Response to Referee #2 (Hartwig Deneke)

In the following, our responses are in black, reviewer comments in *blue italics.*

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.

We are pleased to hear that and want to thank the reviewer for the constructive, detailed and helpful comments.

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.*

The methods developed here for the 1.6µm channel are not yet sufficient for all solar channels. We think it is justified to focus on this channel, as it provides important additional information, compared to visible channels, and is available on many existing satellite instruments. While we would like to keep the 1.6µm channel as the focus of our study, we fully agree with the reviewer that it would be useful to discuss for which other channels our methods are useful (without going too much into detail). For this purpose, we have evaluated our method for other instruments and channels.

As an example, the root mean squared profile simplification error for all purely solar channels (we will not consider channels with thermal contributions like 3.7-3.9µm channels) of the current and next-generation EUMETSAT satellite imagers is shown in the figure below (as circles). This figure gives a good indication on which channels are already usable with the current method. From the figure it is obvious that all channels with wavelengths up to 0.7µm and many channels with larger wavelengths up to 2.2µm have errors similar to or smaller than the 1.6µm channel. In particular, the stronger absorption by clouds in case of the 2.2µm does not seem to be a problem for our method. While stronger Rayleigh scattering could in principle pose an additional problem for channels with small wavelengths, the error for these channels is actually smaller than for the 1.6µm channels. It seems that the cloud top pressure and surface pressure input variables originally introduced to quantify the absorption by CO_2 and CH_4 characterize also the Rayleigh scattering sufficiently well. For those cases with a profile simplification RMSE smaller than 0.01 we trained also neural networks and the full reflectance RMSE (due to profile simplification and imperfect network training) is shown as crosses in the figure. We did not train all networks for the same number of epochs and in some cases we used networks with only 2000 and not 5000 parameters, which results in some variation in the additional error related to the network training. In none of the cases the full reflectance RMSE exceeds 0.012 and further optimizations seem possible.

The channels for which the simplification error is rather large are the ones for which absorption by water vapour is stronger than for the 1.6µm MSG channel and channels sensitive to certain trace gases like the 0.76µm oxygen A band channel of MetOp SG. The influence of the spectral response function explains the difference between the 0.8µm channels on FCI, MetImage, SEVIRI and AVHRR. On the older instruments the channels are wider and overlap more with water vapour absorption bands, whereas on the newer instruments the channels are more narrow and experience less absorption. Quantifying the impact of water vapour with just one input parameter is sufficient for the newer instruments with their weak water vapor sensitivity, but not for the more sensitive older ones. For all channels with higher simplification error, additional input parameters are required to better quantify the influence of water vapour or trace gases. In future studies we will investigate how to add suitable input parameters for these channels.

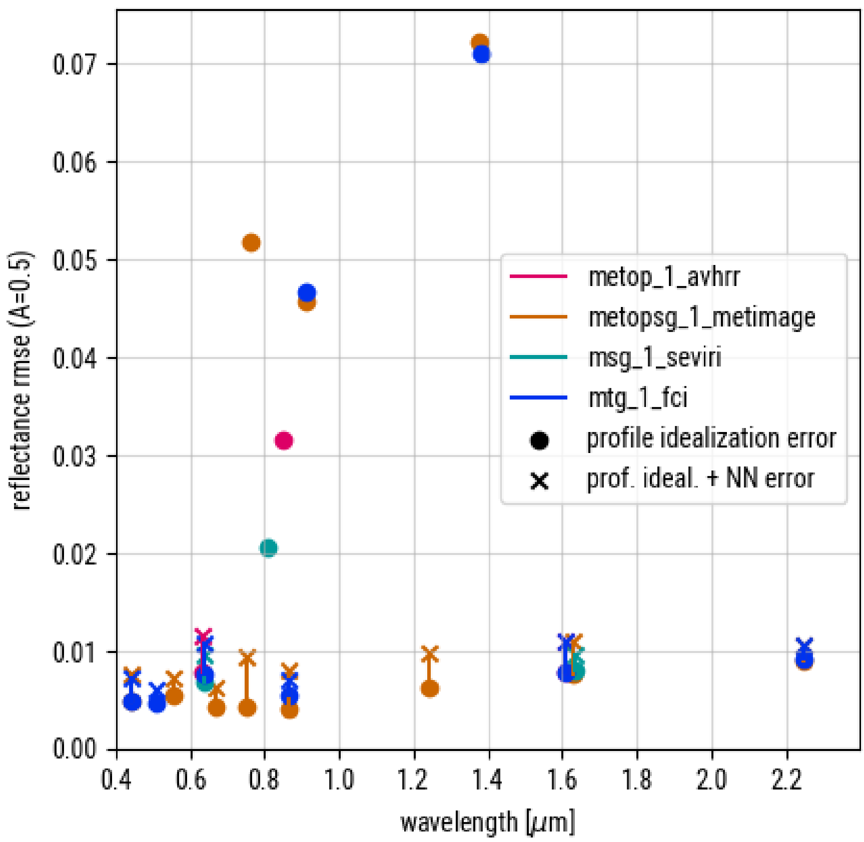

Figure 1: Root mean squared reflectance error due to profile simplification (circles) and, where available, full error (profile simplification and network training error, crosses) for all purely solar channels of the current and next generation of geostationary and polar orbiting imagers by EUMETSAT.

We will include the figure and the information given in this reply in the revised version.

** 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...).*

We will adjust the terminology, avoiding to call 0.8µm "visible".

** 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?*

Exact measurements can already be found in Scheck 2021 (Fig. A12) for the 5000 parameter network and in Scheck et al. 2016 for the LUT-based method. We will include updated information in the manuscript.

** 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.*

The method presented in our study is intended to speed up the computation of reflectance for cloud profiles from current global and regional NWP models. These models use either parameterizations or two-moment microphysics schemes for determining the effective radius profiles and so this is also what we are using for testing the method. As the method is primarily aimed at data assimilation and model evaluation we think this approach is sufficient for the current study.

While we agree that it would be interesting to check how well our method works for retrieved profiles, which could allow for additional applications, we think this is beyond the scope of this work and should be investigated in a separate, future study. Retrieved profiles may include features that are quite different from what is contained in model profiles (e.g. sharp gradients, extreme values) and these features and the distribution of integral quantities may not always be more realistic, but could also be related to limitations of the retrievals (see e.g. https://amt.copernicus.org/preprints/amt-2023-49/). Understanding the A-Train data set and selecting a suitable set of profiles could be a significant part of a future study, as well as approaches similar to the one in [1].

Concerning the question how well the parameterizations capture the variability of effective radii in the hindcasts, we agree that this should be discussed in more detail -- see response to the comment on Sect. 2.3 below.

We will add a more in-depth discussion on the aims and limitations of the method and future extensions. We will stress that the method is primarily aimed at data assimilation and model evaluation and discuss in the outlook that evaluation with retrieved profiles could be an interesting next step that may also lead to further improvements in out method.

** 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”).*

We will try to address these issues.

Specific comments:

Abstract:
Given the paper content, I think the abstract can be clarified and improved to better describe the paper contents!

We agree that the abstract should be improved.

** L2: "with improved accuracy": the baseline for the "improved accuracy" should be clarified.*

We will clarify that we are aiming at an accuracy that is comparable to the one of existing operators for visible channels.

** L6: "vertical gradients": Gradients implies linearity, I therefore would prefer "vertical variations"*

We agree and will change that in a revised version.

** 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!*

What was meant was that different parameterizations were used for testing the method. We will adjust the sentence.

** 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?*

The mean absolute error includes the error caused by profile simplification. The errors related to different profile simplification methods and the NN training are provided in Table 3 and discussed in detail in Section 5.1. We will adjust the abstract to make clear that the simplification error is dominating over the imperfect NN training error.

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.*

In this study we just aim for replacing 1D radiative transfer methods like DOM by a faster alternative and 3D effects are not considered, as we will state more clearly in the manuscript. The 3D extension for MFASIS (Scheck et al. 2018) was developed for the 0.6µm channels and has not been tested for 1.6µm channel. In principle it should be possible to adjust the method for the 1.6µm channel, but this has to be investigated in a future study.

** 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.*

We will add a reference to the manuscript.

** 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])*

We agree that this statement is not true for all cases (as visible e.g. in Fig. 4 of https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2018JD029772). We will formulate more cautiously that the 1.6µm can provide information that is helpful for distinguishing water from ice clouds.

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.*

"Too large" is indeed not well-defined and "significantly larger" is a good replacement. What we meant is that the errors are not an order of magnitude smaller than the assumed observation error in VIS assimilation. For other applications like just producing an image for forecasters to look at the errors of the old method may already be acceptable.

** 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!*

The reviewer is right -- it is only the derivative of reflectance with respect to the effective particle radius for constant optical depth that motivated this statement and we agree that the derivative for constant water/ice content is more relevant. While for water there is basically no difference between VIS006 and NIR016, the sensitivity to ice particle radii for constant ice content is actually somewhat larger for NIR016 (to give an example, if the radius is increase from 30µm to 60µm reflectance decreases by about 40% for VIS006 and 55% by NIR016).

We will reformulate the paragraph to make clear that the sensitivity to particle radius is not a decisive difference between VIS006 and NIR016. We will also change formulations in the introduction related to this question.

** L111: role of water vapor absorption for SEVIRIs 1.6um channel could be described more clearly.*

We will give a more quantitative description. Reflectance decreases approximately linearly with the water vapor mass. To give an example, for a relatively high column integrated water vapor content of 50mm and solar and satellite zenith angles of 60° the reduction is about 5%.

** 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?*

It is indeed Eq. 6 that should have been referenced. If albedo was just used as an additional input parameter, the neural network would have to learn Eq. 6, and of course it would not be able to perfectly reproduce it. The errors related to this imperfect learning can be avoided by training the network to generate reflectances for three different albedo values and using Eq. 6. This has also the useful side effect that it is very cheap to compute reflectances for additional albedo values using Eq. 6. The need to do this arises e.g. in model columns containing land/sea or sea/sea-ice boundaries.

RTTOV14 will allow for specifying multiple albedo values (plus corresponding weights) and use Eq. 6 to compute the correct mean reflectance.

** 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?*

We agree that the discussion needs to be improved. We will add something like

"The effective radii from the two-moment scheme (Fig. 4) show qualitatively different dependences on the optical depths compared to those obtained with parameterizations (Fig. 2). For the two-moment calculations the mean effective droplet radius reaches a maximum at tau_w=100 and decreases for higher optical depths, whereas for the parameterization there a further increase for tau_w>100. The mean effective ice from the two-moment scheme are more similar to the Wyser (Fig. 2c) than to the McFarquhar (Fig. 2d) results, but show a minimum at tau=10, where for Wyser there is a minimum. Even more obvious differences between parameterized and two-moment radii are found for the spread. The parameterizations mostly depend on quantities strongly correlated with the optical depth, like LWC and IWC, and are therefore quite well-defined functions of the optical depth with a small spread. The only exception is the Wyser parameterization (Fig. 2d), which has an additional dependency on temperature and therefore a larger spread. In contrast, the two-moment radii always show a spread that is considerably larger than for the parameterizations for both water and ice clouds, as is to be expected for more realistic radii."

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!)*

We will investigate this problem and correct the colors.

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

We will cite Scheck et al. 2020 here.

** L490: "the 1.6um provides": add channels*

Ok.

** 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?*

Here we talk about the full mean absolut errors including simplification and NN training errors, i.e. the rows in Tab. 3 labeled NN2/5/8k. The point we want to make is that these errors are an order of magnitude smaller than the assumed observation errors in the VIS006 assimilation.

We would replace the sentence starting with "In all cases" by something like
"For all network sizes and test data sets the mean absolute errors were found to be about

an order of magnitude lower than typical observation errors assumed for the assimilation of visible channels, which are in the range 0.1-0.15 for Scheck et al. 2020 and also for the operational assimilation at DWD.”

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