Manuscript number: EGUSPHERE-2023-2885
Old title: Algorithm for continual monitoring of fog life cycles based on geostationary satellite imagery as a basis for solar energy forecasting
New title: Algorithm for continual monitoring of fog life cycles based on geostationary satellite imagery
Submitted to: Atmospheric Measurement Techniques (AMT)

Dear anonymous reviewer,

On behalf of all authors, we would like to kindly thank you for your very constructive comments and suggestions, and as well for spending your precious time on reviewing our manuscript.

Please find below our answers to your inquiries.

Note that the list of references mentioned in this document is given on the last page.

Sincerely yours,
Babak Jahani, Jan Cermak

Referee #1

Summary

- The authors try to develop a IR-only fog detection using channels at 8.7, 9.7, 10.8, 12, 13.4 microns of SEVIRI. The best combinations of these channels for fog detection are inferred by XGBoost. Before this work, the same group developed another fog detection algorithm, named SOFO, based on multiple tests (cloud, ice, snow, droplets) using IR-vis channels at 0.6, 0.8, 1.6, 3.9, 8.7, 10.8, 12 microns from the same satellite instrument. SOFO is more physically based but requires more channels to operate. They show that the XGBoost-based method has a POD of ~75-80% and an FAR of ~20-30%.
- The authors argue that their XGBoost-based method is expandable because while their validation/test datasets contain regimes that are not included in the training dataset, the POD and FAR performances on those regimes remain comparable to the training (except test4 which has fog frequencies < 1%).
- Overall the presentation is clear. The motivation of developing a new method of IR-only fog detection is discussed. The steps creating the training dataset, test1, test2, test3, and test4 are outlined. These details make sure their work is reproducible.

Comments

C#1.1

My biggest concern is that the FAR is well above 20%. Statistically, a method would be deemed useful if it has a FAR less than 5%-10%. My concern also applies to their previous method, SOFOS, which has an even greater FAR (as high as 40%). The XGBoost-based method may seem to be better but note that the POD of XGBoost-based is 10% less than SOFOS. So based on the POD and FAR, in my opinion, both methods do not seem to be practical. A potential problem is that a part of the training dataset has been based on the SOFOS method to create the fog/low stratus labels. Therefore, errors in SOFOS would propagate into the training dataset and eventually upset the training of XGBoost.

A#1.1

Thank you for raising these valuable points. Please let us clarify this matter:

1. You are correct that the XGBoost-based algorithm shows a lower FAR, but also a lower POD compared to SOFOS, which translates into a similar accuracy level for both products. In fact, this is in line with the hypothesis and objective of the study: “spectral LIR data can be used to differentiate FLS from cloud-free land and non-FLS clouds with comparable accuracy to an existing state-of-the-art daytime FLS detection algorithm”. This fact has been highlighted in the revised manuscript at lines 64-66, 81-84 (Section 1), 351-353 (Section 4) and at lines 432-434 (Section 5). Here we are not aiming at proposing a technique that has better accuracy compared to SOFOS. Instead, the novelty here is to have a technique that is applicable over day and night, as well as the day-night transition times, and has an accuracy comparable to an existing technique making use of a wider part of the

electromagnetic spectrum (i.e. visible-range channels). The idea here is not to develop a perfect technique for all situations, but one that works around the clock.

2. As the error metrics *POD*, *FAR*, *CIS* and *BS* are calculated as relative statistics, the absolute number of FLS/non-FLS cases included in the datasets used for validation can make a big difference in the results obtained. This is particularly important here because they are calculated relative to a small subset of the data: the FLS-positive cases (as identified by the truth or predicted product) which are inherently low in number. As a result of the relatively small denominator, they can show a relatively high degree of sensitivity to a few misclassifications. Additionally, they do not provide a global image about the overall classification accuracy of the product, as they do not account for the "*true negative*" instances, which are very large in number for FLS. For these reasons, although the metrics *POD*, *FAR*, *CIS* and *BS* provide essential and detailed information on the product's performance, they need to be interpreted with care. To account for these two matters the error metrics *ACC* and *PFD* were introduced. As the denominator of these metrics are rather large and take the "*true negative*" instances into consideration, they are expected to be better suited for showing the product's overall performance over the whole dataset. As can be seen from Figure 1 presented in the manuscript, both products show comparably good performances in terms of these two metrics over all datasets (ACC and PFD are both above 0.97). This argument has now been included in the revised manuscript at lines 385-395 (Section 4).
3. Another matter to point out is that the METAR data used in algorithm evaluation itself does not perfectly capture ground truth. One important aspect is that METAR data are taken from the ground, looking up, whereas the satellite takes the opposite perspective. Plus, the METAR stations belong to different organizations, managed by different organizations and do not use unified instrumentations. Additionally, as mentioned in the manuscript, what is observed from a point ground measurement does not necessarily match with what can be seen by a space-born observer that has a pixel size of 16 to 64km^2 over Europe. Please see sections 3.1, 3.2, and lines 475-480 of the revised manuscript where the resulting problems are discussed. We have tried minimizing the effect of these variables by imposing the quality control protocols described in the sections 2.3 and 3.2 of the revised manuscript and correcting for the non-FLS clouds passaging through the satellite's line of sight (described in Section 3.1 of the revised manuscript) -this is especially relevant in Europe, where about 30% of FLS events are obscured by other clouds (Cermak, 2018). Of course, this cannot be perfect and has its own limitations, leading to the leak of unwanted datapoints to the training and test sets.

We hope this response addresses your concerns. Thank you for prompting us to clarify these important points.

Please be advised that in line with this comment and C#2.1 (first comment from referee #2), we have added a new figure to the manuscript (i.e., Figure 3) which shows 1) applicability and consistency of the algorithm over day and night as well as during day-night transitions, 2) applicability over water and land/water transition regions, and 3) consistency of the classifications with the observed radiances.

Furthermore, for the same objectives we had provided an animation presented in the supplementary material S1 (downloadable from https://zenodo.org/records/10244714) in the initial submission. In this animation, the left-hand panel shows a false-color RGB image constructed based on the SEVIRI raw channel data with the red, green, and blue channels being $BT_{12.0}$ - $BT_{13.4}$, $BT_{8.7}$ - $BT_{12.0}$, and $BT_{10.8}$ - $BT_{12.0}$, respectively. In this panel, the green color represents the high clouds, and the light and dark red colors represent the clear-sky and FLS, respectively. The right-hand panel of this animation also shows the outputs of the ML FLS detection algorithm developed in the present study.

- -

C#1.2

There is a lack of physical explanation why BT12.0, BT8.7 - BT12.0, BT10.8 - BT12.0, and BT12.0 - BT13.4 would have been "chosen" by XGBoost. Their searching process (randomizing the combinations of the channels and find which minimal set of combinations give a desirable result) is typical of modern machine learning approaches. But in applied sciences, the interpretation of the results is as important as the method itself.

A#1.2

Thank you for pointing it out. Very brief information as provided in the initial submission (at lines 271-274 in the initial submission). In line with your comment, the explanations were further extended. Additional explanations were added to Section 3.4 of the revised manuscript at lines 269-281.

- -

C#1.3

Most of the discussions of PV in the abstract and the text are irrelavant to the study. At least the discussions of PV in the abstract should be removed.

A#1.3

In line with your comment, the discussions of PV in the abstract were removed.

- -

C#1.4

In addition, the term "life-cycle" in the title is misleading because the current manuscript does not study the life-cycle of fog/low stratus.

A#1.4

In line with your comment, the title has been modified to “Algorithm for continual monitoring of fog based on geostationary satellite imagery”.

- -

C#1.5

A "train dataset" should be a "training dataset".

A#1.5

The suggested change has been applied to the manuscript.

- -

References

Cermak, J.: Fog and Low Cloud Frequency and Properties from Active-Sensor Satellite Data, Remote Sens., 10, 1209, https://doi.org/10.3390/rs10081209, 2018.