Review comments on “A WRF-Chem study of the greenhouse gas column and in situ surface mole fractions observed at Xianghe, China. Part 2: Sensitivity of carbon dioxide (CO2) simulations to critical model parameters” by Callewaert et al.

The manuscript presents a WRF-GHG simulation framework tailored to Xianghe, evaluating both column-averaged XCO2 and near-surface CO2. The model reproduces the temporal variability of XCO2 with good skill (r≈0.7), despite a persistent background bias, and robustly captures a July 2019 heatwave anomaly. Near the surface, the model performs well in the afternoon (r≈0.75, MBE≈–1.65 ppm after bias correction), while nighttime CO2 is consistently overestimated (MBE≈6.5 ppm), yielding an exaggerated diurnal amplitude (~22 ppm observed, overestimated by ~4.6 ppm). Sensitivity experiments show that using elevated anthropogenic emission profiles reduces nighttime overestimation and improves diurnal amplitude agreement, with particularly strong improvements in March and July (e.g., amplitude overestimation reduced from ~22.7 ppm to ~1.7 ppm in March; ~14.2 ppm to ~–6.0 ppm in July). These changes have minimal impact on XCO2. Land cover choice and VPRM parameter adjustments also affect nighttime near-surface CO2 and seasonal behavior, with industry and energy dominating anthropogenic contributions and biogenic fluxes providing a moderate summer sink. The manuscript demonstrates strong diagnostic capability (e.g., tracer partitioning, diurnal PBL analysis), and provides plausible explanations for observed/model discrepancies.

The paper is well structured, clearly written, and the experiments are well designed. The tracer framework and sensitivity analyses are valuable, and the case study is compelling. Most important, the authors reorganize the limitations of the model setup and have sufficient discussion around those limitations. A few important clarifications and additions—particularly regarding vertical sampling, transport evaluation, and the vertical distribution of CO2—would substantially strengthen the interpretation and generality of the conclusions.

Recommend publication after minor revisions. The comments below aim to improve clarity, document assumptions, and bolster the evidence base for key claims.

General comments:

1. The striking difference in biosphere contributions between XCO2 and in situ CO2 in Figure 5 warrants explicit analysis of the vertical distribution of simulated CO2 and tracer components. Please consider to include some or all the following items in the revision :
   - Vertical profiles and/or cross-sections of total CO2 and tracer-specific contributions (background, anthropogenic sectors, biosphere) during representative periods (e.g., the July event and a typical spring day).
   - PBLH overlays and stability indicators to relate vertical gradients to mixing state.
   - A cross-section similar to Figure A1 focused on the biosphere tracer, and ideally analogous profiles for anthropogenic tracers, to show how vertical gradients translate into different column vs. surface signatures.

   These figures will strengthen your explanation of partitioning differences and also provide visual evidence for the roles of PBL mixing and emission release height discussed elsewhere.

2. Please clarify the sampling level used to compare model outputs to in situ measurements:
   - Is the comparison at the first model level, interpolated to the instrument height, or sampled as a layer-average?
   - Consider to provide a short sensitivity test on vertical sampling (e.g., first vs. second model level; interpolation to sensor height) to demonstrate robustness. This is conceptually parallel to the sensitivity to emission release levels and may influence nighttime biases in stable conditions.
3. Transport is often the dominant source of bias in CO2 simulations. While you diagnose biases from initial conditions and fluxes, a targeted transport evaluation would help establish confidence in the dynamics:
   - Compare simulated meteorology against observations: near-surface wind speed/direction, temperature, humidity, and especially PBL height (ceilometer/radiosonde/reanalysis if available) if they are available

   Demonstrating good transport fidelity will substantiate your focus on flux and boundary-condition uncertainties.

Specific comments:

1. WRF-GHG has 60 vertical levels. Please report how many reside within the typical PBL (<2 km) over Xianghe and the lowest-level thicknesses. Dense resolution in the PBL is crucial for resolving steep CO2 gradients. If possible, discuss whether vertical resolution could influence nighttime bias.
2. Why was a two-week spin-up chosen? Please clarify CO2 initial conditions (e.g., from CAMS vs. homogeneous/zero fields). If initialized from non-physical fields, demonstrate that the domain (including vertical extent) reaches a dynamically consistent state post spin-up (e.g., by showing domain-mean CO2 convergence, vertical profile stabilization).
3. Just wanted to acknowledge: Good approach to deseasonalize the CO2 time series prior to correlation analysis
4. There is a gap in in situ measurements between July and August 2019, coincident with the XCO2 anomaly analyzed in Section 3.3. Please explain the data gap (instrument downtime, QA/QC filtering, etc.) and discuss any implications.
5. Table 2 reports nighttime and afternoon metrics. Please also include morning statistics (e.g., 08:00–12:00 LT), as morning transition periods are critical for entrainment and can reveal transport/flux issues.
6. Consider combining Table 2 and 3 and using parentheses to show values after bias correction, which will make the presentation more compact and reader-friendly.
7. You apply a mean bias correction derived from TCCON sites in 30–50°N. Please justify this choice. Given strong spatial gradients in XCO2, a site-specific CAMS bias at Xianghe may be more appropriate.
8. The discrepancy between XCO2 and in situ biosphere contributions is puzzling given that other tracers track similarly across perspectives. Please add a figure showing vertical gradients (profiles or cross-sections) of biosphere and other tracer CO2 to substantiate the explanation. A cross-section anchored on the region highlighted in Figure A1 would be ideal.

9. Line 232: Remove the phrase “which is linked to the northern hemisphere’s growing season and increased photosynthesis.”
10. Lines 235–244: Please clarify how XCO2 and its standard deviations were computed for the “before,” “during,” and “after” periods. Define the exact time windows, averaging procedure (hourly vs. daily means), and whether uncertainty reflects temporal variability, sampling, or retrieval error.
11. It would be informative to overlay prevalent wind direction in Figure 6c to connect transport pathways with tracer anomalies.
12. Figure A3: This is a useful figure but appears not to be referenced in the main text. Please add a citation and incorporate its interpretation where relevant.
13. The authors claimed that the high biosphere signal is attributed to sources in the Gobi Desert/Inner Mongolia, but, in Figure A1, winds appear southerly and Xianghe is near a high-pressure center, consistent with elevated temperatures. Please revisit the source attribution in light of the wind fields and temperature pattern shown.
14. Consider adding a vertical gradient plot of total CO2 and relevant tracers to support statements about PBL mixing and vertical distribution.
15. Many PBL/surface-layer schemes can produce sharp near-surface gradients under stable conditions. Please check whether this artifact occurs in your simulation and whether it contributes to nighttime bias.
16. For the sentence stating improved agreement with elevated emissions “particularly in March and July,” please add reasoning or speculation (seasonal stability, emission sector timing, boundary-layer depth) if direct evidence is limited.
17. Consider including spatial maps of VPRM fluxes in Figure A1, which would make the biosphere signal interpretation more transparent.
18. Line 390: Please confirm whether this refers to Table A3 or Figure A3 and correct accordingly.

Overall, the manuscript is strong and close to publication. Addressing the vertical distribution, transport evaluation, and a few documentation gaps will substantially enhance clarity and confidence in the conclusions.