Response to Anonymous Referee #2

We thank the referee for their constructive comments on our manuscript.

Below, all comments are repeated in italics, followed by our response typeset upright. Changes to the manuscript are highlighted in blue colour.

General comments

Overall, this is an excellent and exciting paper. It demonstrates a novel application of UAS for atmospheric science and adds to an exciting literature concerning the new horizons this sampling platform offers. It is a proof-of-concept study, intended to demonstrate the potential use of UAS in CO2 biospheric respiration measurement. It identifies the challenge and importance of nocturnal respiration measurements and the gap that EC methods (and limited spatial scale of chambers) cannot fill. It proposes a mass balancing approach suited to night-time measurement, taking advantage of the assumption of a stable boundary layer. Given that this is an initial study, intended to open up a new direction in this field, some of the questions about the validity of the flux method itself (see specific comments) should be seen in that context, i.e. that this paper identifies a problem and suggests an innovative approach that can be built on and refined in future work. I believe the paper would be of great interest to readers of AMT and the quality of presentation, figures etc is excellent. I specifically praise the way the authors have carefully considered the specific challenges of rotary UAS sampling (i.e. the influence of downwash, instrument response time, etc) and proposes a solution to only use descent profiles to avoid disturbance and take into account response time. These factors are often overlooked and this paper serves as excellent guidance. The paper also compares UAS results with chambers and raises some interesting questions. I do have some important comments though. These concern the UAS flux approach and the way in which surface footprint and vertical mixing scales have been derived (see specific comments). I hope that these comments can be addressed or answered in a revised version of the paper. I see this method as something that can be improved upon in future work and perhaps the most important edits to the text could highlight the remaining uncertainties and challenges to the approach.

We appreciate that the referee sees our work as a valuable contribution to the scientific community. We share the view that the pilot study presented in our manuscript cannot fully answer all questions about the accuracy of the derived fluxes and associated footprints, but is a foundation that future works can build upon.

Specific Comments

1/ Use of STILT to define footprint: I have sympathy with the approach and I do not have a good alternative solution to accurate night-time footprint evaluation, however Lagrangian trajectories near to the surface are known to be subject to significant error/uncertainty. Surface trajectories tend to hug the surface and follow (typically) the 10 m wind vector suggested in the reanalysis met data used to drive the model (in this case ECMWF 0.1 IFS), i.e. upward/downward motions are supressed. How many vertical levels does this version of ECMWF have and what resolution in the vertical domain used in the study?

The vertical resolution of the meteorological data is specified in the original manuscript at the end of Sect. 3.4: 'The vertical resolution of the meteorological data depends on height above ground. The lowest layer extends from the ground to 10 m height, the following 5 layers extend from the top of the next lower layer to 31 m, 55 m, 80 m, 108 m and 138 m, respectively.' The total number of layers is 89. In the revised manuscript we add the likewise relevant fact that 'The temporal resolution of the ECMWF IFS data is 3 h.'.

In order to test the hypothesis that vertical motions are suppressed in the transport model, we plotted the height of 200 out of the 10 000 particles as they travel backwards from the Fendt site. As an example we chose the particles released inside the nocturnal boundary layer at $z = 10\,\mathrm{m}$ and particles released above the nocturnal boundary layer at $z = 100\,\mathrm{m}$ on July 6 21:00 UTC (Fig. 1 and 2) and July 9 21:00 UTC (Fig. 3 and 4). Note that these are a small subset of the particle trajectories based on which the footprints presented in Fig. 13 and 14 in the discussion paper were calculated.

A prominent feature of the trajectories for 6 July is the absence of a stable boundary layer for travel distances greater than 5 km, which approximately corresponds to the time period before 20:00 UTC. In contrast, the trajectories for 9 July indicate a stable boundary layer throughout the time period between 18:00 UTC and 21:00 UTC. This would mean that accumulation in a shallow layer near the ground would have happened during a period of 1 h on 6 July and during a period of 3 h on 9 July. While the profiles measured by COCAP (Fig. 6 and 7 in the discussion paper) do suggest a weaker inversion layer on 6 July, a threefold difference in the accumulation seems too high. Additionally, the measurements at the 9 m mast presented in Fig. 8 in the discussion paper show that CO_2 has accumulated near the ground as early as 19:10 UTC on 6 July, an observation that cannot be explained without a stable boundary layer. This demonstrates that meteorological datasets are an imperfect description of the atmosphere and more generally underlines the referee's point that transport modelling in the nocturnal boundary layer is subject to considerable uncertainties.

The hypothesis that STILT suppresses vertical motion, however, is clearly refuted by Fig. 1 through 4. Even within the strong inversion on 9 July the particles are frequently redistributed between the two lowermost layers of the meteorological dataset.

The approach used here is to release 10000 particles per time-step at very small increments in height up to some assumed mixing height (see comment below). I would raise some concerns with this approach. Perhaps an improvement may be to run STILT in ensemble mode – to perturb each trajectory with some assigned uncertainty (diagnosed from the ECMWF data or obtained by drone-based wind measurement variability in future) to the wind vector to examine advective uncertainty - Section 5 nicely acknowledges the future role of wind measurement. A set of releases at different heights is unlikely to recreate any meaningful 2D footprint as the trajectories will cluster along one singular wind vector (as Figure 13 tends to show) extracted from the ECMWF model grid (0.1 deg is $\sim$ 10km of fetch after all), whereas an ensemble may at least give a better qualitative indication of the possible extremes of the fetch/footprint. This is likely to be the biggest source of uncertainty in any Lagrangian budgeting approach and I think it may be important to state this in the paper, even if an ensemble approach is not used in any revision.

The STILT model is stochastic in the sense that different realisations of vertical turbulent transport are used for each of the particles released, which is illustrated by Fig. 1 through 4. The particles reside at different vertical levels of the meteorological dataset for different periods of time, hence experiencing different advective transport. The horizontal dispersion resulting from this mechanism can be seen in Fig. 5. Over a travel distance of 10 km these 10 randomly chosen trajectories for 6 July and 9 July spread out over 1.5 km and 5 km, respectively. We do not have a reference at hand to compare to, but this spreading seems reasonable for a statically stable boundary layer. STILT makes use of the atmospheric stability as well as the wind variability between grid cells, vertical levels and time steps. We do not know which further uncertainty could be extracted from the meteorological dataset alone. On the other hand we fully agree that wind measurements alongside the NBL profiling would provide an estimate of the model error and could be used to determine the uncertainty of the footprints.

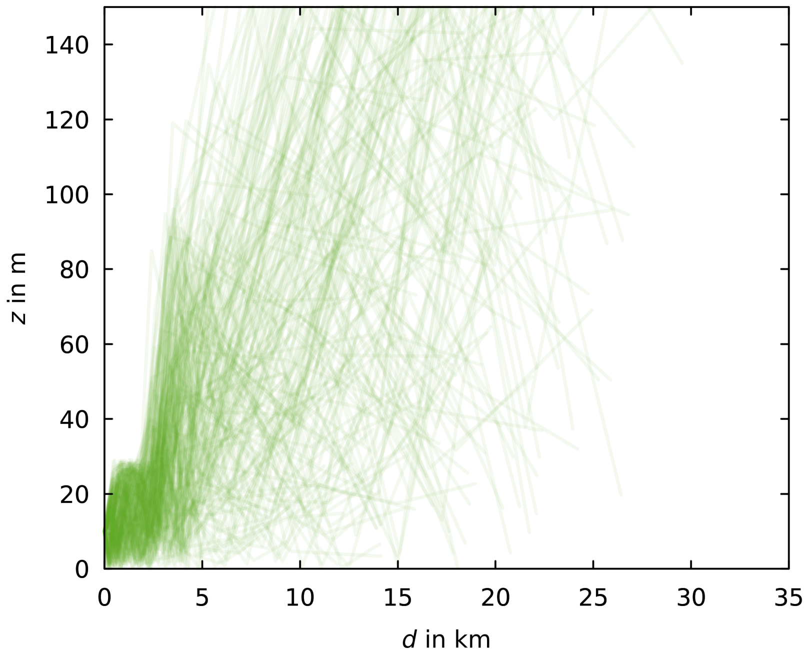

Figure 1: Height z versus distance travelled d of 200 particles released on 6 July at 21:00 UTC at a height of 10 m as they travel back in time until t_0 =18:00 UTC

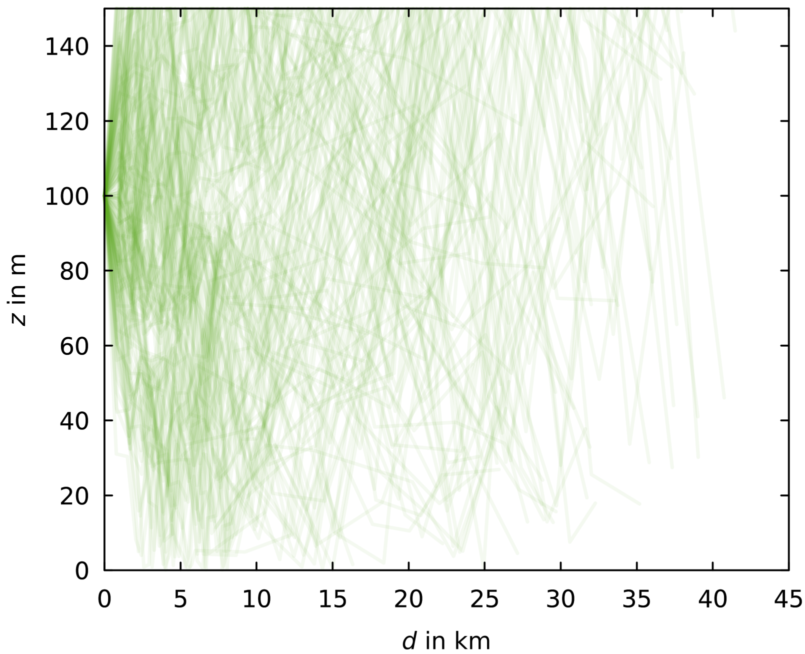

Figure 2: Same as Fig. 1, but for a release height of 100 m

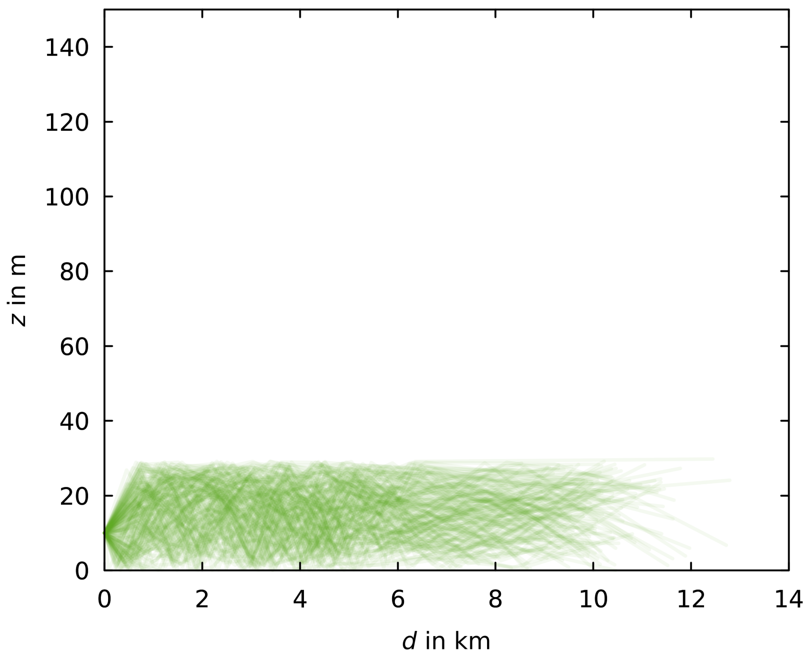

Figure 3: Same as Fig. 1, but for particle release on 9 July at 21:00 UTC

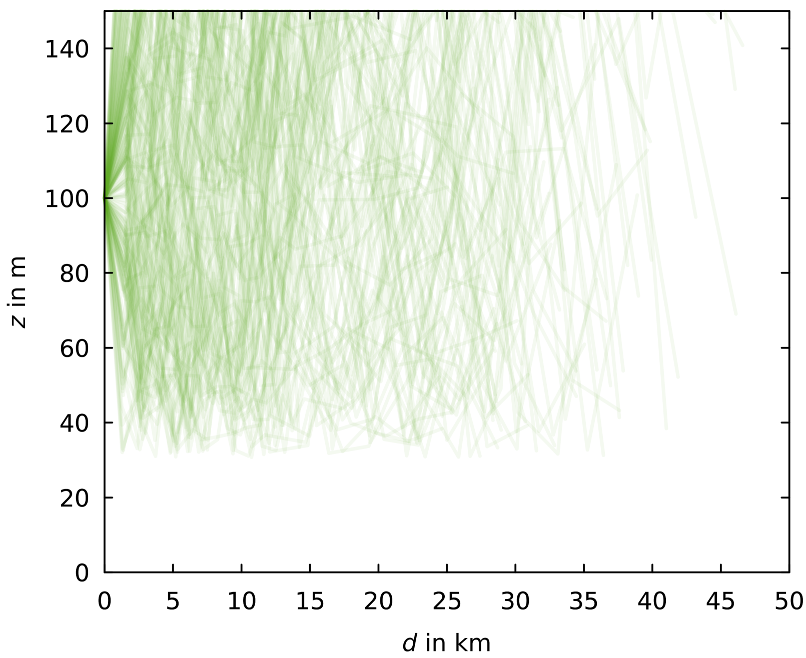

Figure 4: Same as Fig. 1, but for particle release on 9 July at 21:00 UTC at a height of 100 m

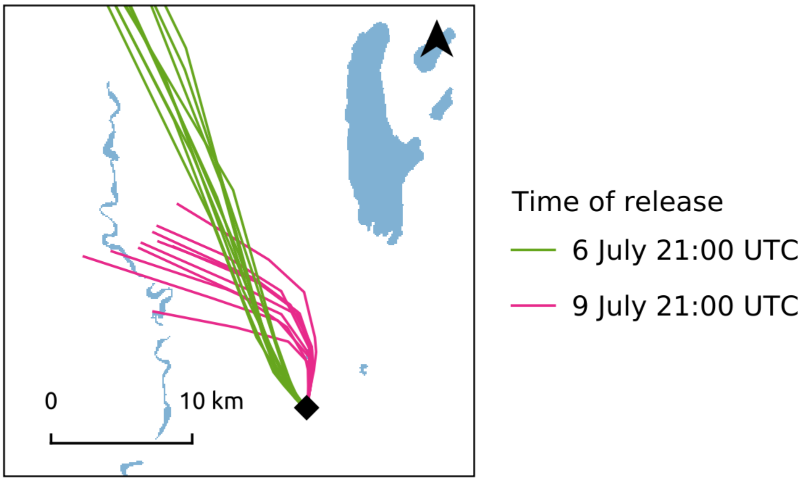

Figure 5: Trajectories of particles released at a height of 10 m at 21:00 UTC on two different days as they travel backwards until t_0 =18:00 UTC of the same day

I realise that footprinting is extremely difficult but it would be useful to acknowledge just how difficult and error- prone it is. The same is true of EC footprints in topographically-variable environments of course.

We agree that the calculation of footprints is challenging, especially at the scale relevant for the NBL budgets. The discussion paper lists the limitations of our approach in Sect. 3.4, 4.7 and 5. We add to Sect. 4.7: 'Variability of the horizontal wind component within a grid cell and on time scales below three hours is neglected, possibly resulting in an underestimation of the footprint size. Likewise, terrain features that are smaller than a grid cell are not represented in the meteorological model.'

2/ P.12 line 1 – why is it expected that "Surface fluxes are expected to be diluted into a column that extends from the surface to 1/2 this height in each time step"? This seems rather arbitrary? Why is this expected? How was it derived from ECMWF data?

The respective paragraph describes how the STILT model represents vertical mixing. Like any model, it uses a simplified description of reality. In the case depicted in Fig. 3, for example, the STILT model has determined a boundary layer height of approximately 30 m. Whenever a particle resides below 1/2 this height, i.e. below 15 m, it is considered to be in contact with surface fluxes and a finite sensitivity to surface fluxes is assigned to this point of the particle's trajectory. The boundary layer height is calculated using a modified Richardson number method (Lin et al., 2003). The threshold 1/2 has been chosen for computational efficiency; thresholds between 10 % and 100 % of the boundary layer height have been found to have insignificant effects on the footprints (Gerbig et al., 2003). We revised the description in Sect. 3.4: 'To do so, the height up to which mixing occurs is estimated from the meteorological data using a modified Richardson number method (Lin et al., 2003). Surface fluxes influence air parcels within a column that extends from the surface to 1/2 this height in each time step (Gerbig et al., 2003).'

In a stable night-time boundary layer, what is the vertical mixing process assumed to reach this quantitative mixing-height value? In stable NBLs, vertical dilution is dominated by diffusion with some small residual vertical turbulence, e.g. the "fanning" Pasquill stability class.

The profiles depicted in Fig. 6 and 7 illustrate that vertical mixing does take place within a stable boundary layer. Profile #20, for example, taken on 9 July at 21:10 UTC, exhibits pronounced gradients both in virtual potential temperature and CO_2 dry air mole fraction up to a height of $z = 50\,\mathrm{m}$. Both rapid radiative cooling and the emission of CO_2 take place at the ground, so these gradients are the signature of vertical mixing. Molecular diffusion is slow; a CO_2 molecule in air at a temperature of $20\,^{\circ}\mathrm{C}$ travels on average $1.6\,\mathrm{m}$ in a whole day (Karion et al., 2010). The main mechanism for vertical mixing in the stable NBL is turbulence generated by vertical wind shear due to friction at the surface.

Given that assumed vertical mixing timescales (and horizontal footprint) are key to deriving flux per unit area in the footprint using the proposed method, these quantities are key. This (and comments below and above) cause me to start to question the overall flux method as it stands.

This is a misunderstanding. The fluxes are derived from the NBL budgets by means of Equation 11 in the discussion paper, which does not contain any quantity that depends on the calculated footprint. The flux footprint is calculated from Equation 14 and is unitless. The value in each grid cell of the footprint is a measure of how much the flux in this grid cell influenced the flux derived from the NBL budget at the Fendt site.

Wouldn't a much more conceptual and simple approach simply be to look at the temporal gradient in CO2 throughout the NBL throughout the night and assume a fetch equivalent to the length scale of advection over that timescale (e.g. treating the NBL like a large-scale vented flux chamber, so long as footprint can be defined)? Such a concept would negate a diagnosis of any spatial heterogeneity in flux (arriving at a bulk net flux for a defined airmass volume) but I don't have any confidence that the proposed approach can do anything better than this in reality (without a fleet of drones that is). In summary, I'm not convinced that any useful 2D footprint can be obtained, so averaging the accumulated NBL mass over any surface area is problematic, so a simplified NBL bulk net flux approach may be more meaningful?

The fluxes derived from NBL budgets are spatial and temporal averages, but this averaging takes place due to the transport in the atmosphere and the physical accumulation in the NBL. We do not apply averaging in the data processing. The role of the footprints is flux attribution to an area, not flux calculation. We agree that transport modelling is subject to errors and it is especially challenging in case of a shallow stable boundary layer. However, as the ECMWF IFS data used to drive the STILT model contains horizontally, vertically and temporally resolved wind vectors, we are confident that our approach yields a more realistic footprint than an estimate based on a single mean wind vector.

3/ Other sources of flux uncertainty: These include the assumed background CO2, any variability in upwind sources of CO2 (i.e. variability in the background airmass entering the footprint over the time frame of the measurements), measurement error/precision, wind speed and direction variability etc. Section 4.2 and 4.6 addresses measurement error nicely and explores sensitivity, but not the other sources of flux error. Perhaps it would be good to note these in the paper, even if they cannot be determined or budgeted in this work, so that others following or improving on the work are aware.

The two sources for uncertainty of the assumed background CO_2 dry air mole fraction are measurement error (covered by sensitivity check 1) and spatial variability. The flux error stemming from spatial variability of the background air mass is covered by sensitivity check 2 as detailed in Sect. 4.6. Subgrid and sub-time-step variability of the wind are not represented in the model, except for the vertical turbulence parametrisation. We state this more clearly in Sect. 4.7 of the revised manuscript as detailed above.

4/ Use of w from ECMWF and the nature of night-time lifting or subsidence (page 11): I'm not sure that large scale vertical motions need to be considered in the proposed flux approach. The effect of subsidence is to suppress the night-time boundary layer, i.e. to move the night-time inversion lower. Lifting would act to lift the inversion and entrain air from above (diluting the

NBL and therefore XCO2). Since this approach treats the NBL as a flux chamber (in effect), this motion seems not to be important and implicit (i.e. manifest) in the concentration measurements themselves. Or have I interpreted this incorrectly?

Subsidence (lifting) is intrinsically tied to horizontal divergence (convergence) of air, which does affect the NBL budgets. Imagine a case with flat terrain and no advection at the measurement location. Without subsidence or lifting, the NBL-derived flux equals the surface flux at the measurement site. If lifting takes place, surface emissions originating from the vicinity of the site 'pile up' at the measurement location and the flux estimates will be too high. Taking into account that subsidence and lifting are relatively slow processes, we do not expect strong mixing and entrainment at the border between the NBL and the residual layer above. Conversely, in case of subsidence, some fraction of the local emissions are dispersed horizontally and not included in the NBL budget, resulting in too low flux estimates.

Technical comments

Remember to add spaces between quantities and units (e.g. 100km^2 on line 11) and other instances.

We use protected thin spaces to between quantities and units, which depending on the PDF viewer and zoom setting might be occasionally overlooked. We have checked again the typesetting of quantities and units in the manuscript and made it more consistent at several locations.

References

Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S., and Grainger, C. A.: Toward Constraining Regional-Scale Fluxes of CO2 with Atmospheric Observations over a Continent: 2. Analysis of COBRA Data Using a Receptor-Oriented Framework, Journal of Geophysical Research-Atmospheres, 108, 4757, doi:10.1029/2003JD003770, wOS:000187858300011, 2003.

Karion, A., Sweeney, C., Tans, P., and Newberger, T.: AirCore: An Innovative Atmospheric Sampling System, Journal of Atmospheric and Oceanic Technology, 27, 1839–1853, doi:10.1175/2010JTECHA1448.1, 2010.

Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger, C. A.: A Near-Field Tool for Simulating the Upstream Influence of Atmospheric Observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) Model, Journal of Geophysical Research: Atmospheres, 108, 4493, doi:10.1029/2002JD003161, 2003.