Response to Reviewers

Thanks to both reviewers for their time and comments. We have made numerous changes to improve the manuscript in response to the comments. The point-by-point reviewer comments (in *blue italics*) and our corresponding responses (in black normal font) are below.

Reviewer #1

there is no mention of the instrument used and the problems documented in the literature associated with sensor wetting.

Addressed below in response to more detailed comments on this topic.

The description of how detrained and entrained masses at specific levels are calculated is not clear.

We have added a step-by-step method for how the calculation is done. See Section 2.4

1. p21787, l10: Murphy et al did not specifically mention vertical transport. Is Stainforth et al 2005 a better reference?

Agreed. We've removed the Murphy et al. reference.

2. Line 17: It is not clear what process the authors believe detrainment can potentially dominate. The reference to Wang and Geerts seems inappropriate.

The term "latter" refers to the second mode of vertical transport introduced in the previous sentence. We've edited the sentence for more clarity.

We believe that Wang and Geerts is appropriate because, as they state (p. 323 of their paper): "[vertical transport of boundary layer air] is mainly the debris of individual Cu towers rather than detrainment from long-lived convection." This is exactly the mode that we describe may "potentially dominate".

3. p21788, lines 9-10: The location of the studies, the environment, type and depth of clouds studied should be mentioned.

Good idea. Done.

4. Line 19: It is not clear what is meant by "similar results for two clouds."

Edited for clarity. The sentence now reads: "Raymond et al. (1991) combined aircraft and radar observations of summertime thunderstorm clouds over New Mexico (cloud depths ranging between 6 and 12 km) and found a similar vertical pattern of detrainment predominantly in the upper portion of clouds."

5. Lines 19-21: The statement does not summarize the main results of the study that are of most relevance to the current work.

Fair point. We have edited to sentence to read: "Barnes et al. (1996) found that the net entrainment or detrainment mass flux into or out of a layer within a cloud 0.6 to 1 km thick is typically within a factor of two of the mean vertical mass flux into that layer. They also found that detrainment varied greatly with time, with the same layer changing from net entrainment to net detrainment, or vice versa, on the order of a few minutes."

6. Lines 26-29: It should be stated that the cloud system studied was capped by an inversion.

We did not add this detail since this is generally true, and indeed is true of most of these studies, so it would seem odd to single out this particular study to make this comment.

7. p21789, lines 4-9: The authors should make their arguments clearer. Raymond et al for example mention rapidly rising tops.

Edited for clarity.

8. Lines 9-15: Since calculations were performed at each level using radar data it seems, presumably the statement here simply means on one side of the cloud.

The Carpenter et al. (1998) paper referred to here is primarily a modeling study, so this discussion is relevant to the whole cloud, not just one side.

9. p21790, line 1: The statement does not account for the presence of downdrafts.

Fair enough. We have edited the statement to add: "but not ruling out the possibility of localized entrainment that is then transported to other regions by, e.g., the descending outer shell."

10. p21791, lines 25-28: The sample rate of the chilled mirror dew point hydrometer should be given.

Sample rate is 1 Hz. This has been added, along with its accuracy of 0.2 C.

What instrument was used to measure temperature?

A Rosemount 102E4AL temperature sensor, with stated accuracy of 0.4 C was used. These details have been added to the text.

The problem of wetting should be discussed. There could be large errors in the quantities if there was a wetting problem.

In-cloud wetting of the aircraft probes does not appear to affect the thermodynamic measurements. Based on observations in stratocumulus clouds using the same instrumentation over many years, profiles of equivalent potential temperature (θ_e) can be calculated. On those days when the boundary layer appears to be well-mixed based on constant total water with height, we can check to see if the calculated in-cloud values of θ_e agree with the sub-cloud values. Data from numerous flights shows no sudden jump in calculated θ_e at cloud base, nor any vertical trend in θ_e beyond the expected constant θ_e profile. This leads us to believe that the temperature probe can yield accurate temperatures in cloudy conditions.

This discussion has been published as part of a previous study (Small, Chuang, and Feingold, GRL, 2009), so we have added a condensed description and a reference to this previous study.

Furthermore, cloud regions may not be saturated as assumed. A criterion has been described in the literature using the cloud droplet probes.

It's true that there may be holes of various sizes. We have investigated each cloud penetration to make sure there are no obvious holes at the 5 m averaging length scale, and none have been found. This does not rule out holes at smaller length scales, obviously, but given that the results are presented as averages for full cloud penetrations (minimum length 330 m, mean length of 660 m), we believe it is unlikely that deviations from this assumption will greatly affect the results.

Also the instrument errors should be given along with an estimation of the errors in the analysis.

Instrument errors have been added. Estimation of errors in the analysis are not easily propagated from the measurement uncertainties, however. Our approach here is to consider the optimization residuals as some estimate of uncertainty. The residuals are discussed in the results section.

11. line 30 - p21792, line 1: The temperature of typical cloud bases and tops should be given. How were the altitudes of cloud top and base measured?

Typical cloud base and cloud top temperatures are 8 to 15 C and 18 to 22 C, respectively. Altitudes are determined visually by the pilot, who is asked to fly through cloud base and tops for a period of time. Both points have been added to the manuscript.

12. p21792, End of 2.1: Details should be given in Section 2.1 about the calculation of the environmental, in-cloud and cloud-base values of moist static energy and the assumptions and errors.

This sentence is added in Section 2.3 (right after moist static energy is introduced): Typical uncertainties in calculated MSE are a few tenths of a percent based on instrumental uncertainties. The assumptions and errors are discussed in Sect. 2.1 when measurements of temperature and water vapor mixing ratio are described.

13. p21793, lines 1-6. I think it would be better to delete the paragraph and simply say that the clouds did not precipitate. The first part is obvious and the second is speculation that requires further analysis.

A reasonable point. We have edited down this section to read:

Precipitation could affect cloud properties, but the focus of this study is on non-precipitating clouds, so this is not an important consideration. The clouds sampled did not precipitate due to the combination of polluted aerosol conditions from the Houston region and the limited depth of the clouds which limits cloud liquid water path.

14. Lines 7-18: The paragraph should be shortened considerably.

Done. The paragraph now reads:

Net emitted radiation from a cloud causes cooling and therefore decreases MSE, while net absorption warms. During the daytime (when the research flights took place), the net radiative balance for each cloud is determined by the difference between longwave cooling and shortwave heating, which tend to be similar in magnitude. We will assume no net change due to radiation. The bias in cloud temperature, and hence MSE, caused by this assumption is likely to be very small. If we assume a 20 $\mathrm{W\,m^{-2}}$ imbalance, and a mean cloud lifetime of 30 min, the mean temperature change for a 1-km deep cloud will be a few hundredths of a Kelvin and thus unlikely to be a large source of uncertainty in this analysis.

15. p21794, lines 9-21: The text should say that MSE is only approximately conserved in a moist adiabatic process.

Done.

16. p21795, Assumptions 1 and 2: The assumptions perhaps paint a simplistic picture. The net effect of entrainment may be lateral in an Earth frame, but air is likely to have traveled vertically with respect to the ascending turret. It is well known that clouds have downdrafts at their edges.

Recent LES work (see de Rooy et al., QJRMS, DOI:10.1002/qj.1959, 2013 for a review) indicates that lateral mixing is dominant. The de Rooy et al. 2013 review states in the abstract: "A highlight of the fundamental studies resolves a long-lasting controversy by showing that lateral entrainment is the dominant mixing mechanism in comparison with the cloud-top entrainment in shallow cumulus convection." This isn't necessarily the last word on this topic (we would bet not), and is subject to the assumption that LES is accurately representing shallow cumulus, but it lends substantial credence to our use of lateral entrainment assumption.

We have substantially expanded the discussion on lateral entrainment to reflect these recent studies. In particular, we have edited the section on this page to read:

> Entrainment occurs perfectly laterally, so that all the entrained air in the cloud at the aircraft sampling altitude originates from clear air at the same altitude. A recent review paper (de Rooy et al., 2013) states that "lateral entrainment is the dominant mixing mechanism in comparison with the cloud-top entrainment in shallow cumulus convection," an idea with a long history (see references and discussion in de Rooy et al., 2013) supported by recent LES-based studies (Heus et al., 2008; Yeo and Romps, 2013).

We also edited the discussion of the sensitivity of our results to the lateral entrainment assumption (Section 3.2.2), which now reads:

> We previously made the assumption that entrainment occurs only laterally at each sampling level. Although this is an oversimplification of the entrainment process, and thus is a limitation of this model, there exists justification for this assumption. As discussed above (Section 2.4), support for lateral entrainment as the primary mechanism has gained substantial support (de Rooy et al., 2013).

We performed sensitivity tests of our model to the assumed source level of entrained air. In simulations of cumulus congestus with cloud height of 8 km, Yeo and Romp (2013) find that entrained air within the cloud at each height can be traced to air in the environment at an altitude of 1 to 2 km lower, at least during the mature and dissipating stages. If we assume self-similarity in the vertical direction, then for the clouds in this study (with depths of 1 to 2 km), the equivalent entrainment altitude is a few hundred meters below the sampling level. Thus, we test the sensitivity of our results by performing the optimization was using MSE and q_t soundings that are shifted upwards or downwards in altitude by 400 m.

17. p21795, lines 26-27: It is not clear what is meant by only applying to the cloud slice and not the entire cloud. The mass at a given level depends on entrainment and detrainment that occurred at all levels en-route to the level of interest. It is not clear how the method allows detrainment or entrainment at a particular level to be determined.

As stated in lines 19-22 on this page, and illustrated in Fig. 2 (as mentioned on line 21) we are using the in situ measurements of the cloudy air for each aircraft penetration to deduce entrainment and detrainment. Thus, the data inform only the cloud at a given level. While we thought this was stated clearly, we have edited it to hopefully improve clarity. It now reads: "Thus, the analysis results apply only to each cloud at the level of aircraft sampling, as illustrated in Figure 2, and not to the entire cloud. "

We agree that "*mass at a given level depends on entrainment and detrainment that occurred at all levels en-route to the level of interest*". That this method must assume a single level for entrainment and detrainment is clearly a limitation. We're very clear that this is a limitation of the method (e.g. lines 1 to 6 on p. 21795).

18. p21796, line 24: These points do not necessarily represent unmixed air that can be taken as the "cloud base" point. There have been several papers showing that the properties of updrafts below cloud base are different from those away from a cloud.

While that may be true, the constraint of this method is that it only permits one end member for the "adiabatic" air, and we have chosen the mean surface layer air as representative. As you point out, it would probably be more physically realistic to use air that is slightly warmer than the mean value since this would represent the positively buoyant air that will rise through the surface layer to create the cumuli. If we were to do so, the results are unlikely to change by much, as a $\sim$0.5 K warming yields a moist static energy increase of 2 kJ/kg, which is about 0.5% of the mean value. Realistic moisture biases are probably less important than temperature biases because of the magnitude of their impact on θ_v, i.e. updrafting air is more likely to be positively buoyant because of their positive temperature anomalies rather than their positive moisture anomalies.

19. p21798, Figs 3-8. Are all the figures necessary? Perhaps it would be better to show only 2. The main results are shown in Figs 10 and 12.

A good point. We've taken most of them out, leaving just two as examples.

20. p21799, line 21: The errors in the normalization should be given.

The normalization uncertainty is likely quite low, since the LCL and inversion height are well known from in situ profiles by the aircraft. We have added the following sentences to express this point: "The uncertainty in $\hat{z}$ on a day-by-day basis is likely small compared to the $\hat{z}$ bin spacing. Cloud base altitude is easily determined within $\sim$100 m from in situ measurements. Cloud top altitude is less easily determined by the pilot, but the uncertainty is likely modest compared to the total cloud layer depth as cloud top is usually constrained by a temperature and/or humidity inversion."

21. p21800, lines 14-15: Detrainment often gives rise to thin cloudy detrainment layers. Is it possible that pilots avoided these levels in order to get a better view of the clouds?

The levels were chosen to be approximately evenly spaced between cloud top and cloud bottom, so there was likely no bias against detrainment layers.

22. p21801. The discussion on this page should come after the assumptions have been tested and errors quantified. For example, it is discussed that the observed larger amount of entrainment in the upper portion of the cloud is consistent with the shedding thermal model, whereas it is assumed in the calculations that air is entrained laterally at each aircraft level.

We think the discussion of results is better here, as it more naturally follows the results. We agree that our findings from the following section, where we test assumptions, are also important for interpreting the results, but keeping the main results together with the discussion of them seems like the better of the two editorial options.

23. Lines 3-9: The authors should clarify why a moister environment leads to a larger value of mo.

To clarify this, we have added the following sentence: "In a drier environment, a large entrainment fraction would lead to the complete dissipation of the cloud. "

24. Lines 11-13: Lu et al found that shallow clouds "exhibit quasi- adiabatic regions extending from cloud base up to 0.5 - 1 x cloud depth (H)." The idea of cores existing for a few diameters should be discussed.

This particular statement in Lu et al. 2008 refers to clouds with depth 400 to 500 m, as stated later in that sentence. Lu et al. goes on to say that "for deep clouds (>1700 m thickness), the quasi-adiabatic region extends only a few hundred meters above cloud base". These "deep clouds" are more representative of those studied here, as the range of cloud depths in this study (see Table 1) is between 1 and 2.5 km. Thus, our results are consistent with those of Lu et al.

25. Lines 20-24: Is the entrained air uniformly mixed with the adiabatic mixed-layer air?

Our method does not provide us with that information, as it provides only averages at a single cloud level. We do get a clue into this question by looking at the inferred mass fractions from only the cloud edges. As discussed in Section 3.2.1, we found no substantial change in the results when we used only data from the outermost 50 m of the cloud.

26. p21802, lines 3-5: Also if the aircraft pass was across shear and perhaps only the growing part of the cloud was sampled.

We agree that there are other reasons for biases in the data set. We have edited this section to reflect other possible bias sources. The start of the paragraph now reads: "A straight-line penetration of a cloud can potentially misrepresent the area-averaged cloud properties by biasing the measurements in a number of ways. As described in section 3.2.1, one such bias is to emphasize the interior of the cloud at the expense of cloud edge."

27. p21803, lines 2-5: There are only a few points in the left-hand diagram of Fig 13 that are significantly different from zero. Is it meaningful to discuss means?

As seen in Fig. 9, this is what the results from the base case scenario (without shifted entrainment altitudes) also look like, so it's meaningful in the sense that we are processing the data in the same way and thus it is the appropriate way to compare the scenarios. That said, it's probably just as useful to compare the individual data points with those in Fig. 5. We have edited the text to add: "...(compare these results with Fig. 4)" (removal of some of the figures has renumbered the original Fig. 5 to Fig. 4) in order to emphasize this latter point.

28. p21804, lines 15-16: The statement should be qualified: a limited number of levels were sampled and the clouds were from a particular environment.

This statement was not intended to be a global statement, but rather one discussing the difference in the results from using different assumptions, which is the subject of this section. To clarify our intention, we have edited this sentence to read "The overall picture is consistent between these two analyses: detrainment is generally a weak process in these summertime shallow cumulus clouds."

29. Figs 15 and 16: It is interesting that the only negative buoyancy occurs in the upper levels.

The data points that the reviewer is interpreting as negative buoyancy in these figures lie very close (a few tenths of a Kelvin or a few tens of meters) to the environmental sounding. Thus, we would not way to over-interpret their deviation from neutral buoyancy as statistically-significant negative buoyancy.

It is also surprising that the mean positive buoyancy is so large (Fig 15, $z = 2100$ m).

Simulations from Heus et al. JGR-Atmos, 2009 show that the largest positive buoyancy perturbations within the cloud can be at or near cloud top during the growth phase of the cloud (e.g. Figs. 4, 5, 11, 12, 13 in their paper). Our physical explanation of this is that latent heat release accumulates as a parcel is lofted to higher altitudes, creating strong warm anomalies near cloud top.

It would be good to show a few time series with temperature, liquid water content and updraft speed. This is also relevant to the point raised about the wetting of the temperature probe.

The reviewer doesn't really give a clear rationale for this request. We have addressed the wetting issue separately. We don't think it's a necessary figure, so for now we have not added it.

30. p21805, lines 19-20: If air descended until it was neutrally buoyant and then detrained as suggested by Carpenter et al, why are the values of md not larger?

The short answer is we don't know, but we can speculate. It could be that these clouds, with their modest size and fairly high humidity environment, don't generate particularly strong negative buoyancy through evaporative cooling due to entrainment. It may also be that the idea of collapsing towers as per Carpenter et al. is strongest closer to the end of a cloud's lifetime, and that pilots are visually biased towards sampling clouds that are more vigorous looking and thus more in the early and middle stages of their lifetime.

Reviewer #2

The method is proposed as a novel one (see beginning of conclusion section) but it is not well described. I strongly suggest illustrating this method by one example (one single cloud penetration).

We have greatly edited and expanded the methods section to be more explicit. In particular, we have listed the step-by-step procedure that we use for generating the mass fractions.

Figures 3 to 8 could be merged in a clever way but showing six different profiles of the same parameters does not provide a conclusive picture.

As per comments from Reviewer #1, we have removed 4 of the 6 figures for specific days since the results are summarized in (old) Figures 9 and 11 (now Figures 5 and 7).

If you mention that there are "a number of approaches that could be used to address this important problem" why don't you apply these and compare with your new method?

Detrainment has been addressed using at least three main approaches: model simulations, radar observations and aircraft observations, sometimes in combination. Since this study proposes a method using only aircraft observations, some of the other approaches are obviously not relevant to this study.

Many of the aircraft observational methods involved flight patterns different from those conducted during this experiment. Barnes et al. 1996 required two aircraft to conduct multiple penetrations through a single cloud at an altitude separation of 750 m. Studies from Raymond and Wilkening (1982 and 1985) require flying closed "boxes" around a cloud to deduce the detrainment flux.

The point about the Raymond and Wilkening studies was already in the introduction ("using aircraft flying closed circuits around individual cumulus"), and we have added details about the Barnes et al. 1996 study ("using two coordinated aircraft flying at different altitudes") to alleviate any misunderstanding.

Another general suggestion is to compare your experimental results with LES. I think that there are many LES cases including fields of shallow cumuli available and you could apply your technique to these data in order to *compare your results and see how robust your results are. I know that this would need some more time but it would really strengthen your results significantly.*

We agree this would strengthen this study, but we believe the scope of the current study is sufficiently broad. It is far from trivial to deduce these types of detrainment amounts from LES results (e.g. Romps 2010). These are not simple products reported during a typical LES run like the T, RH or liquid water fields. One would have to do substantial model development in order to produce cloud detrainment fluxes. Such a large effort seems appropriate for a follow-up study.

A few figure captions/labels/legends are of poor quality; for example I am not able to read the legends of Fig 3. In general the font size of the figure labels should be increased.

We agree. We've substantially increased font sizes for number of the figures.

Specific Comments:

Abstract: page 21786, line 14: please be more specific, what means "small" cumulus clouds? Can you provide at least a mean diameter?

The manuscript text has the specific details, including Table 1. We do not think this is critical information for the abstract, however, so we are not revising it.

If you close your abstract with a statement like "findings are consistent with previous studies" one could easily argue: "so what is new? and why is your data worth to be published?"

The method is completely new, as stated in the Abstract. Because detrainment has not been extensively studied, we believe that any studies, particularly those based on observations, are useful to building a more comprehensive understanding of this important process under a variety of cloud conditions.

The part that "findings are consistent with previous studies" refers to is ONLY the earlier part of the sentence (regarding the level of neutral buoyancy), not all of the results that are described. We have edited this last sentence for increased clarity, and also to emphasize that the "previous studies" are based on models. It now reads: "Entrained and detrained air mass fractions both increased with altitude, consistent with some previous observational studies. The largest detrainment events were almost all associated with air that was at their level of neutral buoyancy, which has been hypothesized to occur based on previous modeling studies."

I suggest finishing the abstract with something more positive like "this new method allows for..."

As suggested, we have added a final sentence: "This new method could be readily used with other previous aircraft studies to expand our understanding of detrainment for a variety of cloud systems."

Introduction: page 21789, line 4:if you don't discuss the budget equation in detail most people will not understand what you mean with "accumulation term",

We believe that we do clearly explain what this means in the rest of the sentence, where we state: "i.e. the cloud is at steady state with respect to mass."

furthermore, what do you mean with "qualitative" picture?

We didn't think that point was necessary, so we removed the sentence that contained this phrase.

I think in particular the detrainment part of the introduction could be improved because - as you mention – most studies are about entrainment.

We believe that the Introduction does discuss detrainment fairly extensively – currently this discussion is about 800 words long. Obviously one could discuss the existing literature more deeply, but this is not a review paper, and we do believe we have more than adequately set up the context for the current study, which should be the main goal of the Introduction. There is also a recent review by de Rooy et al. (2013) on this topic.

On page 21790 , line 7 ff: I don't understand the sentence starting with "Because this method..." Furthermore, the last part of the last sentence in the introduction is difficult to interpret. If you want to introduce a new *method than it is not really important if previous studies are based on "bigger clouds". Same as with the abstract, I would finish this introduction with something more positive – you want to convince the reader to continue with reading your paper!*

We've deleted this sentence since you found it confusing and the point is made elsewhere as well.

I think your statement in the last sentence is not really important because you don't compare your results with previous studies.

We do think it is a relevant consideration when putting our results into context of previous studies, but we agree that this was not a very effective way to end this section. We have deleted this sentence. The same point is made early in the Introduction and again in the Conclusions.

Method section 2.1:

There are not too many details about the sensor used in your analysis, however, there are some critical concerns in particular about temperature readings in clouds and the time response of a dew-point mirror. The dew-point mirror has typically a time response of a few Kelvin per second. There is no information about the difference of the dewpoint inside cloud and the environment but the spatial resolution of the humidity observations is probably in the order of hundred meter or so. It is most critical to discuss this in terms of your analysis. The same with temperature: it is well known that typical airborne temperature measurements in clouds have serious problems with sensor wetting (see papers of Lawson/Rodi, 1992 & Lawson/Cooper, 1990 & Lenschow/Pannell, 1974). This is still an actual problem and the error in temperature readings can be as large as a few Kelvin, even with housings such as Rosemount or so. That is, total water and static energy estimates are subject to big errors, which will or might influence the results. A thorough discussion is mandatory for this paper!

A good point, and one that Reviewer #1 also brought up. The below is the same response regarding wetting as given to Reviewer #1.

In-cloud wetting of the aircraft probes does not appear to affect the thermodynamic measurements. Based on observations in stratocumulus clouds using the same instrumentation over many years, profiles of equivalent potential temperature (θ_e) can be calculated. On those days when the boundary layer appears to be well-mixed based on constant total water with height, we can check to see if the calculated in-cloud values of θ_e agree with the sub-cloud values. Data from numerous flights shows no sudden jump in calculated θ_e at cloud base, nor any vertical trend in θ_e beyond the expected constant θ_e profile. This leads us to believe that the temperature probe can yield accurate temperatures in cloudy conditions.

This discussion has been published as part of a previous study (Small et al., GRL, 2009), so we have added a condensed description and a reference to this previous study.

page 21794, line 11, please provide a value/range of used heat capacities.

We realize there was a typo previously, as we assume c_p is a function of water vapor content and not of temperature. We ignore any condensed phase as those terms are quite small. We have added the requested information into the text as: "...the heat capacity of moist air $c_p = c_p(q_v) = c_{pd}(1 + 0.9q_v)$ where c_{pd} is the heat capacity of dry air (assumed to be a constant value 1005 J/kg K),..."

Same page line 19ff: please avoid repetitions: If you start a sentence with "We note again..." you can probably avoid this statement.

Fair enough. We have re-worded it to remove repetition, and synthesize the earlier discussion (in Section 2.2) with that in the current section. The new wording is: "We have argued above (Section 2.2) that processes such as precipitation and net radiation flux divergence that can cause MSE to not be conserved are likely negligible in this study. "

Page 21795, line 12: one of your assumptions is that detrainment occurs mainly for adiabatic conditions. It is well known that adiabatic conditions are only found in the cloud core region of actively growing clouds. This assumption seems to be inconsistent with the conclusions (line 25) where you state that detrainment increases *with height and in the same sentence you mention that adiabaticity decreases with height. This is at least confusing and should be better explained.*

The reviewer is mistaken regarding our assumption. We do not assume that "detrainment occurs mainly for adiabatic conditions". This sentence is describing two extreme (or end member) scenarios within a range of possibilities, of which one is detrainment of air with adiabatic properties. As we state in the following sentence, our assumed scenario is the <u>mean</u> of these two end-members.

You mention – at several places in the entire text but particular in Sec 3.2.1 – that there might be a possible bias in your observations due to the sampling strategy; mainly it is questionable that one single penetration cannot really represent a single cloud and the edge region might be oversampled. I suggest discussing this in more detail, please have a look at : F. Hoffmann, H. Siebert, J. Schumacher, T. Riechelmann, J. Katzwinkel, B. Kumar, P. Götzfried, and S. Raasch. Entrainment and mixing at the interface of shallow cumulus clouds: Results from a combination of observations ans simulations. Meteorologische Zeitschrift, 23(DOI: 10.1127/0941-2948/2014/0597):349 – 368, 2014. In this paper such sampling problems are simulated and discussed.

Thanks for pointing out this reference. We have added it into the manuscript when we first discuss the sampling bias as it is directly relevant. We believe that we dealt with this potential bias in the most direct way possible in this study, which was to perform the analysis for only the observations at cloud edge, so further discussion is not necessary in our opinion.

Page 21801 line 1 to 25: this part is highly speculative and should be carefully rewritten

We disagree that this section is "highly speculative". We seek to put our observations of vertical trends of entrainment and detrainment in the context of processes that govern such trends and past observations. We agree that we can not prove that these are the definitive explanations for the observed trends, but they are not presented as such.

Page 21802,line 14ff: A more thorough analysis of this issue would make your paper much stronger. Why not testing other methods? Why not trying a comparison with LES data? Without such a comparison the results are somewhat weak and the conclusions are not really convincing.

The issue of lateral entrainment was also brought up by Reviewer #1. The below response is copied from this discussion.

Recent LES work (see de Rooy et al., QJRMS, DOI:10.1002/qj.1959, 2013 for a review) indicates that lateral mixing is dominant. The de Rooy et al. 2013 review states in the abstract: "A highlight of the fundamental studies resolves a long-lasting controversy by showing that lateral entrainment is the dominant mixing mechanism in comparison with the cloud-top entrainment in shallow cumulus convection." This isn't necessarily the last word on this topic (we would bet not), and is subject to the assumption that LES is accurately representing shallow cumulus, but it lends substantial credence to our use of lateral entrainment assumption.

We have substantially expanded the discussion on lateral entrainment to reflect these recent studies. In particular, we have edited the section on this page to read:

> Entrainment occurs perfectly laterally, so that all the entrained air in the cloud at the aircraft sampling altitude originates from clear air at the same altitude. A recent review paper (de Rooy et al., 2013) states that "lateral entrainment is the dominant mixing mechanism in comparison with the cloud-top entrainment in shallow cumulus convection," an idea with a long history (see references and discussion in de Rooy et al., 2013) supported by recent LES-based studies (Heus et al., 2008; Yeo and Romps, 2013).

We also edited the discussion of the sensitivity of our results to the lateral entrainment assumption (Section 3.2.2), which now reads:

> We previously made the assumption that entrainment occurs only laterally at each sampling level. Although this is an oversimplification of the entrainment process, and thus is a limitation of this model, there exists justification for this assumption. As discussed above (Section 2.4), support for lateral entrainment as the primary mechanism has gained substantial support (de Rooy et al., 2013).
>
> We performed sensitivity tests of our model to the assumed source level of entrained air. In simulations of cumulus congestus with cloud height of 8 km, Yeo and Romp (2013) find that entrained air within the cloud at each height can be traced to air in the environment at an altitude of 1 to 2 km lower, at least during the mature and dissipating stages. If we assume self-similarity in the vertical direction, then for the clouds in this study (with depths of 1 to 2 km), the equivalent entrainment altitude is a few hundred meters below the sampling level. Thus, we test the sensitivity of our results by performing the optimization was using MSE and q_t soundings that are shifted upwards or downwards in altitude by 400 m.

Page 21083 line 14ff: this conclusion is not convincing. Phrases such as " easily understood reason." In line 16 should be avoided or much more specific but this threelines are really generic.

Agreed. We have changed the phrase to "in the expected manner".

We disagree that the lines are generic given the context of the previous paragraph, where we have quantified the sensitivity of our analysis to the shifted sounding. We have moved the three lines back into the previous paragraph (instead of standing as a separate paragraph as before) to emphasize their connection, and edited them to alleviate the reviewer's concern to read: "These tests suggest that our analysis is robust with respect to our assumption of lateral entrainment. Detrainment mass fractions change rather little, while entrainment mass fractions change moderately in the expected manner."

Lines 21 to 24: This statement is also not convincing. It sound like the observed effect is small so there should not be a big difference if we change the under laying model?

We agree that these lines need to be improved for the reason stated by the reviewer. We deleted the offending sentence as it was unnecessary and misleading. We let the remaining paragraph make the real point, which is that changing the underlying model does not change the qualitative conclusions.

Page 21804, line 12 ff: Why do you consider your model to be more realistic. Please give a detailed argument.

The argument is presented earlier in the paragraph (lines 24 to 28 on p. 21803 in the original manuscript). This paragraph is meant to summarize and close the discussion on this topic, and thus does not repeat the argument.

If you close this section with a statement like: The overall picture "detrainment is a weak process." the reader might ask why you think this paper contributes to an important topic. Here it's more the wording, which makes your argument weak. The conclusion of section 4 remains unclear to me. What is the physical picture behind this observation? You only mention that your qualitative picture agrees with one publication.

We are puzzled by the reviewer's comment. Just because our study finds that a process is weak does not imply the study's importance is low, or that our argument for reaching this conclusion is weak. If a study on the health benefits of a new drug finds that it is weak, does that mean we consider the study to be of little importance?

In this case, there are studies that suggest detrainment is a strong process, and, yes, there may be few that suggest it is weak, but isn't this the way science works? We study a problem, present the results as honestly as we can, and if we do so, the community will build understanding over time. That we disagree with previous studies isn't a fundamental problem, especially given the variability of cloud types and environments and the limited range of previous studies.

Perhaps one source of concern is the (unintended) general statement at the end of this section "detrainment is generally a weak process". We had intended for it to be a statement about our sensitivity tests, not an overall claim. To clarify this, we have edited the sentence to read: "detrainment is generally a weak process in these summertime shallow cumulus clouds.", i.e. to be clear that this is a statement about the clouds in this study alone, and not a more general statement.

We have also edited the text by joining the last paragraph onto the previous paragraph to make clearer that this discussion is referring to the sensitivity tests that are the subject of this section. These lines now read: "These sensitivity tests show that our results do depend on the assumed detrained air properties, mainly in the fraction of large m_d events, although we consider our base case analysis to be more realistic regarding detrainment than the model used in this sensitivity analysis. The overall picture is consistent between these two analyses: detrainment is generally a weak process in these summertime shallow cumulus clouds. "

You should also define the buoyancy by an equation, than you can avoid ambiguous explanations such as around line 18 or 11/12.

Buoyancy is a well-established concept in atmospheric science so we do not see the need for a definition. We have edited line 18 to improve clarity. It now reads: "Previous studies have suggested that detrainment is related to cloud buoyancy profiles."

For the discussion following that line (including lines 11/12 on the following page), we specify virtual potential temperature θ_v as the relevant variable, so we don't believe that there is ambiguity regarding the physical parameter in question.

Conclusion section: you spent a lot of time arguing why your results do not agree with other observations ("These low values..." Starting at line 13). Following your arguments starting in line 17 one might conclude that natural variability is dominating and the effect od detrainment is not very clear.

That is quite close the idea we are trying to convey, although we are not exactly suggesting "natural variability is dominating". Rather we believe that there may be a number of environmental factors that control the process in ways that as of yet are unclear, and thus there is no good theory for prediction of detrainment rates.

We have re-written this section to clarify our view point on this issue.

Page 21807, line 1. "...it is well-known..:" than please provide a reference

Done. We also changed "well-known" to "common".

Line 8: "...positively buoyant relative to the environment" Buoyancy is always defined as a temperature difference and the reference temperature is the temperature of the environment. See previous comment, define buoyancy and you can avoid phrase like "..relative to the environment" which are repetitive.

We agree (not surprisingly) with your statement about how buoyancy is defined. However, one can use the term "buoyant" without it specifically meaning the scientific quantity, which is how we're using it here. In that case, we don't agree that "relative to the environment" is repetitive. We did find a different occurrence where your comment applied (p. 21804, line 22 in the original submitted manuscript) and we have removed the repetition.

Line 15: if it is possible to develop more complex model... why don't you do it? It would really strengthen your conclusions, which are a little bit weak so far. The same with the last sentence of the conclusion.

A fair comment. Upon further reflection, this was a poorly thought-out statement. We're going to remove it.

Regarding the last sentence of the conclusion, what you construe as "weak", we would claim is "not overselling the results of one study." We have edited this last paragraph to make the message clearer, but it does not go so far as to make forceful claims about what this study means in the larger picture. We believe that it is one modest piece in a large puzzle that has many large blank areas, which is what we are trying to convey.

The end of this last paragraph now reads: "The dearth of previous studies of gross detrainment hampers our ability to evaluate these results within a broader context, especially when we expect detrainment to depend on cloud type and environmental conditions. Developing a deeper understanding of detrainment from clouds, and its controlling parameters, will likely require combining a variety of approaches, of which this study is one example, in a variety of settings."