Response to Reviewer 1

We would like to thank the reviewer for their work and their helpful comments and their positive evaluation of our work. Point-by-point responses to reviewer comments are provided below.

We adopt the following notation: Our responses are written in italics and changes to the manuscript are reproduced in ***bold.***

Updates to the manuscript (except for the correction of technical errors such as typos or punctuation) can also be reviewed in the uploaded version of the manuscript with tracked changes.

Main comment

This study investigates synoptic weather system related to CO2 transport by applying PRIM to airborne observations from the ACT-America summer 2016 and winter 2017 campaigns, and aims at quantifying the WRF-Chem model uncertainties associated with specific atmospheric characteristics to improve atmospheric inversion processes.

The present study appears to be relevant, novel and well within the scope of ACP. However, I believe that some improvements could be made. They are summarized here, and they are further expanded in the detailed/line-by-line comments:

1. The manuscript is overall well-structured, but the method subsection that illustrates the PRIM (2.3) should be reworked to be clearer for also readers that are unfamiliar with this method. Furthermore, the section that describes the atmospheric conditions in ACT (3.1) is very short, and I suggest either to expand it (e.g. by highlighting facts that are relevant to the results later presented in the paper) or by integrating it in another subsection.
    - *We agree with the reviewer that this section is rather short as it is not the main result of the manuscript. However, this section provides context to PRIM results and is not technically part of the methods section as it presents results from data analysis conducted for this manuscript. Therefore, we decided to keep the structure of the results section as is and have added the below clarification:*
      "To provide context for interpreting PRIM results, an overview of atmospheric conditions during the ACT Summer 2016 and Winter 2017 campaigns is shown in Figure 2."
2. The manuscript has some typos, and there are phrases throughout the text that could be improved in clarity
    - *We have addressed the indicated comments and carefully revised the original manuscript*
3. This study's results could be made more robust by adding another data analysis method, as some of the results could be due to PRIM's features, rather than being actual properties of the considered atmospheric characteristics. A conventional, simple clustering method would suffice, and this could also be included in the SI. In my opinion this will make the results presented more robust in the case the two methods agree (and if they do not then the cause could be guessed). While I don't think that this aspect is critical (which is the reason for suggesting minor revision), I strongly believe that this aspect should be investigated further.
    - *We concur with the reviewer that this would be a very useful addition to the manuscript that would further strengthen our argument. However and unfortunately, the ACT-America project has ended and we do not have the resources to implement this at this time. We also appreciate the reviewer's evaluation of our results as sound without this additional analysis.*

Detailed/line-by-line comments

Line 15: The introduction is very simple and clear, covers the issue at hand effectively and states the study's goal effectively. I do suggest only a couple of changes in this section.

- *We have addressed the specific comments below.*

Line 56-57: Explanation of what OCO-2 v9 MIP is should be included here.

- *The sentence was modified to increase clarity and flow:* ***""... and to infer systematic underestimation of flux-seasonality in the inversion models examined during the Orbiting Carbon Observatory-2 (OCO-2) version 9 Model Intercomparison Project (Cui et al., 2021, 2022)."***

Line 72: The method subsections 2.1 and 2.2 seem good, but 2.3 needs quite some adjustments in my opinion, as many concepts are not explained clearly.

- *Thank you for the comment. We have addressed the specifics comments for this section below to improve structure and clarity of Section 2.3*

Line 87: "To facilitate our analysis, we exclusively use data from level-leg flight segments". Why would this facilitate the analysis?

- *ACT-America flights executed several distinct flight patterns including profiling spirals and level-leg flights (which make up the majority of data) maintaining a steady altitude above ground. Similar to Gerken et al. (2021), we exclusively focus on level-legs and have excluded profiles to simplify classification into ABL, LFT, and HFT categories. It should be noted that flight planning was conducted with the ABL, LFT, and HFT categories in mind such that level-leg flights were conducted to avoid align with altitude classes and to avoid edge cases. We clarified the text to:* ***"Similar to Gerken et al. (2021), we exclusively use data from level-leg flight segments ... "***

Line 101-102: "with unusually high [CO2], indicative of CO2 point sources, ([CO2]>430ppm)". Could this be justified better? How are you sure that you are not canceling any actual feature when excluding these kind of outliers?

- *The level of 430 ppm was chosen because it was substantially higher than what was typically found in the dataset (see Figure 2 of the original manuscript), such that few data points were eliminated. This processing was necessary because there were several instances when the plane flew in the vicinity of industrial plants of fossil fuel power plants with [CO_2] reaching up to and exceeding 800 ppm.*

Line 111: As far as I know, residuals are usually defined as "observed minus modelled", rather than the other way around

- *We defined residuals in this work in the same way as done in Gerken et al. (2021) to maintain consistency.*

Line 119: Please review this subsection carefully, as terms here appear to be used interchangeably and could result in some confusion.
Line 121-123: "Simple rules ... higher than usual frequency." This sentence is not clear. What are "simple rules about input variables"? What does it mean that a "designated variable of interest occurs at a higher than usual frequency"?
Line 123-125: "PRIM rules ... (Hadka et al., 2015)" This sentence seems a little bit off here, and I suggest to move it later in the section.
Line 126-130: "The PRIM ... inside the box." This is not very clear - what does higher mean value mans? What is a target?
Line 143-149: This example makes things more clear than the explanation what was provided before, so I'd try to restructure the section to present this earlier in order to explain what PRIM is - I also suggest using this toy

example to clarify definitions (e.g. writing here in parentheses what is a target/variable of interest/frequency, and so on).

- *Thank you for these comments. Because all are about section 2.3 we address these jointly. We agree with the reviewer that this section introduces a substantial amount of terminology and concepts that are likely not familiar to a reader without prior knowledge of PRIM. We also agree that there are overlaps in terminology. Unfortunately, this is partially rooted in the original terminology chosen by the creators of PRIM. In response to the comments we have restructured the section as follows:*
    - *We moved the example forward in the section and clarified definitions as requested*
    - *Clarified the definition of the target case*
    - *Restructured this section to emphasize PRIM differences and advantages compared to strict clustering methods*

  Due to the length of this section, we do not reproduce changes here and refer to Section 2.3 in the revised manuscript.

Line 156: Is the threshold 0.75 arbitrary? While sounding reasonable, this choice should be justified.

- *There is no "objective" method to select a coverage level. The threshold of 0.75 was chosen to reflect the trade-off between density and coverage inherent to the PRIM method such that we include the majority of observations for a target-class while excluding extreme conditions. We have clarified the text to:*

  "PRIM output for each case is saved and a coverage level of 0.75,which reflects the trade-off between coverage and density inherent to PRIM and includes the majority of target observation for each case while excluding extreme environmental conditions, is selected for analysis."

Line 171-172: One concern of mine here is related to how the ACT dataset is handled. Shouldn't the PRIM classification capabilities be tested on an external dataset? i.e. determine the parameter subspace for each classes on a subset of the ACT dataset ("calibration/training dataset"), and then apply those rules to the rest of the data and see how they match with the expert designations ("validation dataset" – whose results should be displayed in Figure 3 and referred in this section).

- *Thank you for this comment. The reviewer is correct that it is customary for classification methods to perform a train-test split of the data or to use data the model has not seen before for evaluation to ensure that the model has predictive skill. In this case, PRIM is used as a data-mining method to extract information about the target-cases from an existing dataset, which is different from classification and more akin to clustering where a ground truth is not known. Importantly, we do not claim that PRIM has predictive skill, but that PRIM is able to identify high-density regions associated with the target class within the observational dataset.*

Line 268-272: I'd rather say that PRIM's inability is in line with the hypothesis that upper tropospheric air represents background conditions. The way it was phrased makes it seem like PRIM's inability discern frontal warm and cold sectors is a proof for upper tropospheric air representing background conditions (which cannot be proven this way and would need more evidences).

- *We reformulated the text in line with the reviewer's sentiment to:*

  "PRIM's inability to distinguish near frontal warm and cold sector air from fair weather conditions in the higher free troposphere is in line with the hypothesis that synoptic systems have limited impacts on upper tropospheric air. HFT air would thus represent background conditions (Parazoo et al., 2021; Baier et al., 2020) with respect to CO_2 while terrestrial carbon fluxes and vertical transport associated with synoptic systems act on vertically homogeneous coastal inflows (Sweeney et al., 2015; Campbell et al., 2020) to produce the observed vertical CO_2 gradients."

Typos/unclear sentences

Line 2: “of of” → *corrected*

Line 10-12: “Using the PRIM … are less typical.” Unclear sentence

- *Clarified to:* ***"Analysis of the the parameter space constrained by PRIM, …"***

Line 12: “that that” → *corrected*

Line 60-62: “model … (Gerken et al., 2021).” Unclear sentence

- *Clarified to:* ***"… model biases were strongly related to season and synoptic conditions such that warm sector airmasses near frontal boundaries were associated with larger magnitude model-observation differences compared to fair weather air (Gerken et al. 2021)."***

Line 65: Useful tool? → *corrected*

Line 156-166: “There is notably are much larger”?

- *Corrected to:* ***"There is notably a much larger…"***

Line 191: “1% 1%” → *corrected*

Line 250: “to to” → *corrected*

Line 258-260: “While PRIM … by PRIM” Unclear sentence.

- *Clarified to:* ***"While PRIM had difficulties […], PRIM achieved good separation for near frontal warm and cold sector air."***

Figures and tables

Figure 1: y-axis labels overlap

- *We updated the figure accordingly.*

Figure 4: I think it would be better if y-axis limits were consistent throughout ABL-LFT-HFT

- *Thank you for the comment. During the data analysis stage of the manuscript, we prepared a version of Figure 4 with consistent y-axis. We settled on the current version of the figure, because of the large difference in range between ABL, LFT, and HFT values which made it difficult to discern details in the distributions. We updated the figure caption to alert the reader to the differing y-scale between subplots:* ***"Please note that the y-scale between subplots varies to account for differences in range between ABL, LFT, and HFT levels."***

Figure S1: figure resolution should be higher

- *We updated the figure accordingly*.

Figure 5 and figure 6: a plot like the one in figure 6 feels more useful than the one figure 5 (which could be as well SI in my opinion). Is there a reason why this is just ABL and not LFT and HFT?

- *Thank you for this comment. In our view, Figure 6 provides an integrated view of Figure 5 and is thus of value within the main body of the manuscript. We chose to focus on ABL rather than all levels, because large magnitude residuals associated with fronts are most prevalent within the ABL. This is likely due to the fact that frontal systems integrate the effect of terrestrial carbon fluxes over large areas such that they generate large cross frontal [CO_2] differences, whereas this effect is less pronounced in the LFT and mostly absent in the HFT which is hypothesized to be akin to a continental background.*

References

Gerken, T., Feng, S., Keller, K., Lauvaux, T., DiGangi, J. P., Choi, Y., Baier, B., and Davis, K. J.: Examining CO_2 Model Observation Residuals Using ACT-America Data, Journal of Geophysical Research: Atmospheres, 126, e2020JD034 481, https://doi.org/10.1029/2020JD034481, 2021.

Response to Reviewer 2

We would like to thank the reviewer for their work and comments to improve the clarity of this manuscript. Point-by-point responses to reviewer comments are provided below.

We adopt the following notation: Our responses are written in italics and changes to the manuscript are reproduced in ***bold.***

Updates to the manuscript (except for the correction of technical errors such as typos or punctuation) can also be reviewed in the uploaded version of the manuscript with tracked changes.

This MS used PRIM method to evaluate the model-observation difference during synoptic weather systems, and identified the influencing factors. Honestly speaking, I am good at modeling and observation the GHG concentration and also inversing approach, not a expert of PRIM method, but as a normal reader, it seems vary hard to understand the logical expression and follow authors' main idea. The difference between observation and simulations can be treated mainly caused by both atmospheric transport model and prior CO2 flux, where the atmospheric inversion approach usually give an known uncertainty extent to transport model and attribute the difference of observed and simulated GHG concentration (or enhancement) to spatial-temporal bias in CO2 flux. Compared with use of PRIM method, direct comparisons between spatial-temporal difference between model and simulations can better give us potential reasons and locations of such difference.

- *Thank you for the comment. We agree with the reviewer that direct comparison of spatial-temporal differences between model and simulations is necessary to diagnose specific causes for encountered differences. Our previous work (Gerken et al. 2021) describes the overall structure and statistics of model observation differences and links the occurrence of large magnitude errors to frontal systems. This analysis could be expanded into a case-by-case analysis.*

  Notwithstanding, the this manuscript had a different focus, which we think is also valuable. Specifically, we used PRIM to:

  1. *Examine whether the expert designated groupings (ABL, LFT, HFT) and frontal designations are useful categories for analysis. This is confirmed by PRIM results (e.g. Figures 2+3 of the manuscript).*
  2. *Investigate the association of high magnitude residuals to unusual atmospheric conditions. This expands on results in Gerken et al. (2021), showing that highest magnitude residuals during frontal passage are linked to unusual atmospheric conditions. This highlights potential issues with models (e.g. model tuning towards improving the mean state rather than extremes)during such rare conditions, which nevertheless may be impactful due to the observed large residuals (e.g. Figures 5-6).*

  In summary, we are convinced that the present manuscript is scientifically sound and adds value by suggesting PRIM as an additional tool. Importantly, we do not disagree with the reviewer that other issues such as diagnosing prior flux vs transport error or methods including direct comparison are not also valuable tools to improve inversion modeling.

In general, the description of PRIM method is not well displayed or explained, which (I think) is also vary hard for other audience to follow this MS.

- *In response to comments from both reviewers, we have restructured Section 2.3 that describes the PRIM method. Specifically we have:*
  - *We moved the example forward in the section and clarified definitions as requested*
  - *Clarified the definition of the target case*

- *Restructured this section to emphasize PRIM differences and advantages compared to strict clustering methods*

Due to the length of this section, we do not reproduce changes here and refer to Section 2.3 in the revised manuscript.

the readers may want to know what's the main reasons between observations and simulations during different synoptic weather systems or contributors (values) behind them.

- *We agree with the reviewer, but do not see how this could be addressed within the scope of this manuscript and the manuscript goals outlined in our response to a comment above.*

However, in the 5. conclusion section, no values are displayed that was concluded in this MS, the authors mainly describe the background, meaning of this study and the advantage of PRIM method, which seems unnormal as a research paper.

- *We are unsure how to interpret this reviewer comment. In our view, the reviewer appears tp request the addition of quantitative results to the conclusion, which are contained in the results section of the manuscript and then placed in context in the discussion section. In our view the conclusions are supported by results and present a broader perspective including additional implications.*

there are also some Logical jump that make me confused i.e. line 8 "PRIM results separate winter- and summertime observations as well as observations from ABL, LFT, and HFT with enrichment factors of 4–20", what does with enrichment factors of mean? please explain it at the first time when it appear, and I also can not find how does 4-20 come from in the main text, from the numbers in Table 2, it should be 4.0-20.5 or 4-21.

- *Thank you for this comment. To clarify we added additional information about the PRIM method to the abstract:*

  "Here, we apply the Patient Rule Induction Method (PRIM), a data-mining method to identify high-density regions for a target-class within an input parameter space, to airborne observations [...] Analysis of the parameter space constrained by PRIM, ..."

  We also updated the numeric values in the abstract to ***"4.0-20.5"*** *as requested.*

instead of using description on lines 120-155, the diagram figure that explain how PRIM method works is strongly suggested.

- *In addition to the improved description of PRIM, we updated the manuscript to better highlight the content of Figure S1. The manuscript now reads:* ***"PRIM is further described in the Supplemental Information, and Supplemental Figure S1 shows an example of how PRIM boxes are constructed in a multidimensional space and their relation to density and coverage levels and associated input variable ranges."***

typo as "This association suggests that that PRIM" $\rightarrow$ *Thank you for noticing, corrected*

References

Gerken, T., Feng, S., Keller, K., Lauvaux, T., DiGangi, J. P., Choi, Y., Baier, B., and Davis, K. J.: Examining CO_2 Model Observation Residuals Using ACT-America Data, Journal of Geophysical Research: Atmospheres, 126, e2020JD034 481, https://doi.org/10.1029/2020JD034481, 2021.