Comments on "Tight physics and chemistry coupling improves air quality forecasts: a case study of the 2020 North American wildfires with the Rapid Refresh model coupled to Chemistry (RAP-Chem v1.0)" (egusphere-2026-4786), by Schnell et al.

General Comments

This manuscript presents RAP-Chem v1.0 and evaluates its performance during the 2020 North American wildfire season. The topic is relevant, and the model developments are potentially useful for improving air quality forecasts. However, I found several parts of the manuscript difficult to follow, particularly the presentation and interpretation of the sensitivity experiments. In several places, the interpretation of the results is not clearly linked to the supporting figures and observations, making it difficult to assess the reported improvements from the different coupling developments. In addition, the novelty of some developments is not sufficiently clear, and several statements about model performance, including phrases such as "exceptional skill," "extremely well," and "accurately," appear stronger than the evidence presented in the figures and tables.
I believe the manuscript requires substantial clarification and a more careful interpretation of the results. I therefore recommend major revision. My detailed comments are provided below.

Major Comments

1. The Introduction provides a useful discussion of several chemistry–physics inconsistencies in WRF-Chem and the evolution from RAP-Smoke toward RAP-Chem. However, I found it difficult to distinguish which coupling developments are new or modified in RAP-Chem v1.0 and which are inherited from WRF-Chem, RAP/HRRR-Smoke, or previous studies. It would be helpful to add a short summary near the end of the Introduction that clearly identifies the main new or modified couplings in RAP-Chem v1.0, distinguishes them from previous developments, and clarifies which are specifically evaluated in this study. The technical details can remain in Section 2. This would make the novelty and scope of the present work easier to follow.

2. **Section 3.1:** The meteorological evaluation against NOAA ISD observations (Fig. S3) is presented only for the control simulation. Although the effects of direct aerosol feedback on temperature over Washington (Fig. 6b) and surface radiation (Fig. 7) are evaluated, it would be helpful to also evaluate no_dirfdb against the same ISD observations (e.g., 2-m temperature, dew point, and visibility), particularly in smoke-affected regions. For example, Fig. S3a shows a warm 2-m temperature bias over parts of the western US. Comparing both simulations against observations would help determine whether direct aerosol feedback reduces this bias. This would strengthen the evidence for the meteorological improvements described in the Conclusions.

3. **Table 2:**
   (a) The evaluation in Table 2 is based only on NMB. It would be helpful to provide a more comprehensive comparison of model performance across the EPA regions and experiments. For example, a Taylor diagram showing correlation, normalized standard deviation, and centered RMSE, alongside the NMB values in Table 2, could make the results easier to compare.

(b) I also noticed that the O3 NMB in Region 6 is −13.6% for newmix, compared with −4.8% for both oldmix and enhmix, while newmix and enhmix have nearly identical values in most other regions. Please check this value and, if correct, briefly explain the relatively large difference in Region 6.

4. **Section 3.2:** The negative AOD bias at AERONET sites is attributed partly to the "low bias in simulated PM2.5 (Fig. 1b)." However, Fig. 1b presents PM2.5 observations from EPA AirNow sites, while Fig. 1c presents AOD observations from 20 AERONET sites. It is therefore unclear whether the PM2.5 and AOD biases are spatially consistent. Moreover, PM2.5 is overestimated in seven of the ten EPA regions (Regions 1–7, Table 2). Please clarify how the PM2.5 biases at AirNow sites support the proposed explanation for the AOD underestimation, considering the different spatial sampling and the distinction between surface PM2.5 and column AOD. A comparison with satellite AOD could provide useful spatial context.

5. **Section 3.3:**
   (a) The discussion of the wildfire-affected regions needs clearer links between the reported improvements and the supporting model results. For example, the reduction in O3 in Region 10 during 9–14 September is attributed to lower temperatures and photolysis rates caused by overhead smoke. Although Fig. S5 shows that control better captures the observed O3 decrease than no_dirfdb, it does not directly establish the proposed mechanisms. The temperature response is shown only for Washington (Fig. 6b), and the radiation comparison (Fig. 7) is averaged over all sites, while photolysis-rate changes are not shown. Please provide supporting evidence for these mechanisms and clarify how they explain the simulated O3 response.

   (b) Similarly, the interpretation of the O3-PM2.5 bias correlations in Fig. S8 relies on several proposed mechanisms, including excessive surface insolation, changes in photochemistry, and errors in smoke and precursor concentrations. However, Fig. S8 alone does not establish these mechanisms. Please provide supporting model diagnostics or relevant figure references, or revise the discussion to better reflect the uncertainty in these explanations.

   (c) More generally, the discussion in this section is somewhat difficult to follow because it frequently switches between O3 and PM2.5 and among the three regions. I suggest reorganizing the discussion to address one species at a time across Regions 8–10, with more consistent references to the relevant figures and tables and clearer interpretation of the regional results. Please also ensure that the relevant figures and tables are consistently referenced throughout the results discussion. In addition, since the wildfire-affected regions are central to this case study, please consider moving one or more key figures from the Supplement (e.g., Figs. S4, S5) to the main text to better support the discussion.

   (d) Several descriptions of model performance in this section appear stronger than the results support. For example, O3 in Region 10 is described as "mostly simulated extremely well," although Table 2 shows an O3 NMB of 22.4% for control, the largest regional value, and Fig. S5 indicates that O3 remains overestimated during the September smoke event. Similarly, the descriptions of "excellent PM2.5 predictions" in Region 8 and "relatively good agreement overall" in Region 9 should be qualified in light of the biases in Table 2 and the missed PM2.5 peaks in Fig. S4. The statement that the August PM2.5 peaks in Region 8 are "much better simulated by control" is also not clearly supported by Fig. S4, where the

two simulations show relatively small differences during these peaks. Please clarify or revise these statements to better reflect the model performance.

6. **Section 3.4.1:** The abstract states that the updated mixing "tends to improve predictions across species," but the observational evidence for this conclusion is not sufficiently clear. For example, oldmix and newmix have similar CONUS statistics for O3 and PM2.5 in Table 2, while the changes in regional NMB are mixed, with improvements in some regions and worse performance in others. Similarly, Figs. S9–S10 show generally small and spatially variable changes in the correlations with AIRS CO between the mixing experiments, and these comparisons evaluate CO only. Although the authors discuss surface CO observations and AIRS vertical profiles, it remains difficult to assess whether the updated mixing consistently improves model performance. Please clarify which species and observational comparisons support the conclusion in the abstract, and whether the small, spatially variable changes in the AIRS correlations indicate a meaningful improvement in model performance. The statement that "the correlation with AIRS CO improves in newmix and oldmix" should also be clarified, since oldmix is the reference simulation for evaluating the updated mixing.

7. **Section 3.4.2:** The results show that including MYNN-EDMF subgrid clouds reduces photolysis rates and lowers O3, particularly in the eastern US. However, the simulated cloud fields and their radiative effects are not evaluated against observations. Although Fig. 5d shows closer agreement with observed O3 at a single site (United Ave) on some days, this alone does not establish that the simulated cloud effects on photolysis are accurately represented. Furthermore, Table 2 indicates that the effects on O3 performance vary by region: while the coupling reduces positive biases in some regions, it increases the magnitude of negative biases in others (e.g., Regions 6 and 7). Please provide observational evidence or relevant previous evaluations supporting the simulated cloud representation and its radiative effects or discuss this limitation and the mixed regional effects when interpreting the O3 results.

8. **Section 3.5:** The evaluation of direct aerosol feedback needs clearer quantitative support and a more detailed explanation of the reported improvements.

   (a) The authors state that CONUS-averaged surface O3 shows improvements in MB and RMSE (from 2.4 to 1.4 and from 13.2 to 12.2, respectively), but these statistics are not presented in the tables or figures. Table 2 reports NMB rather than MB or RMSE. Please provide the supporting statistics, clarify whether they are based on hourly or MDA8 O3, and include the regional RMSE values supporting the statement that control has a lower RMSE than no_dirfdb in 7 of 10 EPA regions.

   (b) Section 3.3 reports an O3 overestimate of almost 70 ppb for no_dirfdb in Region 10, while Section 3.5 reports a reduction of approximately 40 ppb due to aerosol feedback "over EPA Region 10." However, Fig. 6c presents averages over Washington state only. Please clarify the spatial and temporal basis of these values and distinguish the model bias relative to observations from the difference between simulations. Although control substantially reduces the O3 overestimation during the September smoke event, Fig. 6c still shows an overestimation during approximately 12–20 September. Please revise the statement that control "accurately" captures the observed O3 reduction to better reflect the model performance.

(c) The authors report that including direct aerosol feedback increases simulated surface PM2.5 over Washington by approximately 50 μg m^{-3} during the September smoke event (Fig. 6a). Although the general effects of aerosol feedback on boundary-layer stability are described, the processes responsible for this difference between control and no_dirfdb are not examined in detail. Please briefly discuss the potential contributions of changes in boundary-layer mixing or transport, and provide supporting model results, such as PBL-height differences, if available.

(d) Figure 7 presents radiation results averaged across all SURFRAD and SOLRAD sites, which may make it difficult to assess model performance under heavy-smoke conditions. The authors state that irradiance biases increase with aerosol loading, but the supporting results are not shown. Please provide evidence for this relationship, particularly during high-aerosol-loading periods, and briefly discuss the remaining positive bias in direct irradiance in control (61.3 W m^{-2}).

Minor comments

1. **Section 2.1:** The description of convective removal is unclear. The Grell–Freitas scheme is described as being used with "convective removal and parameterized aqueous chemistry turned off," whereas Sect. 2.1.2 states that aerosol wet scavenging by convective precipitation is active (conv_tr_wetscav = 1) and occurs within the Grell–Freitas scheme. Please clarify whether these statements refer to different processes and specify which convective removal processes are enabled in the simulations.

2. **Section 2.1:** The abstract mentions online biogenic emissions, but Sect. 2.1 describes anthropogenic, wildfire, sea salt, and dust emissions without specifying the biogenic emission scheme. The Code and data availability section also mentions biogenic emission files. Please specify which biogenic emission scheme is used, whether the emissions are calculated online, and provide an appropriate reference.

3. **Section 2.1.1:** The enhanced mixing is applied when the PBL height is below 100 m, with the enhancement depending on NO emissions and/or FRP. Please briefly explain the basis for the 100 m PBL-height threshold and the NO-emission threshold of 10 mol km^{-2} h^{-1}.

4. **Section 2.2 and Table 1:** Although the experiments are described in the text and Table 1, it is not immediately clear which experiment comparisons are used to evaluate each coupling development. The name phtmix is also somewhat misleading, since the PBL-cloud photolysis coupling is disabled in this experiment. Please consider adding a column or note to Table 1 identifying the relevant experiment comparisons (e.g., control versus no_dirfdb, newmix versus oldmix, enhmix versus newmix, and enhmix versus phtmix) and specifying which coupling development each comparison evaluates. Please also use consistent experiment names throughout the manuscript (e.g., no_dirfdb vs. no_fdb, and *mix vs. *_mix).

5. **Section 3.1, Lines 280–281:** The statement that the visibility and temperature biases "indicate large underestimations in near-surface smoke concentrations" seems stronger than the evidence presented, as visibility biases may also be influenced by uncertainties in aerosol optical properties and relative humidity. Please provide supporting evidence or references for this interpretation, or revise the statement to reflect this uncertainty.

6. **Section 3.2:** Since Table 2 and the discussion frequently refer to EPA Regions 1–10, it would be helpful to indicate these regions in Fig. S1(a) or provide a separate map in the Supplement. This would make the regional comparisons easier to follow, particularly for readers who may not be familiar with the EPA regions.

   The discussion of the regional results could be more comprehensive. Results are presented for all 10 EPA regions, but the text discusses mainly Regions 4 and 5 individually. Other regions with relatively large biases in Table 2 (e.g., PM2.5 NMBs of 89.5% and 53.9% in Regions 2 and 3, respectively, for control) are only described collectively as part of the overestimation in Regions 1–7. A brief discussion of the main regional differences, including possible reasons for these large PM2.5 overestimations, would be helpful.

7. **Section 3.4.1:** Given the relatively coarse vertical resolution of the AIRS CO retrievals and their limited sensitivity near the surface, please clarify whether and how these limitations were considered when interpreting the vertical CO differences between the simulations, particularly within the boundary layer.

8. **Section 3.4.1:** The *mix simulations are described as “in general biased low” for O3 with reference to Table 2. However, the CONUS NMBs in Table 2 are positive for all four simulations (4.3–8.2%), with negative values occurring only in Regions 6–8. Please correct or clarify this statement. In addition, the text refers to Table 3, which does not appear to be included in the manuscript.

9. **Section 3.4.1:** The authors report some improvement in nighttime AIRS CO correlations over urban areas with enhanced mixing, but also note poorer agreement with AIRS CO aloft. Please briefly discuss how these findings affect the assessment of the enhanced-mixing treatment.

10. **Section 3.4.2:** The small O3 increase over the Mojave Desert is attributed to the Apple Fire. However, since phtmix and enhmix differ only in the PBL cloud–photolysis coupling, and relatively few PBL clouds are simulated over the western US, it is unclear how this coupling produces an O3 increase near the fire. Please briefly explain the possible mechanism or clarify the attribution to the Apple Fire.

11. “Grell-Devyani” should be “Grell-Dévényi” (lines 86, 88 and 446).

12. **Section 4:** Some statements in the Conclusions do not fully reflect the findings presented in Section 3. The O3 reduction associated with the PBL cloud–photolysis coupling is reported as 1–5 ppb in the Conclusions, compared with approximately 1–8 ppb in Section 3.4.2. Please clarify this difference. In addition, the Conclusions state that the updated mixing reduces primary pollutant abundances closer to the surface, whereas Section 3.4.1 shows that surface CO is “elevated over urban locations and wildfires” (Fig. 2), including an increase of approximately 5 ppb near Phoenix (Fig. 3) and only very small surface differences near Atlanta (Fig. 4). Please revise this statement to better reflect the spatial variability of the results. The final statement about improved weather and atmospheric chemistry predictions should also be revised to reflect the improvements actually demonstrated by the sensitivity experiments.

13. **Figures 3 and 4:** The text refers to the CO differences as Figs. 3e and 3f, whereas the Fig. 3 caption identifies them as panels (f) and (g), with panel (e) showing water vapor. Please check and correct the panel references. In addition, the water vapor panels (Figs. 3e and 4e) are not discussed in the text. Please briefly explain their relevance to the discussion of the mixing schemes.

14. **Figure 5:** Figures 5b–c show differences calculated as phtmix − enhmix (without minus with the PBL cloud–photolysis coupling), whereas Fig. 2 uses newmix − oldmix (with minus without the updated mixing treatment). Consequently, the O3 reductions discussed in Sect. 3.4.2 appear as positive differences in Fig. 5c. Please consider using a consistent sign convention across these figures (e.g., enhmix − phtmix in Fig. 5) to make the effect of including the coupling easier to interpret.

    There are also two inconsistencies in Fig. 5a. First, Section 3.4.2 states that the cloud field is from phtmix, whereas the figure caption identifies newmix. Please clarify which simulation was used. Second, the text and caption describe a total cloud column, but the color-bar label reads “Surface PBL Clouds” with units of kg kg^{-1}. Please clarify whether the plotted quantity represents a surface cloud mixing ratio, a vertically averaged mixing ratio, or a column-integrated cloud amount, and ensure that the description, label, and units are consistent.

15. **Figure 6b:** The temperature time series are difficult to interpret because the observations and simulations overlap considerably. This makes it hard to assess the reported improvement from direct aerosol feedback. Please consider adding a zoomed-in view of the September smoke event or a panel showing the temperature differences between the simulations.

16. **Figure 7:** The caption states that the diurnal cycle is shown in UTC, whereas the x-axis is labeled “Hour of Day [LT].” Please clarify which time convention was used and correct the caption or axis label accordingly.

17. **Technical corrections**: “RRMTG” should be “RRTMG” (Sect. 2.1); “disassociation” should be “dissociation” (Introduction); “(kg1-)” should be “kg^{-1}” (Sect. 2.1.2).