We thank Reviewer #1 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 #1

General Comments

The WRF-STILT model results at 4 and 1.3 km resolutions compared well with observation. But the differences between WRF-STILT model results and observations became unexpected large at 12 km resolution (Fig. 3), with poorer agreement than the much lower resolution models (GDAS-STILT and CarbonTracker, both at 1 degree resolution). The problem seems to be the modeled PBL. The 12 km WRF-STILT model mean PBL shown in Figs. 7, 8 and 9 are quite different from the WRF-1.3, WRF-4 and GDAS-1 deg results. Similarly the GDAS-STILT results exhibit odd behavior (see below).

The results from the modeled PBL scheme actually do not differ appreciably in the three WRF configurations. This was made much clearer by our addition of new figures for the Discussion paper (Figs. 8, 10, 12) showing the time series of PBL heights (and trajectories) with respect to above-ground level (AGL). The key difference is the smoothing of the terrain in the coarser simulations. We have added new text to clarify this point:

" An alternative perspective is to view the trajectory and PBL heights relative to the ground surface ("AGL") instead of above sea level, at each time step backward in time from the receptor (Figs. 8, 10, 12). These figures highlight the fact that while PBL dynamics in the three WRF configurations are similar, the heights of the trajectories relative to the PBL height differ. The trajectory exits above the nocturnal PBL one hour backward in time, on average, while the WRF-12km trajectory spends several hours within the PBL.

The difference in the trajectory behavior can be explained by the differing terrain. In mountainous terrain, PBL heights generally follow the terrain elevations, albeit with attenuated amplitude (Steyn et al., 2013). Thus in WRF-1.3km and 4km, the more highly resolved terrain produced shallow nocturnal PBL height that descend in the valley (Fig. 7) while the corresponding trajectory hovers above it. Viewed relative to the ground surface (Fig. 8), the trajectory originating from HDP appears to have exited above the nocturnal PBL in WRF-1.3km and 4km. In contrast, due to the significantly "flattened" mountains in WRF-12km and in GDAS, the PBL heights exhibit less spatial variation near the mountaintop receptor, since the terrain itself was smoothed. Consequently, WRF-12km trajectories, unlike the WRF-1.3km or -4km cases, travel closer to the ground surface, within the nighttime PBL, even as it is advected away from the three RACCOON sites (Figs. 7, 8). This resulted in stronger nighttime footprints in WRF-12km as seen in Figs. 4 and 5. Another effect of the proximity of the air parcels to the model's ground surface is the slower windspeeds from surface drag, causing the air parcel trajectories to remain close to the 3 sites until the previous day; for HDP and SPL, the mean trajectories spiral toward the site at the surface, following an "Ekman wind spiral" pattern (Holton, 1992). In WRF-1.3km and WRF-4km, the measurement sites are at significantly higher elevations above the resolved valleys in the area surrounding the sites, and the air parcels are found above the shallow nocturnal boundary layer hugging the valley floor, on average (Fig. 7)."

This raises many questions:
(1) Is there a problem with the modeling of PBL in the WRF-12km model?
No--please see response above.

(2) Is there a problem with the modeling of PBL in the GDAS-STILT model (Figs. 10, S10, S12)? Figs. S10 and S12 show HDP and NWR PBL much higher (5000m) compared to PBL from WRF-STILT runs.
We believe the considerable difference in behavior within GDAS is due to both the significantly coarser vertical and horizontal resolutions.

We now explain the impact of this coarse resolution:

"Another noticeable difference in GDAS-ASL trajectory was the significantly higher daytime PBL heights (Figs. 8, 10, 12). We suspect this is because of the greatly reduced vertical resolution within GDAS (23 levels versus 41 levels in WRF): since STILT diagnoses the PBL height to correspond to a model level, a higher PBL height was chosen for GDAS because of the thicker vertical level. Another subtle artifact of the coarse resolution within GDAS can be seen in the anomalously low daytime PBL height just in the vicinity of HDP (Figs. 13, S10). It appears that the GDAS model set an entire $1^{o}\times1^{o}$ grid box near HDP to be water body (the Great Salt Lake), thereby suppressing the PBL height."

Fig. 10 appears mislabeled and not consistent with the discussion in the text, therefore very difficult to understand, see specific comment.
We believe that the impression of Fig. 10 being mislabeled likely resulted from the Reviewer looking at an older version of the manuscript instead of the version published as the Discussion paper. See below ("Specific Comments") for details.

(3) How realistic are the PBL results in these models (WRF-1.3km, WRF-4km, WRF-12km, GDAS)? How were the PBL results evaluated (for mountaintop conditions)?
(4) How realistic is PBL model for mountaintop simulations? How does the PBL model account for the difference between flat plain and mountainous conditions (as the model domain contains such surface conditions)?

The authors should provide more technical details on the PBL model (e.g. dependence on atmospheric conditions and topographic variations, daytime and nighttime variations), particularly how it changes for the different (met data) models (to cause the unusual results for WRF-12km and GDAS) in this study. Then present more analysis and discussion on the uncertainties or errors in the PBL and consequently footprints in this study.

The aforementioned questions (3) and (4) are linked and currently difficult to answer. How to carry out PBL simulations in mountainous terrain and how to evaluate them very much remains unsettled. The requisite meteorological and tracer observations (in addition to CO_2) are limited for the sites examined in this paper. We fully recognize this difficulty and have already included a recommendation for additional observations in future mountaintop sites:

" 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)."

Specific comments

Lines 413-414: 'Due to the significantly "flattened" mountains in WRF-12km and in GDAS, the PBL height exhibits less spatial variation.' This is not true for WRF-12km in Figure 7.

This actually still holds for WRF-12km in Fig. 7, but is true, strictly speaking, only for the key area near the receptor. We have now clarified this point to:

" In contrast, due to the significantly "flattened" mountains in WRF-12km and in GDAS, the PBL heights exhibit less spatial variation near the mountaintop receptor, since the terrain itself was smoothed."

Lines 469-472: 'Focusing on the three-dimensional plots at the hours of 0800 and 1100 MST (Fig. 10), when the simulated peaks are found at SPL and both NWR/HDP, respectively, the peaks coincide with times when average trajectories are found within a relatively shallow morning PBL.' This statement seems to conflict with Fig. 10, which shows the HDP site with a deep morning average PBL. There seems to be error(s) in Fig. 10. The HDP plot domain is the same as shown for NWR in Fig. 9. The plot domain for NWR is as for HDP in Fig. 7. The plotted curves for the two upper plots with nearby domains seem to show drastic difference for PBL, the PBL for the HDP plot is definitely not 'shallow morning PBL' at _6000m.

Reviewer #1 appears to be looking at an earlier version of paper and not the version published as the Discussion paper. Lines 469-472 have different text now, and we believe the text in the published Discussion paper matches what is shown in the figures.

Lines 734-736: 'Red portions of the trajectory refer to the nighttime (1900_0700 MST), while pink portions indicate the daytime (0700_1900 MST).' The red and pink portions are hard to distinguish, use more contrasting colors.

We thank the Reviewer for pointing this out. We have followed the suggestion and changed the colors to show more contrast.

Lines 738-739: 'Fig. 8 Similar to Fig. 7, but for the Storm Peak Laboratory (SPL) site.' The blue PBL line appears to have dark and light portions not explained. Again, the

contrast is hard to distinguish, use more contrasting colors. This may apply to Figs. 7, 9, 10 too.
We thank the Reviewer for pointing this out. We have followed the suggestion and changed the colors to show more contrast.

Figure 7 (also 8, 9, 10, and corresponding figures in supplemental material): add color scale to facilitate figure comparisons.
We do not feel that the color scale is necessary, since the colors represent elevation and the same information is also represented by the z-axis, as well as the three-dimensional terrain representation. Furthermore, these figures already incorporate a lot of information, so adding another colorscale could make the figures look too "busy".

Supplemental material

Figure S3: it is not clear if the flux reversal correction takes into account the longer daytime than nighttime during the summer.
Yes--the longer daytime can be accommodated in our flux reversal correction algorithm. The algorithm preserves the area between the flux time series and the flux = 0 line. So to the extent that the diurnal pattern in the uncorrected CarbonTracker fluxes shows a longer period of release or uptake, this would show up in the corrected fluxes too. In other words, the algorithm preserves the relative length of carbon uptake or release in the original CarbonTracker fluxes, but any nighttime uptake would be moved to the daytime.

Figure S5: 'CO.obs...' to 'CO2.obs...'
We thank the reviewer for pointing out this error. This figure has now been fixed.

Figs. S11, S12: PBL for GDAS higher than 6000m seems unusually high.
We believe the high PBL within GDAS is mainly due to the coarse vertical grid spacing. We have added an explanation of this point:

"Another noticeable difference in GDAS-ASL trajectory was the significantly higher daytime PBL heights (Figs. 8, 10, 12). We suspect this is because of the greatly reduced vertical resolution within GDAS (23 levels versus 41 levels in WRF): since STILT diagnoses the PBL height to correspond to a model level, a higher PBL height was chosen for GDAS because of the thicker vertical level."