Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2018-168-RC1, 2018

Atmospheric Measurement Techniques Discussions

Open Access

AMTD

Interactive comment

Interactive comment on "Application of High-Dimensional Fuzzy K-means Cluster Analysis to CALIOP/CALIPSO Version 4.1 Cloud-Aerosol Discrimination" *by* Shan Zeng et al.

Anonymous Referee #1

Received and published: 23 July 2018

Review of "Application of High-Dimensional Fuzzy K-means Cluster Analysis to CALIOP/CALIPSO Version 4.1 Cloud-Aerosol Discrimination" Shang Zheng, et al.

This manuscript describes a cluster analysis technique applied to CALIPSO data products as an alternative to the standard cloud-aerosol discrimination (CAD) algorithm(s). The stated objective of this activity is to validate the CAD algorithm and better understand what is important in the classification process.

I believe the elements of a good paper are here, but the logical flow needs some significant work. I need to really dig to understand why you did this in the first place, and what specifically you learned when you were done. The authors used a large number

Printer-friendly version

Discussion paper

of statistical and otherwise assessment techniques, but often gave little justification for the choices that were made during the assessment, and minimal description of the results. I think it would have been better to not describe every single analysis that was performed for this work, and instead focus on some of the most important and try for a better explanation of what you were doing, why you picked that technique, what you expected to find, and how the differences between expectations and results are significant to the CAD algorithms. The significance should be summarized both in the conclusions and the abstract. Before I start addressing specific portions of the manuscript, I have a few general questions and comments:

1. It seems to me that hierarchical clustering, in one of its many forms, would be useful to CALIPSO data. Presumably, clouds are very different from aerosols, but ice and water clouds, while different, are much more similar to each other than to aerosols. This is not the topic of the paper, of course, by if one is to spend time looking into clustering algorithms as an alternative for CAD, why wasn't this considered?

2. At no point do you discuss the possibility of the FKM acting as a potential replacement for the standard CAD algorithms. Why not? In the last paragraph, you mention that the "FKM method is much more time consuming than the operational algorithm." I don't understand this. Does this mean it is time consuming if one attempts to recreate the FKM for the entire dataset? I would think one would create the cluster centers with a subset of data and apply to the rest, which I couldn't imagine would be slower than the CAD approach. Other than speed, I see no discussion about why FKM would be less desirable than CAD. In some ways (less dependence on non scene information like latitude), I would think it would be preferable.

3. Much of your validation relies on a comparison of various versions of CAD and FKM. In some cases this is just for a few specific scenes. Once we make the leap of faith that those scenes are representative of all scenes, the fact remains that you don't know truth. So, is agreement between CAD and FKM the best metric when they could both be wrong? It seems the implicit assumption is that FKM could never reach CAD levels

of correctness, but when if FKM does correctly identify the scene but CAD does not? In that case, ‘agreement’ isn’t appropriate as a means to validate the results. One way of addressing this problem is to apply both algorithms to synthetically generated scenes, where one knows the ‘truth’ and can verify its identification.

On to specific comments:

Abstract: As mentioned previously, the abstract gives minimal details about why you’re undertaken this study, other than the rather vague “provide new insights” and validation of CAD. What are the new insights? Does it validate as expected?

Page 2, line 23: OK, so a difference between FKM and CAD is that the latter uses latitude as an input. I would think the arbitrary nature of the use of latitude is undesirable, so if FKM is successful than it’s ability to perform without the use of latitude is very important and should be highlighted more.

Page 2: here you mention the difference between CAD V1 and CAD V4, but later on you validate against V3 and V4, and even mention V2 at some point. It seems overly complicated to compare against anything other than the latest version, but if you must you need to describe what is in each of the versions, the important differences, and why you need to validate against 3 & 4.

Page 4: OK, great, FKM doesn’t use Latitude. Why not also skip altitude? Or try with and without? Altitude to me also seems like an arbitrary input that may bias your results, although perhaps somewhat more justifiable than latitude.

Page 5, last paragraph: You’re using a random distribution of initial class memberships. How sensitive are you to that randomly selected distribution? Is there any difference in the results between one random seed and another? Also, is there any potential benefit in starting with class centers corresponding to preconceived notions of the class centers?

Figure 1, step 4: You have change in the norm of m as a metric to stop iteration (or max iteration number). Could other metrics also be used, such as change in c, or d? I've seen iterative methods that use multiple means of assessing when no further improvement can be provided as a means to reduce the number of time the max iteration number boundary is used.

Page 7, paragraph 1: could the FKM be used to improve the PDF used in CAD for the 'arbitrary' classification inputs (latitude and altitude)?

Page 7, second paragraph: "…sample is used to determine the optimal number of classes and fuzzy exponent required for classification…" How is this done? This is an important detail to skip.

Figure 2: how were these PDF's defined? Also, caption needs to spell out HOI and ROI

Table 1: Based on the PDF's in Figure 2, it seems the filter criteria for AB is much tighter than what was selected for DR and CR. Why? Shouldn't they be similar?

Page 9, paragraph 1: "Mahalanobis distance can be used for correlated variables…" This seems to imply that you expect AB, DR and CR to be correlated, but can you make it clearer if they are or not? If they are (and I suspect this is the case), this has implications for the uncertainty analysis later.

Figure 3: (a) xaxis title should have caps (b) what is NCE? And more generally how do we know from these figures that the ideal # of classes is 3 or 4 and corresponding fuzzy exponents 1.4 or 1.6. It is not clear what I should be looking for in these plots.

Page 11, second paragraph: mentioned before, but again: is it essential to compare against both V3 and V4? It seems to be unnecessarily complicated.

Page 11, line 17 "…around 74S are misclassified…" confused by this statement since the lowest latitude on the figure is 71.44S

Page 13, line 15 "Note the value (1-c1)x100 for the 2-class FKM algorithm is equivalent to the CADFKM score" isn't it equivalent to the absolute value of the CADFKM score?

Page 14, line 6 "It is evident..." While I agree they do look this way, there are much more rigorous comparison 'statistic' out there than eyeballing a figure.

Figure 7: it appears there is a big difference at high latitudes between FKM and CAD for clouds. Considering that CAD uses latitude, isn't this a very significant difference that should be highlighted and discussed further?

Table 2: What should I be looking for in the C and A columns and rows? This would be more meaningful if I knew what the actual expectations for percentage C and A should be.

Figures 4, 8, 9, 10, 11 would be easier to understand if the colorbar/label at the bottom indicated which side was 'aerosol like' and which was 'cloud like'.

Page 19, lines 12-13. This is an important point that must be emphasized! Use of altitude leads to mis-classification!

Page 21, line 6: this is the first time we hear that a reason for the use of V4 is improved calibration coefficients. Version differences for CAD should be described in more detail in an earlier section.

Figure 10 caption: spell out PSC, STS and NAT acronyms

Page 23, line 14: this is a confusing statement – you're using FKM to rebuild PDFs?

Page 23, line 20: under what circumstances does classification degrade with the addition of additional dimensions?

Figure 11: What is HH? I'm assuming altitude, but you use H elsewhere for that. It's probably best to spell out AB, DR, CR, HH in caption.

Page 25: I really don't understand why you are avoiding high latitude and altitude regions. This is where FKM could presumably help, and differences with CAD might indicate problems with the latter's use of altitude and latitude as dimensions. Or, it might indicate that FKM without altitude and latitude still can't resolve well, in which case those dimensions are needed with CAD. This seems like a in important issue you're sidestepping.

Figure 12: How can you expect anybody to understand this figure? After a bit of effort, I think I understand what you're trying to show, but even if I am correct there's no way for me to differentiate the 15+ different colors. What should this figure look like in a perfect case? I think this figure and the corresponding text are an example of something that should be cut so more focus can be given to other sections.

Equation 16: It would be nice if you said a sentence about what this matrix should look like (ie square pxp matrix that is I in a perfect case)

Equation 18: What kind of norm is this?

Table 3. I'm having difficulty interpreting this. We want low Wilks' lambda for best classification, right? So lowest values are for backscatter alone for 2 class, and backscatter and depol for 3 class. So, does higher lambda when adding other parameters mean classification become worse? Or is it that wilks lambda can't be compared when the dimensionality is different? This is an example of an analysis that seems lacking in its description of what you are looking for, and what the results mean.

Section 4.3: why not just do PCA on the input parameters?

Figure 13: I'm confused by the distribution here. Are we to understand that aerosols are a class completely (or mostly) surrounded by other classes? Also, why these axis ranges, at least make PCA(2) from -4 to 4 so we can see what is going on.

Page 30, line 11: so, you're using one second of observations for the error propagation? I understand the computational limitations but this seems exceedingly limited.

Section 4.4 Aren't your dimensions correlated to some extent, such that you should expect uncertainties to be related (correlated) as well?

Equation 19: What kind of norm?

Figure 14: why not do this with actual CALIOP uncertainties, since these have been assessed? Also, why not use all three uncertainties simultaneously, like the real world?

Figure 15: why is the 2 class case the only one investigated this way?

Page 33, like 28: "While the two algorithms use largely identical inputs. . ." I heartily disagree with this statement. CAD uses altitude and latitude, which your analysis have shown to be important (even if you don't emphasize as much as I would like).

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2018-168, 2018.