Referee 1:

General comments:

1. Citation: "Remarkably little discussion on the susceptibility of the method to measurement uncertainty is included in the manuscript. No error estimates are included in the graphics (no for measurements or for the inferred values). The authors fail to address how uncertainties, for example, in cloud base estimates and in the sizing of the droplets by the cloud probe translate into uncertainty of their inferred RH_best or Na values. In general, the description of the measurements is overall inadequate."
   Response: We added a section (3.6) and a double-panel figure (6) that show and discuss the results of the sensitivity tests conducted.

2. Citation: "The authors do not show how their proposed method (which involves direct measurements of droplet concentration, N) compares to simpler estimation methods, for example, estimating Na as the average N measured near the cloud base (where entrainment effects are expected to be weak), or from relatively undiluted data points trough the cloud (i.e., selecting a subset of samples with an AF higher than a given threshold). The authors emphasize the importance of the derived Na values, but nowhere in the manuscript do they perform comparisons of this number with the actual measured N profiles trough the sampled clouds. Are simple averages of the measured droplets concentrations also strongly correlated with below cloud CCN?"
   Response: The first paragraph in Sect. 3 describes N_a as a macro-physical property of a cloud or set of clouds in a homogeneous area and addresses the relationship to N (which is extremely variable in the convective clouds). There are many ways to derive some "representative droplet concentration", all of them are expected to correlate reasonably with each other and even with the CCN concentrations below the cloud base. However, in this manuscript we describe a methodology to derive N_a that takes into account the effect of the entrainment and mixing on the droplet concentrations and sizes, which is not sensitive to the exact path flown and which should also work in case the cloud base can not be documented or determined precisely (Fig. 6b) and even when there are no "adiabatic" measurements. We do not believe that there is a much more simple and accurate way to do that. Any straight-forward attempt to calculate N_a by using AF as filter and measurements close to cloud base, would be very sensitive to variations/errors in cloud base, as the adiabatic liquid water mixing ratio increases rapidly (in a relative sense) close to cloud base.
   The comparisons with the CCN measurements given in Fig. 8 are not there for showing that N_a has the highest correlation with the CCN concentrations of all other N estimates, but rather that it correlates reasonably with the CCN concentration measured independently in each campaign separately despite suspected calibration issues that make the slopes so different.

3. Citation: "The lack of thermodynamic measurements of environmental air (relative humidity and temperatures profiles) harms the strength of the conclusions reached by

the authors in the manuscript. It would be also interesting to see a more thorough discussion of the impact that has the use of 1Hz data (corresponding roughly to 100 m) to study effects of entrainment on droplet sizes. As stated by Lehman et al. 2009, there exists a scale above which entrainment will be mostly inhomogeneous and under which it occurs homogeneously."

Response: The phrasing of the conclusions has been slightly "softened" and the discussion regarding the lack of reliable RH measurements around and close to the penetrated clouds was slightly extended. In addition reference to the mixing type at smaller scales (Lehmann et. al, 2009) and a short discussion about it have been added to Sect. 4 as well as referred to in the third conclusion.

4. Citation: "Despite the authors efforts in making the proposed algorithm clear, important details in the calculations were left out of the description, and this makes following the proposed method difficult."

   Response: See next item.

5. Citation: "To improve the clarity of the manuscript, I suggest the authors to reserve section 3 to the detailed description of the proposed algorithm (including specifics of the calculations involved), and to remove or displace some of the paragraphs that are devoted to speculations on possible uses and applications of these steps to other sections"

   Response: Section 3 has been rebuilt to improve the clarity of the manuscript and make it easier for the reader to understand the process of deriving N_a and reproduce it with his own code (see also items 1 and 3 in response to referee #2).

Specific Comments:

1- Introduction

Citation: "Line 26 - p9675: "…as the supersaturation has to be controlled and adjusted". I suggest the authors to either extend their explanation of this or remove it from the manuscript. What do the authors mean by controlling and adjusting the supersaturation?"

Response: Controlling and adjusting the super-saturation within the CCN counter allows the derivation of the CCN activity (CCN spectra). Remote sensing techniques can, of course, not control the super-saturation. We have omitted the sentence.

Citation: "Line 24 – p9676: "here we introduce a methodology for deriving Na of convective clouds, regardless of the exact knowledge of the Earth's surface radiative properties" – This phrase should be removed. The manuscript only makes use of in-situ measurements with a cloud probe, and no remote-sensing techniques are used, so the relevance of this phrase is unclear."

Response: This sentence related to previous studies that derived N_a based on remote-sensing techniques (and mentioned in the previous section), and was meant to highlight the fact that the methodology in the current study does not use remote sensing. However, it has been omitted from the manuscript.

2- Entrainment-mixing process

Citation: "Line 6 – p9677: St and Sc are not defined in the manuscript."
Response: Added definition for St and Sc

Citation: "Line 13 – p9677: Remove "until it is saturated". This is not always possible and it is not always the case.
Response: Removed "until it is saturated".

3 – Methods

Citation: "The 5-step algorithm described in this section would be much clear to the readers if the exact equations applied to the measurements were shown. The authors devote some time to rather trivial equations (1, 2, 3 and 4), but do not include the equations leading to the mixing diagrams of Figure 1 (of great importance in the proposed approach). It would be definitely helpful to include these equations and to show exactly how they were applied to the data. Before discussing the algorithm to estimate Na, it would be useful to first describe the available data and instrumentation used.
Response: The three latter "trivial" equations (with alpha and R_e) were moved to the new subsection (3.5) discussing the inter-changeability of R_e and R_v (see also item 1 in response to referee #2).
Including all equations from the script to calculate the mixing diagrams, would overload the manuscript with many more trivial equations of calculating vapor saturation mixing ratios and such. These can be found in textbooks and should be easily accessible. However, when first mentioning Fig. 1 we added a direct reference to the discussion of the differences between the mixing scenarios, so it should be clear that in homogeneous mixing the number of droplets remains constant (concentration reduction only due to dilution) but they become smaller - just enough to saturate the mixed air.
There are references to relevant papers that describe the data and instrumentation used in the examples. The main idea of this paper is to present and describe the methodology for deriving N_a rather then showing/analyzing data.

Citation: "Line 4 - p9682: "Re_a" is not defined yet. It would perhaps be useful to explicitly state that the sub-index "a" will in general denote adiabatic values of the cloud parameters."
Response: A footnote that explicitly states that a subscript "a" denotes adiabatic values for LWC, R_v and R_e has been added.

Citation: "Line 21 - p9682: The research flights and available data are not described yet."

Response: Figures 2-4 show now a single case (see item 2 in response to referee #2), for which the reference that describes the instrumentation and flight pattern has been added.

Citation: "Line 23 - p9683: "In theory […] Na is sensitive to LWC_a". This should be rephrased since it has no physical meaning. Perhaps the authors meant that the inferred value of Na depends on the (also inferred) adiabatic liquid water content (lapse rate?). In this case and some others in the manuscript the authors tend to confound physical variables with inferred quantities."

Response: Added "derived" and "inferred" to point out that these values are not strictly physically measured values here and in other relevant places in the manuscript.

Citation: "Line 25 to 30 - p9683: The authors imply here that the pressure and temperature at cloud base could be inferred from a linear fit to equation 4 by "prescribing different cloud base properties until the fit crosses the origin". However, no support for this is included in the manuscript and this assertion should be removed. Are the linear fits included in Fig. 2 one-parameter fits in which only the slope of the fitting line is allowed to change while the intercept with the axis is set to zero? In line 26, where the authors explain how this fitting is done, it is difficult to tell if they are suggesting a possible use for their method or if they used this method in their calculations."

Response: This paragraph has been rewritten. It is clearly stated (Sect. 3.1) that the best fit line has to be forced through the origin if its slope is used for calculating $N_{a\text{-}init}$.

Citation: "Line 16 – p9683: I found this part of the manuscript confusing. In line 18 the authors assert that the derived Na would be too sensitive to errors in Re and LWCa for a too high AF threshold, but immediately after this, in line21 it is mentioned that the final derivation of Na is not sensitive to the AF threshold chosen. This is a little confusing for the reader. Is the AF threshold only used in step 3.1 or are those data points with AF values lower than the threshold also dropped out of the calculations of RH_best in the following steps?"

Response: $N_{a\text{-}init}$ was by mistake replaced with N_a in line 18 and appears to be the source of this confusion. $N_{a\text{-}init}$ may be sensitive to small changes in R_v and LWC_a in case it is based on too few data points from the lower part of the cloud. The idea is to start with as "good" $N_{a\text{-}init}$ as possible to minimize the number of iterations. However, the derived N_a at the end of the process is independent of $N_{a\text{-}init}$. This typo has been corrected.

All AF values larger than 0.1 are used for RH_{best} calculations and the final N_a derivation. This is stated now in Fig. 3 (and suggested in Sect. 3.6).

Citation: "Line 21 – p9684: "…the mean RH_best is smaller than 100%". Perhaps this phrase should be more elaborated. I assume the authors restrict the fitting parameter RH_best to be at most 100%, so this is not very meaningful since the only way the average RH_best could be 100% is if every penetration gives as a result the maximum allowed value."

Response: RH_{best} should be 100% in case of extremely inhomogeneous mixing or when entrained air is saturated. This is not the case because the mean RH_{best} is smaller than 100% and R_v shows dependence on AF. This sentence has been rephrased.

Citation: "Line 20 – p9685: It is not clear to me exactly of which best fitting curve are the MPR minimized for. Is it for the Dv-AF diagrams similar to those in Figure 3? It would be useful to state this explicitly."

Response: This sentence has been rephrased for clarification. In addition the values of MPR have been added to each panel in Fig. 3.

4 – Results and discussions

Citation: "As a general comment for this section, no discussion of how much improvement is attained with RH_best, i.e., what is the average change in the MPR. Also, the sensitivity of RH_best to sizing errors of the probe is not discussed."

Response: The absolute values of MPR or the relative change in them is not very meaningful, although it can be assessed from the provided examples (including the ones online). The MPR minimum is sought for every profile in order to derive the best N_a (and the corresponding RH_{best} values).

RH_{best} is more like a bi-product in the derivation of N_a. There is discussion about N_a sensitivity various errors in Sect. 3.6, and there are couple of sentences stating how RH_{best} is affected by errors/variations in N_a.

Citation: "Line 21 - p9687: "… therefore, it is important to correct this drift", It is unclear to me if the authors did this to their data or if they are suggesting others to follow this procedure (or both)."

Response: We routinely test our data for any drift in concentration calibration by comparing the integrated LWC to the HotWire data and applying corrections in case such drifts are detected. This can be done if the sizing calibration remains stable and correct. Because the derived N_a is affected by concentration errors we point out that a correction should be applied. But again, the methodology is the important point of the paper and we mention that unreasonably high N_a values are derived for the Amazon data (Fig. 8) because of probable under-sizing by the FSSP. Trying to correct sizing errors may be "dangerous" because the derived N_a is quite sensitive to such errors (Sect. 3.6).

Technical corrections

All technical corrections have been addressed.

Referee 2:

Main comments:

1. Citation: ”The use of both effective radius Re and volume mean radius Rv is seems to make things much more complicated than necessary. Please change to using only Rv only since that's what naturally comes out of the equations. Re is primarily used for radiation calculations. I believe a short section near the end of the paper could address the relationship between Rv and Re to connect these two concepts, but throwing α into the mix and switching from Re to Rv in different plots in the course of describing the method just adds far more confusion than needed.”
   Response: Section 3 that describes the methodology has been modified so R_v is exclusively used in the equations and figures to make things simpler. A new subsection (Sect. 3.5) as well as a Figure (5) have been added after the description of the methodology for discussing and showing the inter-changeability of R_v and R_e.

2. Citation: ”Why do the data sets used change from Fig 2 to Fig 3? The perception is that something is being hidden from the reader. To alleviate this, use the same data set for all steps so the reader can see how each step transforms an initial data set.”
   Response: Originally we wanted to show how the methodology applies for different cases and therefore showed a different example for each step. However, we accept the referee's comment and modified the manuscript that now includes a single case for allowing the reader to see how each step advances toward deriving a more realistic N_a. Figures 2-4 for three additional profiles are available as online supporting material.

3. Citation: “The authors state that it is difficult to do this, but essentially the method is an exercise in curve fitting so I think it's possible and important to do. I believe it's important because there does seem to be substantial sensitivity of the final result to the method, as even small changes in RH_best lead to large (order 10%) changes in Na. Expressing this seems important. I understand that some of the uncertainties related to the measurements themselves are hard to tackle, but the Lance et al. 2010 paper in AMTD which characterizes the CDP should be useful in this regard. At the very least, the uncertainty associated strictly with this method should be quantified, i.e. uncertainty assuming the measurements are perfect.”
   Response: A new subsection (3.6) and Figure (6) that discuss the uncertainties and sensitivity to errors have been added, as well as a reference to Lance et. al, 2010 (see more in response to referee #1, item 1).

4. Citation: “It's unclear if the authors really took the coalescence problem as serious as they should. They say (p. 9691, lines 23-24) that no more than 5% of the liquid water should have converted to hydro-meteors. How do they define hydro-meteors? What is the size threshold? How was this measured (since the CDP doesn't measure drops in the larger size range)? I think a much more rigorous filtering process to make sure that these really were clouds where collision-coalescence was not active is needed.”

Response: There are two places in Sect. 3 that now state that the data with R_e>13 μm were filtered out, because this is an indication for strong coalescence and onset of the formation of precipitation-sized particles. We discuss it in more depth in a follow-up paper, whose reference is given and that has just been submitted, where a DMT CIP is used to measure the precipitation sized particles and to distinguish samples with precipitation falling from above rather than forming in place. In addition the sentence that the referee refers to has been slightly rephrased.

5. Citation: "Why do Figs 1 and 3 have adiabatic fraction AF on the x-axis, but Fig 2 has it on the y-axis? It seems like these should stay consistent. Also, I strongly believe that Figs 1 and 3 should be plotted in log-space so that the fitted curves become straight lines. Or alternatively, plot mean drop volume Rv3 (as in Fig 4) since this will also produce a straight line. It's easy to do and also the right way to present data, in my opinion. And why does Fig 4 plot Rv3 where the others plot Rv?
Response: Some of the suggested corrections have been made. However, the y-axis in Fig. 2 denotes the adiabatic liquid water mixing ratio and not AF. It refers to the vertical dimension and therefore should be in the y-axis in our opinion. In addition, switching Figs. 1 and 3 to the log space would not make the RH_{best} fit-lines straight, as the R_v to AF relationship in homogeneous mixing cannot be described by a single equation - It involves quite a few processes and depends on a few parameters. It is $R_{v\text{-}a}^3$ that changes linearly with LWC_a and not R_v^3.

6. Citation: "There's some inconsistency in the plots as to whether they are 1-sec data, or penetration averages, or daily averages. And it's not always clear which is being plotted. Figs 2, 4, 6 and 7 have no statement for what length of time each point represents, and it's hard to figure out what they really represent. Figs 4 and 5 give averages over single penetrations, which seems like a reasonable choice for a fundamental averaging period. Why do Figs 6 and 7 have much less data, then? How are these data averaged?
Response: Each data point in Figs. 6 and 7 (now 8 and 9) stands for an entire profile; this is why there are fewer of them. It is now stated clearly in each figure caption what every data point stands for 1Hz / a cloud pass / entire profile.

7. Citation: "Does Fig 3(b) imply that all 1-sec data points always fall along a single homogeneous mixing curve, i.e. if one were to plot the points on a similar plot to Burnet and Brenguier 2007, that all the points would sit along a single curve? I find this surprising since no other study finds such behaviour. Usually there is substantial scatter in 1-sec data within such a diagram, indicating that no single choice for an effective "homogeneous mixing" RH is applicable. See Burnet and Brenguier 2007 (Fig 8) and Lehmann et al. 2009 (Fig 5) for Cu examples that clearly behave very differently. What makes these data behave so much better than these previous studies?
Response: There are a few possible reasons for why our data "behave so much better": 1) the studies mentioned did not separate for different altitudes and/or cloud passes. Fig. 3b shows that each penetration has a slightly different RH_{best} (the value of RH_{best} does not mean that there was actually homogeneous mixing with ambient air at that

RH. It is actually more of an indication that the mixing tends toward the extreme inhomogeneous as RH_{best} is remarkably larger than the ambient RH). There is no single RH_{best} that all 1-Hz data from an entire profile fit to. Figure 6 suggests that RH_{best} may have some altitude dependence (and can be quite variable), therefore separating the data by altitude may help the data look better.
2) Our dataset is based on 1 Hz measurements compared to 10 Hz in Burnet and Brenguier (2007). According to Lehmann et al (2009) mixing is expected to become more homogeneous at smaller scales. This may have an effect on how the data appears on a mixing diagram.
3) The way we plot the "mixing diagrams" (Figs. 1 and 3), i.e. R_v in the y-axis and AF in the x-axis, uses the parameter space more efficiently. Since R_v is not cubed in our mixing diagrams, small changes in it do not show up as large scatter. If we throw the data from an entire flight on the mixing diagram of Burnet and Brenguier, it looks just as "bad".

8. Citation: "The short section on alpha needs to be fixed. The value 62.03 clearly has units but it's never given, which makes it all the more confusing once that same value is used for computing alpha."
   Response: Section 3 has been reorganized and sub-section 3.5 has been added to make things simpler to understand and more physical. The units of alpha are now given in sub-section 3.5

9. Citation: "Lastly, I'm not sure the data in this study are enough to fully support the last two points in the conclusions. The second sentence in point #3 "It appears like the entrained air..." is addressed but, I don't think, with enough thoroughness to warrant a concluding statement. Same for point #4 (lines 20 to 22)."
   Response: We slightly rephrased the last two conclusions to make them less deterministic by replacing "is" with "may" and adding "According to our dataset" and "The results presented here" to emphasize that those statements are not absolute truth (see also item 3 in response to referee #1).