Atmospheric Measurement Techniques Discussions

Atmos. Meas. Tech. Discuss., referee comment RC1
https://doi.org/10.5194/amt-2021-179-RC1, 2021

Comment on amt-2021-179

Anonymous Referee #1

Referee comment on "Twenty-four-hour cloud cover calculation using a ground-based imager with machine learning" by Bu-Yo Kim et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2021-179-RC1, 2021

This manuscript describes the use of several Machine Learning techniques in order to estimate the cloud cover from a series of statistical characteristics derived from all sky images. This study hardly shows any new concept, as analysis of all sky images have been conducted for about 15 years so far, including a great number of studies that use machine learning techniques. However, I tend to be favorable to the publication of this manuscript, as the explanation is quite clear, the number of analyzed ML techniques is high, the data set used is comprehensive (including day and night images), the way of reducing the information contained in an image is stimulating, and the results are relevant enough.

In any case, I have some few minor comments and suggestions that may be helpful to improve the paper.

Line 25: I wouldn't say "to date", as it's been a while since several automated systems have been introduced, maybe not for standard (official) observations, but for uses of research or of solar plants management.
Line 29. I would say that the use of “longer-period” observations is ambiguous here. I understand that you mean at a lower temporal resolution.
Lines 42-42: I think that these two references regard only to camera systems; they could fit better in other places of the introduction (line 50, or line 57). Here there are other more general papers that I think are more adequate: Boers, R., M. J. De Haij, W. M. F. Wauben, H. K. Baltink, L. H. Van Ulft, M. Savenije, and C. N. Long, 2010: Optimized fractional cloudiness determination from five ground-based remote sensing techniques. J. Geophys. Res. Atmos., 115, 1–16, doi:10.1029/2010JD014661.
Line 46: regarding the use of ceilometers for cloud characterizations, this reference is also adequate here: Costa-Surós, M., J. Calbó, J. a. González, and C. N. Long, 2014: Comparing the cloud vertical structure derived from several methods based on radiosonde profiles and ground-based remote sensing measurements. Atmos. Meas. Tech., 7, 2757–2773, doi:10.5194/amt-7-2757-2014.
Line 58: I think that when you say “cloud cover” you mean “pixel”, here.
Line 64: regarding the difficulties for distinguishing clouds from clear sky depending on some conditions, this reference may be useful: Calbó, J., C. N. Long, J. González, J.

Augustine, and A. Mccomiskey, 2017: The thin border between cloud and aerosol : Sensitivity of several ground based observation techniques. Atmos. Res., 196, 248–260, doi:10.1016/j.atmosres.2017.06.010.
Lines 76-77: there is some text which is repeated from line 74 in the same sentence.
Line 119. I understand that this number of 7,402 images refers to images and concurrent human observations.
Line 147-148: this method of masking (brightness greater than 240) may be valid for bright objects (Sun) but not for trees or building, if I understand correctly.
Lines 151-152. I would say that the calculation of Y should be presented in section 2.2, not here.
Line 155. I suggest stressing the fact that only the statistical characteristics (along with time information) are used as inputs for the ML methods.
Eq. (7-11). I would write the equations much closer to the place where they are mentioned in the text (in line 146) not here after line 160.
Lines 168-169. You should specify that these values of 0, 10, and 5 tenths refer to the sample images in Fig. 2.
Section 3. This section is difficult to judge by me, as I'm an atmospheric scientist, not a computational science researcher. However, I would suggest presenting the ML methods in increasing complexity order. That is MLR as the first method, then K-nearest, SVR, etc.
Fig 4 and fig 6: I think that the color code is not clear. Of course, cloud cover classes which are more populated get higher percentages (i.e., colors tending to orange and red). But what is relevant is, for each cloud cover class (given by the reference, DROM), how are the ACOS predictions distributed. In other words, I would color the boxes regarding the percentage relative to each class, not to the overall number of cases in the whole dataset.
Fig 5. I find this figure quite interesting, as it allows some physical interpretation of the results shown. So, despite some comments are made in lines 293-298, I would suggest adding more detailed comments if possible. Do you think that RBR ad RBD are "redundant" information? Why date and hour have a minor role?...
Line 346. When you say "in general", do you mean "in other, previous studies"?
Line 380-384: I don't see how come you can get physical characteristics of clouds, as this study only predicts clouds cover, but no spatial characteristics. You could extend a little this suggestion for further work.