

Interactive comment on “APOLLO_NG – a probabilistic interpretation of the APOLLO legacy for AVHRR heritage channels” by L. Klüser et al.

Anonymous Referee #1

Received and published: 26 May 2015

This paper outlines alterations made to the APOLLO cloud masking algorithm. These aim to generalise the algorithm to accept input from any satellite-based radiometer (that observes the AVHRR-heritage channels) and extend the yes/no flagging to a probabilistic evaluation. Such extensions allow the APOLLO results to be useful to a wider range of users as they can select an appropriate probability at which to call a pixel ‘cloudy’ and consider any instrument observing the AVHRR-heritage channels.

I cannot recommend this paper for publication in its current form. The English is very difficult to understand, with numerous very long and ambiguous sentences. The procedures described in Sections 2 and 4 do not clearly establish how the methodology

C1245

relates to previous work (i.e. what was new, what was altered, and what remained unchanged) and the explanations are basically incomprehensible without a familiarity of Kriebel et al. (2003). The greatest weakness of the paper, though, is that it provides no justification for any of the changes to the algorithm. The reasons for desiring such improvements are clear, but why these changes in particular? Some rationale needs to be provided for these choices either from theoretical concerns, a demonstration of practical utility, or (ideally) both.

I appreciate that this paper aims to outline an improvement to an existing algorithm. I don’t expect the entire methodology to be justified, but the new components (such as the probability estimation by linear equations) need a rational explanation. I also wouldn’t require a rigorous validation; that can be in a future paper. However, some demonstration of the impact of the changes to the algorithm’s output should be presented and shown to be advantageous. I encourage the authors to redraft the piece and resubmit it as the ideas deserve to be published, but not in this immature form. I outline below some additional areas of particular note they may wish to consider during that process. (4415, L15 indicates that the comment refers to text on line 15 of page 4415.)

4415, L15 Holzer-Popp et al. (2002) applied APOLLO to ATSR-2. AATSR hadn’t been launched yet.

4415, L18 I fail to see any mention of APOLLO within Schroedter-Hornscheidt et al. (2013). Regardless, please explain what ‘solar radiation issues’ means as this is overly ambiguous.

4416, L5 ‘One major goal of the next generation method is to be applicable with any satellite sensor...’ Strictly, don’t you mean that the goal is to produce consistent results across the data record and to achieve that you choose to limit yourself to the AVHRR heritage channels? It is feasible to have a method that can be applied

C1246

consistently to each satellite but, by using differing channels, achieve inconsistent results.

- 4417, L15-27**
1. Interpreting a distance as a probability is not problematic. See [dx.doi.org/10.5194/acp-13-2195-2013](https://doi.org/10.5194/acp-13-2195-2013) or [dx.doi.org/10.5194/amt-5-73-2012](https://doi.org/10.5194/amt-5-73-2012) for usage of the Mahalanobis distance.
 2. While one can interpret a binary threshold within the concept of confidence intervals as in Fig. 1, this doesn't imply that a threshold alone is analogous to a probability. A threshold answers the question 'Do I think this observation is consistent with <feature>?' and you choose how conservative to be in that judgement. It isn't a probability until you have some model of the statistical distribution of states around the selected threshold.
 3. By weighting the various tests, your original algorithm isn't selecting probabilities of 0 or 1. They are 0 and whatever the weight is (neglecting that they can then sum to greater than unity).

Fig. 2 This is not a helpful diagram. I was hoping for something like Fig. 2 in Kriebel et al. (2003) that summarises the numerous if/then/else clauses you outline in the text of Section 2. Those are particularly difficult to follow.

- 4421, L18**
1. If you said 'knowledge about the cloud' I would agree. Cloudiness is, in the manner you discuss throughout the paper, a binary value indicating the presence of cloud in an observation.
 2. The Shannon information content is informing you of how many bits of information about the cloud can be extracted. Measurements with a high information content are well suited to numerical retrieval of cloud properties. However, including a test with low probability amongst high probability tests does not convey no additional information. It simply conveys no additional information about cloud properties within the model you defined. That data

C1247

will contain information; it just won't necessarily be about cloud. For this reason, please delete the sentence on L22-24.

3. Regardless, I am highly wary of using information content analysis on these probabilities. Without a physical meaning behind them, how can the information content be determined? The discussion around Fig. 5 is reasonable enough, but I have no reason to believe the values have any objective meaning.

Sec. 2.2 Is there any physical reason to believe that the probability (that an observation is of cloud) scales linearly? It's a very straightforward technique, and I've no opposition to its use, but there needs to be some discussion of how accurately this represents actual cloud. The fact you have to ignore any $P(x|y) = 0$ implies that a linear scaling isn't actually sensible.

Sec. 2.2 Further, your assumption in Eq. (3) that x is a binary variable doesn't represent how the five tests were designed to work. Each test is sensitive to different types/features of a cloud (e.g. the T45 test is sensitive to cirrus while the IGT test is most successful with thick clouds). A simple Bayesian processing of these probabilities seems inappropriate as they aren't quantifying the probability of the same event x , but the probability for different types of cloud (broadly speaking).

Eq. (6) More explanation of this equation is necessary as it is not clear to me how it follows from Kriebel et al. (2003), especially the first 0.2 in the denominator. It is also not clear how the multiplication of the two values follows from Eq. (3).

Sec. 2.3 Considering how often you cite values from Frey et al. (2008) and Kriebel et al. (2003), it would be very useful to reproduce these in a table in this paper.

4423, L19 Why 0.95? Why not 0.99 or 0.9 or any other number?

4424, L16 As this explanation occurs at the beginning of each subsection, why not move it to the beginning of the section?

C1248

Sec. 2.3.5 Please describe this envelope of conditions (or simply state that the technique is unaltered from Kriebel et al. (2003)).

4427, L9 You forgot the underscore before the v in $\cos(\Theta_v)$.

4430, L4 Don't you mean 'large errors'? Uncertainty describes your knowledge of the error while error is the difference between your measurement and reality. If you neglect the thermal emission, you will be wrong but you won't necessarily know how wrong you are.

4432, L15 Use \exp rather than e in the equation.

4433, L13-14 The LPF inequalities do not match between these two lines.

4436, L1 Fig. 4 shows the algorithm results in a cloud mask. You do nothing to show it is valuable.

4436, L22-24 In line with a previous comment, I believe you should swap 'uncertainty' and 'error'.

4436, L28 Yes, but are they of a similar magnitude? Doubling the uncertainty is rather important.

Overall Though your vocabulary is exemplary, your grammar produces sentences where it is often very difficult to work out what is the subject for the verb. Please use fewer interrupting statements (4423, L5-8 is a particularly egregious example). As I recommend redrafting the paper, the only specific English correction I will mention is to point out that 'respective' cannot be used as a conjunction, as you frequently do. For example, your usage on 4422, L17 is correct but your usage at 4424, L10 is not.