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 #3
Received and published: 12 July 2011

General comments

In this study the authors explore long-term (i.e. multi-year) measurements of CF and TCF at various permanent ARM sites in the world, and discuss their potential use in the evaluation of the cloud-radiative model climate of the GCMs participating in IPCC AR4. Given the significant uncertainty in numerical predictions of future climate that is related to cloud representation in models, this topic is very relevant for the climate science community. In addition, any attempt to involve the wealth of available ARM data that could be used to this purpose should also be encouraged. In this study the authors make use of simultaneous measurements of cloud presence by multiple independent instruments. They go to great lengths in inter-comparing these different datasets, and provide a detailed analysis of their differences and the possible causes behind them. The introduction contains a good and quite thorough review of the topic of the evaluation of model clouds against observations, and the written text on the instrumentation and the measurements is quite clear and accessible also for non-experts in observational meteorology. All of this is commendable.

In my view two results of this study stand out, as they are the most relevant for the climate science community and in particular for model evaluation purposes. However, as presented in the current manuscript, these results either i) are still somewhat complicated or ii) their implications are not elaborated on as fully as possible. I think addressing both shortcomings is necessary, as they would significantly increase the value and relevance of this paper.

1) An important result is that the considerable differences that exist between the various independent measurements of CF and TC on a daily basis seem to reduce significantly in the monthly and annual means. The authors in principle do well in explaining all known causes for these differences. I am not convinced though that long time-averaging completely solves all comparability problems, as the authors seem to suggest when they then use the long-term means for model evaluation. Just establishing that the long-term means converge and get pretty close is not sufficient; while long time-averaging might perhaps solve the ergodic side of the problem (i.e. time-averaging gets equal to spatial averaging), compensating errors might still exist between say the effective detection threshold of the instruments and the cloud side effect in wide FOV instruments, just to name two examples. I don't see why the last two effects could not affect the monthly or annual means. For example, the cloud side effect in wide FOV instruments will be significant in deeper boundary-layer cumulus cloud fields; averaging over a month long of exclusively cumulus days will still show this impact. The same goes for threshold differences. So what is lacking in the current manuscript are some arguments for why the effects other than the ergodic one will not affect the longer-term means; can the authors provide those? My guess is that the application of instrument simulators are the only way to eliminate all comparability problems. Am I right in assuming that this was not an option because the IPCC AR4 runs did not

require simulator output? And what do the authors think about the assumptions that still go into simulator models?

An additional reason (besides the ergodic hypothesis) that monthly and annual means show smaller differences is that many of the differences in individual observations are likely cloud-type dependent. For a short time period (e.g., a day), a single cloud-type might dominate the observations, but as you move to longer time averages you also get contributions from multiple cloud types, and these compensating errors help reduce the overall bias of each instrument. However, we did not intend to suggest that long-term averaging completely solves all comparability problems between the model and observations, and there could still be biases in the observations. We have added the following comment to the discussion in Section 3:

"*Although the three estimates of CF converge to < 15% for multi-year monthly means, this does not necessarily mean that the estimates of measured CF are unbiased, only that averaging over a longer time period and a multiplicity of cloud types tends to offset detection differences between the different instruments.*"

Yes, the application of instrument simulators was not an option because the IPCC AR4 runs did not require simulator output. We thank the referee for bringing up this interesting point. Instrument simulators may help reduce comparability problems, especially with regards to threshold detection issues, but they will not eliminate the problems completely because there are still multiple assumptions made in the simulators – the sub column generator, the microphysical and optical assumptions, etc. Even had the IPCC AR4 runs required simulator output, the simulator would have been designed for satellite instruments (e.g., the Cloud Feedback Model Intercomparison Poject Observation Simulator Package (COSP, Bodas-Salcedo et al., 2011)), not ground-based instruments and thus the output would not have been useful for this study. While the COSP radar simulator would be fairly easy to be adapted for ground-based observations (primarily requiring consideration of attenuation between the ground and cloud rather than between TOA and cloud), a simulator for the TSI and TSK measurements would be more challenging. We have added a couple of sentences regarding instrument simulators to the introduction and discussion, respectively.

"*Because of the difficulty in relating model variables to quantities retrieved from remote sensing observations, instrument simulators which use model output to directly simulate the signal that an instrument would observe have been developed in recent years. For climate models, these simulators have focused primarily on satellite observations to date. Development of techniques to simulate ground-based remote sensing observations of the type used in this study would be useful to alleviate some of the uncertainties in the model/observation comparisons.*"

"Inclusion of a ground-based radar and lidar simulator in the model will allow more direct assessment of cloud overlap, vertical structure, fall velocity, cloud phase, and cloud microphysical assumptions against the ARM radar observations. New radar observations and techniques such as radar spectra measurements provide vertical velocity statistics which can be used to examine assumptions in convective parameterizations (Kollias and Albrecht 2010), better identification of regions with multi-modal characteristics such as mixed phase regions (Shupe et

al., 2004), and better discrimination between cloud and drizzle (Kollias et al. 2011) which will be useful for investigation of autoconversion rates."

A. Bodas-Salcedo, M. J. Webb, S. Bony, H. Chepfer, J.-L. Dufresne, S. A. Klein, Y. Zhang, R. Marchand, J. M. Haynes, R. Pincus, V. O. John, COSP: satellite simulation software for model assessment, Bulletin of the American Meteorological Society, DOI:10.1175/2011BAMS2856.1,2011.

2) The authors manage to establish from the confrontation of long-term means of model cloud fraction and radiative fluxes with their observed equivalents that compensating errors are present in the interaction of clouds and radiative transfer as represented in most of the GCMs. Unfortunately, from these results we do not (and can not) learn exactly where in the system of interacting parameterizations this compensation takes place. This is also acknowledged by the authors in the final section (p.14963 lines 18-19). While I realize that this may be beyond the scope of this study, what is important though is to think about how this insight might actually be obtained, and to also discuss this. In my view, this exercise is key to solving this important problem in climate science. Signalling the problem of compensating errors is of course great, but in this case is actually not new; previous studies have shown this, and by now most modellers are already aware that such a compensation takes place in the cloud-radiation interaction. The actual reason for the existence of compensating errors is that each individual parameterization (e.g. cloud overlap, inhomogeneity, vertical structure, etc.) has not yet been sufficiently constrained by observations. Could the authors come up with a list of all these processes, and perhaps make proposals on how to cover each with observations? I think the addition of a discussion on this topic, probably in the summary section 6, would benefit this paper. It would certainly be constructive, as it would suggest a way forward and could also act as an outreach to the modelling community.

This is an excellent point. It is well known the representation of cloud is the most uncertain element in climate models and it is an extremely challenging task that may take several decades of effort just to improve the cloud scheme in a single GCM. This study aims to evaluate a group of GCMs rather than one specific GCM with a particular cloud scheme in simulating cloud and cloud-radiation scheme. For this reason, we are not able to go deeply to investigate the cause of biases of each model. On the other hand, only CF and surface radiation data are used in this study. Besides cloud overlap assumption, the cloud thickness and water content, height and shape of clouds, ice/liquid water fraction, etc., all of which could affect the surface radiation, are not investigated in this study. Without comprehensively analyzing all those fields, we are not able to attribute the bias of surface radiation to cloud overlap assumption scheme or other factors. All of those works can be done in future study.

This comment has been taken into account in the revised manuscript. We have added the following paragraphs to the summary in Section 6:

"*In future work, we plan to do such experiments for the physics parameterizations used in the Community Atmosphere Model (CAM5). Colleagues at PNNL have implemented the CAM5 physics package in the weather research and forecasting (WRF) model [Jerome Fast, personal communication, 2011], allowing examination of the range of behavior of the physics parameterizations over a range of scales, including those closer to the scale of the ARM* *observations. The WRF model can be run using all of the CAM5 physics or only individual components, which will allow exploration of the effects of individual parameterizations on the resulting cloud fraction and radiation relationships. The WRF model can also be run at very high spatial resolution, which can reduce the uncertainty in model evaluation induced by the different spatial coverage between model and measurement. Forcing the model with reanalysis data will also reduce the potential discrepancies in large-scale dynamics between the model and observations that can exist in free-running climate models, and this can be one cause of model/observation disagreement. We will also save hourly output from model to allow us to investigate the diurnal cycle of cloudiness in parameterization scheme.*

Inclusion of a ground-based radar and lidar simulator in the model will allow more direct assessment of cloud overlap, vertical structure, fall velocity, cloud phase, and cloud microphysical assumptions against the ARM radar observations. New radar observations and techniques such as radar spectra measurements provide vertical velocity statistics which can be used to examine assumptions in convective parameterizations (Kollias and Albrecht 2010), better identification of regions with multi-modal characteristics such as mixed phase regions (Shupe et al., 2004), and better discrimination between cloud and drizzle (Kollias et al. 2011) which will be useful for investigation of autoconversion rates. The satellite simulator has been installed in some of the IPCC AR5 GCMs; we may repeat the analysis for the IPCC AR5 GCMs and compare the results with those from this study."

Kollias, P., J. Remillard, E. Luke, W. Szyrmer, 2011, Cloud radar Doppler spectra in drizzling stratiform clouds: 1. Forward modeling and remote sensing applications. J. Geophys. Res., 116, D13201, 14 PP., doi:10.1029/2010JD015237.
Kollias, P., and B. Albrecht (2010), Vertical velocity statistics in fair weather cumuli at the ARM TWP Nauru climate research facility, J. Clim., 23, 6590–6604, doi:10.1175/2010JCLI3449.1
Shupe, M., P. Kollias, S. Matrosov, and T. Schneider, 2004, Deriving Mixed-Phase Cloud Properties from Doppler Radar Spectra. J. Atmos. Tech., 21, 660-670.

To summarize, given i) the relevance of the topic, ii) the good use of available observational data, and iii) the fair scientific quality of the study (both in writing and content) I think this manuscript should be acceptable for publication after some major revisions as mentioned above. Specific comments

1) p14934, line 13-14: The cloud overlap assumption should also be mentioned here. Is it known what the overlap assumptions were in all participating GCMs?

In Section 2.1 we added a few paragraphs (attached below) briefly describing how CF and TCF are calculate and overlap assumptions made in IPCC AR4 GCMs. It might not be realistic to provide detailed description for each cloud scheme including overlap assumption used in all GCMs, given the lengthiness of this paper has already exceeded a normal one. Instead, we provided a brief summary on the major cloud overlap assumptions used in IPCC AR4 GCMs. Here below is what we added.

"*CF is a critical variable in climate models for determining the radiative fluxes through the atmosphere and at the surface. Depending on the complexity of the model, CF may also be used* *in many other physics parameterizations in the model such as cloud microphysics, aerosol wet removal and convective transport. In this study, we focus on the role of CF in radiation, where the area-averaged CF is used. As discussed in Brooks et al. (2005), although CF produced by most cloud schemes is volume-averaged, most GCMs assume that the cloudy area of a grid box fills the entire grid box in the vertical, thus essentially assuming area-averaged CF is the same as the volume-averaged CF. In GCMs, CF can be parameterized using statistic, diagnostic or prognostic approaches. Due to space constraints, we just summarize the CF parameterization schemes for all GCMs used in this study in Table 1; for more details of each cloud scheme, including references, see http://www-pcmdi.llnl.gov/ipcc/model_documentation/ipcc_model_documentation.php.*

In GCMs, the vertical correlations between cloud layers have to be prescribed because cloud elements are often smaller than a typical GCM grid cell and there is no general theory for how different cloud systems should overlap (Collins, 2001). Assumptions about vertical overlap of clouds can affect the exchange of energy between the atmosphere and other components in the model, influencing not only radiative heating rates but also atmospheric temperature and hydrological processes (Collins, 2001). In the IPCC AR4 models, the most common overlap assumptions are maximum/random (Geleyn and Hollingsworth 1979). One type of maximum/random assumption has maximum cloud overlap in each of three regions representing the lower, middle, and upper troposphere, and random overlap between these regions (e.g., Chou et al. 1998). A second type of maximum/random overlap scheme has maximum overlap between clouds in adjacent levels and random overlap between groups of clouds separated by one or more clear layers (e.g., Zdunkowski et al. 1982). The latter form of maximum/random overlap was found to be more consistent with a statistical analysis of observed cloud distributions (Tian and Curry 1989). "

Chou, M.-D., M. J. Suarez, C.-H. Ho, M. M.-H. Yan, and K.-T. Lee, 1998: Parameterization of cloud overlapping and shortwave single-scattering properties for use in general circulation and cloud ensemble models. J. Climate, 11, 202–214.
Collins, W. D., 2001: Parameterization of generalized cloud overlap for radiative calculations in general circulation models. J. Atmos. Sci., 58, 3224–3242.
Geleyn, J.-F., and A. Hollingsworth, 1979: An economical analytical method for the computation of the interaction between scattering and line absorption of radiation. Beitr. Phys. Atmos., 52, 1–16.
Tian, L., and J. A. Curry, 1989: Cloud overlap statistics. J. Geophys. Res., 94, 9925–9935.
Zdunkowski, W. G., W.-G. Panhans, R. M. Welch, and G. J. Korb, 1982: A radiation scheme for circulation and climate models. Contrib. Atmos. Phys., 55, 215–238.

2) p14934, line 15: "very minimal". Use of the word "very" is not scientific, please rephrase.

Have removed "very".

3) p14938, lines 8-10: This would be a good point to discuss the potential use of instrument simulators.

We have added the following statement to this section: "*Because of the difficulty in relating model variables to quantities retrieved from remote sensing observations, instrument simulators which use model output to directly simulate the signal that an instrument would observe have been developed in recent years. For climate models, these simulators have focused primarily on satellite observations to date. Development of techniques to simulate ground-based remote sensing observations of the type used in this study would be useful to alleviate some of the uncertainties in the model/observation comparisons.*"

4) p14938, lines 19-21: "... examine ... at different time-scales". The evaluation at daily time-scale as announced here is in conflict with what is mentioned on the previous page (p14937, lines 19-21: "... we do not perform direct hour-by-hour or daily comparisons, ..."). This is confusing. Also, if your intention is to avoid daily comparisons from the start, why then still dedicate such a big part of the manuscript (Figs 1-2) to this?

P14937, lines 19-21: we do not perform direct hour-by-hour or daily comparisons between model and observation because free-running GCM is not possible to simulate exact weather system during a specific time period. P14938, lines 19-21, in Sect. 3 we compared three CF-related datasets among themselves (rather than against model result) at different time scales from hourly to monthly. The reason we currently include this part is to quantify the differences among the three observational datasets at hourly or daily scale and to provide evidences for caution in evaluating climate or weather forecasting models using the hourly or daily ARM CF datasets. Comparison results also provided confidence for using ARM multi-year averaged monthly data to evaluate CF in climate models.

5) p14939, section 2.1: Could you spend some sentences on what CF actually represents in the models? Is it the cloud fraction as used in the radiation scheme, or is it maybe the cloud fraction as part of a vertical transport model? These are not equivalent; the fraction as used in the radiation code should carry information on sub-grid scale cloud overlap, while the transport scheme cloud fraction does not (area-averaged versus volume-averaged; see e.g. Brooks et al., JAS, 2005). Accordingly, the vertical resolution also plays a role here. Would it not be better to evaluate the CF in models on the same vertical grid as the observations?

As described in the introduction, "Climate models typically interpret CF as the horizontal area fraction covered by clouds as viewed from nadir". CF information is mainly used in radiation scheme, but can also be used in many other places such as cloud microphysics, aerosol wet removal and convective transport depending on the complexity of a climate model. In this study, we mainly focus on the role of CF in radiation, where the area-averaged CF is used. As discussed in Brooks et al. (2005, JAS), although CF produced by most cloud schemes is volume-averaged, most GCM assumes that the cloudy area of a grid box fills the entire grid box in the vertical, thus essentially assuming area-averaged CF is the same as the volume-averaged CF. We have added more discussions on this issue and cited Brooks et al. (2005) in Section 2.1 in the revised manuscript. Please also see response to Specific Comment 1.

Brooks ME, Hogan RJ, Illingworth AJ. 2005. Parameterizing the difference in cloud fraction defined by area and by volume as observed by radar and lidar. J. Atmos. Sci. 62: 2248–2260.

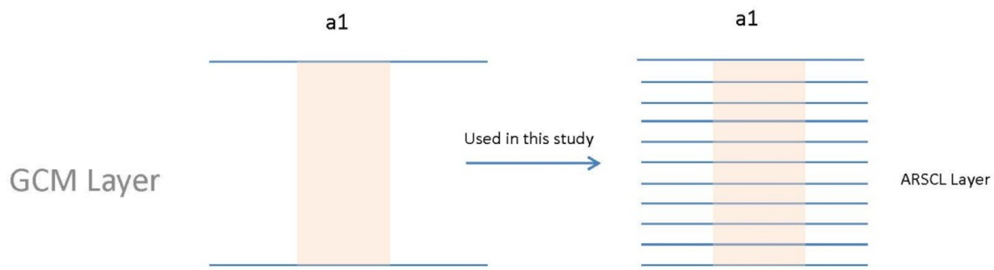
Yes, we evaluate the CF in models on the same vertical grid as the observations. We did vertical interpolation/mapping of the CF from each GCM into the ARSCL vertical grid in our original manuscript, so the comparisons presented in Figures 9-14 are based on finer ARSCL vertical grid. In the original manuscript, we discussed about uncertainty associated with the layer thickness of ARSCL data, but that was not meant for the CF that was not compared at the same vertical grid layer. We apologize for the confusion of description in the original manuscript.

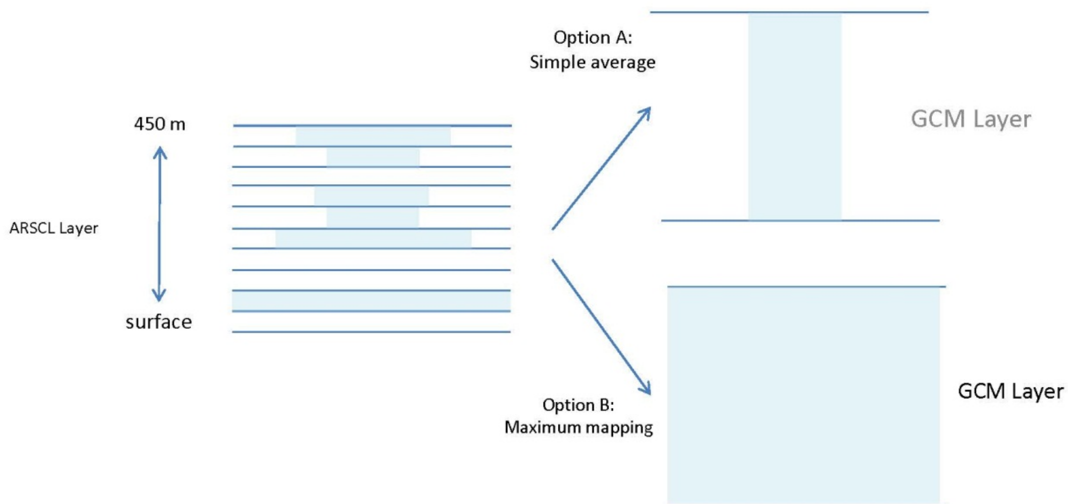
Since CF in GCMs is assumed vertically constant within each grid layer, in this study we evenly distribute the CF at each GCM layer (hundreds of meters thick) into the much finer ARSCL layer (45 m thick), which is illustrated in the top panel of the attached figure below. There are two reasons why we mapped the CF from model vertical grid to ARSCL grid. First, if we map the CF from ARSCL to model, we need to determine which GCM vertical grids should be used to represent all GCMs since the vertical resolution of each GCM is different. Otherwise, we would have to do a separate mapping for each of the GCMs. Moreover, mapping the CF from finer (ARSCL) to coarser (model) vertical grid may smooth out some meaningful vertical variability of CF.

Second, which is actually a more important one, is about how we vertically map the CF. The bottom panel of figure below demonstrates two, out of many, different vertical mapping approaches. Assuming that the depth of one GCM layer is 450 m, spanning over 10 45-m ARSCL uniform layers. CF may vary in the 10 ARSCL layers. To derive the CF in that GCM layer, one can take the option A to make a simple average, i.e. (c1+c2+…+c10)/10, or option B to take the maximum CF (100% in this case) in the 10 ARSCL layers, which physically make sense sometimes because a thick layer will be horizontally fully covered by cloud even if a thin layer embedded within that thick layer is horizontally fully covered by cloud. Because the CF is assumed to be vertically constant within each GCM grid layer, option B apparently overestimates the mean CF in GCM, especially for the purpose of radiation calculation. We believe that option A is more reasonable when mapping the observational ARSCL CF into GCM layer. The result of the option A (simple average) is similar to that using the approach currently applied in this study (i.e., evenly map the CF from coarse GCM to fine ARSCL vertical grid).

Therefore we believe that the comparison of model and observational CF currently presented is the “best” that we can do within the framework/scope of this study (i.e. we are not able to change the original vertical resolutions of either models or ARSCL observation).

We have added one new subsection (2.2.d) at the end of Section 2.2 rather than in the Conclusion and Discussion part to explain how the model CFs are vertically interpolated/mapped to the same vertical grids as observations in this study.





6) p14942, lines 22-23 "... averaged only for daytime hours between 8 and 17 LT": Would it make sense to preserve the diurnal time-dependence in the evaluation of the monthly and annual means that follows later? This might be informative, as it can indicate at which point of the day biases are biggest. In turn, this could help in attributing biases to specific parameterization schemes (boundary layer, deep convection, etc.).

This is a very constructive comment. Yes, diurnal cycle information of cloud could be very helpful in attributing biases to specific parameterization schemes. The reason only the data between 8 and 17 LT are averaged in figure 1 is that some instruments like TSI are not able to provide the CF observation during nighttime or when the solar height angle is very low. So we use an average only for daytime hours (8-17 LT) for all three datasets to make a fair comparison in this plot. On the model side, none of the GCMs provided the hourly CF so we are not able to compare the diurnal cycles of observed and simulated CF. In future study that we coordinate and focus on one GCM such as CAM5, we could preserve and save the hourly outputs of clouds to allow us to investigate the diurnal cycle of cloudiness. We have added this proposal in the paragraph about future work plan in the discussion section.

7) p14943, line 23 "... affected by cloud sides ..": See also my first general comment above. Although you mention and explain it here, after this point no attention is given to the potential impact of this effect on longer-term means.

See response to general comment # 1.

8) p14957, line 15-16: This is pretty much the only point in the text where the overlap assumption is mentioned as a possible contributor to model-obs differences. In broken and irregular cloud fields this could actually be significant. Also, I don't think the ARSCL product can detect the true overlap, as i) the lidar can not penetrate through optically thick clouds, so that it misses all clouds behind the first, and ii) the cloud radar might be blind for small droplets in warm clouds.

We agree that the observed overlap is not the true overlap because the lidar does not penetrate through thick clouds, and the radar does not detect thin layers or cloud top. We have rewritten the sentence at P. 14957, Line 14-16

"*Meanwhile, the TCF simulated in CCSM is close to that in ARSCL (Tables 2 and 3), which indicates that a different cloud overlap scheme from the true overlap seen in the ARSCL observations is probably used in CCSM.*"

into:

"*Meanwhile, the TCF simulated in CCSM is close to that in ARSCL (Tables 2 and 3), which indicates that the cloud overlap scheme in CCSM produces a result similar to the observation. However, we should keep in mind that the overlap derived from the ARSCL observations is not necessarily the true overlap due to limitations in the measurements (such as the difficulty of lidar to penetrate through thick clouds and the lack of detection of thin layers by radar).*"

9) p14964, first paragraph: The proposition to compare retrievals of cloud presence by ground-based and satellite-based instrumentation is interesting. Would the technique of cloud height - optical thickness histograms as originally used for the ISCCP dataset be worth mentioning here?

Although the ISCCP cloud height-optical thickness histograms are a very useful way of looking at cloud distributions, they may not be the best method for evaluating the vertical cloud fraction from the active sensors. ISCCP (and other passive sensors) are known to underestimate cloud top height for high clouds compared to active sensors and also don't see the full column if optically thick cloud exists. CloudSat and CALIPSO are likely better comparisons. For CloudSat/ARM radar data, it would be useful to compare cloud frequency by altitude (CFAD) diagrams (Yuter and Houze, 1995) which look at the distribution of reflectivity as a function of height. To compare MPL/CALIPSO vertical profiles would be more difficult because of the attenuation of each, and a simpler comparison of the PDF of highest cloud detection might be more useful (Thorsen et al., 2011; Dupont et al., 2010; Liu et al., 2010).

Yuter, S. E., and R. A. Houze Jr., 1995: Three-dimensional kinematic and microphysical evolution of Florida cumulonimbus. Part III: Vertical mass transport, mass divergence, and synthesis. Mon. Wea. Rev., 123, 1964–1983.

Thorsen, T., Q. Fu, and J. M. Comstock (2011): Comparison of the CALIPSO satellite and ground-based observations of cirrus clouds at the ARM TWP sites, J. Geophys. Res., In Press.

Dupont, J.-C., M. Haeffelin, Y. Morille, V. Noël, P. Keckhut, D. Winker, J. Comstock, P. Chervet, and A. Roblin (2010), Macrophysical and optical properties of midlatitude cirrus clouds from four ground-based lidars and collocated CALIOP observations, J. Geophys. Res., 115, D00H24, doi:10.1029/2009JD011943.

Liu, Z., R. Marchand, and T. Ackerman (2010), A comparison of observations in the tropical western Pacific from ground-based and satellite millimeter-wavelength cloud radars, J. Geophys. Res., 115, D24206, doi:10.1029/2009JD013575