We appreciate Ref1 comments, they are a good contribution to improve the overall understanding of the document. Review comments below are reproduced in blue and responses are in black.

The authors examined a methodology to obtain snowpack reanalysis using ICAR and data assimilation methodology over data scarce regions. The work is valued since the authors take the advantage of using atmospheric reanalysis data together with remote sensing and integrating techniques of downscaling and data assimilation. There are number of analysis in comparison with ground truth (AWS data sets) and remote sensed observations (fSCA of MODIS) in the study. In this regard, the study addresses an interesting topic which fits well with the scope of HESS and it is also an important contribution to the snow science especially for the mountainous regions where there is a data scarcity.

The authors replied a number of comments in the previous review step which has already added value to the study. There are still couple of important aspects that need to be refined so that the possible impact of the study would further increase.

Each process has its own contribution with different advantages, WRF improves the resolution compared to ERA5 and ICAR improves WRF results even more with computational efficiency. It would be remarkable to see the contribution of ERA5, WRF and ICAR in temperature and precipitation reanalysis separately with simply showing the same error analysis provided in Figure 3 and Figure 4. Even if ICAR will not improve WRF in terms of statistical performance it is still required for any analyzes/processes need a better resolution.

We have included the WRF outputs in the Figure 3 and 4, for a better understanding of the contribution of each process in the workflow. However, such comparison should be taken with care. The lack of ground-based data and short length of the AWS series makes difficult to extract conclusions in the comparison of the gridded products, as the ICAR/WRF/ERA5 sub-cell variability is not properly represented by the AWS due to the very complex topography.

According to the text, ICAR snowpack reanalysis data is acquired as a result of ICAR output. This step should be indicated in the flow chart as an output (in Figure 2). Then, the authors improve ICAR snowpack reanalysis since these are not very consistent with the observations (Observed and ICAR in Figure 5), so in the next step they prefer to implement data assimilation using satellite snow cover data. The authors should clarify this part and refine the role of Flexible Snow Model (FSM) in the study (which seems also to be the concern of reviewers previously); it looks as if there are two snowpack data sets: one as a direct output of ICAR reanalysis (ICAR in Figure 5) and the other is an output of FSM where the forcing variables are ICAR precipitation and temperature reanalysis data without data assimilation (which is not provided in the text or figures). The assimilation of MODIS fSCA is applied to the later as indicated, ICAR_assim, in Figure 2 and Figure 5. If this is so, to make a fair comparison, my suggestion is to use FSM results without data assimilation instead of ICAR snowpack reanalysis output (ICAR) in Figure 5 and also in the statistical performance analysis. This will also change the conceptualization in the study and objectives since ICAR_assim is not a direct output of ICAR snowpack reanalysis as authors emphasize that it is an ensemble output of FSM as a result of perturbation of ICAR reanalysis forcings.

ICAR snowpack reanalysis (ICAR_assim) is generated fusing the fSCA MODIS retrievals with FSM outputs forced by perturbed ICAR surface variables. It is already pointed in Fig2, we can not add any new step there. The use of a decoupled snowpack model (FSM) is mandatory, as an ensemble of snowpack simulations is required to implement the PBS. We have added the FSM outputs forced by ICAR unperturbed surface variables in the Figure5 as Ref1 requested. FSM and ICAR performs similarly during most of the situations, with differences by the end of some seasons where FSM has exhibited a faster decline. As the uncertainty associated to using different snowpack models, is lower than the uncertainty induced by the forcing, we hypothesize that the differences between ICAR_Noah and ICAR_FSM are mainly as consequence of the precipitation phase partitioning parametrization used in this work, that is less robust than the Thompson microphysics scheme implemented in ICAR. These difficulties on simulate the precipitation phase are typically found in the mild climates, where the spring snowfalls occur close to the zero-isotherm elevation.

We have included the following paragraph in the text, to explain the contribution of FSM to ICAR_assim:

[*The use of FSM to generate the ensemble of simulations, introduced some uncertainties in the workflow. Some water years showed earlier snow melts using the FSM forced by ICAR, compared with the ICAR snow outputs. As the uncertainty of the snow models associated to the forcing is higher* *than the uncertainty associated by the use of different model parameterizations and model structures (Günther et al., 2019), we hypothesize that such differences were caused by the differences in the precipitation phase partitioning, which is challenging to simulate in the areas that remain close to 0 °C during the snow season (Fayad and Gascoin, 2020). The lack of spring snowfalls in some years may have deep implications in the snowpack simulation that are not limited to its effect in the mass balance and the releasing of latent heat by refreezing the liquid precipitation. It leads to lower albedos, which combined with the high short-wave radiation of Lebanon due to its latitude causes earlier snow melts. However, such discrepancies are greatly minimized in ICAR_assim, by the assimilation of the fSCA retrievals.*]

There is an important impact of PBS assimilation on the results but this is rather unrealistic for some cases where there is almost no snow water equivalent before assimilation (Figure 5). Providing FSM results with forcing of ICAR before and after data assimilation as suggested above might better explain this remarkable contribution of data assimilation by PBS. If the description above (on direct ICAR snowpack output and FSM snowpack output without data assimilation) is not accurate then the authors should better explain the reasoning behind such a spectacular change in SWE with data assimilation through the use of fSCA of MODIS. Since there is a remark on NDSI reformulation using Theia in fSCA analysis to improve the results compared to the ordinary equation, it should also be important to deal with cloud removal process and give more explanation in section 3.2.1.

The use of FSM forced by the unperturbed surface meteorological variables of ICAR does not improve the snow simulations compared to ICAR_Noah, actually the implementation of FSM suffered from a less robust precipitation phase partitioning scheme, as explained above. It should be noticed that the comparison between the 1km grid scale of the simulations and the AWS exhibited differences due to the very variable nature of the snow at the point scale of the AWS, which is obviously not represented by the 1km resolution. The impact of smoother-based data assimilation schemes on snow simulations is well-known in the literature. Specifically, the use of PBS to assimilate fSCA observations into a snowpack model achieved similar or even better performances comparing with in-situ observations in the Californian Sierra Nevada (Margulis et al., 2016). Similar improvements in the simulations were found after the implementation of the PBS to assimilate fSCA retrievals in the Andes (Cortés et al., 2016), Switzerland (Fiddes et al., 2019), and Svalbard (Aalstad et al., 2018). We would like to emphasize that these implementations of the PBS and other smoother schemes work by correcting the forcing (mainly biases in solid precipitation) for a given water year through a Bayesian updating of the ensemble of modelled fSCA trajectories using the observed fSCA trajectory from MODIS. This, in turn, updates the entire SWE trajectory in a batch (i.e. for the entire water year, rather than sequentially) through the posterior forcing time series. Such updates can be considerable. For example, if the prior mean melts out too early relative to the observed fSCA, then the posterior mean SWE will typically be higher than the prior mean since it is only the trajectories with considerably higher snowfall that can reproduce the observed fSCA depletion. The converse case occurs when the prior mean melts out too late. This hopefully explains the "spectacular change" in SWE that can occur with the PBS. Such an update would not be possible with the widely used (for snow DA) filtering algorithms which directly update SWE sequentially in time, since with filters the information from observations can not propagate backwards in time. We did not follow any specific procedure for the cloud removal, as the cloud masks are already processed and included in the remote sensing products used in this study (MODIS and Theia snow).

References:

Aalstad, K., Westermann, S., Schuler, T. V., Boike, J. and Bertino, L.: Ensemble-based assimilation of fractional snow-covered area satellite retrievals to estimate the snow distribution at Arctic sites, Cryosphere, 12(1), 247–270, doi:10.5194/tc-12-247-2018, 2018.

Cortés, G., Girotto, M. and Margulis, S.: Snow process estimation over the extratropical Andes using a data assimilation framework integrating MERRA data and Landsat imagery, Water Resour. Res., 52(4), 2582–2600, doi:10.1002/2015WR018376, 2016.

Fayad, A. and Gascoin, S.: The role of liquid water percolation representation in estimating snow water equivalent in a Mediterranean mountain region (Mount Lebanon), Hydrol. Earth Syst. Sci., 24(3), 1527–1542, doi:10.5194/hess-24-1527-2020, 2020.

Fiddes, J., Aalstad, K. and Westermann, S.: Hyper-resolution ensemble-based snow reanalysis in mountain regions using clustering, Hydrol. Earth Syst. Sci., 23(11), 4717–4736, doi:10.5194/hess-23-4717-2019, 2019.

Günther, D., Marke, T., Essery, R. and Strasser, U.: Uncertainties in Snowpack Simulations—Assessing the Impact of Model Structure, Parameter Choice, and Forcing Data Error on Point-Scale Energy Balance Snow Model Performance, Water Resour. Res., 55(4), 2779–2800, doi:10.1029/2018WR023403, 2019.

Margulis, S. A., Cortés, G., Girotto, M. and Durand, M.: A landsat-era Sierra Nevada snow reanalysis (1985-2015), J. Hydrometeorol., 17(4), 1203–1221, doi:10.1175/JHM-D-15-0177.1, 2016.

The authors thank Ref.2 for their valuable comments. We provide a point by point answer bellow. Review comments below are reproduced in blue and responses are in black.

Line 51: Please provide a reference to this new statement.

Authors: We have moved the (López-Moreno and García-Ruiz 2004) reference to include this new statement.

Line 325: How was the snow depth measured in the AWS? I asked before and I think I did not see the answer.

It's a Campbell SR50A acoustic gauge. We have added the model of the sensor in the text:

“[...]information derived from a Campbell SR50A acoustic gauge at the three AWS.”

Figures 3, 4, 5, 6: I appreciate the authors changed the format “Date since” for actual values. However, I still miss the ticks in the x-axis with the exact location of the dates in some of the figures (You have done that in the y-axis or in x-axis of Figure 6). In addition, more dates (smaller time scales: months) would be appreciated for those sites with less analysed years. As I commented before a grid would help to interpret these figures.

We have added tick marks to the end/begging of the years as well as new reference lines to these figures. We have preferred to keep the same data scales in all the plots in the benefit of the consistency.

Figure 10: Why did the red dotted line change in this version of the manuscript?

The difference was caused by an inconsistency in our framework caused by applying the suggested changes by the previous review. It is fixed now using a more robust routine, thanks.

Supplementary 1: What does the colour scale represent? I understand it is number of pixels that fall within each grid. However, if that the case they seem less than 40% of 58400, but maybe I am wrong. In any case, I would appreciate if the authors included the metrics (RMSE and MAE) in the figure. Moreover, could the authors also add the value of determination coefficient for both linear fitting? I think is a better metric in this case. I would also recommend adding the same plot for the validation period.

The colour scale represent the relative density of the scatter plot in a log10 scale (added to the caption). The colour scale was designed to improve the interpretation of the figure. The RMSE and MAE metrics are since last version included in the 4.2 section, we apologize but the routine used to estimate the linear fit was primally designed to optimize the linear and non-linear functions, for which the R2 would not be relevant. To obtain the r2 it would be necessary to develop deep modifications in the code.