

Interactive comment on “Leveraging spatial textures, through machine learning, to identify aerosol and distinct cloud types from multispectral observations” by Willem J. Marais et al.

Anonymous Referee #1

Received and published: 20 May 2020

General comments

This very informative paper presents a novel solution to the outstanding problem of differentiating thin cloud from thick aerosol features in satellite datasets derived from visible imagers. A machine learning (ML) approach based on Convolutional Neural Network (CNN) is presented and discussed. The training datasets are derived from images that are not from radiometers, but are random images from the internet for the initial pre-training and from True Color images classified by hand for the fine tuning of the training. The latter dataset is also used for verification. The reason why this works

C1

is that the algorithm operates based on spatial smoothness of the features and not necessarily on the fact that the images represent radiometric information per se. This is a point of strength of the ML approaches. A series of test cases is presented.

Results show that the the CCN algorithms (bot the fine-tuned with 25x25 pixels and the pre-trained with 100x100 pixels) performed better than the standard deviation method in correctly classifying cloud versus aerosol features. The algorithms do not perform well for the cases of cloud or aerosol egdes, for thin cirrus with optical depth lower than three and thin aerosol cases. The fact that the algorithm does not perform well in thin aerosol situations which were purposefully not included in the human-labelled training dataset highlights the importance of a good design of the training datasets. Any “biases” present in the training datasets will also reflect in the output of the ML algorithms which are basically ignorant of the physics of the problem and only deal with certain characteristics of the image itself.

Overall the paper is well written, robust, well documented with plenty of references and the research presented is of high interest for the community. I have some minor comments below. I would also recommend a bit of rewriting of the algorithm description which I found difficult to follow. I encourage the authors to pursue further this work to improve on the current findings and be able to address the remaining “sticky” points.

Minor comments/typos

line 80: can be labelled

line 95: correctly identify

line 185 :I do not see L and K defined. At this point of the paper, the reader gets a bit lost in the maths (at least I did). Please see if you can keep it more discursive and easier to read. I would suggest putting the mathematical details of the algorithms in an appendix.

line 218: I think this is a really important point. If there are any biases in the training

C2

datasets the algorithms will likely reproduce those biases.

line 223: It might be rather arbitrary to recognize clouds at different altitudes from the True Color images. Could you comment on that. Further down, figure 6 makes this point: how can you be sure that that's a cirrus cloud?

line 283: there's a period in an odd place, should be a comma

line 289: for how

line 292: unfinished sentence

Section 2.5 is difficult to read. Some details could perhaps be given in an appendix and the text in the paper made more fluid.

Section 2.5.3 is a bit too concise. Please expand.

Line 365: yes, the edges are problematic. Please spend a few words in saying how you will address this problem in future work

Line 376: this is a difficult case even for a human-based recognition.

Line 423. "However" instead of "But" to start the sentence

Line 428. This is not the focus of this paper, but are there other studies confirming the inconsistency that you see between MODIS and CALIOP data?

Line 447: this also highlights that it is really important to choose well the training dataset.

Line 450: this is indeed a sticky point.

Line 476. You do not need the question mark at the end.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2020-74, 2020.