

General Comments:

The authors describe the development of an atmospheric chemistry modeling framework based on an MPAS-A dynamical core coupled with online gas-phase chemistry using a conventional KPP-based approach. The study also examines the resolution dependence of terrain representation, which can substantially influence ozone transport and near-surface concentrations over complex terrain. The authors further use the higher-resolution configuration to investigate ozone formation and its controlling physical and chemical processes over North China. However, I have significant concerns regarding the definition of the model development, the relationship between the present version and previous iAMAS/MPAS-A developments, the verification of the newly implemented chemistry, the experimental design, and the interpretation of the resolution-dependent ozone responses.

In its current form, I find it difficult to identify precisely what constitutes the new model development presented in this paper. The manuscript mixes descriptions of the inherited MPAS-A dynamical core, capabilities already documented in previous iAMAS versions, newly introduced chemistry components, model evaluation, and a scientific application to North China. In addition, the paper mixes model description, chemistry development, global evaluation, and a detailed North China application, making the main contribution difficult to follow. The manuscript should be reorganized around the new model development and verification, with the regional application presented more clearly as an evaluation or demonstration case. Consequently, the novelty and technical contribution expected for a GMD model-development paper are not sufficiently clear. GMD emphasizes precise model/version specification, reproducibility, and evidence that newly developed code implements the intended mathematical and physical formulation.

I think this manuscript requires substantial revision before the scientific conclusions can be adequately evaluated. I therefore do not think the manuscript is suitable for publication in GMD in its current form. The issues below require substantial clarification, additional verification, and in several cases additional experiments or diagnostics. Please find my concerns and comments listed below:

Major concerns and comments:

- 1) The new development relative to previous iAMAS versions is unclear. Previous iAMAS publications already describe an MPAS-A-based global variable-resolution model with online emissions, tracer transport, diffusion, vertical mixing, dry and wet deposition, and aerosol–radiation/aerosol–cloud processes. Therefore, terms such as “emissions,” “transport,” and “deposition” by themselves cannot be considered new developments of the present study. From my reading, the genuinely new components appear to be mainly: coupling SAPRC-99 gas-phase chemistry through KPP; incorporation/coupling of Fast-J photolysis; implementation of Linoz for stratospheric ozone? If that is correct, the manuscript should state this explicitly. I strongly recommend adding a table containing what is the new implementation in the current version compared to the previous version.

Without such a comparison, it is not possible to judge the novelty of the present model version.

- 2) The relevant MPAS-A literature should be cited directly where inherited algorithms are described, rather than allowing the reader to interpret inherited MPAS functionality as a new iAMAS contribution. This includes, where appropriate, Skamarock et al. (2012), Ringler et al. for SCVT, and the relevant MPAS transport/numerical references. Similarly, the absence of lateral boundary conditions is fundamentally a consequence of the global MPAS/SCVT architecture and should not be presented as if it were introduced by the new chemistry development.
- 3) The iAMAS version history appears internally inconsistent. This needs to be resolved before the technical contribution can be evaluated. A separate 2026 GMD preprint describes iAMAS v2.6.3 and states that anthropogenic emissions and SAPRC-99 gas-phase chemistry advanced iAMAS to v2.6, while the TVD–FFSL tracer transport scheme introduced in that study was implemented in v2.6.3 (<https://egusphere.copernicus.org/preprints/2026/egusphere-2026-3149/>). The present manuscript, however, is labelled iAMAS v2.5.1, yet appears to employ TVD–FFSL transport and contains the SAPRC-99 development. The authors need to explain: the chronological relationship among v1.0, v2.5.1, v2.6, and v2.6.3; whether v2.5.1 and v2.6.x represent different development branches; whether TVD–FFSL was backported into v2.5.1; which exact code commit/tag produced the simulations; which components belong to each version. At present the version numbering makes the development history difficult to reproduce and raises uncertainty about which code is actually evaluated in this manuscript.
- 4) For a GMD model-development paper, important implementation details remain unclear, including how SAPRC-99/KPP, photolysis, emissions, transport, and deposition are coupled, how species are mapped, and how numerical consistency is maintained. Basic benchmark tests against an established SAPRC-99 implementation would also help verify the new chemistry framework.
- 5) The model evaluation has important limitations. The chemical spin-up is only a few days, which is likely insufficient for species with relatively long lifetimes such as CO, O₃, and some reservoir species. In addition, the ozone field used for initialization appears closely related to the dataset later used for evaluating the vertical ozone distribution, so this comparison is not fully independent. The authors should provide a longer spin-up or demonstrate that the results are insensitive to initialization, and independent observations such as ozonesondes would provide a more convincing evaluation. Since the regional mechanism also relies strongly on PBL height, mixing, and terrain-induced circulations, these meteorological processes should be evaluated more directly as well. Therefore the spectacular R=0.99 zonal-mean ozone agreement is not a convincing independent validation, particularly so soon after initialization. I would either run a much longer chemistry spin-up/restart, or substantially soften the claims around Fig. 3. Independent ozonesondes would be much more meaningful.
- 6) The main concern is that the current evaluation is not sufficient to support the strength of the conclusions. Surface ozone performance is evaluated largely using highly averaged diurnal profiles, which can mask substantial spatial and temporal errors. More detailed

station-level and hourly evaluation is needed. In addition, the model does not include biogenic VOC emissions (I understand author stated as limitation and future model directions), however modeled FNR is used to diagnose photochemical regimes and interpret ozone responses over North China. In fact, there are a number of studies that summertime ozone largely attributed to BVOCs in East Asia. Thus, this substantially limits confidence in the conclusions regarding VOC- and NO_x-limited chemistry. The applicability of the selected FNR thresholds and their interpretation across the full diurnal cycle also require reconsideration. I also have concerns about the FNR methodology itself. The manuscript applies fixed thresholds of $FNR < 1$ for VOC-limited conditions and $FNR > 2$ for NO_x-limited conditions and interprets the full diurnal cycle using these thresholds. However, such thresholds are strongly region dependent, and their applicability to North China should be justified much more carefully. More importantly, FNR is fundamentally a photochemical-regime diagnostic, but Figure 11 applies and interprets the same thresholds throughout nighttime as well. I do not think nighttime FNR should be interpreted in the same manner as midday photochemical sensitivity. The cited literature and the threshold definitions should therefore be revisited (it is not consistent with how the author described the FNR, and how the cited work described), and the conclusions based on Figure 11 should be substantially reassessed. If these comments cannot be addressed adequately, respectfully, the author cannot argue or state that they established the modeling framework works for ozone chemistry in East Asia.

- 7) The treatment of nighttime chemistry is incomplete. The manuscript attributes nocturnal O₃ changes mainly to NO_x accumulation and NO titration, but the model does not include aerosol-gas heterogeneous chemistry, particularly N₂O₅ uptake, which can be an important nighttime NO_x sink over the North China Plain and can influence subsequent ozone chemistry. This omission should be explicitly discussed, and the related mechanistic conclusions should be qualified or supported by additional sensitivity analysis. Also, dry deposition should be considered in the nighttime ozone mechanism
- 8) The 60-km versus 4-km comparison does not isolate the effects of resolution. The attribution also appears stronger than what the current experiments can establish. Resolution changes simultaneously affect terrain, emissions, transport, boundary-layer structure, and numerical behavior, while these contributions are not isolated. The proposed terrain-circulation mechanism is further supported mainly by model diagnostics without independent meteorological evaluation. The interpretation of emission heterogeneity, urban and rural classifications, and TPI analysis also requires greater caution. For me this should not be left entirely to “future work.” At least one or two targeted experiments would transform the paper: for example, V4km with the 60-km-smoothed emission field, and V4km with smoothed/coarse orography. Then they could quantitatively say how much of ΔO_3 comes from emissions versus dynamics.
- 9) Several chemical mechanisms are inferred rather than directly diagnosed. The manuscript often uses correlations or proxy variables to infer ozone production, titration, and vertical transport. Direct ozone-budget diagnostics, including chemical production/loss, NO titration, transport, mixing, and deposition, would provide much stronger support for the proposed mechanisms.

- 10) In addition, the same ozone climatology is used for model initialization and evaluation, which limits the independence of the global ozone assessment. Finally, if NO titration is central to the nighttime ozone interpretation, it should be diagnosed more directly using modeled NO, Ox, or relevant chemical loss rates rather than inferred primarily from NO₂.

Specific comments:

The manuscript repeatedly motivates iAMAS using high-performance computing and the feasibility of global high-resolution atmospheric chemistry. Yet I do not see a meaningful assessment of the computational cost introduced by the new chemistry.

Line 24-26: no need to say U60km, V4km here, confusing. What does U and V stand for? Just need to define where it was first introduced with description.

Line 25: You already say O₃, and immediately say ozone...? Should be the other way around.

Line 28: same, ozone -> O₃. several places in abstract, please use ozone (O₃) in the first place and other places just use O₃.

Abstract: hyphen length is different. ex) cross-scale, and physical-chemical, plz check other places as well.

Abstract: I read twice, but really don't think that the abstract is well written. The first few sentences are okay, but you start with your model evaluating large-scale ozone and precursors, I expect to see a performance summary. And suddenly the next sentence moves with a 4km mesh that reproduces the O₃ pattern...? And then suddenly shift from physics-chemistry coupling to dynamics-chemistry..?. Please consider re-write abstract with more logical progression.

Line 36: I believe you want to say atmospheric chemistry composition, or did you mean overall composition including meteorological variables as well?

Line 39: local scale includes less than kilometer scale.

Line 40: CAM-chem, full name, next few lines, WRF-Chem, CMAQ. You need to spell out.

Line 41: many GEOS-chem work, especially, these days, higher than 100km resolution.

And also would be more appropriate to say regional and local. Not just local. And also, CMAQ and WRF-Chem, is not always run in convection allowing scale. You can't just generalize simply here. Many CMAQ and WRF-Chem reported with lower than 8, 12km existed. And also it is not so clear that these regional chemical transport models, are you saying it has prescribed boundary conditions from nesting? Or are you saying prescribed boundary conditions from the global model such as GEOS-Chem, WACCM?? Those two are clearly different. Need a better description

Line 55: again, weird hyphen.

Line 56: same, hyphen.

Introduction: respectfully, I do not understand how this manuscript got the minor revision during the first round. Introduction needs substantial improvement. I stopped documenting line by line review here, b/c number of problems from first line of the introduction to the line 70. First of all, acronyms, no consistency at all. Sometimes defined, sometimes not. Second, no logical progression, current writing is very confusing. I understand that author wants to start from the well-known limitations from the traditional global and regional scale, but when iAMAS is introduced, limitations - previously established work - so what are the remaining, what are the uncertainties - so what are the things that we've done here logic is not well established. And the heart of the work, dycore, "Model for Prediction Across Scales" MPAS developed from NCAR, not even spelled out?

Line 82:..? Reuse the full name of iAMAS here again..?

Line 93: Thompson microphysics, I don't think this is a start state-of-the-art microphysics scheme. And you wrote a famous scheme for convection, mp, radiation, and extra, I feel that spell these out makes much more sense.

Section 2.1 also needs substantial rewriting.

Line 116: Do you mean inline-emission processing? And the sentence is very weird, gas phase chemistry framework include gas phase chemistry..?

Line 130: you said established performance, but only one citation? There are many studies using SAPRC in East Asia, consider citing it more. I don't think it has to be "SAPRC99", SAPRC with different versions well established for reproducing O3 formation in regional models.

Line 149: NOx long live..? Now I'm suspicious and question the robustness of the entire research.

Line 151: Better explanation needed here. Please list or describe how you lump this.

From Line 203 to the end of section 2.4.5. Consider moving this to SI. None of these are the heart of this work.

Line 230: Is this classification correct? Where is the suburb? And in most cases, urban, sub-urban, and rural are classified using population density. Authors need to cross validate to use their own NO2 driven classification.

Line 232: The station distribution is shown in Figure S1 -> The stations are shown in Figure S1

Line 233: Spell out MB, NMB, RMSE. And ozone in fact has performance criteria for modeling using MNB, and MNE, why using R, MB, NMB, and RMSE?

Line 246: You already spelled out TCO before.

Line 246-253: Why does the author try not to focus on East Asia? It is okay to do a general performance overview since it covers global regions. But the main point that the author wants to make is largely for East Asia. Only lines 263-268 describe East Asia, consider expanding it.

Line 284: Suddenly describe there are several WRF-Chem outputs in the SI. Consider moving those details in Text S1 to the main text. And if the author really wants to keep it in SI, SI text should be guided first then describe the Figure S3-S6.

Line 297: not so convinced that you can say “localized”.

Line 299: This is basically wrong. First of all, Figure 7 itself, you compare the model with observation, and why did you cite another work..? Observational evidence is already in Figure 7. And second, in fact, those regions that you explicitly wrote have high o3 concentration. The model basically shows significant bias, at most I see from observations going up to “maximum” 42-48 ppb, while your model reproduces ozone concentration more than 78 ppb (I’m very suspicious it is higher than your highest color bar scale). Then how come? It is first of all not supported by your own evaluation and thus I completely disagree that you somehow bring the other work here and cite and suddenly jump on to Figure 8. In fact your global resolution with 60km performs better. For panel g-i, I assume this is averaged all over the stations. If you do statistics over just those complex terrain regions, I don’t expect your statistics to be that good.

Line 305: not so convinced you can use WRF-chem simulation here, in fact your U60km perform way better than WRF-Chem, and how come the logic suddenly becomes the bias is a general problem with coarse resolution?

General comment for Section 3.2 title: It would make much more sense to say “Coupled effects of complex terrain and associated emission” or else. The current title is very confusing.

Line 322: This is such a generalized problem. Regridding emission to the model grid. Nothing new. Consider moving this to SI.

Line 343: what is DEM

Line 370: Author described “Section 3.2 established that mesh refinement primarily improves nighttime surface O₃ simulation over complex terrain through resolved terrain circulations and emission redistribution. We now turn to the diurnal cycle of surface O₃ across the full North China refinement zone, examining how the same resolution increase (~60 km to ~4 km) modifies the photochemical and boundary layer processes that govern the daily O₃ evolution ”

You did not establish that your model improves the nighttime ozone chemistry. See my previous comments.

Line 387: This is entirely wrong. Ren et al. 2022, did a very comprehensive analysis on the photochemical ozone formation regime across many regions in China. And in fact they diagnose it varies significantly by region. Now for the BTH region, they proposed a transitional regime in the 3-3.8 range of FNR. And also for regions that this study uses to argue their ozone regime experience significant spatial heterogeneity. See Ren et al., Figure 4 in their article, VOC-limited, transitional, NO_x-limited are all collocated in the region where this work presents their result. And more importantly, FNR has never been established using nighttime satellite retrieval, and thus, nighttime ozone chemistry is not so simply explained by modeled FNR during nighttime which is even average over many regions.

Figure 1: What is the root cause of the straight line in panel a from U60km? See blue triangle. I think more explanation is needed. Does this mean the U60km has largely failed in reproducing terrain height? But I do not see those in Figure 8b, white line.

Figure 11: consider putting satellite retrieval driven FNR in panel d-f. The surface NO₂ evaluation and model predicted FNR doesn't really help readers to understand the model performance on reproducing photochemical regime of O₃. see my other comments for FNR.

SI Line 22: you cite WRF-Chem paper for general description and development. Both Grell et al. and Fast et al., I do not think they evaluate WRF-Chem in East Asia nor globally.

SI Line 39: For qualitative comparison, it is okay to use different versions of inventory, but authors should at least describe what those differences are, not just write versions that are different. And did you not include biogenic emission in WRF-Chem?

Figure S4: not described at all both in main text and SI. This seems to me WRF-Chem performs way better than U60km.

Figure S7: WRF-Chem..? Used Fig S3-S6 are from WRF-Chem. Figure S7 is from which model? The main text describes it as a VR model. Which one is right? And spatial distribution of ozone looks very weird too. Does it have those contour lines in the urban area? Not even a single hot spot? And it would make much more sense to plot with the same colorbar scale for both day and nighttime ozone. You often compare day and nighttime ozone in the main text.

Figure S9: How does NO₂ perform in general? I believe stations that measure O₃, likely have NO₂ as well. And colorbar for panel a, b and c. unit should use superscript correctly not using "Λ". And also I noticed that resolutions for absdiff and reldiff are not so good. Consider plotting it with higher resolution (at least 300dpi). Also there are grey lines in panels which are not province boundaries, what are those?