Review of manuscript submitted to ACP:

Title: An AeroCom/AeroSat study: Intercomparison of Satellite AOD Datasets for Aerosol Model Evaluation

Authors: Nick Schutgens et al.

General Comments

In this paper, the authors evaluate and compare fourteen satellite (seven morning and seven afternoon) products, representing 9 different retrieval algorithm families using observations from 5 different sensors on 6 different platforms. The five sensors are MODIS (Terra and Aqua), AATSR, SeaWiFS, AVHRR, and OMI. The evaluation is done for three years (2006, 2008, and 2010) by using AERONET (AErosol RObotic NETwork) and MAN (Maritime Aerosol Network) data as "ground truth". The results show that aggregated satellite AOD agrees much better than the spatial coverage (often driven by cloud masks) and that up to 50% of the difference between satellite AOD is attributed to cloud contamination.

Relative diversity, defined as the standard deviation divided by mean, is calculated at each grid-box from the 3-year average AOD values of all seven morning or afternoon products. The point is being made that this diversity may be used as an indication of AOD uncertainty.

The paper is well organized, well written, and the results are presented in a clear and concise manner. Nevertheless, the paper could be strengthened in the following ways:

1. In view of the fact that cloud masking seems to be one of the primary reasons for the discrepancies between AOD results, a discussion of how the various algorithms deal with cloud screening would be of great interest.

2. Another important reason for the reported discrepancies is attributed to the treatment of reflection by the underlying surface. This issue deserves some attention in the paper.

3. There is no discussion of the various algorithms, which is a weakness because the average reader that the authors may want to reach will not be familiar with these algorithms. Therefore, a brief description of each of the algorithms is highly recommended to be included in Section 2 of the paper. This addition would make the paper of interest to a much wider audience, and it should include a description of how each algorithm deals with cloud screening and surface reflection.

Specific Comments

- On page 4 the authors state: "We will provide evidence that cloud masking is the dominant factor....". Please be more specific about where in the paper such evidence is provided.

- On page 5 the authors state: "Most ocean boxes with observations will be in coastal regions, with some over isolated islands." Please explain the reason for this restriction.

The MAN web-site indicates that data are obtained by Mircrotops deployed on ships all over the world. Why are those data not used in this study?

- On page 6, the following sentence appears: "Bootstrapping has been shown to be reliable even for relatively small sample sizes." A reference in support of this statement seems to be required.

- on page 9 the authors remark: "It is not surprising that some products will have better cloud masking than others ...". Some explanation would be useful here.

- On page 10 the authors write: "Terra/Aqua-DT, by the way, sometimes produces negative AOD leading to e.g. very low values for averaged AOD over Australia." What is the reason for this problem?

- On bottom of page 10: "The contrasts in the differences over land and neighbouring ocean (e.g. African outflow for Terra-DT with AATSR-SU or AATSR-ADV, or Aqua-DT with OMAERUV, or AVHRR with SeaWiFS) may likewise be driven by albedo treatment." Some more discussion of this "albedo treatment" issue would be useful.

- On page 11 the authors write: "The three products based on the DeepBlue algorithm (Aqua-DB, AVHRR and SeaWiFS) suggest that already small algorithmic differences can yield significant differences." Please be more specific about what is meant by "small algorithmic differences".

- On page 12, the authors state: "Diversity is generally lowest over ocean, never reaching over 30% while over land values of 100% are possible." Please explain this result.

- On page 13, the authors state: "We see that the diversity goes down when the mean AOD increases, and goes up when the uncertainty in cloud masking increases. This is as one would expect." Please elaborate on why this result is to be expected.

Technical Corrections

In general this paper is well written, and I did not spot any typographical or grammar mistakes, except for the following:

at the top of page 3, the following text appears:

...the data usually come in different spatio-temporal grids. In addition, data has often....

The noun "data" is plural (from the singular latin word "datum"). In the first appearance "data usually come" it is considered to be plural, while in the second appearance "data has" it is considered to be singular. I would recommend that "data" be considered to be plural, and that the whole paper be searched and changed to amend this inconsistency.