Response to RC1 on egusphere-2023-1953

The Chalmers Cloud Ice Climatology: Retrieval implementation and validation
Preprint https://doi.org/10.5194/egusphere-2023-1953

Adrià Amell   Simon Pfreundschuh*   Patrick Eriksson

April 24, 2024

Text from the Anonymous Referee is presented in gray and ours in black.

We would like to thank the reviewer for their thorough review of our work and the constructive comments. We are convinced that the comments helped us to improve our manuscript. We hope that we could successful address the reviewer's concerns.

Changes to the manuscript

While the manuscript was in review we discovered a mistake in the radar retrievals from the Palaiseau cloud radar, which used a tropical instead of a mid-latitude PSD parametrization. For the revised manuscript, we have updated the results of the ground-based TIWP retrievals. This did not change the results considerably.

In addition to the issues highlighted by the reviewers, we have also corrected a number of smaller mistakes in the figures included in the manuscript.

Major comments

1. While the paper is well-written and includes multiple statistical examples demonstrating the efficacy of the machine learning technique, it lacks maps and curtain plots illustrating the geographic representation of CCIC retrievals. To convincingly demonstrate the representativeness of these estimates, such examples are essential. The authors could add for example:

    - Monthly global IWP maps showing the CCIC against the cloudsat estimates (no need to subsample the CCIC, just show that the global distribution is as expected)
    - Percentage difference in these types of maps.
    - A latitude – altitude cross section of cloud ice fraction, again comparing versus the cloudsat one.
    - global maps of IWC at different levels showcasing the variation with height

*simon.pfreundschuh@colostate.edu

- Cross sections of IWC through different longitudes

We agree with the reviewer that the initial manuscript provided insufficient evidence of CCIC's capability to capture the spatial distribution of TIWP and TIWC.

To address this, we will extend our analysis of the CCIC retrieval results on the test dataset, which comprises one full year of collocated geostationary observations and corresponding CloudSat measurements. Due to the sparse sampling of the CloudSat observations, we have decided against assessing the spatial distributions of the retrieval results and errors by month as the results would be extremely noisy. We will extend the manuscript with the figures shown in Fig. RC1.1 and Fig. RC1.2, which display the distribution of retrieved and reference TIWP and the zonal means and biases of the TIWC retrieval, respectively.

figs/global_dist_tiwp.pdf

Figure RC1.1: Spatial distribution of retrieved TIWP and 2C-ICE-based reference TIWP for the CPCIR and GridSat test datasets over the domain covered by CCIC. Panels (a) and (b) show the retrieved TIWP aggregated to a resolution of 5°. Panels (c) and (d) show the corresponding distributions of the reference TIWP measurements. Panels (e) and (f) show the biases relative.

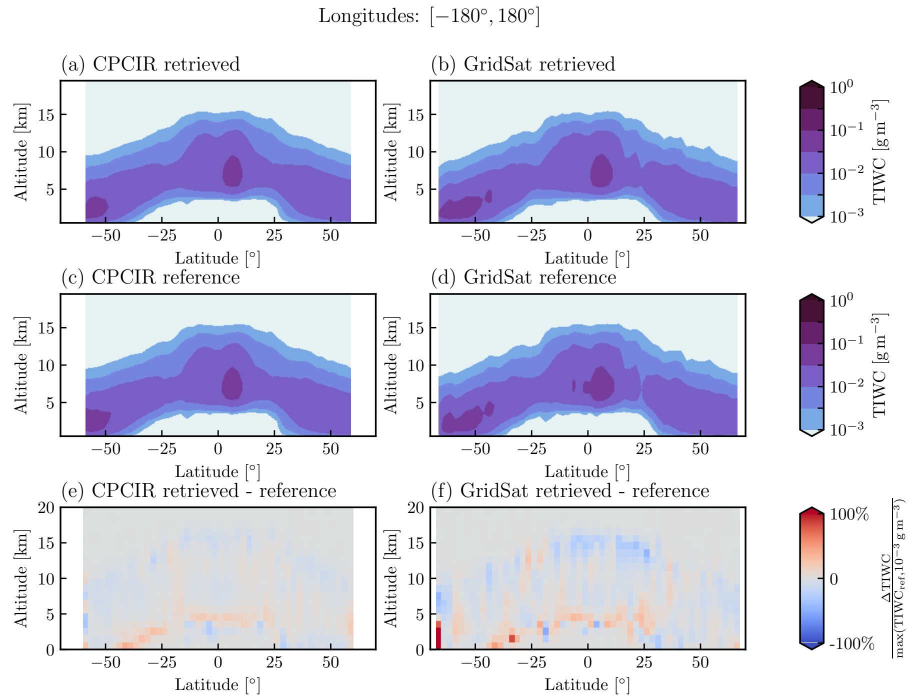

Figure RC1.2: Zonally averaged distributions of retrieved TIWC and 2C-ICE-based reference TIWC for the CPCIR and GridSat test datasets over the domain covered by CCIC. Panels (a) and (b) show the retrieved and reference TIWC for the CPCIR observations aggregated to a resolution of 2.5°. Panel (c) show the truncated relative bias of the retrievals. Panels (d), (e), and (f) show the corresponding distributions of the GridSat-based retrieval.

We also provide maps of retrieved and reference TIWC at discrete altitude levels and zonal averages of retrieved and reference TIWC for different longitude bands in Fig. RC1.3 to Fig. RC1.7. While these results provide further evidence of the ability of CCIC to capture the three-dimensional distribution of TIWC in the atmosphere, we do not plan to include them in the manuscript as it already contains a large number of figures.

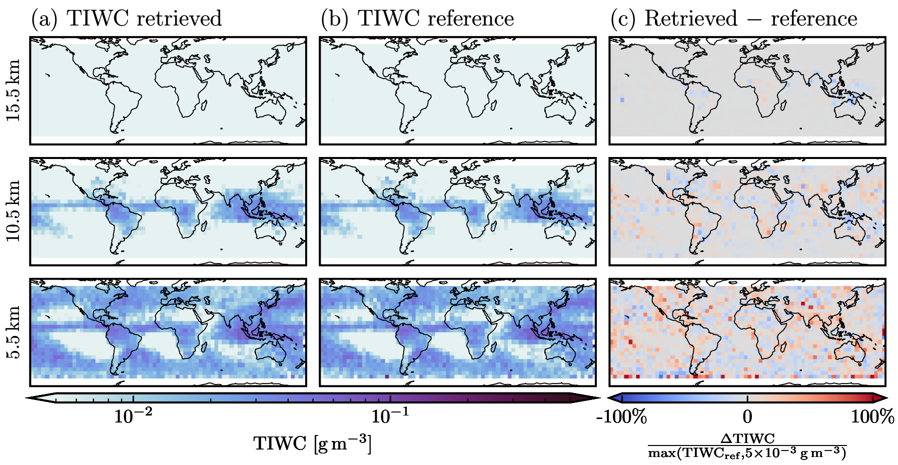

Figure RC1.3: Distribution of retrieved and 2C-ICE-based reference TIWC measurements for CCIC CPCIR retrievals at three atmospheric levels. Column 1 displays the mean retrieved TIWC. Column 2 displays the corresponding reference TIWC. Column 3 displays the truncated relatived error. Rows contain the results for the atmospheric levels at altitudes of 5.5, 10.5 and 15.5 km, respectively.

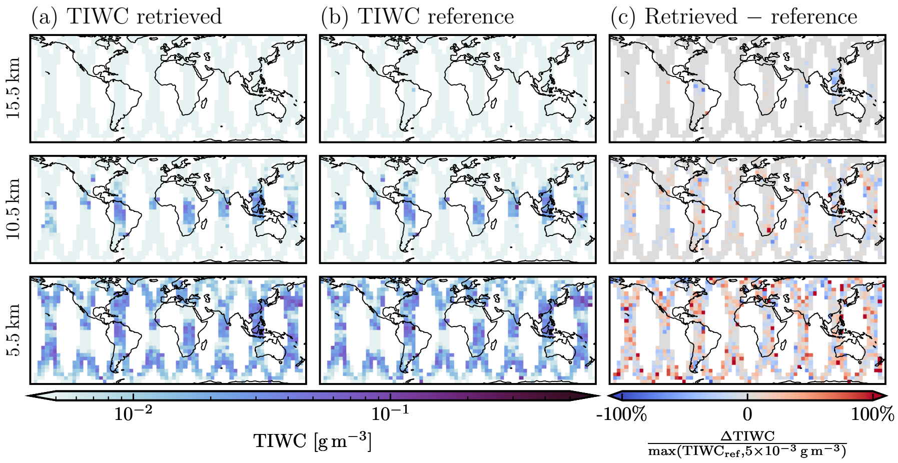

Figure RC1.4: As Fig. RC1.3 but for GridSat-based CCIC retrievals.

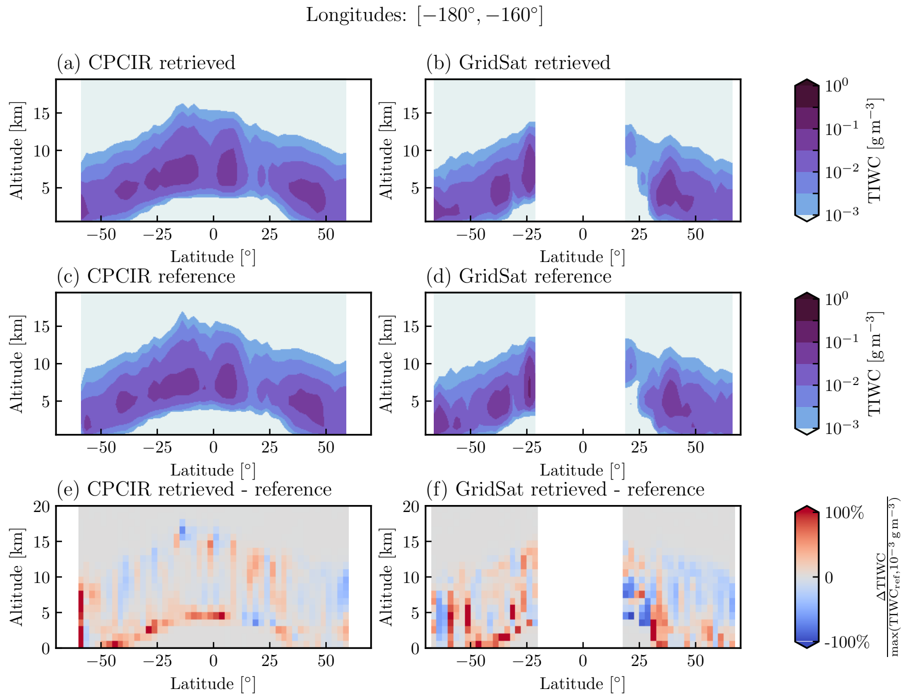

Figure RC1.5: As Fig. RC1.2 but for longitudes within $-180°$ and $-160°$.

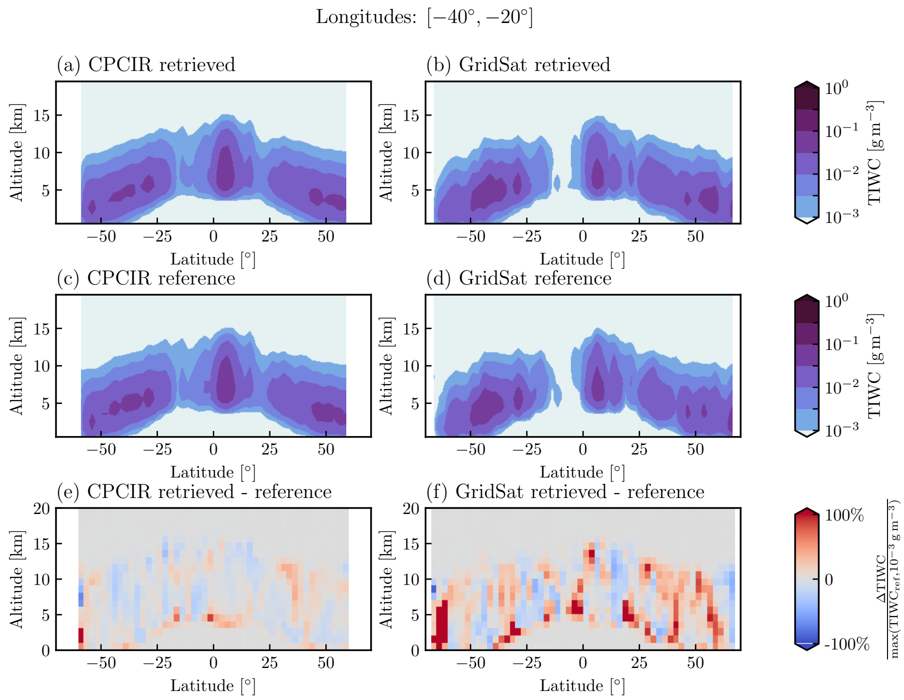

Figure RC1.6: As Fig. RC1.2 but for longitudes within $-40°$ and $-20°$.

Longitudes: $[20°, 40°]$

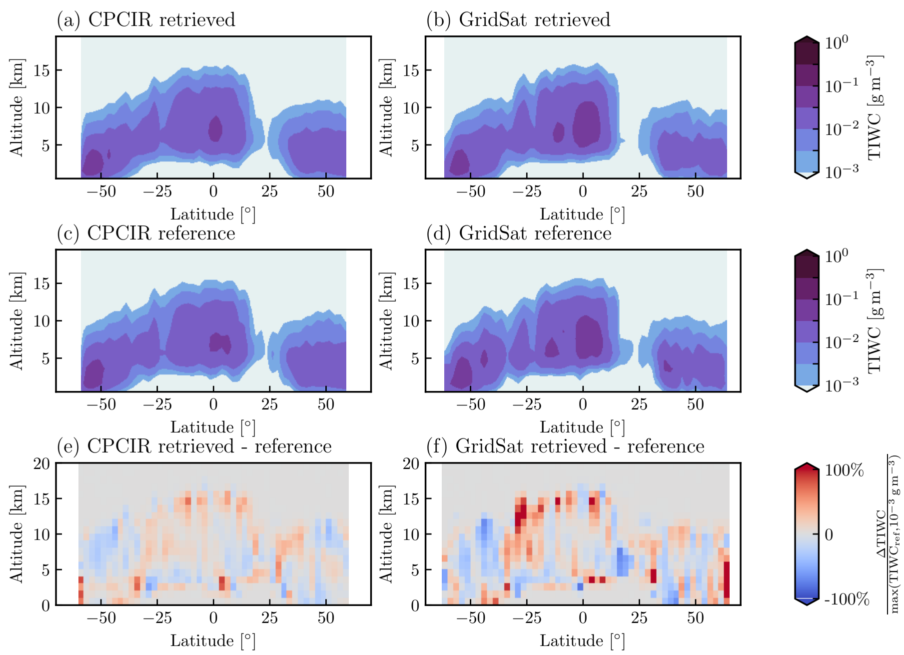

Figure RC1.7: As Fig. RC1.2 but for longitudes within 20° and 40°.

Specific comments

1. The title is not representative of the context of the manuscript. There is no mention of constructing a climatology or anything of that sort.

   We thank the reviewer for pointing out that the introduction failed to properly present the CCIC project and the scope of the manuscript. We have reformulated the abstract to state that CCIC is a novel cloud-property datasets. Moreover, we have reformulated the two paragraphs starting in l. 65 of the revised manuscript to clearly state that the presented manuscript describes the first step towards the production of an ice water path climate record using the presented retrieval.

2. Line 41: "For the study of processes on annual and decadal scales it is therefore necessary to find ways to make better use of observations with a long record of availability". The authors should mention that several IWP records exist with annual and even decadal scales, such as the ones from MODIS, Aura MLS, Odin SMR, CloudSat, etc. As currently written, the introduction implies that such records do not exist.

   We thank the reviewer for pointing out this shortcoming of our manuscript. We will extend the introduction with a list of currently available records of comparable TIWP estimates and their respective shortcomings.

3. Further since CCIC provides IWC, the authors could compare partial IWP versus those records matching their respective altitude coverage. The comparison versus the campaigns is limited to a few periods and it is limited geographically.

   As per the reviewer's suggestion, we will extended the analysis of the CCIC retrieval results to provide a more detailed assessment of the three-dimensional distribution or retrieved TIWC, as detailed in the response to major comment 1. The analysis is based on a full year of collocations with 2C-ICE estimates and thus covers the full geographical extent of the CCIC retrievals.

   While the comment seems to suggest to also extend our comparison to estimates from limb-sounding instruments such as Aura MLS or Odin SMR, we choose not to do this as we do not think that this would offer any benefit over the comparison against the CloudSat/CALIPSO-based estimates.

   Finally, we would like to point out that, while the individual field campaigns used in the validation are naturally limited in their spatial and temporal coverage, they cover both tropical and mid-latitude climate regimes and extend from instantaneous to seasonal time scales. With this, they exceed the scope of most validation efforts of TIWP/TIWC products that were able to find in published literature (Deng et al., 2010, 2013; Eriksson et al., 2008; Wu et al., 2008; Barker et al., 2008).

4. Line 17: "considerable skill" is a qualitative description please provide a more quantitative description.

   We will reformulate the sentence in question to provide quantitative accuracy estimate derived from independent test data.

5. Line 20: “first order” is a qualitative description please provide a more quantitative description.

   We will remove the formulation 'first order' and instead list the linear correlation coefficient and bias of the TIWC estimates in the previous paragraph.

6. Line 45: please describe the rationale behind only using the 11micron channel. Presumably additional channels could provide more information.

   We chose the 11 µm channel because it provides the best temporal and spatial coverage throughout the available record of geostationary satellite observations. While the GridSat B1 product also includes visible and water vapor imagery, which could likely help to improve the retrieval, they are not always available and therefore not considered in the current CCIC retrievals. We will revise the paragraph in question to include the motivation for this design choice.

7. Line 55: “Estimates of TIWP differ widely between”. Please give the ranges, this would allow you to later show how well (or bad) CCIC estimates are.

   We will rewrite the relevant parts of the introduction and include the ranges of disagreement between satellite-based IWP estimates.

8. Line 87: why not use lat lon info as well? And day of the year?

   We do not include spatial and/or seasonal context in the retrieval input data as we want the underlying neural network to learn relations between the satellite observations and the corresponding cloud properties and not their variability with respect to geographical location and season. Since the training data period is limited from mid-2006 through 2009, including geographical coordinates and seasonal information could limit the retrieval's ability to reproduct changes in the regional or seasonal variability of clouds outside of the training data period.

9. Line 108: The use of “2D” here is confusing since the authors are talking about profiles, I suggest deleting it.

   We will replace ‘2D’ with ‘horizontal’.

10. Line 113 – Line 119: A schematic of this entire procedure will be appreciated. Also, what is the treatment for the uncertainties in 2C-ICE

    We acknowledge the importance of making our training-data-generation process transparent and reproducible. To of clouds this end, we have published all relevant code in the repository accompanying this manuscript. However, since the manuscript already contains a large number of figures, we chose not to include a schematic of the data extraction process as we consider it of minor interest for the general audience.

    We will add a sentence referring the interested reader to the relevant code to the revised manuscript.

    Regarding the uncertainties of the 2C-ICE product: We treat the 2C-ICE estimates as ground-truth and do not make any effort to model the associated uncertainties. Synetergistic radar-lidar retrievals of ice hydrometeor concentrations have to be regarded as the most most accurate global measurements of TIWC and TIWP. Although even these combined radar-lidar estimates remain affected by significant systematic uncertainties, largely due to the underlying microphysical assumptions on particle shape and distribution, these uncertainties are not well characterized due to the limited amount of work that compares them with in-situ measurements (Deng et al. (2010, 2013) are the only studies that we are aware of). Since what ultimately matters to future users of CCIC is the uncertainty in the estimates provided by CCIC, we chose to extensively validate the resulting CCIC estimates instead of trying to handle uncertainties in the 2C-ICE product upfront.

11. Line 136: "Training scenes of 384×384 pixels". Is the geographical size of this scene important? Is there an impact for using smaller or bigger scenes? Why this particular size.

    We chose this size since it allows use to extract randomly rotated crops of size 256×256 pixels without generating invalid values. The scene size of 256×256, which is ultimately used in the retrieval, was chosen because it corresponds to scenes covering more than 900 km in zonal and meridional extent. The resulting scenes should thus contain information on the mesoscale and, to limited extent, synoptic-scale context of the retrieval.

    To make this point clear, we have reformulated the paragraph in question to provide a better description of the training-dataset generation and the underlying motivation.

12. Line 136: "The process involved randomly selecting a pixel with valid reference data as the starting point and then adding a random zonal offset." I don't really understand this please clarify

    We have reformulated the paragraph in question to describe the training-data generation process more clearly.

13. Line 140: lower → coarser

    We will adopt this suggestion in the revised version of the manuscript.

14. Figure 2 caption: Why is the brightness temperature normalized?

    Normalizing the inputs to neural networks is common practice and generally leads to better results and faster training convergence.

15. Line 155: Good is the enemy of great. The authors should explore tuning of these parameters or rephrase this sentence to state that minimal tuning was required.

    It is generally acknowledged that exhaustive tuning of all architecture-related hyperparameters of a neural network model is too resource consuming to be practically feasible and principled architecture search remains and activate area of machine-learning research (Ren et al., 2021). We therefore consider the reviewer's request out of scope for our work.

16. Figure 3, and 4 captions: specify the locations. for example, Aug 2015, and Jul-Aug 2018 flights were over the US and the US nearest oceans. Or,

Flights took place over the Olympic Peninsula in the Pacific Northwest of the United States.

We will add the requested flight locations to the captions of Fig. 3 and 4.

17. Line 229: Which of the four periods is the Darwin campaign?

    We will extend the sentence in question to state that the Darwin campaign is the first set of flights of the HAIC-HIWC campaign.

18. Line 237: Specify frequency

    Following the reviewer's suggestion, we have replaced 'W-band' by '94-GHz'.

19. Line 247: Which year?

    The year was 2019. We will include this information in the revised manuscript.

20. Figure 5: panels g and h should say Cloud classification and cloud classification (retrieved) respectively.

    We will update the panel titles in the revised manuscript.

21. Figure 6 caption should mention cloudsat somewhere, as well as the period use for this comparison.

    We will add the requested information to the figure caption.

22. Section 3.2.1. It is not clear which period this comparison cover.

    To make it clear to the reader that all results in this sub-section are derived using the independent test data, we add an introductory sentence to Section 3.2. Furthermore, we now state that the distributions were computed from the test dataset.

23. Line 297: This should be shown as a separate subsection to emphasize its importacnce: Zero order comparison of IWP and IWC (for example)

    Following the reviewers suggestion, we have added a new section titled 'Consistency of retrieved TIWP and TIWC profiles' and moved the discussion of the consistency of the retrieved TIWP and TIWC there.

24. I think the whole classification is barely working and the authors should just not show any of those results.

    We respectfully disagree with the reviewer's comment but acknowledge that results presented in the first version of the manuscript may give an overly negative view of the retrieval's capabilities. We would like to point out, however, that the classification results from the case study presented in Fig. 5 demonstrate the ability of the retrieval to distinguish the types of the principal cloud systems in the scene. To provide a more comprehensive picture of the retrieval's classification skill, we will add curtain plots (conceptually similar to the ones the reviewer requested to demonstrate the skill of TIWC retrieval) showing the spatial distributions of the retrieved and reference cloud classes with respect to latitude and altitude. The results, shown in Fig. RC1.8 below, show good agreement between the distributions of the retrieved and reference cloud classes for all cloud classes except the stratus (St) class, whose frequency in the training dataset is only 0.03 %. We consider this compelling evidence that the CCIC retrieval, in fact, has skill in classifying different cloud types.

figs/spatial_cloud_class_distributions.pdf

Figure RC1.8: Spatial distribution of retrieved cloud classes and 2B-CLDCLASS-based reference cloud classes for the test samples from the year 2010. Each row of panels shows the distribution of one of the 8 cloud classes distinguished by the 2B-CLDCLASS product. The first column shows the results retrieved from CPCIR observations while the second column shows the corresponding reference distribution. Column three and four show the corresponding results for retrievals based on GridSat observations.

In addition to adding the curtain plots to the manuscript, we added text that discusses the results and points out that the confusion matrix shown in Fig. 10 assess the classification of cloud layers at a vertical resolution of 1 km, which leads to high uncertainties in the classification of individual layers.

25. Figure 13 is missing the conditional mean line

    We have made the conscious decision to not show conditional mean lines for the validation results as some of the campaigns have very few samples causing the conditional mean lines to become very noisy. While the HAIC-HIWC in-situ data used in Fig. 13 contains sufficient samples to show the conditional mean lines (see Fig. RC1.9 below) we feel that we would have to add conditional mean lines to all following scatter plots to be consistent. Since we do not think that the conditional mean lines add significant information to the scatter plots, we have decided against showing conditional mean lines for the validation results.

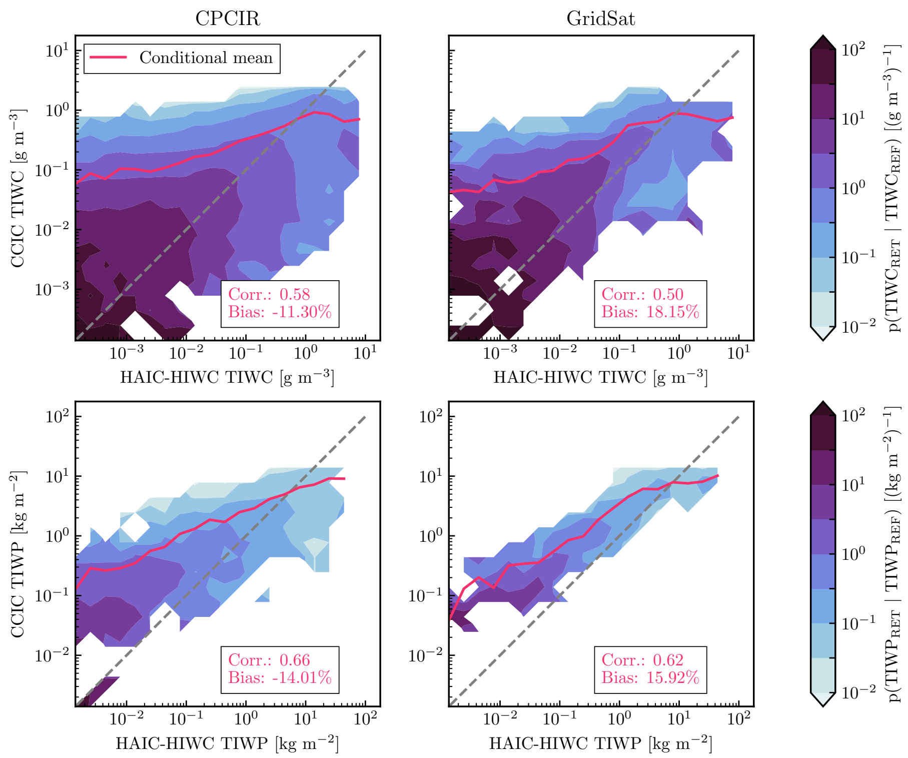

Figure RC1.9: As in Fig. 13 from the preprint, but with conditional mean lines.

26. Line 417: CCIC is not representing the diurnal variability well, it is really flat.

    While the CCIC results certainly do not represent the diurnal variability perfectly, the retrieved and reference diurnal cycles show a high degree of correlation. To make this point clear we will add a table containing the relative biases and linear correlation coefficients of the retrieved mean TIWP to the manuscript. These results show that the correlation of the diurnal cycles calculated over the full year is is 0.86 (0.97) for CPCIR-based (GridSat-based) retrievals and does not fall below 0.74 for any of the assessed three-month periods.

27. Line 438: provide an estimate of the uncertainties associated with the estimates of the ice hydrometeors.

Table RC1.9: Relative bias and linear correlation coefficient of the diurnal cycles retrieved using CCIC compared to those derived from ground-based cloud-radar observations.

	CPCIR		GridSat	
Time period	Bias [%]	Correlation coeff.	Bias [%]	Correlation coeff.
All year	27.43	0.86	34.59	0.97
DJF	37.97	0.92	23.58	0.97
MAM	53.04	0.81	73.47	0.92
JJA	45.76	0.74	30.92	0.96
SON	-0.66	0.84	25.79	0.92

We have extended the sentence in question to reference the validation study of A-train-based IWC retrievals by Deng et al. (2013) and mention the biases of up to 59 % compared to in-situ measurements.

28. Line 456: "Despite these encouraging results, CCIC should still be considered a proof of concept. CCIC's principal objective remains to explore the potential of modern deep-learning techniques to expand the observational climate record of ice clouds". This should be mentioned upfront in the abstract and the introduction.

    Since we have in the mean time processed the full observational record of available geostationary observations, it is not adequate anymore to consider CCIC merely a proof of concept. We will therefore the sentence from the revised manuscript.

29. Table A1: "Cloudy pixel" Cloudsat or retrieved? how come the cloudy pixel is 40ish while the no cloud is 97 %, please clarify

    "Cloudy pixel" refers to cloudy profiles containing a cloud anywhere at the 20 vertical levels used by CCIC. The 97 %, on the other, hand refer to the fraction of non-cloudy levels. To make this point clearer we will replace 'cloudy pixel' with 'cloudy profile' in Table 1 and update the caption.

References

H. W. Barker, A. V. Korolev, D. R. Hudak, J. W. Strapp, K. B. Strawbridge, and M. Wolde. A comparison between cloudsat and aircraft data for a multilayer, mixed phase cloud system during the canadian cloudsat-calipso validation project. *Journal of Geophysical Research: Atmospheres*, 113(D8), 2008. doi: https://doi.org/10.1029/2008JD009971. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JD009971`.

M. Deng, G. G. Mace, Z. Wang, and H. Okamoto. Tropical composition, cloud and climate coupling experiment validation for cirrus cloud profiling retrieval using cloudsat radar and calipso lidar. *Journal of Geophysical Research: Atmospheres*, 115(D10), 2010. doi: 10.1029/2009JD013104.

M. Deng, G. G. Mace, Z. Wang, and R. P. Lawson. Evaluation of several a-train ice cloud retrieval products with in situ measurements collected during the sparticus campaign. *Journal of Applied Meteorology and Climatology*, 52(4):1014 – 1030, 2013. doi: 10.1175/JAMC-D-12-054.1. URL https://journals.ametsoc.org/view/journals/apme/52/4/jamc-d-12-054.1.xml.

P. Eriksson, M. Ekström, B. Rydberg, D. L. Wu, R. T. Austin, and D. P. Murtagh. Comparison between early odin-smr, aura mls and cloudsat retrievals of cloud ice mass in the upper tropical troposphere. *Atmospheric Chemistry and Physics*, 8(7):1937–1948, 2008. doi: 10.5194/acp-8-1937-2008. URL https://acp.copernicus.org/articles/8/1937/2008/.

P. Ren, Y. Xiao, X. Chang, P.-y. Huang, Z. Li, X. Chen, and X. Wang. A comprehensive survey of neural architecture search: Challenges and solutions. *ACM Comput. Surv.*, 54(4), may 2021. ISSN 0360-0300. doi: 10.1145/3447582. URL https://doi.org/10.1145/3447582.

D. L. Wu, J. H. Jiang, W. G. Read, R. T. Austin, C. P. Davis, A. Lambert, G. L. Stephens, D. G. Vane, and J. W. Waters. Validation of the aura mls cloud ice water content measurements. *Journal of Geophysical Research: Atmospheres*, 113(D15), 2008. doi: https://doi.org/10.1029/2007JD008931. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JD008931.