Authors: We would like to thank the anonymous referee for his/her interest and the comments on our manuscript. Bellow we provide a point by point answer to the issues raised by referee #1

Ref.1: It is indicated in the article that "ICAR model was used to obtain a finer 1 km x 1km spatial grid atmospheric simulation nested in the aforementioned WRF simulation domain.", and the comparisons and error analysis are done for ICAR and ERA5 for temperature and precipitation (Figure 2 and 3), showing that ICAR better performs compared to ERA5, which is rather expected concerning the scaling and processes. It would also be important to see the performance of WRF in comparison with ICAR so that one can be sure that ICAR is superior to WRF and it is worth to do such a downscaling process even though it is rather preferable compared to fine scale WRF simulation. This is also valid for the comments on reproduction of snowpack by ICARand ERA5.

Authors: It is mostly true that there is an expected improvement of the ICAR simulation compared with ERA5 as consequence of its higher resolution. This is obvious concerning the temperature because of the smoothing of the topography caused by the coarser resolution of ERA5 while it is necessary to remark that there is an added value of using ICAR for simulating the precipitation which is a much more uncertain variable. The main objective of ICAR was to provide the forcing for the data assimilation scheme that is described later. An intercomparison between ICAR, ERA5 and WRF is a valuable exercise that should be carried out, but it falls outside of the scope of the current manuscript. In our study the comparison between ICAR and the automatic weather stations is performed mostly to define the parameters of the prior probability distribution functions used to perturb the members of FSM2 ensemble of simulations, as is underlined in the text:
"This validation provides a range of uncertainty estimates to generate the probability density functions for the perturbations of the ensemble"
An intercomparison of models should be developed over much more well instrumented areas. We do not consider it appropriate o develop an in depth intercomparison herein, as the results could be extremely constricted by the very low availability of data. In the case of snow, ERA5 is not able to even simulate any snowpack for our domain (as highlighted in the text) as a consequence of the coarse resolution. WRF simulation is able to simulate a very marginal snowpack, as mount Lebanon is too small to reproduce the snowpack at 10 km spatial resolution. Thus, the comparison between models will not show similar results, but this does not mean that each model is not working as expected.

Ref.1: It is better to include topographical and climatological characteristics of AWS (e.g. altitude, aspect, annual average values etc) and comment on these since there are differences in comparison results (e.g. the error difference is less (Figure 2) in the second AWS, it could be assigned to the topographic similarity or just the short period of comparison, but the errors are rather high for the same station in precipitation comparison). This would also be helpful for SWE comparisons in Figure 4.

Authors: We have included a new table summarizing the topographical characteristics of the automatic weather stations and the pixel elevation of ICAR.

Ref.1: The comparison in observed and simulated SWE values is very valuable and worth further discussion. The authors give some details on the inconsistency of comparisons in the third AWS for 2011/2012 which indicates that the observed SWE values might have rather higher values. On the other hand, this inconsistency is also valid for independent snow cover comparison in Figure 5 for the same year, which might indicate some other problems for that year. The consistencies are rather high for the first AWS may be due to its higher altitude, however especially for the second AWS, neither ICAR nor ICAR_assim provides a good performance, for the third AWS, there are varying comparison results and the scale of SWE

(due to extreme value in 2011/2012)makes the graphic rather difficult to interpret. Questions arise on the differences in ICAR and ICAR_assim; assimilation process changes ICAR results dramatically in some years (the second AWS, both years but especially 2015/2016; the third AWS,2010/2011, 2013/2014) while not much for the other years. In some years, assimilation yields significant amount of SWE from almost no snow condition (e.g, the third AWS,first year). On the other hand, assimilation shows very well performance on the first AWS for 2014/2015. Would it be possible to give some explanations on such a big andvarying impact of assimilation?

Authors: There is a very big difference of scale between the ICAR/ICAR_assim simulations and the point-scale AWS observations. Much of the inconsistencies could be explained by this scale mismatch, as the snowpack varies at much finer resolutions at the local scale of the AWS's as explained in paragraph 4.2 Fractional snow cover assimilation.
We hypothesize that The 2011/2012 inconsistency between ICAR_assim and MODIS gap-filled snow cover extent could be explained by the fact that MODIS gapfilled products will be biased during the snow seasons with persistent cloud covers (as the 2011/2012 season), as the gapfilling algorithm will have just a few observations. Thus, the ICAR_assim snow cover exhibits higher values than MODIS as consequence of the low elevation snowpacks. While MODIS gapfilled products will not be able to detect properly such low elevation and very variable snow covers as a consequence of the cloud cover, the particle batch smoother is able to propagate the few fSCA observations through the whole season. However, it is very surprising that the independent observations of Koeniger et al., 2017 highlight the extraordinary snowpack accumulations of the 2011/2012 snow season (as can be observed in the ICAR_assim reanalyses), while it is not observed in the MODIS gapfilled products. To improve the discussion about this topic, we have added the following sentences to the manuscript.
"In addition, the MODIS snow cover products should be considered less accurate over areas of fast melting (Gascoin et al., 2015). Such effect combined with the fact that 2011/2012 snow season showed persistent cloud covers related with its exceptional snowpack, could explain the biases in the Figure 5 2011/2012 snow season, as the gapfilling algorithm had less information to fill the MODIS snow cover time series, while the PBS had propagated the fSCA information through the whole season from the few available observations."
The inconsistencies observed between the AWS and ICAR_assim, are similar to those found in Fayad and Gascoin (2020) using the MICROMET + SNOWMODEL framework. They found that it was not trivial to simulate the snowpack at the AWS locations, even using meteorological observational data from the AWS itself. They hypothesize that the inconsistencies could be related to the partitioning of the precipitation phase, because of the relatively warm conditions close to the 0ºC. In addition, some local effects are probably affecting the AWS data, but unfortunately there is not enough information to study such effects and the inconsistencies should be considered as part of the total uncertainty. Actually, the ICAR_assim and AWS SWE comparison (as for any other grided numerical model) should be taken with care, as the ICAR_assim represents an averaged region (i.e. model grid cell). Thus, the good results showed on the first AWS for 2014/2015 snow season could be completely different if the AWS were at a different location just few meters way, as reported by a manual inspection by Fayad and Gascoin (2020) at 15 of January 2016, cited in the manuscript as follows
"For example, Fayad and Gascoin (2020), reported large differences with the AWS data from insitu measurements on 15 of January 2016, when they measured snow depths up to 258 cm on the surroundings of the third AWS location (Figure 4; bottom panel), while the AWS sensor itself detected 7.5 cm."

Ref1: Concerning a rather constant (or slightly decreasing) relative area in Figure 9 and rather constant SWE values above 2500 m a.s.l. in Figure 8, it is surprising to see an increase in total water storage at 2800 m a.s.l. so it would be nice to give attention to this part.

Authors: Such an increment is caused by the accumulated surface over 2800 masl, combined with the very high values of SWE at the higher elevations. Actually, in the figures 9 and 10 the relative area above 2800 masl is slightly higher than the previous elevations. We have modified the figures 9 and 10 to include the label >2800 to clarify this.

Specific comments:

Ref1: Since ICAR_assim is already produced by assimilating MODIS through ICAR, the comparison in Figure 5 might include ICAR directly instead of ICAR_assim and/or more statistical results can be given on both.

Authors: Figure 5 was designed to show the performance of the PBS, that is why we consider it is better to show ICAR_assim snow cover extent compared with the MODIS gapfilled snow cover extent. To highlight the improvement of the performance after the PBS implementation we have added the statistics of ICAR compared to MODIS gapfilled products.

Ref1: *In section 4.3, the time period (2010-2017?) should be indicated instead of "recent years" for the explanation of Figure 7.*

Authors: Change accepted, thanks.

Ref1: In paragraph with code "515" there is a repetition for two sentences which should be avoided.

Authors: Corrected, thanks