Review of the Manuscript “A neural network-based method for generating synthetic 1.6 μm near-infrared satellite images” Baur et al., doi:10.5194/egusphere-2023-353

The paper describes an extension of the existing MFASIS-NN satellite radiance forward operator

The paper is generally well-written, describes a novel scientific algorithm, and lies within the scope of AMT. As such, I recommend publication of the paper after addressing the following comments which summarizes my concerns about the present version of the manuscript.

General comments:

* I miss some discussion on the difficulties expected adapting this approach to other NIR/SWIR channels with cloud absorption, e.g. 2.2um or 3.7um, as well as the influence of the spectral response function (e.g. interaciton of gas absorption vs. droplet absorption). Is this just a matter of re-training the NN with different DOM forward simulations? If so, why has this not been done? Given that cloud particle absorption is comparatively weak at 1.6um vs. 2.2um, would this aspect influence the accuracy of the given approach? While I realize that fully covering this aspect would significantly expand the length of paper length, it seems worthwile to cover this point at least to some degree. At this stage, the focus on a single wavelength and instrument seems to unnecessarily limit the scope of the paper.

* Terminology: I have some reservations with the names used for the SEVIRI spectral channels. Frequently, the 1.6um and 2.2um channels are referred to as SWIR, and 0.8um is termed NIR[1]. While this might be a matter of taste, refering to 0.8um as VIS channel is misleading, as this wavelength is not within the range of human vision (even if this is the terminology used by EUMETSAT...).

* Performance: it would be good to give some more concrete indication of performance, beyond the two numbers given in the present manuscript. You state that “MFASIS-NN is an order of magnitude faster than MFASIS”, and MFASIS is orders of magnitudes faster than running DOM. Maybe you can add a table of execution times of each algorithm in terms of pixels/profile calculations per second&CPU?

* The vertical variation of effective radius/ ice crystal size is purely based on parametrizations. What if these parametrizations are unrealistic? One could use A-Train profiles instead of IFS profiles to avoid this constrain. An alternative approach/extension could be to develop a set of representative basis profiles for different conditions / cloud types (e.g. similar to [1]). How well do these parametrizations capture the variability in effective particle size e.g. versus the ICON model hindcasts? I would really like to see this aspect/limitations discussed more in-depth, including possible ways improving this point in future research. Note that the treatment of vertical variations in cloud microphysics could also be used in cloud retrievals, giving guidance on selecting a target parameter set / limited number of degrees of freedom.

* Language: while the article is generally well-written some sentences would benefit from either being split or at least separating different aspects using a comma, and adding hyphens between words (e.g. L480, “machine learning based approach” => “machine learning-based approach”).

Specific comments:

Abstract:
Given the paper content, I think the abstract can be clarified and improved to better describe the paper contents!
* L2: “with improved accuracy”: the baseline for the “improved accuracy” should be clarified.
* L6: “vertical gradients”: Gradients implies linearity, I therefore would prefer “vertical variations”
* L10: Sentence starting: “Additionally, a different parametrization … was used for testing”. This sentence is suprising/unclea: please clarify explicitely the role of the “other” parametrization!
* L14: “in all cases, the mean absolute reflectance error achieved is about 0.01 or smaller”. Is this with or without the “profile simplications” mentioned before? Can you add representative error estimates for the individual steps, e.g. going from DOM with fully known profiles to DOM with simplified profiles to MFASIS-NN?

Sec 1, Intro:
* L46: “An extension of MFASIS to account for the most important 3D effects….”. Are these extensions applicable to the 1.6um capabilities presented in this paper? If not, what is the impact of 3D effects for the accuracy of the described method? In particular, it should be made clear that the chosen evaluation approach does not include an estimate of the resulting uncertainty.
* Paragraph starting at L49: I would recommend adding at least some context of the use of VIS channels plus the 1.6um channel in Nakajima-King style retrievals, and the fact that some of the challenges addressed in the present work are highly relevant for the resulting cloud products.
* L57: “because at this wavelength water clouds can be distinguished from ice clouds”. I believe this statement is not true, there is an intermediate range were reflectances (best reference I can find is this comment on a preprint in ACP, which raises concerns about the separability [2])

Sec2, Data and Methods
* L101: “they remain too large”: How is “too large “determined? This statement implies an objective target accuracy, whose origin and magnitude should either be explicitely mentioned, or the statement should be reworded (e.g. “errors are significantly larger”), to make it clear that this is a subjective statement.
* L102: “Sensitivity to the effective particle radii is higher”: it remains unclear how sensitivity is defined here. Given the link between effective radius, optical depth and liquid water path, this statement only holds if optical depth is kept constant, not if liquid water path is kept constant!
* L111: role of water vapor absorption for SEVIRIs 1.6um channel could be described more clearly.
* L128: see Eq.2 in Scheck 2021. The equation reference seems to be wrong! The aspect of surface albedo also raises another interesting question: while this equation (referenced to Jonkerheid in Scheck 2021) can be used, why has the neural network not been trained to take surface albedo as input, and learn this equation? Maybe the authors can comment on this?
* Sec2.3: I find the discussion if differences in effective particle size for water/ice clouds between ICON and the parametrizations too short and qualitative. What does e.g. “somewhat smaller” mean?

Sec.3, Selecting Input parameters
* Figure 6: there seem to be a problem with the color bar, both online in Firefox and in my PDF viewer (okular), the color bar does not show a similar range of colors to the one visible in the figure! (Also applies to Figure 9 and 10!)

Sec.6 Conclusions
* L488: “that have been assumed in the assimilation”: is there a reference for this number?

* L490: “the 1.6um provides”: add channels

* L487: “in all cases, the mean … errors was about 0.01 or lower”. Why is the number 0.01 given here? How does this number relate to the value given in L470 (“the errors of NN5k are predominantely below 0.04)? If I understand correctly, 0.01 refers to the comparison of DOM with simplified cloud profiles vs. the NN. For applications, isn’t the larger number more relevant, which includes the error contribution resulting from the profile simplification?

[1] https://doi.org/10.5194/acp-23-2729-2023
[2] https://acp.copernicus.org/preprints/3/S1548/2003/acpd-3-S1548-2003.pdf