Review of the manuscript 'Characterizing clouds with the CCClim dataset, a machine learning cloud class climatology'

Peter Kuma

Department of Meteorology (MISU), Stockholm University, Stockholm SE-106 91, Sweden

7 December 2023

Dear editor and authors,

The authors present a dataset of cloud types derived from satellite data using supervised machine learning previously described by Kaps et al. (2023a). This complements existing cloud type datasets such as ISCCP and those derived using unsupervised machine learning. I find the study well-presented and the discussion and conclusions meaningful. However, there are some factual errors, some details require further clarification, all limitations should be properly discussed, and potentially more validation would be beneficial. My comments are detailed below.

Kind regards,

Dr. Peter Kuma

General comments

Calling the CC-L cloud types WMO cloud types is misleading. The WMO cloud genera are defined using ground-based visual observations, as standardised by the WMO International Cloud Atlas. While CC-L definitions of cloud types aim to be similar to the WMO cloud genera, they are not the same. The deep convective (Dc) type is also not a WMO cloud genus. I think that the authors should be more clear about this.

There is a considerable overlap between the presented work and Kaps et al. (2023a). It is not entirely clear if the authors use the exact same methods as in the previous paper, or if anything is different. I think that the authors do not necessarily have to duplicate the description of the methods in this manuscript, and can instead reference the previous paper, which is really a pre-requisite for understanding this manuscript.

L177–178, Fig. 4 caption: The Cloud_cci AVHRR L3U PM daily data are a composite of instantaneous values and not a daily mean, because the satellite overpass time is always at about the same local time, and the data have not been corrected with a diurnal cycle model. Stengel et al. (2020): 'All data are collected on two processing levels: (a) Level-3U, which represents daily composites of non-averaged data collected on a global latitude–longitude grid with 0.05° resolution and'. Unless some more sophisticated processing is done on the input data, the CCClim dataset also represents daily composites of instantaneous values.

It is not clear whether the results only represent daytime clouds or nighttime as well. The authors say that the CUMULO dataset is only available for daytime, but they do not say whether they apply the RF on nighttime grid cells from Cloud_cci and ICON-A. If they only apply it on daytime, it should be made more clear that the results can be biased toward daytime clouds and not representative of polar night conditions. If they apply it on nighttime too, this should be mentioned as a limitation due to the fact that this is assuming that the statistical links between the input physical quantities and cloud types applying during daytime also apply during nighttime (reference training data for nighttime are lacking in this case).

Because the comparison between cloud types derived from ICON-A and Cloud_cci is between 3-hourly instantaneous output of the model and daily composites of AVHRR measurements, it is not necessarily comparing the same time of day at local time. Because clouds tend to exhibit a diurnal cycle, this can introduce biases. This should be at least mentioned as a limitation. Also comparing 35 years of CCClim from AVHRR with 2 years from ICON-A is not necessarily a good idea because the global climate has changed during the last 35 years. Comparing the same 2 years of both would be better.

One potential limitation that the authors do not mention is that cloud-related quantities such as cloud top pressure and cloud optical depth are not always directly comparable with those derived from radiance measured by a satellite instrument. For this reason, simulators for ISCCP and MODIS such as COSP exist. Because the RF takes these quantities as a input from the observations and the model without using an instrument simulator, an artificial bias can be introduced in the comparison.

The authors should include short sections in the Methods describing the Cumulo dataset and ICON-A. The section on CC-L should at least briefly describe how the CloudSat–CALIPSO cloud types are defined in the original dataset.

A global regular angular (longitude—latitude) grid has unequal area of grid cells by latitude. This means that grid-cell averages are calculated over larger areas over polar latitudes than equatorial latitudes. This can have an effect on a predictor, such as the RF used on the coarse-grained Cloud_cci data or ICON-A data. I think that this should be at least discussed.

I think it would be possible to use a single machine learning model to determine the cloud types instead of two (IResNet on the fine scale followed by RF on the coarse scale). One could train a CNN to predict coarse grid cell cloud type fractions by training on grid cells where CC-L data are available (averaged over the coarse grid cell and represented as fractions of cloud types). It seems to me that the two step approach can be both more complex and less stable. Can the authors comment on this?

'CloudSat labels' in the text and figures should be really called CC-L labels or CloudSat–CALIPSO labels, since they are based on measurements from both satellites.

I suggest that the authors use terms 'cloud type' and 'cloud class' consistently throughout the manuscript. They seem to be used interchangeably, but it would be better to use either only one of them, or define the distinction between them.

The active satellites sensors used as a reference for the CC-L cloud types have known problems with detecting low clouds. The CloudSat radar is affected by ground clutter, and the CALIPSO lidar signal is often attenuated by overlapping clouds before reaching mid and low-level clouds. These problems in turn also affect the presented dataset, and should be mentioned as a limitation.

The authors train the RF on MODIS but apply it to Cloud_cci. It is not clear if this produces any side-effects. Ideally, the authors should compare the results of applying the RF on MODIS and Cloud_cci for the same testing time period (excluded from training).

Fig. 4: Long-term AVHRR datasets are usually not very well suited for determining trends due to discontinuities in the orbital parameters of the satellite series. This limitation can also translate to any derived dataset. Can the authors comment on whether the slopes shown in the figure are reliable? The authors mention this on L339, but it would be better to also caution the readers in the context of Fig. 4, and also on L8–10.

L14–15: **or its radiative effects**: This could be a problem with reanalyses which usually parametrise clouds, which can result in radiation fluxes not corresponding to the actual cloud cover.

L224: **ice-free oceans**: It would be better to say 'ice-free (SST > 275 K) oceans', since the authors are not actually using sea ice information to exclude grid cells, but a temperature condition which also excludes some ice-free ocean grid cells.

L275: **This waveband corresponds closely to the absorbing band used for the MODIS retrievals of cod and cer.**: But what about AVHRR? Since the comparison here is between cloud types derived from ICON-A and from AVHRR.

L290–291: **With respect to CCClim the simulation also exhibits an increase in high cloud RFO (Ci, Ns)**: Ns is classified as a middle level cloud in the WMO International Cloud Atlas.

L351–352: **there is arguably no better way to obtain cloud observations from space than combining radar, lidar and passive instruments.**: I think that this is merely stating the obvious. Since there are currently no other satellite instruments relevant to cloud measurements other than lidar, radar and passive instruments, combining them necessarily has to give the best opportunity to obtain cloud measurements. One could also consider other kinds of passive measurements such as those by geosynchronous satellites, multi-angle measurements (MISR), passive microwave, Doppler lidar (Aeolus) and passive instruments on deep space satellites (DSCOVR). Including those in a combined analysis could produce more accurate results. The fuzzy logic classifier used in CC-L could have many different alternatives, for example using unsupervised learning or by supervised learning using ground-based observations. I do not think that one could say that it is the best, but rather one of many options of how to classify clouds.

Technical comments

Greater attention is needed to make the plots accessible to people with colour blindness. For example, in Fig. 1, 2, 6 and 7 some of the cloud type colours could not be distinguished. I recommend testing this for example with KMag on Linux.

L44: Missing space in 'CCClim(Kaps et al.,'.

L105: I think that the official name of the dataset is 2B-CLDCLASS-LIDAR.

L133: Parenthesis not closed.

L141: **grid boxes**: Probably better to say 'grid cells' for consistency with the rest of the manuscript.

L190: **Fig.6**: Missing space.

L240, 268, 344: **Tropics**: Usually not capitalised.

L277: **sensors(Eliasson et al.)**: Missing space and missing year.

L294: **as most cells contained >90% Ci have an ice water path** $iwp < 10\frac{g}{m^2}$: as most cells *which/that* contained...

L355: This section should contain links to all of the datasets used, including Cumulo, ERA5, MODIS, 2B-CLDCLASS-LIDAR, Cloud_cci and ICON-A.

L372: It should be made clear that the DOI is a DOI.

Table 1: **cloud type**: Should be capitalised to be consistent with Table 2.

Table 1, Fig. 6: Capitalisation of the first word in the caption.

Fig. 2: This is more of a style issue, but pie charts are usually not considered a good visualisation method because the proportions are hard to compare. Bar charts usually provide a much better experience.

Fig. 3, 8, 9: Ideally, all colour bars should have labels.

Fig. 4: The plots should have y-axis labels.

Fig. 5: Plots should have a y-axis label RFO.

Fig. 5: **see 4**: 'see Section 4'.

Fig. 10: The plots are probably too small. I recommend dedicating more space to the figure. The axes should have better labels, i.e. not just the cloud type name, but also that it is RFO on the x-axis, and probability density on the y-axis.