

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 #2

Ref.2: Line 45: Could the authors be more precise with deep and long lasting snowpacks? How deep and how long? Are those common characteristics of snowpack in all Mediterranean mountains?

Authors: There is not an obvious answer to these questions. The depth and duration of the snowpack are often related to the elevation. As is stated in the text, the Mediterranean snowpacks are deep and persistent, as long as the range has a sufficient elevation, due to the wintertime distribution of the precipitation of Mediterranean climates. The references supporting such a statement (Alonso-González et al., 2020; Fayad et al., 2017b) are much more extensive with several data at different locations and elevations. We have added to the following to the text:

“...snowpacks accumulating more than 3 meters and lasting more than 5 months at the summit areas”

Ref.2: Line 47 and 48 and Line 51: Does the interannual variability referred in Line 51 any influence in the reshape of the hydrographs? I might think yes.

Authors: Yes it does. We have modified the text as follows:

“Mediterranean snowpacks are characterized by a high interannual variability, which affect the amount and seasonality of river flows”

Ref.2: Lines 100-102: There are many works in combining numerical modelling and remote sensing using data assimilation techniques. Please try to be less categorical in your statement.

Authors: We have removed the following sentence:

“However, less often, numerical modeling and remote sensing have been combined in a data assimilation framework to study the multiyear snowpack dynamics.”

Ref.2: Lines 107-126: I miss the aim of the research in this paragraph, what is relevant scientific question addressed by this work? The paragraph seems more a summary of a methodological section.

Authors: We have added the following sentence to the text:

“The objectives here are: i) to explore the potential of a methodology to develop a snowpack reanalysis over data scarce regions and ii) to describe the main snowpack dynamics over the Lebanese mountains being the first use of ICAR for this approach”

Ref.2: Line 129: Could the authors provide some data or references about this typical Mediter-ranean climatology?

Authors: We have added the following reference to the text:

Peel, M. C., Finlayson, B. L., and McMahon, T. A.: Updated world map of the Köppen-Geiger climate classification, Hydrol. Earth Syst. Sci., 11, 1633–1644, <https://doi.org/10.5194/hess-11-1633-2007>, 2007.

Ref.2: Line 139: The authors mentioned here that the mountainous ranges act as a barrier to humidity advected from the sea. Could you provide some data about physiographic (elevation, slopes, land covers) and meteorological characteristics that differentiate both mountainous ranges?

Authors: We consider that Fig1 describes elevations and slopes and the Jomaa et al. (2019) support such a statement about the orographic precipitation. We have added the following sentence referring to the land cover:

“Lebanese mountains are highly karstified encouraging the infiltration of rainfall and snowmelt. The land cover is mostly composed of bare rocks and soils with irregularly distributed patches of shrubland, oak and pine forest.”

Ref.2: Line 158: Could the authors clarify what are they referring with previously?

Authors: We have change “Previously” by “First”.

Ref.2: Line 162: Why do the authors chose 35 levels and 50hPa?

Authors: This is the regular WRF configuration. As there is limited information in our domain, we had to choose a regular model set up. We have added the following sentence to the text:

“similarly to other studies over Mediterranean climate (Arasa et al (2016))”

Arasa, R. , Porras, I. , Domingo-Dalmau, A. , Picanyol, M. , Codina, B. , González, M. and Piñón, J. (2016) Defining a Standard Methodology to Obtain Optimum WRF Configuration for Operational Forecast: Application over the Port of Huelva (Southern Spain). Atmospheric and Climate Sciences, 6, 329-350. doi: 10.4236/acs.2016.62028.

Ref.2: Line 166: Which version of WRF is been used?

Authors: We have added “3.8 version” to the text.

Ref.2: Line 172-180: The authors justify the use of a specific parameterization schemes in the WRF simulation base on Ikeda et al., (2010) a study performed over Colorado. I can understand similitudes regarding topography of both areas. However, the sizes of both mountainous ranges, the proximity to the sea are different. What are the influences of having choose the same parameterization? Since the lack of data does not allow a deeper analysis, could you explain a bit deeper the physical reasons behind the selection of this atmospheric parameterization?

Authors: As referee.3 highlights there is not a way to test different WRF parameterizations here. Thus, there is a need for assumptions. However, the reason for using an atmospheric model as forcing is actually the lack of observational data allowing to work over areas and times where it is not possible to find any information. There is probably not any physical reason to choose a specific WRF set up. Most of the studies looking for perfect WRF configurations are factorial experiments over well monitored/instrumented areas, as it is not easy to offer physically based explanations about why a particular parameterization performs better than others.

Despite thefact that all the parameterizations used in the WRF simulation (as well as its alternatives) are physically based, there are many empirical components inside them that are impossible to avoid. Thus, we can not justify choosing a parameterization over the region different from what the literature recommends, as all the parameterizations concern physics and at some point over empirical approximations. This is where the importance of the data assimilation becomes crucial, correcting the uncertainty caused by parameterizations and observations, exploiting the strengths and weaknesses of both.

Ref.2: Line 204: Do the authors mean they are not considering any convective process in their simulation? What are the implications?

Authors: Convection can not be represented by the linear theory simplification and therefore by ICAR. The convective schemes of ICAR are highly experimental and in most cases becomes the model unstable making it crash at 1km resolution. The implications for winter precipitation are probably related to the amount of precipitation, but are much less significant than during the summer season. Such effects should be compensated by the PBS if it has some impact on the snowpack. We have added the following clarification to the text:

“The lack of convection could have some impact on the total amount of precipitation, and therefore on the seasonal snowpack. However, such deviations in the total amount of precipitation are compensated by the PBS.”

Ref.2: Line 220: What is the temporal resolution of Theia? And therefore, how many days of overlapping between Theia and MODIS do you have? Is this overlapping constant during the year? Could that introduce errors in the transformation function?

Authors: The revisit period of Sentinel-2 is at least 5 days since the launch of Sentinel-2B (i.e. after march 2017). It can be even less in areas where successive swaths overlap laterally (every 2 and 3 days). As written in the manuscript we have used a total of 645 Sentinel-2 snow images. This corresponds to all available images from 03 Sep 2017 to 24 Dec 2018 over Lebanon (five Sentinel-2 tiles: T36SYB, T36SYC, T36SYD, T37SBT, T37SBU). For every Sentinel-2 image we can match a MODIS image since there is a MODIS image every day over Lebanon during the same period. However, the number of Sentinel-2/MODIS images is a bit misleading since large parts of a single image can be covered by cloud, or correspond to the sea surface. In addition we only extracted Sentinel-2 pixels where MODIS NDSI is strictly positive (i.e. MODIS snow covered pixels) to establish the relationship between MODIS NDSI and Sentinel-2 fSCA. Therefore we think it is more informative to provide the number of pixels that were actually used to optimize the fSCA function (5.84e4). We will clarify this in the text accordingly.

Ref.2: Line 226-228: How do the authors choose this 40% of the data? Why do the authors use a bigger number of data for calibration than for validation? Could the authors show the same errors in the calibration phase to see differences? I think the reader would be interested in see the fitting graph

Authors: In fact we used 40% for calibration (L226), therefore we used a bigger number of data for validation. By using a larger fraction of data for validation we expect to have a more robust estimate of the model accuracy. We will include the graph of the model calibration in supplement of the revised manuscript.

Ref.2: Line 234: The authors mentioned here the difference between revisiting times of Aqua and Sentinel-2, but what is the Terra revisiting time in the area?

Thank you for this comment, it is approximately 10:30 A.M. local time daily, i.e. similar to Sentinel-2. We will add this information in the revised manuscript.

Ref.2: Line 240: When the authors say “empty” are they referring to a non-snow cell or a non-information cell? Why have the authors chosen that option instead of an interpolation with nearest cells?

Authors: We consider it a non-information cell (changed in the text for clarification). The reason for this is to not propagate into the reanalysis information derived from interpolations. As a smoother, the PBS can propagate the information over the whole season (forward and backward in time), with this information being the trajectory of the fSCA the variable that is assimilated. There is not any

added value on including a few more noise cells derived from incomplete observations, especially in a Mediterranean area like Lebanon where persistent cloud cover is not expected.

Ref.2: Line 259: How do the authors apply the FSM2 snow model in a distributed way? In the last paragraph of this section it seems the authors use some depletion curve for that. However, it is not clear if that is just part of the assimilation or plays a role in the actual snow modelling. Could the authors add a sentence in this paragraph specifying how that is done?

Authors: There is not any specific FSM set-up to implement it in a distributed way. What we did was implement the PBS grid cell by grid cell, generating the distributed reanalysis. The subgrid depletion curve was used to translate the grid cell scale FSM outputs to fSCA (within each cell) to make it possible to assimilate the MODIS information. This is the regular way to assimilate fSCA into snow models: independently for each grid cell and snow season. The methodology is explained in Line 290, see also Line 300:

“The PBS was implemented over the fSCA ensemble over each grid cell and season independently”

Ref.2: Line 270: Why do the authors chose a log-normal and a normal gaussian probability density function? Are just precipitation and temperature the inputs/forcing variables of the FSM2 snow model? If they are more than precipitation and temperature, how are you perturbing them in the assimilation scheme?

Authors: We use a lognormally distributed multiplicative parameter to perturb the precipitation and a normally (Gaussian) distributed additive parameter to perturb the air temperature. A lognormal distribution, which only has positive support, is chosen for the multiplicative precipitation perturbation parameter since precipitation can't be negative, while a normal distribution, which has both negative and positive support, is chosen for the additive temperature perturbation parameter to allow for both positive and negative additive perturbations. So, aligned with the Bayesian underpinnings of data assimilation (Wikle and Berliner, 2007), we are selecting the distributions for these uncertain parameters based on physical constraints and prior knowledge. Note that these forcing perturbations are closely in line with previous applications of the PBS for snow reanalysis (e.g. Margulis et al., 2015; Cortes et al., 2015; Fiddes et al., 2019). The energy and mass balances in FSM are driven by standard hydrometeorological forcing variables; i.e. near surface air temperature, wind speed, specific humidity, precipitation, and incoming longwave and shortwave radiation. The reason that we do not perturb more forcing parameters is that by doing so we would enlarge the dimensions of the parameter space which, due to the curse of dimensionality, would make degeneracy more likely with the PBS especially since we are assimilating a relatively large number of independent observations (van Leeuwen 2009; Margulis et al., 2015). Our choice of perturbing precipitation (whose phase is controlled by air temperature) in particular is justified by the fact that precipitation bias is often the key uncertain factor controlling physically-based snow models (Raleigh et al., 2015).

New references:

Wikle and Berliner (2007), A Bayesian Tutorial for Data Assimilation, *Physica D*, <https://doi.org/10.1016/j.physd.2006.09.017>

Raleigh et al. (2015), Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework, *HESS*, <https://doi.org/10.5194/hess-19-3153-2015>

Ref.2: Line 319: How was the snow depth measured? Why not to do the comparison in term of snow depths avoiding to use a constant density value? Reading Essery (2015)FMS2 provides snow depth as an output.

Authors: As explained in the text, we wanted to compare the snow output of the ICAR model directly, which is provided just in terms of SWE.

Ref.2: Lines 328-342: The methodology explained here is not clear. Neither the reasoning behind nor the way SWE is compared against satellite observation. Are you using SWE measuring using remote sensing?

Authors: We have split the paragraph in two different ones at line 330. It helps to clarify this as the SWE comparison is not related with the remote sensing part.

Ref.2: Lines 345-381: I miss numbers supporting the statements throughout the section. Here a few examples: "Figure 2 shows how the ICAR model was able to improve the 2 m air temperature data, compared with ERA5 reanalysis", "ICAR reduces the spread of the daily precipitation errors". Moreover, I think it could be interesting to analyse a deeper when the error between observations and simulations occurs. Are they bigger in winter than in summer? Is there any dependence with the total precipitation of the hydrological year (dry or wet)? It may have large impact in your results.

Authors: We have added numbers to the statements of the section. We agree that a deeper error assessment of ICAR should be done as it is a very new regional atmospheric model under continuous development. However, the mountains of Lebanon are not an appropriate location to do this in due to the limited data availability. As for the suggestions: i) here there is no observed information in summer (Fig. 2 and 3) and ii) it is not possible to define the dry and wet years due to the very short length of the observed series (Fig.3).

Ref.2: Figure 3: In general, ICAR precipitation values seems to be higher than ERA5 precipitation values. However, the bias in ERA5 are positive and bigger than ICAR. How do the authors explain that?

Authors: Exactly, the ICAR precipitation values are higher than ERA5 values, but the difference between ERA5 and the observations is bigger than between ICAR and observations. ERA5 is too dry over the area, likely due to the lack of orographical precipitation as consequence of the smooth of the topography of the ERA5 spatial resolution.

Ref.2: Figure 4: How do the authors explain the heterogeneous differences in the assimilation results between years? The authors gave some explanation about one of the years in Lines 399-410, however, could you give deeper explanations about the differences between years in the whole period?

Authors: Fig.4 as well as Fig3 and 2 should be used with caution as highlighted in the text. There is a big scale mismatch between the point scale AWS information and the reanalysis. The reanalysis seems to perform better at higher elevations (Fig4 A and C), suggesting difficulties to model the precipitation partition phase, as other authors have shown over the region (Line 280). It is not possible to offer any convincing reason about the (mostly small) differences in the performance between years. The snow wind redistribution processes strongly controls the snowdepth at the point scale resolution of the AWS making comparisons complicated. In addition, previous studies have highlighted the very high snow depth spatial variability over the area as remarked in the text.

Ref.2: Lines 421-448 and Figure 5: If I read well, here you are comparing the results of your assimilation (ICAR_assim) with the assimilated variable (Obs). This is a prove that your assimilation scheme works well, and therefore, the obtained metrics should be interpreted as

that. The real impact of the assimilation scheme on snow dynamics is the show in the comparison with the independent variable SWE, not assimilated during the process.

Authors: Exactly, to clarify this point we have added the following to the text:
“...showing the potential of fSCA assimilation through the PBS in improving the ICAR SWE products.”

Ref.2: Section 4.3: All section is written as if the simulated values were a “ground truth”, I would indicate some of the limitations of the performed simulations and all the sources of uncertainty and errors that are conditioning these statements.

Authors: ICAR limitations are already described in lines 349-355. In addition, we have added the following sentences to the section 4.3 :

“ICAR_assim exhibits some limitations that should be considered. First, despite the high resolution of the reanalysis the regional nature of the simulations prevent the representation of some processes like wind or avalanche snow redistribution. In addition, there are some other sources of uncertainty involved in the development of the reanalysis, like the depletion curve, the fSCA derived from MODIS or the structural uncertainty associated with each model. However, ICAR_assim has been shown to be consistent with the few observations providing a valuable resource in the data scarce context of the Lebanese mountains.

Ref.2: Figures 9 and 10: What does the relative area referring to (snow area over the area of the band or area of the band over all area of the mountainous ranges)? It would be interesting to see these two graphs in both mountainous ranges.

Authors: It is already described in the text (Line 501):
“The relative area lying at each elevation compared with the total elevation over 1300 m a.s.l...”
We have added the Lebanon and Anti-Lebanon ranges SWE and accumulated water partition.

Lines 504-508: Could you elaborate more the reasoning in this paragraph?

Authors: We have added the following:
“This suggest that the mean peak SWE series at lower elevations could hide a large variation in mass due to the wider areas at lower elevations *where many different peak SWE values can coexist,...*”

Ref.2: Figure 10: How do you explain that the total storage at 2800 m a.s.l. increases?

Authors: This is because the storage at 2800m a.s.l. integrates all the surface over 2800. We have modified Fig,10 to clarify it.

Technical comments

Ref.2: Figure 1. What are A and B, could the authors specify it in the body text lines 58-62 and in the figure caption.

Authors: Fig1 legend indicates the meaning of A (WRF domain) and B (ICAR domain). We think that in this way it remains clear to readers.
The topic of Lines 58-62 do not match the atmospheric models domains.

Ref.2: A figure with a scheme of the implementation process would help to better understand the complexity of the flow chart followed.

Authors: We have added a schematic flowchart to the Section 3.2.2 summarizing the whole process.

Ref.2: Figures 2, 3 and 4: It is difficult to know in which season of the year you are with the format “Days since”. I propose to add actual dates in x-axis of these figures. Moreover, it is complicate to see differences between the 3 represented variables, especially in the precipitation graph. Finally, it is difficult to see what the values of the boxplot are represented, I would recommend here to change the y-axis limits, add, y-axis values and/or a grid.

Authors: We have added the suggested changes.