COST short term scientific mission report

Verena Schenzinger
with: Stelios Kazadzis, Natalia Kouremeti, Akriti Masoom

April 11, 2023

1 Introduction

Aerosol Optical Depth (AOD) can be measured with various techniques, from ground-based instruments via in-situ measurements to satellite observations. Each of these methods has their underlying assumptions, and therefore uncertainties. Of all the possibilities, ground-based sun photometry provides total column AOD measurements with the lowest uncertainty, which is why it serves as the ground truth in satellite observations and for validation of other measurements.
To ensure the high quality of the data, one crucial step is to identify AOD variations that are due to clouds, while keeping signals that are due to changes in aerosols, like dust or volcanic ash. Ideally this quality control can be done fully automated, and in a consistent manner for different instruments and measurement sites. While AERONET recently changed their data processing to not require manual checking anymore [Giles et al., 2018], their algorithm cannot be used for precision filter radiometers (PFRs) used e.g. in the GAW-PFR network, which do not provide one of the necessary measurements (aureole scan). Furthermore, the measurement frequency of PFR is much higher than for the CIMEL instruments used in AERONET, meaning a lot of the parameters have to be adapted.
In general, cloud flagging algorithms are based on the change of the radiation field. This can be achieved by limiting the variation of AOD within consecutive measurements (triplet criterion, Smirnov et al. [2000]) or controlling for change of direct irradiance with air mass compared with a cloudless sky model. Changes in the Angstrom exponent α, which is very sensitive in any direct sun changes, are usually taken into account as well, however at low AOD this parameter gets noisy as the relative uncertainty gets larger. This means that in situations with rapid changes of the radiation field due to aerosols (e.g. dust events) or little change in the presence of very thin cirrus clouds, accurate flagging can be difficult for any algorithm.

Schenzinger and Kreuter [2021] suggested an alternative algorithm to identify clouds in high frequency (1 minute) PFR measurements based on a nearest-neighbour clustering method. This approach takes into account the AOD at 500nm, its temporal variation, as well as the Angstrom exponents α and γ (similar to $\delta\alpha$ used in Gobbi et al. [2007]). As their algorithm seemed promising for cloud flagging PFR data in general, this short term scientific mission expanded their analysis to the stations in Davos and Izaña which have different aerosol conditions than Innsbruck, where it was originally tested.

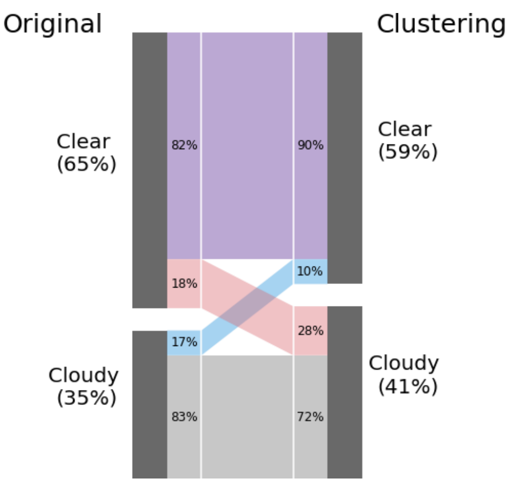
2 Data and methods

We used one full year (2019) of AOD measurements from Davos (PFR and CIMEL) and Izaña (PFR) measurement stations. The data was already quality controlled, including cloud flagging, which is what will be referred to as "original" flags for the rest of this report. This data was analysed with the clustering algorithm described in Schenzinger and Kreuter [2021] as well (using the level 1.0 data from CIMEL as starting point), resulting in the "clustering" cloudflag. We compared the two types of cloud flagging quantitatively, as well as checking the implications of using either on daily means and yearly statistics. Furthermore, we took a closer look at the days where the disagreement between the original and the clustering cloudflag were most pronounced.

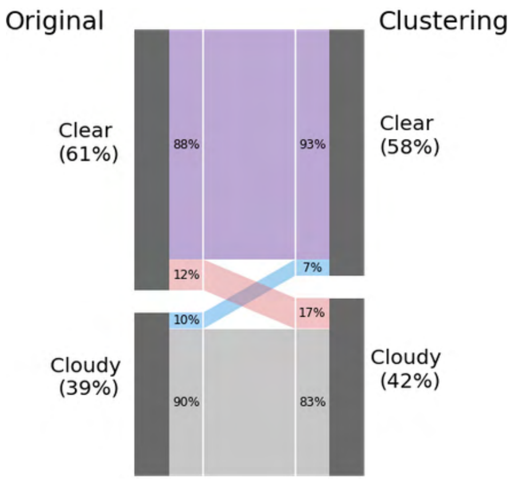
3 Results

3.1 Davos

In Davos, the timeseries of 2019 of two different instruments were analysed: a CIMEL operating as part of AERONET, and a PFR. As the CIMEL data does not explicitely have information about cloud/nocloud, the cloudflag was inferred from comparing level 1.0 and level 2.0 data and regarding datapoints that do not pass the quality assessment from 1.0 to 2.0 as cloudy. Note that this is not data points not passing the quality assurance for any reason are marked as cloudy in this process, which can be one reason for differences between the original and the clustering cloudflag. For the CIMEL, which measures with a frequency of 15 minutes, we have a dataset of 7.3k measurements. Of those, the original quality assurance deems 65% as clear, and 35% as cloudy (see figure 1(a)). Compared to that, the clustering algorithm (which was used with a limit value of 0.042 due to the lower measurement frequency of the CIMEL), flags some more points as cloudy, resulting in a split of 59% clear and 41% cloudy.
A similar analysis was done for the PFR (88k measurements), for which the cloud



(a) CIMEL



(b) PFR

Figure 1: Comparison of the amount of datapoints which get flagged by the original and clustering algorithm (dark grey bars). The height of the bars is proportional to the number of points in the category. The grey area signifies where both algorithms agree on the point being cloudy, and the purple where both agree on the points being clear. Red and blue show differences in flagging, with the percentages referring to the partition of the cloudy/clear points of the respective algorithm.

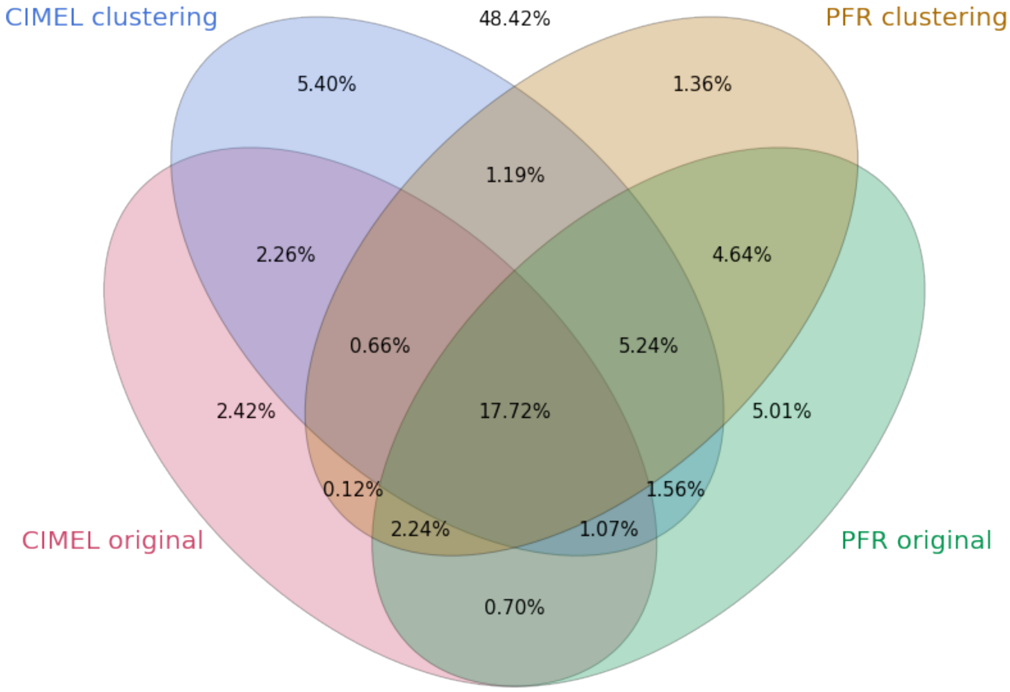


Figure 2: Percentage of points which get flagged by the original algorithm and the clustering algorithm for the CIMEL and the PFR instruments. The proportion of cloudfree points, which both algorithms and instruments agree on, is indicated on top outside the ellipses.
Beware that only coincidal points in the CIMEL and the PFR are included here, i.e. the percentages in the ellipses do not necessarily add up to the numbers in figure 1.

information is recorded separately. The same parameters as in Schenzinger and Kreuter [2021] were used for the clustering algorithm, resulting in a similar overall split of clear and cloudy points as for the CIMEL (58%/42%). Clustering still flags slightly more than the original algorithm, however the split into clear and cloudy is not that different (61%/39%). Compared to the original CIMEL dataset, the amount of cloudy points is higher, however from the quantitative analysis alone it is not evident which instrument is better in the assessment. Given that the measurement frequency differs (15min vs. 1min), and some types of clouds are short-lived, actually both might be correct for their assessment.
Some of the differences in cloudflagging for the CIMEL could be due to its lower measurement frequency compared to the PFR, for which the clustering algorithm was originally developed. Figure 3 exemplifies this with subsampling the timeseries from the original 1 minute to 5 minute and 10 minute frequency, running the clustering algorithm with identical parameters, and comparing the cloudflag for points which coincide in all 3 (i.e. 10min resolution overall, 8.8k points in total). While nearly all points which are flagged at a 1 minute resolution remain flagged at the lower resolutions as well, a lot of additional points get flagged when subsampling to 5 and 10 minute resolution. This leads to the amount of cloudy points going from 38.5% at one minute resolution to 59.7% in 10min resolution, which is not realistic for Davos. Furthermore, of the nearly 12% of points that just get flagged in the 10min resolution, only 15.7% would be regarded as cloudy in the original algorithm, which means that most of them are likely misclassified.
The reason for increased flagging lies in the lower overall density of the points, as outlined in Schenzinger and Kreuter [2021], and can be counteracted to some extent by applying a higher limit in the cloudscreening algorithm. However, there might be further adaptations needed, e.g. regarding the weight of the parameters.

Figure 2 summarizes the two panels of figure 1 into one Venn diagram, using datapoints that appear in both the CIMEL and the PFR instruments: the numbers in the coloured ellipses indicate the amount of points identified as cloudy by the respective instrument and algorithm. The number on top, i.e. 48.42%, represents the number of datapoints which for each instrument and algorithm pass the quality assurance, including cloud flagging. Similarly, 17.72% of the measurements, are deemed cloudy in all scenarios, which means that there is more agreement than disagreement between instruments and algorithms.
While this is an encouraging result, we still did not look further into a comparison of the CIMEL and PFR data, especially at the ∼35% of data where there are differences between the algorithms or instruments, as some of them might be due to the nature of measurement and quality assurance itself, rather than specific for clouds. However, it highlights that consistent data recording even in terms of

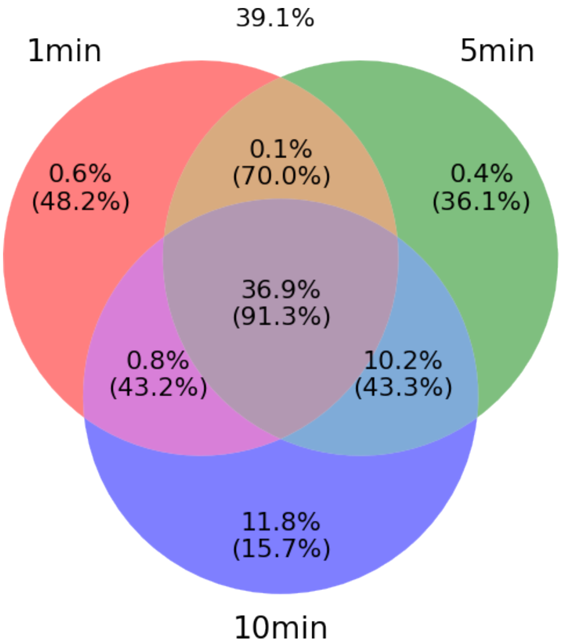


Figure 3: Percentage of all points (8.8k) which get flagged when using PFR data with the indicated time resolution (red:1min, green: 5min, blue: 10min). The percentage of points which are identified as clear regardless of time resolution is indicated on top outside the circles. The percentage in brackets refers to the fraction of the points in the respective category which would also be considered cloudy by the original algorithm.

quality assurance criteria can be beneficial for intercomparisons of instruments. Therefore, we focus on the performance of the clustering algorithm for PFR data. Figure 4 shows the timeseries of agreement/disagreement of the clustering and original cloudflag for days which had at least 10 clear datapoints. As in figure 1, the grey and purple bars highlight agreement on cloud/clear, whereas blue and red bars show disagreement. Overall, the two algorithms only disagree on $\sim$11% of the datapoints, and there are no days where the disagreement is higher than the agreement (indicated by the ratio being below 1 for the whole year). Regarding daily mean AOD of cloudfree data, these differences are next to negligible: There are very few days where a difference can be seen in the timeseries, and it is not necessarily on the days where the disagreement is comparably high. It is somewhat more noticeable when looking at the Angstrom exponent α, but hardly significant. Often, the differences in Angstrom exponent are bigger when the AOD is low, i.e. when the error α itself is high. When looking at the distributions of these two parameters of the respective cloudfree datapoints (see figure 5), the median AOD at 500nm gets slightly lowered and the distribution gets a bit narrower, whereas for α, only a slightly lower median is evident.

3.2 Izaña

The same analysis for the Davos PFR was done for the PFR in Izaña, which is a good testing place for a cloudflagging algorithm as it experiences dust events regularly, which should not be mistaken as clouds.

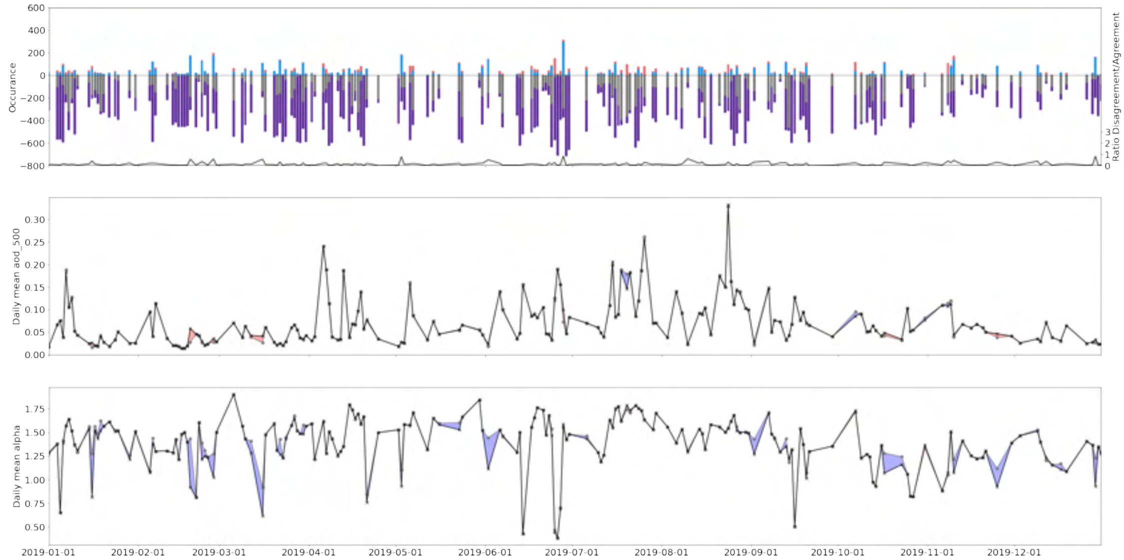


Figure 4: Timeseries of all days in 2019 with at least 10 clear datapoints for Davos. Top: Number of points where the algorithms disagree (above the zero line) and agree (below zero line). The colours are identical to those in figure 1. The black line indicates the ratio of disagreement/agreement.
Middle/Bottom: Timeseries of daily mean AOD/α at 500nm using the original cloudflag (grey, with circles) and the clustering cloudflag (black, with crosses). The coloured areas highlight the differences between the algorithm; red area: original is lower, blue area: clustering is lower.

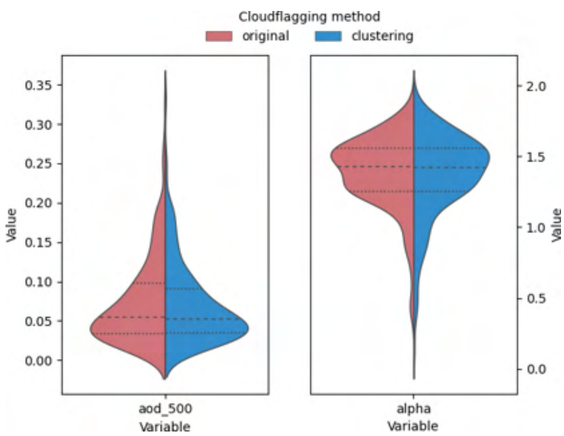


Figure 5: Distribution (kernel density estimate) of values for AOD at 500nm and α identified as cloudfree by the original (red) and clustering (blue) algorithm. The solid line indicates the median, the dotted lines the 25th and 75th percentile of the distrubution.

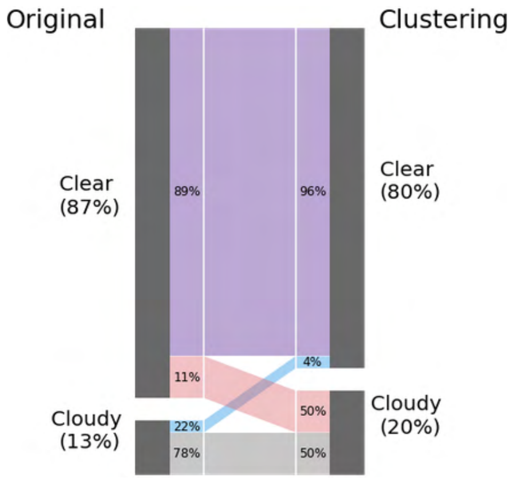


Figure 6: As figure 1, but for PFR data from Izaña.

Figure 6 shows the amount of clear and cloudy points for each of the cloudscreening methods. Overall, Izaña has more clear data than Davos, regardless of the algorithm, and the agreement is still high (about 87% of the data). However, half of the points which are identified as cloudy by the clustering algorithm would pass the original quality assurance, which could point to possible misidentification of dust. The timeseries (figure 7) reveals that there are only a few days with comparatively large disagreements, and even these do not have a big influence on the respective daily means. The distribution of AOD is basically unaffected by the change of the algorithm as well albeit its median is slightly higher, whereas the distribution of α gets shifted a bit to lower values, which would, however, indicate that thin clouds do not get appropriately flagged instead of dust being mistaken as cloud.
Overall the agreement between the algorithms in Izaña is very good, however this is partly due to the location being above the clouds most of the time, therefore experiencing cloudless skies, and the agreement being mainly about the clear sky (see figure 6).

3.3 Case studies

While the overall agreement between both cloud flagging algorithms is good, there are a few cases when the daily mean AOD differs by more than 0.01 (as comparison: the mean daily mean is 0.071, the median daily mean 0.055 in Davos and 0.044/0.027 for Izaña). Figures 9 and 10 show the timeseries of the measurements on two of the six days in Davos, including images from the sky camera next to the instrument.
While on 2019/07/20 (figure 9) the daily mean AOD was lower when using the clustering algorithm, on 2019/02/19 (figure 10), it was the other way round. The sky camera give some insight into both days: While it was sunny with some cirrus

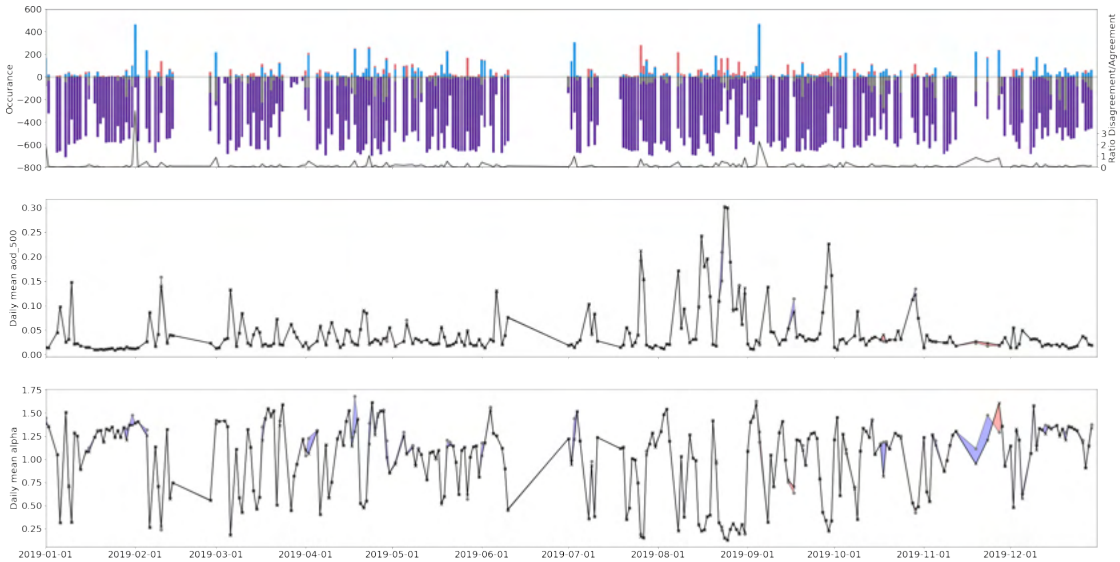


Figure 7: As figure 4, but for PFR data from Izaña.

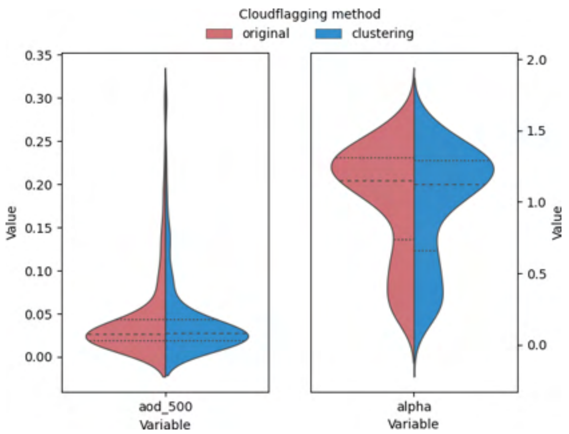
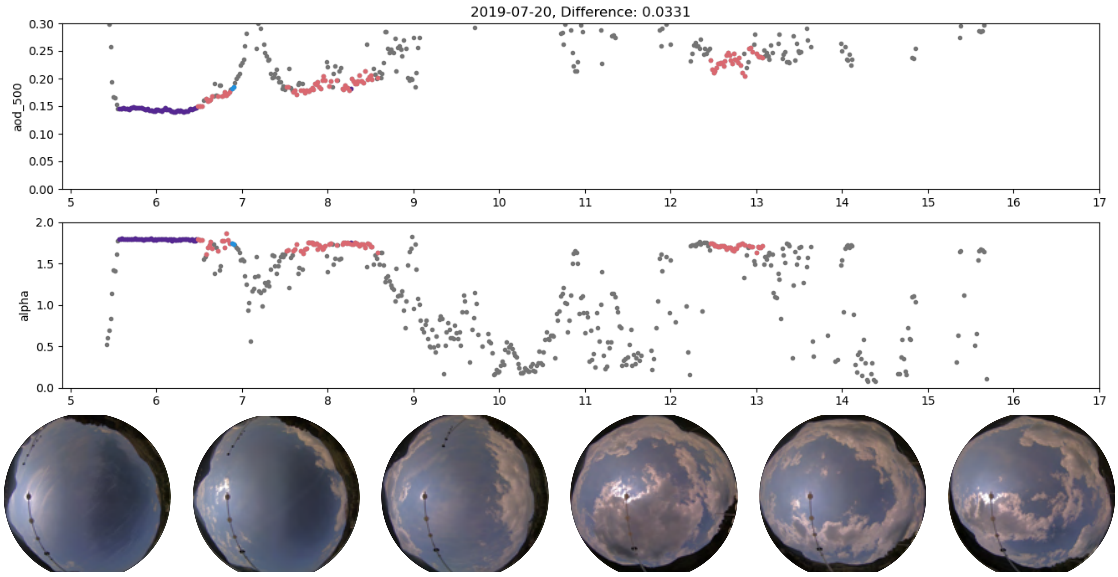


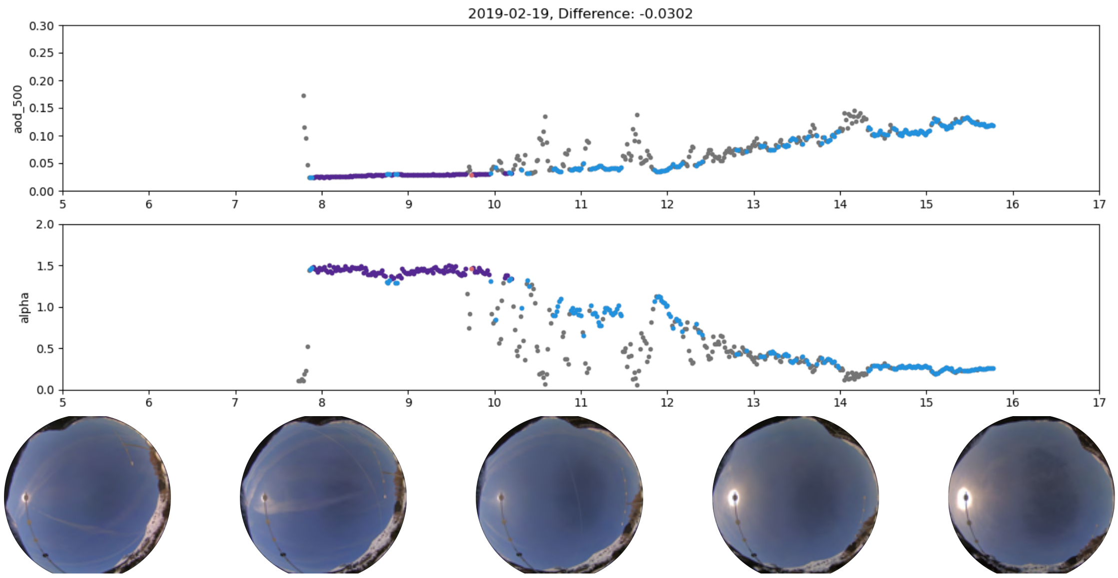
Figure 8: As figure 5, but for PFR data from Izaña.

in the morning of 2019/07/20, thicker broken clouds moved in during midday. Around 7am, both algorithms identify the thin clouds as such, as well as the thick clouds at midday. The cloud gap around 13:00 however only gets marked as clear by the original algorithm, and could still be affected by thin clouds; however, the camera pictures are inconclusive. Overall, this day is characterised by very high AOD (compare the distribution in figure 5), which gets even more exaggerated by the original cloudflagging.
In contrast to that, on 2019/02/19 the daily mean AOD is low, and some of the thin clouds in the afternoon do not correctly get picked up by the clustering algorithm, while the change in AOD and α are attributed to the clouds in the original algorithm and thus flagged. Unfortunately even in sum these changes seem to be below the threshold of detection for the clustering algorithm, which shows room for improving it, maybe by using relative AOD change instead of absolute (i.e. $\frac{\Delta\mathrm{AOD}}{\mathrm{AOD}\Delta t}$ instead of $\frac{\Delta\mathrm{AOD}}{\Delta t}$) to take into account the low AOD, or introducing a similar variable for Angstrom α as well.
Unfortunately, the webcam availability in Izaña limited this kind of analysis there. Figure 11 also shows the timeseries of AOD and α for one day, and corresponding sky camera pictures for the morning. Those pictures indicate that some thin clouds might have been mistaken as clear sky by the original algorithm, leading to an overall higher mean AOD (0.114 compared to 0.088).
What makes Izaña an interesting location to test a cloud screening algorithm is the occurance of dust events, i.e. periods of high and variable AOD compared to the mean at the station. Compared to clouds, however, the variation in the Angstrom α is not as high, giving the possibility to distinguish dust from clouds. Figure 12 shows the timeseries of the AOD and α measurements, as well as the assessment on each of both the cloudflagging algorithms. While they do not agree on every single point (overall 71.2% agreement), the daily mean of the cloudfree points does not differ significantly: 0.298 for the original, 0.300 for the clustering algorithm. Figure 13 illustrates the aerosol conditions a bit further in an α-γ-diagram (similar to the ones suggested by Gobbi et al. [2007]). With average particle diameter between 0.15μm and 0.2μm, and a fine mode fraction between 10% and 30%, these measurements strongly indicate a dust event, which gets correctly identified as non-cloud by both algorithms. As a confirmation, figure 14 shows the HySplit back trajectories for 25th of August, confirming Sahara dust transport to Izaña.



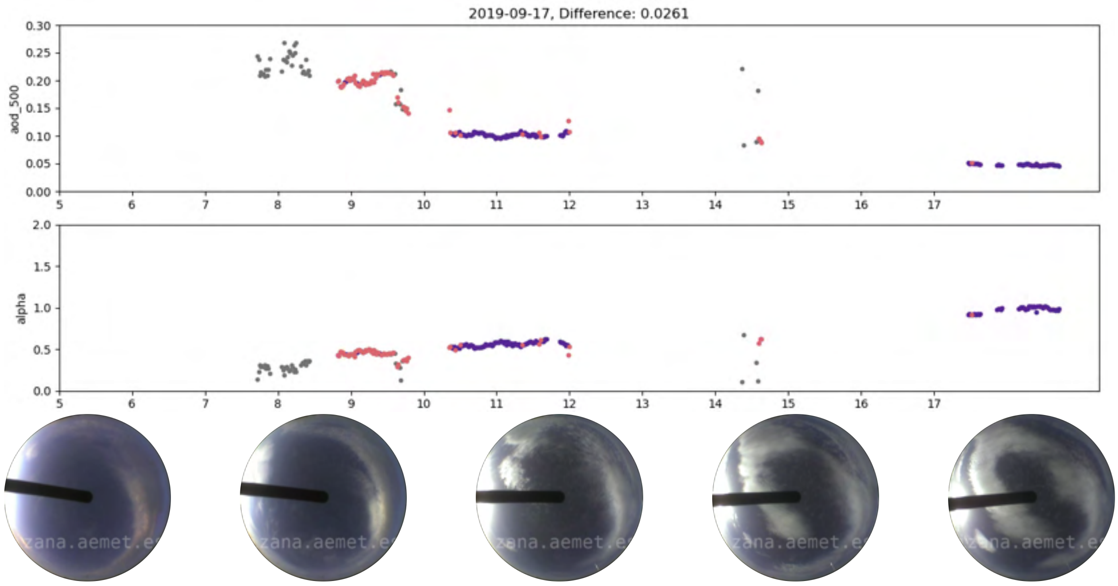
(a) 07:00 (b) 08:00 (c) 09:00 (d) 12:00 (e) 13:00 (f) 14:00

Figure 9: Timeseries of AOD at 500nm and Angstrom α for 2019/07/20. Like in figure 4, purple/grey indicates points both algorithms see as clear/cloudy whereas blue/red means they are clear only for the clustering/original algorithm. The sky camera pictures illustrate the cloud conditions on this day (note that they are not lined up with the time on the plot).



(a) 10:00 (b) 11:00 (c) 12:00 (d) 13:00 (e) 14:00

Figure 10: Like figure 9, but for 2019/02/19.



(a) 07:00 (b) 07:20 (c) 08:00 (d) 08:30 (e) 08:50

Figure 11: Like figure 9, but for 2019/09/17 in Izaña.

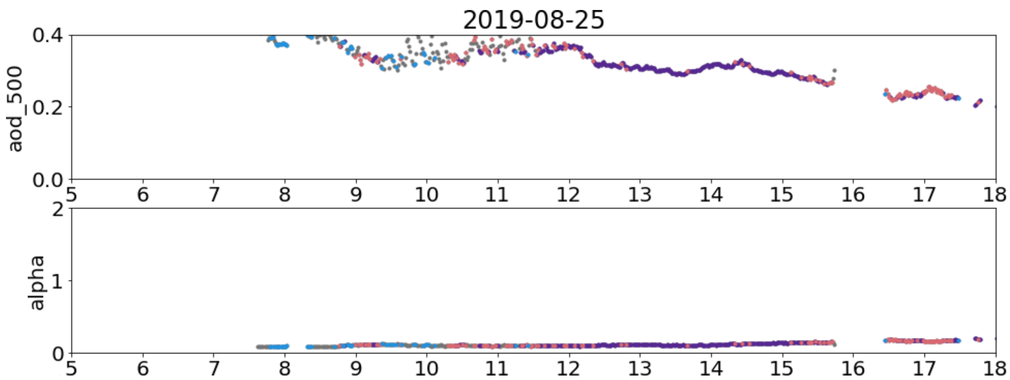


Figure 12: Like figure 9, but for 2019/08/25 in Izaña.

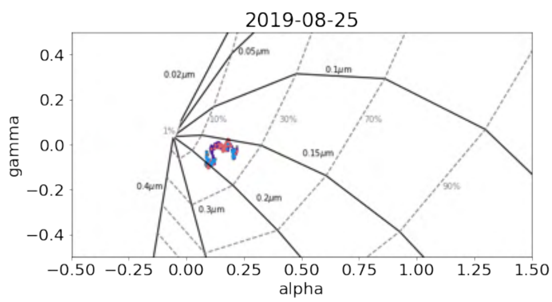


Figure 13: Angstrom parameters α vs. γ for 2019/08/25 in Izaña, colouring identical to 12. The lines indicate different particle size (solid black) and fine mode fraction (dashed grey).

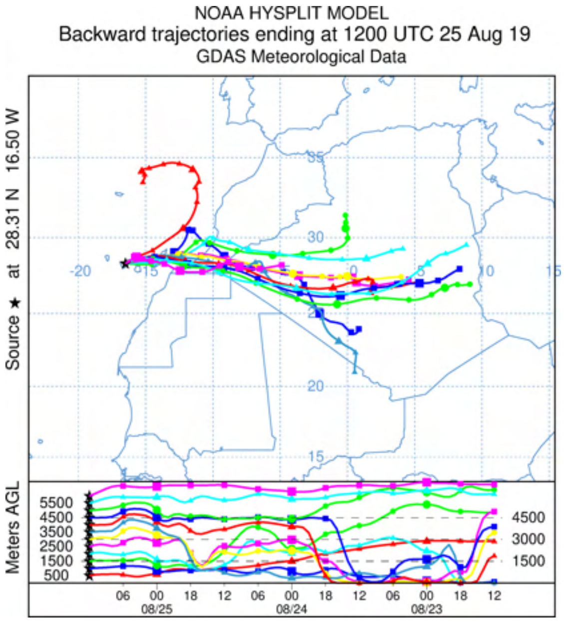


Figure 14: Backward trajectories for 2019/08/25, showing Sahara dust has been transported to the station in Izaña from the Sahara.

4 Conclusions

Overall, the clustering cloud screening algorithm performs very well, even in different environments than originally tested. While Davos arguably still has similar aerosol conditions to Innsbruck, albeit with lower AOD overall, Izaña with its dust events can be challenging. In the overall distribution of AOD values, little difference between the original and the new algorithm can be identified, which is positive given that the PFR data does get automatically and manually checked, whereas clustering is a purely automated algorithm.
Still, figure 10 highlights that in low AOD conditions, thin clouds do not reliably get identified, giving hints as towards improvement potential. Similarly, the algorithm cannot yet be implemented to CIMEL instruments as their lower measurement frequency makes adaptation necessary (see figure 3).
The clustering algorithm so far tends to flag more points than the original algorithm, which is unproblematic for PFR instruments as their high measurement frequency means that false positives (i.e. clear sky that gets mistaken as cloudy) do not influence the data quality as negatively as false negatives (unflagged clouds) would. However, for less frequent measurements, especially at high altitude sites with low solar elevation, a more careful assessment is needed to not throw away data unnecessarily.

References

David Giles, Aliaksandr Sinyuk, Mikhail Sorokin, Joel Schafer, Alexander Smirnov, Ilya Slutsker, Thomas Eck, Brent Holben, Jasper Lewis, James Campbell, Ellsworth Welton, Sergey Korkin, and Alexei Lyapustin. Advancements in the aerosol robotic network (AERONET) version 3 database – automated near real-time quality control algorithm with improved cloud screening for sun photometer aerosol optical depth (AOD) measurements. *Atmospheric Measurement Techniques Discussions*, pages 1–78, 09 2018. doi: 10.5194/amt-2018-272.

G. P. Gobbi, Y. J. Kaufman, I. Koren, and T. F. Eck. Classification of aerosol properties derived from AERONET direct sun data. *Atmospheric Chemistry and Physics*, 7(2):453–458, 2007. doi: 10.5194/acp-7-453-2007. URL `https://www.atmos-chem-phys.net/7/453/2007/`.

V. Schenzinger and A. Kreuter. Reducing cloud contamination in aerosol optical depth (AOD) measurements. *Atmospheric Measurement Techniques*, 14(4): 2787–2798, 2021.

A. Smirnov, B.N. Holben, T.F. Eck, O. Dubovik, and I. Slutsker. Cloud-screening and quality control algorithms for the AERONET database. *Remote Sensing of Environment*, 73(3):337–349, 2000. ISSN 0034-4257. doi: https://doi.org/10.1016/S0034-4257(00)00109-7.