Response to reviewer 1

Dear reviewer,
Thank you for thoroughly reading the manuscript and your enthusiastic response. The feedback has been valuable in making changes and improvements.

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.

Specific responses:

1. The text in figures would, ideally, be similar in size to the text within the captions. While those reading a PDF can zoom in, those of us foolish enough to print off papers to get some time away from a screen cannot read any of the labels in the figures of this paper. Is it possible to regenerate the diagrams with text scaled to their printed size?
   See above.
2. Though it isn't necessary, the paper may be improved by a qualitative description of some of the mathematical operations in S2.3 as most researchers in atmospheric science do not have much training in the formalism of mathematics. For example, "nearest neighbour distances between samples" would likely be interpreted to refer to the physical distance between observations rather than the distance in state space. The references given are accurate but not necessarily useful for the audience of this journal; a textbook or lecture course may be useful (as was done for the description of the copula).
   I have included the mention of "sample space" over state space to the description in Sec. 2.3. I have also added a reference to a book by James V. Stone (Information theory: a tutorial introduction) to Sect. 2.2 to direct readers to an easy to digest introduction to information theory.
3. L68: It's not clear to me why optimising the correlation coefficient biases the results *high*. This may be a matter of definition, as I'm reading this line to say "we are preferentially selecting results where one is consistently larger than the other" as that is how 'bias' tends to be used in atmospheric science. You may have meant "we are preferentially selecting results that are large" (in magnitude) or "we are preferentially selecting results that resemble each other". My instinct is that the sample will be biased but with unpredictable sign. There is every chance this is a standard result in statistics of which I am unaware.
   The description in the introduction has been changed to reflect that it is the value of the correlation coefficient itself which is biased high (towards "better" values), not the bias between the values being compared.

4. L161 and L163: The values given in the text differ slightly from those given in Fig. 2. Please check which is correct.

   I have updated the values referring to figures 2 and 3 since the figures have been regenerated.

5. L270: I agree that, in the 2-D presentation of Fig. 4, Cloudnet data is a point. However, in 3-D space, the satellite swath is a (mostly vertical) surface and the site is a vertical line. I'm not sure the document would be improved by making this point, but the pedant in me could not leave this distinction unmentioned.
   I have addressed this by adding "Projected onto the Earth's surface,...", which hopefully clarifies the point.

6. L401: It is not obvious to me that the number of profiles scales with \cosh(R). I would have expected this to come from the rule of cosines giving the length of a bisector of a circle to be $R\sqrt{2(1-\cos\gamma)}$; hence linear R dependence. Regardless, if you are correct, isn't $\lim_{R \to \infty} cosh(R) = \infty$ rather than R?
   I agree, and I've realised the mistake I made. The distance between a satellite and a fixed point (the Cloudnet site) follows a hyperbolic curve of the form
   $(\frac{r(t)}{r_{min}})^2 - (\frac{v(t-t_0)}{r_{min}})^2 = 1$, with t-t_0 being the temporal displacement of the satellite from being at the minimum separation of r_min. However, N_profiles scales as the length of the cord traced within the circle of radius R by the satellite ground path. This can instead be derived through Pythagoras, and gives a cord length
   $s = 2\sqrt{R^2 - r_{min}^2}$. This function N_profiles \propto s(R) is hyperbolic, but not cosh(R). In the limit of R >> r_min, this is approximately linear in R. The text has been changed.

7. These are two very good questions:
   Fig. 5 and L422: The discontinuity in significant solutions deserves some more comment to guide readers in how to communicate their confidence in the selection of optimal parameters. A continuous region of significance can sensibly be communicated through traditional uncertainty notation (even if not strictly appropriate), but this second solution is much harder to communicate.

   a. Would the authors expect this degeneracy to resolve as more data is added to the problem (i.e. wait for a longer dataset) or is this an unavoidable aspect of the problem?

      This is a very good question, and unfortunately, hard to answer due to the nature of the datasets being used. I would personally expect that the degeneracy would lift and the region of possible optimised co-location parametrisations would become simply connected. However, as you elude to, this could only really be tested in this scenario by waiting for more data. This would be an interesting target for a simulation to determine.

   b. What would you do if the difference in mutual information between the two solutions was negligible?

In this scenario – where two co-location parametrisations provide the near-identical mutual information between them – I would choose the one with the smaller standard deviation estimate for the mutual information. If the two co-locations were similar in both magnitude and variance of the mutual information estimate, I would personally choose to use the co-location that maximises the available data for further analysis, although I would argue that this then becomes the researcher's choice, and that this framework provides no explicit way to choose amongst the candidate optimised co-location parametrisations.

As for how to communicate the connfidence/uncertainty in the region of candidate optimised co-lcoation parametrisations, I agree that for simply connected regions, traditional uncertainty notation (although not stricty proper) would give an appropriate way to describe the region. For disconnected regions, this could be extended to a set of traditional uncertainty descriptions, with each element corresponding to one of the disconnected regions.

8. L425: I understand this section to argue that I(p) decreases as you move away from \hat{p}, but I'm unsure that 'unimodal' is the correct word for that. Unimodal means the surface is single-valued at each point – which is true by construction for I. Collapsing the 2-D surface into a function of distance from \hat{p} clearly *isn't* unimodal because the surfaces clearly aren't isotropic about that point. 'Largely monotonic' strikes me as more appropriate but several options are available.

   I believe that unimodal is the correct term here, refering to a function with a singular maximum value.

9. L778: If the authors think there is any general utility to this simulation, it may be worth mentioning within the main body that $N > 10^4$ achieves consistency in these estimators. Some guide on the number of observations necessary to apply this method would be useful to most readers.

   I included the simulations at the end to visually demonstrate the idea of estimating mutual information in data-limitted and data-contaminated regimes. I think the idea of a simulation like this being used in advance of an analysis to (for instance) bound the values of \vec{p} that are checked is very interesting. However, I think that the model in App. A is too simplified to provide a rule of thumb for how many data points analyses should use in general. Extending this work would be interesting, but I don't have the scope to do that at present.

10. App. C: This is probably just a difference between the language of a mathematician and a physicist, but I would call these three points approximations rather than assumptions that are broken.
    Re-reading App. C, I prefer the idea of calling them approximations. I have changed the language to reflect this.

11. Sec. 3.6.2: There is no way you could have known this, but section 4 of https://doi.org/10.1007/s10712-025-09898-4 published after your submission

explores some of the issues with collocation you discuss at the end of this section. Further, in the aerosol community there is precedent in the publications of Nick Schutgens for the conclusion that a one-size-fits-all collocation criteria is sub-optimal, such as Fig. 1 of https://doi.org/10.5194/acp-20-12431-2020 or the conclusions of https://doi.org/10.5194/acp-17-9761-2017. Admittedly, Nick aims to minimise the bias, which is exactly what you argue against doing, but you might find the comparison could be useful.
Thank you for bringing the mentioned papers to my attention, they are all very interesting reads. I have included references to Schutgens et al. 2017 and Langsdale et al. 2025 in the manuscript.

Technical corrections:

- L16: validation and constraining

  corrected
- L62: expectation to the comparison

  corrected
- Tab. 1: There should be a space before N_{profiles}.

  corrected
- L598: these different quantities

  corrected
- L710: There should be a space after the comma.

  corrected
- Eq. C10: Shouldn't it be \lambda_{21} here? If I'm correct, it might be clearer for C11 to show \cos\lambda_{21} and switch to \cos\lambda_{12} for C12.

  corrected
- Maybe C14 should come before C13 as the more natural equation to follow from C5?

  corrected
- L888: this bearing into an across-track

  corrected
- I randomly tried a few of the DOI links in the references; Crameri 2023 and Palm 2021a were dead links for me.

  corrected