Anonymous Referee #1

Thank you very much for the positive feedback and all your relevant remarks and suggestions. We will try to answer all your questions in this document. The line numbers in our answers correspond to the numbers in the new corrected version.

Line 142: The results can be very sensitive to the choice of such thresholds. Did you investigate this? If so, please add a sentence on the reason of your choice in the article.

There are several thresholds in our methodology: the SZA threshold (SZA < 84°), the cloud fraction threshold for cloud detection (CF > 2 %) and the neighborhood size (5x5 pixels). The following figures shows the Fig. 10a of the article along with equivalent figures with different values of these thresholds, in order to see how the results can be impacted.

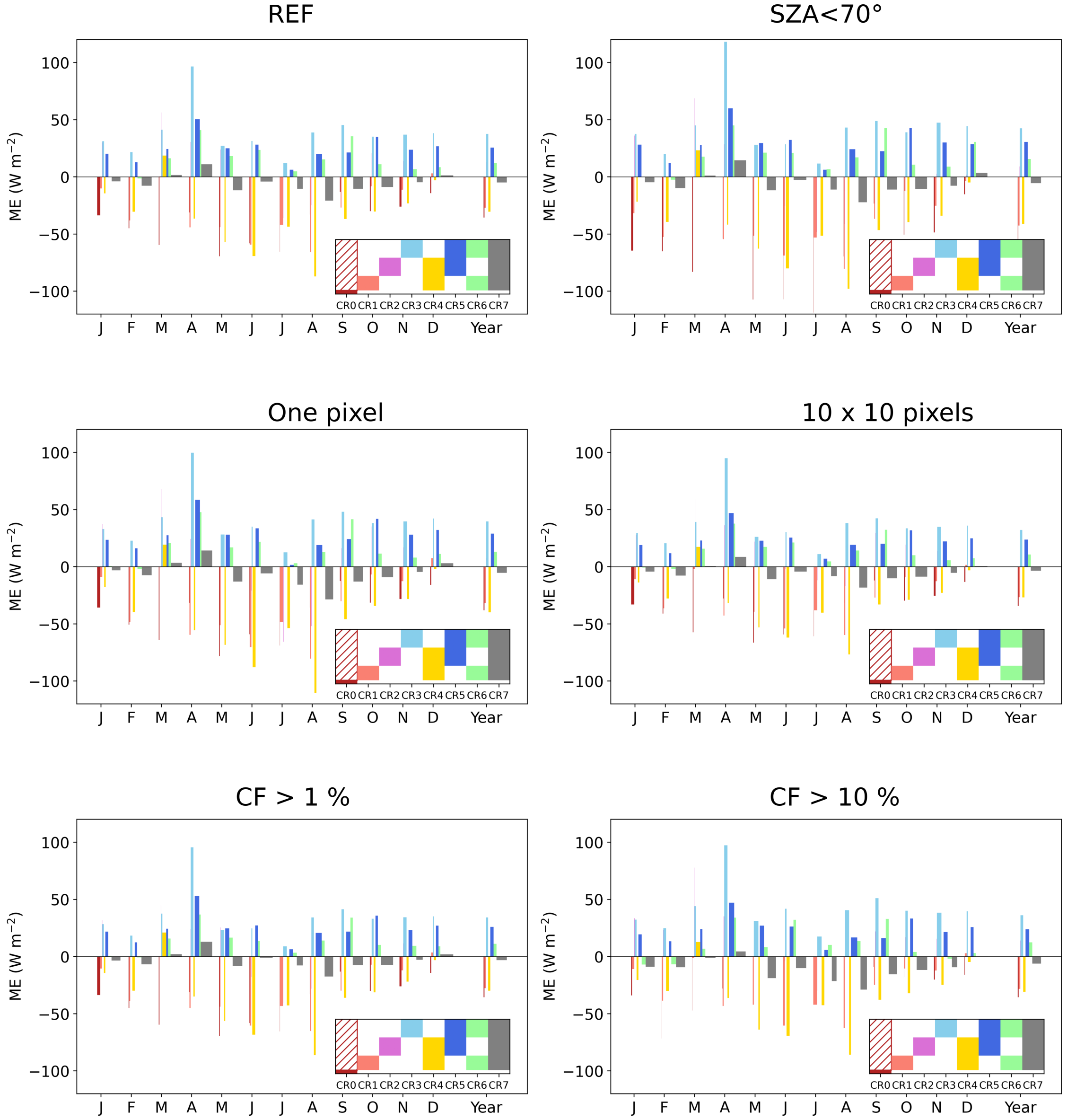

Figure 1: Monthly mean SWD bias (bar height, in W m^2) and relative frequency (bar width) for all modeled cloud regimes in 2020 and for different thresholds.

The Figure 1 shows that our results are barely sensitive to the chosen thresholds, and, in particular, biases keep the same signs.

In practice, we have selected the cos(SZA)>0.1 (SZA<84.3°) threshold to keep most of the observations. Although at high SZA observations are less reliable, the corresponding SWD are low, so the impact on biaises and SDE is low. Note that Tuononen et al. (2018) and Ahlgrimm et al. (2012) also chose a poorly restrictive threshold of respectively 5 and 1 W m^{-2}.

In the same idea, we tried more than one value for neighborhood size (neighborhood size of 10 x 10 pixels and without neighborhood size (1 x 1 pixel) by just selecting the closest point in the model), and likewise the systematic errors remain similar. Based on that, we chose the most relevant value 5 x 5 pixels, which is comparable to the satellite pixel size.

Regarding the cloud fraction threshold, the same value is chosen for cloud detection and cloud regime definition. Tuononen et al. (2018) chose a threshold for cloud detection of 5%, and Van Weverberg et al. (2018) chose a threshold of 0.5%. We have chosen an intermediate value of 2%, as in Ahlgrimm et al. (2012). Again, changing this threshold from 1 % and 10 % does not qualitatively change our results .

To be more accurate about the different threshold chosen, we did the following modifications :
L150, we added: “The neighborhood size is **chosen to be** comparable to the satellite pixel size.”
L165, we removed: “The neighborhood size is comparable to the satellite pixel size.”
L172, we added: “For each region, a 2 % threshold, **similar to the threshold value for cloud occurrence**, is set [...]”
L240, we added: “Note that several tests were performed by changing the neighborhood size **(1 pixel and 10 x 10 pixels)**, the cloud detection threshold **(1 %, 5 %, 10 %) and the SZA threshold (70°)**, which did not qualitively change these results.”
L369, we added: “5 %”

Line 174: Another threshold is introduced here. Again, as far as I understand, the sensitivity to this parameter has already been investigated. A short comment here would be helpful.

See previous comment.

Line 147-150: While neighborhood strategies are well-suited for verification purposes, this double penalty problem is exactly what PV power forecast have to struggle with. One could argue that this strategy is not suitable for solving the problems targeted in this study. An alternative strategy would be the usage of ensemble forecasts. As I understood from the following, you tried more than one value for neighborhood size and the results were not sensitive to this choice? Please clarify.

We agree that PV producers expect the forecast to be accurate for the exact location of the PV plant.

However, the effective resolution of AROME is about 8 km. For AROME and for other NWP models, we have no reason to trust a forecast at a very fine scale (that of a PV plant), at a specific point and at a specific moment. By using NWP models, we accept that it can’t be accurate at individual grid points. For more accurate very short term forecasts, other forecasting techniques are commonly use, such as machine learning approaches or all-sky imagers/satellite-based methods. It also depends on the size of the PV plants. For large ones, it's less important to know where exactly a cloud is locating as the production is aggregated over all PV panels. In any case, the double penalty does not prevent us from understanding SWD errors even if the methodology proposed does not solve this major limitation of NWP models.

We didn’t look at ensemble forecast, mostly due to excessive data volume, and we already have a lot of information with operational deterministic forecasts. It is an interesting venue that should be further investigated. However handling ensemble forecasts would not be trivial and alternative metrics would be necessary to evaluate the performances of such ensemble forecasts.

We added in L144: “While such a neighborhood strategy may seem unsatisfactory to PV producers who are concerned about SWD at the exact location of the PV plant, these users should keep in mind that NWP forecasts are not expected to be accurate at individual grid points.”

Line 173 (Figure 2) : In the following, several terms are introduced that describe a combination of several cloud regimes. For example, "low clouds" which are cloud classes CR0, CR1, CR4. Another example is geometrically thin clouds which are CR1, CR2, CR3. A table here that summarizes these groupings of cloud regimes would be helpful. It would also help to avoid inconsistencies like for geometrically thick clouds which are CR4, CR5 and CR7 in line 286 and 438, but only CR5 and CR7 in line 464.

We added a table (Table 1) that summarizes these groupings of cloud regimes.
The text was changed as follows at L457 to be more consistent and we added “high”: “For optically thick **high** clouds that most likely correspond to CR5, CR7 and HC”.

Line 173 (Figure 2): Furthermore, a more descriptive naming of the cloud regimes might be beneficial for the readability of the manuscript.

See previous comment.

Line 177 It is not clear which special cloud regime is meant. CR7?

This special cloud regime is just here to explain why the sum of relative frequencies is not totally equal to one in the figure 3a. As it was confusing, we changed the text in Line 177 : "Note that the sum of the relative frequencies of clear skies and all cloud regimes is not exactly equal to one in Fig. 3a. This is because there are situations, with very low relative frequencies, with a total cloud fraction larger than 2 % (implying that the scene is considered cloudy), but with a cloud fraction for each of the 3 vertical slices of the troposphere lower than 2 % (implying that it is not associated to any CR)."

Line 195: This description is a bit misleading and on first reading it sounds like (SWD_mod/SWD_obs)/SWD_modclear. What you meant is probably SWD_mod/SWD_modclear for the model and SWD_obs/SWD_modclear for observations. One (or two) short formulas are always clearer than a description in words.

We added an equation and changed the sentence Line 199 by:
"To begin with, we use the clear sky index (CSI), which is defined as :

$$\mathrm{CSI}_{\mathrm{mod}} = \frac{\mathrm{SWD}_{\mathrm{mod}}}{\mathrm{SWD}_{\mathrm{clear,mod}}} \ ; \ \mathrm{CSI}_{\mathrm{obs}} = \frac{\mathrm{SWD}_{\mathrm{obs}}}{\mathrm{SWD}_{\mathrm{clear,mod}}},$$

with SWD_{mod} the SWD in the model, SWD_{obs} the SWD in the observations, and $SWD_{clear,mod}$ the theoretical SWD under clear sky conditions from the model. The histograms of CSI are used to […]"

Line 201: This clear-sky underestimation of SWD is not surprising. There is a decreasing aerosol trend in Central Europe in the last decades. Older climatologies (like Tegen) are thus overestimating present-day aerosol loadings in Central Europe.

It is a good point. We added a reference and a sentence at line 207:
"In the observations the CSI can greatly exceed 1, as reported by Nielsen et al. (2018). This can be explained by an underestimation of the model clear sky SWD, which can be due to a reduction in aerosol emissions over the last decades (Wild et al., 2009) while the aerosol climatology used in AROME is older (Tegen et al., 1997) and thus overestimates present-day aerosol loadings. These values exceeding 1 can also be due to cloud enhancement effects (e.g. Gueymard et al., 2017), which occur under broken cloud conditions."

Line 247: As mentioned before, there is also a positive bias in Tegen AOD for Central Europe.

Actually we found an overall positive bias of SWD, which can not be attributed to this positive AOD bias. Hence we don't mention this in this section.

Line 416: "...and the SWD higher (15%)" Do you mean SD?

Yes, corrected.

Line 428-431: I did not fully understand this paragraph. It starts with the description of the second pair of values, the conclusion seems to be rather suitable for the third pair of values which is not described anywhere else.

Finally, we removed the Fig.13 and changed the text of this section. Please refer to our reply to Anonymous Referee #2 to check these modifications.

Line 446: For this section where you verify the satellite-based cloud classification of your approach, SYNOP data on high, mid and low cloud fraction might be helpful as an independent dataset looking from a different perspective. Have you considered this?

We believe that you refer to telemeters data or human observations. We indeed consider both types of observations, however the data are not available at all pyranometer stations used in this article. In addition, for human observations, cloud altitude can be very difficult to estimate, hence the information is highly subjective and not adapted for a quantitative study. Furthermore, cloud regime evaluation wasn't our primary objective. Nevertheless, in a companion study (that may be published in the future), we are comparing modeled cloud regimes with data from a highly instrumented site where lidar and radar observations are available.

Line 481: Did you separate between low and high clouds here? For low clouds, the consideration of snow might even further pronounce the already existing bias.

No, we didn't separate low and high clouds here.
We agree that taking snow into account could accentuate the existing bias. However, in an NWP model, there would be no reason to account for snow in high clouds only, and not low clouds.
The figures below show the profiles of the monthly snow mass normalized by the cloud hydrometeors mass for each cloud regime, in February 2020 (a) and in August 2020 (b) in AROME forecasts.

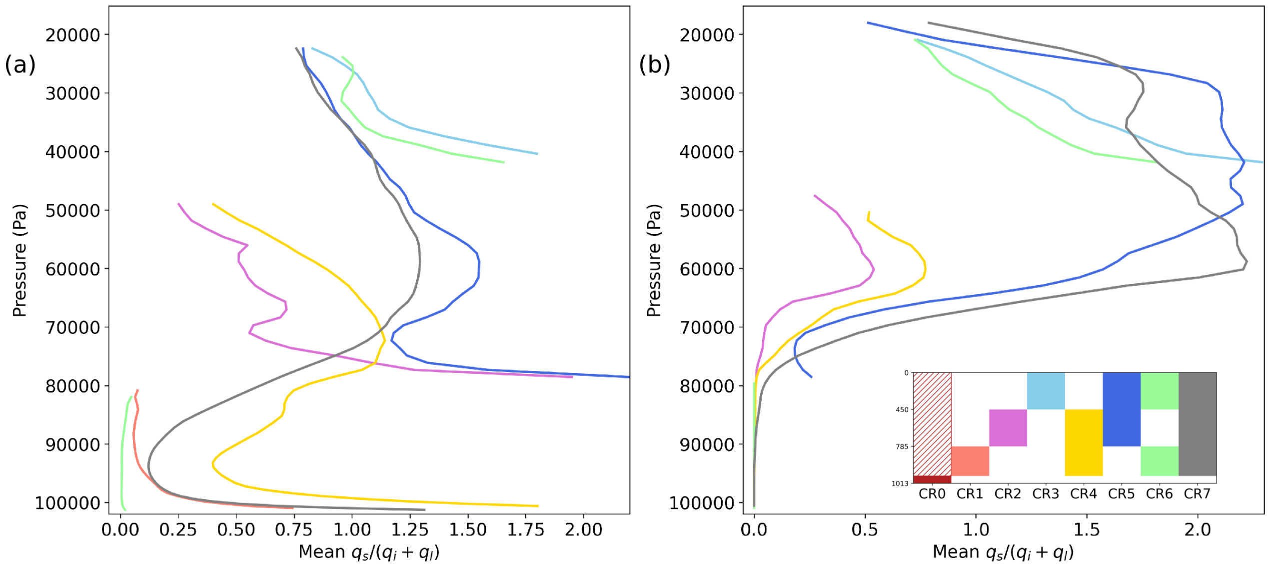

Figure 2 : Profils of mean $q_{snow}/(q_{ice}+q_{liq})$ (where q_{snow} is the mass of snow, q_{ice} is the mass of cloud ice and q_{liq} the mass of cloud droplets) in (a) February 2020 and (b) August 2020 in the model AROME and for each simulated cloud regime.

It shows, indeed, that taking snow into account could accentuate the existing SWD bias, in particular for CR4 in Winter. To further investigate this, we run simulations using a more recent version of AROME without (named: New version of AROME) and with accounting for snow (named: +Snow) for only two months (February and August 2020). The results are shown in Fig. 3, in terms of bias and SDE for each cloud regime in overcast conditions.

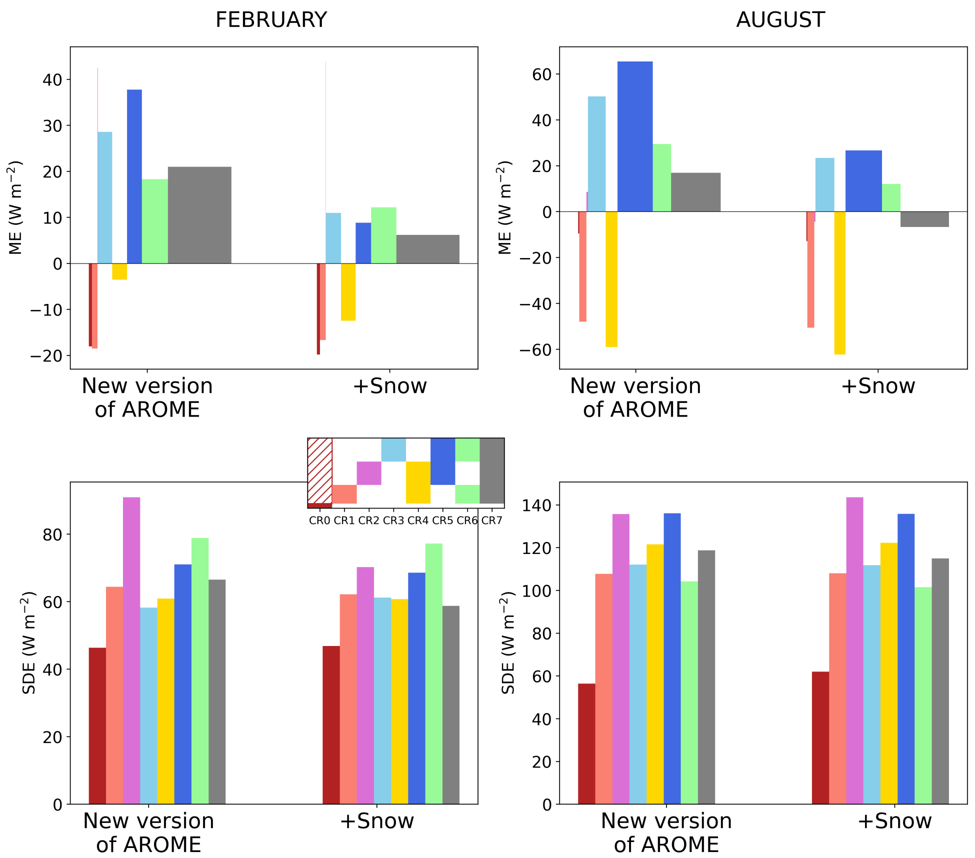

Figure 3: For a new version of AROME and for the simulation +Snow: in the first line, monthly mean SWD bias (bar height, in W m^2) and relative frequency (bar width) for all modeled cloud regimes for February and August 2020. In the second line, monthly SDE (in W m^2) for all modeled cloud regimes.

It confirms that the SWD bias of CR4 is deteriorated when snow is taken into account in the radiation scheme. However, the overall error is reduced even if the bias for some cloud regimes is accentuated, as the bias and SDE are greatly improved for some cloud regimes. In any case, this issue needs further investigation. For information, the snow is already taken into account in the radiative scheme of the IFS model, as its radiative impact is not negligible.

We added at line 479: "Note however that it could also deteriorate the bias for already too opaque low clouds (e.g. CR4 in winter)."

References:

Ahlgrimm, M. and Forbes, R.: The Impact of Low Clouds on Surface Shortwave Radiation in the ECMWF Model, Monthly Weather Review, 140, 3783–3794, https://doi.org/10.1175/mwr-d-11-00316.1, 2012.

Gueymard, C. A.: Cloud and albedo enhancement impacts on solar irradiance using high-frequency measurements from thermopile and photodiode radiometers. Part 1: Impacts on global horizontal irradiance, Solar Energy, 153, 755–765, 2017.

Nielsen, K. P. and Gleeson, E.: Using shortwave radiation to evaluate the HARMONIE-AROME weather model, Atmosphere, 9, 163, 2018.

Tegen, I., Hollrig, P., Chin, M., Fung, I., Jacob, D., and Penner, J.: Contribution of different aerosol species to the global aerosol extinction optical thickness: Estimates from model results, Journal of Geophysical Research: Atmospheres, 102, 23 895–23 915, https://doi.org/DOI:10.1029/97JD01864, 1997.

Tuononen: Evaluating solar radiation forecast uncertainty, Atmospheric Chemistry and Physics, 19, 1985–2000, https://doi.org/10.5194/acp-19-1985-2019, 2019.

Weverberg, K. V., Morcrette, C. J., Petch, J., Klein, S. A., Ma, H.-Y., Zhang, C., Xie, S., Tang, Q., Gustafson, W. I., Qian, Y., Berg, L. K., Liu, Y., Huang, M., Ahlgrimm, M., Forbes, R., Bazile, E., Roehrig, R., Cole, J., Merryfield, W., Lee, W.-S., Cheruy, F., Mellul, L., Wang, Y.-C., Johnson, K., and Thieman, M. M.: CAUSES: Attribution of Surface Radiation Biases in NWP and Climate Models near the U.S. Southern Great Plains, Journal of Geophysical Research: Atmospheres, 123, 3612–3644, https://doi.org/https://doi.org/10.1002/2017JD027188, 2018.

Wild, M.: Global dimming and brightening: A review, Journal of Geophysical Research: Atmospheres, 114, 2009.