

Interactive comment on “Detailed characterisation of AVHRR global cloud detection performance of the CM SAF CLARA-A2 climate data record based on CALIPSO-CALIOP cloud information” by Karl-Göran Karlsson and Nina Håkansson

H. Deneke (Referee)

deneke@tropos.de

Received and published: 22 November 2017

The manuscript provides an in-depth investigation of the cloud detection performance of the algorithm employed in the CLARA climate data record, utilizing CALIOP lidar observation as reference. The topic of the paper is interesting, presents novel results, and the approach is scientifically sound, hence I do recommend the paper for publication in AMT.

There are however a number of general comments/concerns which I'd like to see ad-

C1

dressed/at least discussed in the manuscript before publication, which will further clarify the relevance of the results for readers. I also added a number of specific minor points/language corrections below, which is likely incomplete. I do recommend proof-reading of the manuscript by a native English speaker.

General comments:

- Title: “Detailed characterisation” => from my point of view, the term “characterisation” mainly refers to a characterisation of performance in terms of CALIOP cloud optical thickness, I'd recommend adding COT to the title (e.g. “based on CALIPSO-CALIOP cloud optical thickness”), this is more specific than “cloud information” (what other information do you use?). I would also prefer the term “sensitivity” over “Performance”, but that is definitely a matter of taste. Hence please consider modifying the title, taking these points into account.

- The authors should describe in more detail the cloud detection scheme and the changes between the CLARA-A1 and A2 data records, in particular with respect to cloud masking. The short paragraphs at the end of Section 2.1. seems somewhat too brief, considering that the aim of the paper is to characterize the performance of that scheme, and the findings might be different for other cloud screening methods. Has the cloud mask algorithm been changed/improved between the two versions of CLARA? Are changes in cloud detection performance expected, is it possible to quantify such changes using the validation approach? Do the calibration updates affect the cloud mask performance? Has the analysis of Karlsson et al., 2013, helped to improve the algorithm, i.e. have you been able to tune the algorithm based on the results of the previous validation study? Do you expect that your results are specific to this cloud masking method, or do you expect them to be linked to fundamental characteristics of the AVHRR observations you are using, so your findings would apply similarly to other AVHRR-based cloud detection algorithms? If the latter, how would this translate to other sensors as e.g. MODIS/SUOMI NPP/geostationary observations?

C2

-In general, I find the approach of looking at the COT regardless of observing conditions somewhat too simple. I expect the detection performance to be very different during daylight/nighttime conditions, and also depend on cloud type/phase (viewing angle might be another important influencing factor). Additionally, the cloud detection scheme relies on a combination of tests, which will show different sensitivities to thin/thick/low/high clouds (it might be interesting to look at the sensitivity for each individual test separately). While it is nice to quantify the geographic variation of detection performance, what are the dominating factors for those variations (I guess surface albedo, cloud type?). Here, I urge the authors to discuss their results with more focus on the underlying physical effects (suggested plot: using a global surface albedo map e.g. from MODIS, show an x-y plot of threshold COT vs. surface albedo), and at least discuss if considering day/night different cloud types separately would add new insights.

-Due to GAC sampling, the comparability of CALIOP and AVHRR observations likely suffers. Can you quantify this effect using spatially complete data, e.g. by use of MODIS data to simulate GAC sub-sampling, in particular for those regions where clouds with significant small-scale variability are expected (i.e. the sub-tropical ocean). Even an analysis on limited data might shed some more insights in the context of the rather speculative discussion on page 10 ("We believe"...).

-In the conclusions, the author's stress that long-term availability of active observations from space would be beneficial in the conclusions. While I generally support this point, due to the inherent value of active observations, I am not convinced that this indeed adds value to the aims of this paper. Do the authors expect the performance of the cloud mask to change over time? If so, what factors could change? Why is not a once-only characterization sufficient?

-Finally, I do think that the language/wording of the article can be significantly improved, both in terms of English language use and in terms of being stricter/more consistent in terminology (some examples: use of terms "parameters" vs. "scores", "performance"

C3

vs. "sensitivity", "cloud screening" vs. "cloud detection" vs. "cloud masking", using the abstract term "detection sensitivity" instead of COT). Please do revise the paper once more carefully with respect to this points.

Detailed/language comments (disclaimer: I am not a native speaker myself...):

-L10 : "including their global distribution" => "regional variation"(?) (results is unspecific, so it remains unclear what a "distribution" of results actually refers to)

-L11 "sensitivity of the results" => which results? This opens up the possibility for misunderstanding, please change "the results" to "the cloud detection performance" or name the statistical score you are referring to.

-L 11: "cloud optical thicknesses" => "thickness"

-L 21: "sensitivities . . . were larger than 0.2" => please make it clear that COT is used as measure for sensitivity, and hence 0.2 is value of COT!

-L22 "over Sahara" => "over the Sahara"

-L23-L24: "The validation method", "validation results are proposed". This is fairly unspecific. Why not mention explicitly "It is suggested to also quantify the detection performance of other CDRs in terms of a sensitivity threshold of cloud optical thickness which can be estimated using active lidar observations"

-L28: "appear increasingly important", do not use "appear", or do the author's doubt the value of their own work?

-L29: "cloud description and . . . feedback processes" => suggested re-phrasing "the parametrization of cloud processes and cloud-aerosol interactions including related climate feedbacks."

-L37: I suggest to drop the part "in combination with ...", I do think satellite observations have sufficient value even without complementary ground-based observations

C4

-L41: “the global view” => “their global coverage”

-L57: “Aqua train” => I have never heard this term, all references I can come up with translate A-Train to “Afternoon train”

-L162: “A very strict definition” => I do not think this is a definition, but a characterization (this point also applies to other similar uses later in the manuscript)

-L235: “behave in a strange way” => maybe “introduce distortions”

-L341/342: places=> regions/locations

-L442: performance parameters => be more consistent in terminology, do you mean skill scores, or the threshold in COT?

-L448: “The method . . . is not . . . valid for the CLARA-2 . . . method”: from my reading, this statement seems to invalidate the whole paper, and does not make sense. Do the authors mean: “The method of using CALIOP data as reference is applicable”

-L449-450: “Because of this...”: I do not understand the meaning of this sentence, please clarify it.

-L495: “A specific problem with the current method”: its not an inherent problem of the method, but of data availability of active observations, I would thus suggest to use a different wording.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2017-307, 2017.