Dear Editor,

We thank the three reviewers for taking the time to carefully read the paper and for providing us with valuable recommendations. In this reply letter, all comments of all three reviewers are summarized

The revised version of the paper (with changes in bold) is given below (after the reply letter).

Some general changes have been made, based on the reviewers'comments:

- **The title has been changed, no longer 'Role of updrafts…', … and 'layered clouds'**

- **An overview table (Table 1), with lidar-derived products (aerosol, wind, and cloud parameters) has been introduced which provides also an overview of the uncertainties in these parameters, and information on the basic lidar signal resolution and product-retrieval resolution.**

- **Figure 1 (sketch of our lidar approach to investigate aerosol-cloud interaction) has been improved, gives more details now.**

- **Discussion of the results is improved, and the discussion of the impact of retrieval uncertainties on the findings is extended and more carefully done.**

Because all three reviewers have their doubts that our findings are statistically significant we want to give a general statement in the beginning, … before we do a step-by-step answering of all the review comments.

The reviewers have the same doubts….

Reviewer #1:One of the big issues in quantifying aerosol-cloud interactions is the lack of significant number of samples. I would like the authors to do an error/uncertainty characterization of the reported results. I understand that they have done it in Schmidt et al. (2014 JGR), but the sample size here is little bigger. A simple t-test should suffice to test whether the differences in ACIN, vertical velocity, cloud drop number concentration between updrafts and downdrafts are statistically significant.

Reviewer #2: Has data analysis been done properly and are results statistically significant for the authors. I believe if the authors perform proper statistical tests, all (e.g., Figure 6) would fail to pass the 95% confidence level, except the red bar at 30–70 m. In summary, I am afraid that I fail to see how such a small dataset and a lack of rigorous statistical analysis presented in the manuscript can be scientifically appropriate to draw meaningful conclusions.

Reveiwer #3: I am concerned about the lack of rigorous analysis of measurement errors and uncertainties, and the propagation of those uncertainties into retrieved parameters. The uncertainty in the retrieved parameters is crucial to making a convincing argument concerning the manuscript conclusions. The authors reference the Schmidt et al 2013 Applied Optics paper (and a Ph.D. dissertation that I do not have access to) regarding the measurement uncertainties and error analysis, but do not discuss those uncertainties in the context of these results. I think that by doing so, the authors' arguments will be much stronger and more convincing.

Our answers are always in bold:
First of all: We did not include any further statistical test. We do not know, why we should do that. Sure, most tests would tell us: most of the results are statistically not significant! Especially in all cases, where vertical wind information is ignored. Atmospheric variability kills the significance when playing around with just 13-29 cloud cases.

We would like to emphasis in this context: we believe 90% of all published papers regarding ACI have the same problem with strongly scattering data sets.

Nevertheless, the important message of the paper is another one: we show a statistical analysis of ACI values based on <u>well-defined liquid-water cloud layers</u> with a promising <u>innovative approach</u>!

…. and we have a <u>clear result</u>: If we include the <u>updraft information</u> (obtained in our unique approach from the Doppler lidar) in the analysis of aerosol-cloud interactions, we get <u>a strong aerosol-cloud-interaction</u> signal! We defined three cloud layers above cloud base, and in all three cloud layers, we clearly see a strong increase in ACI! That's it! That is the central message of the paper.

And if you say (as a reviewer) that is not just new …. then we would like to answer, but this has never been shown so clearly as in this paper.

And just another thing regarding the value of statistical significance: satellite retrievals may show statistically significant results, but definitely use the wrong approach (scale problem, as explained in section 4) and so the obtained results are rather rather poor.

Back to our reply (more details to all points, step by step):

The facts (what we did, what we improved):

General:

(1) We performed and presented a detailed, rigorous and conservative uncertainty analysis of the basic retrieval products in the foregoing papers (Schmidt et al., 2013, 2014). We will not repeat that here. Instead we provide a new Table 1 summarizing the typical retrieval uncertainties for all basic lidar-derived parameters in the new Table 1.

(2) The ACI values are obtained from linear regressions applied to scattered atmospheric data. Scattering is caused by retrieval uncertainties, sure! BUT to a very large part, scattering is caused by atmospheric variability. We observed young as well as aged cloud systems. Aged clouds (maybe dissolving evaporating clouds) do not show a strong aerosol effect anymore. Furthermore, as long as one ignores vertical motion in the ACI retrieval, the standard deviation of the ACI value is almost fully controlled by atmospheric variability. This is at least the conclusion from our 2008-2012 observations and the literature review. Most of the serious literature values (airborne in situ observations) show this large scatter in the data pairs (aerosol-vs-cloud property), too.

(3) We found such a clear indication that updraft occurrence has a strong impact on ACI that we never came to the idea: Is that statistically significant?

We defined three layers above cloud base (0-30m, 30-70m, 70-120m), and we found a huge increase in ACI *for all three layers*, …. when separating simply updraft from downdraft periods… and using only the updraft periods! We left out to provide a statistical analysis in the revised version.

So, in conclusion, what do we now present: ACI values, ACI standard deviations. We enlarge the discussion on the impact of retrieval uncertainties, as recommended, and we give more arguments to convince the reader that our results are trustworthy (section 3). But we did not make any attempt towards more statistical tests.

To be complete, we performed some statistical tests (t-test), as recommended by reviewer #1, but we have only 10 cloud layers where we can really apply a t-test. This data set is too small, we do not trust the result although they are pointing to the right direction. We found, e.g., no dependence between particle ext. (x) and CDNC (y) in the case when wind info is ignored, and a clear x versus y dependence with 70-95% probability for the 70-120m layer (worst case) when updraft-only data pairs are used…., as shown in Figure 6.

Details: Step-by-step …

Referee #1

………….. The conclusions drawn from the analyses seem little far-fetched. Hence, I recommend this article for publication only after they have addressed my concerns below. I am particularly concerned about (c) below.

a) The authors have not shown any statistics of cloud boundaries, phase etc. I suggest the authors to characterize the cloud base heights, cloud top heights, cloud thickness, cloud fraction etc. The aerosol cloud interactions are highly dependent on cloud characteristics, the authors mention warm clouds yet do not provide any evidence. Additionally, I suggest the authors to tabulate the mean and standard deviation of these and the aerosol properties for each case. The location of your site suggests that most of the clouds might be mixed-phase and possibly that is the reason why the Doppler lidar is able to penetrated 100 m above cloud base. This issue needs to be fully discussed.

We are sure that the remaining 29 cloud layers are purely liquid. We used polarization lidar (ice crystal detection) and Doppler radar (drizzle and ice detection) to identify pure liquid-water clouds. We state that clearly in section 2 and in the beginning of section 3, how we selected the remaining set of data. And we know cloud top temperature. So, we removed all critical (mixed-phase clouds) carefully.

We provide Tables 2 and 3 with statistics on cloud geometrical and optical properties. They were already given in the submitted version.

b) One of the big issues in quantifying aerosol-cloud interactions is the lack of significant number of samples. I would like the authors to do an error/uncertainty characterization of the reported results. I understand that they have done it in Schmidt et al. (2014 JGR), but the sample size here is little bigger. A simple t-test should suffice to test whether the differences in ACIN, vertical velocity, cloud drop number concentration between updrafts and downdrafts are statistically significant.

See our answer above!

c) The cloud dynamics needs to be characterized properly. The authors have stated that the resolution of Doppler Lidar is 70 m (Line 23, Page 31413). So then how were they able to report vertical velocities at a resolution less than that (0-30 m) in Fig 5 and 6. This makes me question the data and the data processing technique itself. Please Iist the instrumentation along with the resolution at which they operate. Also section 2 should include how you have calculated the statistics. How were the updrafts defined? By a simple sign of some threshold (0.25 m/s and -0.25 m/s) was applied.

We now state more clearly in section 2, that the Doppler lidar was used to separate regions with updraft and downdraft motions at cloud base (first height bin influenced by cloud backscatter). That's all! We did not include any threshold in our analysis, nothing, we just considered: negative or positive vertical wind at cloud base. And this Doppler lidar observation are performed with 70m resolution. They are conducted fully independently from the dual-FOV Raman lidar measurement (with 30-50m height resolution in the clouds). We provide Doppler lidar information (resolution, uncertainty in the vertical wind measurement) in the new Table 1.

…………………………………………………………………………………………………

Referee #2

1) Has data analysis been done properly and are results statistically significant for the authors to address the question shown in the title "Role of updrafts in aerosol-cloud interactions"?

Although the authors analysed 2-year long night time data, the final sample size for this manuscript is 29 cases. For some reason, only 26 cases were analysed/plotted, and the authors didn't explain why the other 3 cases were not suitable for the analysis. I was hoping to see some statements indicating that each case actually contained quite a few profiles (data points) so the data analysis here was based on a considerable number of data points. Since I don't see such statements in the manuscript, I assume that the authors use 26 data points for their analysis. When these 26 data points are further stratified by vertical velocity, as shown in the manuscript, "the role of updrafts" is then discussed based on very limited samples. I believe if the authors perform proper statistical tests, all (e.g., Figure 6) would fail to pass the 95% confidence level, except the red bar at 30–70 m. Additionally, it is unclear if seasonal variability is properly taken care of in the data analysis. In summary, I am afraid that I fail to see how such a small dataset and a lack of rigorous statistical analysis presented in the manuscript can be scientifically appropriate to draw meaningful conclusions.

We used 26-29 cloud systems for our analysis. We observed more than 200 cloud cases, but we omitted the majority because they showed the presence of ice crystals or drizzle. Now we have the problem with the statistical significance….

Such a careful selection of proper cloud layers has certainly not be done in most of the published papers on aircraft observations, because such a selection was simply not

possible. So, maybe the number of cloud samples was nicely high in these airborne studies, and the statistical significance was perfectly given, but maybe the published results are simply bad and erroneous…

We provide more information of the measured cloud cases at the beginning of section 3 (second paragraph, 29, 26, 13, 10 different subgroups of clouds).

We did not observe a seasonal dependence.

Concerning the mentioned lack of rigorous statistical analysis, see discussion above.

We changed the title, skipped 'role of updrafts…

2) Has the manuscript provided any new insights for quantifying aerosol-cloud interactions?

I am afraid that the intercomparison and discussions in Section 4 are not sufficiently critical to provide any new understanding of aerosol-cloud interactions. I also feel that some conclusions really lack supporting evidence and rigorous justification. For example, Case A is used to indicate that the low ACI_N is in the right direction and is consistent with the past estimates over the continents. What is Case A? Is Case A in Figure 5 at all? Why is it OK to assume a simplified cloud droplet number concentra- tion profile and conveniently choose an integrated number concentration, rather than certain "penetration heights" like results shown in Section 3? If Case A is consistent with the past estimates over the continents, estimates presented in Section 3 are ALL for the continents, so why are they so different from 0.1?

Section 4 provides an extended overview of the ACI related literature. Such a large, complex, and almost complete review has never been given before. We state that now very clearly. Such a compact overview as in Figure 7 has never been shown before. That's why we produced this figure. Foreced by the reviewer's comments, we re-checked the discussion keeping the comment in mind (… discussions in Section 4 are not sufficiently critical to provide any new understanding of aerosol-cloud interactions) **and omitted several statements. We improved the discussion where possible.**

But let us ask the reviewer…: Ok, if you, as an expert, have the time to check all these 100 and more papers to this ACI topic, you may then be able to conclude: nothing new in section 4! But who else has the time or is just willing to read all these 100+ papers, sometimes with very confusing and not easy-to-extract results, as you did or as we had to do? That's the main reason for this section 4. It is so difficult to get such an overview as in section 4 when checking all the available literature, we are strongly convinced that this section 4 is a very important and an extremely valuable contribution to the literature.

We removed case A, to avoid long explanations. We deleted the Case A bar in Figure7.

Specific comments:

1) Concern about statements/conclusions about mixing near cloud top: While dual (or multiple) field-of-view lidar measurements allow cloud optical depth retrievals, my understanding is that information on cloud geometric thickness will be still missing. I wonder how the authors can be so confident about the locations of cloud tops based on these lidar measurements. Is any additional instrumentation used?

We mention now, that we use cloud radar data in addition for cloud top height detection (section 2). We mention also (section 2), that there are always optically less thick cloud parcels. Usually, lidar detects both, base and top height. We do not need the top heights in our ACI investigations when focussing on particle extinction versus CDNC correlation.

2) Title: I am not sure that the title is appropriate. First, as explained, the sample is too small to conduct a meaningful analysis to investigate the role of updrafts. Second, I am not sure what "layered" really means here. Can the dual FOV Raman lidar provide information on the number of cloud layers? It looks like most of cases are probably single layered. If that is the case, "layered" in the title may confuse readers.

We changed the title accordingly, no longer 'layered clouds'.

3) Section 2: Could the authors please explain how ice clouds are excluded? Addition- ally, please provide brief information on vertical and temporal resolution here (rather than referring back to Schmidt et al. (2014a)). This information is important for readers to understand how many data points have been used for calculating error bars shown in figures.

We have polarization lidar for ice crystal detection, and also the cloud radar (which typically detects big ice crystals or drizzle drops only)!

We introduced a new Table, with all the information on lidar vertical resolution, products, and retrieval uncertainty.

4) Section 3: Figure 1 doesn't bring in any additional information that the text hasn't provided; I don't feel it serves any purpose. The text and captions are full of arbitrary thresholds – justifications of these thresholds are needed. It is important to comment on how sensitive results are to the choice of these thresholds? Also, in Section 3.2, do the authors really mean 10–90 min for signal averaging? Is it for wind retrievals only?

We improved Figure 1, provide more and detailed information, and we believe that such a sketch is needed for all the non-lidar scientists.

Concerning…. text and figure captions are full of arbitrary thresholds?... **we have no clear idea, what is meant here. Maybe the overall revision solved this problem. We don't know.**

5) Very minor – Page 31418, Line 12: I understand what the authors mean by "cloud penetration-depth effect", but clouds don't have such an effect. I would suggest writing this sentence in more precise words.

We substituted 'cloud penetration depth' by 'Height range above cloud base' in the Figures, and also in the text. We compeletly avoid 'cloud penetration depth' statements.

…………………………………………………………………………………………………..

Referee #3

……….. However, I am concerned about the lack of rigorous analysis of measurement errors and uncertainties, and the propagation of those uncertainties into retrieved parameters. The uncertainty in the retrieved parameters is crucial to making a convincing argument concerning the manuscript conclusions. The authors reference the Schmidt et al 2013 Applied Optics paper (and a Ph.D. dissertation that I do not have access to) regarding the measurement uncertainties and error analysis, but do not discuss those uncertainties in the context of these results. I think that by doing so, the authors' arguments will be much stronger and more convincing. For these reasons, I recommend that this paper be accepted only after this major issue has been addressed. I have supplied some specific comments below, which should also be ad- dressed before accepted.

See uncertainty discussion in the beginning (general statements)…

We summarize typical errors of all retrieved products in a new Table 1. The error analysis is presented in the mentioned foregoing papers. We do not repeat that here. The error overview in the paper in Table 1 must be sufficient. We discuss the uncertainties of all shown results in more detail now (section 3).

Specific Comments:

1) Introduction: I think one of the primary motivations for making long-term observations is that they provide necessary constraints for processes that are difficult to represent in models. The processes examined in this manuscript are active on the sub-grid scale relative to the GCM grid scale. A large number of observations are required to produce statistically significant constraints on sub-grid scale parameterizations, many of which are developed based on a few cases studies. This is an important motivation that should be emphasized in the introduction. While you are examining cases over
2-yr, it is only 29 cases. Is this number statistically significant? Also in the introduction, suggest also referencing ARM since it has a much longer continuous record than CLOUDNET and was established before CLOUDNET.

We mention this sub-grid scale issue now in the introduction.

We sampled more than 200 cloud layer in all these years, and then we checked all these cloud cases to come up with the final list of 29 well-defined liquid cloud layers.

We think if we have 29 well-defined pure-liquid cloud layers then we have enough data to draw conclusions. We did not make any attempt regarding…. whether these results are statistically significant or not. The atmospheric variability has a strong impact on the standard deviation of our results.

We see a clear effect (this updraft effect on ACI in all three defined cloud levels) which is in consistency with what we expect! This should be sufficient to draw solid conclusions.

Nevertheless, we discuss the large scatter in our data in more detail. We discuss the uncertainties in more detail but also state, that we see clear ACI results

We mention the ARM activities in the introduction now.

2) The cloud cases are chosen only for altocumulus clouds, which can often have ice virga falling out of the cloud. Your retrieval of LWC and effective radius relies on the assumption that the clouds are liquid. What steps are taken to ensure that ice conditions are not included in the dataset?

As mentioned above, we have polarization channels in the dual-FOV Raman lidar and can thus easily detect ice virga, and we take the cloud radar data and check them for drizzle and ice crystal presence. We state that in the manuscript (section 3.1).

3) Section 2: Please provide a short summary about the cloud properties retrieval and the uncertainty of parameters used in the study. What is the uncertainty in the ACI indices that are computed using these parameters? Are your results robust given these uncertainties? How is the number concentration (N) retrieved? I did not see this in the referenced Schmidt et al 2013 paper. What is the uncertainty on the Doppler lidar updraft velocity measurements in cloud?

We provide Tables 2 and 3 (in section 3) with aerosol and cloud properties (already in the submitted version). The retrieval of CDNC (drop number concentration) is described in Schmidt et al., JGR, 2014).

We performed an extended error analysis in the foregoing papers. The ACI values are then obtained by linear regressions fitted to the noisy data sets. We obtain in this way the ACI values together with the standard deviations (which are caused by retrieval uncertainties and atmospheric variability). This is what we can do, more is not possible.

As mentioned above, we find convincing results in agreement with our hypothesis (ACI enhancement at all three cloud levels when considering updraft periods). And this effect is close to 0.8 (which is close to the maximum of 1.0) and in good agreement with the most reasonable airborne observations. So, this is satisfactory to us.

Sure, we need more observations. We always need more measurements as we have in hand.

Doppler lidar uncertainty is about 15 cm/s, but it is , in fact, much better (about 5 cm/s) if we clearly point the Doppler lidar to the zenith what we do. This is given in the new Table 1.

4) Last sentence in Section 3.4: I don’t think that you can make any concrete con- clusions about downdrafts, turbulent mixing and entrainment processes with out using model simulations to support your conclusions.

If the ACI effect decreases with height, what else may be the reason? … if not turbulent motions, mixing, and dry air entrainment, what is then left as a reason…? Because the findings are at least consistent with the well-known effects of turbulent mixing and entrainment, we leave the statement in. It makes sense to us.

5) Figure 7 – can you annotate the figure to show which references include vertical wind in their analysis?

The few publications that considered vertical wind information are mentioned and discussed in section 4. Figure 7 shows practically only literature values without taking updraft occurrence into account.

6) Your discussion of spatial scales in Section 4 (Literature Review) is key to the significance of your findings. It would be useful to quantify the subgrid scale variability and the impact of this variability on the ACI conclusions and package it in a way that can be used to constrain model simulations and parameterizations. It really is not all that sur- prising that the influence of aerosol will be enhanced by stronger updrafts. Quantifying this phenomenon will increase the impact of your results.

We may think about this aspect in future analysis. We feel unable to do that at the moment.

7) Figures 3 and 4: the error bars are huge (orders of magnitude) and correlations and between parameters (i.e. R-squared Fig 6) are very small. In Fig. 5 the ACI index is 0.5 with +/- error bar of 0.4. This lends question to the robustness of your results/conclusions. Please provide a more thorough discussion of these error bars. It may help to compare with the uncertainties in other studies discussed in the literature review, which currently is not very quantitative in nature (in terms of uncertainty).

We provide more discussion on the uncertainties, as mentioned several times above. The scatter in the data sets (aerosol parameter versus cloud parameter) is always large and also documented in all papers we found (and used in section 4, literature overview), provided they show these scatter plots. Many publications avoid to show such scatter plots. It is not only a question of retrieval uncertainties that the variability in the findings is large!