Atmospheric Measurement Techniques Discussions

Atmos. Meas. Tech. Discuss., referee comment RC2
https://doi.org/10.5194/amt-2022-45-RC2, 2022

Comment on amt-2022-45

Anonymous Referee #2

Referee comment on "Segmentation-Based Multi-Pixel Cloud Optical Thickness Retrieval Using a Convolutional Neural Network" by Vikas Nataraja et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2022-45-RC2, 2022

Overview:

The authors have developed an efficient way of taking a continuous monochromatic radiance field (6.4 x 6.4 km^2 image) formed by sunlight reflected off clouds and mapping it to the underlying cloud optical thickness (COT) field at the pixel scale (0.1 km). More precisely, each pixel is assigned to one of 36 predefined classes of COT defined as intervals. This is done using a convolutional neural net (CNN) trained on synthetic imagery of LES clouds using a forward 3D RT model. The LES clouds of course come with known ground truth in COT. As is customary in machine learning (ML), 1/5 of the synthetic data (radiances and COTs) is set aside for CNN performance evaluation. Specifically, the authors ask how much better does the CNN do compared to the independent pixel approximation (IPA).

Although this is not the first study to use ML to go from radiance fields to COT maps, it is still an emerging method, and the authors are commended for pushing this envelop in a new direction by using a discretized ("segmented") COT scale. The paper will likely become an important contribution to the emerging literature in cloud remote sensing where the 3D variability of clouds is embraced rather than ignored. However, to get there I recommend a major revision of the current manuscript.

Major concerns:

As a reader/reviewer with very little knowledge about ML, I approached this paper with a strong desire to learn more. And about how ML can be applied to a an endeavor that I care about. The fact that none of the authors come with an affiliation in computer science made my expectation even greater. I was, however, disappointed. The key Section 3 was not easy to read. I came out of it feeling that that there was either too much or not enough detail. In particular, what I guess is ML jargon was often not explained.

I strongly recommend that the authors make that Section 3 into an Appendix with improvements suggested below (mostly more details), and leave in the main text (thinking of it as "mandatory reading") a well-crafted high-level summary. That summary of what's going on in the "black box" should be just enough to leap into the interesting results presented in Section 4.

Sequential comments:

* Fig. 1: The radiance scale is missing (best to use "BFR" units, pi I / mu_0 F_0).

* Fig. 1: To better show the IPA underestimation, maybe add a 4th panel: same as (c) but with a stretched scale.

* Fig. 1 and elsewhere: Avoid the "rainbow" color scale that does not work well in B&W print, nor for color-blind persons. Hint: the default "green-yellow" scale in python avoids these pitfalls.

* Section 1.2: The history of efforts to mitigate 3D RT effects is interesting to read. Someday it would be nice to have a more exhaustive version, but here at least one approach antedates BL95 and is worth mentioning, namely:

Cahalan, R.F., 1994. Bounded cascade clouds: Albedo and effective thickness. Nonlinear Processes in Geophysics, 1(2/3), pp. 156-167.

Interestingly, Cahalan's solution involves a multiplicative prefactor, as in (4), rather than BL95's scaling exponent.

* Section 1.3: This too is an interesting read that contrasts (physics-based) cloud tomography and (statistics-based) neural nets. The former is a pretty recent development with the real breakthrough paper being:

Levis, A., Schechner, Y.Y., Aides, A. and Davis, A.B., 2015. Airborne three-dimensional cloud tomography. In Proceedings of the IEEE International Conference on Computer

Vision (pp. 3379-3387).

A paper of special interest here that uses a CNN rather than 3D RT in the cloud tomography per se is:

Sde-Chen, Y., Schechner, Y.Y., Holodovsky, V. and Eytan, E., 2021. 3DeepCT: Learning volumetric scattering tomography of clouds. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 5671-5682).

* Fig. 2: What happens to the IPA when true COT > 40? We need to see that for a fair visual comparison with CNN.

* Fig. 2: How well does the prediction in (4) for the slope work here, in comparison with the empirical IPA vs true COT slope?

* Fig. 2: To the eye, it looks like, although biased low, the dispersion around the IPA retrieval is much smaller than around the (unbiased) CNN retrieval. Why? This looks like an opportunity to get the best of both approaches.

* lines 190-193: I couldn't comprehend the interchangeable use of "level of coarsening" and "aspect ratio" (AR) until I downloaded and browsed BL95. As I understand, BL95 is based on cloud models generated from 2D (Landsat) imagery. So, there is a user choice of how much geometric thickness (h) to assign to the clouds. In that case, talking about AR makes sense, and BL95 modulated it by varying h. But here the LES-generated clouds are inherently 3D. So, talking about coarse-graining (in the horizontal plane) makes sense. But the AR is something different, unless all that is taken from the LES is the 2D COT field and cloud thickness in the 3rd dimension is assigned and held constant, like in BL95. If that is the case, I missed it.

* line 208: Not an expert here, but I thought that LES microphysics schemes were either "bulk" or "bin" and, in the _former_ case, they can be either 1- or 2-moment. Please clarify "two-moment _bin_ microphysics".

* Section 2.1.2: Can I suggest one figure here to visualize the important differences with the Sulu Sea LES clouds? Something like Fig. 3 for the Sulu Sea simulations.

* Eq. (5): I think you mean r_cloud, not r_water, and maybe "." like in (6), not "*".

* Eqs. (5-6): Why not use the more common "q_lw" and "q_wv" for your mixing ratios? And, accordingly, the usual "Q_ext" for the Mie efficiency factor?

* line 286: Up to 9 km? Is is 6.4 x √2? If so, say it.

* Fig. 4: Great start for understanding the ML technique used here! However, still too many questions and undefined concepts for the non-cognoscente:
How do you get the 64 layers from a single one in step #1? Can it be another number? (I understand the subsequent doubling and halving.)
What is ReLU Activation?
What is Batch Normalization?
(Maybe better to have different colors for these two operations?)

* Section 3.2: "cross-entropy" is explained, but not "one-hot encoding", nor is "softmax activation".

* Eq. (11): Does alpha depend on i or c? If so, which and how? (And add the appropriate subscript.) If not, it can be factored out.

* Fig. 5c: What is the top layer? Looks like a binary cloud mask resulting from all the COT classes.

* line 403: Delete "resolution" (it is the domain size).

* Below Fig. 6: Please tell us a little about the "Adams" optimizer.

* Above Eq. (16): One COT bin (#27) is finally given explicitly. What about the others? Are they linearly sampled? Logarithmically? Surely this discretization of the COT scale also has to be somehow optimized.

* Section 4.1, and below: Better to use "coarsening factor" than "aspect ratio" (see comment above for lines 190-193).

* line 447 and Figs. ≥7: Cloud Variability is an interesting non-dimensional quantity. It seems to have an upper bound of 2, but that isn't clear from the definition. Please clarify.

* Section 4.2: Why is the number of scenes used for training described as "cloud morphology"? (That term does comes up later on, in p. 25 in Sect. 4.4 where clouds from different regions are contrasted.)

* Fig. 9, caption: What are the red dots?

* line 576: typo in "erroneous"

* line 605: Maybe the contradiction found here with BL95 has to do with the key difference (discussed previously) between their use of "aspect ratio" and the "coarsening level" that it is equated to here?