Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2017-362-RC1, 2017

Atmospheric Measurement Techniques Discussions

Interactive comment on "Comparison of dust layer heights from active and passive satellite sensors" *by* Arve Kylling et al.

Anonymous Referee #1

Received and published: 2 November 2017

This paper compares dust aerosol layer height from passive sensors (different IASI algorithms, as well as GOME-2 and SCIAMACHY) against active observations from the CALIPSO lidar. This work is important and relevant to the scope of AMT, as aerosol height affects factors like radiative effects, transport, and possible contribution to surface air quality, yet aerosol height is less readily available from satellite sensors compared to quantities like AOD. Previous work mostly considered gridded data or more limited case studies, so this expands the volume and resolution of data considered. As the thermal sensor (IASI) is only sensitive to coarse aerosols, the analysis is restricted to dust layers. I like that the authors examine different reasonable definitions of 'aerosol layer height' from CALIPSO (half cumulative extinction vs. geometric mean), and that they attempt to account for movement caused by time differences between

CALIPSO (early afternoon orbit) and the other sensors (mid-morning orbits). I also like Table 1, which provides a direct comparison of various key features of the algorithm and references. This made it easy to see how the algorithms are similar and how they are different. The conclusions seem to be that the typical level of error in these heights is of order 1-2 km, but exact errors and biases depend on how the CALIPSO height is defined, underlying surface, and (for IASI) whether the scene is daytime or nighttime.

I don't have any major scientific problems with this paper (aside from more information about FLEXTRA being needed and an omission of discussion of uncertainty estimates; see later). I think it hangs together quite well, and leaves open questions for future study in the Conclusions. I recommend re-review after major revisions, in large part to improve the presentation of the paper (Figures and Tables are hard to follow and need reorganising), because I feel I need to see the redrawn Figures and Tables to make sure that my interpretation of what is shown and discussed is correct. My specific comments are below (PX LY refers to page X, line Y of the submitted manuscript):

Abstract: the quantitative discussion here only really covers the IASI data sets. A sentence or two about GOME-2 and SCIAMACHY performance could be added as well.

Table 1: I suggest changing the 'institute' column to cover both the institute and the algorithm name, as it is algorithm names (e.g. MAPIR, IMARS) which are often referred to in other publications and may be most memorable.

P5L9: the MAPIR data presented are stated to be from two different versions, v3.2 and 3.4. However, it is hard to tell how significant the differences are. From the text it seems to only affect the a priori temperatures used in the retrieval, but it is not clear if this is the only difference. This should be clarified in the text. If the difference in retrievals from the change to the prior is negligible, then this should be stated and the sentence reworded to avoid potential reader confusion/concern. My suggestion, both for simplicity and consistency (both internal and with any future available data set) would be to just reprocess all the data with the latest version of the algorithm (v3.4).

P7, general: some of this information can be removed for readability, since there is already a reference for the algorithm in table 1 and some of this information is probably not directly necessary for the interpretation of the results here. For example I don't think the section from L7-L18 about the a priori can be shortened or removed, unless the authors feel this is significantly different from what is described in the prior algorithm paper or is somehow crucial for the understanding.

Section 2.2.4: this algorithm description can also be shortened, as Table 1 provides a reference and some details for the interested reader, and much of the information provided is not necessary for the interpretation of the results. For example P10L2-7, P10L18-22.

P13L10-24: from this it seems like the movement did not make a large systematic change to the dust height (0.02 km or less). Since the standard deviation of the height change was 0.25 km, this also implies that the error which would result from not attempting to account for the time difference is small, and that the altitude of dust layers is fairly stable over the time period of (as I understand) up to 5 hours. So that in itself is an interesting result as it suggests temporal sampling errors are not a big problem for this type of comparison, at least when looking at the big picture (since 0.25 km is somewhat smaller than the other retrieval uncertainties). The other side of the coin, which I think should be discussed more, is whether FLEXTRA's assessment of the transport during this time period is accurate. I would like to see some quantification on the reliability of FLEXTRA here as I assume it depends in part on the resolution and quality of the meteorology data ingested, since this is what will determine the horizontal and vertical motion from the trajectory. So for example which ECMWF data version was used? Has FLEXTRA been quantitatively evaluated? If you use another meteorological data set or change some other input parameter is the solution stable? I know that e.g. HYSPLIT lets you run an ensemble of trajectories as one way to assess stability; I don't know if FLEXTRA does this too. The bottom line is that the comparability between

CALIPSO and other sensors rests on the use of FLEXTRA to assess and correct for the change in aerosol location between observations, and this is not covered in much detail in the present version of the manuscript, so more discussion is needed.

Figures 2-7: I understand the intention here, but find these figures hard to interpret. In the top panel the grey is hard to see unless you zoom a lot, and the blue lines are somewhat similar to tones in the colour table. For the bottom panels, the background lidar curtain is the same between the IASI figures but there are too many different coloured symbols overlaid to quickly and easily compare and get the picture of what is going on. I wonder if, for the IASI plots at least, these could be redrawn. For example, make the top panels of Figures 2-5 their own figure, since these are all the same date and same map bounds, so we can directly compare the maps of coverage and heights between the different algorithms for IASI. Then also make the bottom panels their own figure and simplify in some way for clarity. For example we probably don't need to see the lidar curtain (although this could be added as a panel by itself at the top, or as a separate figure) since what is being compared is the effective heights, and the curtain just adds 'noise' when the eye is trying to see coloured symbols. The deep pink pluses and red crosses appear similar in tone, again taking the lidar curtain off would allow one to use a more contrasting color for one such as blue or black. The points for CALIPSO AOD could also be removed since this is in black and naturally the most eye-catching, yet it is the least relevant since it is one set of points which does not represent a height. (In my mind the think you want the reader to focus on should probably be black or red since these stick out.) Alternatively the CALIPSO AOD could be just overlaid on the lidar curtain. The point here is that when looking at figures like this I want to answer two questions: (1) how similar are the spatial coverage and distribution of heights from the data sets and (2) how do they compare to CALIPSO? The present Figures do not allow me to do this effectively.

I realise that Figures 6 and 7 are a different date so can't be combined with the IASI figures for my suggested redrawing. Is there no date with all instruments providing data? If so that would be better to show. Since both have much sparser coverage than the IASI plots, though (both these Figures are essentially a lot of white space and then overlapping symbols which are hard to distinguish without zooming), unless a better common case can be found, perhaps these two examples could be removed. I understand that from parity you probably want to show one case from each algorithm, but surely there must be a more instructive case than this; there is so little data that it's hard to assess looking at it whether these data sets are reasonable or not. And the summary statistics are in Table 3, so if the data are always sparse I don't know that we need to see GOME/SCIAMACHY case studies.

Table 3: as with Table 1, I'd add algorithm name here in the column headers. I'd also add instrument name, for ease of reference. I also have some formatting questions/comments here. In general this table is not well organised because the values in the headers only seem to refer to some of the statistics (I guess the top four of the eight boxes in each subset?) and there are too many colours. I think this could be much simplified, and the clarity improved, by condensing things and labelling individual rows rather than relying on colour-coding and large column headers. If I interpret the legend right, the top left and top right numbers in each column are the mean and standard deviation of height difference. So this could be represented as one entry, e.g. 0.590 +/- 1.213 km for BIRA-IASB. And the row legend would just read "height difference, km". Then you'd have other rows for "number of points" and "inlay". Labelling rows (which would then repeat for the all points, day/night, land/ocean splits) would remove the need for the complicated colour coding, and the header would only say instrument and institute/algorithm name. The statistics for cumulative vs. geometric heights could be split as left-right subcolumns for each algorithm instead. Does this make sense? Changing to that layout would make each row/column's content unambiguous, and dispense with the need to have 10 different colours (plus white) to code the table. As it stands, again, it is very difficult to pick out the key numbers from the table.

Figure 8 (and later): if these are scatter density plots on the first two rows then shouldn't

they be shown as filled rectangles rather than clouds of points? There is also no colour scale to indicate what is shown (absolute or relative frequency, how many points are we looking at)? I suggest redrawing. For the bottom panels, it would be clearer if a vertical line for a height difference of 0 km is added. This will again aid in direct comparisons of the data sets. I would take off the Normal distribution fits; they don't add anything that we can't already see from the bins, and don't appear to be great fits in some cases anyway (the actual distributions seem to have higher kurtosis than a Normal distribution, at least for IASI).

Figures 11, 12: I don't think these add anything to the discussion, and should be deleted. This is essentially another visualisation of the data already presented in Figures 8-10 and Table 3, albeit also sliced by time. The lines are also too faint to see. The authors can just modify the discussion on P24, L12-16 to note that optical characteristics may have been different but no clear temporal variation was found when the data were examined. We don't need to see the plots.

P24L19-23: From the 'inlay' columns it seems overall that this has values between 5.3% (LISA, ocean, night, cumulative) and 68.9% (LMD, land, night, cumulative), with typical values being something of order 30%. That means that typically two thirds or so of retrieved heights are entirely outside the dust layer (i.e. the retrievals put the dust somewhere totally without dust). Is that interpretation really correct? If so, that sounds pretty bad, and seems at odds with the other statistics presented, which show a standard deviation of around 1 km or so (and to me seems like a good result). I would double-check the calculation of this "inlay" statistic or reword the text if I have misunderstood what it means. The only way I can think of to reconcile this discrepancy is if the vertical extent of the CALIPSO dust layers is typically significantly smaller than the roughly 1 km IASI retrieval error. So perhaps some statistics about the CALIPSO dust layer geometric thickness should be presented here. Either way, there appears to be some unresolved issue in this statistic or its interpretation.

P27L14-15: Table 1 indicates that several of the algorithms use the Optimal Estimation Method (OEM) or similar techniques, which should be able to provide pixel-level uncertainty estimates on retrieved aerosol height. It would be instructive to compare these estimates in a statistical sense to the level of agreement with CALIPSO, to see whether these uncertainties are reasonable. This does not appear to have been done; I suggest adding it in the revised manuscript, since this analysis provides a useful way to 'validate' the uncertainty estimates. I realise that this can't be done for all the data sets, but since this is a big advantage of OEM, it makes sense to use it! For the OEM retrievals, it should be there already. If it is not, why not? It is definitely naturally within the scope of the existing study.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2017-362, 2017.