Reviewer comments for "CLAAS-3: the third edition of the CM SAF cloud data record based on SEVIRI observations" by Nikos Benas, et al.

General impression: Major Revisions.

General Comments: I get a repeated impression that very old data was used for comparisons, especially for MODIS. Since this entire paper is about comparing data records, getting some serious clarification about the MODIS data sources, especially those used in images, is absolutely in order. Studies that were based on data as old as MODIS Collection 4, created before Aqua was even completed, let alone launched, are quoted in the comparisons. MODIS Collection 4 suffered from some severe deficiencies, especially for 3.7um retrieval. Things got a bit better for C5 and finally mostly remedied in C6.1. That'd be about all there is to say at this point.

Specific Comments:

Figure 1: maybe add some clarifying text as to the time periods the backups were used. I believe the thin data strips are the backup periods for the ones with the more continuous lines, but it might not be clear to the user. Also what are the implications of using backups? For example, I imagine Meteosat-8 and Meteosat-9 would show some deterioration in data after that many years.

Line 90-95: What about the thermal channels? Was the calibration slope determined for both main and backups?

Line 110: what do you do with the data that falls between 75 and 95 degrees SZA?

Line 143: so does this new cloud mask algorithm actually perform better for marine stratus off Africa at sunrise and sunset? Or the issue is still there and that is why you're cutting out your solar zenith that way?

Line 151: Your training dataset is based on CALIOP, which means *all* your training data is early afternoon. Yet you apply the algorithm to all observation times. How does that affect the accuracy of your neural network? What are your uncertainties outside of 1:30pm local time? You say that you calculate them, but what kind of value ranges are you getting?

Line 175: Your cloud microphysics retrievals methods are eerily similar to SEV06-CLD. I am wondering why that product is not mentioned in any way.

Line 194: replace "… on board the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellite" with "on board the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) spacecraft". Repetitive.

Line 220: could you just drop some ballpark uncertainty values that you encounter with CALIOP? For more discussion, of course the reference, but a few numbers here would do some good.

Line 260: If you would've looked at SEV06-CLD product that is the MODIS C6.1 code running on SEVIRI, you would have an exact uncertainty comparison due to instrument differences rather than waving of hands you are doing here.

Line 320: CER actually is a rather mixed bag, heavily influenced by above-cloud aerosol. You can't really make a statement that cloud drops get larger towards the cloud top. Bennartz and Rausch are careful to indicate that it's not always true. Moreover if there is any ice cloud in the scene at all, 3.7um will always be smaller than 1.6um. Anyways, it's more complicated. I don't know what impact if any that has on your conclusions, but still.

Line 335: Again, that is a very old study that was moreover influenced by outright bugs in the 3.7um retrieval algorithm at the time. The issues were not corrected until Data Collection 5 was released. Please don't quote that paper. What impact this has on your use of DARDAR data, I couldn't tell you, but I would not base a single thing on that paper, particularly because you're making statements about penetration depth and retrieval values. Retrieval values were garbage, differences in penetration depth completely aside.

Line 385: Can you show an image same as figure 2, that the bias actually decreases with additional screening criteria applied? That -30%. Oooof.

Figure 4: It seems from this image that CLAAS-3 is picking up persistent aerosol loads as clouds. It is very dusty in the areas with the highest positive biases. Can you explain?

Figure 11: Don't tell me you used Data Collection 5 MODIS data! For Collection 6.1 each retrieval pair has its own optical thickness reported. This kind of difference in count between tau and re was only present in C5 data and older. Please explain.

Figure 14: Same thing. What data collection did you use here? This doesn't look right at all, if you used C6.1. Number of successful retrievals for optical thickness and effective radius is identical and each band combination has its own tau, re and water path. They've been entirely decoupled in the latest MODIS data collection.