Interactive comment on "Evaluation of cloud fraction and its radiative effect simulated by IPCC AR4 global models against ARM surface observations" by Y. Qian et al.

Anonymous Referee #2
Received and published: 1 July 2011

Review of paper: Evaluation of cloud fraction and its radiative effect simulated by IPCC AR4 global models against ARM surface observations.

This study evaluates the representation cloud fraction and related radiative effects in a number of climate models used in the IPCC fourth assessment report (AR4). Clouds pose a major source of uncertainty in climate models and their projections of future climate evolution, and therefore a comprehensive assessment of the ability of the models to reproduce clouds under present day conditions is a worthwhile (and necessary) endeavor. While many studies use satellite data to assess clouds in climate models, here the authors focus on surface based cloud observations, by including 3 observation sits from ARM in different climate regimes with comprehensive instrumentation. The authors make extensive use of the information on cloud characteristics available at the ARM sites compiled by 3 different observational methods. The comparison of these different methods as done by the authors is useful as it gives a good idea on the uncertainty in cloud fraction estimates as seen from the surface observation. I would have liked if also the classical synop observations of cloud fraction by human observers could have been added, as this is by far the most widespread and abundant information on clouds from the surface. In case such observations would be available at the ARM site, it would be interesting to get an idea how they perform compared to these more sophisticated methods, as a means to better interpret the many studies that have used synop cloud data before.

One of reasons why we include the Total Sky Imager (TSI) data in this study is the similar nature of the TSI and the human observations of CF that is more classic and wide-spread. The TSI in essence takes hemispheric "fish eye" color digital pictures of the sky each 30 seconds during daylight hours from a camera mounted looking down on a curved mirror. These images are then processed to infer what fraction of the sky view contains cloud elements, or fractional sky cover. One advantage of sky imagers over human observations is consistency of the retrieved results, where the subjective nature that affects human observations is removed. An overview and examples of this processing methodology are presented in Long (2010). Comparisons with TSK give overall agreement at better than 10% (Long et al., 2006) and with the Scripps Whole Sky Imager at the same level (Long et al., 2001).

Because the ARM Program fields a suite of sophisticated and consistent cloud instruments, and the subjective nature of human observations, no human observations are included in ARM observational strategy. A comparison between TSK and human observations from several sites of the Australian Bureau of Meteorology gives agreement at about the 10% level (Long et al., 2006), and by extension one can speculate about the same agreement for sky imager retrievals as well. We have added above information including references about TSI and its similarity with the human observations in Section 2.2.b.

Long, C. N., D. W. Slater, and T. Tooman, (2001): Total Sky Imager (TSI) Model 880 Status and Testing Results, Atmospheric Radiation Measurement Program Technical Report, ARM TR-006, 36 pp., Available via http://www.arm.gov.
Long C. N, T. P. Ackerman, K. L. Gaustad, and J. N. Cole. 2006. "Estimation of Fractional Sky Cover from Broadband Shortwave Radiometer Measurements." Journal of Geophysical Research. D. (Atmospheres) 111(D11):D11204, doi: 10.1029/2005JD006475.
Long, C. N. (2010): Correcting for Circumsolar and Near-Horizon Errors in Sky Cover Retrievals from Sky Images, TOASJ, 4, 45-52,doi: 10.2174/1874282301004010045.

On the modeling side of course one can argue that the models used in this study from the 4th assessment report are by now becoming somewhat outdated, reflecting the state of climate modeling at least 5-10 years ago, which somewhat limits the usefulness of this study. But as we experience such delays with the runs for the 5th IPCC assessment report, which are only now gradually becoming available, we can expect publications based on the 4th IPCC report models still to be published for a while. While I think it is too much to ask the authors to repeat their study with the new CMIP5/IPCC AR5 models for the present paper, I still would encourage them to do so as soon as a comprehensive set of model simulations will become available, to provide an assessment of more up to date models, which may then have a more immediate impact on model development. If already now the authors would be in a position to make any statements with respect to the applicability of their conclusions to the AR5 models, this would enhance the visibility of this study.

It seems that the final release of 5th IPCC simulations will be much behind the schedule. Based on the information at the PCMDI website, the first model output of IPCC AR5 was expected to be available for analysis by February 2011 and there is no date yet when all data are available, see http://cmip-pcmdi.llnl.gov/cmip5/.

Nevertheless, this is an excellent suggestion. We will consider repeating similar analysis as soon as the 5th IPCC model simulations are available and comparing with the results from IPCC AR4 models. We have added a few sentences regarding this in the paragraph on “future work plan” in Discussion section.

Specific comments:
p14939, L4. “more than a dozen”. I would say it is rather “two dozen GCMs” that provide data to PCMDI. All of them provide cloud fraction and all sky solar fluxes, most of them also clear sky solar fluxes. Therefore I slightly wonder why the authors only used 11 GCMs.

Yes there are two dozen GCMs providing data to PCMDI and most of them claim their data includes cloud fraction and all-sky and clear-sky solar fluxes. However, when we went to data archive at PCMDI server, we cannot find these variables for quite a few models. I have double-checked this. Certainly we prefer to have as much model results as possible.

P14940, L21 SkyRad, GndRad, and QCRad are not generally known acronyms and should be explained.

SkyRad stands for “sky radiation”system and are the downwelling broadband radiometer measurements. GndRad stands for”ground radiation” system are the upwelling broadband radiometer measurements. QCRad stands for ”Quality Control for Radiation measurements”. It is an ARM Value Added Product and is the recommended measure for all ARM site broadband radiation measurements. We have added explanations in the manuscript (Section 2.2.a).

Section 2.2: The observational methods used in this study could benefit from a more elaborate description, e.g. the TSI method is not explained at all.

We have added one paragraph (one subsection as Section 2.2.b) describing TSI method. See attached below.

b) Total Sky Imager (TSI)

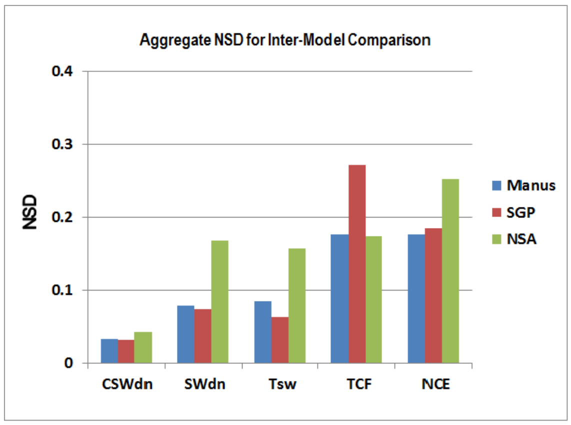
“The ARM observational strategy does not include human observations, however the TSI is the instrument most similar to a traditional human observation of cloud cover. The TSI takes hemispheric “fish eye” color digital pictures of the sky every 30 seconds during daylight hours from a camera mounted looking down on a curved mirror. These images are then processed to infer what fraction of the sky view contains cloud elements, or fractional sky cover. The processing uses the ratio of red to blue color values for each pixel in the sky image, except for that part of the image that is masked for the camera arm and sun blocking strip on the rotating mirror. One advantage of sky imagers over human observations is consistency of the retrieved results, where the subjective nature that affects human observations is removed. An overview and examples of this processing methodology are presented in Long (2010). Comparisons with TSK give overall agreement at better than 10% (Long at al., 2006) and with the Scripps Whole Sky Imager at the same level (Long et al., 2001).”

Section 4: How did you compare the gridded simulated fields with the point observations: e.g., taking the surrounding gridpoints, or the nearest gridpoint? For example at the coastal site Barrow, has it been ensured that only simulated land points are taken into account? Manus is a pure ocean point in the model, while observed cloud formation may be affected by the island. Is this taken into account in the comparison?

The GCM grid variables represent the average in a grid cell (e.g., 200 km x 200 km, which is a typical horizontal grid of AR4 GCMs). The fractions of land/ocean and different land surface types within each grid cell are accounted to calculate the grid cell mean, so the GCM output variables include the weighted contributions from ocean and land (or from different land types) within each grid cell. As other studies usually do, we first identify the grid cell (e.g., 200 km x 200 km) in which the observational site is located, and then use the model result at that grid cell to compare with the observations. Even though some GCM sub-grid variables have fractional values, without the geographical location for sub-gird ocean/land or land type information, we are not able to get the model output that can separate the contribution from land or ocean within each grid cell. This is one of fundamental problems in comparing the point observations with model results that have coarse spatial resolution, we have added a comment about this being an additional uncertainty in the comparisons to the conclusion section.

P14947 L27ff: why is this comparison only done for Manus, it would be interesting to have similar information from the other sites. There are a number of figures which could profit from a redesign:

We have added the results including the aggregate NSD (normalized standard deviation) for both inter-model and model-measurement differences from the other two sites (i.e., SGP and NAS) in Figure 5 (see attached below). We have also added discussions about those two sites Section 4.1.



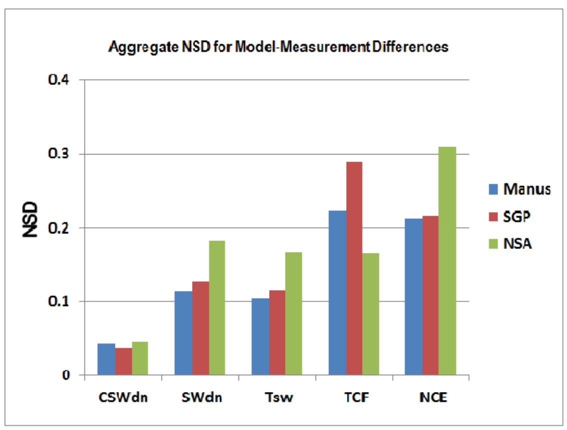


Figure 5

Figure 1: Horizontal Axis (years) is not readable

Large fonts are used, and this figure is modified to address comments by Reviewer #1.

Figure 3: Vertical axis should not exceed one, which would also allow to better differentiate between the different models.

Changed.

Figure 4: Why not put all 3 methods into one figure. The values for TSK are repetitive in the top and bottom figure.

We only chose the days in which both ARSCL and TSK or both TSI and TSK are available to calculate the frequency for ARSCL/TSK or TSI/TSK as shown in the original Figure 4. Because of different time periods with missing data in the three datasets, the numbers of sampling days used for TSK in ARSLC/TSK and TSI/TSK are different, so the calculated values of frequency for TSK in these two plots are slightly different, that's why we didn't combine the two plots into one figure.

In the new Figure 4, we compared the frequency of ARSCL and TSI, and TSI and TSK (the information of ARSCL/TSK can be inferred from ARSCL/TSI and TSI/TSK), and added two more plots for SGP and NSA. More discussions are added in Section 3.

Figure 6: The curve representing the GCM means is not well discernible, it could be represented as a thick line for example. Again the vertical axis should only go to one, which would allow a better separation of the different GCM results. Legends may then be outside the figure.

We have modified this plot. The GCM mean is represented by a different type of thick line. Vertical axis only goes to one, and legends are only presented in one panel. We will reminder technical editor to keep reasonable size for this figure to ensure quality as the paper being formally published.

Figure 7: Why is only one type of observational method (TSK) shown here?

First of all, TSK has longer records than the other two. Here PDF is calculated based on monthly mean TCF since we only have monthly mean TCF from models. Usually larger samples are required to construct a more representative PDF, so we just use TSK that has longest records. Secondly, the difference of PDF among the three datasets is expected to be small at monthly mean level (much smaller than that at daily level as shown in Fig. 4). So we only choose one set of observations rather than all three to compare with model results. We have added explanations about this in Data introduction section.

Figure 9: Figure caption not detailed enough, it should be explained what is shown in the left and right panel, respectively. The number of years that went into the averages should also be mentioned. This applies also to some of the other figures.

Suggestions have been taken. Thanks.

Typos/small changes:

P14935 L25 should be clouds, similarly P14936 L1 Clouds are also: : :

Done.

P14937 L19 replace "difficulty" by "impossibility"

Done.

P14940 L7 clouds and their radiative forcings

Changed.

P14945 L 14, should be (Fig. 3, middle), or add a), b) c) into the Figure, if you refer to them in this way.

Changed to (Fig. 3, middle).

P 14948 L2: one of TWP: something missing

Changed “one of the TWP sites (Manus)” to “the Manus site”.

P 14948 L10 Fig. 5a should be 5 (top), or add a) and b) into the Figure, if you refer to them in this way.

We have modified the caption for this figure.

P14949 L8, should be Fig. 5, not 4, and bottom, not a)

Changed.

P14951 L16: “too frequent overprediction”, is too much. Better “too frequent large cloud cover”

Changed. Good suggestion!

P14951 L25 has a similar

Done.

P14958 L21 lower levels

Changed.

P14959 L20, should be Fig. 14, not 13.

Changed.

P14961 L 7, with “transient” do you mean “daily”?

We mean daily or hourly or any shorter time periods than monthly. We added “hourly or daily mean” to make it clear.

P14962 L10 near the surface

Done.

P14963 L12 should be “from the 1990 at most of the sites”

Changed.