms were inherited or adapted from CAMx and CMAQ, whose foundational contributions are explicitly acknowledged. Building on these foundations, the EPICC development team designed a modular and plug-in architecture and independently integrated several specialized chemistry treatments, including sulfate formation, heterogeneous HONO production, and* N_2O_5 *hydrolysis.*

Comment 2:

L31: "maximum biases" should perhaps be "maximum bias".

Response to comment 2:

We agree and have corrected "maximum biases" to "maximum bias" in the revised manuscript.

Comment 3:

L40: some indication of what specific aspect of boundary layer representation improvement is suggested would be useful. As written, the statement is too broad and does not convey much.

Response to comment 3:

We agree that the original description was too broad. We have revised the sentence to identify vertical turbulent mixing under stable boundary-layer conditions as the specific aspect requiring further evaluation, without attributing the model biases to a particular boundary-layer process in the absence of direct observational evidence. The revised sentence is as follows:

(Lines 37-40) *Future improvements should focus on refining secondary aerosol chemistry and emission inventories and further evaluating the representation of vertical turbulent mixing under stable boundary-layer conditions.*

Comment 4:

L47: "As a typical secondary pollutant, PM2.5 is influenced by primary emissions" is contradictory and awkwardly stated. The sentence should be rewritten to convey that ambient PM2.5 is composed of constituents that have both primary emissions and are *secondarily produced.*

Response to comment 4:

The original sentence conflated $PM_{2.5}$ as a whole with its secondary components. We have revised the text to distinguish directly emitted primary particles from secondary components formed through atmospheric transformation. The revised text is as follows:

(Lines 47-50) *Specifically, $PM_{2.5}$ comprises both directly emitted particles and secondary components formed from SO_2, NOx, NH_3, and VOCs, with its levels also affected by regional transport and meteorology (Zhang et al., 2012; Jing et al., 2020; Zhai et al., 2021; Bao et al., 2023).*

Comment 5:

L48-49: There are more seminal references that establish these underlying processes regulating $PM_{2.5}$ accumulation and distributions in the troposphere and should probably be included.

Response to comment 5:

We appreciate the reviewer's suggestion. To strengthen the scientific basis for the processes governing $PM_{2.5}$ accumulation and distribution, we have supplemented the cited literature with recent studies addressing the increasingly important role of particulate nitrate in haze pollution over China. Zhai et al. (2021) demonstrated that particulate nitrate has become an increasingly important component of wintertime $PM_{2.5}$ pollution in China and explained its persistence in terms of changes in aerosol thermodynamic partitioning and deposition. More recently, Bao et al. (2023) provided observation-based evidence that rapid nitrate formation can directly contribute to the development of severe haze, with daytime gas-phase oxidation and nighttime aqueous-phase processing both contributing to nitrate production. These recent studies complement the earlier mechanistic evidence concerning secondary aerosol formation and provide a more up-to-date basis for the relevant statement in the Introduction. The references are as follows:

[1] Zhai, S., Jacob, D. J., Wang, X., Liu, Z., Wen, T., Shah, V., Li, K., Moch, J. M., Bates, K. H., Song, S., Shen, L., Zhang, Y., Luo, G., Yu, F., Sun, Y., Wang, L., Qi, M., Tao, J., Gui, K., Xu, H., Zhang, Q., Zhao, T., Wang, Y., Lee, H. C., Choi, H., and Liao, H.: Control of particulate nitrate air pollution in China, Nature Geoscience, 14, 389-395, https://doi.org/10.1038/s41561-021-00726-z, 2021.

[2] Bao, Z., Zhang, X., Li, Q., Zhou, J., Shi, G., Zhou, L., Yang, F., Xie, S., Zhang, D., Zhai, C., Li, Z., Peng, C., and Chen, Y.: Measurement report: Intensive biomass burning emissions and rapid nitrate formation drive severe haze formation in the Sichuan Basin, China – insights from aerosol mass spectrometry, *Atmospheric Chemistry and Physics*, 23, 1147–1167, https://doi.org/10.5194/acp-23-1147-2023, 2023.

Comment 6:

L53-54: Again, the dependence of O3 formation on NOx and VOC and on meteorological factors has been well established dating back several decades (e.g., Hagen Smit and Fox, 1956) and should be reflected by the citations.

Response to comment 6:

Thank the reviewer for this valuable suggestion. We agree that the fundamental dependence of O_3 formation on NO_x and VOCs was established several decades ago and should be supported by foundational references. Accordingly, we have added Haagen-Smit and Fox (1956), who demonstrated the photochemical formation of O_3 from organic compounds in the presence of nitrogen oxides, and Sillman (1999), who systematically described the nonlinear O_3–NO_x–VOC relationship and the distinction between NO_x-sensitive and VOC-sensitive regimes. We have retained the more recent studies to support the effects of meteorological factors under contemporary conditions in China. The relevant sentence and reference list have been revised accordingly. The revised text reads as follows:

(Lines 52-58) *In contrast, O_3 is formed through photochemical reactions involving NO_x and VOCs, and its production exhibits nonlinear responses to precursor levels, characterized by NO_x-sensitive and VOC-sensitive regimes (Haagen-Smit and Fox, 1956; Sillman, 1999). Ambient O_3 levels are also strongly influenced by meteorological factors, including solar radiation, temperature, humidity, and boundary-layer dynamics (Li et al., 2019b; Gao et al., 2022).*

Comment 7:

L54: This statement (use of "despite") implies some dependence on O3 and PM levels which may not always occur because of different trends in their precursors. Perhaps the statement should just convey that in many Chinese urban clusters PM2.5 levels have reduced but MDA8O3 has increased.

Response to comment 7:

The reviewer correctly notes that the use of “despite” may imply a direct relationship between the trends in $PM_{2.5}$ and O_3, although the two pollutants have different precursors and formation mechanisms. To avoid this unintended implication, the sentence has been revised to present the observed changes in $PM_{2.5}$ and MDA8 O_3 as two parallel but independent trends. The revised text reads as follows:

(Lines 58-60) *In many Chinese urban clusters, $PM_{2.5}$ concentrations have decreased, whereas warm-season maximum daily 8 h average (MDA8) O_3 concentrations increased at a rate of 1.2 ± 1.3 ppb yr^{-1} during 2013–2022 (Wang et al., 2024).*

Comment 8:

L64-68: References for these models should be provided here.

Response to comment 8:

Following the reviewer's suggestion, appropriate references have now been provided for the representative chemical transport models introduced in this paragraph. Specifically, references have been added for CMAQ (Byun and Schere, 2006; Foley et al., 2010; Appel et al., 2017, 2021), CAMx (Ramboll, 2024), WRF-Chem (Grell et al., 2005), and GEOS-Chem (Eastham et al., 2018; The International GEOS-Chem User Community, 2025). These additions provide readers with the relevant model descriptions and development histories. The corresponding paragraph and reference list have been updated in the revised manuscript. The revised text reads as follows:

(Lines 66-73) *Representative CTMs include the Community Multiscale Air Quality Modeling System (CMAQ) developed by the U.S. Environmental Protection Agency (EPA) (Byun and Schere, 2006; Foley et al., 2010; Appel et al., 2017, 2021), the Comprehensive Air Quality Model with Extensions (CAMx) developed by Ramboll (Ramboll, 2024), WRF-Chem for regional-scale simulations (Grell et al., 2005), and GEOS-Chem, which is widely applied for global-scale studies (Eastham et al., 2018; The International GEOS-Chem User Community, 2025).*

Comment 9:

L83: Structural differences in context of models is often used to describe differences in model structure - architecture, code, design, and not necessarily process differences. Suggest changing "structural differences" to "differences".

Response to comment 9:

The reviewer raises an important point regarding the terminology used in the model context. In this manuscript, "structural differences" was intended to refer broadly to differences in process representations and model configurations rather than to differences in model architecture, code structure, or software design. To avoid ambiguity, we have replaced "structural differences" with "differences" in the revised manuscript. The revised text reads as follows:

(Lines 86-89) *Multi-model inter-comparison not only helps reveal differences among models in pollutant transport, chemical reactions, and meteorological feedbacks, but also serves as a key approach to quantifying simulation uncertainties and improving model robustness and reliability (Carmichael et al., 2008).*

Comment 10:

Table 1: Though based on Fast-J, Binkowski et al. (2007; https://doi.org/10.1175/JAM2531.1) is perhaps a more appropriate reference for the photolysis approach in CMAQ.

Response to comment 10:

The reference for the CMAQ photolysis treatment in Table 1 has been revised as recommended by the reviewer. Binkowski et al. (2007; https://doi.org/10.1175/JAM2531.1) provides a direct description and evaluation of the online photolysis module implemented in CMAQ and is therefore more appropriate for the information presented in the table. The previous citation has been replaced, and the corresponding full reference has been added to the reference list.

Comment 11:

L139-140: To my knowledge there is nothing that specifies 14 layers for CMAQ or CAMx. Why did the authors also not configure CMAQ and CAMx with the so-called "superior" 20-layer configuration or the default WRF layer structure?

Response to comment 11:

We appreciate the reviewer's important comment regarding the vertical configurations. We agree that neither CMAQ nor CAMx requires the use of 14 vertical layers. In this study, "operational configurations" refers to the model setups actually implemented in the respective routine air-quality forecasting systems during the study period. These configurations had been established for routine forecasting by considering computational cost, forecast timeliness, numerical stability, and compatibility with the existing preprocessing and forecasting workflows. They were not intended to represent the default or theoretically optimal vertical configurations of CMAQ or CAMx.
The objective of this study was to evaluate the overall performance of the three forecasting systems under their actual operational setups, rather than to conduct a controlled sensitivity experiment in which the vertical-layer configuration was standardized. Reconfiguring CMAQ and CAMx with 20 layers or retaining the complete native WRF vertical structure would have produced configurations different from those used operationally and would also have required corresponding changes to emissions processing, initial and boundary conditions, and other model inputs.
In addition, the EPICC-Model uses a terrain-following height-based coordinate, whereas the CMAQ and CAMx configurations use terrain-following pressure-based coordinates. Therefore, using the same number of layers would not necessarily result in equivalent layer interfaces or vertical resolution among the three models. We have revised the manuscript to clarify that the 14-layer configurations are study-specific operational choices rather than model requirements, and we have removed the implication that the 20-layer configuration represents an inherent advantage of the EPICC-Model. The revised text reads as follows:

(Lines 149-151) *In the operational configurations adopted in this study, the EPICC-Model employs 20σ_z layers layers, whereas CMAQ and CAMx use 14σ_p layers.*

Comment 12:

L142-143: Though not employed in the current study, it should be noted that for consistency, CMAQ applications typically employ WRF fields also simulated using ACM in WRF.

Response to comment 12:

We appreciate the reviewer for highlighting this relevant aspect of WRF-CMAQ coupling. To provide a more complete description of the CMAQ configuration, we have added a brief note that WRF meteorological fields generated with the ACM2 scheme are commonly used in CMAQ applications to maintain consistency in boundary-layer treatment. The revised text reads as follows:

(Lines 152-155) *CMAQ uses the Asymmetric Convective Model version 2 (ACM2) for vertical diffusion. WRF meteorological fields generated with the ACM2 scheme are commonly used in CMAQ applications to maintain consistency in boundary-layer treatment.*

Comment 13:

L151-153: Given that models are continuously evolving, it is important to qualify these statements with the specific version of the models used in the inter-comparison. Though the CMAQ version used here may have been configured with CB6r3, newer versions do include CB6r5 and variants.

Response to comment 13:

We appreciate the reviewer's reminder that descriptions of chemical mechanisms should be qualified by model version. We have therefore specified that CB6r3 refers to the CMAQ v5.3.3 configuration used in this study and noted that newer CMAQ versions include CB6r5 and its variants. The revised sentences are as follows:
(Lines 174-179) *The CMAQ v5.3.3 configuration used in this study employs the CB6r3 mechanism in combination with the AE7/VBS module to represent SOA formation, and from version 5.2 onward, incorporates additional parameterizations such as Potential Combustion SOA (PCSOA) to compensate for combustion-related SVOC and IVOC emissions not captured in current inventories (Murphy et al., 2017). Newer CMAQ versions include CB6r5 and its variants.*

Comment 14:

L164-166: Default LBC profiles for CMAQ are only provided to facilitate model setup and initial testing. By definition, the LBCs are dependent on the geographical extent and location of the limited area model application. Thus, it is a bit surprising that the authors did not consider harmonizing the LBCs across the 3 modeling systems. Acknowledgement of this somewhat fundamental difference in configuration setting should be provided.

Response to comment 14:

We appreciate the reviewer's careful observation. We agree that the term "default boundary conditions" was inappropriate because the lateral boundary conditions (LBCs) supplied with CMAQ are intended primarily to facilitate model setup and testing and are not necessarily representative of a specific regional modeling domain.

In the configurations evaluated here, the initial and lateral boundary conditions were not harmonized across the three modeling systems. The EPICC-Model used fields derived from the Model for Ozone and Related Chemical Tracers, version 4 (MOZART-4), whereas the CMAQ boundary conditions were generated from a CB6r3 profile file, and the CAMx initial and boundary conditions were prepared using the icbcprep procedure with the corresponding profile configuration. These settings were retained because the study was designed to evaluate the three modeling systems under the configurations used in their respective forecasting applications. The relevant profile files have been provided in the Zenodo repository.

We agree that this represents a fundamental difference in the model configurations. Because no sensitivity simulations with harmonized initial and lateral boundary conditions were conducted, their contributions to the inter-model differences cannot be quantified or excluded. In particular, differences in the boundary fields may influence simulated regional background concentrations and areas affected by inflow, especially for relatively long-lived species such as O_3. Therefore, the model-performance differences discussed in the manuscript should be interpreted as the combined effects of internal model formulations and configuration differences, including the initial and lateral boundary conditions, rather than being attributed exclusively to differences in chemical or physical process representations. This limitation has now been explicitly acknowledged in the revised manuscript. The revised text reads as follows:

(Lines 185-196) *Regarding boundary conditions, the EPICC-Model incorporates MOZART outputs as the initial and lateral background fields, while CMAQ and CAMx use their respective boundary-condition preprocessing approaches. Specifically, CMAQ boundary conditions were generated based on the CB6r3 profile file, and CAMx initial and boundary conditions were prepared using the icbcprep procedure with the corresponding profile configuration, which can be found in the Zenodo repository (https://doi.org/10.5281/zenodo.17055755). These initial and lateral boundary conditions were not harmonized across the three models, and their individual influence was not isolated through sensitivity simulations. Consequently, differences among the model results may reflect the combined effects of internal model formulations and initial and lateral boundary conditions, particularly for regional background concentrations and areas influenced by boundary inflow.*

Comment 15:

L265: Is the Kz_min setting too low relative to other models or what is physically expected - if the latter what is the basis to support the value? What fraction of the time during the January-May period are stable conditions encountered?.

Response to comment 15:

The original statement was intended to propose a possible explanation rather than establish a causal relationship between Kz_min and the intermittent $PM_{2.5}$ overestimation. This hypothesis was motivated by the CMAQ sensitivity experiments of Kim et al. (2024), which demonstrated that surface $PM_{2.5}$ concentrations over China can respond substantially to the prescribed minimum vertical eddy diffusivity. The phrase "set too low", however, could be interpreted as suggesting that 0.01 $m^2\ s^{-1}$ is below a universally accepted physical threshold, which was not the intended meaning. The text has therefore been revised to retain the hypothesis while avoiding this implication. Specifically, the revised manuscript states that the intermittent overestimation may be associated with the sensitivity of near-surface to the Kz_min treatment in ACM2. Because no quantitative attribution to stable-condition frequency is made in the revised text, the January–May frequency i $PM_{2.5}$s not used as evidence for this interpretation. The revised sentences are as follows:

(Lines 294-298) *In addition, CMAQ exhibited intermittent overestimations during several periods, which may partly reflect the sensitivity of near-surface $PM_{2.5}$ concentrations to the treatment of minimum vertical eddy diffusivity (Kz_min) in ACM2, as indicated by previous CMAQ sensitivity experiments over China (Kim et al., 2024).*

Comment 16:

L278-279: This sentence is vague as it covers everything possible and thus not insightful – which process may be the dominant in driving the noted overestimation?

Response to comment 16:

The reviewer correctly notes that the original description was too broad to identify the process most relevant to the CMAQ overestimation. The model-component evaluation presented in Figure 6 provides more specific evidence: CMAQ simulated substantially higher OC concentrations than both the observations and the other two models, particularly during the polluted periods. This positive OC bias directly increased the simulated total $PM_{2.5}$ concentration and is therefore considered the most likely dominant contributor to the CMAQ overestimation discussed in this paragraph. Differences in vertical mixing may additionally affect the accumulation of near-surface pollutants, but the available results identify the elevated OC concentration more directly than the other candidate processes. The sentence has been revised accordingly. The revised sentences are as follows:

(Lines 306-314) *During the heating season, observed $PM_{2.5}$ concentrations increased substantially and exhibited greater temporal variability. CAMx maintained relatively good agreement with the observations, whereas the positive bias of CMAQ became more pronounced. The component evaluation in Figure 6 indicates that the substantially elevated OC concentration simulated by CMAQ was the most likely dominant contributor to this $PM_{2.5}$ overestimation, while differences in vertical mixing may have additionally influenced near-surface pollutant accumulation. The performance of the EPICC-Model was intermediate between the two.*

Comment 17:

Figure 1 caption: please clarify if daily PM2.5 implies daily-average PM2.5?

Response to comment 17:

Following this suggestion, the term "daily $PM_{2.5}$" in the caption of Figure1 has been clarified as "daily $PM_{2.5}$ (i.e., daily mean $PM_{2.5}$ concentrations)." The values shown in the time series were calculated by averaging the hourly $PM_{2.5}$ concentrations over each calendar day.

Comment 18:

L296-297: please clarify what "lower background concentrations" imply. Is the background low or is it that ventilation is strong in the summer and thus concentrations are lower?

Response to comment 18:

The phrase "lower background concentrations" was not intended to indicate a distinct or independently quantified background component. It referred to the generally lower ambient $PM_{2.5}$ concentrations during spring and summer. As discussed earlier in this section, the lower summertime concentrations were associated mainly with increased boundary-layer height, enhanced turbulent ventilation, and more frequent precipitation, which promoted pollutant dilution and wet removal. To avoid ambiguity, "lower background concentration conditions" has been replaced with "generally lower ambient $PM_{2.5}$ concentrations during spring and summer." The revised sentences are as follows:

(Lines 329-333) *Its simulations during spring and summer are more consistent with CAMx, indicating similar performance under the generally lower ambient $PM_{2.5}$ concentrations during these seasons, particularly in summer when enhanced boundary-layer ventilation and more frequent precipitation favor pollutant dilution and wet removal.*

Comment 19:

L299-301: Based on Figure 1 alone it is difficult to say whether compared to the other two, one model does better in striking better balance between accurately simulating extreme pollution and sustaining strong annual performance, since these are regional average values. Would the same conclusion also hold at individual locations?

Response to comment 19:

The statement was intended to describe the national-scale aggregate performance shown in Figure 1, rather than uniform behavior at every monitoring site. At this scale, the EPICC-Model reproduced the seasonal transition of $PM_{2.5}$ and responded to high-concentration episodes while maintaining reasonably strong annual statistics. Site-level performance inevitably varies with local emissions, meteorology, topography, and monitoring-site representativeness. The spatial evaluation presented later in the manuscript provides additional information on regional variations, but it does not imply identical performance at all individual sites. The original comparative expression "a better balance" has therefore been replaced with a scale-specific and more objective description. The revised sentences are as follows:

(Lines 335-338) *At the national aggregate scale represented in Figure 1, the EPICC-Model reproduced the seasonal transition of $PM_{2.5}$ concentrations and responded to high-concentration pollution episodes while maintaining reasonably strong annual statistical performance.*

Comment 20:

L309: Reduced reaction activity during the cooler months arises from both decreased photolysis as well as lower reaction rates associated with thermal chemistry.

Response to comment 20:

This distinction has been incorporated into the revised manuscript. The discussion now attributes the reduced photochemical activity during the colder months to both weaker photolysis under reduced solar radiation and slower temperature-dependent thermochemical reaction rates. The revised sentences are as follows:

(Lines 344-347) …*while O_3 concentrations remain relatively low during the colder months because weaker solar radiation suppresses photolysis and lower temperatures reduce thermochemical reaction rates, collectively weakening photochemical activity (Wang et al., 2019).*

Comment 21:

Do the scatter plots in Figure 2b show the same data as in 2a, i.e., average across all sites or is each point the MDA8O3 at individual observation location?.

Response to comment 21:

The scatterplots in Figure 2b use the same nationally averaged daily data as the time series in Figure 2a. Each point represents one day in 2021: the x-coordinate is the observed MDA8 O_3 concentration averaged across the 1,644 monitoring sites, and the y-coordinate is the corresponding simulated concentration averaged across the model grid cells associated with these sites. Thus, the points do not represent individual monitoring locations.

Comment 22:

L317-319: References should be provided for the heterogeneous HONO formation and nitrate photolysis pathways. To my knowledge, though uncertain, the nitrate photolysis primarily occurs in marine environments (e.g., Ye et al., 2017; Reed et al 2017; Kasibhatla et al., 2018; Shah et al., 2023; Sarwar et al., 2024; Sarwar et al., 2025) or environments impacted by marine air masses, since halide ions are highly surface active and promote the surface propensity of coexisting nitrate anions (DOI: 10.1039/d1ea00087j). How then does it influence inland environments, especially since *the large scale O3 in EPICC is expected to be modulated by the MOZART LBCs? Did the MOZART simulations also include nitrate photolysis? How different are the OH concentrations across the 3 models?*

Response to comment 22:

We appreciate the reviewer's thoughtful comment on the applicability of these nitrous acid (HONO)-related pathways. To clarify their relevance to inland environments and their implementation in the EPICC-Model, we have supplemented the discussion with both regional evidence from China and recent EPICC-specific sensitivity analyses.

Previous observational and modeling studies conducted in Beijing, the North China Plain, and other inland regions of China have demonstrated that heterogeneous HONO formation and particulate nitrate photolysis can contribute to radical production and O_3 formation under inland atmospheric conditions (Bao et al., 2018; Zhang et al., 2022; Li et al., 2024). More directly, sensitivity experiments reported by the EPICC Model Working Group (2026) showed that the HONO-related heterogeneous chemistry implemented in the EPICC-Model, including particulate nitrate photolysis, enhances simulated O_3 and contributes to improved O_3 simulations over China. Zhang et al. (2025) further evaluated different HONO sources in the EPICC-Model and found that heterogeneous reactions on aerosol and ground surfaces were major contributors to the uncertainty in simulated HONO, while particulate nitrate photolysis made an important contribution during daytime. These studies provide direct model-specific support for the relevance of the included HONO pathways.

Regarding the reviewer's question about the Model for Ozone and Related Chemical Tracers, version 4 (MOZART-4), these fields were used to provide initial and lateral boundary concentrations. They do not contain the particulate nitrate photolysis parameterization implemented within the EPICC-Model. However, because the initial and lateral boundary conditions were not harmonized among the three models, their possible influence on regional background concentrations cannot be excluded or separated from that of the internal chemical mechanisms in the present intercomparison.

In response to the question concerning hydroxyl radicals (OH), the wording has also been revised to describe enhanced radical production as a mechanism supported by previous process and sensitivity studies, rather than implying that OH concentrations were directly compared among the three models. The revised manuscript text reads as follows:

(Lines 354-361) *The improved representation of high-O_3 episodes by the EPICC-Model may be partly related to its inclusion of heterogeneous HONO formation and particulate nitrate photolysis, which can enhance radical production and O_3 formation in inland China (Bao et al., 2018; Zhang et al., 2022; Li et al., 2024). Recent EPICC-specific sensitivity and source-diagnosis studies further demonstrate that these HONO pathways, including daytime particulate nitrate photolysis, contribute to simulated HONO and O3 formation and improve the representation of O3 over China (Zhang et al., 2025; EPICC Model Working Group, 2026).*

Comment 23:

L325: Is the isoprene chemistry representation in cb6r3 and cb6r5 that different to impact the ozone forming potential as speculated? It would be useful to further discuss the differences in the isoprene chemistry leading to the noted O3 differences.

Response to comment 23:

We appreciate the reviewer's careful observation regarding the differences between CB6r3 and CB6r5. Further examination of the mechanism updates indicates that CB6r5 primarily revises selected reaction rate constants, photolysis rates, and product yields relative to CB6r3, rather than introducing a major restructuring of isoprene oxidation chemistry. Available sensitivity evaluations also indicate that the overall effects of the CB6r5 updates on simulated O_3 are generally small and spatially variable (Yarwood et al., 2020). Therefore, the difference in isoprene chemistry alone does not provide sufficient support for the original interpretation. The corresponding statement attributing the higher O_3 formation potential of the EPICC-Model to a more comprehensive representation of isoprene chemistry in CB6r5 has been removed.

Comment 24:

L328: An overly deep PBL should result in more dilution and systematic low (negative) bias as opposed to positive bias. I am not sure I understand the rationale here?

Response to comment 24:

We appreciate the reviewer's insightful comment, which helps us clarify the effects of boundary-layer mixing on surface O_3. A deeper boundary layer generally enhances the dilution of near-surface pollutants and therefore cannot, by itself, explain a positive O_3 bias. Vertical mixing can nevertheless affect surface O_3 through several competing processes, including dilution, downward transport of O_3 from aloft, and changes in near-surface NO titration. We have therefore revised the sentence to avoid linking the positive bias directly to an excessively deep boundary layer, while retaining vertical redistribution and photochemical production as possible contributing processes. The revised sentences are as follows:

(Lines 361-364) *At lower O_3 concentration levels, CMAQ reproduced the general temporal variation but exhibited a systematic positive bias, which may reflect the combined influence of vertical redistribution and photochemical production.*

Comment 25:

L336: "well-controlled absolute bias despite its systematic underestimation" is confusing and somewhat contradictory. If the model is systematically low biased, how can its absolute bias be well controlled?

Response to comment 25:

We appreciate the reviewer's careful observation regarding the terminology. The phrase "absolute bias was well controlled" was intended to indicate that CAMx had a relatively small overall error magnitude, as reflected by its RMSE of 9.15 μg m^{-3}, despite its modest negative mean bias (NMB = −5.34%). To avoid suggesting that the negative bias and absolute bias are contradictory measures, we have replaced the qualitative expression with the corresponding statistical values. The revised sentence now distinguishes temporal correlation, mean bias, and overall error magnitude explicitly. The revised sentences are as follows:

(Lines 368-370) *Statistical analysis indicates that CAMx achieved the highest temporal correlation (R = 0.94) and the lowest RMSE (9.15 μg m^{-3}), while exhibiting a modest negative mean bias (NMB = −5.34%).*

Comment 26:

L345-346: How does one quantify the better balance between overestimation and underestimation. Based on the analyses presented, it is hard to conclude this in any compelling manner. At best the performance is comparable across 3 models with one model performing better than the other two under certain conditions.

Response to comment 26:

The expression "a better balance" was intended to refer to the smaller annual mean bias of the EPICC-Model, but it was not defined by an independent quantitative metric and could therefore overstate the comparison. We have replaced this expression with a direct comparison of the reported statistics. The revised text indicates that the three models show comparable overall performance, while exhibiting different strengths: the EPICC-Model has the smallest absolute NMB, CAMx has the highest correlation and lowest RMSE, and all three models have FAC2 values close to 1. The revised sentences are as follows:

(Lines 374-379) *The scatterplots in Figure 2b show that all three models have FAC2 values close to 1, indicating comparable overall performance. The models nevertheless exhibit different statistical characteristics: CMAQ tends to overestimate MDA8 O_3, CAMx achieves the highest correlation and lowest RMSE despite a modest negative bias, and the EPICC-Model has the smallest absolute NMB.*

Comment 27:

L404: It is not apparent to me what "multi-centered and scattered distribution pattern" implies? Is there a specific physical process that drives the Springtime distribution? This discussion would benefit from additional clarification.

Response to comment 27:

We appreciate the reviewer's request for clarification. The phrase "multi-center and dispersed distribution pattern" was intended to describe several geographically separated $PM_{2.5}$ concentration maxima, but it did not identify either their locations or the associated processes clearly. We have replaced this phrase with an explicit description of the major springtime high-concentration regions. The revised discussion attributes the pattern to the combined effects of northwesterly transport of dust from arid regions, local anthropogenic emissions over North China, and pollutant accumulation under relatively limited ventilation in basin regions. The revised sentences are as follows:

(Lines 390-394) *In spring, elevated $PM_{2.5}$ concentrations occurred in several geographically separated regions, including North China, the arid basins of Northwest China, and the Sichuan Basin. This spatial pattern reflects the combined influences of northwesterly dust transport, local anthropogenic emissions, and pollutant accumulation under relatively limited ventilation in basin regions.*

Comment 28:

L408: Please elaborate on the reasons why coarse vertical resolution reduces mixing? Also, what altitude O3 are the near surface precursors being mixed with? Why does the EPICC layer structure facilitate "better" mixing representation and under what conditions? What prevented CMAQ and CAMx from also being configured with the same layer structure (and implicitly same meteorology fields) as EPICC?

Response to comment 28:

We appreciate the reviewer's detailed comment. A coarser vertical grid does not necessarily reduce mixing, but it may affect the numerical representation of concentration gradients and exchange within the boundary layer, particularly under conditions with strong vertical stratification. The 20-layer configuration used by the EPICC-Model provides finer vertical resolution in the present simulations, but this is a configuration-specific feature rather than an inherent model advantage; CAMx and CMAQ can also employ alternative vertical-layer structures. We have therefore revised the text to avoid attributing the CAMx underestimation directly to its 14-layer configuration and instead refer more generally to the combined effects of chemical and vertical-transport representations. The revised sentences are as follows:

(Lines 441-444) *CAMx generally underestimated O_3 across China, which may reflect the combined effects of its gas-phase chemistry and representation of vertical transport (Ren et al., 2022; Tang et al., 2011).*

Comment 29:

L414-417: The authors suggest that regional O3 levels and differences between models are primarily controlled by photochemical production. How does one rule out the role of large-scale transport as represented by the differing boundary conditions? Also, given that the layer structure of the native meteorological fields and that of the 3 models is different, and also different amongst the CTMs, I do not see how it can be claimed that identical meteorological fields were used? The starting WRF output is the same *but it's translation to the grid structures of the individual CTMs (Table 1) is likely different, resulting in not "identical" meteorological fields driving the CTM calculations.*

Response to comment 29:

We appreciate the reviewer's careful distinction between common meteorological forcing and identical CTM-ready meteorological fields. The three models were driven by meteorological inputs derived from the same WRF simulation; however, these fields were processed according to the vertical structures and input requirements of each CTM and therefore should not be described as identical. We have corrected this wording accordingly. We also recognize that the present comparison does not exclude the influence of lateral boundary conditions and large-scale transport on regional O_3. The original statement has therefore been revised to describe the simulated O_3 distribution as reflecting the combined effects of regional photochemistry, meteorological transport, and lateral boundary conditions. The strong summertime photochemical environment remains relevant to the seasonal O_3 maximum, but it is no longer presented as the exclusive control on the inter-model differences. The revised sentences are as follows:

(Lines 449-453) *All three models were driven by meteorological inputs derived from the same WRF simulation but processed for their respective vertical structures. The simulated summer O_3 distributions therefore reflect the combined influences of regional photochemistry, meteorological transport, and lateral boundary conditions (Li et al., 2019a).*

Comment 30:

L417: Perhaps "nation-wide O3" is better usage than "national O3".

Response to comment 30:

We appreciate the reviewer's helpful wording suggestion. "National O_3" has been replaced with "nation-wide O_3 concentrations" to describe the spatial extent more clearly.

Comment 31:

L425: What shifts southward - peak values or just relative abundance? What processes drive this seasonal change in distribution and how does the ability of a model to capture it link back to its process formulation?

Response to comment 31:

We appreciate the reviewer's request for clarification. The phrase "shifted southward" refers to a southward displacement of the relative seasonal O_3 concentration maximum, rather than the southward transport of a specific O_3 plume or an increase in the relative abundance of O_3.

During winter, the stronger reduction in solar radiation and temperature over northern China suppresses photochemical O_3 production. At the same time, relatively high NO_x emissions favor stronger near-surface NO titration, particularly under weak boundary-layer mixing. In contrast, southern coastal China retains comparatively stronger solar radiation and photochemical activity, causing the relative seasonal O_3 maximum to occur farther south.

We have also re-examined the model distributions in Figure 4 and corrected our description of the CMAQ bias. The most evident wintertime O_3 overestimation by CMAQ occurs primarily over the middle and lower reaches of the Yangtze River and the Sichuan Basin, rather than over northern China as stated previously. The EPICC-Model reproduced the relative O_3 maximum over southern coastal China, whereas CAMx displaced this maximum farther southwest.

The ability of a model to reproduce this seasonal redistribution depends on the combined representation of photochemical production, NO titration, vertical and regional transport, and the effects of clouds on photolysis. Previous satellite observations have shown pronounced regional and seasonal variations in cloud optical depth over China, with southern China exhibiting a seasonal pattern different from that of northern China (Li et al., 2019). The CMAQ overestimation in the middle and lower reaches of the Yangtze River and the Sichuan Basin may therefore partly reflect uncertainties in the online representation of cloud effects on photolysis. However, because photolysis rates and cloud-related sensitivities were not evaluated separately in this study, this interpretation is presented as a possible contributing factor rather than a demonstrated attribution. The revised sentences are as follows:

(Lines 458-463) *In winter, nationwide O_3 concentrations decrease to their annual minimum, and the relative seasonal concentration maximum shifts toward southern coastal China. This redistribution reflects weaker photochemical production and stronger NO titration over northern China, whereas comparatively stronger solar radiation sustains greater photochemical activity in the south.*

(Lines 463-469) *The EPICC-Model reproduced the relative O_3 maximum over the southern coastal region, whereas CAMx displaced it farther southwest. CMAQ reproduced O_3 reasonably well over North China but overestimated concentrations mainly over the middle and lower reaches of the Yangtze River and the Sichuan Basin. These differences reflect the models' combined representations of photochemistry, NO titration, and atmospheric transport.*

Comment 32:

L430-431: Can just examination of a single IOA metric be regarded as "comprehensive evaluation"? Suggest replacing "comprehensive evaluation" with just "evaluation".

Response to comment 32:

We appreciate the reviewer's helpful suggestion. Because this paragraph evaluates the seasonal spatial performance primarily using the IOA metric, "comprehensive evaluation" has been replaced with "evaluation" to describe the scope of the analysis more precisely.

Comment 33:

L486: The underestimation of SO4 may also contribute to the overestimation of NO3 assuming NH3 emissions are well constrained..

Response to comment 33:

We appreciate the reviewer's helpful observation regarding the thermodynamic coupling among sulfate, nitrate, and ammonium. When NH_3 availability is sufficiently constrained, underestimated sulfate formation can reduce the amount of NH_3 consumed in ammonium sulfate formation, leaving more NH_3 available to partition with HNO_3 and form particulate NH_4NO_3. Sulfate underestimation may therefore contribute to nitrate overestimation in some regions. This mechanism has been added to the revised discussion. The revised sentences are as follows:

(Lines 529-531) *The reduced formation of ammonium sulfate may also leave more NH_3 available to partition with HNO_3, potentially enhancing NH_4NO_3 formation and contributing to the overestimation of particulate NO_3^- in some regions.*

Comment 34:

L510-511: Since one would expect artificial dilution of emissions over grid cell volumes to likely result in underestimation of concentrations of a primary species, does the BC overestimation imply overestimation of BC emissions or could the representativeness of the monitor locations also influence the noted overestimation?

Response to comment 34:

We thank the reviewer for this important comment. We agree that grid-cell averaging tends to smooth near-source concentration gradients and may cause an underestimation of a primary pollutant at an urban site. We also clarify that the "slight overestimations" described in the original text referred primarily to the relative contribution of BC, rather than necessarily to its absolute concentration. We therefore examined both the absolute BC concentrations and relative contributions in Beijing, Tianjin, and Zhengzhou (Table R1).

Table R1. Comparison of observed and simulated BC concentrations and relative contributions.

City	Model	BC concentration: Obs./Sim. (bias)	BC fraction: Obs./Sim. (difference)	Component-sum bias
BJ	EPICC-Model	0.828/1.264 (+52.5%)	3.46/6.14 (+2.67 pp)	−14.0%
	CAMx	0.828/1.482 (+78.9%)	3.46/6.61 (+3.15 pp)	−6.3%
	CMAQ	0.828/1.316 (+58.9%)	3.46/6.61 (+3.15 pp)	−16.8%
TJ	EPICC-Model	1.372/1.393 (+1.5%)	3.32/5.98 (+2.65 pp)	−43.5%
	CAMx	1.372/1.568 (+14.4%)	3.32/6.03 (+2.71 pp)	−37.0%

City	Model	BC concentration: Obs./Sim. (bias)	BC fraction: Obs./Sim. (difference)	Component-sum bias
	CMAQ	1.372/1.504 (+9.6%)	3.32/6.12 (+2.80 pp)	−40.5%
ZZ	EPICC-Model	0.750/1.850 (+146.7%)	2.38/5.36 (+2.98 pp)	+9.6%
	CAMx	0.750/1.977 (+163.7%)	2.38/5.98 (+3.60 pp)	+5.0%
	CMAQ	0.750/1.918 (+155.7%)	2.38/6.31 (+3.93 pp)	−3.6%

Note: BC concentrations are in μg m^{-3}, and BC fractions are in %. "pp" denotes percentage points. The component sum includes NO_3^-, SO_4^{2-}, NH_4^+, OC, and BC.

In Beijing, all three models overestimated the absolute BC concentration by 52.5%-78.9%; in Zhengzhou, the overestimation was 146.7%-163.7%. Thus, the elevated BC fractions in these two cities were accompanied by clear positive biases in absolute BC. Uncertainties in the magnitude or spatial allocation of BC emissions are therefore a plausible contributor, but the comparison alone cannot demonstrate that the emission inventory is the sole cause, because boundary-layer mixing, transport, and dry and wet removal also affect surface BC concentrations. In Tianjin, in contrast, the absolute BC biases were much smaller (1.5%-14.4%), whereas the total concentration of the five measured components was underestimated by 37%-44%, particularly because of underestimated OC. The elevated BC fraction in Tianjin was therefore mainly a denominator effect and does not indicate a comparable overestimation of BC emissions.

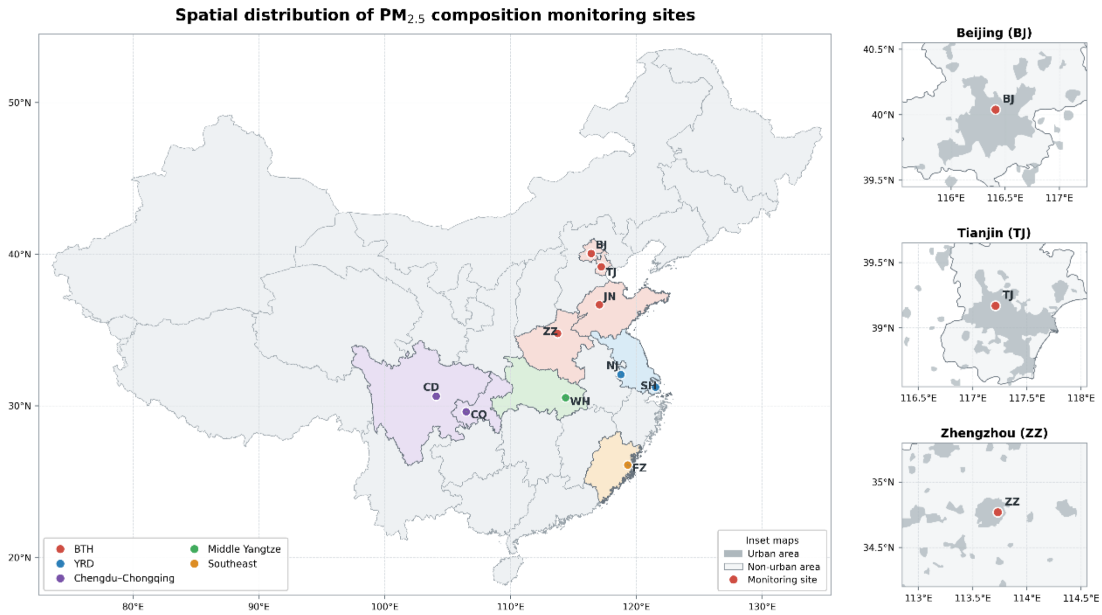

Figure R1. Spatial distribution of the ten $PM_{2.5}$ composition monitoring sites. The enlarged panels show the locations of the Beijing, Tianjin, and Zhengzhou sites relative to the mapped urban extents. Model concentrations were extracted from the grid cells containing the corresponding monitoring sites.

To assess the possible influence of site representativeness, we further examined the spatial locations of the composition monitoring sites. The Beijing, Tianjin, and Zhengzhou sites are located within mapped urban areas, whereas the model values represent averages over the grid cells containing the sites (Figure R1). This point-to-grid mismatch introduces representativeness uncertainty. However, its direction cannot be determined from the available spatial information alone; indeed, a site strongly influenced by local urban emissions could have a higher concentration than the grid-cell average, which would tend to produce a model underestimation rather than explain a positive bias. We therefore identify site representativeness as an uncertainty in bias attribution, rather than as a demonstrated cause of the BC overestimation. The manuscript has been revised accordingly. The revised sentences are as follows:

(Lines 550-558) *In Beijing and Zhengzhou, the elevated fractions are accompanied by overestimated absolute BC concentrations, suggesting possible uncertainties in BC emission magnitude or spatial allocation, as well as in mixing, transport, and removal processes. In Tianjin, however, the absolute BC bias is small, and the elevated fraction mainly results from underestimation of the total measured-component concentration. Differences in spatial representativeness between individual urban sites and grid-cell-averaged concentrations may introduce additional uncertainty, although their influence on the direction of the bias cannot be determined from the available data (Hong et al., 2017).*

Comment 35:

L570-571: It is not clear to me what either "consistency in species simulations" or "stable performance of RMSE and NMB" implies? Are the authors suggesting that EPICC simulated PM composition is similar to the other models? Both RMSE and NMB in Figure 7 exhibit some spread across models for several species - it is not apparent to me what stable performance of these metrics implies?

Response to comment 35:

We appreciate the reviewer's careful observation regarding these qualitative expressions. "Consistency in species simulations" was not intended to indicate that the three models produced identical $PM_{2.5}$ composition, and "stable performance" was not defined as a separate statistical measure. To avoid ambiguity, these expressions have been replaced with a metric-specific description of Figure 7. The revised text now states that CMAQ generally shows higher correlations for NO_3^- and NH_4^+ but tends to overestimate OC, whereas the relative performance of the EPICC-Model and CAMx varies among chemical components and cities. The inter-model dispersion in RMSE and NMB is also explicitly acknowledged. The revised sentences are as follows:

(Lines 612-614) *The performance of the EPICC-Model varies among chemical components and cities, with its RMSE and NMB generally falling within the ranges of the other two models.*

Comment 36:

L576-579: Full references to the documents mentioned here should be provided.

Response to comment 36:

We appreciate the reviewer's helpful suggestion. Full bibliographic information has now been added for the Technical Guideline for Numerical Forecasting of Ambient Air Quality (HJ 1130–2020) and the Technical Regulation for the Assessment of Urban Ambient Air Quality Index (AQI) Forecasts (Trial). The corresponding institutional citations in the main text have also been standardized. The references are as follows:

[1] MEE (Ministry of Ecology and Environment of the People's Republic of China): Technical Guideline for Numerical Forecasting of Ambient Air Quality (HJ 1130–2020), Beijing, China, 2020. Available at: https://www.mee.gov.cn/ywgz/fgbz/bz/bzwb/jcffbz/202005/t20200518_779714.shtml (last access: 8 August 2026).

[2] CNEMC (China National Environmental Monitoring Centre): Technical Regulation for the Assessment of Urban Ambient Air Quality Index (AQI) Forecasts (Trial), Beijing, China, 2021. Available at: https://www.cnemc.cn/zzjj/jgsz/ybzx/gzdt_ybzx/202112/t20211203_962949.shtml (last access: 8 August 2026).

Comment 37:

L582: If heavy and severe AQI are dictated by significant wind-blown dust (WBD) events and if WBD emissions are not included in the models, what is the value of conducting this AQI comparison, since it should be expected that lower accuracy is expected at higher observed AQI. In other words, if the models are not configured for an application, what is the value in "evaluating the applicability" for graded air quality forecasts. Do the measurement locations examined in Figure 6 also report crustal contribution?

Response to comment 37:

We appreciate the reviewer's important comment. We agree that, if the heavy and severe AQI samples were predominantly associated with springtime wind-blown dust events in northwestern China, evaluating models without wind-blown dust emissions for these categories would have limited interpretive value. Following the reviewer's suggestion, we therefore examined the spatial and seasonal distribution of the observed heavy and severe $PM_{2.5}$ events.

Heavy and severe pollution was defined as a daily $PM_{2.5}$ concentration greater than 150 μg m^{-3}. We performed the analysis at both the monitoring-station and city levels. Spring was defined as March-May, and the conventional northwestern region (NW5) included Shaanxi, Gansu, Qinghai, Ningxia, and Xinjiang. Because Inner Mongolia is also affected by frequent dust events, an expanded definition including Inner Mongolia was examined as a sensitivity test.

Table R2. Spatial and seasonal distribution of observed heavy and severe $PM_{2.5}$ pollution events in 2021.

Spatial–seasonal category	Station-days	Percentage of station-days	City-days	Percentage of city-days
All heavy/severe events	4,761	100.0%	902	100.0%

Spatial–seasonal category	Station-days	Percentage of station-days	City-days	Percentage of city-days
Spring and NW5	412	8.7%	139	15.4%
Spring and outside NW5	875	18.4%	124	13.7%
NW5 and outside spring	865	18.2%	214	23.7%
Neither spring nor NW5	2,609	54.8%	425	47.1%
Outside the joint spring–NW5 window	4,349	91.3%	763	84.6%
Spring and expanded NW region	551	11.6%	168	18.6%
Outside the joint spring–expanded-NW window	4,210	88.4%	734	81.4%

As shown in Table R2 and Figure R2, 4,761 observed heavy and severe station-days were identified. Only 412 station-days (8.7%) occurred simultaneously during spring and in NW5, whereas 4,349 station-days (91.3%) occurred outside this joint spatial–seasonal window. Even under the stricter criterion requiring observations to occur in neither spring nor NW5, 2,609 station-days (54.8%) remained. When Inner Mongolia was included in the expanded dust-prone region, 88.4% of the station-day samples still occurred outside the joint spring–northwest window.

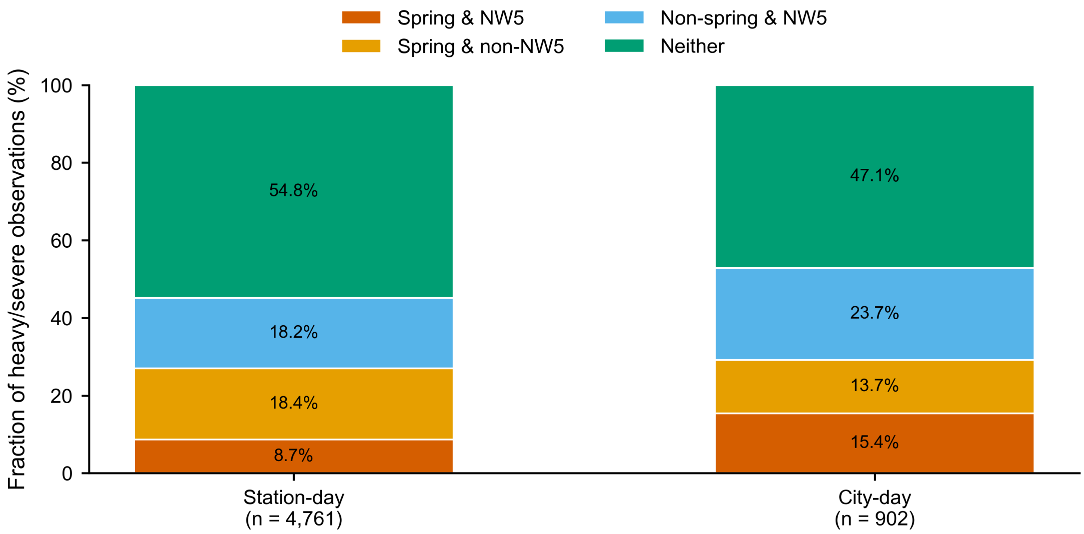

Figure R2. Spatial and seasonal classification of observed heavy and severe $PM_{2.5}$ pollution events in 2021.

The city-day analysis yielded the same qualitative conclusion. Of the 902 observed heavy/severe city-days, 139 (15.4%) occurred during spring in NW5, while 763 (84.6%) occurred outside this joint window. Moreover, 425 city-days (47.1%) occurred in neither spring nor NW5. After Inner Mongolia was included, 734 city-days (81.4%) still occurred outside the joint spring–northwest window. The monthly distribution in Table R3 and Figure R3 further shows that the events were distributed throughout the year and were especially frequent in January, March, and February, rather than being concentrated exclusively in spring.

Table R3. Monthly distribution of observed heavy and severe $PM_{2.5}$ pollution events in 2021.

Month	Station-days	City-days	Number of affected cities	Percentage of all city-days
January	1,852	340	107	37.7%
February	684	112	54	12.4%
March	996	193	82	21.4%
April	147	33	20	3.7%
May	144	37	17	4.1%
June	9	4	4	0.4%
July	7	3	1	0.3%
August	12	4	2	0.4%
September	7	3	1	0.3%
October	124	31	14	3.4%
November	293	58	30	6.4%
December	486	84	50	9.3%

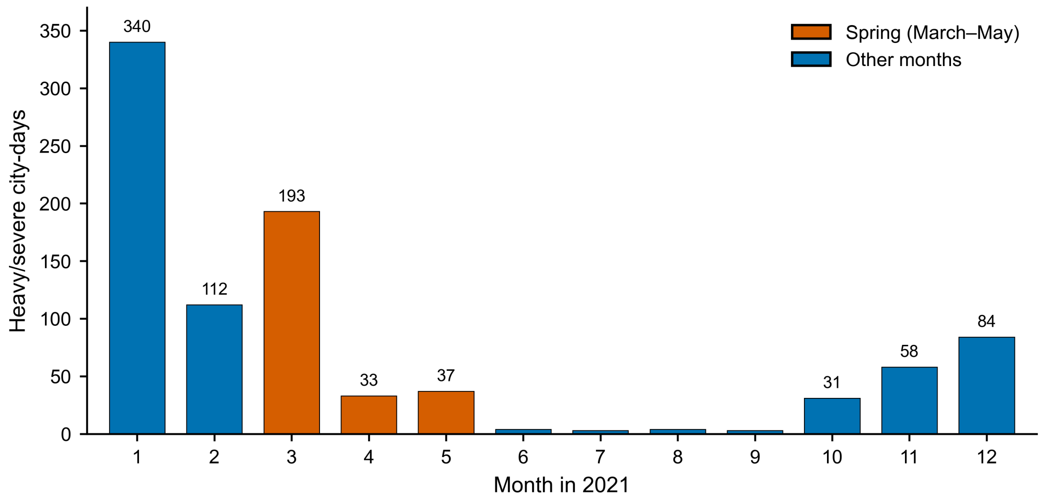

Figure R3. Monthly distribution of observed heavy and severe $PM_{2.5}$ pollution events in 2021.

We further evaluated the models using the unambiguous event threshold of $PM_{2.5} > 150$ μg m^{-3}. As shown in Table R4 and Figure R4, none of the models captured a spring–NW5 heavy-pollution station-day. However, limited hit rates also persisted outside this joint window: 13.4% for the EPICC-Model, 6.0% for CAMx, and 29.4% for CMAQ. For observations occurring in neither spring nor NW5, the hit rates increased to 21.5%, 9.7%, and 43.7%, respectively, but substantial underprediction remained.

Table R4. Model hit rates for observed heavy and severe $PM_{2.5}$ pollution events in different spatial–seasonal subsets.

Observed-event subset	Number of station-days	EPICC-Model	CAMx	CMAQ
All events with $PM_{2.5} > 150$ μg m^{-3}	4,761	12.2%	5.4%	26.9%
Spring and NW5	412	0.0%	0.0%	0.0%
Outside the joint spring–NW5 window	4,349	13.4%	6.0%	29.4%

Observed-event subset	Number of station-days	EPICC-Model	CAMx	CMAQ
Neither spring nor NW5	2,609	21.5%	9.7%	43.7%

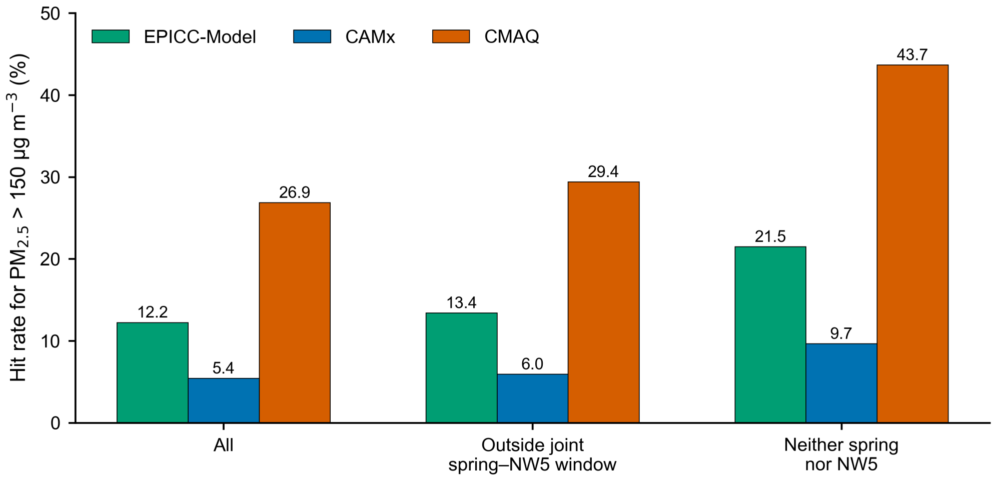

Figure R4. Model hit rates for observed heavy and severe $PM_{2.5}$ pollution events in different spatial–seasonal subsets.

These results demonstrate that the heavy/severe $PM_{2.5}$ statistics are not dictated primarily by the canonical springtime northwestern dust regime. The omission of wind-blown dust emissions is likely an important limitation for the spring–northwest subset, but it is insufficient to explain the model behavior for the majority of the observed high-$PM_{2.5}$ events. Errors outside this subset may additionally reflect limitations in the representation of anthropogenic haze, secondary aerosol formation, emissions, and meteorological processes under high-pollution conditions.

We have therefore retained the highest $PM_{2.5}$ category as a diagnostic benchmark of model performance under the present configurations, while revising the manuscript to avoid interpreting it as an evaluation of dust-forecasting capability or attributing all errors in this category to omitted wind-blown dust. The component-monitoring sites used in Figure 6 provide measurements of sulfate, nitrate, ammonium, OC, and BC, but do not provide directly comparable measurements of crustal material. Therefore, the crustal contribution to individual high-AQI events could not be quantified directly.

The revised sentences are as follows:

(Lines 626-633) *Because the present model configurations do not include wind-blown dust emissions, the combined heavy/severe $PM_{2.5}$ category is retained as a diagnostic evaluation of model performance under the configurations used in this study, rather than as an evaluation of dust-forecasting capability. To determine whether the statistics for this category were dominated by typical dust events, the spatial and seasonal distribution of the observed heavy/severe $PM_{2.5}$ events was additionally examined. The detailed analysis is presented in Section S4 and Table S1 of the Supplement.*

Comment 38:

L591-596: This is a potentially useful finding but based on the results presented it is *somewhat speculative. Based on comparisons of concentrations of secondary inorganic aerosol constituents by the 3 models for selected such episodes, can the authors demonstrate that EPICC indeed predicts different and more accurate "rapid growth" of secondary inorganic aerosols?*

Response to comment 38:

We appreciate the reviewer's recognition of the potential value of this finding and the constructive suggestion for providing more direct evidence. The original interpretation was based on the categorical performance summarized in Tables 5-7. We agree that categorical AQI and IAQI accuracy alone is insufficient to evaluate the temporal buildup of individual aerosol components. To strengthen the process-level evidence, we additionally examined aerosol-component evolution during the representative 20-28 January 2021 $PM_{2.5}$ episode shown in Figure 10. Beijing (BJ) and Zhengzhou (ZZ) were used because these were the two cities in Figure 10 for which daily aerosol-component observations were available. Secondary inorganic aerosol (SIA) was calculated as the sum of nitrate (NO_3^-), sulfate (SO_4^{2-}), and ammonium (NH_4^+).

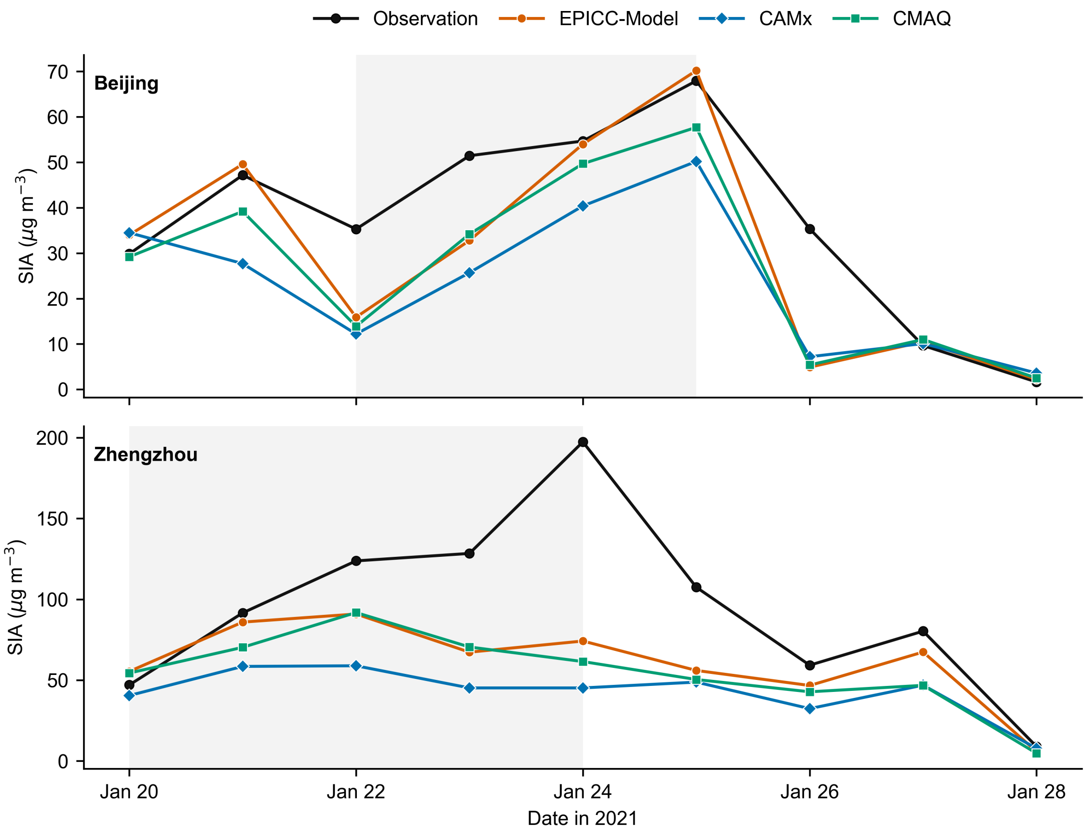

Figure R5. Observed and simulated daily SIA concentrations in Beijing and Zhengzhou during the 20-28 January 2021 $PM_{2.5}$ episode. SIA is calculated as the sum of nitrate, sulfate, and ammonium. The shaded areas indicate the observed buildup periods used in the quantitative evaluation: 22-25 January in Beijing and 20-24 January in Zhengzhou.

Table R5. Performance of the three models for daily total SIA during the 20-28 January 2021 $PM_{2.5}$ episode.

City	Model	Event R	Event RMSE	Event NMB	Observed peak	Simulated peak	Peak timing error	Observed buildup	Simulated buildup	Buildup RMSE
BJ	EPICC-Model	0.848	13.67	−17.7%	67.94	70.20	0 d	32.64	54.30	13.51
	CAMx	0.807	17.99	−36.5%	67.94	50.20	0 d	32.64	38.00	20.71
	CMAQ	0.863	14.34	−27.1%	67.94	57.70	0 d	32.64	43.80	14.88
ZZ	EPICC-Model	0.702	50.62	−34.9%	197.35	90.90	−2 d	150.33	18.80	63.37
	CAMx	0.630	67.24	−54.5%	197.35	58.90	−2 d	150.33	4.70	84.15
	CMAQ	0.671	55.78	−41.6%	197.35	91.80	−2 d	150.33	7.10	68.31

For total SIA, the EPICC-Model showed the lowest event-period RMSE and buildup-stage RMSE in both cities (Figure R5 and Table R5). In Beijing, the event-period RMSE values of the EPICC-Model, CAMx, and CMAQ were 13.67, 17.99, and 14.34 μg m^{-3}, respectively. The corresponding values in Zhengzhou were 50.62, 67.24, and 55.78 μg m^{-3}. Averaged across the two cities, the event-period RMSE was 32.14 μg m^{-3} for the EPICC-Model, compared with 42.62 μg m^{-3} for CAMx and 35.06 μg m^{-3} for CMAQ. The EPICC-Model also had the smallest mean absolute NMB, at 26.3%, compared with 45.5% for CAMx and 34.3% for CMAQ. These results indicate that the EPICC-Model achieved comparatively better overall agreement for the SIA concentration evolution during the selected episode.

The individual-component comparison further illustrates the model responses associated with different secondary aerosol pathways (Figure R6 and Table R6). Manganese-catalyzed sulfur dioxide (SO_2) oxidation is directly relevant to sulfate formation. The EPICC-Model had the lowest sulfate RMSE in Zhengzhou and substantially reduced the sulfate underestimation relative to CAMx and CMAQ, while its sulfate concentration in Beijing was higher than observed. Heterogeneous hydrolysis of dinitrogen pentoxide (N_2O_5) is directly relevant to nitrate formation. For nitrate, the EPICC-Model performed better than CAMx and was close to CMAQ in Zhengzhou, whereas CMAQ had the lowest nitrate RMSE in both cities. The EPICC-Model had the lowest ammonium RMSE in both cities, consistent with its improved representation of the combined inorganic aerosol system. Heterogeneous nitrous acid (HONO) formation may affect secondary aerosol production through its influence on atmospheric oxidation capacity, but its individual contribution cannot be evaluated directly because HONO observations were not available.

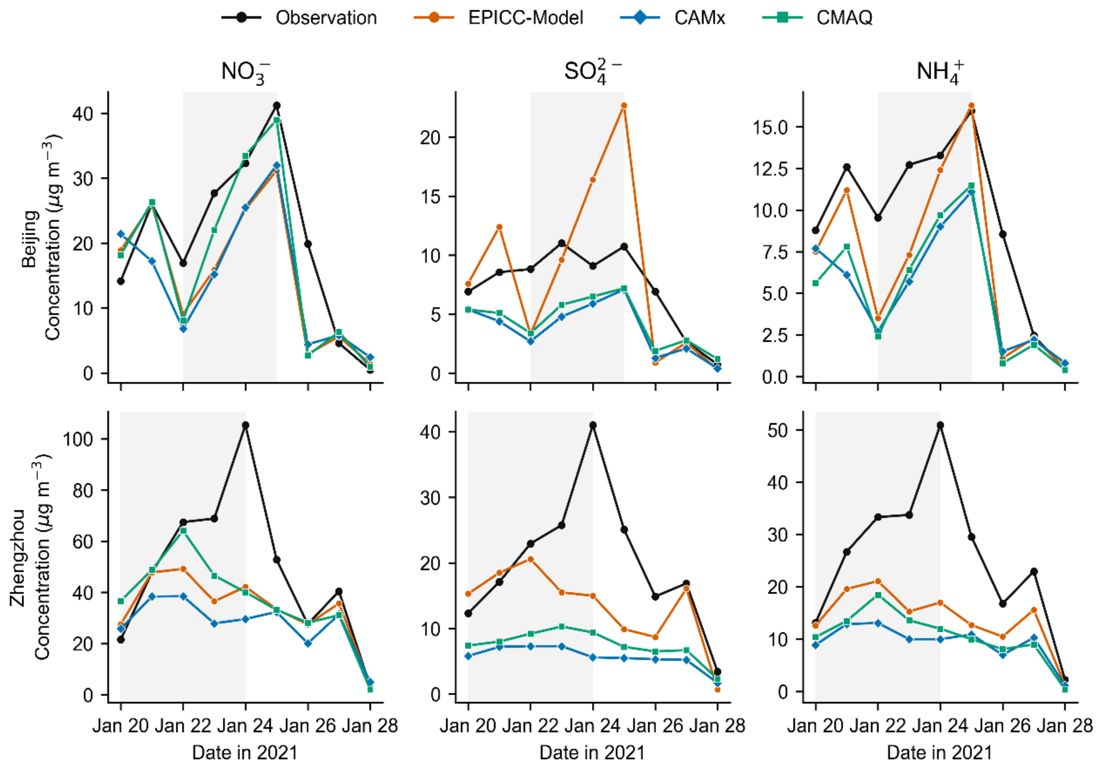

Figure R6. Observed and simulated daily nitrate, sulfate, and ammonium concentrations in Beijing and Zhengzhou during the 20-28 January 2021 $PM_{2.5}$ episode. The shaded areas indicate the observed buildup periods used in the quantitative evaluation.

Table R6. Event-period and buildup-stage performance for individual secondary inorganic aerosol components.

City	Component	Model	Event RMSE	Event NMB	Buildup RMSE
BJ	NO_3^-	EPICC-Model	8.59	−25.7%	9.38
		CAMx	9.22	−28.7%	9.88
		CMAQ	6.92	−14.3%	5.41
	SO_4^{2-}	EPICC-Model	5.59	+15.8%	7.57
		CAMx	4.10	−47.9%	4.99
		CMAQ	3.59	−40.0%	4.36
	NH_4^+	EPICC-Model	3.74	−26.3%	4.09
		CAMx	5.07	−44.5%	5.88
		CMAQ	4.92	−44.9%	5.56
ZZ	NO_3^-	EPICC-Model	25.40	−30.2%	32.89
		CAMx	31.49	−42.8%	40.91
		CMAQ	24.65	−24.0%	31.61
	SO_4^{2-}	EPICC-Model	10.94	−33.0%	12.65
		CAMx	16.99	−71.6%	19.91

City	Component	Model	Event RMSE	Event NMB	Buildup RMSE
		CMAQ	15.03	−62.7%	17.53
		EPICC-Model	15.16	−45.2%	18.39
	NH_4^+	CAMx	19.61	−63.3%	23.90
		CMAQ	18.19	−58.4%	21.55

Taken together, the categorical and process-oriented results support the interpretation that the EPICC-Model has a competitive capability for representing secondary aerosol evolution. The comparatively better agreement for total SIA is physically consistent with the heterogeneous pathways represented in the EPICC-Model, including manganese-catalyzed SO_2 oxidation, N_2O_5 heterogeneous hydrolysis, and heterogeneous HONO formation. The revised sentences are as follows:

(Lines 638-643) *The EPICC-Model showed competitive performance in several light and moderate pollution categories. This performance is physically consistent with the heterogeneous pathways represented in the model, including manganese-catalyzed SO_2 oxidation, N_2O_5 heterogeneous hydrolysis, and heterogeneous HONO formation (Wang et al., 2021; Zhang et al., 2022; Yang et al., 2024; EPICC-Model Working Group, 2026).*

Comment 39:

L696-700: How does one quantify spatial progression from central to south China to assess which model simulated this progression "best". Figure 11 only enables a qualitative and subjective assessment.

Response to comment 39:

We appreciate the reviewer's important comment regarding the need for a quantitative assessment of the spatial progression. To supplement the qualitative distributions shown in Figure 11, the daily spatial index of agreement (IOA) for MDA8 O_3 was calculated across monitoring stations in central and eastern China during 3-14 June 2021. The same set of valid stations was used for the observations and all three models on each day.

As shown in Figure R7 and Table R7, the mean daily spatial IOA values were 0.752, 0.745, and 0.748 for the EPICC-Model, the Comprehensive Air Quality Model with Extensions (CAMx), and the Community Multiscale Air Quality Modeling System (CMAQ), respectively. However, each model achieved the highest daily spatial IOA on four of the twelve days. For the four representative dates displayed in Figure 11, the model with the highest IOA varied among the EPICC-Model, CAMx, and CMAQ. Moreover, when the MDA8 O_3 concentrations were first averaged over the entire episode at each station, CMAQ showed the highest spatial IOA.

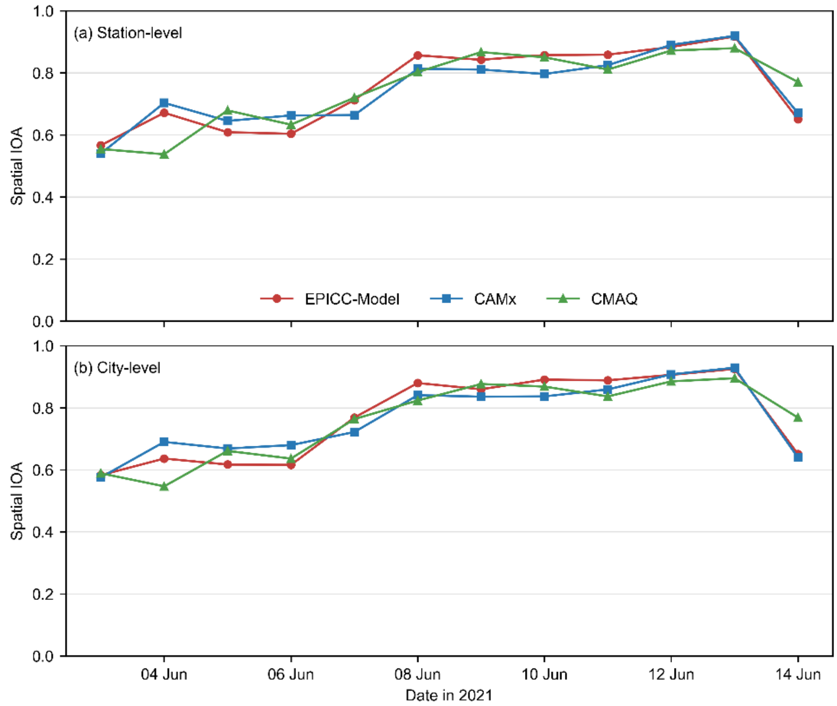

Figure R7. Daily spatial IOA for maximum daily 8 h average ozone at monitoring stations in central and eastern China during 3-14 June 2021. The city-level results are included as a sensitivity analysis.

Table R7. Quantitative evaluation of the spatial performance for the O_3 episode over central and eastern China during 3-14 June 2021. Spatial IOA was calculated using monitoring sites with valid observations and predictions from all three models.

Metric	EPICC-Model	CAMx	CMAQ
Mean daily spatial IOA	0.752	0.745	0.748
Spatial IOA of event-mean MDA8 O_3	0.822	0.779	0.834
Days with the highest spatial IOA	4	4	4
Mean absolute error in polluted-city coverage, percentage points	5.5	12.2	12.3
Mean absolute error in polluted-city coverage during 5-9 June, percentage points	5.8	20.4	21.5
Spatial mean bias on 6 June, $\mu g\ m^{-3}$	−12.3	−27.7	−29.9

These results indicate that the EPICC-Model maintained competitive spatial agreement throughout the episode but do not support the interpretation that it reproduced the spatial progression best on every day or according to every spatial metric. We have therefore removed "best reproduced" and revised the manuscript to distinguish the quantitative spatial agreement from the qualitative description of the progression.

Comment 40:

L700: Would be useful to elaborate on how excessively strong aerosol-radiation feedback led to O3 overestimation in southeastern China and on which day?

Response to comment 40:

We appreciate the reviewer's request for clarification. The local O_3 overestimation referred specifically to the peak stage on 6 June, primarily over southeastern China. We agree that the original attribution to an excessively strong aerosol-radiation feedback

was overly specific and was not directly demonstrated by the analyses presented in this study. Aerosol radiative effects may influence O_3 through competing changes in photolysis rates, boundary-layer development, precursor accumulation, and regional transport. Therefore, the direction and magnitude of the resulting O_3 response cannot be inferred from the strength of the aerosol-radiation feedback alone.
We have removed this specific attribution, explicitly identified 6 June as the relevant date, and revised the discussion to state that the local overestimation may reflect the combined effects of uncertainties in simulated photochemistry and transport. The revised sentence is as follows:

(Lines 741-745) *The EPICC-Model reproduced the expansion of the elevated-O_3 area from central China toward southeastern and southern China and the subsequent clearing, although it locally overestimated O_3 over southeastern China during the peak stage on 6 June, potentially reflecting uncertainties in simulated photochemistry and regional transport.*

Comment 41:

L701: Visually while CAMx and CMAQ appear to underestimate relative to the EPICC fields, both CAMx and CMAQ also appear to better capture the gradients in O3 - high O3 at monitors interspersed with monitors with low O3 while EPICC depicts more regionally elevated O3.

Response to comment 41:

We appreciate the reviewer's careful observation of the local spatial gradients shown in Figure 11. We agree that the three models exhibit different strengths in representing the spatial characteristics of this episode. CAMx and CMAQ generally simulated lower O_3 concentrations in the observed high-concentration regions, particularly on 6 and 10 June, but reproduced the local contrasts between neighboring high- and low-O_3 monitoring sites more clearly. In comparison, the EPICC-Model reproduced a broader region of elevated O_3, but its spatial field was smoother and more extensive, leading to local overestimation in parts of central and southeastern China.
We have incorporated this distinction into the revised manuscript to provide a more balanced interpretation and to avoid implying that the broader elevated-O_3 field simulated by the EPICC-Model necessarily represents better performance for local-scale gradients. The revised sentences are as follows:

(Lines 745-750) *CAMx and CMAQ generally underestimated O_3 in the observed high-concentration regions, particularly on 6 and 10 June, but reproduced the local contrasts between neighboring high-O_3 and low-O_3 monitoring sites more clearly. In comparison, the EPICC-Model reproduced a broader region of elevated O_3 but produced a smoother and more spatially extensive enhancement, resulting in local overestimation in parts of central and southeastern China.*

Comment 42:

L706-708: In my assessment Figure 11 in its current form does not conclusively support the claim conveyed in this statement that EPICC outperformed the other two models in capturing the spatiotemporal patterns.

Response to comment 42:

We appreciate the reviewer's careful assessment of Figure 11. In response, the comparison was supplemented with quantitative metrics for both spatial agreement and the regional extent of elevated O_3. As shown in Table R7, the mean daily spatial index of agreement (IOA) was slightly higher for the EPICC-Model (0.752) than for CAMx (0.745) and CMAQ (0.748), although no model achieved the highest spatial IOA consistently throughout the episode.

The regional extent of the episode was further quantified using the fraction of cities with maximum daily 8 h average ozone (MDA8 O_3) exceeding 160 μg m^{-3}. As shown in Figure R8 and Table R7, the mean absolute error in this fraction during 3-14 June was 5.5 percentage points for the EPICC-Model, compared with 12.2 and 12.3 percentage points for CAMx and CMAQ, respectively. During the main development and peak period of 5-9 June, the corresponding errors were 5.8, 20.4, and 21.5 percentage points. On 6 June, the spatial mean biases were −12.3 μg m^{-3} for the EPICC-Model, −27.7 μg m^{-3} for CAMx, and −29.9 μg m^{-3} for CMAQ.

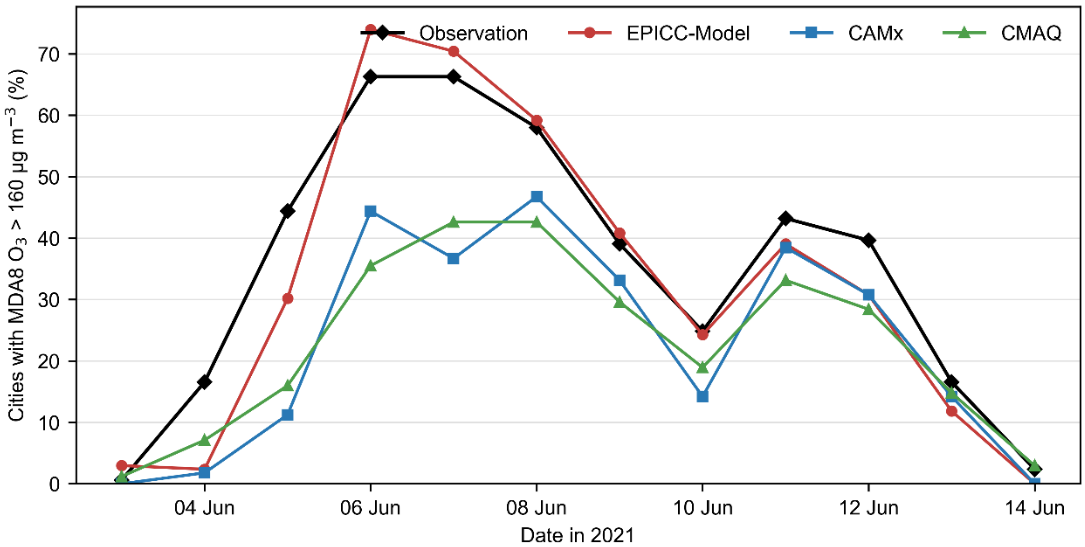

Figure R8. Observed and simulated fractions of cities with maximum daily 8 h average ozone exceeding 160 μg m^{-3} over central and eastern China during 3-14 June 2021.

These results provide direct quantitative support for an advantage of the EPICC-Model in representing the overall regional extent and magnitude of the elevated-O_3 area during the development and peak stages. However, they do not indicate uniform superiority in all aspects of the spatial pattern. We have therefore replaced the original broad statement that the EPICC-Model "outperformed" the other two models with a more specific conclusion concerning the regional-scale extent and overall magnitude of the event, while retaining the limitations evident in Figure 11.

The revised sentences are as follows:

(Lines 750-752) *Overall, the EPICC-Model showed an advantage in representing the regional extent of the elevated-O_3 area during this episode, although no model consistently reproduced all aspects of the spatial pattern better than the others.*

Comment 43:

L725-726: I am not sure what "unified WRF fields" implies. Also as noted earlier, given the different layer structures, it is also difficult to ascertain whether the 3D fields are identical or even consistent across models. Thus, probably should be reworded to: "using meteorological fields derived from the same WRF simulation".

Response to comment 43:

We appreciate the reviewer's helpful clarification. The expression "unified WRF meteorological fields" was intended to indicate that the meteorological inputs for the three CTMs originated from the same WRF simulation, rather than that the processed three-dimensional fields were identical. Because the WRF outputs were converted according to the vertical structures and input requirements of the individual CTMs, we have replaced this expression with "meteorological fields derived from the same WRF simulation." The revised sentences are as follows:

(Lines 768-771) *This study presents a systematic evaluation of three CTMs (EPICC-Model, CAMx, and CMAQ) for $PM_{2.5}$ and MDA8 O_3 simulations over China in 2021, using meteorological fields derived from the same WRF simulation and multi-source emission inventories.*

Comment 44:

L731: "seasonal turning point" is vague – sentence should be reworded.

Response to comment 44:

We appreciate the reviewer's helpful comment on this wording. The phrase "seasonal turning points" was intended to describe the seasonal change in the direction of the $PM_{2.5}$ simulation bias, but it did not convey this meaning clearly. We have replaced it with a direct description stating that the EPICC-Model tends to underestimate $PM_{2.5}$ in spring and summer and overestimate it in autumn and winter. The revised sentences are as follows:

(Lines 775-776) *The $PM_{2.5}$ bias of the EPICC-Model varies seasonally, with a tendency toward underestimation in spring and summer and overestimation in autumn and winter.*

Comment 45:

L745-746: No comparisons with observed boundary layer heights were presented to support the statement that ACM2/YSU schemes underestimate mixing layer heights. The statement should be adequately qualified.

Response to comment 45:

We appreciate the reviewer's careful comment on the scope of this interpretation. The discussion of mixing-layer height was intended as a possible explanation based on previously reported behavior of the ACM2/YSU schemes under stable conditions, rather than as a conclusion derived from a direct evaluation of boundary-layer height in this study. We have therefore qualified the statement accordingly. The revised text indicates that insufficient nocturnal mixing may contribute to the autumn–winter $PM_{2.5}$ overestimation, consistent with previous evaluations of these boundary-layer schemes, without implying that their mixing-layer-height bias was directly established here. The revised sentences are as follows:

(Lines 786-789) *Conversely, autumn-winter $PM_{2.5}$ overestimations occur in North China and the Sichuan Basin. Insufficient nocturnal mixing may contribute to these biases, consistent with previously reported tendencies of the ACM2/YSU boundary-layer schemes under stable conditions (Jia et al., 2023).*

Comment 46:

L800: It is stated that this work "provides a clear pathway for improving next generation of atmospheric CTMs". I feel this is a somewhat lofty claim as no new or specific process or structural deficiency was identified. There may certainly be specific enhancements identified for the EPICC system, but I am not sure the analysis identified a "clear path" for next generation of models. Suggest deleting this sentence as the study and results do not substantiate it.

Response to comment 46:

We appreciate the reviewer's careful assessment of the scope of this conclusion. The expression "provides a clear pathway for improving the next generation of atmospheric CTMs" was broader than warranted by the operational evaluation presented in this study. Our intention was to emphasize that the comparison identifies practical priorities for subsequent model development, including the representation of natural dust, boundary-layer mixing, secondary aerosol formation, and emission inputs. We have therefore replaced the original statement with a more specific and appropriately scoped conclusion. The revised sentences are as follows:

(Lines 839-844) *This study establishes a multi-model evaluation framework for systematically comparing the performance of the EPICC-Model with established CTMs over China. By identifying model-specific limitations and challenges shared across the three models, the comparison highlights priorities for future model development, including improved representations of natural dust, boundary-layer mixing, secondary aerosol formation, and emission inputs.*

Comment 47:

L802-804: In my assessment this study at best provides a comparison of operational *evaluation metrics of 3 models. There is no specific analysis presented to support air pollution control or policy making. Thus, I feel this statement should be reworded or deleted.*

Response to comment 47:

We appreciate the reviewer's careful comment on the policy relevance of the study. The present work evaluates model performance rather than directly assessing emission-control strategies or policy scenarios. Therefore, the statement that the results provide direct scientific support for air-pollution control and policy-making has been revised. The new wording emphasizes the demonstrated contribution of the study: providing benchmark information for the evaluation, selection, and further development of regional air-quality modeling systems in China and other regions facing similar modeling challenges. The revised sentences are as follows:

(Lines 844-848) *The results provide benchmark information for evaluating, selecting, and further developing regional air-quality modeling systems in China. The multi-model evaluation framework and identified performance characteristics may also serve as a reference for model applications in other regions facing similar air-quality modeling challenges.*