

Response to reviewer 2

Dear reviewer,

Thank you for reading my manuscript and giving thorough feedback. I have appreciated reading through your responses, they have been very helpful for identifying areas in the manuscript that could be clarified and improved upon.

One common theme of the responses from both reviewers was about the figures having too small text. To this end, we have doubled the font size across all figures, and will work with the editors to make sure that the figures appear properly in any final copy that is produced. The figures have also been enlarged within the document, allowing more space for them to be viewed on the page.

Reviewer comments are provided in blue, with author responses given in black below.

General comments

1. The Introduction is very informative but I would consider it too long. I would recommend to make it a bit more compact while keeping all relevant references in. I believe the topic of the manuscript has a lot of nuance that may be unfamiliar to readers new to the field. I have tried to justify each part of the motivation for the framework, which has lead to the introduction being two pages long. I worry that shortening the introduction too much risks the loss of important details. I have changed some of the writing to try to make some sentences easier to understand.
2. The paper is lengthy, so I would recommend to try out ways to reduce its length as it could be a big discouraging for a potential reader to read it, even though its topic of discussion is of a great significance. Towards this aim, I would recommend to move some from the Appendices to a Supplement, which will come in a separate document. To my opinion, Appendix B is a good candidate to be moved to the Supplement. Moreover, Appendix C could become another section in the supplement. If you still decide to keep it in the manuscript, please try to make it shorter. I believe that Appendix A should still stay in the manuscript. I agree that the paper is long, and this could be discouraging. We did try to keep the manuscript shorter, but the nuances of the framework and results needed explaining and this increased the length of the manuscript. We submitted to AMT specifically because the journal allows for longer articles, which allowed us to go into sufficient detail to explain and justify the new framework.
3. The section 4 “Discussion” is also too lengthy. I would not create separate subsections for 4.1, 4.2 and 4.3. I would just be merging it to one while explaining the important parts of the frameworks “The choice of mutual information estimator”, “Physical interpretation of \hat{p} ” and “Choice of co-location scheme”.

Having considered this suggestion, I believe keeping the headings 4.1–4.3 in the manuscript is important, as they represent specific and overarching questions about the use of the framework in future studies. Condensing them under a single subheading would make the information harder to find for a reader referring back to the paper.

4. Section 3.6.3 contains some weak statements as the comparison of the two extreme parameterizations p_{00} and p_{11} is not that strong. Try to make this section compact and more meaningful.

I have made some changes to 3.6.3 to hopefully make the meaning of the comparison more clear.

5. I would recommend to completely remove Section 3.7 too. I do not clearly see any strong point to keep it. On the contrary, the manuscript will benefit from the change as it will become shorter.

Section 3.7 was added as a response from co-authors that the manuscript needed to show a “real-world” comparison and the effects of using different co-location parametrisations on the outcomes of the comparison. It reinforces the point made in the introduction that the selection of a co-location parametrisation that optimises the value of a comparison metric (in this case the mean bias and variance of the bias distribution) would not necessarily agree with a choice of co-location parametrisation that optimises for the mutual information between the measurand distributions.

6. The Figures have some issues with the axis labels. The font size is way too small. When the reader prints the manuscript, the axes are not even readable. Even on a typical-size laptop screen, the axes are not visible and only with zooming in you can read them well. But this interrupts the flow of the reading. I would recommend to simply increase the font size of all axes in the figures.

See above

Specific comments

1. The reference (Loew et al, 2017) appears too frequent in the Introduction; precisely four times and then again in the second sentence of the Section 2 “Framework”. It seems like a repetition.

I agree that I have heavily used the reference Loew et al. (2017). It is an important and comprehensive review of satellite validation methodology, and was a primary source of inspiration for this paper. I have changed the fourth reference to the newly published Verhoelst et al. (2026), as it is also applicable.

2. Line 60 – I would not start the sentence “This should be avoided”. Please rephrase it. This has been rephrased.

3. Introduce numbering items for the co-location scheme, criteria, parameterization and event. Their definition is good as it introduces early enough the terminology to the reader. It is very much appreciated.

The definitions have been numbered.

4. Figure 1 is very good introductory of the potential schemes. Please add main references for each scheme to the figure legend. There are many references in the text but it is still nice to have the visualization of the scheme together with the main reference(s).

I tried several ways to introduce the references into the Figure itself, and into the Figure caption. They all cluttered the figure and required additional space for the figure which would make the manuscript even larger. I have instead highlighted in the caption that references can be found in the text in Sect. 2.1, in case a reader would like to find the lists of references.

5. Line 105 – What “best results” mean? Earlier in the introduction (line 60), it was recommended that this should be avoided.

I have changed this to be “better results”. This is in a different context to the comment on line 60. If we co-located data by matching purely based on the same time, instead of including the location of the satellite as co-location criteria, then data could be matched when the satellite is antipodal to the Cloudnet site. This is not physically motivated, and would clearly perform worse than the co-location scheme that is more physically motivated and includes a spatial co-location criteria. This is what I mean by “better results”.

6. Line 122-124 – The sentence “For example, ... ancillary wind data” is disconnected with the previous text. It only makes sense with prior explanation why/how local advection might affect the colocation scheme. It should be further explained and ideally be moved to another location in the text.

This comment isn't to tell a reader exactly how a co-location scheme should be changed to include additional ancillary data, only to highlight that it can be changed, and in a myriad of ways.

To include expectations of the advection into the co-location, the user of the framework could offset the circle of radius R from the ground-based station by the vector $(\langle u \rangle, \langle v \rangle) * \tau * k$, where $\langle q \rangle$ is the average of q over the duration spanned by τ ; u and v are the local wind velocity components; and k is a scaling constant. In this case, k is a new free parameter in the co-location scheme. This is only one way, the user could instead make the circle eccentric based on the mean wind vector and co-locate data within an ellipse instead of a circle, oriented along or perpendicular to the mean wind vector. There is no single correct co-location scheme, so I would not want to prescribe a particular scheme in the manuscript, in case people interpret this as the only way to do it.

7. Figure 3 is very nice. Please make clearer that the preferred situation is depicted in plot 3b.

I have updated the figure caption to highlight this. I have also created a box around 3b to make this obvious.

8. Line 221 – Remove the first bracket from Eq. 3. It should appear as...

The outer brackets have been removed.

9. Line 233 – The vertical (range) resolution is on the order of decimeters (~10–30 cm), determined by timing precision. 30 m should refer to the horizontal segment length,

not the vertical resolution. Do you mean the following: “Photon returns are accumulated into 30 m along-track segments, which are further aggregated over 400 consecutive laser pulses to produce a product with an effective along-track resolution of approximately 240 m.”?

Your comment is partially correct, in that the vertical resolution of individual photon returns is dictated by the timing precision of the ATLAS detector, and is on the order of centimeters to decimeters. However, in order to boost the signal to noise ratio and produce a gridded data product, ATL04 consists of 30 m vertical bins, with any photon return falling within a bin contributing to the photon count of that bin.

The along-track resolution was however misquoted, being 280m instead of 240m. The 280 m along-track resolution comes from the pulse repetition frequency of 10 KHz, and the orbital speed of ICESat-2, yielding an along-track resolution of $(400 \text{ pulses per profile} / 10_000 \text{ pulses per second}) * (\sim 7_000 \text{ m/s}) = 280 \text{ meters per profile}$.

10. Line 278 – There are some problems with Eq.5: (a) The vertical bar | is ambiguous here (it usually means “such that” in set notation or conditional probability), (b) The prime notation j' is unnecessary and makes it harder to read. The easier is to express it like...

This has been implemented.

11. Line 286-287 – “The observed clouds are optically thick enough to attenuate the ATLAS lidar beam throughout the co-location event, so lower-level cloud layers will be missed.”. This is an important information which can have an important effect on the comparisons against the ground-based VCF. However, I would expect that the opposite might happen for the Cloudnet data. The higherlevel cloud layers will be then missed? Or what is the situation exactly?

This is a correct observation. The ATL09 retrieval only consists of lidar measurements, so is sensitive to attenuation due to liquid water in cloud layers. And this will be part of the interest when comparing ATL09 VCF retrievals against Cloudnet retrievals – is there so much attenuation that the ATL09 VCF profiles lose significant information with respect to the Cloudnet profiles. Equally, the attenuation itself can be considered additional information, providing upper bounds on the VCF values at lower altitudes. The Cloudnet retrieval suffers less from this problem as it is synthesised from lidar, microwave radiometer and radar data. The radar in particular experiences less attenuation than the lidar signal, and liquid attenuation acts to degrade the quality of the Cloudnet microphysical classification, but less so the macrophysical classification of hydrometeor presence in the atmosphere. Thus, the Cloudnet retrieval suffers less from missing higher level clouds (although it will still happen) than ATL09 suffers from missing lower level clouds.

- 12.

- a. Fig 4c should indicate a minimum of 18.4km, if I understood well? Please add the information to the plot.

You are correct, I tried adding this to the plot labels but I think it clutters the plot and isn't necessary information.

- b. Also, it might be preferable to show the area which corresponds to the contributing pairs ATL09-Cloudnet at the co-location parameterization

The area of contributing vertical profiles given the parametrisation is the unhatched regions of b-e, and the unshaded region in a. I have made this more explicit in the figure caption.

- c. Fig. 4f shows that the profiles have huge differences for altitude 4km and below. How do the uncertainties look like for the two measurements? Are the uncertainties taken into consideration at all? It would be highly advisable to include the error bars into the VCF profiles.

Computing error bars of an aggregation over a binary mask, such as the arithmetic mean of the cloud mask at a given height is an ill-defined process. I did compute error bars at one point through bootstrapping, but the uncertainties are not used in further analysis and do not meaningfully add to the plot. A full uncertainty analysis of these cloudmasks is outside the scope of this work.

I have added in a line to represent the ATL09 cloud-and-attenuation profile, which provides an upper bound for the amount of cloud detected by ATL09. Regions that are attenuated in the ATL09 data product could be considered to vary between totally cloudy or entirely cloud free. Interestingly, we see that when the ATL09 cloud deviates from the Cloudnet VCF, the ATL09+attenuation profile still correlates with the Cloudnet VCF to much lower altitudes.

- d. Moreover, it seems like there is an overlap issue in the Cloudnet profile. Can you please explain that? The VCF profile increases exponentially from the ground up to the first 1.0-1.5km where it reaches the value of 1.0 (i.e., fully cloudy). At the ATL09 profile the VCF is exactly zero, which could imply that the ATLAS laser beam is completely attenuated at those altitudes. But then, the vertical bins should be removed from the dataset in a pre-processing step. Is this performed? Can you also elaborate on the potential problem of incomplete overlap of the Cloudnet lidar? The incomplete overlap problem refers to the situation in which, at short ranges from a lidar system, the laser beam and the telescope field of view (FOV) do not fully overlap. Remaining incomplete overlap problems in the retrieved vertical profiles could introduce artifacts which should be removed in a kind of pre-processing step.

We can see from Figure 4b that the ATL09 data is attenuated at ground level throughout the entire scene, which explains the 0 VCF value at ground level from ATL09. Removing these bins from the data would however make the analysis infeasible, as each co-location event would be considering ATL09 data in different-dimensional state spaces, which would be incomparable, and no mutual information would be computable between the ATL09 and Cloudnet data.

Interestingly, the lidar overlap problem should not affect the mutual information computed between the Cloudnet and ATL09 data. The overlap problem can be corrected for by a height-dependent multiplicative correction to the measured lidar backscatter signal. This transformation is invertible/reversible/lossless, as it is simply a scaling of the lidar backscatter values at lower altitudes. One property of mutual information is that it is invariant under reversible transformations of the input data distributions. Thus, whether or not the overlap correction is applied to the data would not affect the mutual information analysis (although it would affect subsequent

analysis of the data after determining the co-location parametrisation). I have added in a paragraph to explain this in Sect. 2.2.

This does raise another good point though, in that the Cloudnet categorize data product uses a height grid measured from mean sea level. On top of that, the minimum detection range of the radar is ~100m above the radar, and the Cloudnet stations are all located above 0m msl. The vertical gridding from the homogenisation process has a bin that is constructed from -120m to +120m, which includes ATL09 data, but does not include Cloudnet data, so it is uniformly homogenised to contain 0 cloud fraction. Removing ATL09 data from height bins below the Cloudnet station could result in the loss of information from the ATL09 data if it passes over lower-lying surfaces.

I have removed the lowest data point from the Cloudnet VCF profile in figure 4f, and have also included a line showing the ATL09 cloud + attenuation profile.

13. [Line 301-302 – “Above 5 km in height, both VCF profiles visually correlate with each other, indicating that the co-location ... unobserved by ICESat-2.”: Could it be that simply above 5 km, the ground-based lidar face laser beam attenuation issues? Or this is certainly not a case in such conditions of optically thick clouds present?](#)

If the Cloudnet retrieval faced attenuation issues, we would expect the decrease in the Cloudnet VCF profile with height. However, we would expect it to not agree with the decrease in the ATL09 VCF profile, as they would be making independent observations. Attenuation in the Cloudnet VCF values would then miss cloud layers that we would expect ICESat-2 to observe. As noted above (point 11), the Cloudnet retrieval synthesises both lidar and radar data, which should be able to detect the same higher altitude clouds that ICESat-2 is seeing. Thus, I believe that ICESat-2 and Cloudnet are observing the same cloud layers at higher altitudes.

14. [Line 311 – Can you elaborate more on the selection of k=10?](#)

The selection of k=10 was chosen as a compromise between having a larger value of k (i.e. k=20), which is less subject to variance induced by the input-data, but is more computationally expensive; against a choice of smaller k (i.e. k=1,2,4) which is subject to larger variances dependent on the input-data, but is computationally cheaper. I tested small batches of parametrisations with k=1,2,4,10,20, and found that k=10 was the largest value tested that ran within a sensible timeframe. I have reworded this in the text to hopefully make this clearer.

15. [Line 319-320 – Further, can you elaborate on the \$n_i=10\$ and \$n_{repeats}=20\$ selection?](#)

The numbers $n_i=10$ and $n_{repeats}=20$ we selected fairly arbitrarily. The time taken to compute each MI variance estimate scales linearly with $n_{repeats}$, so 20 is picked to obtain a good number of values for a maximum likelihood estimate whilst not taking too long. $n_i=10$ non-overlapping subsets of the VCF profiles are used in each of the computations as the values computed are expected to follow a chi squared distribution with n_i-1 degrees of freedom. Too large a value of n_i takes a long time to perform many mutual information computations, whilst a too small value of n_i requires more repeats to reliably obtain a maximum likelihood estimate for B.

16. Line 325 – And if null hypothesis is rejected, then p is not a good candidate, right? Probably refer to the t-test statistical test you make. Consider rephrasing to “For each $p \neq \hat{p} \neq \hat{p}$, we apply a Welch’s t-test to evaluate the null hypothesis that the mean estimated mutual information $I^{\text{KSG}}(p)$ does not differ significantly from $I^{\text{KSG}}(\hat{p})$. Parametrizations for which the null hypothesis is not rejected at the 0.05 significance level are considered candidate optimized parametrizations.”.

I like your phrasing about the null hypothesis and when not it is not rejected. This has been implemented.

17. Line 331 – Could you explain how the measurement uncertainties are counted into the selection of partially cloudy category of very small VCF (e.g. $\text{VCF} < 0.1$)?

As mentioned in my response 12c, the uncertainties of the VCF profiles aren’t well defined quantities.

If there were uncertainties on the VCF values, then the distribution would presumably not be degenerate at 0 and 1 (or could be made continuous through a kernel density estimate that used the values and their uncertainties to compute contributions to the joint probability density of the VCF values at arbitrary values). It is the degeneracies at VCF values of 0 and 1 which make the VCF distribution a mixed measure distribution, and prompted the need to classify cases into the nc, pc and tc categories defined.

18. Line 341 – Please revise Eq. 8. Option A (copula in terms of uniform variables): Option B (copula expressed with the original variables): Both are correct; which one you use depends on context and clarity.

I revised this to use Option A.

19. Figure 5 – Plot 5c looks like there is contamination from independent data? It appears very similar to Figure 3d. Please clarify this point. Also, if there are no unique parameterization and the hatching shows several possible good parameterization candidates, then some analysis should be included which shows and explains how much the impact could be.

Figures 3d and 5c are showing completely different quantities. Figure 3d shows a probability distribution between two non-linearly related variables that is contaminated by independent data. As a result, the distribution has a low mutual information value of 0.018 nats encoded in the distribution.

Figure 5c instead shows a collection of mutual information values computed over a grid of co-location parametrizations. At each point on the mutual information surface, you can imagine that there is a plot similar to 3a-d (but in 100 dimensions), from which a mutual information value is computed, determining the height of the surface at that point in the parametrization space.

A sentence has been added to make this distinction clearer.

Given the mutual information across the set of candidate parametrizations should be similar, it would imply that the joint probability distribution between the Cloudnet and ATLO9 VCF distributions is likely similar across the set of candidate parametrizations too. Thus I would expect that the selection of co-location parametrizations from within the set of candidates should not alter the resulting analysis significantly, except from the volume of data available for the analysis as this changes as a function of p .

20. Line 397 – “As with Nevents, the smooth color gradient ... and τ .” But, Nevents have not been expressed as a function of τ . Please clarify this discrepancy.

We can describe $N_{\text{events}}(R, \tau) \propto R$ (i.e. there is no τ dependency, or τ^0). We can also describe $N_{\text{profiles}}(R, \tau) \propto R^2 \tau$. The reasons for this are given in the text. The smooth colour gradient described is a result of the variable being plotted (N_{events} or N_{profiles}) being polynomial in both of the plotting coordinates (R and τ). Either way, this in depth description of the dependency on N_{events} and N_{profiles} to the co-location parametrisations is to show that they are consistent with our expectations, and that N_{events} and N_{profiles} vary across the parameter space in a way that is dissimilar to how the mutual information varies across the parameter space. I have changed the proportionality descriptions to emphasise that $N_{\{\text{events/profiles}\}}$ are functions of the co-location parametrisations.

21. Line 418-419 – I am not sure if I interpreted correctly. Do you imply that the difference of 10 hours (from 8 hours to 18 hours) is significant or not?

The significance described in the text is with respect to the differences in mutual information computed at different parametrisations, not with regards to the values of the parametrisations themselves. Parametrisations with $\tau < 8$ hours will include less data, making the analysis data-limited. Parametrisations with $\tau > 18$ hours will permit data that is independent between Cloudnet and ATL09, enough so that it would degrade the inferred relationship between the VCF distributions from each data source.

22. Figure 6c – The dispersion of the possible optimal co-location parameterization is high. There is some explanation/interpretation in the text which associates this to the flat terrain around the ground-based station. Could you give an estimate of how much such a condition could impact the comparisons over Hyttiala?

I don't fully understand the question being posed here, but I will attempt to answer what I believe it to be – How would selecting different co-location parametrisations for Hyttiala affect the comparison between ATL09 and Cloudnet data, and how does this relate to the terrain?

It is my belief that selecting any of the parametrisations within the hatched region in Figure 6c should produce quantitatively similar final results in a comparison between the Cloudnet and ATL09 data. In theory, the 50-dimensional marginal distributions of the VCF profiles for both ATL09 and Cloudnet do not significantly vary across the set of candidate parametrisations. As such, the only difference in the mutual information between the ATL09 and Cloudnet distributions will come from changes in the relationship between the two distributions, their joint distribution. I struggle to see how over a small range in parameter space that the joint distribution relationship would change drastically in form, such that the mutual information remains similar but that the conclusions that would be drawn from the relationship would significantly change.

The argument relating the terrain to the larger \hat{R} values is that if there were nearby mountains, that could induce changes in the cloud regime in the vicinity of the mountains. So the clouds near the mountains would be affected by different physical processes than the clouds at the ground-based station. Thus, a parametrisation whose R value was big enough to include the mountains would be introducing

observations between physically independent clouds at smaller values of R compared to a location without nearby mountains (or other large changes in orography).

23. Line 427-428 – It appears like there is contamination from independent samples like in Fig. 3d. Could you please clarify this point and probably rephrase the sentence “regions where the input data to the estimator is contaminated with independent samples”?

Figure 6 shows mutual information surfaces, not individual distributions which can be contaminated with independent data (see my response to 19a).

24. Line 434 – The statement “are similar orders of magnitude” is a bit misleading here as the minimum of 4 hours is less than half the maximum of 10 hours. Please rephrase that.

I think 4 and 10 are similar orders of magnitude. I’m considering 10 to be an order of magnitude, in which case the numbers differ logarithmically by ~ 0.4 .

25. Line 435 – There is a statement that the temporal values τ are consistent with other studies but there is no statement about the spatial values R . Are also those consistent with available literature?

I have looked for available literature, I haven’t been able to find much explicitly about the horizontal scales of cloud variability from satellite observations. The closest I could find is Feijt and Jonker, (2000), *Comparison of scaling parameters from spatial and temporal distributions of cloud properties* (DOI: 10.1029/2000JD900414) but this more addresses finer scale spatial variation within a given cloud regime as opposed to the scale of variation between cloud regimes.

26. Line 455 – If I understood correctly, the selection of 100-dimensional joint probability distribution constrains the sampling in the other 3 locations. Probably for each location, depending on the latitude that the ground station lies, other selection should be done? What is the reasoning of having 100 dimensions?

The dimensionality of the joint probability distribution is the sum of the dimensionalities of the individual measurements being compared. In this case, our VCF profiles consist of cloud fraction values recorded over a vertical profile containing 50 distinct vertical levels. As such, the distribution of each VCF measurement is 50 dimensional, and the joint probability distribution is $50+50=100$ dimensional. If we were instead comparing integrated cloud geometric thickness from ICESat-2 (1 dimensional), and mean cloud geometric height and thickness from Cloudnet (2 dimensional), the resulting joint probability distribution would be $1+2=3$ dimensional. The dimensionality depends purely on the measured values being compared, and not on where they are being compared.

A 100 dimensional joint probability distribution is rather horrible. The dimensionality can be reduced by only considering coarser vertical grids for VCFs, but this reduces vertical resolution. If a different vertical resolution were used at each location to account for the different sample density at each location, the results of the subsequent comparisons between the ATL09 and Cloudnet VCFs would be incomparable between locations.

27. Line 476-480 – I cannot clearly see what it is hidden in those sentences. Could you proof that another set of co-location parameterization would not be suitable? Could you think of any validation of the optimal co-location parameterization?

$\hat{p}_{\text{Ny-Ålesund}}$ is provided as an example, but using any of the \hat{p} values from a specific site would induce problems at other Cloudnet sites (by selecting a parametrisation with reduced mutual information for whatever reason), hence the one-size-fits-all approach being flawed.

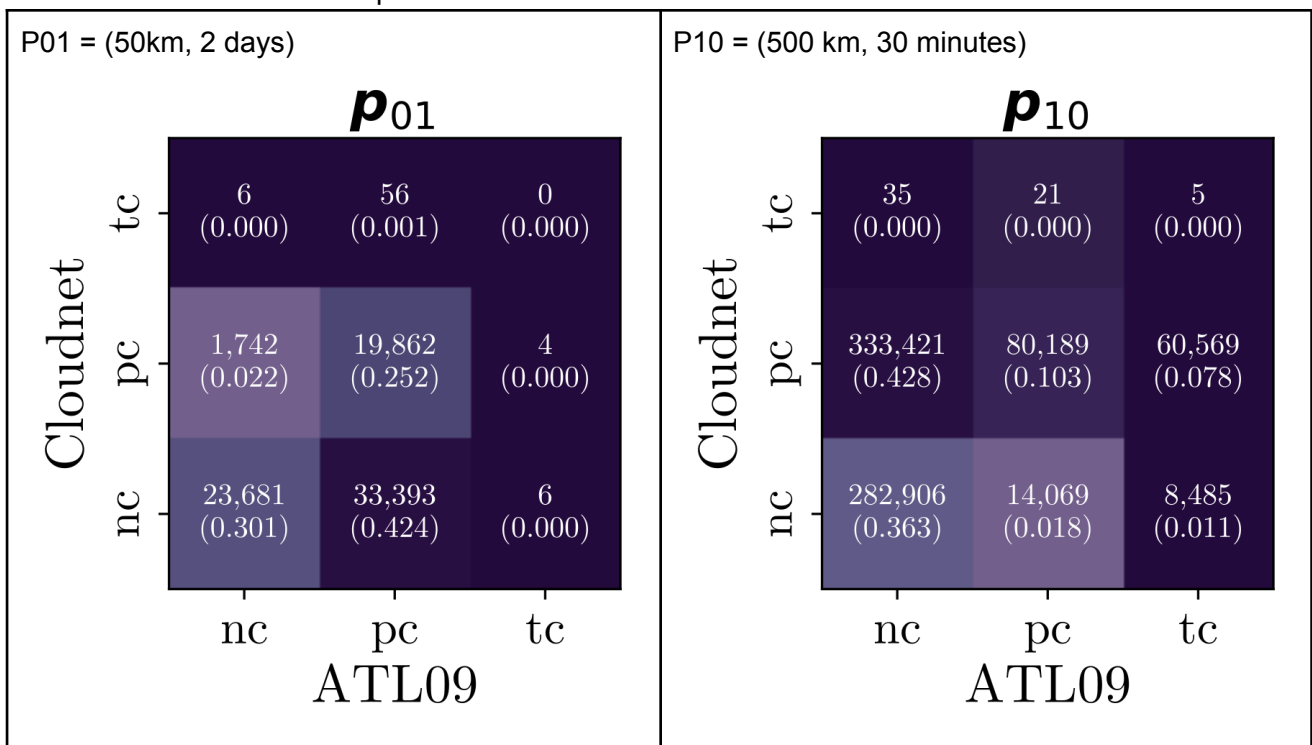
Inherently, there is no proof that a different tested co-location parametrisation would be better as the parametrisation that produces the highest mutual information is the optimal choice by definition within the framework.

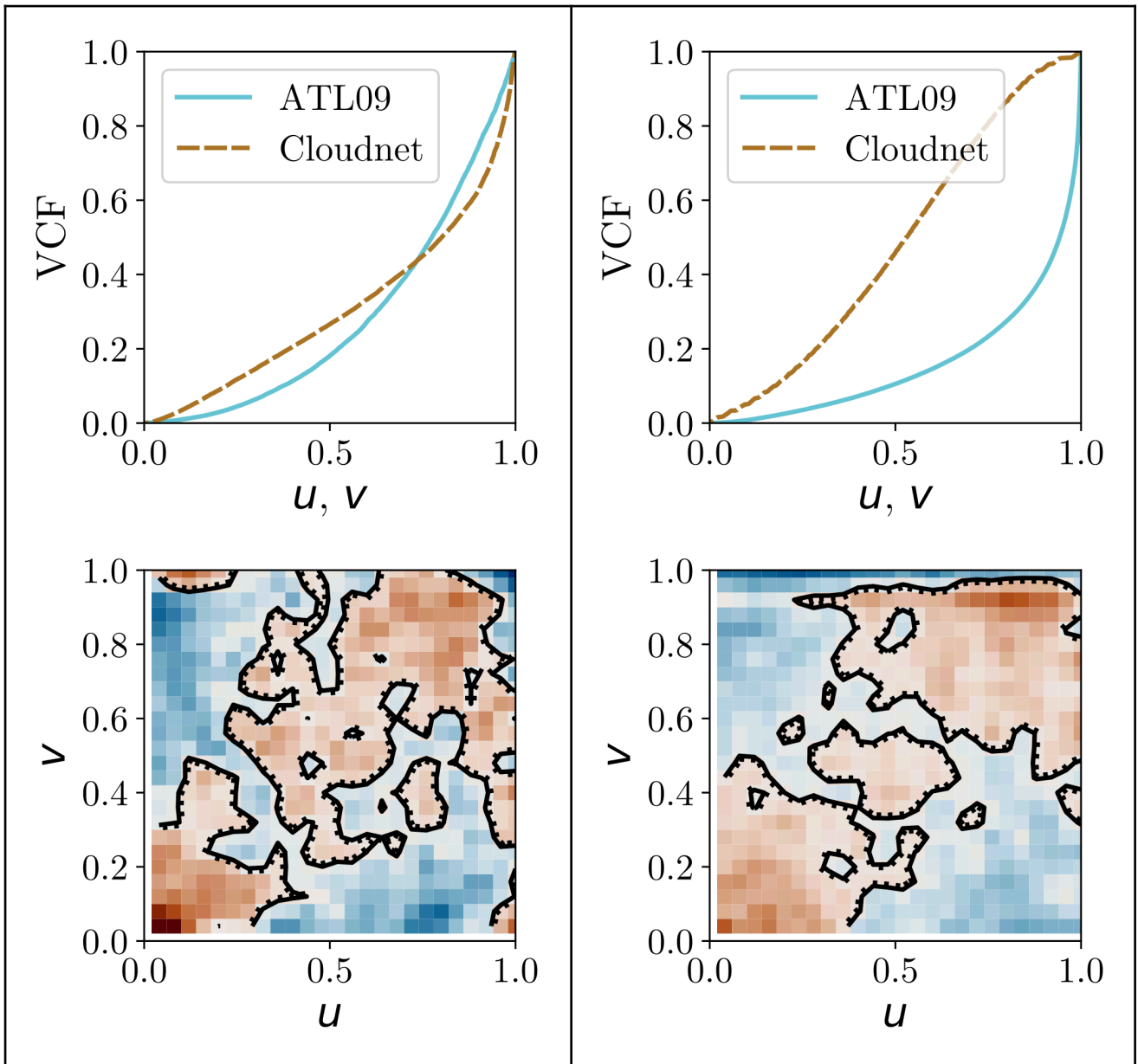
A validation of the optimised co-location parametrisation is possible however, by comparing its performance against other co-location parametrisations using independent metrics. This is demonstrated in Sect. 3.6.3.

28.

- a. Instead of comparing with the two extremes p_{00} and p_{11} , I would also compare the confusion matrices for other potential candidates of a co-location parameterization. Probably, I would pick the one with the smallest temporal window and the one with the largest spatial extend.

This was in fact done, and is the reason why $p_{\{00\}}$ and $p_{\{11\}}$ are not simply $p_{\{0\}}$ and $p_{\{1\}}$. Similarly to $p_{\{11\}}$ and $p_{\{00\}}$, $p_{\{10\}}$ (large R, small tau) and $p_{\{01\}}$ (small R, large tau) underperform across all metrics in Table 2 when compared against p_{lit} and \hat{p} . They were removed to make the manuscript more concise.





parametrisation	R (km)	tau (hours)	ACC	c_min	c_max	c(1,1)	RMSD
p_{01}	50	48	0.553	0.35	2.08	0.35	0.16
p_{10}	500	0.5	0.466	0.50	1.48	0.83	0.13

b. Please also add in the legend what p00 and p11 mean. The parameterization is given in the text but the plots should also be self-informative without the need to read all the text.

I have directed the reader to Table 2 within the caption of Figure 7.

c. I also have a problem with the group of plots e-h: can they be biased due to the first 1.0-1.5 km that the VCF exponentially increases with height (w.r.t/ the incomplete overlap)? Please provide some more cases including also the other Cloudnet stations. If the VCF profiles have the same structure, then

consider to apply some filtering/pre-processing prior to the co-location parameterization selection scheme.

The cumulative distribution functions shown in Fig. 7e–h account for all of the nonzero VCF values across all heights, across all considered VCF profiles from all sites, using the given co-location parametrisation. These are generated by flattening both the ATL09 and Cloudnet VCF profiles, and effectively generating a joint histogram (and then cumulative distribution) from all of the flattened VCF profiles together. Thus, all of the Cloudnet stations are already included in this analysis, and height information is ignored beyond the pairing of values between the Cloudnet and ATL09 VCFs.

29. Line 504-505 – From the statement I understand that p_{lit} is still a good choice. Therefore, the empirical choice of the parameters τ and R as taken from the literature appear as good as the one that you found from the maximum mutual information. Please clarify this point.

If we were only to consider the confusion matrix for the (nc, pc, tc) classifications, then yes, p_{lit} and \hat{p} would appear to perform similarly. The rest of the section goes on to make the argument that if we look at the actual relationship between the ATL09 and Cloudnet VCF values through the copula density, that \hat{p} has metrics that indicate a better relationship between the Cloudnet and ATL09 VCFs when using \hat{p} over p_{lit} .

30. Line 508-509 – “The shapes of ... at lower values.” Could that be due to problems of incomplete overlap? Please clarify this point.

The large distance between ICESat-2 and the highest altitude of the ATL09 data product, relative to the 15km height span of the ATL09 data product, means that incomplete overlap is unlikely to affect the distribution of VCF values determined from the ATL09 data. If there is a multiplicative bias in the ATL09 signal due to incomplete overlap, it would be fairly homogenous throughout the scene, which would have little impact on the shape of the cumulative distribution.

31. Figure 8 – There are essential problems with the visibility of the axis's labels in printed form. Please try to make them also friendlier for people with color vision deficiencies (CVD); certain color combinations are hard or impossible to distinguish and should be avoided.

Thanks for highlighting this. I've been using the `cmcramer` python package (Cramer, 2023) and the Coblis online CVD checking tool (<https://www.color-blindness.com/coblis-color-blindness-simulator/>). I have also made extensive efforts to make plots CVD friendly through other means, such as differentiating line markers and styles. For Figure 8, I'm using a colorblind friendly map (Acton) for the probability density, consistent with the rest of the paper. Plotting lines over the top of a colourblind friendly colourmap will almost necessarily have places with poor contrast (particularly under monochromaticity) when the line goes from lighter to darker parts of the underlying colourmap. I have tested Figure 8 against all given CVD conditions provided on Coblis. I agree that there are places in 8a–d where there is poor contrast between the underlying probability density and the line plotted over the top. As such, I have changed the colours of the lines to ensure

that reasonable contrast is maintained between the lines and the probability distributions in 8a–d, and that there is some contrast in 8e–f, but distinction between the lines can be determined by their plotting styles.

32. Line 528-530 – I would not conclude the session with such a strong sentence. There is no validation which proves the statement, right?

I have changed the end of the section to remove language like “better” and make the statements more objective.

33. Line 534-535 – Could this be related to signal attenuation from the ground-based lidar? Please comment on this.

This is due to the different sensitivities of the retrievals to atmospheric scatterers at different heights. This is partially due to a decrease in sensitivity of the retrievals with separation between the scatterer and receiver (due to the inverse-square law) and it is partially due to attenuation effects. I have mentioned attenuation effects in the manuscript to clarify this point.

34. Line 824 – Equation B1 could be rephrased similarly to Eq.5 at Line 278

This has been implemented.

35. Line 880 – Equations C10-C11 must have a wrong sign. Please check if the minus sign (-) must be present in the aforementioned equations.

Thank you for spotting the sign error. I have changed the equations to rectify this, and have filled in the equations from (what was) C14–C16 (and what is now C13–C17) to show how the full derivation for how the final result is obtained.

36. Figure C1 – Why do you include in the plot also the ISS, EarthCARE and A-Train satellites? Relevant to that, there is some text in lines 901-902 about those satellites. Please provide a statement of why would this be relevant to this manuscript? Do you foreseen to apply this co-location parameterization to other missions?

I include the other satellites in the plot, as they are satellites from which vertically resolved atmospheric measurements are made. The same framework and same co-location scheme described in the manuscript can be applied to retrievals from all of these satellites when comparing them against surface-based observations. I think the framework as a whole can be applied to any and all satellite retrievals, with the outcome being the identification of a co-location parametrisation that is location and data-source dependant.

Technical corrections:

- Line 16 – Typo of “validation an constraining” should be “validation and constraining”. corrected
- Line 62 – Remove the second “to the”. corrected
- Line 103 – Remove the second “be”

corrected

- Line 118 – “Co-location” with capital letter.
corrected
- Line 229 – Please correct the ATLAS acronym to Advanced Topographic Laser Altimeter System (ATLAS)
corrected
- Line 452 – There is an extra “a”. Please remove it.
corrected
- Line 541-543 – This sentence reads like being replicated to the lines 535-537. Please remove or rephrase.
corrected
- Line 562-563 – In case you do not remove the complete section 3.7, please remove the conclusive sentence.
corrected
- Line 776 – “As 1/N ..” with capital A.
corrected
- Line 784 – The sentence does not read well. Remove either the word “is” or add the word “that” after “we see”.
corrected
- The doi of the reference “Herzfeld, U., Palm, S., and Hancock, D.: Ice, Cloud, and Land Elevation Satellite (ICESat-2) Project Algorithm Theoretical Basis Document for the Atmosphere, Part 2: Level 2 and 3 Data Products, version 4, <https://doi.org/10.5067/JT3IX3YFW1RR>, publisher: NASA National 960 Snow and Ice Data Center Distributed Active Archive Center, 2021b.” is wrong.
corrected
- 48. The same for the doi of the reference “Palm, S., Yang, Y., Hertzfeld, U., and Hancock, D.: Ice, Cloud, and Land Elevation Satellite (ICESat-2) Project Algorithm Theoretical Basis Document for the Atmosphere, Part I: Level 2 and 3 Data Products, version 4, <https://doi.org/10.5067/VZIDW16UA4S1>, publisher: NASA National Snow and Ice Data Center Distributed Active Archive Center, 2021a.”
corrected
- 49. The same for the doi of the reference “Vinjamuri, K. S., Vountas, M., Lelli, L., Stengel, M., Shupe, M. D., Ebell, K., and Burrows, J. P.: Validation of the Cloud_CCI (Cloud Climate Change Initiative) cloud products in the Arctic, Atmospheric Measurement Techniques, 16, 2903–2918, <https://doi.org/10.5194/amt-16-1095-2903-2023>, publisher: Copernicus GmbH, 2023.”
corrected