

**Interactive comment on “Total cloud cover from satellite observations and climate models”
by P. Probst et al.**

Anonymous Referee #1 (Received and published: 20 October 2010)

The paper compares CMIP3 model produced monthly mean climatologies of total cloud cover against similar satellite derived measurements. Global annual means, averages for large latitude bands, specific seasons, and the strength of the seasonal cycle are investigated. It is found that most models underestimate total cloud cover by about 10-15% for the global annual mean, with a large intermodel spread. This bias is most pronounced in subtropical and higher latitudes, and to a lesser extent in the inner tropics. However, the models reproduce quite well the zonally varying structure of cloud cover and its typical seasonal cycle. It is also found that the interannual variability of total cloud cover is underestimated by the models.

1. I do not find that the analysis and results presented in the paper are particularly new or interesting. It is well known that the simulation of clouds is a major task for any type of model, and I am actually quite surprised that the mean model bias amounts only to 10-15%.

Fig. 1 in our manuscript shows that most models feature a “remarkable negative bias of the order of 10-15%”. We did not refer to a “mean model bias”. If the Reviewer were to define criteria (and associated references) to back up his statement that a mean model bias amounting to 10-15% is a surprisingly good result, then we would be happy to comment on it.

2. In addition, the main results of the paper, that the CMIP3 (and other) climate models tend to underestimate both total cloud cover and its interannual variability, have already been shown before (Pincus et al. 2007, Figure 2). The amount of new information presented is small and the shown analysis is quite simple, which taken together does not justify publication in a “stand-alone” paper.

We suspect that the reference mentioned is Pincus et al. 2008 (which will be denoted as PBHTG from now on) as we have not found any reference to Pincus et al. 2007. PBHTG is mentioned in our introductory section, without much detail because we were not particularly impressed by it. It is in fact extremely extensive (as it deals with many parameters and many datasets), but there is no discussion of the datasets, with a “user-oriented” attitude that leaves any discussion on the intrinsic limitations of a given dataset to the data providers. The impression one gets is that it is sufficient to take a dataset, apply some careful, but simple, algebraic manipulation to extract basic statistical properties, without the need to understand in depth the limitations of each set. Figure 1 of PBHTG presents the complete results in a semi-quantitative diagram. The best that can be said about Figure 2 is that it is nice looking. It is virtually impossible to extract useful quantitative information on the behaviour of the models under scrutiny since the results are plotted on top of each other and the last data to be plotted relate to the “IPCC mean model”. Moreover only two metrics are shown, namely the standard deviation and the correlation. PBHTG states that the “IPCC mean model” is constructed by averaging the monthly mean fields provided by each model. No indication is given of how the average is obtained, which weights are applied (all models have same weight?). Using a Bayesian language, the predictive uncertainty of the average model is heavily dependent on the cross correlation matrix of the model ensemble. Which is the actual cross correlation among the models that were averaged? Without a quantitative answer much of the discussion in Section 4 of PBHTG could and should be questioned.

Our work shows an analysis of cloud cover fraction of total cloudiness, showing not only the global analysis, as in PBHTG (with the mentioned difficulties to extract useful informations from Fig. 2), but also a zonal and seasonal analysis, unlike the other previously published articles.

3. I also find that some of the interpretations and conclusions of the paper are overly strong, for example that “the models feature a remarkable negative bias” or that

See our reply to statement 1. In any case the adjective “remarkable” can be dropped as it does not add much to the content of Fig. 1.

4. “models seriously underestimate inter-annual variability”.

Figures 4 to 7 were constructed to provide clear evidence, and they show that the models (seriously) underestimate the inter-annual variation. Again there must be caution when using adjectives, but Fig. 4 to 7 show clearly what we meant with “seriously”. In any case the adjective can be dropped with no loss of information since the figures show that the half-horizontal bars of the models are very small compared to those of the observations, especially in Fig. 4 (60N to 60S) Fig. 5 (from Equator to 60N and from Equator to 60S) and Fig.6 (from Equator to 30N and from Equator to 30S).

5. Also, some of the explanations and interpretations given in the paper are common text book knowledge (e.g., that baroclinicity and cyclogenesis is stronger in winter) and should only be mentioned briefly in a scientific paper.

By the same token, we expect that the Reviewer would ask any paper dealing with climatological cloud properties not to mention its relation to the planetary albedo, because this is a far simpler (energy balance) physical process. Again, a classification by the Reviewer of processes worthy and unworthy of being mentioned, with explanation and references to justify the choices, would be welcome. Otherwise, the comment is entirely subjective.

In our paper the explanations and interpretations are described very concisely, but references to other papers and books containing more information are given. Our replies to the Reviewer contain more technical details, whenever it is appropriate.

For these reasons I do not recommend the publication of the paper in ACP.

6. I also have a number of more technical comments. For example, it is mentioned in section two that the ISCCP data have difficulties with overlapping clouds, but does this really matter for the present paper?

The study has taken time, as we have tried to understand the limitations of the cloud products, as they are the results of complex processing. We think it is important to know the limitations and problems when using a dataset, to understand if some unphysical factors may act together to determine spurious statistical properties for the parameter under study.

7. Also, only one observational data set is used in the study, making it impossible to judge how large the observational error is. However, this study is on the very specific and isolated topic of total cloud cover, and one can expect that more attention is devoted to this issue.

The ISCCP dataset is the best known and most widely used cloud dataset for the period analyzed. An estimate of “distance” between ISCCP D2 data and MODIS/Terra Collection 5 cloud fraction estimate, as measured by the standard deviation and cross correlation, is given in PBHTG Figure 2 (luckily the red spot is clearly visible). We do not wish to comment on the use of this “distance” as a measure of observational error.

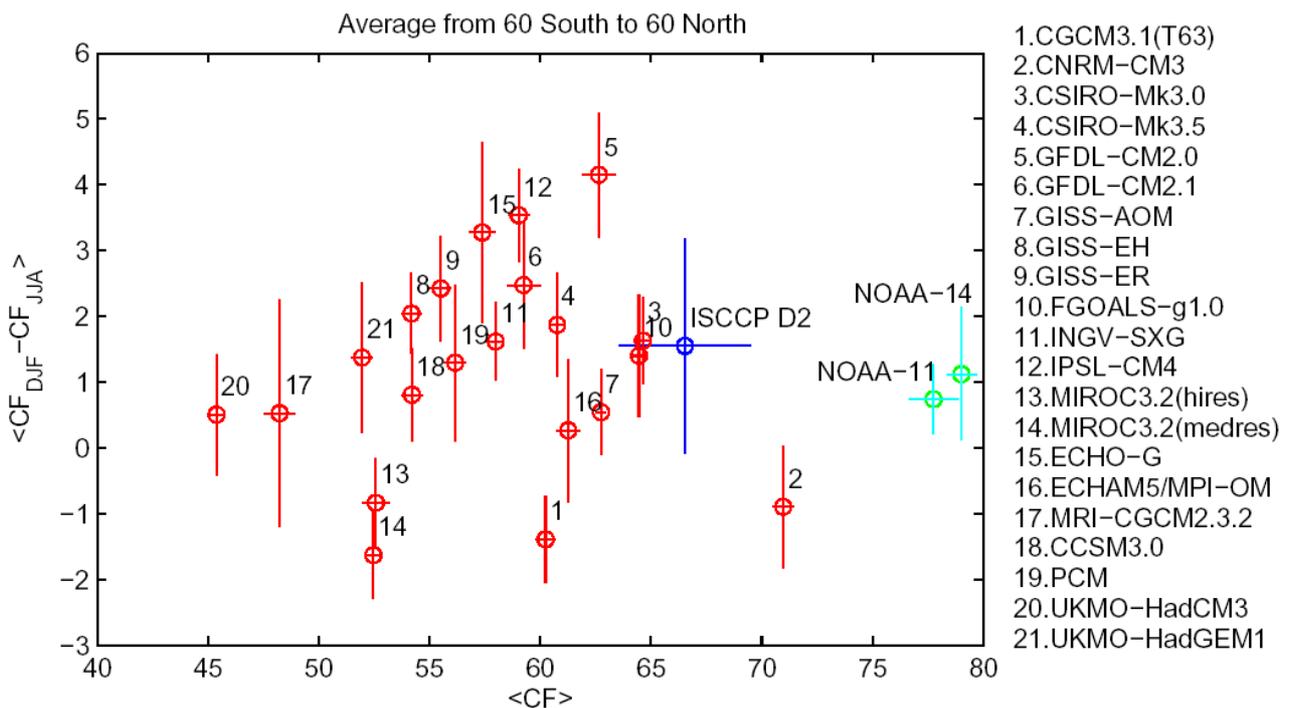
We have compared also the ISCCP products against products obtained with a technique that employs the High-resolution Infrared Radiation Sounder (HIRS) data (Wylie, D., D. L. Jackson, W.

P. Menzel, and J. J. Bates. Trends in global cloud cover in two decades of HIRS observations. *Journal of Climate*, 18 (15):3021–3031, 2005). For example the figure at the end of this reply is same as Fig. 4 of the manuscript, but including the results obtained from the datasets of two different NOAA polar satellites, NOAA-11 and NOAA-14.

However:

- cloud parameters in the HIRS dataset are derived using some of the sounder absorption channels, while the ISCCP products are derived from imager's window data;
- while a typical imager spatial coverage is nearly contiguous, the sounder measurements have gaps (the sampling is not contiguous) with typical distance between two adjacent FOVS of about 40 km on average; hence a number of consequences arise, and among them that the product of HIRS processing is a cloud frequency, not a cloud area fraction as in ISCCP (see reply to statement 8);
- the globe is sampled only twice a day from one polar spacecraft, and the observations and products cannot represent the daily changes;
- we have received from Menzel and colleagues the cloud products for each satellite and combining them into a single dataset is a difficult issue;
- each polar satellite covers a portion of the time interval for which we have ISCCP (and model) products (hence some of the statistical measures, like the inter-annual variability, can hardly be compared).

These are the main reasons why it was decided not to include the HIRS products in our manuscript.



Please note that in the figure above the analysis period is 1984–1999 for ISCCP D2 and CMIP3 models, 1989–1994 for HIRS NOAA-11 and 1996–1999 for NOAA-14.

8. In addition, the paper does not discuss the problem of defining of what actually defines a cloud. It is very likely that models and observations use different criteria and thresholds for their total cloud cover definition, making it likely that some of the shown discrepancies are related to this lack of a common definition.

Many “theoretical aspects” are not detailed in the paper, but only briefly mentioned. For example a complete discussion on how the cloud cover is determined from satellite observations and how the

models simulate them is not dealt with, but several papers are indicated where more detailed information can be found.

To answer the reviewer, the variable of the PCMDI/CMIP3 models is the cloud cover fraction, while that of ISCCP is the cloud amount.

The ISCCP definitions are the following:

- Cloud Amount represents the frequency of occurrence of cloudy conditions in individual satellite image pixels, each of which covers an area of about 4 to 49 square kilometres. Comparisons to other measurements confirm that this quantity also represents the fractional area coverage at any one time for the larger 280 km grid cell areas.
- Cloud cover fraction represents the fractional area covered by clouds as observed from above by satellites. It is estimated by counting the number of satellite fields-of-view (about 5 km across for ISCCP) that are determined to be cloudy and dividing by the total number of fields-of-view in a region about 280 km across.

More information and discussion can be found in Rossow et al. (1993) and Rossow and Schiffer (1999).

9. The paper also mentions several times that there is little consistency amongst errors from different models, and this behavior is interpreted as a negative aspect of model performance. I actually would interpret this discrepancy in the opposite way, since this shows that models are more or less statistically independent and that model errors cancel each other to some extent in the multi-model mean.

The reviewer is adopting the point of view of PBHTG, which we have already discussed in reply to statement 1. A statement like “*models are more or less statistically independent*” should perhaps undergo a close scrutiny. One should consider that the model data set contains runs (with different configurations) of the same model (CSIRO, NASA/GISS, GFDL, Met Office, and ECHAM is the core model of 3 different groups). The “intra-model” results, that are clearly shown in our manuscript, are sometimes very similar and sometimes show appreciable differences. Following a similar line of thought (as statement 9) one could suggest that it would be advisable to simulate the 20th century climate using as many runs of the the same model with different sets of physical parametrisations as necessary to average out random errors and biases. The issues of statistical correlation is an open issue for climate models' runs, we believe.

10. I also find that the introduction and methodology section of the paper are unnecessarily long and detailed (e.g., Table 1 has already been shown in a countless number of previous papers), in particular compared to the little amount of information given in the result section.

We have already replied to the content of first part of statement. Regarding Table 1, unfortunately, it is used as reference for the symbols in the figures and cannot be easily eliminated.

11. Similar comments hold for the conclusion section, which mainly constitutes a long summary of the results. The paper also contains a number of (minor) errors in the English language.

We agree with the comment, and will simplify the conclusion section, that now constitutes a summary of statements discussed previously.

We do our best to use a correct English, but are not mother tongue. One must also consider that ACP(D) is an European journal, not Anglo-Saxon and certainly not American (meaning U.S.).

**Interactive comment on “Total cloud cover from satellite observations and climate models”
by P. Probst et al.**

Anonymous Referee #2 (Received and published: 20 October 2010)

The manuscript compares observed cloud covers (cloud cover fraction, CF) with simulated data in 21 climate models. The authors compare global CF, zonal means, mid latitudes and tropical averages, as well as seasonal cycles. The comparison reveals severe discrepancies between model and observational results.

The study is a useful contribution to model assessment and is certainly of interest for the development of the models included. Unfortunately, the manuscript resembles a technical report.

To address a more general readership, the manuscript should be extended by a more climatological discussion:

1. In addition to the latitude belts, a land/ocean comparison would be useful.

Our work aims to evaluate the performance in simulating the CF of climate models and not the climatological consequences. Several articles dealt with the relevance of a correct cloud simulation, within climate modelling, as shown in the introduction section of our article.

The land/ocean comparison would be useful in case the global simulation results were closer to the “observations”, but we have shown (significant) discrepancies in our annual and seasonal analysis that already merit attention.

2. Why is one model (CNRM-CM3, in part also CSIRO versions) superior to all other models?

The purpose of intercomparison works, and of our work, is to discuss some results that may or may not indicate basic modelling problems. It cannot provide insight on the reason behind particular results obtained by some models. In order to address this problem one needs to apply more accurate diagnostics and specific simulation strategies.

3. What are the consequences for the interpretation of scenario simulations?

As mentioned earlier our work is concerned on the fact that the climate models under scrutiny do not adequately simulate the cloud amount rather than discuss the consequences (several articles deal with the relevance of a correct cloud simulation within climate modelling).