Response to the Reviews

April 9, 2024

Arndt Kaps
Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR)
Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
Email: arndt.kaps@dlr.de

Earth System Science Data
Manuscript Number: essd-2023-424

Dear Editor: thank you for the opportunity to improve our manuscript. We have addressed the reviewer's comments below.

Yours sincerely,
Arndt Kaps on behalf of the author team

Response to Referees

Earth System Science Data

Manuscript:	essd-2023-424
Title:	Characterizing clouds with the CCClim dataset, a machine learning cloud class climatology
Authors:	Arndt Kaps, Axel Lauer, Rèmi Kazeroni, Martin Stengel, Veronika Eyring
Date:	April 9, 2024

We thank the two referees for their helpful comments. In this document, we answer each point raised by the referees. The original referees' comments are given in **blue**, our answers in **black**. All line numbers refer to the "track changes" version of the revised manuscript.

Referee: 3

Review comments for "characterizing clouds with the CCClim dataset, a machine learning cloud class climatology" by Kaps et al. This manuscript describes a cloud class climatology dataset that merges passive and active observations together in a fixed coarsened grid that is suitable for model comparisons. This dataset employs the CloudSat+CALIPSO 8 WMO cloud class as the "truth", and Aqua-MODIS Level 2 physical cloud property retrievals as the inputs for the training in consideration of obtaining large amount of collocated training samples and to make the model interpretable (hence not using MODIS Level 1 measurements). Then the labels predicted from native resolution MODIS data are then coarsened to GCM grid size and training coarse-grained MODIS data to generate another ML model that is finally applied to the ESACCI-AVHRR dataset to generate a 34-year long cloud class dataset. The authors then gave several examples illustrating how to use this dataset for GCM comparison to help identifying issues associated with certain cloud types. Overall, I think there's some scientific merits of this current research. I particularly like the example regarding comparison with ICON-A outputs, and appreciate the "MODISequivalent filter" is applied to make the effort for a true apple-to-apple comparison. However, there are some fundamental issues with the design of the ML architecture that cause inevitable flaws to make this dataset really useful. These issues need to be addressed before publication of this work.

Major issues: (1) The training "truth" CC-L coming from 2B-CLDCLASS only has CloudSat and some MODIS information. If you read the ATBD on CloudSat website, you'll find CALIPSO data is in their plan to use jointly, but was not implemented so far. The 2B-CLDCLASS-LIDAR product has cloud mask from joint CloudSat-CALIPSO observations, however, that dataset doesn't have cloud type classification. Because only CloudSat and limited MODIS information was fed into the "truth", it inherently underestimates cirrus clouds and then propagates this bias into your product.

According to the CloudSat Product Process Description and Interface Control Document (PDICD [1]), the 2B-CLDCLASS-LIDAR dataset used here provides cloud type information based on both, CPR and CALIOP data. We agree with the reviewer that the radar-only dataset 2B-CLDCLASS does not contain this information, but this dataset is not used here.

(2) Aqua-MODIS always collocates with CloudSat at "nadir" view. That means the off-nadir correction must be made in order to not "overpredicting" cloud masks because of the slantwise integration length making the off-nadir view easier to detect cloud. This correction factor was never discussed in the current manuscript. The overprediction of the overall cloudiness in your CCClim dataset might likely attributes largely to this factor.

When creating CCClim, we followed the approach typically taken in existing literature regarding nadir/off-nadir effects, which is to prefer coverage over precision. But the reviewer has a good point.

In order to estimate possible effects of the viewing angle on the cloud classes in CCClim, we created a second version of the dataset by discarding pixels that are presumably affected by large viewing-angle-related uncertainties. For this, we removed the 300 outermost pixels on each side when training the Random Forest (i.e. in each across-track line 600 of 1354 pixels are removed). As expected, this leads to an increase in the

[1] `https://www.cloudsat.cira.colostate.edu/cloudsat-static/info/dl/2b-cldclass-lidar/2B-CLDCLASS-LIDAR_PDICD.P1_R05.rev0_.pdf`

fraction of “undetermined" pixels per training grid cell, causing a reduction in performance for the other eight classes as visible in their mean R^2-Score of $R^2_{cropped} = 0.79$, compared to the original score of $R^2_{orig} = 0.84$, despite applying weights in their favor (see Kaps et al., 2023). We then performed a comparison equivalent to Table VI in Kaps et al. (2023), which assesses how well CCClim captures the average global distribution of cloud types as in CC-L. For every cloud type, the correlation with CC-L is much lower than for the original CCClim. This suggests that the inclusion of all available training data, even though they are affected by viewing angle effects to some degree, is beneficial overall.

A correction of the viewing angle effects as proposed by the reviewer depends on cloud type and cannot be easily done since scattering angle and solar zenith angle are not trivially related for each specific pixel (Maddux et al., 2010; Painemal et al., 2021). There are no correction methods proposed in literature for the cloud property retrievals. Examining a possible correction method would require re-deriving the level-2 MODIS products (Bennartz and Rausch, 2017) and is therefore beyond the scope of this paper.

In the previous version of the manuscript we mention that CCClim is produced under the assumption that off-nadir retrievals are of similar quality than the nadir-near retrievals. The point raised by the reviewer made it clear that this statement is oversimplified. We therefore added the following paragraphs to the revised manuscript:

- (lines 134-140) While viewing geometry is part of the retrievals for the cloud optical properties (Platnick et al., 2017), across-track angular dependencies have been identified as a source of uncertainties in the MODIS Collection 5 data (Horvath et al., 2014; Maddux et al., 2010). Similar effects are expected for the Collection 6 data used here, but their magnitude is unclear and no published correction methods are available. These effects are, however, expected to be less relevant for long-term averages (Maddux et al., 2010). Following the data usage in many other studies (e.g. Bennartz and Rausch, 2017; Cho et al., 2021; Oreopoulos et al., 2016) we have therefore opted to make the above assumption instead of introducing new uncertainties by trying to correct these non-trivial effects.
- (lines 151-155) Sensitivity tests with the outermost 300 pixels on each side discarded during training of the RF to reduce the viewing angle effects (not shown) show a decreased performance for all of the eight cloud classes as measured by their mean R^2-Score and a smaller correlation with the CC-L ground-truth. We therefore used the complete breadth of the swath to generate CCClim despite the quality degradation towards the edges of the swath.

We furthermore now mention explicitly that viewing angle effects introduce additional uncertainties in CCClim by adding lines 351-356:

- Also, since CC-L labels are only available for nadir-near MODIS pixels, uncertainties might be introduced by labeling off-nadir pixels in the classification stage. A correction, however, is non-trivial (e.g. Painemal et al., 2021) and no concrete solutions are available in literature. Simply discarding pixels that are "far" away from the center reduces the data coverage and results in a significant reduction in performance. Even though the viewing angle effect is very hard to quantify, it is expected to be at least somewhat mitigated through the spatial and temporal aggregation in CCClim (e.g. Bennartz and Rausch, 2017; Maddux et al., 2010).

(3) For ancillary data from ERA5 reanalysis, I don’t understand why temperature and water vapor profiles are not included. Aren’t they the closest atmospheric variables to determine whether to form a cloud or not?
The ERA5 data used in this work are only an example of how CCClim can be complemented by additional data for further analysis. No ERA5 data are provided as ancillary variables with CCClim. The example is meant to show that the CCClim cloud types hold relevant information comparable to that of other techniques when categorizing clouds by dynamical regime using ω_{500} as a proxy as done by Bony et al. (2004). This classification of the dynamical regime does not rely on other parameters such as water vapor or temperature profiles. Additionally, these variables are not particularly informative for the cloud *type*, but more for a cloud mask, as indicated by the reviewer.
(4) It is not explained why you can apply two-step ML models trained by MODIS data directly to AVHRR data. Admittedly you use similar L2 cloud retrieval products for training and prediction, but AVHRR has so

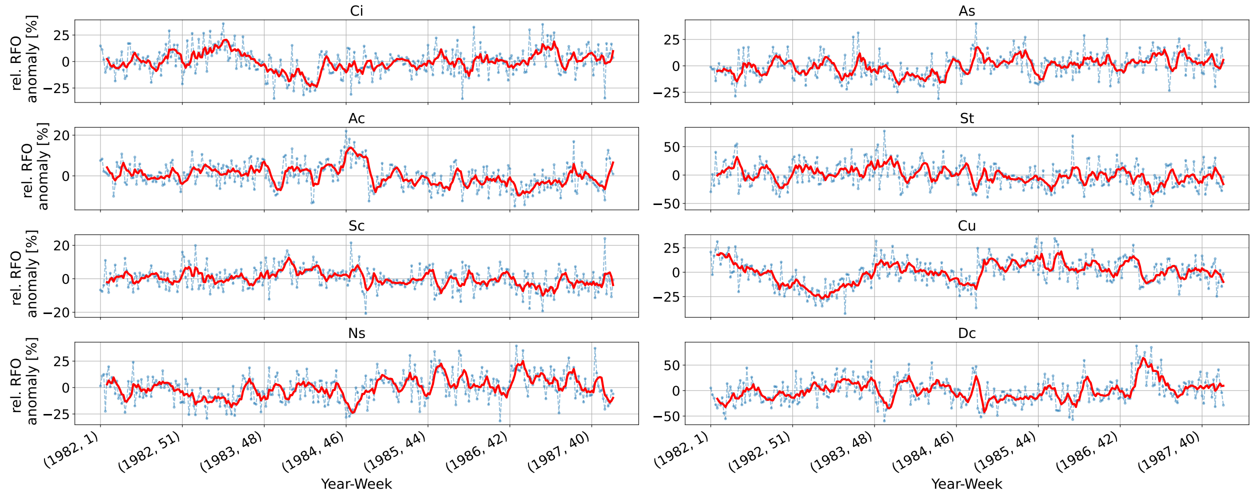

Figure 1: Relative deviations from average annual cycle (blue) averaged over the region $(30,60)^{\circ}N,(60,0)^{\circ}W$ with rolling average (red) showing the interannual variability.

few bands (literally only one visible band), so the products are not quite comparable. Even the Cloud_CCI project had published a paper illustrating their discrepancies (https://doi.org/10.5194/essd-9-881-2017).

The paper mentioned by the reviewer illustrates that the two datasets deviate in particular in their overall/liquid cloud fraction, which are not included in the values used by us. We showed in (Kaps et al., 2023) that predictions obtained from AVHRR data with our method are in good statistical agreement with the CC-L ground truth. This suggests that the differences between MODIS and AVHRR do not seem to play a major role. Being able to apply the ML-model to independent other datasets is a prerequisite for application to climate model data for model evaluation, an important aim of this ML framework. This is mentioned in lines 149-151.

(5) By training on multi-year collocated MODIS-CloudSat data, I don't quite understand why the interannual variability is not learnt by the ML model, resulting in no interannual variation (e.g., ENSO) in your timeseries shown in Fig. 4. Although it is clarified later on the manuscript that this dataset is not suitable for trend study, it is never claimed that it is not suitable for interannual variability study either. Ultimately, if a 34-yr long dataset is not intended for studying inter-annual variability, why product that? Why not just stop at Step-2 model and produce a MODIS cloud class that suits every application presented in this manuscript.

Interannual variability is learned by the ML model as we can see in our results. Interannual variability is, however, very hard to see in Fig. 4 as we show simply an average over all ocean grid cells in the Southern Hemisphere to illustrate the daily mean RFO and its standard deviation for different cloud types. For interannual variability studies the average seasonal cycle is typically subtracted from the time series and often smaller regions are investigated. As an example to show that interannual variability can indeed be studied with CCClim, we present a time series of the anomalies in RFO (i.e. the average seasonal cycle is subtracted) averaged over the north Atlantic ($(30,60)^{\circ}N,(60,0)^{\circ}W$), for the 3 years from Jan. 1982 to Dec. 1987. Figure 1 shows that the interannual variability of cloud types can be seen in CCClim with periods of positive and negative anomalies (visible for example in the anomaly for Ci).

(6) As also notified in this manuscript, MODIS (all passive sensors) have issues distinguishing clouds against snow-cover surface in polar regions. However, the statistics (e.g., Fig. 2, Fig. 5) were summarized globally. I'd strongly suggest you exclude polar areas in computing your statistics.

This is a sensible suggestion and for the example analyses we show, the samples are cleared of possibly ice-covered ocean, as is stated in the manuscript. However, this manuscript is intended to introduce the CCClim dataset and not meant as a scientific analysis. We therefore believe it is more helpful to show global statistics rather than masked values. We added lines 366-369 to inform users of this dataset of this potential error source.

Minor issues: Fig. 3 - recommend adding a map from CCL for straightforward visual comparison.
Such a figure is already shown in (Kaps et al., 2023) (their figure 8) and we would prefer to not repeat this figure here.
For example, I don't see the Gill model distribution in the Western Pacific cirrus clouds (might be obscured by the annual cycle or your coarse colorbar). Same recommendation for Fig. 5 (i.e., if using CCL annual cycle, do you still see the same bias suggested by your CCClim product?)
We think in order to study atmosphere-ocean interactions such as the ones mentioned by the reviewer, one would have to look at anomalies rather than the absolute values of the seasonal cycle. Such an analysis could possibly be done with CCClim, but is in our opinion beyond the scope of this paper that aims at introducing CCClim and providing a few examples of what could be done with the CCClim dataset. As the CC-L data are very sparse, we do not expect that a robust annual cycle can be calculated that would be needed for a meaningful comparison.
Fig. 10 - It's not understandable for cirrus clouds, how can ICON model produces a bunch of thin cirrus (i.e., IWP low) with large size ice particles? Please double check your graphing codes.
The parameterization of the effective radius of ice particles $r_{e,i}$ in the ICON-A version used here depends only on the in-cloud ice water content q_i as there is no information on the ice crystal number concentration: $r_{e,i} = 83.8 \cdot q_i^{0.216}$. This means that the vertically integrated cloud ice water content shown in Fig. 10 (gridbox-averages, output of ICON-A) can be small if the cloud fraction is small and/or cloud ice is only present in one or two model layers. Yet, $r_{e,i}$ at cloud top shown in Figure 10 can have relatively large values that are typical for cirrus clouds.

Referee: Peter Kuma

We would also like to thank Peter Kuma for helping us to improve the manuscript.

L88: 'While these classes are defined to be consistent with the WMO definitions, misclassifications can occur, caused for example by the small footprint of the active sensors.': I think this is still overstating the consistency between the CC-L classes and WMO genera. The CC-L classes are defined based on a set of relatively synthetic (rule-based or fuzzy logic) thresholds. They are not expected to be matching statistically when compared to ground-based observations of clouds which are defined relatively vaguely based on a set of features determined visually by a person. Therefore, misclassifications are expected purely because the definition and the viewpoint are not the same.

The fuzzy logic algorithm is designed in a way that the physical features of these cloud genera are within the range of what is shown by observations of these cloud types over many decades (Wang, 2019). We agree that cloud types from CC-L and ground-based observations are not exactly the same because of the very different approach of both methods. Here, it helps that active sensors are sometimes able to determine also the cloud base height, which is expected to help with providing at least some consistency with the surface-based classifications. Following the point of the reviewer, we rephrased the corresponding statement to make clear that even though the CC-L cloud classes are not matching the WMO-classes, they have been designed in a way that the retrieved physical properties are within the range of typical values for these cloud classes.

Also on L327: 'Using multiple cloud properties to define the classes makes them physically consistent and well-aligned with the morphological WMO genera.' And on L114: 'comparable': I am not sure if this is a good term to use. It would suggest that one could for example compare the CC-L RFOs with WMO cloud genera RFOs from a ground station and expect them to have comparable statistics. But that is not the case because the definition is different. It might be better to say 'analogous' or 'similar'.

Following the suggestions of the reviewer, we rephrased the sentence using 'similar'.

One reason why training on daytime samples and applying the algorithm on nighttime might lead to biases is because nighttime passive retrievals are based on the infrared spectrum bands only, and lack any information provided by the visible spectrum bands. Therefore, there could be limitations on the minimal detectable cloud optical thickness and so on, and the results could have statistical biases compared to the daytime retrievals. The Cloud_cci nighttime products are considered of experimental quality, and they use different thresholds for day, night and twilight, according to Stengel et al., 2020: 'Night-time COT and CER retrievals are considered to be experimental products and only included in Level-3U products.', 'Please note that retrievals of CER, COT, CWP and CLA are also provided during night-time, although as experimental products.'.

We agree with the reviewer that the nighttime products are experimental and thus subject to larger uncertainty than their daytime counterparts. We emphasized this by explicitly mentioning in the manuscript (lines 379-380) that the nighttime retrievals are experimental.

L106: 'high temporal resolution': It is not clear what high resolution means. It is better to be more specific.

Changed to “daily resolution”.

Fig. 4: 'dailymean RFO': As mentioned previously, I think that this is misleading because it is not representative of a daily temporal mean, but rather is a daily composite with an incomplete diurnal coverage.

Agreed, changed to “daily RFO values”.

Fig. 7: It might be good (but not necessary) to mention in the caption that the dashed line is SW CRE = LW CRE. 'Latitude Range: $0°N/S - 90°N/S$' is a bit unclear to me, but I guess it simply means global (?)

In order to avoid confusion, we removed the title and added the dashed line to the legend.

References

Bennartz, R. and J. Rausch (2017). "Global and regional estimates of warm cloud droplet number concentration based on 13 years of AQUA-MODIS observations". In: *Atmospheric Chemistry and Physics* 17.16, pp. 9815–9836. ISSN: 1680-7324. DOI: 10.5194/acp-17-9815-2017.

Bony, S., J.-L. Dufresne, H. L. Treut, J.-J. Morcrette, and C. Senior (2004). "On dynamic and thermodynamic components of cloud changes". In: *Climate Dynamics* 22.2-3, pp. 71–86. DOI: 10.1007/s00382-003-0369-6.

Cho, N., J. Tan, and L. Oreopoulos (2021). "Classifying planetary cloudiness with an updated set of MODIS Cloud Regimes". In: *Journal of Applied Meteorology and Climatology*. DOI: 10.1175/jamc-d-20-0247.1.

Horvath, Ã., C. Seethala, and H. Deneke (2014). "View angle dependence of MODIS liquid water path retrievals in warm oceanic clouds". In: *Journal of Geophysical Research: Atmospheres* 119.13, pp. 8304–8328. ISSN: 2169-8996. DOI: 10.1002/2013jd021355.

Kaps, A., A. Lauer, G. Camps-Valls, P. Gentine, L. Gomez-Chova, and V. Eyring (2023). "Machine-Learned Cloud Classes From Satellite Data for Process-Oriented Climate Model Evaluation". In: *IEEE Transactions on Geoscience and Remote Sensing* 61, pp. 1–15. DOI: 10.1109/TGRS.2023.3237008.

Maddux, B. C., S. A. Ackerman, and S. Platnick (2010). "Viewing Geometry Dependencies in MODIS Cloud Products". In: *Journal of Atmospheric and Oceanic Technology* 27.9, pp. 1519–1528. ISSN: 0739-0572. DOI: 10.1175/2010jtecha1432.1.

Oreopoulos, L., N. Cho, D. Lee, and S. Kato (2016). "Radiative effects of global MODIS cloud regimes". In: *Journal of Geophysical Research: Atmospheres* 121.5, pp. 2299–2317. DOI: 10.1002/2015jd024502.

Painemal, D. et al. (2021). "Evaluation of satellite retrievals of liquid clouds from the GOES-13 imager and MODIS over the midlatitude North Atlantic during the NAAMES campaign". In: *Atmospheric Measurement Techniques* 14.10, pp. 6633–6646. ISSN: 1867-8548. DOI: 10.5194/amt-14-6633-2021.

Platnick, S. et al. (2017). "The MODIS Cloud Optical and Microphysical Products: Collection 6 Updates and Examples From Terra and Aqua". In: *IEEE Transactions on Geoscience and Remote Sensing* 55.1, pp. 502–525. DOI: 10.1109/tgrs.2016.2610522.

Wang, Z. (2019). *CloudSat 2B-CLDCLASS-LIDAR Product ProcessDescription and Interface Control Document*. Version p1_R05. URL: https://www.cloudsat.cira.colostate.edu/data-products/2b-cldclass-lidar.