CGMS-35 EUM-WP-25
v1, 4 September 2007
Prepared by EUMETSAT
Agenda Item: II/6
Discussed in WG II

REPORT OF THE CLOUD PARAMETER RETRIEVAL WORKSHOP

In response to CGMS action 34.26

A satellite cloud parameter retrieval Workshop was held near Norrköping, Sweden 17-19 May 2006. The Workshop was organized by SMHI and sponsored by EUMETSAT. Mainly European experts working with operational cloud parameter retrieval schemes were invited to participate in the workshop and to contribute results of their algorithms to a cloud parameter inter-comparison campaign. The algorithm intercomparison was focussing on MSG cloud products for one day of data, 17 January 2006. Intercomparison work was carried out as a visiting scientist activity by Andy Walther within the SAF for Nowcasting and Very Short Range Forecasting.

This paper has been provided by Dr. Anke Thoss from the Swedish Meteorological and Hydrological Institute.

Report of the Cloud Parameter Retrieval Workshop

17 – 19 May 2006, Norrköping

Anke Thoss, Swedish Meteoroligical and Hydrological Institute

1 Background and Objectives

A satellite cloud parameter retrieval Workshop was held near Norrköping, Sweden 17-19 May 2006. The Workshop was organized by SMHI and sponsored by EUMETSAT. We invited mainly European experts working with operational cloud parameter retrieval schemes to participate in the workshop and to contribute results of their algorithms to a cloud parameter inter-comparison campaign. The algorithm intercomparison was focussing on MSG cloud products for one day of data, 17 January 2006. Intercomparison work was carried out as a visiting scientist activity by Andy Walther within the SAF for Nowcasting and Very Short Range Forecasting.

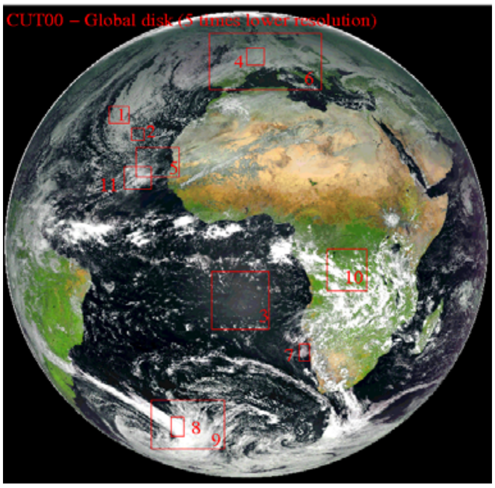
With new generations of satellite sensors, advances in methodology and new user requirements on the horizon, it was felt that it was time to discuss how operational cloud parameter retrieval algorithms can and should evolve while ensuring continuity of products and service to users.

The workshop addressed the following topics:

- Clarification of most important goals for cloud processing - now and in the future (*presentations of algorithm developers and users, discussions*).
- Overview of existing methods with honest appraisal of strengths and weaknesses *(as presentations of the algorithm developers)*
- Discussion of first inter-comparison results *(three working groups: WG1: Cloud mask, WG2: Cloud top products, WG3: Cloud microphysical products)*
- Identification of areas to be addressed in future work, potential for collaboration, future inter-comparison campaigns etc.(*in working groups and panel discussion*)

2 Results of intercomparison campaign

An algorithm intercomparison of twelve MSG cloud algorithms was carried out for 17 January 2006. When comparing data for just one day, it has to be kept in mind that this just presents a first intercomparison and not a proper validation, which would not be statistically significant for one day of data anyhow. It is however possible to identify in which situations and for which parameters algorithms agree well, and to also identify some problem areas which need to be addressed in further investigations.



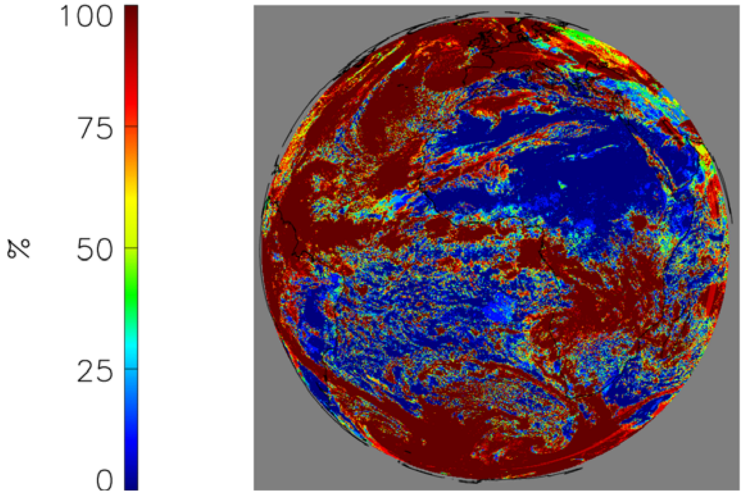


Figure 1: Areas selected for intercomparison

Figure2: Percent of algorithms indicating cloud cover 17 January 2006 12GMT.

2.1 Cloud Mask results

- Disagreement between cloud masks was mainly found in known problem areas (for example high viewing angles, extremely low surface temperatures, snow cover, temperature inversions, over desert areas, thin cirrus)
- Some thresholds had to be set for algorithms supplying effective cloud cover, cloud cover likelihood etc..This resulted in differences for cloud edges.
- One to one intercomparisons of algorithms are useful to understand differences and improve algorithms further.
- Cloud mask is easy to verify against synop – verification recommendations were formulated to make intercomparisons of individual validation results possible.
- Other cloud parameters should only be compared where cloud masks agree.

Figure 2 presents the amount of agreement for the 9 cloud masks which provided data for the entire MSG disk. Dark blue indicates that all algorithms report cloud free, brownish red indicates that all algorithms report cloudy.

2.2 Cloud top product results

- Cloud top temperature disagrees mainly for high clouds, likely due to different treatment of semitransparency.
- Cloud top pressure disagreements might additionally be caused by following factors:

1. differences in strategies on how to search fit to NWP profile (top down or bottom up)
2. Differences in NWP profiles used, especially with regard to tropopause treatment and inversions
3. different RTM models used

- Conceptionally similar algorithms might give different results. This which needs further investigations.
- There is an apparent need for thoroughly planned validation campaign to resolve open issues

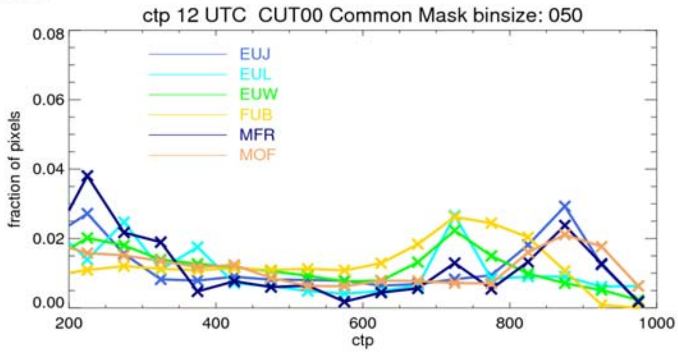


Figure3: Cloud top pressure histogram for global data. Cloud top pressures disagree for both high and low clouds.

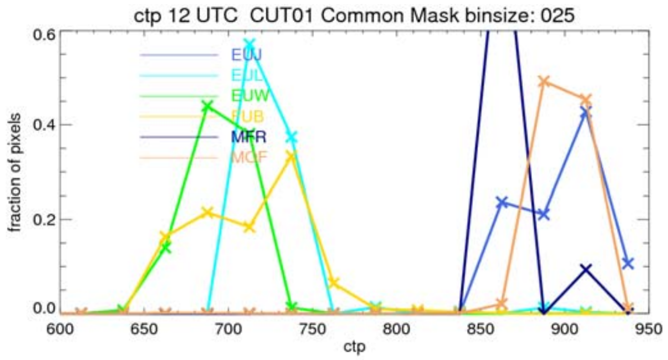


Figure4: Cloud top pressure histogram for a stratocumulus area. Cloud top pressures fall into two distinct clusters, depending on the strategy of how to search the NWP profile for a fit (bottom up or top down).

2.3 Results for microphysical parameters

The Working group on microphysical parameters noticed differences which might have been due to intercomparison methodology, such as initially not stratifying after cloud phase. Also a shift of data of one pixel for one of the algorithms was detected. Another algorithm supplied a new dataset to eliminate a time difference of 15 minutes to the other datasets. Some new areas of interest where recommended for further investigation. The recommendations of the cloud microphysics working group with regards to the intercomparison were to a large extend implemented in a rerun of the intercomparison. The new results significantly reduced the scatter between different algorithms. Stratifying after cloud phase did however not suffice to explain all systematic differences. The new results would need further in depth discussion.

- Effective radius shows a fairly good agreement with correlations between the algorithms of 0.79 to 0.84 and a bias of 0.1 to 2.1.
- Effective radius for ice phase is stronger correlated between algorithms than for water phase. However very close agreement of effective radius was found for stratocumulus cases
- Optical depth: when studying specific situations systematic differences between algorithms are prominent.
- For two different stratocumulus cases agreement between algorithms varied significantly. This might be due to different algorithm performance at higher viewing angles.
- The dataset is a good starting point for in-depth analysis. It is expected that algorithms can be improved with findings from this analysis.

- A dedicated AVHRR intercomparison would be useful

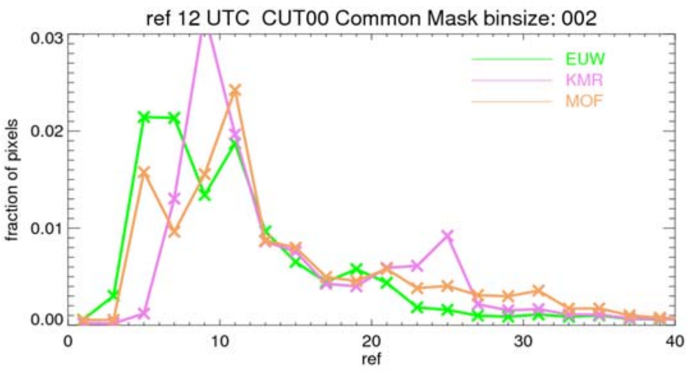


Figure5: histogram of cloud effective radius for the MSG full disk

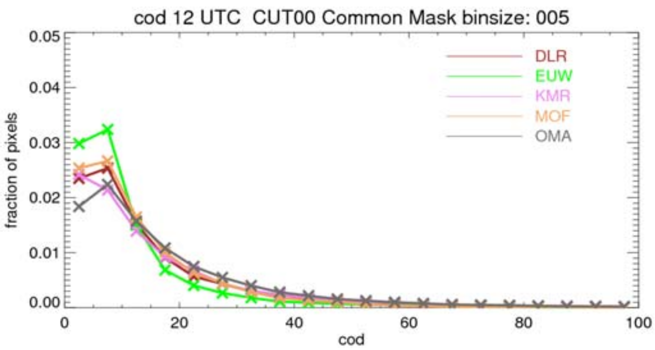


Figure 6: histogram of cloud optical depth for the full MSG disk.

5 Conclusions and Recommendations

- The Working groups felt that the workshop and intercomparison was very useful and would like to continue to meet at about two yearly intervals.

- The group would welcome an effort of EUMETSAT to coordinate validation and intercomparisons of MPEF and SAF cloud products in the future, even with possibility of other operational algorithms to participate. A dedicated effort to validate cloud products against CloudNet data and CloudSat- CALIPSO data would benefit algorithm development greatly. Common validation efforts are specifically needed for cloud top and microphysical products.
- It should be investigated whether cloud parameter retrievals can be contributed to CloudNet and standardized.
- All operational cloud algorithms (also outside Europe) should be invited to future meetings and intercomparisons .

Annex: Working group reports (for Eumetsat internal use)

Annex 1: Report of the working group 1 on Cloud Mask
Rapporteur: Paul de Valk

The group identified the need for a ground truth. The Synops reports are accepted as a ground truth. The group is aware of the short comings of the Synops reports, like increasing automated stations missing sometimes high clouds, manual observations at night are not perfect no good quality outside Europe, limited coverage over sea.
In order to get comparable statistics the group decided on the use for the synops. In the first step of the validation only the 0,1 and 7,8 octa reporting synops station are used. The synops drives the selection of cases. (So all participants should have the same number of cases, provided the overlap in area is complete). A collocation area of 5x5 MSG pixels around the reporting station is chosen. There is a preference to use the HH:45 images for the comparison, as these have the smallest observation time difference with the synops.
From the selected cases the contingency table and all related scores can be determined, like the kuyper skill score, POD, correct clear, correct cloudy scores, bias (using all synops reports 0-8).

Commitment: DWD and KNMI will on a best effort basis provide hourly synops reports for January 17, 2006, including information on location within the Meteosat 8 image (columns lines) This will be ftp-eed to the EUMETSAT/CWS site.

In the future there will be exchange of synops information to provide a ground “truth”.

The group reached consensus on a one to one comparison of cloud masks results and discuss the differences. This is the only way to understand the differences and to improve the cloud masking. RGB composite combining 3 different cloud masks is an easy tool to highlight the agreement and the differences.
Also polar satellite cloud masks should be compared to geostationary cloud masks.

The group is aware of the calypso and cloudsat data becoming in 2006. It is welcomed but at the time of the workshop, May 2006, it is not clear how we will benefit from this data.

The group is in favour of a future workshop. An agreement should be reached on the comparison dates (e.g. one week in each season, leading to four weeks in total). For cloud masking there is an easy accessible ground truth, synops. This makes it easier to compare for longer periods oppose to e.g. ctth and ctp.

Recommendation to EUMETSAT
The group is very positive about this comparison activities achieved in this workshop. They would welcome an effort of EUMETSAT to coordinate validation and comparisons also for SAF products in a more substantial manner in the future.

The group would welcome continuation of this workshop, possibly as a side meeting to (EUMETSAT) conferences.

Annex 2: Report from Working Group 2 on Cloud top products

Cloud Workshop Norrköping 17-19 May 2006
Karl-Göran Karlsson, 'rapporteur'

Other group members:
Hans-Joachim Lutz
Herve Le Gleau
Bryan Baum
Sheldon Johnston
Ruth Taylor

Question 1: *How do we best exploit the existing dataset?*

It is evident from results that cloud top results from the different algorithms differ to a very large degree. However, since a large part of the deviation comes from areas where the cloud masks differ (mainly because of different methods when detecting fractional or thin clouds) it is suggested that one should primarily start with an analysis of results where the majority of methods agree on cloud occurrence. This means in practice that the capacity of the various algorithms to treat opaque or thick clouds is the most obvious starting point for a deeper analysis.

A follow-on step could then be to try to make some conclusions about the treatment of semi-transparent or fractional clouds. However, since there is generally much less agreement here between the algorithms regarding the identification of these particular clouds the present dataset is perhaps not the most suitable for such studies. For future studies this problem has to be specifically addressed (see further discussion regarding Question 5 below).

Question 2: *Are there any specific conclusions (differences and reasons)?*

Results are indeed very different for some algorithms. Furthermore, even results based on the same method differ substantially (in this case the operational MPEF method - denoted EUJ - and the same method applied in a development environment – denoted EUL). From this it is clear that the influence of some of the basic tools (e.g., RTM methods) and the associated auxiliary input data (here mainly NWP profiles) can be as important as actual algorithm differences.

The group generally agreed that some of the differences we see in the results most probably come from differences in the used NWP input data. However, the actual method which is chosen to relate radiances or temperatures to the reference NWP vertical profiles might be even more important. For example, in cases with temperature inversions (if present in model profiles) it is evident that one gets different results when one compares with temperature profiles starting from the surface and going upward compared to if one is starting the matching from upper levels going downward. The group was of the opinion that such method differences very well could be responsible for a large part of the differences seen in the results. In addition, even if there are no particular problems with temperature inversions there is still always the question how clouds close to the tropopause level are treated by the methods. For example, in the case of overshooting tops of deep convective clouds– will these be assigned a realistic cloud top height or pressure? It very much depends on the specific matching method.

However, to be sure about this aspect of the problem some deeper studies appear necessary. A recommendation for a final report or presentation regarding the cloud top products could be to at least list the basic approaches for the matching to vertical reference profiles (NWP or alternative, e.g. climatological profiles) for each participating algorithm. Also an intercomparison of our various definitions of tropopause height based on NWP model profiles could be valuable as well as an intercomparison of our various definitions of lower tropopause inversion detection. For example, how strong a temperature inversion, say 0.5 K, has to be present to justify flagging the presence of an inversion?

The group briefly discussed the degree of dependence of NWP model information. It was stressed that current NWP models still have problems in correctly depicting near-surface thermal inversions. Especially, the cold extremes are still not well simulated by NWP models which tend to produce too weak inversions and too warm surface temperatures. This could in turn be a consequence of incorrect cloud descriptions in the NWP model which shows how sensitive this topic really is in this context!
Noteworthy is that there are also methods (FUB) which do not rely on any modelled profile information. However, at this stage it is impossible to state which of the methods is the most realistic one since we have no validation data available.

A final comment is that it is striking to see how the results go from a relatively good agreement for the cloud top temperature product (except for differences related to the correction of semi-transparent clouds) into a broad disagreement for the cloud top pressure product. Once again, this shows how sensitive the cloud top pressure interpretation is to the used reference profile and the method to match to the reference profile.

Question 3: *How do we want to present inter-comparison results to a wider scientific public?*

In the short-term range (e.g. concerning the presentation at the EUMETSAT conference in Helsinki), the group recommended the following content to be emphasized concerning the cloud top product information:

1. First show results for cloud top temperatures where we have general agreement except for some differences (probably due to different treatment of semi-transparent clouds)
2. Then show cloud top pressure results where disagreement is pronounced.
3. Try to point at major agreements-disagreements in the common cloud mask (i.e., where most algorithms agree on cloud occurrence).
4. Say something about the most likely causes for creating this disagreement (e.g., NWP model input differences, different methods to match to profiles – down-up or up-down, etc.)

Concerning the more general question whether we could continue later on with additional scientific publications based on this dataset, the group was a little bit pessimistic. First of all, we only have data from one particular day and this naturally limits any prospects of making firm conclusions in the statistically significant sense. Secondly, the lack of validating observations means that we really don't know which of the algorithms that actually compares best to the real cloud situation. This situation means that it is really not recommended at present to use any subset of the data as a reference or baseline for any specific scientific publications.

On the other hand, it was still felt that the present dataset contains a lot of interesting information that could be used further. Consequently, any further presentation of the results depends to some extent on any additional efforts in analysing the dataset more thoroughly (discussed in next section).

Question 4: *Who will commit to which work?*

This question did not get any clear answer in the group discussions. It was felt that it is now up to each responsible (for algorithms) team to further explore this dataset within their given resources. More work will definitely be done but no one made any specific statement on particular studies associated with real commitments.

The only commitment that was agreed upon from all in the group was the wish to participate also in future inter-comparison activities. This is discussed further below.

Question 5: *Future inter-comparison: Why? What? How? Who?*

Everybody in the group was of the opinion that inter-comparison activities are very valuable. If results are used wisely, they lead to a better understanding of weaknesses and strengths of algorithms which in turn will lead to an overall improvement of product quality in the end.

Several in the group wanted an inter-comparison activity to be a regular event to be done approximately every second year. These events should not be done too often (e.g. every year) since there must be time enough to take advantage of results and improve algorithms before the next inter-comparison event.

Many in the group expressed that the next inter-comparison activity should also include a certain amount of validating observations. For the cloud top products it would mean that the additional product *cloud top height* has to be included in the study since this is the normal observed quantity from e.g. cloud radars and cloud lidars.

An important aspect of future inter-comparisons would be to see if a more organised and strict use of NWP model information could be applied because of its profound influence of product quality. Ideally, all algorithms should preferably be using the same NWP model input. If this is not possible, one should at least try to present how a particular method’s NWP data is comparing to a particular reference data set. For example, a reference surface temperature map could be defined. Possibly also similar reference profiles of both temperature and moisture could be defined. Each participant would then either have to use these reference data or at least describe the corresponding departures from this reference dataset for their particular algorithm.
However, Ruth added an important point in this context: If a reference NWP profile is used with all algorithms it may lead to an even wider spread of results, as algorithms could contain tuneable thresholds and other parameters that attempt to compensate for deficiencies in the particular NWP profiles that they have been developed to use. So perhaps there should be caveat that NWP as a source of inconsistency can't necessarily be cleanly removed. This circumstance must also be taken into account (if possible) in new inter-comparisons. At least one could think of the procedure that each participant delivers two results: One based on the preferred NWP model and one based on the chosen reference model. In that sense it would be possible to test the algorithms dependence on one particular NWP model.

Bryan suggested also that other types of reference data sets could be considered. Especially he pointed at the possibility for using a common surface emissivity map. He suggested that people already now could start looking at the following web site describing such an IR emissivity map:

http://cimss.ssec.wisc.edu/iremis/

This data set contains emissivities at 6 wavelengths located at inflection points, and then a Fortran routine is provided that shows you how to interpolate between the inflection points to obtain a more complete set of wavelengths. This is what is available now but in the near future (anticipated for late June) a high spectral resolution data set will be made available on this web page.

The current inter-comparison reveals that it is difficult to make any firm conclusions on the ability of different algorithms to correct for semi-transparency of clouds. Maybe it would be appropriate for a future study to try to identify a set of cases with truly semi-transparent clouds (as identified also using appropriate ground-based observations) which could be tested separately.

Concerning the content of data for the next study it was clear that it should naturally contain data from other (European) seasons than just the winter season. However, it was not clear whether it was still preferred to stay with single day information or whether longer periods were needed. Everybody in the group realised that the data amount increases rapidly if expanding the covered period.

Finally, regarding the question *WHO?*, the general feeling was that this should preferably be organised centrally from somewhere which naturally points at EUMETSAT as the most suitable candidate. The reason is that it would be difficult for any particular member state institute or university institute to find the necessary resources to cope with the organisation of such a study. However, this does not rule out solutions similar to the current inter-comparison where the organisational task is distributed to a member state with financial support from EUMETSAT.

However, some different views on exactly how to do it were heard in the group. Rather than just promoting follow-on studies to the present ones, many in the group wanted a heavier commitment to the validation issue besides the continuation of the present inter-comparison concept. Hans-Joachim Lutz mentioned that it is currently discussed at EUMETSAT to define a call for an invitation to tender (ITT) concerning the use of CloudSat – CALIPSO data for validation purposes (and other applications). This would be very interesting in this context and the group recommended the following cloud parameters or cloud-related features to be specially studied:

- Cloud motion vectors
- Cloud top height
- Cloud boundaries
- Cloud phase
- Depolarisation effects (for cloud-aerosol interactions, e.g. related to volcanic activity)
- Cloud extinction profiles (for studies of geometrically thick but optically thin clouds)

The group also suggested that EUMETSAT should consider yet another ITT more specifically focussed on cloud validation in a more general and continuous manner. This would be an extremely valuable help for the ongoing algorithm development. Somebody should be given the task to collect time-series of simultaneous observations and satellite data and do the validation on a more regular basis. Especially the task to collect this 'Universal Matchup Database' should at least cover the current CLOUDNET sites and its closest surroundings due to its importance for cloud top product validation. It was even stressed that such regular validation work should not be done by developers but instead by independent groups or institutes. This would ensure that different algorithms are treated the same way which would facilitate the interpretation of results. The group recommends EUMETSAT to consider this proposal very seriously.

Appendix 3: Report of working group 3 on Cloud Optical Thickness, Particle Size and Phase

Participants: Pete Francis, Andy Heidinger, Alexandro Ipe, Rob Roebeling, Phill Watts and Ulrika Willen
Rapporteur: Rob Roebeling

1. How do best exploit the data we have?

The data set as it is now is a good staring point for intercomparison studies. However, the first analysis results consisted of some bugs and inconsistencies. Before presenting the results to a wider community the following modifications are proposed:

- Fix bugs in analysis. Need a technical iteration to ensure data was used correctly.
- Apply better filters/stratification on the analyzed products, e.g. separate water from ice clouds.
- Was the quality flag, for those groups that provided it, used in the analysis? (according to Andy this was done)
- It was proposed to do a comparison with a common cloud mask for all data-sets instead of using a cloud mask that is just common between the two data-sets (groups).
- In addition to the scatter plots with linear scales the results should also be presented on a log scale to stretch out low optical thickness range.
- Use a common phase to stratify results.
- The MOF images have a one pixel shift compared to the other groups.
- For the KMR retrievals the 11:45 observation was used instead of the 12:00 hr observation. This is most probably the explanation for the large scatter between the cloud properties of KMR and the other groups.

2. How do we explain the differences?

Despite for the differences due to the analysis method the intercomparison results provided a wealth of information for discussion on the differences between the results of the different groups. The scatter plots and frequency distribution were very useful to identify biases and variability between the retrievals of the groups. The images of the different CUTs could be used well for a more detailed analysis of spatial differences between the retrievals. Going through the intercomparisons results of CTT, COT and particle size the following remarks were made:

- Part of the differences may result from calibration differences. To rule out uncertainties about the calibration the reflectances and brightness temperatures that the groups used in their retrievals should be compared over a limited area. In addition it needs to be verified if the same angle computations and sun-earth distance approaches are used.
- The retrieval of cloud properties does largely depend on the accuracy and the parameterization (e.g. aerosols and water vapor profiles) of the forward model (radiative transfer model). Because the retrievals are very sensitive to differences in the forward model simulations it would be useful to compare simulated reflectances of the different groups for some cloud cases.
- The scatterplots of COT and particle size seem to show a clear difference between ice and water cloud (bimodal grouping). Since the phase is so important for the retrievals of cloud microphysical properties it would be useful to include the comparison of Cloud Phase in this study as well.
-

- Beside the accuracy of the forward model, some patterns in the scatter plots may be explained by Lookup table fidelity and interpolation.
- The differences observed in Cut 5- phase may explain the bimodal grouping in the scatter plots.
- Check time of images are really 12Z. Rob thinks his data is 11:45Z and Peter thinks he processed 12:00Z.
- EUW uses an offset between particle size for water clouds and particle size for ice crystals. It seems that this offset was not removed in the comparison of particle size products of EUW. Fig 4.88 shows that the areas of EUW histograms are too small, which indicates that a large portion of the particle sizes is outside plotted range.

3. How to present to a wider audience?

The results of the intercomparison were considered very interesting. Due to this intercomparison it is much clearer what differences can be expected between the cloud property retrievals from the different groups. It would be worthwhile to work towards a publication of the results in the future. This would help growth of confidence among the users. However, before doing so some modifications of the intercomparison analysis are needed (as described above in this document). A good list with concrete activities needs to be prepared to keep Andi Walther on the project.

4. Further Work Required.

Based on the analysis of the results of Andi we came up with some suggestions to make the comparison better:

- Concentrate on an easier case. For example, include a more favourable region such as stratocumulus off of African Coast. Good candidate is south-east of sun-glint near Namibia. Keep all solar zenith angles less than 60.
- Include a multi-layer scene as another scene.
- Include a southern ocean frontal system
- Include a tropical convective case.
- Conduct phase comparison

Further we proposed some work to be done by the data providers:

- Check calibration and angles
- Generate high-level summary of algorithmic approaches and assumptions.
- It would be nice to develop an electronic system for sharing data and comments. For example, use a WIKI based web-site.

5. Who will commit to do the Work?

The group agreed on the importance and needed continuation of this work. It would be best if the contract of Andi could be extended either through an other VS activity (co-sponsored by one of the SAFs) or by a fellowship of EUMETSAT, so that the work done for this CWS can be completed. Further, regular repetitions of a similar type of intercompaison activities (e.g. within two/three years) would be useful.

6. Future intercomparisons.

Finally some ideas came up for future intercomparisons:

- Could involve A-train observations where A-train is collocated with SEVIRI
- Rob's archive of SEVIRI observations collocated with CloudNet data can be provided to test performance over a long time period over a small area. Data is available upon request from CloudNet.
- Could do an AVHRR only comparison to include NOAA results. (Adam, Rob and Andy)
- We should include IWP and LWP comparisons as well. This removes potential differences due to differing definitions of effective particle size for ice crystals. Moreover, CLWP is one of the few cloud properties that can be validated with accurate cloud observations from ground based observations.
- Include more groups involved in cloud property retrievals to this intercomparions activity (e.g. Pat Minnis's SEVIRI work with the French).
- See if we can contribute to the Cloudnet data base and standardize our comparisons.

7. Artifacts that look anomalous

- EUW looks too low in term of optical depth especially in CUT 1.
- DLR or OMA bimodal effects in scatter plots warrants looking at.
- EUW added 100 micron to ice particle sizes. Was this properly accounted for?
- EUW's histograms of particle size all seem to end at approximately 20 microns.
- MOF products are shifted one pixel
- KMR products differ 15 minutes (11:45) from the products of the other groups (12:00)