Response to Referee Comment (RC1) on

Understanding the model representation of clouds based on visible and infrared satellite observations (https://doi.org/10.5194/acp-2021-5)

We appreciate the detailed comments of the reviewer and have revised the manuscript to address all remarks. Our response is provided in black and the review in blue.

The paper discusses biases in the representation of clouds in convection-permitting simulations with the ICON-D2 model. The authors use a combination of visible satellite reflectances and infrared brightness temperatures to derive model shortcomings. Using satellite forward operators, observation equivalents are computed from model data which allow for a direct comparison with observations. The authors contrast uncertainties in the visible forward operator to sensitivities in model parameter setting. Based on their analysis result, the authors emphasize that the assumptions on subgrid-scale water clouds are the primary source for model biases in the visible spectrum and that the representation of these clouds need to be carefully revised to make further improvement possible.

I think the present study will become a valuable resource and I recommend the publication of the manuscript in ACP after major revision.

General Remarks

The paper is in general well written and structured. The objectives are clearly stated and all arguments are well supported. I don't see that language usage is of any concern. The manuscript discusses a relevant topic in atmospheric research, advanced analysis techniques are applied and the resulting scientific outcome is of interest for a wider audience.

General Comments

** Relationship to ISCCP-style analysis?: Combining VIS and IR as joint histograms or PDFs is not a new technique. There are a lot of different examples in literature which use joint histograms of cloud-optical thickness and cloud-top height to assess quality of climate model, global and regional weather forecasts. Two respective examples are: Zhang, M. H., et al. (2005), Comparing clouds and their seasonal variations in 10 atmospheric general circulation models with satellite measurements, J. Geophys. Res., 110, D15S02, doi:10.1029/2004JD005021. & Otkin, J. A., & Greenwald, T. J. (2008). Comparison of WRF model-simulated and MODIS-derived cloud data. Monthly Weather Review, 136(6), 1957-1970. and much more references therein and also based on these papers. It feels like you had completely overlooked this branch of studies and their* *relationship to your research. Please add a comprehensive discussion on this topic in your introduction and in your results section (where it is appropriate).*

Thank you for pointing to these additional references, which we clearly should have cited. We have added a discussion of these approaches to the introduction and cite also some of the studies in the results section.

While these studies do indeed use information derived from infrared and visible channels, all of them seem to be based on retrieved quantities. The novel aspect of our study is to perform the analysis directly in observation space. We performed another literature search, but could not find previous studies using this observation-space approach for visible and IR observations.

We believe that using forward operators is a favourable approach over the use of retrievals as characterizing the retrieval errors can be problematic (Errico, R. M., Bauer, P., & Mahfouf, J., 2007, https://journals.ametsoc.org/view/journals/atsc/64/11/2006jas2044.1.xml). Obviously a good characterization of errors is required also for the evaluation of changes in the model and we see this as a clear advantage of our approach.

Actually, our approach is fully in line with the recommendations from the 4th ECMWF workshop on assimilating satellite cloud and precipitation observations for NWP (https://www.ecmwf.int/sites/default/files/medialibrary/2020-06/Working_Group_Summaries_2020.pdf), where it is argued that it is easier to quantify errors for forward operators than for retrievals and that it is recommended to perform a more comprehensive and systematic evaluation of the errors in forward operators.

** Connection to solar power prediction: To my opinion, the analysis that tries to establish a connection between satellite data and global irradiance measured at surface is the weakest part of your manuscript. I guess you try to make the argument that solar power prediction would improve if the representation of clouds (measure from space) is becoming more realistic. However, your analysis and the presented arguments do not support such a conclusion by now. I recommend you to revise the analysis in Sect. 3.3. It would be beneficial for the reader to show how biases in GHI are correlated with the biases in VIS and IR108. One would expect that lower GHI biases coincide with lower VIS biases which would support the conclusion that the use of visible satellite data is beneficial for ground-based irradiance predictions.*

We see the point that section 3.3. was not well-connected to the rest of the paper and did not provide a detailed investigation of this connection. Thus, we removed 3.3. instead added version of the GHI-VIS / GHI-IR correlation plot containing only observations as a motivation in Section 2.

** Figure Quality: Please make sure that font sizes in your figures (e.g. axis labels, legends) are sufficiently large. Text in figures should not be significantly smaller than the text in the figure caption. Please update your figures accordingly!*

All plots were updated.

Detailed Comments

L. 11: "modified ...settings": Please rephrase to make more clear that both, variations in model settings and forward operator uncertainties, have been considered.

We have rephrased the abstract and introduction to address this.

L. 16: "VIS solar reflectance and global horizontal irradiance": Please make clear that the former in measured at TOA and the latter at surface.

We removed the statement.

L. 17: "will coincide" -> "can enable"?

We removed the statement.

L. 35: "are usually ... smaller" - Please support this statement with references!

We removed the statement.

L. 45: "minimization" -> reduction

Changed accordingly: minimization → reduction

L. 46/47: "Unfortunately, ..." Statement is very general and for sure not true for all current NWP systems. Please make it more specific and supported by references!

We added two references on this.

L. 51: "meteorological sensitive areas": Unclear what this means.

We added atmospheric instability for clarification.

Fig. 1: Does not appear to be referenced in the right order. Labels are too small.

We added a reference in section 2.1; all Figures were updated.

L. 58: "solar irradiance fluctuation" + "at surface" (or at ground). Also this statement needs to be support by a reference.

We added a reference in the rephrased introduction.

L. 65 full paragraph: This needs like a conclusion paragraph and is not in the right place here. Please rephrase and complete paragraph. This is the place where you can state your research question and outline how approach your research goal.

We have rephrased the introduction, including this paragraph.

L. 78: "cloud climatology": Here, and everywhere else: Please avoid the term cloud climatology which is mis-leading because it refers to long-term (!) cloud statistics which is not the case in your study. Please use "time-mean statistics" (or "time-average") instead.

We do agree and have revised accordingly in the whole paper. Instead of climatology we use "cloud statistics for a full test period" and "statistics for the full period".

L. 86: "ICON-D2". Please name the model version here.

We added the information (a development version based on version 2.6.1).

L. 96: "We performed six" + "additional"
Revised accordingly.

L. 98: "The objective was to ..." Please rephrase sentence.

We rephrased the sentence.

eq. (1): needs more explanation! I guess this is only a partial contribution to total cloud cover (might be indicated by subscript cc_turb). Is this cc contribution just added to the other contributions? What is q_sat? And where does the scheme come from (reference) and how should it be interpreted? Parameter B needs to be explained as well.

A more extensive description of the cloud-cover scheme has been added to the paper in Sect. 2.1, exchange point 2

L. 125: "like the operational one." -> "like the operational one-moment scheme."

Revised accordingly: one → one-moment scheme

L. 129: "cloud-concentration number" -> cloud droplet number concentration"

Revised accordingly: cloud-concentration number → cloud droplet number concentration

*L: 135: 2*10^4 to 4*10^4 hPa: This is definitely too large! Wrong units?*

Thanks for spotting this error, the unit was indeed wrong (Pa instead of hPa).

L. 146: "visible 0.6 um channel" Please specify if the visible reflectance is corrected by solar zenith angle. If yes, comparison in Fig. 8 would be inappropriate because GHI is scaled by a constant.

Thanks for this point. Now, GHI is corrected by solar zenith angle (added it in the caption and x-label of the Figure) and we only show the observed dependency as a motivation.

Figure 8 → Figure 3

L. 151: "TCW" / TCI": I would prefer "LWP" and "IWP", liquid-water path & ice-water path is more commonly used.

Revised accordingly to LWP and IWP

Section 2.3 misses to tell how aerosol is treated.

We added a sentence for clarification

Eq. (2):

** Consistency of symbols: You use low-case q in eq. (1) for content. And you use capital R as reflectance later. I suggest to use consistent symbols.*

Revised accordingly, also in the text;

** Is this equation consistently applied to visible and infrared? Please comment on this aspect.*

We applied this "effective effective radius" only in the visible forward operator, as it was not possible to use an external effective radius for water with RTTOV 12.1, which was used for the infrared calculations. Interestingly, we have recently performed tests with the new version of RTTOV (13.0), which allows for using the model effective radius. The differences in the histograms arising from using different effective radii were significantly smaller in the infrared channels than in the visible ones.

** How does this method compares to the generalized effective diameter in Senf and Deneke (2017), AR, eq. (B.3)?*

As in Senf and Deneke, the generalized effective radius is based on the addition of the volume extinction coefficient from the different phases. Our more simple formula is the result from assuming the same shape for ice and snow.These details are now included in the text.

Changes to the text:

new Line 207: […] and snow rs,eff. The effective radii for ice and snow are calculated under the assumption that both hydrometeors behave as randomly-oriented needles, and using the mass-size relationships, size distributions and number concentrations from the microphysics (for details see Fu et al. 1997 and Muskatel et al. 2021).

new Line 211: [...] snow phases, similar as Senf and Deneke (2017).

References:

Fu, Q.; Liou, K.N.; Cribb, M.C.; Charlock, T.P.; Grossman, A. Multiple Scattering Parameterization in Thermal Infrared Radiative Transfer. *J. Atmos. Sci.* **1997**, *54*, 2799–2812

Muskatel, H.B.; Blahak, U.; Khain, P.; Levi, Y.; Fu, Q. Parametrizations of Liquid and Ice Clouds' Optical Properties in Operational Numerical Weather Prediction Models. *Atmosphere* **2021**,*12*, 89. https://doi.org/10.3390/ atmos12010089

Senf, F., & Deneke, H. (2017). Uncertainties in synthetic Meteosat SEVIRI infrared brightness temperatures in the presence of cirrus clouds and implications for evaluation of cloud microphysics. Atmospheric Research, 183, 113-129.

L. 192/193: This is much too short! SGS clouds play an important role in your analysis. Please be much more explicit about your treatment of SGS clouds. What are the assumptions about microphysics (effective radius, adiabaticity) of SGS clouds? How does this impact cloud-optical thickness?

Both the microphysics and cloud-cover schemes produce a mass concentration qc/qi. The diagnostic qc/qi combines the mass concentration from both schemes. We calculate the effective radii using these mass concentrations and the assumptions for number concentration, probability distribution function and particle mass-size relation form the microphysics scheme. Therefore there are no differences in the assumptions for the calculation of the effective radius for grid clouds, subgrid clouds or the combination of them. Since we have a consistent calculation for grid and subgrid clouds we have not made any sensitivity study about the optical properties of subgrid clouds.

Changes to the text:

Line 214: […] water or ice. We assume no differences in the microphysical and optical properties of grid and subgrid clouds, so that the effective radius calculation is the same for both cases.

L. 197: "calibration offset" To my opinion, you are removing a systematic bias from the simulation which is fine in general. However, I would phrase it in that way.

You are right. We changed it in the text, accordingly.

L. 201: "spatial resolution" -> Please move to Sect. 2.2.

The sentence on spatial resolution was moved to Sect. 2.2

Sect. 2.4: What is the accuracy of GHI measurements?

This depends on the sensor:

1. CM11 (21 stations), CM21 (5 stations): WMO secondary standard instrument, where error should be less than 2 %
2. SCAPP (96 stations) deviates less than 10 % from secondary standard instrument

We think for motivating the work, the sensors are accurate enough.

L. 215: "... without coarsening and thinning" -> unclear

We removed this sentence.

L. 220: I don't understand why you don't take the observation as a reference: eps = P(SIM) - P(OBS)?

We want to study the effect in simulated reflectances. Observations are only of secondary interest here.

Violin plots: I would recommend to skip the distribution outside a certain range (<10th and >90th percentiles) to increase readability of the plots in Fig. 12. Otherwise these plots are dominated by the extremes.

Good suggestion, the y-axis is now limited to (-2,2).

L. 224: CFAD -> reference

Revised accordingly: Added a reference for CFADs and we also mention the ISCCP-approach here.

L. 224ff "Standard atmosphere .... ": Don't understand why you choose this distinction. Much more natural would be <273 K, (273 K... 243 K), <243K which would separate liquid, mixed-phase and ice clouds.

The reason is that we wanted to be close to WMOs cloud type classification: with an altitude of low clouds < 2000 m, middle clouds < 6000 m and high clouds > 6000m.

Fig. 4 + 5: Please use same projection as in Fig. 1 or the other way around. Please avoid histograms and use PDFs instead as you introduced PDFs as verifcation metrics.

We use the same projection for all Figures now.

We only show PDFs now.

L. 248: Fig. 4a & 4b -> wrong reference, 4b shows BTs.

Revised accordingly b → c
(Fig. 4 → Fig. 5)

Fig. 6 caption: White lines: What do they mean? "normalized by the sum" -> confusing. If you show PDFs then normalization is not a matter of choice: \int P(BT, R) dBT dR = 1!

You are right! We show the PDF and the caption was wrong. We removed the white lines from the plot.

Fig. 7: Observed BTs are higher than 300 K. Is the range > 300 K considered in the normalization of the PDFs?

We have extended the range to 310 K.
(includes all observed/simulated BTs in [200, 310] K)

Sect 3.2. Again, avoid the term "climatology".

Revised accordingly

L. 282 / L.284: There is a duplicate statement: "findings from previous studies"; "found in other studies" Please rephrase the two sentences.

We have rephrased the sentences and added some additional references, where the obs to model approach was used.

Sect 3.3:

** Please see my general comment. What is the general idea of this analysis? I guess you like to show that GHI forecasts can improve when VIS / BT forecast are more realistic, right? Why don't you show the bias in GHI vs. the bias in VIS? Otherwise, the reader get the feeling that plotting hourly average GHI values against instantaneous VIS observations is rather inappropriate (see L. 312 - 14).*

See above: Sect. 3.3 is now in Sect. 2.2

** Caption Fig. 8 "number of matches"-> unclear.*

We replaced "matches" by "collocated observations" in the caption.
[number of collocated surface GHI observations @Pyranometer stations and reflectance (TOA)]

** Meaning and usefulness of lines in Fig. 8 is also unclear.*

We removed the lines.

** Is scaling of GHI consistent with scaling of VIS radiances? See above.*

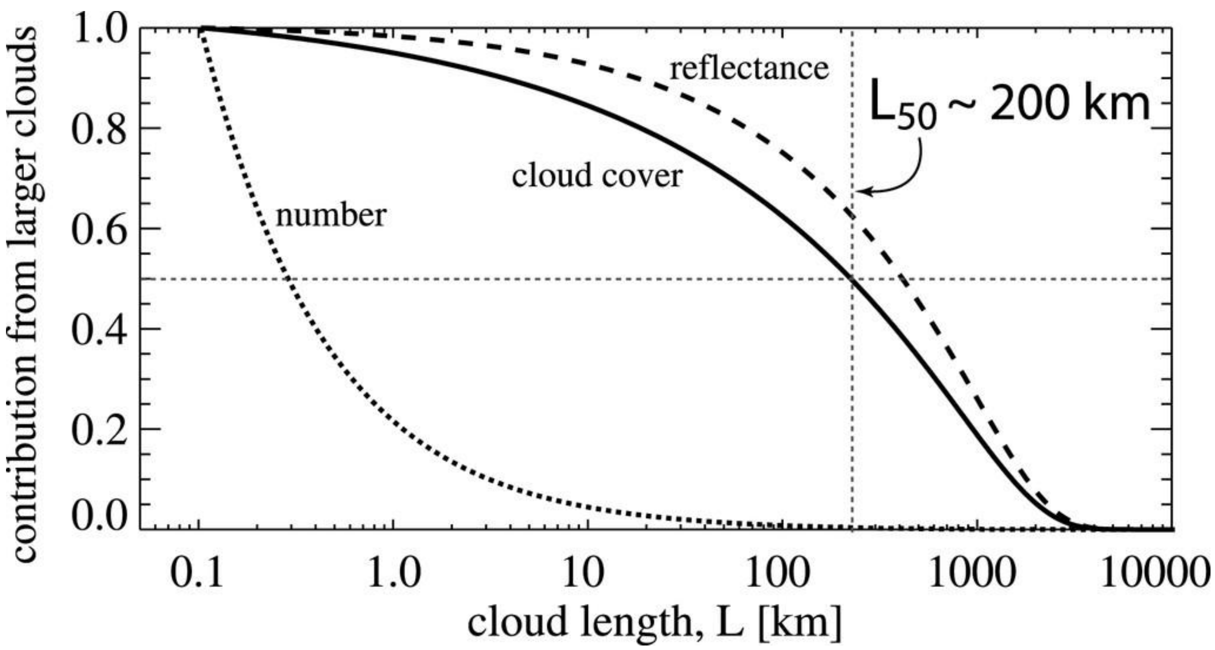
We made it consistent (see above)

L. 335: "imperfect parameterization" Again, a clearer description of the microphysics of SGS clouds would help.

A more detailed description of the parameterization is provided in the new version of the paper (see comment above). Given its simplicity, we think it is an obvious imperfect parameterization.

L. 337: flat plateau for grid-scale clouds: Would this mean that this VIS bias can be resolved by proceeding to even higher resolutions, e.g. hecto-scale simulations? Could you comment on this? Are there any indications in the literature?

According to Wood & Field (2011), Fig. 6, which is based on high-resolution satellite observations, 85% of global cloud cover comes from clouds larger than 10km and the cloud cover contribution from clouds smaller than a few 100m is very small:



[from https://journals.ametsoc.org/view/journals/clim/24/18/2011jcli4056.1.xml]

Therefore, subgrid cloud parameterization should become unnecessary for hecto-scale simulations. In fact, Heinze et al. (2017) neglected subgrid clouds in their ICON experiments with 150m resolution and found an improved representation of clouds, compared to km-scale models.

However, cloud representation depends on nearly every parameterization in the model (Zhang, M. H., et al. (2005): ,Otkin, J. A., & Greenwald, T. J. (2008), references you suggested and Webb et al., (2001)). We think that subgrid-scale cloud representation is currently the main challenge for VIS bias.

L. 347: "it seems that the subgrid water cloud parameterisation needs to be improved" -> or its coupling to the VISOP?

The visible forward operator sees the same clouds as the model, including the subgrid clouds. This is clearly a failure of ICON representing these subgrid-scale clouds correctly, in particular of subgrid-scale clouds in the boundary layer (see Fig. 8).

L. 353: "missing RT effects" -> unclear

We have clarified this statement: *"missing RT effects"* -> "missing 3D RT effects"

L. 382ff: In this paragraph, it is not clear how you treat aerosols in the reference run.

We have added a sentence in Sect. 2.3

L. 387: "Aerosol can scatter photons ..." Sentence reads weird. Please rephrase.

We have rephrased the sentence.

Interpretation of Fig. 11: It seems that carefully chosen aerosol can bring simulated VIS006 closest to the observation. Is this conclusion correct?

For improving the low reflectance part of the histogram adding aerosols is certainly the most important measure. However, for mid to high reflectances they are not that helpful. They have some influence on the maximum reflectance of the clouds, but the latter depends on many other factors.

L. 398: "ice habit is thus not likely to cause large uncertainties..." This is only true because your scenes have a very high low-level cloud cover and semi-transparent cirrus overlays lower clouds, right?

We agree on that and made this clear in the text.

L. 403: "simulation" -> "simulations"
We changed accordingly.

L. 411 & l. 415: "pixels" I find "pixel" inappropriate for model data.

We would agree if we compare in model space. However, the evaluation is carried out in observation space and we think pixel is also appropriate for synthetic images.

L. 420: "experiment VI" -> do you rather mean VII?
You are right. The text was revised accordingly.

L. 424 / 425: "cloud top height is an important additional parameter" I guess you mean in addition to cloud-optical thickness? Please make this clear!

We agree and revised the text to make this clear.

Fig. 12:

** It is hard to see the differences here. The plot are dominated by the extremes. Please trim the range of the PDFs e.g. within (-2, 2).*

** Panel a & b should have the same size.*

** y labels should be eps_n,d consistent with Sect. 2.5*

We modified Fig. 12 accordingly.

L. 450: "to eliminate the errors of the reference run" + "in the IR108 channel"
We added "in the IR108 channel "

L: 485: "solar satellite observations are novel for model evaluation". This might be true for RTTOV, but not for other forward operator methods. Please be more specific and discuss, if applicable, already existing advancements by others (e.g. in CRTM).

You are right. We meant operational applicable forward operators for model evaluation and data assimilation. A clarification has been added to the paper and we removed the word "novel".

L. 490: "well-suited to improve model cloud parameterisations for better PW power production forecasts" This statement could be better supported by your analysis. To my feeling, you can show that better VIS / IR108 forecasts ultimately lead to improvements in GHI predictions

We have removed this paragraph from the conclusion.