Atmos. Meas. Tech. Discuss., 7, C1992–C1998, 2014
www.atmos-meas-tech-discuss.net/7/C1992/2014/

Atmospheric Measurement Techniques Discussions
Open Access

Interactive comment on "Cloud mask via cumulative discriminant analysis applied to satellite infrared observations: scientific basis and initial evaluation" *by* U. Amato et al.

U. Amato et al.

serio@unibas.it

Received and published: 31 July 2014

We thank the referee for the appreciation of this study and the in depth review that will improve the final version of the paper.

1 Item-by-item reply to referee comments

1. The climatic zones are defined in terms of latitudinal belts and of the canonical four seasons: Spring, Summer, Autumn, Winter. The two data sets are represen-

tative of the four seasons, although they do not cover all the days of the year. This is not particularly important to the study. For the objective of the study it is more important to have data which are representative as much as possible of all types of cloud. In this respect, our training data set could be lacking information about the wide distribution of cloud types. However, consider that the present analysis is mostly intended to present the methodology and show preliminary results. We are now developing new training and validation data sets capable of being representative of all cloud types. This discussion will be summarized in the conclusion of the final version of the paper.

2. SEVIRI data are used for validation. We use two data sets, one corresponding to 22-23 July 2007 and the second, smaller, from 25 September to 4 October 2012. Both are silver standard since they rely on the SEVIRI operational cloud mask. However, as said in the paper, the second, smaller, data set was itself validated with ground-truth observations and the final score was better than 95%. We have clarified the matter in section 2.3.

3. We do not agree with the referee. The uncertainty of CLAVR-x we deal with in the paper has been assessed in Heidinger et al 2012 independently of the CMS cloud mask. A comparison of CMS and CLAVR-x shows an agreement within 90%, hence the accuracy of CMS and CLAVR-x are comparable, which is what we say in the paper. There is no sentence where we say that CLAVR-x has been validated with CMS.

4. The droplet radius of water clouds has a distribution, and the peak of the distribution is normally in the short-wave range. The fact that clouds are brighter in the short wave than in the long wave is an experimental fact which is hard to argue with. For most of water clouds the ratio between BTs at 5 and 12 μmwill work. It is possible that it fails for some clouds. However, this is why we have a set of statistics.

AMTD
7, C1992–C1998, 2014

5. In clear sky condition is what we mean. For sure the BT at 833 cm^{-1} is not the brightest one for cirrus clouds.

6. We will add new figures to show this difference. They are also shown below as Figure 1 and Figure 2.

7. Eq. (4), as stated, is independent of the size of the data set and of each of the two classes in particular. By construction (and in the limit case of continuous density functions) it is EI=EII at the minimum of the cost function, as can be seen from Fig. 3 of the paper or Fig. 3 of the present reply. The reason is that if, for example, EI>EII for a certain threshold, then the method tends to move the threshold in order to reduce EI; in doing this, EII increases (recall that the cumulative functions are monotone). The movement of the threshold continues and stops when EII reaches EI (EI=EII), because further reduction of EI would imply an increase of EII, which would become greater than EI and therefore the methodology would decrease EII (which implies an increase of EI). As mentioned in the paper, seen from the Classical Discriminant Analysis theory, this means that we are giving equal weight to the clear and cloudy condition error; in other words we discard information on the size of the cloudy and clear data sets (in statistical terms we assume equal prior probabilities on the classes). If we would intend to take account of the numerosities of the clear and cloudy training data sets, then the corresponding prior probabilities (which are the frequencies of clear and cloudy pixels in the training data set) could be introduced in Eq. (4). In this case the minimum of the cost function would occur when the ratio EI/EII is equal to the ratio of the prior probabilities.

8. We agree with the referee, however we have checked that the normalization does not improve results with respect to the non-normalized case.

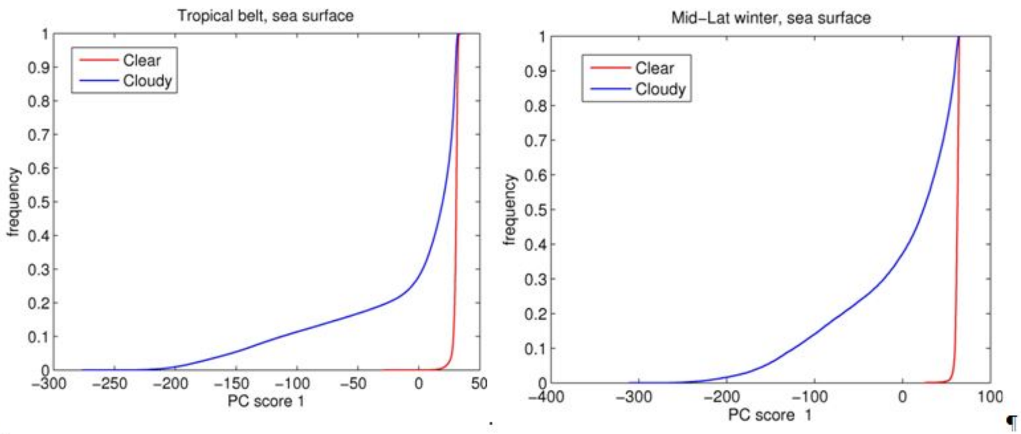
9. This was done in such a way to have a fraction of false positive below 20% or, in case also this rule was not working, to have a total score higher than 65%. These

AMTD
7, C1992–C1998, 2014

were ad hoc rules and the thresholds were obtained by trial and error. Rather than a limitation of the methods itself, the failure to find a solution according to the cost function of Eq, (4) means that clear and cloudy skies are not effectively discriminated with the class of statistics at hand. This effect has to do in part with the loss of thermal contrast between surface and clouds, but it is also an effect of the difficulty to get and effective AVHRR reference/training cloud mask for sea and land ice/snow. In the final revised manuscript the results section will be expanded to better detail and comment about findings.

Interactive comment on Atmos. Meas. Tech. Discuss., 7, 5601, 2014.

¶



¶

Figure 1: (will be shown also in the paper) Exemplifying the cumulative probability density function of the 1-st PC score of the statistic-vector **x** (Eq. 10 in the text) for **sea surface**. Panel on the left applies to the tropical belt and the distributions for clear and cloudy sky have been obtained from DIFA2 training data set; panel on the right applies to mid-latitude winter climatic zone, DIFA1 training data set.¶

¶

Fig. 1.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

AMTD
7, C1992–C1998, 2014

Interactive Comment

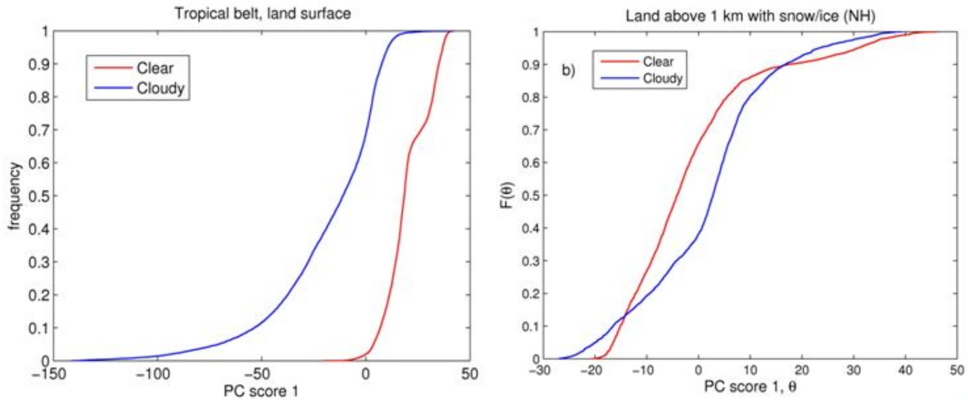
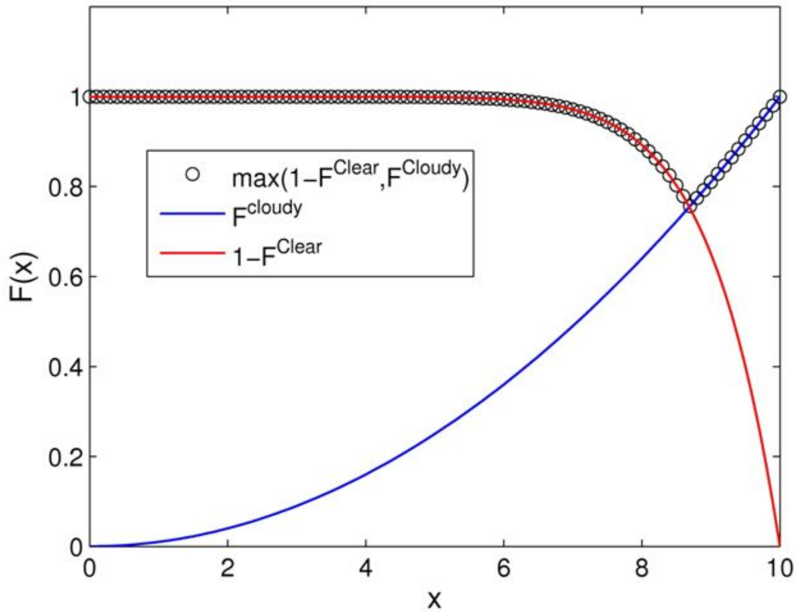


Figure 2: (will be shown also in the paper) Exemplifying the cumulative probability density function of the 1-st PC score of the statistic-vector **x** (Eq. 10 in the text) for **land surface**. Panel on the left applies to the tropical belt and the distributions for clear and cloudy sky have been obtained from DIFA2 training data set; panel on the right applies to Land above 1 km with snow/ice (NH), DIFA1 training data set, and exemplifies one case with completely overlapping cdf..

Fig. 2.

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

¶



¶

¶

Figure 3 (not shown in the paper): Illustrating the rationale behind the choice of the cost function of Eq. (4) (see text). At its minimum we have $E^{I}=E^{II}$ as it is graphically illustrated in the figure.¶

¶

¶

Fig. 3.