Review of Earth System Science Datasets manuscript by Zhuge et al. entitled

” Introduction to the NJIAS Himawari-8/9 cloud feature dataset for climate and typhoon research"

General impression and recommendations

This manuscript describes a new cloud climatology (NJIAS) based on data from the AHI imager onboard the Himawari 8+9 geostationary satellites. The covered region is the East Asia and North Pacific part of the Himawari disc and the temporal period covered is from 2016 to 2022. The methods for deriving various cloud parameters are described in detail. Validation results based on comparisons with data from the cloud lidar CALIOP on the CALIPSO satellite, with corresponding products from MODIS and with and operational AHI-based products from JAXA (Japan) are presented. The climatology is found to produce results compatible or, for some parameters, even better than the reference datasets from MODIS and Jaxa.

I find the study interesting because of the comprehensive use of existing methods (thus, making use of long-term international experience) but also the addition of new features which improve results even further. The manuscript is well-written, figures and tables are very appropriate and well designed, and the English language is actually surprisingly good for coming from non-native authors.

I do recommend acceptance of this manuscript but I also have a number of questions, mainly related to the methods used for validating the results. I have identified some problems with the methods and also some weaknesses in the validation methods and I think it is appropriate that authors at least discuss these weaknesses in an updated manuscript before the paper is published.

Detailed remarks and questions follow below and at the end some editorial remarks.

General remarks and questions:

Page 4, line 112: Is the temporal resolution of 6 hours from NCEP reanalyses really sufficient considering the observed rapid temperature changes of land surfaces in some regions and during some times of the day (e.g. during morning hours)? I fear that you will get some errors because of not describing diurnal temperature changes accurately.

Page 6, line 146: You could consider to add a comment on how the threshold values have been chosen. Are they empirically derived or pre-calculated/pre-scribed from theory?

Page 8, line 161: Why the word "seeds"? Can't see the logic. It is also a bit confusing: If neighbours have similar reflectance characteristics and are labelled cloudy, why isn't the pixel itself also labelled cloudy in the first place? What is the additional information here? Or, are you looking for the case when the pixel is almost similar to its neighbours? Please explain!

Page 8, line 187, sunglint modelling: How accurate are those model simulations? Do they take into account wind/roughness/wave heights at sea which are very important for sunglint strength and occurrence? Thus, are you confident that the model simulations work sufficiently well?
Also, I guess you cannot rule out a sunglint-like enhancement of reflection also from low water clouds (known for having strong anisotropic reflection behaviour).

Page 8, line 187, ratio description: I fear that this is misleading. I mean, it seems wrong to relate the brightness temperature difference between a shortwave-infrared channel and an infrared channel to the reflectance in a shortwave channel. Shouldn't you first estimate the reflectance in the 3.9 micron channel and then calculate the ratio to the reflectance in the 0.64 micron channel? That would make more sense.

Page 11, lines 237-238: Difficult to understand the logics here. Stratus or fog clouds are normally trapped in a temperature inversion as a contrast to e.g., Cumulus clouds. Thus, the temperature difference to the surface may be small or even reversed (i.e., cloud top warmer than surface), especially in winter conditions over land surfaces. I fear that a lot of these clouds will be missed by these tests. Please explain!
(Maybe I misunderstand what you mean by "greater". Do you mean "warmer than" i.e., less negative in the difference?)

Page 15, line 302: It is a bit unclear which cloud product has really been used in the validation process. Is it one of the earlier mentioned cloud products at the 0.04 degree or 0.02 degree resolution of the cloud information. or is it products (i.e., level-2 products) with the nominal AHI pixel resolution (about 2 km)?

Page 15, lines 304-305: Isn't it a rather serious restriction to only select the cases when the CALIOP product is unchanged within an AHI pixel? This means that you throw away a lot of cases and only retain clouds (or clear sections) which have the scale of at least 2 or 4 km size (depending on what product you evaluate). Wouldn't it be fairer to use all cases (more representative for true conditions)? Now you don't have a clue how your schemes behave when there is sub-pixel cloudiness. For example, will these cases be considered cloudy or clear by your method and in correct proportions as given by CALIOP sub-pixel cloud information?

Page 15, line 306: 15 minutes is a large observation time difference. A cloud can move quite some distance in windy conditions and fast-growing cumulus clouds will not be well described. With an AHI scanning frequency of 10 minutes you should be able to at least match CALIOP within 5 minutes, shouldn't you? This would still give a large number of samples. Why was this not done instead?

Pages 15-16, lines 309-310: Why not CALIOP cloud tops? They should be more reliable (and more independent/objective) than any other estimation from passive imagery.

Page 17, line 339: I am a bit skeptical to these results since you have basically thrown away all cases with sub-pixel cloudiness and also severely underestimated the amount of thin cirrus in the CALIOP (1 km) dataset. For example, any of the investigated schemes could have misinterpreted the sub-pixel cloudy cases as either completely cloud free or completely cloudy. For example, this could have biased the estimation of all the scores in any direction. Can you give some estimation of what this restriction means in relation to the case when using all existing clouds in the CALIOP dataset? Notice that there are studies which have been able to make use of the entire CALIOP dataset (i.e., 5 km products complemented with single-shot statistics at CALIOP FOV resolution of 300 m). Two good examples are the following:

Karlsson, K.-G. and Håkansson, N.: Characterization of AVHRR global cloud detection sensitivity based on CALIPSO-CALIOP cloud optical thickness information: demonstration of results based on the CM SAF CLARA-A2 climate data record, Atmos. Meas. Tech., 11, 633–649, https://doi.org/10.5194/amt-11-633-2018, 2018.

Karlsson, K.-G.; Devasthale, A.; Eliasson, S. Global Cloudiness and Cloud Top Information from AVHRR in the 42-Year CLARA-A3 Climate Data Record Covering the Period 1979–2020. Remote Sens. 2023, 15, 3044. https://doi.org/10.3390/rs15123044

Page 18, line 348: Strange reasoning. Validation should always be made against reference data which are closest to the truth. To compare to another dataset based on passive imagery means that you compare with something that is very likely to suffer from the same weaknesses as your own product. Thus, results are likely to be "too good to be true". At least you should mention this and possibly state what errors can be expected. For example, MODIS product have also been evaluated against CALIOP so there is some knowledge of what differences you can expect. Notice also in the previous reference Karlsson et al., 2023 that MODIS-derived cloud tops show quite some problems when compared to CALIOP (and AVHRR) estimates.

Page 19, line 363: From where comes this improvement? I mean, can you point out exactly what differences in the algorithms are responsible for this improvement? It is not clear to me since your description of improvements of cloud top properties deals mainly with the cloud phase determination. Is it simply so that the JAXA algorithm has specific problems in relation to MODIS and NJIAS algorithms?

Page 23, line 430, cloud phase results: Again, I wonder how results would change if a larger fraction of true very thin cirrus clouds would have been present (i.e., by using the 5 km CALIOP datasets). Now only the thick and moderately thin cirrus clouds are included in the study. Can you comment this? For example, what happens in case of very thin cirrus overlying thick low-level water clouds? What phase is reported?

Page 25, line 463: Yes, further studies are needed to better understand the differences but these results do not need to be reported here.

Pages 27 and 29, Figs 12 and 13: Very nice figures showing the differences between the methods!

Page 30, line 558: In the summary and conclusions section, you present POD and FAR results as if they were general (i.e. for all cases). But you should state that you only studied the part of CALIOP data that showed clouds on the scale of AHI. Thus, not including effects of sub-pixel cloudiness or very thin cirrus. I think this must be commented here (and perhaps also in the abstract).

Editorial remarks:

Page 1, Abstract, line 19: Unnecessary information since Dr. Zhuge and colleagues are already listed as authors, and thus responsible for the work. Remove.

Page 1, Abstract, line 20: Add CALIPSO in brackets.

Page 1, Abstract, line 20: Add MODIS in brackets.

Page 1, Abstract, line 24: Remove "Then," and start sentence with “Two applications examples...”

Page 2, line 45: Add “CALIPSO” in brackets.

Page 2, line 45: Add “CPR” in brackets.

Page 2, line 47: Remove “often”.

Page 2, line 53: Maybe a better formulation is "have resulted in the generation of"?

Page 3, lines 64-68: Add several acronyms in brackets here!

Page 3, line 74: Add acronym CAPCOM!

Page 3, line 85: I suggest to replace "clocks" with "hours".

Page 3, line 87: Change “objects” to “objectives”.

Page 5, table 1, CldHeight, unit: Do you mean meters above ground level (AGL)? If so, maybe it should be explained somewhere.

Page 6, line 131: Change “have” to “has”**.**

Page 11, line 211: Change “dusk” to “dust”.

Page 11, line 212: Change “dusk” to “dust”.

Page 17, caption Table 5: Explain in caption the meaning of the bold numbers (even if it can be assumed that it is the "winners").

Page 23, line 432: Change “Colletction” to “Collection”.

Page 23, line 440: Write “The correlation coefficient (CC) of ….”

Page 28, line 526: Change “Infa” to “In-Fa”.

Page 30, line 536: I think it would not hurt (i.e., would be valuable for the reader) if you also mention which channels are used in the two RGBs.

Page 30, line 539: Clarify in the caption or in the text that the content of panels g-h in Figure 13 are weather radar rain rates from the Zhejiang province.