

Response to Reviewers 1:

We thank reviewer for taking his/her time to carefully read the manuscript and for the insightful comments which helped us to add significant information and improve discussion of the algorithm description.

Comment:

Key points:

1. The new screening algorithm revolves primarily around First Differences (FD) of the 3.3° measurement time series but it is unclear exactly what is being differenced. Most crucially, it is not clear if the data is detrended via a polynomial fit before FD is calculated.

Answer: CCS measurements at 3.3° scattering angle were differenced in time with no previous detrending. Discussion is added at lines 220-221:

“ It should be noted that no preliminary detrending using a polynomial fit was used.”

2. The FD part of the algorithm is described in Section 4 "Algorithm Description" but the AE threshold aspect of the procedure is not mentioned until the following section entitled "Algorithm applications at selected AERONET sites". So that a clear, concise summary of the method is easily accessible, I would recommend describing the algorithm fully within a single section.

Answer: the information on selection of the TRS threshold (standard deviation value estimated for cloudless days) was added to Section 4 with reference to Section 5 where the detailed discussion is presented. Discussion is added at line 247-248:

“The threshold for TRS was selected from comparison of FD algorithm results to MPLNET cloud detection and found to be 3.0, as described in Section 6.”

3. The iterative scheme (remove largest magnitude in each FD pair, recompute STD, repeat) could be sensitive to the order of removal and to gaps it creates in the time series. Once you remove a point, the next FD pair connects what were previously non-adjacent measurements. If FD is calculated in a way to avoid these introduce artifacts the approach should be described. If not, the possible influence of these artifacts needs to be further explored.

Answer: If during iterations, FD connects measurements separated by temporal gaps, and the FD value is larger than TRS, the algorithm will then remove the AOD corresponding to CCS measurement with the largest magnitude. These situations are conceptually similar to the cases of temporally adjusted measurements and should not create any additional problems. Discussion is added at lines 234-237:

“If during iterations, FD connects measurements separated by temporal gaps, and the FD value is larger than TRS, the algorithm will then remove the AOD value corresponding to the CCS measurement with the largest magnitude. These situations are conceptually similar to the cases of temporally adjusted measurements and should not create any additional problems.”

4. The writing needs significant editing. There are numerous grammatical issues ("themselves" for "themselves" and "loaden" for "laden"). "Zenobo.org" should be "Zenodo.org" in the Yang et al. reference. The paper is also quite long for the amount of new content — the review of the V3 curvature algorithm (Section 2) could be substantially trimmed since it's already published in Giles et al. 2019.

Answer: We have reviewed and corrected grammatical errors. As for the Giles 2019 paper, the curvature algorithm description there is very formal and somewhat lacking in physical/mathematical explanation. We therefore felt that more physically based explanation is needed and will be useful for the scientific community.

5. The paper has 34 figures but in many cases they seem to convey redundant information. Many of these could be merged into single figures making the manuscript much easier to follow.

Answer: We reorganized and replotted some figures. As a result, the total number of figures was reduced to 26.

Specific Comments:

In 115: The explanation of curvature is helpful here but could be made clearer and more concise. I'm particularly struggling to follow the parentheticals in this and the following sentence.

Answer: we added additional discussion which we hope will bring better understanding of the curvature concept: lines 118-201.

“It can be better understood using the radius of curvature which is the reciprocal of the curvature and equals to the radius of the circular arc which best approximates the curve at this point (Wikipedia). In this interpretation, for the straight line the radius of the curvature is an infinity, and the curvature is equal to zero.”

In 141: The exact variables should be stated explicitly here (the magnitude of the curvature decreases?), rather than describing the peak as "flatter" which is ambiguous.

Answer: the correction is added to lines 142-145.

“ It can be seen that the **magnitude of the slope of the curvature is increasing with increasing cloud fraction but still stands below the curvature slope threshold of 4.55 for dust. Another change is the decreasing of **the slope of the curvature magnitude** at the smallest scattering angle which is the consequence **of the forward peak being shifted into the narrower range of the scattering angles** as cloud fraction increases”.**

In 142 Figure 5: Is "change" here meant to refer to change with respect to cloud fraction? The seemingly redundant "as [a] function of cloud fraction" later in the sentence confuses the meaning, especially in the context of so many other derivatives and slopes (e.g., W.R.T. scattering angle) discussed in this section.

Answer: correction is added at line 145.

“Figure 2b shows the change in the magnitude of the curvature at the smallest scattering angle **with** cloud fraction”

In 143/Figure 5: Technically this is a function of case number which is not monotonic with cloud fraction.

Answer: yes, we agree. The correction at line 145 makes clear that the dependence is not on the cloud fraction, rather it shows the change with cloud fraction. The y axis captions define the dependence as a function of case numbers.

In ~175: This is very difficult to follow across four separate figures. Given that the x-axes are identical, could they all be merged into a single figure?

Answer: We merge the figures: Figures 3 (a) and (b).

Figures 12, 13, 15, 16: The label the y-axis as "CCS, 3.3°", but lack physical units. The text indicates these are sky radiances at 1020 nm but the appropriate radiance units need to be provided.

Answer: Units are added to these figures.

Figure 12: The prior plots used time while this plot and the ones after it use day fraction. It would be easier on the reader to stick with one x-axis scheme throughout the manuscript.

Answer: corrected.

In 207: This sentence is a bit contradictory. Is the smooth signature coming from SZA variation's impact on aerosol scattering?

Answer: it means that radiative transfer simulations will produce smooth CCS dependencies as a function of SZA, which we mention in this sentence.

In 208: The caption of the corresponding figure 13 says July 18th, not the 19th.

Answer: corrected.

Section 4: The manuscript introduces TRS as a threshold for both the standard deviation of the FD and the magnitude of the individual FD elements themselves. Are these FD values calculated from the raw 3.3° radiances or

a detrended version of the data using polynomial fits like those shown in Figure 12? If it is the former, which is what the text seems to imply, variability in geometry and aerosol conditions could cause large values of FD without any clouds present. For example, Figure 12(b) shows FD much larger than 3.0 at the beginning and end of the day that are likely driven predominantly by SZA change.

Answer:

Yes, FD are calculated from raw 3.3° radiances.

We agree that there is FD variability in the absence of clouds. It is predominately due to variability in SZA.

However, TRS for cloud free 3.3-degree sky radiances already includes at least in part this variability. We added a new plot (Figure 7) to the manuscript, and a new paragraph (lines 239-249) to illustrate and emphasize the point.

“It should be noted that there is diurnal variability in FD even in cloud free conditions which is predominately due to the variability in the SZA as can be seen from Figure 4 for polynomial fits of CCS measurements versus time. However, TRS for cloud free 3.3-degree sky radiances already includes this SZA variability. Figure 7 shows distributions of FD calculated from CCS measurements at the DEWA AERONET site: 1)18/08/1019 (cloud free conditions, Figure 4b) (red line), 2) polynomial fit to these measurements (blue line), and 3) 04/07/2019 (partial cloudy day, Figure 5) (green line). As can be seen from the plot STDs in cases (1) and (2) are finite due to SZA variability. The spikes at 3.3° measured radiances in case (3) resulted in the shoulders in corresponding FD distribution which exceeded the STD for both cases (1) and (2) and consequently will be removed by the algorithm. The STDs corresponding to the different FD distributions are shown on the plot in corresponding colors. The threshold for TRS was selected from comparison of FD algorithm results to MPLNET cloud detection and found to be 3.0, as described in Section 6. This value, on average, accounts for the natural FD variability in cloud free conditions. “

Ln 225/Figure 14: The text states that if the daily standard deviation (STD) exceeds the clear-sky threshold (TRS), the algorithm marks individual first differences above the static TRS as cloud-contaminated. However, Figure 14 instructs to "Mark all the FD above STD as cloud contaminated," implying the use of the dynamic daily STD. Please clarify in the text and figure which value is actively used to cull the data points.

Answer:

It should be "Mark all the FD above **TRS** as cloud contaminated,". Corrected.

Ln 225: The manuscript introduces TRS as a threshold for both the standard deviation of the First Differences (FD) and the magnitude of the individual FD elements themselves. Applying a dispersion threshold directly to a raw magnitude mathematically assumes the FD distribution is centered exactly at zero. While this is a reasonable assumption for perfectly stable, flat conditions, the mean of the FD will naturally shift away from zero throughout the day due to changing aerosol loading or the continuous variation of sky radiances with the solar zenith angle. Could the authors elaborate on the motivation for using a single absolute threshold to clip both std(FD) and the FD magnitudes themselves? It would be helpful to clarify why a standard outlier detection method centered around the mean (e.g., removing points outside the mean +/- TRS) was not implemented.

Answer:

Due to the symmetry in FD dependence on SZA, the average value will typically be close to zero, as can be seen from Figure 7. It is true even in the c) case where large irregular spikes in 3.3° radiances are present (the mean value is ~ 0.06). Therefore, the mean $\pm$ TRS threshold was not implemented. Discussion is added at lines 249-251.

“Figure 7 also shows that the mean value of the FD diurnal variability is close to zero due to the symmetry in FD diurnal dependence on SZA. It is true even in Figure 7 case (3) where large irregular spikes in 3.3° radiances are present (the mean value is ~ 0.06).”

Ln 275: Regarding the conclusion that the FD algorithm removes cloud-free observations at Capo Verde due to a lack of AE correlation, could this simply be an artifact of the very low baseline AE of coarse-mode dust? Since both dust and cirrus clouds have very low Angstrom exponents, removing cloud contamination at this site might inherently produce negligible shifts in AE, rather than indicating a failure of the algorithm.

Answer: Discussion is added at lines 294-298 to emphasize that in this particular case the algorithm is failing and removes cloud -free data.

“In the dust aerosol case, the aerosol/clouds separation becomes very challenging. In the particular case considered (Figure 14a), the AE during algorithm iteration stays practically constant. In addition, inspection of 3.3° radiances diurnal variability for this day is relatively smooth and does not exhibit any sharp features. Therefore, it was concluded that in this case the algorithm was removing aerosols and was considered algorithm failure.”

Section 5: Building on my Ln 275 comment, the proposed AE regression slope threshold evaluates the absolute rate of change in AE to distinguish between cloud and aerosol removal. However, this absolute slope is highly sensitive to the initial baseline AE. At fine-mode dominated sites like Thule, removing large ice crystals yields a steep absolute change in AE. Conversely, at dust-dominated sites like Capo Verde, the initial AE is already low ; thus, the maximum possible absolute change in AE upon removing cirrus is inherently constrained. A fixed, universal threshold of 0.01 biases the algorithm to systematically reject cloud screening in coarse-mode environments. That said, throttling the cloud mask in coarse-mode environments is practically understandable, as the optical similarity between large dust particles and cirrus ice crystals makes definitive separation physically difficult. If this is the goal, the authors should explicitly state this radiative transfer limitation as the justification for a less aggressive screening approach at these sites, rather than framing the AE threshold solely as an empirical fix. Furthermore, the authors should discuss whether a relative threshold (e.g., normalized by the initial AE) was considered.

Answer:

We explored the AE regression slope after normalization by the first AE value. The results showed that in both cases (Figure 17) slopes increased after normalization but only slightly: $6.5\text{e-}5$ to $2\text{e-}4$ for Capo Verde and 0.061 to 0.075 for Thule. Therefore, we concluded that normalization of AE would not bring significant changes in algorithm performance. Discussion is added at lines 323-327.

“Due to the slope sensitivity to the initial baseline AE, we also explored the option of the AE regression slope after normalization by the first AE value. The results showed that in both cases (Figure 17) slopes increased

after normalization but only slightly: $6.5\text{e-}5$ to $2\text{e-}4$ for Capo Verde and 0.061 to 0.075 for Thule. Therefore, we concluded that normalization of AE would not provide significant improvements in algorithm performance.”

Regarding empirical fix. It is indeed an empirical fix. The rationale for this is the following: one needs to remember that FD algorithm is applied to the results of V3 cloud screening algorithm, which effectively removed the majority of clouds. Therefore, we assumed that situations with low dynamics of AE variability correspond on average to algorithm failure. The empirical fix will restore the wrongly eliminated data. Unfortunately, this is compromise we are making in situations where separation between large particles and cloud is inheritably difficult. Discussion is added at lines: 354-360.

“Employing the slope of AE regression as an additional parameter controlling the removal of the AOD data could be considered as an empirical fix in situations when the removal of the large amount of data is in doubt. The rationale for this is the following: one needs to remember that FD algorithm is applied to the results of V3 cloud screening algorithm, which effectively removed the majority of clouds. Therefore, it is assumed that situations with low dynamics of AE variability correspond on average to algorithm failure. The empirical fix will restore the wrongly eliminated data. This is the compromise which was made in situations where separation between large particles and clouds is inheritably difficult. “

Figure 30: What does the blue track represent on panel (b)?

Answer: **The blue line is the Aqua MODIS satellite track. This has been added to the figure XX caption (new figure number).**

Ln 433: The slope is stated to be 0.1 here but generally described as an order of magnitude lower in Section 5. Please clarify.

Answer: it should be 0.01, corrected.

Response to Reviewers2:

We thank Lorraine Remer for taking her time to carefully read the manuscript and for the insightful comments which helped us to add significant information and improve discussion of the algorithm description.

Discussion point:

1. The final results, as shown in Figure 34 show an increase of True Positives by what? 1% ? Is that important? Figure 33 shows MCC increasing by as much as 0.05 (Barcelona 30°). Is that important? These numbers need to be put into context.

Answer:

The FD algorithm is applied to Level 1.5 of the V3 data, which has been cleared previously by applying the V3 cloud screening algorithm. The FD goal is to detect clouds, mostly cirrus, that were previously missed. The number of these additionally detected clouds is expected to be small and vary geographically. Therefore, the increase in MCC after applying the FD algorithm is important. As long as MCC increases, the FD algorithm performs satisfactorily in removing cloud contaminated observations.

However, we agree that the increase in MCC and its categories should be put in the context of the number of removed observations as well as the changes in the values of AOD and AE.

Increasing the True Positives by 1% means that applying FD detected 1% more clouds than V3 algorithm alone in the analyzed array of collocated AOD/Lidar data. It is important as long as FD detects clouds missed by cloud screening algorithm of V3. The FD algorithm is after that of V3 which has already removed the majority of the clouds. In general, the percentage of additional removal will vary geographically. We added Figure 25 which shows the geographical distribution of the percentage of data points which did not pass the FD test from V3 L1.5 with an average number of about 7.5%. We note that the cirrus screening algorithm in V3 called the curvature algorithm results in 4.5% of data screened globally. Therefore, even small percentages of additional cloud removal globally or regionally are significant especially in the context of field campaigns or investigations involving limited number of days of sites located in regions with high cirrus cloud fractions.

To put it in context, all the four categories contributing to MCC should be analyzed although the two most important are True Positive and False Negative whose variability primarily determining the value of MCC. For example, in the Barcelona case, MCC increase by 0.05 corresponds to the increase of the True Positives by 9.7% and decrease of the False Negatives by 12.8%.

- Is there a better way than the bar graph to show the difference between the new method and Version 3 in the four categories ? At least give the quantitative numbers in the text.

Answer:

We think that plotting relative (to the total number of matches) contribution of the four MMC categories to the total MMC provides good visualization of FD performance relative to the V3 cloud screening algorithm. On average, applying the FD algorithm results in increasing True Positives and decreasing False

Negatives by 1.4 and 1.42 percent respectively. In terms of the relative cloud identification this corresponds to an increase in the number of clouds detected by 350 from 24736 matches with lidar.

- Can we put the final validation results into context. What is the effect of a 1% increase in cloud identification on overall AOD and AE? Does it matter?

Does a 0.05 increase in MCC make a difference? I looked up the formula for MCC, and it is complex. Using MCC to show the scattering angle dependence is useful, but I need some context to evaluate whether a change in MCC of 0.01 or 0.05 is significant. This is especially so when MCC varies between locations by 0.25.

Answer:

The MCC is the statistical way to characterize the difference between two arrays, therefore, how big a difference in AOD and AE corresponds to 0.05 increase in MCC depends on the size of the analyzed data set. If the data set is small, 0.05 MCC change can result in small change in both AOD and AE and vice versa. However, changes in the average AOD and AE are not good metrics for evaluating the FD performance. The FD, being an addition to the V3 cloud screening algorithm, removes relatively small number of AOD measurements to significantly affect the AOD and AE. For example, for Barcelona data set corresponding to the 30 degrees SZA, MCC increase of 0.045 does not correspond to any changes in AE and AOD due to only 6% increase in True Positives.

In general, quantitative changes in MCC should be analyzed in the context of the specific data sets which one compares. Figure 26 shows the geographical distribution of AOD and AE changes due to FD application.

Discussion was added at the end of section 4.

“The net change in MCC can be put in context of both the number of additional removals of AOD measurements by FD algorithm and the net changes in average AOD and AE. In general, the number of additional removals will vary geographically. Figure 25 shows the geographical distribution of the percentage of data points which did not pass the FD test from V3 L1.5 with an average number of about 3.1%. Because the FD algorithm is after that of V3 which has already removed the majority of the clouds, even small percentages of additional cloud removal globally or regionally are significant especially in the context of field campaigns or investigations involving limited number of days of sites located in regions with high cirrus cloud fractions. The magnitude of the net changes in AOD and AE depends on the size of the data set analyzed. However, changes in the average AOD and AE are not good metrics for evaluating the FD performance. The FD, being an addition to the V3 cloud screening algorithm, removes relatively small number of AOD measurements to significantly affect the AOD and AE. For example, for Barcelona data set corresponding to the 30 degrees SZA, MCC increase of 0.045 does not correspond to any changes in AE and AOD due to only 6% increase in TP. Figure 26 shows the geographical distribution of AOD and AE changes due to FD application. The net changes in MCC values should be analyzed considering all the four categories contributing to MCC although the two most important are TP and FN whose variability primarily determines the value of MCC. For example, in the Barcelona case, MCC increase by 0.05 corresponds to the increase of the TP of 9.7% and decrease of the F N by 12.8%.”

Questions:

1. I noticed that none of the chosen validation locations include the environment that caused the issue in the first place, high latitudes with cold clouds and small crystal sizes. This algorithm change might seem less incremental if we could see it working at high latitude. I noticed an active MPL at King George Island in

Antarctica that seems to be collocated with the Escudero AERONET site. I may be wrong, but if you could look at validation in the situation that caused the problem in the first place it would strengthen the significance of the work that you did.

Answer:

We do not have AERONET site at King George Island, therefore no collocated data are available.

2. Lines 164-165. I didn't understand how the threshold on FD was chosen. *"The lower threshold for the slope of curvature was selected at 4.2 and the upper threshold for the first point at $2 * 10^{-5}$."* Shouldn't the threshold line be drawn as to separate cirrus from aerosol? Drawing the slope of curvature at 4.2 puts cirrus and cirrus mixtures of 50/50 on one side of the line and pure aerosol and less cirrus in the mixture on the other side of the line for Figures 4 and 6 but allows dust to be characterized as cirrus in Figure 3. For FD, a threshold of $2 * 10^{-5}$ makes no sense to me at all because ALL values of FD lie above that line in Figures 5 and 7. All values, from cirrus up to pure aerosol.

Answer:

the sentence . *"The lower threshold for the slope of curvature was selected at 4.2 and the upper threshold for the first point at $2 * 10^{-5}$."* This refers to the thresholds for the curvature part of the V3 algorithm and corresponding plots are presented in Giles, 2019 paper.

3. Lines 274-275. *"As can be seen, there is no obvious correlation between changes in AE and the reduction in the number of measurements meaning that the FD algorithm removes cloud free observations at this site."* What if the FD algorithm was CORRECTLY removing ONLY cloud contamination, cloud and dust AE were similar. You wouldn't see a relationship between removed contamination and changes in AE. You only see the correlation between changes in AE and reduction in number of measurements when the AE of the cloud and aerosol are very different. AE may not be so different with dust.

Answer:

we agree that dust/clouds separation is very challenging, however we found evidence that in the Capo Verde case, the FD algorithm really removes aerosols. Discussion is added at lines 294-298 to emphasize that in this particular case the algorithm is failing and removes cloud -free data.

"In the dust aerosol case, the aerosol/clouds separation becomes very challenging. In the particular case considered (Figure 14a), the AE during algorithm iteration stays practically constant. In addition, inspection of 3.3° radiances diurnal variability for this day is relatively smooth and does not exhibit any sharp features. Therefore, it was concluded that in this case the algorithm was removing aerosols and was considered an algorithm failure."

4. Figures 23, 24. Lines 281-283 and Line 285. I had overall difficulty in following the main points in this paragraph. I had difficulty in understanding the histograms, especially the blue line in the graphs and what the ordered pairs (6,3) and (23,16) etc. mean. The caption needs to be much more descriptive. This statement, *"The AE difference variability within the peak is ~ 0.01 while the total reduction in the number of measurements is about 60% which is obtained by summing the histogram points corresponding to the relative decrease in the AOD measurements number within AE difference peak."* didn't make sense to me either. Finally, when it gets down to *"-0.006, -0.0045, -0.01"* this progression is not monotonically decreasing with iteration, I'm sure I'm missing something.

Answer:

These figures show dynamics in changing AE and the number of observations with iterations. Current captions tell that the blue line is the relative change in the number of AOD observations. The goal is to show how the dynamics is different for dust dominated or cloud contaminated sites and the sites dominated by fine mode aerosols. The sentence , *“The AE difference variability within the peak is ~ 0.01 while the total reduction in the number of measurements is about 60% which is obtained by summing the histogram points corresponding to the relative decrease in the AOD measurements number within AE difference peak.”* gives quantitative description of the graph relating the changes in AE and the relative number of measurements. The first number in the ordered pair is the number of AOD measurements before the iteration and the second number is the one after. We also agree that the progression “-0.006, -0.0045, -0.01” is not monotonic, but we do not see the reason it has to be. We added changes in the captions to the corresponding figures.

“The first number in the ordered pair in parentheses is the number of AOD measurements before the iteration and the second number is the one after.”

5. Figure 26. What is the green line? As I understand it, “regression slope” mentioned in Line 301 refers to the regression slope as plotted in Figure 25, resulting from the relationship between AE and iteration. Then how in Figure 26 you can form histograms of this slope as a function of iteration. The independent variable in the regression line IS the iteration.

Answer:

The green line corresponds to the FD threshold selected for the regression slope. The histograms were formed for the whole year of data at corresponding sites. For the whole dataset all the regression slopes after each iteration were combined to form a histogram. We agree that the regression slope is the function of the iteration, therefore we have separate plots for each iteration.

6. Lines 311- Be a bit careful. This threshold of 0.01 refers to a different separation threshold than the ones discussed for slope of curvature and FD back in Lines 164-165. This one is also a threshold on a slope, but this time it is the slope of AE vs iteration. There are two thresholds on “slope”. I did get confused.

Answer:

We agree that it can get confused. We name the slopes as the curvature slope and AE slope. The following changes were added.

“In both cases the FD algorithm was applied without imposing the **AE** regression slope criterion to demonstrate the FD algorithm performance when the **AE** regression slope is above and below the 0.01 threshold.”

“For this case, the **AE** regression slope was estimated to be ~ 0.018 which is above the suggested threshold of 0.01 suggesting some cloud presence.”

7. Lines 347- “*the threshold value above which the clear distribution is dominant.*” Isn’t it BELOW which the clear distribution is dominant. Throughout the discussions on the many thresholds introduced in this manuscripts, the use of “above” and “below” a threshold should be clarified using a single perspective. Are we trying to find CLOUDS or are we trying to find NONClouds? That should determine whether threshold should be consistently addressed as “above” or “below” a threshold. Here in Lines 347-348 it is explicitly stated that the clear distribution is targeted, and that is below the threshold, as I understand it.

Answer:

We agree it should be below. It has been corrected.

8. Lines 446- “*the FD algorithm outperformed V3 L1.5 by about 0.02 in MCC*”. The FD algorithm is an add-on to the V3 L1.5 algorithm. It is applied after the standard algorithm identifies 90% of the clouds, right? I think the statement here sounds as if it is an “either/or” situation when it is really “with or without” situation.

Answer: we agree. The sentence was corrected:

“The comparison showed that, on average, the **addition of** FD algorithm outperformed V3 L1.5 **alone** by about 0.02 in MCC, suggesting consistent improvement in overall cloud detection accuracy “

And now for a list of minor grammatical issues. Usually just missing “a” or “the” or needing a plural.

Answer:

Everything was corrected, thank you.