Response to Reviewer 1

We wish to thank Reviewer 1 for his or her review of our article. The reviewer feels that some revision is required, and we shall do our best to submit a revised version in line with the reviewer's recommendations.

a) The dataset that we present is indeed limited to a single case, but these are the limitations of field work for atmospheric research. It is often not possible to replicate an experiment; in particular, atmospheric research flights are sporadic in nature and are subject to budgetary and operational limitations. We are therefore uneasy with the reviewer's suggestion to increase the number of cases of CALIPSO underpasses with the research aircraft. As far as we are aware, the FAAM aircraft only did two underpasses of CALIPSO: the one discussed in the present article and a flight over Thessaloniki in 2011 (ACEMED experiment). The 2011 flight has been presented at the 26^{th} ILRC, and is being currently thoroughly analysed with the aim of a future publication; that case is sufficiently dissimilar to this one, both in terms of scenario and methodology (the accent in that case being mainly on the in situ characterisation of the aerosols, whereas here we rely on the on-board lidar) to justify separate treatment.

b) We agree that wavelength conversion has its limitations, and we will add a caveat. However, it is the best comparison that we can have given the different wavelength of the two lidars. Wavelength conversion and caveat: see lines 109-118.

Figure 4a) We understand the point raised by the reviewer, because some profiles display very large extinction values. However, each of the profiles has a different horizontal extent and hence a different weight in the averaging. We shall add a sentence to the revised paper, explaining how the mean is calculated. Mean profile and weighing by horizontal extent: see lines 229-232.

Colloquialisms) Ok. See lines 140, 152, and 222.

Conclusions) We do not believe that one has to be ashamed of presenting a limited dataset. Satellite datasets are global in their nature, but then their verification relies on point comparisons which can be much less general. Different approaches exist: from general comparisons with a large number of ground-based observations within a network, to point verifications at one location; from multi-instrumental approaches to single instrument ones. The scene presented in this paper is quite a complex one, due to the presence of broken clouds within an aerosol layer (top of the boundary layer), and due to the oscillatory behaviour of the aerosol subtype. We have done a considerable effort in trying to disentangle the picture, and the findings are original in showing some of the strengths and weaknesses of spaceborne lidar. Moreover, such a scene is likely to be repeatable in the Amazon during the biomass burning season; and there have not been many validation campaigns in the Amazon, and in general in the Southern hemisphere. We believe that these arguments make a good case for this publication.

Response to Mark Vaughan's review

We wish to thank Mark Vaughan for his detailed review of our article. The reviewer supports our main conclusion, i.e. that this is a case where several factors contribute to make things difficult for CALIPSO: the low-level clouds increase the albedo of the scene, thus reducing the signal-to-noise ratio, and moreover their incorrect removal causes the aerosol subtype to oscillate between smoke and polluted dust, thus introducing distortion in the retrieved extinction retrievals. The reviewer, moreover, points out that validation experiments in the Southern hemisphere are rare, and thus that our underflight of the CALIPSO satellite is a good scientific opportunity.

We wish to point out that our aim is not to "criticise" the CALIPSO retrieval scheme: as a matter of fact, we are convinced that such an advanced and fully automated retrieval scheme is a great scientific achievement. Scientists working in the field of lidar know how difficult it can be to infer atmospheric properties from an elastic backscatter system when no additional information is available; more advanced systems, such as HSRL, would greatly simplify the task (this remark also applies to our aircraft lidar). It is therefore admirable that layer detection, cloud-aerosol discrimination, assignment of an aerosol subtype and lidar ratio, and profile inversion can be obtained automatically on a global scale. It is not our intent to diminish the work of the CALIPSO science team in any manner. Our only aim is to describe a critical scene that we have had the chance to observe, with the perspective that it may contribute to perfect the products in the future.

The reviewer points out that our paper is not well-structured and omits critical details. We hope that we can improve it by following all the points raised by him in the annotated manuscript.

Results are shown before methods: the reviewer is correct that we have forgotten to add a description of the airborne lidar and of the data analysis framework before the presentation of results; we shall add this information in the revised version. This has given the reviewer the impression that we assume facts for which we do not have evidence; we thank him for pointing this out so that we can reformulate the paper in a way that does not leave such a negative impression. We have added more information in section 2 (lines 89-118): this should address the problems addressed by the reviewer.

The reviewer remarks that we do not adequately describe the airborne lidar: we will do so in more detail in the new version; however this system is commercial and the amount of details that we can give is limited. Moreover, he feels that our data processing methods are not described adequately. The latter does not reflect the fact that our algorithm has been documented in a previous paper; probably we have to point this out earlier in the article, so as not to leave such an impression. We have added more information in section 2 (lines 89-118): this should address the problems addressed by the reviewer.

The reviewer then gives a detailed discussion on the question of "unstable outward retrievals" for nadir-pointing lidars. We believe that this is an interesting discussion, and we have dedicated an author comment on the ACPD forum to express our thoughts about it[1]. That author comment is to be considered as being part of this response; however the question of outward or inward retrievals is not the main point of our paper, and we are planning to reduce the emphasis on it.

Finally, the reviewer returns an annotated manuscript with several comments. These are addressed in the following pages.

1 http://editor.copernicus.org/index.php/acpd-14-C4349-2014.pdf?_mdl=msover_md&_jrl=10&_lcm=oc108lcm109w&_acm=get_comm_file&_ms=24600&c=75709&salt=89267299348901077

Page 5 of Mark Vaughan's review (comments to abstract)

Page 5

1	reviewer Sticky Note 06/05/14 17:57:55	good point. for clouds to cause the aerosol subtyping to oscillate between smoke and polluted dust, the clouds must have altered the layer integrated depolarization ratio. the most likely scenario for this to happen is if the multiple scattering from undetected water clouds (or cloud fragments) is sufficient to elevate the depolarization ratio. if this is indeed the case, the those layers where cloud detection failed are likely to be (mis)identified as polluted dust. this is something that could be definitively assessed by examining the VFM and the level 1 data for the presence of depolarizing but undetected boundary layer clouds.
2	reviewer Inserted Text 06/05/14 17:42:48	can cause
3	reviewer Highlight 15/05/14 13:38:18	
4	reviewer Sticky Note 05/05/14 16:28:31	why do these two different estimates use two very different formats?

1) We thank the reviewer for his comments which supports our main conclusion. We agree with the reviewer that multiple scattering in dense cloud can give depolarisation, and we actually see this very well in the level 1 volume depolarisation data (not displayed in the paper but available on the online quicklooks). From the level 2 data, we also see aerosol depolarisation in the smoke layer, at the same altitude where the clouds are detected (Figure 2e), and for this reason we infer that cloud depolarisation may have leaked into the aerosol product. A small fraction of undetected cloud can cause this effect. The consequence of this aerosol depolarisation is the mis-identification of the layers as polluted dust, which we observe in Figure 3e. We cannot think of a different cause for the assignment of this aerosol subtype in the CALIPSO product, and therefore we believe that this explanation is sensible. We are happy that the reviewer seems to agree to our main conclusion.
2) Ok. See line 8.
3)
4) There must have been a typo in the figures that we have given for Assesment Report 4: AR4 indicates +0.03 ± 0.12 W m^{-2} (see Chapter 2, Table 2.12); AR5 indicates -0.0 ± 0.2 W m^{-2} (see Chapter 8, Table 8.4). Note that this is the estimated value of the global mean radiative forcing, and is not representative of any particular observing scene. We shall correct the value for AR4 in the reviewed version of the manuscript. See line 18.

Page 6 of Mark Vaughan's review (comments to introduction)

Page 6

1	reviewer Sticky Note 05/05/14 16:30:14	Lopes et al., 2013 used AERONET data
2	reviewer Line 05/05/14 16:29:47	
3	reviewer Inserted Text 05/05/14 16:30:35	; Lopes et al., 2013
4	reviewer Sticky Note 09/05/14 14:51:23	Tsamalis & Chedin is a 2 paragraph EGU abstract. since peer-reviewed journal articles make for better references, perhaps the authors could cite one of these instead: Kim, M.-H., S.-W. Kim, S.-C. Yoon and A. H. Omar, 2014: "Comparison of Aerosol Optical Depth between CALIOP and MODIS-Aqua for CALIOP Aerosol Subtypes over the Ocean", *J. Geophys. Res.*, **118**, 13,241–13,252, doi:10.1002/2013JD019527. Jethva, H., O. Torres, F. Waquet, D. Chand and Y. Hu, 2014: "How do A-train Sensors Inter-Compare in the Retrieval of Above-Cloud Aerosol Optical Depth? A Case Study based Assessment", *Geophys. Res. Lett.*, **41**, 186–192, doi:10.1002/2013GL058405.
5	reviewer Highlight 06/05/14 20:29:55	
6	reviewer Sticky Note 09/05/14 14:53:28	Hunt et al., 2009 is a better reference for the CALIOP instrument description. Hunt, W. H., D. M. Winker, M. A. Vaughan, K. A. Powell, P. L. Lucker, and C. Weimer, 2009: "CALIPSO Lidar Description and Performance Assessment", *J. Atmos. Oceanic Technol.*, **26**, 1214–1228, doi:10.1175/2009JTECHA1223.1.
7	reviewer Highlight 09/05/14 14:51:53	

1) Thanks for this correction. We shall amend the paper. See lines 33-34.
2)
3)
4) Thanks. We shall add these references. See line 35.
5)
6) Thanks. We shall add this reference. See line 39.
7)

Page 7 of Mark Vaughan's review (comments to introduction and section 2)

Page 7

1	reviewer Inserted Text 06/05/14 20:42:53	identified
2	reviewer Inserted Text 09/05/14 14:57:11	and lidar ratio uncertainty that characterize the identified aerosol type,
3	reviewer Highlight 05/05/14 20:29:45	
4	reviewer Sticky Note 07/05/14 15:53:34	yes, it can. the question to be answered in a validation paper is whether or not it actually does. so presumably the analyses below will show some clear evidence of this 'mathematical instability and divergence'? if not, this sentence is a red herring and should be deleted. http://en.wikipedia.org/wiki/Red_herring
5	reviewer Highlight 09/05/14 17:56:45	
6	reviewer Sticky Note 09/05/14 17:57:14	this topic is not pursued anywhere in the text of the paper
7	reviewer Sticky Note 07/05/14 21:13:38	the authors should describe the Leosphere lidar in at least as much detail as the CALIPSO lidar; e.g., at what wavelength(s) does it operate, is it polarization sensitive, how is it calibrated and how does daylight operation affect its SNR? nominal sampling resolutions for both systems should also be presented. this information should be given before any comparisons between measurements are made.

1) Ok. See line 49.
2) Ok. See line 52.
3)
4) We have moved the discussion on the instability of the outward solution to a separate online discussion comment (see link given earlier in this response file). We shall in any case reformulate this sentence. See lines 56-63, where this question is better explained.
5)
6) Thanks for pointing this out. We shall omit the highlighted sentence. See line 68.
7) We shall add a table with the principal characteristics of the Leosphere ALS450. This lidar has been a success for the characterisation of clouds, volcanic ash, desert dust and smoke because it enjoys additional information arising from a full set of in situ and remote sensing instrumentation on the FAAM aircraft, which complement the lidar observations. However, we are not the manufacturers of this instrument and do not have access to its internals in detail to produce a thorough documentation like CALIPSO's. Unfortunately the instrument is not anymore under production, and we acknowledge that its datasheet seems to have been removed from the manufacturer's web site. For this reason, it will be useful to give more information here. The instrument is not calibrated, but calibration is not required for the data inversion method that we use (Marenco, 2013). The instrument is polarisation sensitive, but we have encountered difficulties into using the depolarisation channel in a quantitative sense; therefore we use depolarisation only qualitatively, and we are reluctant to display any plots of depolarisation since we are unable to calibrate it correctly (see Marenco et al, 2011, where this problem is already mentioned). See lines 89-103 and Table 1.

Page 8 of Mark Vaughan's review (comments to sections 2 and 3)

	Page 8	
1	reviewer Sticky Note 07/05/14 15:53:54	please provide a literature citation that describes the design and measurement capabilities of this instrument
2	reviewer Inserted Text 06/05/14 17:39:20	feature throughout the campaign
3	reviewer Cross-Out 06/05/14 17:39:28	
4	reviewer Inserted Text 06/05/14 17:39:35	5-7 km
5	reviewer Sticky Note 06/05/14 20:51:21	provide context; how long did it take CALIPSO to transit the same distance?
6	reviewer Highlight 06/05/14 20:49:43	
7	reviewer Sticky Note 06/05/14 20:52:35	again, provide context; it would be very helpful if the exact coincidence was indicated on the plot and specified in the text
8	reviewer Highlight 06/05/14 20:51:28	
9	reviewer Sticky Note 06/05/14 21:03:44	measurements are shown at 355 nm; why are retrievals shown at 532 nm? is the Leosphere a two-wavelength system? how were the Leosphere extinction coefficients derived? were any assumptions (e.g., lidar ratio) required by the extinction retrieval scheme? this information is has not been provided, yet it is critical for readers who are trying to evaluate the comparisons the authors show in Figure 2. at this point in the paper the authors still have not given the requisite details on the aircraft lidar nor have they made any mention of the technique used to derive extinction coefficients from the aircraft measurements. both of these topics should be addressed before any results are shown.
10	reviewer Highlight	

1) This instrument is documented in two JGR papers by Gerbig et al (1996 and 1999). References will be added in the revised paper. See line 77.
2) Ok. See line 80.
3)
4) Ok. See line 81.

5) The CALIPSO footprint lasted 0.5 min; in the upgraded version, this shall be indicated. See lines 86-87.
6)
7) The coincidence (i.e. when the aircraft is flying closest to CALIPSO) is at the latitude 10.36S. We shall add a vertical line showing it in the revised version of Figure 2. Thanks for the suggestion. See Figure 2.
8)
9) The Leosphere only measures at 355 nm. Aerosol extinction has been converted to 532 nm using the Angstrom formula for a better comparison with CALIPSO. The Angstrom exponent was derived from AERONET. This is explained in the text; in the revised version we will make sure that it is clearer. As already mentioned (our response to comment 7 on page 7 of Mark Vaughan's review), we shall add a paragraph on the Leosphere instrument, and will indicate the method used for the derivation of aerosol extinction and for wavelength scaling. This information was available in the paper at a later stage, but we agree with the reviewer that it should be explained before the results are discussed. See lines 104-118.
10)

Page 9 of Mark Vaughan's review (comments to section 3)

Page 9			
1	reviewer Inserted Text 06/05/14 15:15:34	horizontal	
2	reviewer Highlight 06/05/14 15:17:22		
3	reviewer Sticky Note 06/05/14 21:08:28	I note that Fig. 2c shows the log (base 10) of the attenuated backscatter coefficients. do all the white regions indicate places where the attenuated backscatter values are negative? if yes, this should be explained either in the text or in the figure caption (or, better still, in both)	
4	reviewer Sticky Note 06/05/14 21:09:22	Fig. 2D shows only extinction	
5	reviewer Cross-Out 06/05/14 21:09:05		
6	reviewer Inserted Text 06/05/14 21:09:52	designed	
7	reviewer Sticky Note 06/05/14 21:13:09	a quibble, perhaps, but I see no evidence that a layer detection algorithm has been applied to the aircraft data (in fact, application of a layer detection algorithm to the aircraft data might raise some interesting questions about the presence of boundary layer aerosols. detecting aerosols in 355 nm data can be notoriously difficult.)	
8	reviewer Inserted Text 06/05/14 15:32:15	observed	
9	reviewer Inserted Text 06/05/14 15:30:38	other	

1) Ok. See line 134.
2)
3) The reviewer is correct: the white space in the discussion paper indicates negative data. In the revised version, we shall revert to a linear scale because we believe that the logarithmic scale does not improve the way the information is presented. See Figure 2c.
4) Correct. We shall amend the text. See line 141.
5)
6) Ok. See line 142.
7) No layer detection is applied because, unlike in the CALIPSO retrieval scheme, for the aircraft lidar the whole column is processed at once. As the scene shows broken clouds, we simply omit whole vertical profiles featuring a cloud (this is a solution ad hoc for this scene) but we do not require a layer

detection algorithm.

8) Ok. See line 146.
9) Ok. See line 146.

10	reviewer Sticky Note 06/05/14 21:35:10	the authors do not specify exactly where these gaps are. however, in cases such as the one shown between ~10.6°S and ~9.6°S there is a clear explanation. as described in Vaughan et al., 2010, the CALIPSO version 3 layer detection scheme adds an 'aerosol base extender' algorithm. aerosol layers with bases that are originally detected within 2.5 km of the surface (as in the example shown here) are automatically extended down to the surface UNLESS the 532 nm integrated attenuated backscatter (IAB) in the 'gap' region is negative. since the base altitude here is 1.310 km AMSL, it follows that the CALIPSO IAB between the base altitude and the surface is less than zero; i.e., with respect to the CALIPSO signal, there is no discernible aerosol in this region. Vaughan et al., 2010, Strategies for Improved CALIPSO Aerosol Optical Depth Estimates, 25th International Laser Radar Conference, St. Petersburg, Russia
11	reviewer Cross-Out 06/05/14 15:38:53	
12	reviewer Inserted Text 06/05/14 21:13:48	seem
13	reviewer Inserted Text 06/05/14 15:42:58	seen

10) The gaps are shown in Figure 2d: from 11.35S to 11.1S (whole column); from 10.45S to 9.7S (surface to ~1300 m) and from 9.55S to 9.4S (whole column). We thank the reviewer for this explanation, and we shall add it to the revised paper. The description of the gaps has been added at lines 148-150; the explanation suggested by the reviewer has been added at lines 162-168.
11) Ok. See line 152.
12) Ok. See line 152.
13) Ok. See line 152.

Page 10 of Mark Vaughan's review (comments to section 3)

Page 10

1	reviewer Highlight 06/05/14 21:46:33	
2	reviewer Sticky Note 09/05/14 15:11:16	the only way to use the AVD to conclude that cloud contamination is negligible is to ignore large portions of the information conveyed in the AVD. as the authors explain below, the AVD flags also include information about the **simultaneous** presence of subgrid features (i.e., clouds detected at single shot resolution) in any layer. these are not independent pieces of information - high resolution cloud removal most definitely has a bearing on the layer type classification of the remaining layer fragments - nor should the authors attempt to use them as such.
3	reviewer Highlight 06/05/14 16:40:48	
4	reviewer Sticky Note 06/05/14 22:00:40	if this really is 'surprising', then the CALIPSO team has done a poor job of documenting its data products. however, I note this statement in Winker et al., 2009: "boundary layer clouds and the region of atmosphere beneath them are identified **and removed** at single-shot resolution, allowing the retrieval of aerosols between broken clouds when the gaps between clouds are smaller than the required averaging interval". the scene the authors are assessing is exactly the sort of scene for which the CALIPSO cloud-clearing procedure was devised.
5	reviewer Sticky Note 06/05/14 22:03:04	see my previous comment on this topic; when interpreted correctly (i.e., as a whole, not as several pieces of disjoint information) the AVD flags do indeed tell a consistent story
6	reviewer Highlight 06/05/14 22:01:18	
7	reviewer Sticky Note 06/05/14 22:05:07	all too true; if the layer detection algorithms fail to perform properly, all remaining steps in the processing chain are subject to error
8	reviewer Line 06/05/14 22:03:47	
9	reviewer Line 06/05/14 22:03:40	

1)
2) We thank the reviewer for this clarification, which we will add in the text. Added at lines 186-187.
3)
4) We thank the reviewer for this clarification, and we shall amend the text accordingly. See lines 188-191.
5) We completely agree with the reviewer's comment and this is the idea that we were trying to convey. In the literature we have often seen these products used in isolation, and we were trying to show that this cannot always be done. We will try to give a more effective message in the new version of the paper. Added at lines 186-187.

6)
7) We thank the reviewer for his comments which supports our main conclusion. See also our response to comment 1 on page 5 of Mark Vaughan's review.
8)
9)

10

reviewer
Sticky Note
09/05/14 18:23:14

perhaps, but not in the way the authors appear to think. see my comment in the abstract about the effects of undetected clouds on the aerosol subtyping algorithm. within the context of this scene, I suspect the aerosol subtyping is behaving exactly as it's supposed to, and the errors here are a case of garbage in, garbage out. (see Omar et al., 2009 and Liu et al., 2009; the primary inputs to both the cloud aerosol discrimination algorithm and the aerosol subtyping algorithm are layer properties calculated within the layer detection engine.)

Liu, Z. et al., 2009: "The CALIPSO Lidar Cloud and Aerosol Discrimination: Version 2 Algorithm and Initial Assessment of Performance", *J. Atmos. Oceanic Technol.*, **26**, 1198–1213, doi:10.1175/2009JTECHA1229.1.

Omar, A. et al., 2009: "The CALIPSO Automated Aerosol Classification and Lidar Ratio Selection Algorithm", *J. Atmos. Oceanic Technol.*, **26**, 1994–2014, doi:10.1175/2009JTECHA1231.1.

---

reviewer
Highlight
06/05/14 17:15:03

11

---

10) We do not think in a particular way. We are just displaying the products and trying to connect them together so as to bring out a more complete picture that would come from taking them in isolation. What we observe is a highly variable CALIPSO extinction coefficient, with larger values near the top of the boundary layer. These are the layers where high resolution clouds are apparent in the Level 1 data, and the same layers where a "polluted dust" aerosol subtype is shown. The aerosol subtyping code is doing what it is expected to, given the imperfect removal of cloud signal at a previous stage in the retrieval chain, but this has to be considered within the context of the complete CALIPSO product (and not in isolation). The oscillation of the aerosol subtype does clearly indicate that the problem lies in the incomplete removal of cloud signal, because it is the only mechanism that could have introduced a spurious depolarisation.
11)

Page 11 of Mark Vaughan's review (comments to section 3)

Page 11

1	reviewer Sticky Note 12/05/14 12:21:02	the lidar ratios used in the CALIOP retrieval are not "assumed". instead they are the result of extensive analyses (e.g., Omar et al., 2005, Omar et al., 2009) applied to a large, global data set in order to determine the lidar ratios that best characterize the different CALIOP aerosol types. Omar, A. H. et al., 2005: Development of global aerosol models using cluster analysis of Aerosol Robotic Network (AERONET) measurements, *J. Geophys. Res.*, **110**, D10S14, doi:10.1029/2004JD004874. Omar, A. et al., 2009: The CALIPSO Automated Aerosol Classification and Lidar Ratio Selection Algorithm, *J. Atmos. Oceanic Technol.*, **26**, 1994–2014, doi:10.1175/2009JTECHA1231.1.
2	reviewer Inserted Text 12/05/14 12:07:50	assigned
3	reviewer Highlight 06/05/14 16:57:06	
4	reviewer Sticky Note 06/05/14 22:18:01	I see no justification for this statement in any of the material the authors have presented thus far. the authors plainly recognize that divergence of a forward solution can be prevented by reducing the lidar ratio used. so if anything, the selection of a lidar ratio of 55 sr (i.e., polluted dust) will yield a **more stable solution** than the selection of the correct lidar ratio (70 sr for biomass burning).
5	reviewer Inserted Text 06/05/14 22:42:46	approximation used in the aerosol subtyping algorithm

1) The reviewer is correct. We shall use the more neutral word "assigned". See line 97.
2) Ok. See line 97.
3)
4) We have moved the discussion on the instability of the outward solution to a separate online discussion comment (see link given earlier in this response file). We anyhow plan to remove this statement from the revised article, and leave the question of numerical stability until the conclusions. See line 205.
5) Ok. See line 217.

Page 12 of Mark Vaughan's review (comments to section 3)

Page 12

1	reviewer Highlight 06/05/14 22:43:52	
2	reviewer Sticky Note 08/05/14 13:44:05	this sounds like exactly the right explanation to me
3	reviewer Sticky Note 07/05/14 12:33:06	once again, the authors make this assertion without providing any evidence whatsoever. once a forward solution starts to diverge, it continues to diverge toward positive infinity as with increasing layer penetration depth. now consider the data shown in figures 2d and 4a. the very large extinction values, which might suggest solution instability and divergence, are found at the top of the layers. but the retrievals below the layer are **lower** by as much as an order of magnitude. this behavior cannot happen in an unstable solution that is beginning to diverge.
4	reviewer Highlight 06/05/14 17:01:55	
5	reviewer Sticky Note 07/05/14 12:28:09	please provide more detail. for example, was this treated as a multilayer scene consisting of two or more aerosol types? (e.g., smoke lofted over boundary layer aerosol)
6	reviewer Highlight 07/05/14 12:26:28	
7	reviewer Sticky Note 09/05/14 15:19:17	while this may indeed be the best data available, its use is nevertheless problematic for several reasons... presumably 'located at ~200 km' means that the AERONET site was ~200 km from the aircraft measurement site, yes? if so, this is at the edge of the range for which reliable comparisons can be made; e.g., see this quote from Anderson et al., 2003: "Using, for example, an autocorrelation criterion of 0.8, our data suggest that coherent timescales and space scales for aerosol concentration are less than 10 h and 200 km, respectively — in some cases much less (Fig. 6 and Table 2)." what time/date was this measurement made? was the AERONET measurement time matched with the lidar measurements using the air mass matching procedure described in Lopes et al., 2013? finally, and perhaps most importantly, AERONET measures total column optical depths only. since in this case there appear to be two distinct aerosol species in the column, application of the column-mean extinction color ratio derived from AERONET measurements is likely to generate misleading results for both aerosol types. Anderson et al., 2003: Mesoscale Variations of Tropospheric Aerosols, *J. Atmos. Sci.*, **60**, 119-136, doi:10.1175/1520-0469(2003)060<0119:MVOTA>2.0.CO;2

1)
2) We thank the reviewer for his comments which supports our main conclusion. See also our response to comment 1 on page 5 of Mark Vaughan's review.
3) We have moved the discussion on the instability of the outward solution to a separate online discussion comment (see link given earlier in this response file). We anyhow plan to remove this statement from the revised article, and leave the question of numerical stability until the conclusions. See line 229.
4)
5) No. It is treated as a single layer scene, because we believe (from evidence of a month-long campaign on a multi-instrumented aircraft) that all the

aerosols (in and above the boundary layer) are smoke, although possibly with different degrees of ageing. We shall add more details on how the aircraft data have been treated to section 2, in the next version of the paper. See lines 104-108.

6)

7) The reviewer is correct: the two measurements were separated by 200 km (see map in Figure 1). We agree that it would be better to have co-located measurements (unfortunately unavailable). However, we anticipate that on flights over the Amazon, where the distance covered was more than 200 km, a very large degree of horizontal coherence of the regional haze has been found during SAMBBA, except for fresh plumes just above a fire. Here, the AERONET measurement is only used to (a) verify the consistency of the Marenco (2013) method, where however AERONET is <u>not</u> a requirement of the method; and (b) infer an Angstrom exponent for wavelength conversion. The latter being an "intensive" property of the aerosols, it can be presumed to be consistent over a larger distance, more or less like the lidar ratio is often assumed consistent over scales much larger than 200 km (see lines 114-118, where these considerations have been included in the text). Concerning the question on timing, the AERONET data have been simply interpolated to the time of the CALIPSO overpass (time now given at line 237). Concerning the remark that two different aerosols are present in the scene, we repeat (see response to comment 5 above) that we have good reasons to believe that all the observed aerosols in this scene are smoke, although possibly with different degrees of ageing (see lines 72-82).

8	reviewer Highlight 06/05/14 23:06:24	
9	reviewer Highlight 07/05/14 12:48:54	
10	reviewer Sticky Note 07/05/14 12:50:54	what does 'lowest layers' mean here? per an earlier comment, I see no evidence that the authors have applied a layer detection algorithm to the aircraft measurements. do the authors mean 'lowest range bins' instead of 'lowest layers'.
11	reviewer Sticky Note 07/05/14 12:53:33	be quantitative; was some fixed threshold of attenuated backscatter used to identify 'large peaks'? if so, what value was used?
12	reviewer Highlight 07/05/14 12:51:58	

8)

9)

10) The method is described in Marenco (2013), with its limitations and uncertainties. As discussed above (response to comment 7 of Page 9 of Mark Vaughan's review), we do not apply a layer detection scheme because we treat the whole column at once. By lower layers, we mean the height range nearer the surface (we shall specify this better in the next version; see lines 239-240): this is where integration is started, and a large uncertainty on the starting values exists; when moving inward, however, this uncertainty quickly decreases due to the mathematical stability of the solution, and becomes independent of the boundary value and little dependent on the lidar ratio (this uncertainty is clearly depicted in Fig. 4B of the discussion paper).

11) Any profile presenting at least a point within the 1500-8000 m altitude range, that has attenuated backscatter larger than 60 Mm^{-1} in both the 532 nm and

the 1064 nm channels, is entirely removed before further processing. We could give the thresholds the revised paper if the reviewer feels it is useful, but to be honest we do not believe that these are general thresholds, nor that they should be applied blindly to other scenes. This quantities have only been tested on the small scale of this experiment and not on a general basis, and this is why we are not so keen to release them.

12)

Page 13 of Mark Vaughan's review (comments to section 3 and conclusions)

Page 13		
1	reviewer Inserted Text 07/05/14 19:48:09	averaged
2	reviewer Sticky Note 16/05/14 12:41:01	since the Marenco technique is not (yet?) as well known as the Fernald and Klett methods, a bit more detail is warranted here; e.g., maybe replace the highlighted text with something like "where the far-field boundary condition has been computed by assuming a constant scattering ratio over the 500-1200 m height range"
3	reviewer Highlight 07/05/14 20:05:42	
4	reviewer Sticky Note 07/05/14 20:50:32	since the red and blue profiles were derived using the same retrieval method (albeit applied to data acquired by different instruments over different measurement intervals), the authors might comment on the differences seen between ~2.0 and ~2.8 km, where the uncertainty envelops do not overlap at all. is there some obvious explanation for the disparity there? (e.g., "twilight zone" cloud remnants that persist even after the authors manual cloud clearing?) Koren, I., L. A. Remer, Y. J. Kaufman, Y. Rudich, and J. V. Martins, 2007: On the twilight zone between clouds and aerosols, *Geophys. Res. Lett.*, **34**, L08805, doi:10.1029/2007GL029253.
5	reviewer Sticky Note 15/05/14 20:38:11	I am disappointed that the authors have chosen to terminate this section before discussing the contributions (if any) to the CALIOP and Leosphere extinction error budgets made by the calibration procedures used by two sensors. radiation noise from the South Atlantic Anomaly (SAA) causes huge difficulties in calibrating the CALIOP signals (see Powell et al., 2009), and these data were acquired well within the SAA. also, as these are daytime data, these measurements could be used to assess the effectiveness of the CALIOP night-to-day calibration transfer scheme. if a future version of this manuscript is produced, the authors should strongly consider expanding their analyses to include a section discussing calibration. as demonstrated clearly in Young et al., 2013, accurate calibration of the CALIOP signal is an essential requirement for an accurate CALIOP extinction retrieval.
6	reviewer Inserted Text 06/05/14 17:02:35	can
7	reviewer Inserted Text 07/05/14 20:54:14	amplifies
8	reviewer Inserted Text 07/05/14 20:53:58	increases

1) Ok. See line 249.
2) Ok. See lines 252-253.
3)
4) We have been puzzled by this difference as much as the reviewer, but we have not found an explanation. The difference is much larger than the uncertainty of the retrieval method, and hence is not to be considered due to the choice of the far end reference. For both datasets, cloud screening is done by completely removing whole vertical profiles affected by cloud, and therefore we do not believe that the discrepancy results from this procedure. It is quite possible that this is due to the "twilight zone" consisting of hydrated aerosols with different optical properties than the rest (lidar ratio and wavelength dependence): in that case, it is possible that retrievals at 532 nm and 355 nm yield a profile looking different. We thanks the reviewer for

pointing this out, and we shall add a comment in the revised paper. See lines 160-164.

5) We thank the reviewer for this comment; however, we have never expressed any doubt about CALIPSO's calibration procedure, which we consider very convincing. We have to point out, however, that data inversion using the Marenco (2013) method, similarly to data inversion with the traditional Fernald-Klett method, does not rely on calibrated lidar signals. If the signal is multiplied by a factor of 10, 100, or 1000, the derived extinction coefficient results identical. Therefore, of the three curves in Figure 4b, it is only the green curve that relies on calibration due to the way it is computed in HERA. Since the blue and green curves show quantitative agreement, where both are derived from the CALIOP dataset although with very different methods, and only the latter relies on calibration, we can say that the comparison tends to suggest that CALIPSO's calibration transfer is accurate. One more point has to be mentioned, however: when looking at layers near the surface, the lidar signal is affected by extinction in the overlying layers. In this scene, a layer of AOD 0.03 is missed by the CALIPSO layer classification scheme; this is equivalent to a 6% ($e^{2\tau}$) error on the instrument calibration when processing the lower layers; despite this, the green and blue curves show quantitative agreement. We do not believe that this brief comparison can reveal such small differences in the space lidar calibration, and our conclusion to the reviewer's question is that CALIPSO appears having a good calibration despite the SAA.
6) Ok. This sentence has now been removed: the same caveat is in fact already present in the introduction (lines 66-68). It is clear from the comments from reviewer 1 that this caveat leaves a wrong impression in the reader; therefore we do not think that we should over-emphasize it.
7) Ok. See line 271.
8) Ok. See line 271.

Page 14 of Mark Vaughan's review (comments to conclusions)

Page 14

1	reviewer Highlight 07/05/14 20:55:10	
2	reviewer Sticky Note 07/05/14 21:12:06	the authors have repeated this assertion ad nauseam, but have yet to provide a scintilla of evidence that this alleged 'numerical instability' is actually occurring. analytic expressions describing the expected behavior of the CALIPSO extinction retrieval under various adverse conditions (e.g., calibration error, lidar ratio error, etc.) have been derived in excruciating detail by Young et al., 2013. the authors seem to be suggesting that the behavior of the CALIPSO extinction retrievals in this study somehow exceeds the bounds that would be expected/predicted based on these analytic expressions; i.e., the retrievals have become 'unstable'. it is therefore incumbent upon them to clearly demonstrate cases of this unstable behavior.
3	reviewer Inserted Text 07/05/14 21:14:33	show
4	reviewer Sticky Note 07/05/14 21:18:42	is the aircraft lidar polarization-sensitive? if so, an image of the observed depolarization ratios along the flight track should eliminate any question about the depolarization of the aerosol layer.
5	reviewer Highlight 07/05/14 21:16:05	
6	reviewer Highlight 07/05/14 21:44:11	
7	reviewer Sticky Note 09/05/14 17:54:13	in making this remark, the authors are overlooking a critically important contribution factor. according to Vaughan et al., 2005, "a column albedo of 5% is assumed" for the calculations done to generate the figure the authors reference. the intermittent presence of bright boundary layer clouds beneath the aerosol layer very strongly suggest that the real world albedo – and hence the magnitude of the background noise in the CALIOP signal – significantly exceeds the 5% value used in the simulations. this remark also emphasizes the need (previously stated) for the authors to be specific about the temporal differences in the samples that were acquired. the aircraft data was acquired over 24 minutes, whereas the CALIPSO data was acquired over ~30 seconds. the aircraft data clearly shows that scattering intensity of the aerosol layer varies considerably along track. was CALIPSO observing the aerosol layer during the period of peak scattering intensity? (see the previous comment about adding some notation to figure 2A indicating the point of exact coincidence.)
8	reviewer Sticky Note 07/05/14 21:53:34	if there is a specific point to be made here, provide a brief synopsis of the results and/or conclusions reached by Kacenelenbogen et al.; otherwise delete the reference

1)
2) We have moved the discussion on the instability of the outward solution to a separate online discussion comment (see link given earlier in this response file). We admire the CALIPSO data analysis framework (calibration, layer selection, cloud-aerosol discrimination, aerosol subtyping and inversion) because it is very powerful and it can be applied automatically to large global datasets; this is a capability that we do not have for the aircraft lidar, where we require human intervention in the data processing and interpretation. We have never intended to stir doubt about the hard efforts from the CALIPSO science team to get the best possible product. We have not said that the behaviour of the CALIPSO retrievals exceed the bounds of the estimated uncertainty; we had just added an element of discussion mentioning one of the potential origins of the large horizontal variability that is

observed with CALIPSO (Fig. 2d) and is not picked up with the airborne lidar (Fig. 2b). It has to be taken as an element for discussion rather than a criticism of the CALIPSO data analysis framework. We have now omitted this sentence in order to reduce emphasis on this point; only one sentence on instability is kept at line

3) Ok. See line 278.
4) The instrument is polarisation sensitive, but we have encountered difficulties into using the depolarisation channel in a quantitative sense; therefore we use depolarisation only qualitatively, and we are reluctant to display any plots of depolarisation since we are unable to calibrate it (see Marenco et al, 2011, where this is already mentioned; see also our response to comment 7 on page 7 of Mark Vaughan's review). The problems with depolarisation are now mentioned at lines 93-96.
5)
6)
7) We thank the reviewer for pointing out that the specification for CALIPSO refers to an albedo of 0.05, and we shall add this information in the revised paper. See lines 195-297. We shall also ask a vertical line indicating coincidence in the figure. See Figure 2.
8) Kacenelebogen et al (2011) have found a misclassification of fine and strongly absorbing aerosol as dust and polluted dust; this is similar to our finding, although probably driven by different causes. We feel therefore that it is ok to reference this paper. We shall re-phrase the sentence. Rephrased text at lines 302-308.

9 reviewer
Highlight
07/05/14 21:48:02

10 reviewer
Cross-Out
07/05/14 21:53:57

Page 15 of Mark Vaughan's review (comments to conclusions)

Page 15

1	reviewer Cross-Out 07/05/14 21:54:06	
2	reviewer Sticky Note 07/05/14 21:59:25	I can only assume that the authors are referring to the statement in Tesche et al. that says "investigations of the CALIPSO Science Team revealed that the classification of dust and polluted dust in marine boundary layers that lie beneath other cloud or aerosol layers is due to an error in the CALIPSO retrieval software that will be fixed in future data releases". this refers to an entirely different scenario than the one being discussed in this paper.
3	reviewer Inserted Text 07/05/14 21:50:23	reported
4	reviewer Highlight 09/05/14 18:24:51	
5	reviewer Sticky Note 09/05/14 18:34:43	I am appalled by this last statement. have the authors not even once considered the enormous difference in SNR evidenced by comparing figures 2A and 2C? have they arrived at this conclusion by conducting an error analysis that attempted to apportion the errors and uncertainties according to source? if so, where are the results of this analysis reported? furthermore, it is quite clear that the root cause of the varying aerosol lidar ratios is not the subtyping routine, but instead due to layer detection and cloud clearing errors. why is this not mentioned?
6	reviewer Cross-Out 07/05/14 22:02:13	

1)
2) Tesche et al have found cases of misclassification of marine aerosols as dust or polluted dust, in the presence of clouds. A software bug has been found responsible for this misclassification, and therefore that case, although apparently similar, is different than ours. This what we are saying, and we do not understand what is wrong with our statement. The discussion tries to put our findings into a broader context, and it is appropriate to quote cases that are related or similar, but not identical. We shall re-phrase the sentence. Rephrased text at lines 308-310.
3) Ok. See line 313.
4)
5) We do not understand the reviewer's comment. Please read again our "second remark" (the previous paragraph) where the causes of varying subtype are clearly analysed in the same direction that the reviewer indicates. The present paragraph is to be read together with the previous one. We agree that the subtyping routine is being mislead by a previous stage of the data analysis (the layer detection routine), and this is exactly the message that we are trying to give. We will also mention shot noise, as suggested by the reviewer. See lines 321-323.
6) We have moved the discussion on the instability of the outward solution to a separate online discussion comment (see link given earlier in this response file). We have kept the sentence on instability because we can not rule out that it plays a role in this specific case, and the theoretical framework suggests that it can. In the separate online comment, this is discussed in detail, and all the vertical profiles are displayed. We would be grateful if the reviewer could read that online comment and provide us with his opinion on the matter.

Pages 23, 24 and 25 of Mark Vaughan's review (comments to figures)

Page 23	**Figures**	
2.1	reviewer Line 06/05/14 15:28:21	
2.2	reviewer Line 06/05/14 15:25:12	
2.3	reviewer Line 06/05/14 15:25:38	
Page 24		
3.1	reviewer Inserted Text 12/05/14 12:14:19	assigned
Page 25		
4.1	reviewer Sticky Note 07/05/14 20:03:19	with regard to several previous comments, I note here the agreement between the CALIPSO level 2 results and the authors' AERONET-constrained results derived using the CALIPSO level 1 data for altitudes below 2 km. how do the authors reconcile this good agreement with their numerous warnings about the instability of the CALIPSO extinction retrieval?

2.1,2.2,2.3) The reviewer has added three lines to our figure. We presume that he wants to highlight the bit of the elevated layer that CALIPSO has detected, but it is unclear to us what message he tries to convey.

3.1) Ok. See Figure 3, caption.

4.1) Please note that this figure shows a horizontally averaged profile over an area ~200 km long, whereas individual profiles show a larger variability (Fig. 4a). We have moved the discussion on the instability of the outward solution to a separate online discussion comment; in that comment we also display the profiles in Fig. 4a in an expanded graphical form, so as to make them easier to read (see link given earlier in this response file).