We thank Reviewer #2 for the constructive criticism of the Discussion paper. The reviewer's comments are shown below in *italics*, while our point-by-point responses are indicated as un-italicized.

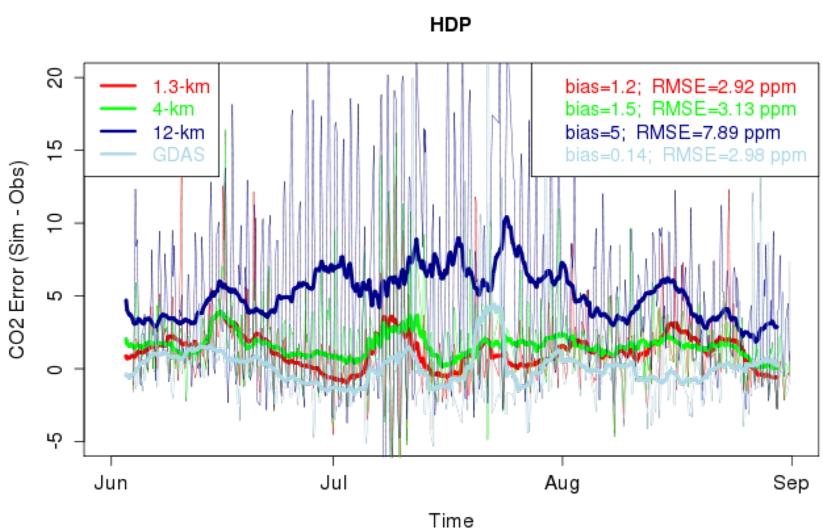
Reviewer #2

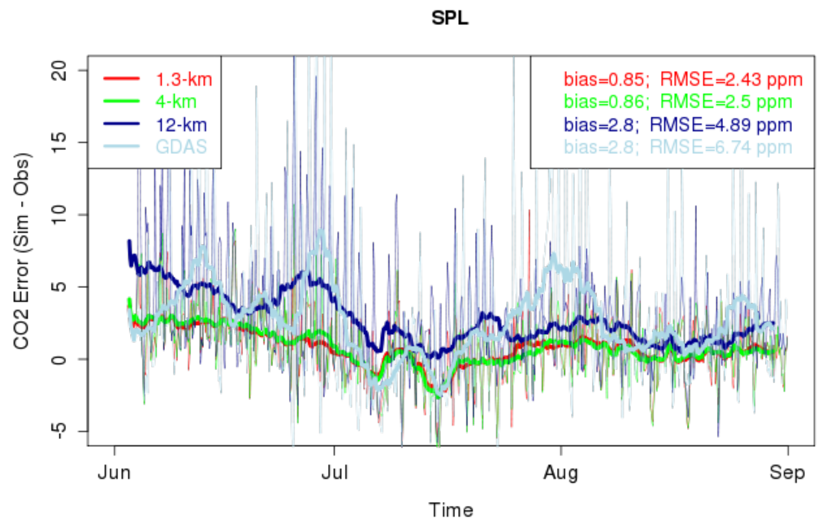
General comments:

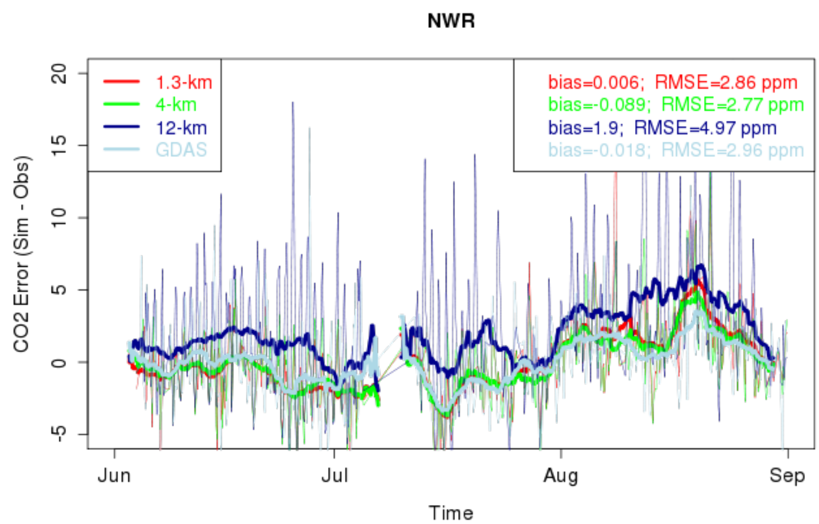
The paper is well-written and the observational datasets, models and methods are properly described and referenced. The authors address an urgent question about the future and current use of observations from mountainous areas or within complex terrains more general. Their analysis of mean daily cycle allows an easy interpretation of mean behaviour, but it would be critical to also investigate if there are episodes (meteorological conditions) for which the mismatch between the different transport model configuration is minimized/maximized. I would suggest to add an analysis of the time-dependent offset of each model compared to the observations.

We agree with the Reviewer that beyond the diurnal timescale, there is day-to-day variability in the model behavior that lead to variations in CO_2 model errors. We have now included in the revised paper the time series of CO_2 model errors at SPL, NWR and HDP (see below) with correlations of the CO_2 errors with various meteorological variables (geopotential height and its E⇔W gradient, U- and V-winds, and the windspeed). The time series plots and a Table of the correlations of the CO_2 errors with these meteorological variables will be added to the Supplement.

Errors at multi-day timescales are the most strongly correlated with different meteorological variables, depending on which is being examined: V-wind, U-wind, and geopotential height gradient for HDP, SPL, and NWR, respectively. The fact that errors are correlated with different meteorological variables depending on the site location points to a complexity that can only be unraveled with a substantial expansion of the paper. This could potentially be a subject for a future paper. This complexity is in contrast to the average diurnal biases that can in large part be linked to the underlying resolution of modeled terrain, which the current paper focuses on.







In general it would be valuable to have more quantitative results and a discussion how generalizable the findings of this study are for other hilly/mountainous areas.

We do attempt to broaden the scope of this study. In particular, we outline various approaches in the Discussion section with regards to making use of mountaintop CO_2 data that can be considered for other mountainous areas as well. However, we believe detailed quantitative results for other mountainous areas would require dedicated modeling efforts. This need for dedicated efforts for individual mountainous sites is hinted at by differences in results between the three sites examined in this study: HDP, SPL, and NWR (Figs. 3, 4). We have pointed to the contrast in elevation at the mountaintop site to the surrounding terrain as a key factor in explaining differences observed at the three sites (Figs. 7, 9, 11). Hopefully these results would stimulate other researchers around the world to also examine the same approaches and factors mentioned in this paper, but the dedicated modeling efforts necessary to do so would be outside the purview of this paper.

Overall this paper also does not fully address the question of "constraining carbon fluxes", but rather how well different model setups can reproduce the atmospheric concentrations of CO2.

To be able to really judge if the models (even the best, 1.3km resolution) are able to e.g. distinguish different prior carbon flux estimates the authors would need to perform a sensitivity study using multiple carbon flux data sets and demonstrate a significant impact at the three sites. Potentially one can consider this study rather a step towards a better use of such data rather than already addressing the question of regional carbon fluxes.
We agree that the paper is the first step towards addressing the question of regional carbon fluxes. However, we believe that it is a critical first step that needs to be taken in order to use mountaintop CO_2 data to constrain regional carbon fluxes.

After the above and below comments are addressed I would definitely recommend this study for publication in ACP as it will help to better understand limitations of such observations and were future model developments should/could be focussed to eventually be able to constrain carbon fluxes in such regions.

Specific comments:
Line 44ff: The claim that nearly 70% of the earth land surface is covered by hills or mountains needs to be better validated. This surely depends on the definition for hill or mountain, which is not given here and the cited publication is hard to access (due to the journal it was published in) and the journal has an impact factor below 1.
We thank the Reviewer for pointing this out. The claim of ~70% of the Earth's land surface as being covered by hills or mountains was attributed to Rotach et al. [2008], but we traced this claim to a book written in the German language. Furthermore, there appears to be no clear definition for what is meant by "hills". Therefore, we decided to revise the statement to just referring to "mountains", which cover about one quarter of the Earth's mountains, citing a readily-accessible UNEP report for this purpose (Blyth et al., 2002).

Blyth, S., B. Groombridge, I. Lysenko, L. Miles, and A. Newton, Mountain Watch: Environmental Change and Sustainable Development in Mountains, UNEP World Conservation Monitoring Centre, 2002.

The authors also mention that carbon fluxes in complex terrain need to be better understood to quantify carbon flux. It seems you are implying that all mountainous or hilly areas are (too) hard to model?
We were not necessarily implying all mountainous or hilly areas are difficult to model. We were suggesting that because mountainous areas cover a large fraction of the Earth's land surface and significant amounts of biomass can be found in mountains (e.g., Fig. 1), a better understanding of this under-sampled region is necessary.

Line 193: The authors refer to a previous publication – nonetheless the key parameters e.g. vertical mixing scheme used should be explicitly given in this publication (e.g. by adding a table in this section).
We have revised the paper to include key parameters regarding the WRF configuration. The revised sentence reads:
"Comprehensive testing of different WRF settings have been carried out as part of a previous publication (Mallia et al., 2015), and these settings were adopted here: i.e., the MYJ, Grell-

Devenyi Ensemble, and Purdue Lin schemes for parameterizing the planetary boundary layer (PBL), cumulus convection, and microphysics, respectively."

Line 216: Please specify if the system allows for two-way nesting or not
Following our testing in Mallia et al. (2015), we have implemented two-way nesting within WRF. This is clarified in the revised text in Sect. 2.2:
"For this study, we ran WRF in a two-way nested mode centered between Utah and Colorado where the RACCOON sites are located (Fig. 2)."

Line 277: A "fix" is mentioned, but not explained at all. Please consider giving a brief description here rather than referring to the supplement. It seems the daily cycle has just been shifted or were there any more complicated adjustments performed?
Yes—only the diurnal pattern has been shifted while preserving the 24-hour integrated carbon flux. We have added more information for the reader in the revised text:
"For this paper, we implemented a fix that removed this artifact by detecting these reversed diurnal patterns, adjusting them while preserving the 24-hour integrated carbon flux. See the Supplement and Fig. S3 for details."

Line 304: Please consider referring to table 1 here so the reader can easily find the height CT data was extracted from.
We have added a reference to Table 1.

Line 503ff: The first question is repeated here "How can mountaintop CO_2 observations be used to constrain regional scale carbon fluxes, : : :." But the 5 subsections following rather discuss IF such data can be used or how they can be better used. It remains unclear if there is a definitive answer on how to use them.
Due to the fact that the errors incurred depends on the model resolution, the relationship of the mountaintop site relative to surrounding terrain and emissions, and the quality of the prior fluxes, the definitive answer depends upon each specific situation. Thus we were hesitant to suggest an answer that overgeneralizes. Also see above for the response regarding statements for other mountainous areas.

Line 530ff: Choosing the appropriate model layer to extract CO_2 does indeed introduce a significant additional degree of freedom. The authors suggest other parameters to avoid creating a fudge factor but do not give specific advice here on which tracers could be useful (^{222}Rn?). Meteorological data is mentioned but looking at table 2 it seems not at all clear that this would be good parameter or what a cut-off would be. Could you suggest how a suitable proxy could be found?
These tracers and meteorological data were mentioned later on in the Discussion section:
"We recommend additional tracers to be measured in conjunction with the mountaintop CO_2 sites. For instance, combustion tracers such as C^{14} and CO (Levin and Karstens, 2007) have been measured alongside CO_2 at mountaintop sites in Europe. Another promising tracer is Rn^{222} (Griffiths et al., 2014), which provides a measure of surface exchange and would help provide constraints on the exchange of air measured at the mountaintop with the surface. Co-located meteorological observations—whether in-situ or remotely-sensed (e.g., radar, sodar, lidar)—to probe atmospheric flows and turbulent mixing would also be of significant value in helping to interpret the tracer observations (Rotach et al., 2014;Banta et al., 2013)."

Line 581ff: Here the authors report on the practice of not using mountaintop data but it is unclear how this is linked to this specific study as no suggestion is made how to e.g. better use Schauinsland data. This section should be considered for the introduction to motivate why Approach 1, 3, 4, 5 need to be improved.
We thank the Reviewer for this valuable suggestion. We have followed the Reviewer's suggestion and moved Approach 2 ("Reject mountaintop data") to the Introduction to provide further motivation for the paper.

Line 595ff: When setting up an inversion system it is common (good) practice to assign proper model errors. This seems not specific to this study and the authors fail to give an estimate of the model error for the three sites discussed here. Please consider removing this section or giving quantitative results for the sites and models investigated here. Of course, the model data mismatch calculated here also depends on flux errors, but the authors can surely use this study to give an upper limit of this combined error (and the difference for different model resolutions).
We agree with the Reviewer that quantitative results for the sites and models investigated here would be useful. An estimate of these errors is the RMSE (root-mean-square-error) calculated for each site and model setup. We now direct the reader to the RMSE values shown in the top right-hand corner of the time series plots found at the beginning of this Response.