Referee#2 comment on the manuscript
"Validation of SCIAMACHY O2 A band cloud heights
using Cloudnet radar/lidar measurements"
by P. Wang and P. Stammes

General comments :

This study intercompares the new version of two SCIAMACHY O2 A band cloud eight products with ground based measurements. There is a distinction in the comparison between single-layer and multiple-layer cases. This distinction and the choice to present results as on Figures 6, 7 and Figure 11 are interesting. This intercomparison is valuable, the scientific interest and quality behind the manuscript are not questionable.

The author draw certainly conclusions from this study. However, they leave too much to the readers the question of the physical explanations behind their results. As it is, the document resembles more like a internal report than an scientific article. Thus physical explanations should be added, as well as a deeper and finer analysis of the results.

For example, the authors write in the abstract on lines 18-20 that "the difference between … depend on the optical thickness of the clouds", but nowhere in the manuscript is it discussed and analyzed. I think it could have been done and it would be valuable information to the reader.

A synthesis is missing between the results issued from Section 4.1 ("Global intercomparison") and from Section 4.2 ("Validation").

Specific comments :

- Title : I am wondering if the title should start more with "Evaluation of " than "Validation of".

- Abstract : it has to be synthesized.

- Page4, line 10 to 12 : add, as an reference's example , Desmons et al, AMT, 6, 2221-2238, 2013. doi:10.5194/amt-6-2221-2013

- Section 2.2 : there is no discussion about the assumption about the cloud phase in the ESA algorithm. Is it accounted for in the algorithm ? It seems to me that it should, as a forward model is used. Could it be interesting to globally compare FRESCO and ESA products per cloud phase (Section 4.1) ?

- Page 11, lines 15 to 17 : the presence of drizzle and rain don't affect the cloud top estimate from the ground ?

- Page 11 line 20 : about Figure 2 : be clearer about what "occurrence" means

- Section 4.1 :
The global comparison of FRESCO and ESA L2 cloud height products could have been more detailed and exploited. For example, the authors write in the abstract on lines 18-20 that "the difference between … depend on the optical thickness of the clouds", but nowhere in the manuscript is it discussed and analyzed. I think it could have been done and it would be valuable information to the reader. There are 6.10^5 collocated points and from it, only the mean and the width of the distributions are compared. It is quite poor.

Also, I don't see any sentence in the conclusion about Section 4.1, nor any cross comment or

lessons learned from the results given in Sections 4.1 and 4.2;

Also, why 4 full days of global data are added to the data coming from one orbit per month ?

In addition, following the short comment given by L. Lelli, 13 nov 2013, the authors should clarify whether or not the quality flags coming with SACURA products, are accounted for, and the implication for it.

- Section 4.2.1 :
The readability of Figure 4 must be improved.
Page 14, line 14 : where does come from the effective cloud fraction (how is it calculated ?)

The readability of Figure 5 must be improved.

- Lack of quantification and physical explanations :
Page15, line 3 : "seems to be better" is not clear; is it better or not ? Can you quantify the distance between ESA and Cloudnet data, and do the same with FRESCO data ?

Page15, lines 6 to 8 : what is the physical explanation for this result ?

Page15, line 20 : "sometimes ..."; add " in particular when clouds are low and thin".

Figure 6 shows that ESA data are more noisy. Discuss why. A comparison in term of cloud optical thickness could be valuable.

- Figure 7 plus the values in the tables are interesting. However, there is a lack of comment of the results (page17, line9) : it would be interesting to add a discussion about the qualities and differences of the two algorithms behind the two compared products, a link with the lessons learned from the global intercomparison of these products (section 4.1).

- Page16, line 22 and Page17, line 1 and 2 : is it the reason why there is an apparent inconsistency between the bars in Figs 7a and 7b in bin ">10" (the blue bar is very small in Fig 7b and lower that the red bar). Please clarify this result.

Technical corrections :

- Page1, line 9 : "compared" instead of "validated" ?
- Page2, line 3 : "Note a good case .. " . To be rephrased.
- Page4, line 2 : redundancy
- Page5, line 17: rephrase "for one year of data"
- Page6, line 2 : rephrase "for 51 cases"
- Page9 line 16 : no space before "1]"
- Page2, line 2 : two "local " peaks
- Page12, line 18 : "not" is missing between "are" and "used".
- Page16, line 12 : typing error
- Page34 : define a correct column width