Anonymous Referee #2

Reviewer comments are in black, author responses are in *blue Italic*. We will submit a revised manuscript by the end of February 2018. All references to page/line numbers and figures refer to the submitted paper (not the revised manuscript).

This paper presents an evaluation of the ability of the snowpack model Crocus to simulate snow depth and snow cover in southern Norway for two winters using different atmospheric driving data: (i) short-range high resolution weather forecast generated by the AROME MetCoOP system and (ii) gridded datasets for precipitation and temperature derived from observations. The authors propose an evaluation of model results using snow depth observations collected at 30 stations across the simulation domain and MODIS snow cover data at 500-m grid spacing. Daily snow depth variations are considered as in Quéno et al. (2016) to discuss more in details the physical processes responsible for differences between simulated and observed snow depth.

The subject of this paper is interesting for the snow and mountain hydrology community because of the growing use of high-resolution weather forecast to drive snowpack models in mountainous terrain (e. g. Bellaire et al., 2011, 2013; Bernier et al., 2012; Carrera et al., 2009; Horton and Jamieson, 2016; Quéno et al, 2016; Vionnet et al, 2016). The analysis of results presented here is similar to the studies by Quéno et al (2016) and Vionnet et al. (2016) and reveals consistent and interesting model behavior between the French and the Norwegian mountains. My main comments about this study concern (i) the comparison between the different atmospheric driving datasets, (ii) the interpolation of AROME forecast on the 1-km grid used for Crocus simulations, (iii) the selection of stations for model evaluation and its impact on the analysis of model results and (iv) the originality of this work compared to other studies using AROME and Crocus in the French mountains. These questions need to be clarified prior to publication in TC. They are listed below as general comments followed by more specific and technical comments.

Author response: Thank you very much for such a detailed review of our paper - we really appreciate your thoughts, comments and suggestions. We will answer to each comment below, which we have numbered for easier reference. We will submit a revised manuscript (based on the comments from both reviews) by the end of February 2018.

Thanks again for taking the time and effort to review our paper.
Kind regards,
Hanneke Luijting

General comments

1. The comparison between simulated snow depth and snow cover using different precipitation and temperature forcing is interesting and illustrate well the strong impact of these variables on simulated snowpack evolution. However, the authors only present the results of snowpack simulations and never compare for example the precipitation forcing in their two experiments; how they differ for different elevation ranges or distance to the sea. Such comparison would be really useful to better understand the differences obtained in the simulated snowpack evolution. For example, Figure 9 shows that the snow cover remains longer at high-elevation in AROME-Crocus compared to GridObs-Crocus. Is it explained by lower precipitation at

high-altitude in the GridObs forcing compared to the AROME forcing? This is not mentioned in Section 3.3 and in the conclusion and should be added to the paper.

Author response: Thank you very much for this good suggestion. We will focus more on the evaluation of the forcing data set in the revised manuscript. We will include a comparison of the precipitation forcing in the two experiments.

2. AROME forecast at 2.5 km are interpolated bilinearly on the 1-km grid used for Crocus simulation. No downscaling is performed to account for the differences between the interpolated terrain height from the 2.5-km grid and the actual terrain height in the 1- km grid. This can potentially lead to large errors in region of complex terrain. For example, the phase partitioning simulated by the AROME cloud microphysical scheme is only valid at the elevation of the grid cell on the 2.5 km grid. A first order correction using a simple lapse rate is required to adjust the phase if the elevation difference is large. Overall, I recommend the author to include simple terrain adjustment routines in their AROME-Crocus simulation to use a meteorological forcing valid on the 1-km grid where are performed the snowpack simulation. This can be done using very simple methods such as the ones used in Bernier et al. (2011).

   Author response: Thank you very much for the considerations and your recommendations about the methodology. We are very much aware of the uncertainty introduced by downscaling the AROME MetCoOp data to 1 km. During the project we discussed the various uncertainties, and found that as a starting point we carried out the simulations as described here.

   However, we believe your recommendations are very interesting and may be studied and examined in future work and future studies. A 1 km resolution dataset is used because operational snow models used by e.g. the national flood forecast service need this resolution. Still, at this time of the paper revision, it is too comprehensive to rerun the 2 years of Surfex/Crocus simulations with a new forcing dataset. We will however add this topic to the discussion (and suggestions for future work) in the revised manuscript.

3. Section 4.1 shows that the authors include in their analysis stations with large differences between the model and the actual height at station location (up to 450 m). They did not make a selection of stations based on a maximal value for the difference between the model and the actual height. In their studies, Vionnet et al (2016) and Quéno et al (2016) used for example a maximal elevation difference of 150 m in absolute value. In this paper, 13 stations among 30 correspond to this criteria. What is the impact of these large elevation differences on the evaluation of model results? For example, for the stations with an elevation difference above 250 m, what is the impact on the evaluation of snow cover duration? What about the wind speed simulated at these stations and used to determine the occurrence of blowing snow days? As mentioned in my previous comment AROME forcing are only interpolated bilinearly on the 1-km grid. Therefore, altitude differences between the station height and the elevation in the interpolated terrain at 1 km from the 2.5-km grid can be potentially even larger. The authors only mention the elevation differences in the discussion (Section 4.1). I think this should be mentioned earlier in the paper; for example in Section 2.3.1 when presenting the snow depth observations. Overall, the effects of these large elevation differences should be better quantified.

   Author response: Thank you for your comment. We used nearly all available snow depth observations from eklima.met.no for our area (except for a few stations that had large gaps in the *snow depth data during 2014-2016). If we had been more strict with for example the maximum height difference between the model and the actual height, we would have had too few stations left for the validation analysis, like you already mentioned. We will add a discussion of this issue to section 2.3.1, and we will calculate statistics for the stations with smaller height difference to be able to say something about the impact of the large elevation differences on our results, and add the results to our revised manuscript.*

4. The simulation framework and evaluation methods presented in this paper are very similar to the ones used by Vionnet et al. (2016) and Quéno et al. (2016) who used AROME to drive Crocus snowpack simulations in the French Alps and the Pyrenees. It is interesting to see that similar results are obtained in a different mountainous environment. However, the author need to better insist on the originality of their study compared to these previous work.

   Author response: We are of the opinion that the originality of our work lies in both the evaluation of the two forcing datasets, and the use of the gridded observations of hourly precipitation and temperature for snow modeling. This gridded dataset has been developed very recently (see Lussana et al, 2017 and 2018, full references below). Most operational snow models for hydrological forecasting in Norway use daily data of precipitation and temperature, while this study was done with hourly data. This is the reason why it is a very interesting dataset to study for hydrological users in Norway. We think that evaluation of using gridded observations with a a temporal resolution of 1 hour and a spatial resolution of 1 km to force SURFEX/Crocus is both interesting and original.

   We will include a section in the introduction of the revised manuscript that better explains the originality of our work compared to previous work, including work done by Vionnet et al. (2016) and Quéno et al. (2016).

   References to papers mentioned:

   Lussana, C., Tveito, O., and Uboldi, F.: Three-dimensional spatial interpolation of two-meter temperature over Norway, Quarterly Journal of the Royal Meteorological Society, https://doi.org/10.1002/qj.3208, http://dx.doi.org/10.1002/qj.3208, qJ-17-0046.R2, Accepted Author Manuscript, 2017

   Lussana, C., Saloranta, T., Skaugen, T., Magnussson, J., Tveito, O. E., and Andersen, J.: seNorge2 daily precipitation, an observational gridded dataset over Norway from 1957 to present days, Earth System Science Data, accepted for publication, 2018

5. It would have been interesting to see additional experiments. For example the authors used a succession of forecast from +3 to +8 to drive Crocus. Vionnet et al. (2016) and Queno et al (2016) combined daily forecast from +6 to +29 issued at 00 UTC to drive Crocus. The impact of these choices on model results has never been discussed and it would be an interesting contribution.

   Author response: We chose to use +3 to +8 to avoid the first hours of spinup of the model, while making use of all available model runs. We believe this is an advantage compared to using +6 to +29 from only the 00 UTC run. Model errors increase with lead time, and our aim was to use the best available model data. A study on the impact of using different lead times from an *atmospheric model to force SURFEX/Crocus would certainly be interesting, but this is beyond the scope of this paper. Our aim was to compare the two forcing datasets. Lead time comparisons would only be relevant for the AROME-Crocus experiment and not for GridObs-Crocus. The gridded observation dataset has a temporal resolution of one hour, and for a fair comparison, we used the best available forcing data from the AROME-MetCoOp model, and this meant using all 4 daily model runs with the shortest possible lead times.*

Changes in manuscript: the following sentence has been added to 2.2.1: "These lead times were chosen to avoid the first hours of a cycle when the model might have spin-up issues, and to make use of all available cycles with the shortest possible lead time (since model error increases with lead time, see for example Homleid and Tveter (2016))."

6. Similarly, the authors mentioned at the end of their paper (P 19 L7-8) the potential importance of blowing snow sublimation for the high-altitude part of their domain. I recommend the authors to carry out an experiment where they test the impact of the parametrization of sublimation loss during blowing snow events implemented in Crocus. The authors could discuss the impact in terms of snow cover duration and compare it with MODIS images. Overall, these additional experiments would bring interesting insights and strengthen the discussion section which is so far very similar to the discussions in Vionnet et al. (2016) and Quéno et al. (2016).

   Author response: Thank you for this suggestion. We will perform a 1D experiment that tests the impact of the parameterization of sublimation loss during snow events, and discuss the results of this experiment in the revised manuscript.

Specific comments

7. **Abstract**: The abstract is rather vague and should present some precise figures such as the overall snow depth bias for the two experiments and the number of stations used for model evaluation. A L11-13, the authors mention the assimilation of snow depth data directly into Crocus. This topic is mentioned here but never discussed in the paper. If the authors want to keep this sentence in the abstract, they need at least to discuss more the assimilation of punctual snow depth data in distributed snowpack simulations in the discussion part.

   Author response: Thank you for your comment. We have included the overall bias and RMSE of the two experiments to the abstract, and have added the number of stations used for the evaluation. Concerning the assimilation of snow depth, we actually do briefly discuss this topic in the first paragraph of the discussions chapter (page 16, line 5). Our argument is that when errors accumulate during the snow season (due to the overestimation of snow in AROME-Crocus), one solution would be to assimilate observed snow depth into Crocus. We agree however that this topic does not belong in the abstract, and have changed the abstract to reflect this.

8. **Introduction**: The current introduction of the paper does not described well enough the context of the study and the scientific questions the authors are investigating. For example, the authors never mention the growing use of high resolution NWP forecast to drive detailed snowpack model in mountainous terrain and the limitations associated with these systems. In particular, previous studies using AROME forecasts to drive Crocus in the French Alps and the Pyrenees (Quéno et al., 2016; Vionnet et al., 2016) are not mentioned in the introduction. Similar studies using other models have also been carried out and are not mentioned in the text. For example, the work done

by Bellaire et al. (2011, 2013) and Jamieson and Horton (2015) with the Canadian GEM model to drive the detailed snowpack model SNOWPACK. The authors should mention in the introduction how their work differs from these previous studies and what is their contribution to this field of mountain snow research.

Author response: Thank you for your comments. We will improve the introduction in the revised manuscript, taking into account your suggestions and discussing previous studies. See also our answer to comment 4 in this review regarding the originality of our work and our contribution to mountain snow research.

9. P 2 L 21: what are the reasons behind the selection of the simulation domain in South Norway? Hydropower forecasting? Avalanche hazard forecasting?

   Author response: We have replaced the sentence "We selected a west-east transect in a mountainous area of South Norway as the study area." with a more detailed explanation of the choice of domain: "The evaluation was done as a part of several research projects within hydropower and flood forecasting. The domain was chosen to cover the mountains in southern Norway and to include a cross-section from west to east that crosses the watershed in this region. The domain also includes several catchment areas that are of interest to hydropower companies."

10. Section 2.1: The description of the configuration of Crocus and SURFEX should be more specific. For example, the following points should be clarified: - how many layers are used in the soil models? - how are determined the soil and surface properties (clay and sand fraction, vegetation type, . . .)? - how large is the simulation domain in km and grid points? - how are initialized the soil and snowpack properties (if any snow is present) on 1st September 2014? Did the authors perform a model spin-up?

    Author response: Thank you for this helpful comment. We will improve the description of the configuration of our SURFEX/Crocus simulations by including the details you mention, and clarify that no snow is present on 1st of September 2014 (as well as on 1st September 2015).

11. P 5 L9-10: how many stations are used to generate the gridded precipitation and temperature products in the region? In particular, are these stations covering a similar altitudinal range compared to the stations used for snow depth evaluation?

    Author response: The number of stations that are part of the dataset is variable with time (new stations are added, sometimes stations are closed down). The number of stations for out SURFEX/Crocus domain: 20-30 stations for hourly precipitation, 90-100 stations for daily precipitation and 70-100 stations for temperature. These stations do cover a similar altitudinal range compared to the stations used in the snow depth elevation: for precipitation the highest station is at 1210 masl., for temperature there are higher stations available with the highest at 1390 masl.

    The following text has been added to 2.2.2: "The number of stations that are included in the gridded dataset is not constant (new stations are added, sometimes stations are closed down). The numbers of stations within the SURFEX/Crocus domain are: 20-30 stations for hourly precipitation, 90-100 stations for daily precipitation and 70-100 stations for temperature. Stations *just outside the domain are included in this estimate as they are used in the interpolation and are therefore part of the gridded dataset used in this study. "*

12. P 6 L 6: on Fig. 1, it seems that the stations are not covering the area of high elevation of the simulation domain. To illustrate this, point, I recommend the author to add on Fig. 2 the histogram of the distribution of elevation in the simulation domain.

    Author response: This is a well known situation in Norway that most weather stations are located at low elevations (at the bottom of valleys), and there are too few stations high up in the mountains. We have included all the available high quality stations that are located within the domain and observe snow depth, from eklima.met.no (a few stations were discarded due to snow depth observations missing for a long period of time within the two years of this study). We will add a histogram of the model elevations to figure 2 in the revised manuscript, and discuss this topic in more detail in section 2.3.1.

13. P 6 L 17: are MODIS snow cover data not available for winter 2015/2016? It would be interesting to compare the evolution of simulated and observed snow cover for this winter as well to see if model results are consistent in between the two winters.

    Author response: Thank you for your suggestion. Unfortunately, the processed MODIS snow cover images for 2015/2016 are not available at this point. It would indeed be interesting to do the same comparison for a second winter season, if images had been available.

14. P 7 L 10-11: where are located the stations used to illustrate model performance? It would be interesting to see their location on Fig. 1. In particular, it would be interesting to see their locations along the West-East transect.

    Author response: Thank you for this very good suggestion. We have added the locations of the stations used in Fig. 4 with a blue color and an indication for the name of the station in Fig. 1. This shows that the 6 stations are quite evenly spread over the domain from west to east. The caption of Fig. 4 has been changed to point to Fig 1. for the locations of the 6 stations, and the same has been done with the text in 3.1 which refers to the 6 stations.

    Indeed, we can expect significant differences in terms of precipitation amount and resulting snow accumulation between the western and the eastern side of the domain due to the proximity with the ocean. In this context, elevation is not the only variable that can explain differences of snow depth from one station to another.

    Author response: This is correct, and the climatology of Norway means a lot more precipitation falls on the western part of the watershed than on the eastern side. We will add climatology information to the study area description to clarify this.

15. P 9 L 7-10: differences of snow depth between GridObs–Crocus and AROME-Crocus are low at Hemsedal II. To support their statement on the best results of GridObs– Crocus compared to AROME-Crocus at this station, I recommend the author to compute bias and RMSE of snow depth at this station for the two simulations.

Author response: Thank you for this good suggestion. The bias at Hemsedal II for GridObs-Crocus is 25 cm vs 30 cm for AROME-Crocus. The RMSE at Hemsedal II for GridObs-Crocus is 27 cm and the RMSE for AROME-Crocus is 33 cm. These differences are not very large, but GridObs-Crocus does perform better than AROME-Crocus at this station. It is also worth keeping in mind that the overall RMSE for GridObs-Crocus (for all stations and for the 2 years combined) is 28 cm (bias: 6 cm) and for AROME-Crocus 68 cm (bias: 42 cm). This means that Hemsedal II is performing much better than most stations in AROME-Crocus. The interesting part is that precipitation is not measured at Hemsedal II, and therefore this station is more representative for the performance of GridObs-Crocus outside the stations that are part of the gridded precipitation forcing dataset. It is therefore interesting that GridObs-Crocus still performs better than AROME-Crocus at this location.

Changes in the manuscript: The sentences "GridObs-Crocus overestimates the snow depth at Hemsedal II, but not to the same extent as AROME-Crocus does. GridObs-Crocus matches the observed pattern of increases and decreases more closely than AROME-Crocus." have been removed.
Instead the following text has been added:
"The bias in snow depth at Hemsedal II for the two seasons combined is 25 cm for GridObs-Crocus (RMSE: 27 cm) and 30 cm for AROME-Crocus (RMSE: 33 cm). When compared to the bias (6 cm for GridObs-Crocus and 42 cm for AROME-Crocus) and RMSE (28 cm for GridObs-Crocus and 68 cm for AROME-Crocus) for all stations for the two seasons combined, it shows that Hemsedal II performs better than most stations in AROME-Crocus. For GridObs-Crocus, the bias at Hemsedal II is larger than at most stations, while the RMSE is slightly better. The fact that GridObs-Crocus still outperforms AROME-Crocus even at a station that is not part of the gridded observation dataset is interesting."

16. P 9 L 12-13: what are the reasons behind this under-estimation of temperature? Is it associated to a large difference between the model and the actual terrain height at station location? Can the authors justified that this underestimation is responsible for an overestimation of the proportion of precipitation falling as snow? From my experience, NWP model can present a negative bias of temperature during clear nights in wintertime. However, this bias does not affect the phase of precipitation during precipitation events characterized by overcast conditions.

    Author response: Thank you for your comment and thoughts on this issue. Midtstova is located at 1297 m in SURFEX/Crocus (this information is now included in figure 4, see comment 32), which is a difference of 135 m with the actual height (1162 m). This is not a very large difference compared to the other stations, so we do not think the underestimation of temperature is related mainly to this difference in height. We will investigate the role of the underestimated temperature on the overestimated snow depth, and discuss this further in the revised version of the manuscript. We will make a case study about Midtstova to further investigate this issue, see also our reply to comment 17 below.

17. P 9 L 17: the beginning of winter 2015 at Midtsova is interesting and shows a net underestimation of snow depth by GridObs–Crocus. Is it associated with an underestimation of precipitation in the GridObs or with errors in the phase of precipitation?

Author response: Thank you for your suggestion. This is indeed an interesting episode to investigate further. In the revised manuscript, we will "zoom in" to this episode at Midtstova and investigate what's going on in the forcing data (precipitation, temperature/precipitation phase).

18. P 9 L 29-30: it is surprising to see that the authors have selected a category that cannot be used to classify observations ([-0.5 0.5] cm). If the snow depth does not change from one day to another, what is the corresponding category? I recommend the author to use a central category that can be used to classify both simulation and observation.

    Author response: This was originally chosen because when we use a category of ([-1 1] cm, a lot of cases fall within this category which then dominates the plot. We agree however that it is not practical to use a category that excludes observations. We will change the central category and include a new figure 5 to the revised manuscript.

19. P 11 Table 2: it would be interesting to see the RMSE for the different variables as well. Maybe make two tables if the number of information is too large.

    Author response: We have added a new table (table 3) which summarizes the RMSE for the different variables. Note that the layout of table 2 has changed in response to a comment from reviewer #1, and therefore table 2 was changed in the same way.

    Changes in manuscript: table 3 was added, and the following text was added to section 3.1: "Table 3 summarizes the RMSE over all stations for the two winter seasons. The RMSE values are significantly larger for AROME-Crocus (compared to GridObs-Crocus) for nearly all variables, except for the date of maximum snow depth for 2015-2016. "

20. P 11 L6 : Quéno et al. (2016) used the same criteria to define blowing snow days but they used the wind speed measured at the stations instead of the wind speed in the atmospheric forcing. Can the author comment on this choice? How accurate is the forecast wind speed for the different stations used in this study? If the wind speed is measured at some stations measuring snow depth as well, it would be interesting to compare the occurrence of blowing snow days with the two wind data to make sure that forecast wind speed can be used to determine the occurrence of blowing snow days.

    Author response: We chose to use model data because only 6 out of the 30 stations measure wind speed. The bias of the forecasted maximum wind speed is 0.3 m/s, which means a slight overestimation of the maximum wind speed by the model. When comparing blowing snow days derived from the observed wind speed at those 6 stations with the forecasted wind speeds from the same 6 stations we find that the model is correct in 94% of the cases (for blowing snow days and non-blowing snow days), with a hit rate of 0.86 (correctly identifying blowing snow days) and a false alarm rate of 0.04 (model data indicates a blowing snow day while observations don't). From this, we conclude that we can use the forecasted wind speed to determine the occurrence of blowing snow days.

    Changed in the manuscript: we've added the following text to section 3.1 : "The modeled wind speed is used because only 6 out of 30 stations used in this study observe wind speed. When comparing the forecasted maximum wind speed from AROME-MetCoOp with the observed maximum wind speed from these 6 stations, we find a slight overestimation by AROME-MetCoOp *(a bias of 0.3 m/s). Blowing snow days and non-blowing snow days are correctly identified in 94% of all days, with a hit rate of 0.86 and a false alarm rate of 0.04."*

21. P 13 L 20-34: the visual comparison of snow cover patterns proposed on Fig. 9 is useful but it should be complemented by a more quantitative analysis. The author could for example compare the temporal evolution of snow cover area in the observations and in the simulations across different altitudinal bands. Similarity metrics such as the Jaccard index or the confusion matrices could be computed as done in previous studies (e.g. Gascoin et al., 2015; Quéno et al., 2016).

    Author response: Thank you for your helpful suggestion. We will perform a quantitative analysis on the snow cover patterns, which will be included and discussed in the revised manuscript.

Technical comments

Text

22. P 1 L 24: remove parenthesis around Bokhorst et al. (2016)

    Author response: Parenthesis are removed from this reference.

23. P2 L 20, L31: the correct reference for the Crocus paper is Vionnet et al. (2012). The authors should refer to the final version of the paper and not the discussion version.

    Author response: The reference has been corrected to the final version from 2012.

24. P 3 L3 and throughout the rest of the paper: units should be written kg m-2 instead of kg/m2.

    Author response: This has been changed throughout the paper, also for other units and for example for table 1.

25. P 4 L 16-17: from the 1800 UTC analysis time, are the authors using the 3-8h lead time or the 3-5h lead time? It is not clear since they mention that they use the 0-8 lead time for the 0000 UTC cycle.

    Author response: Thank you for pointing this out. From the 18 UTC analysis time we indeed use the 3-5 hour lead time.

    The sentence has been changed to "Forcing for our study is taken from the 4 main cycles, with successive 3-8h lead time (0-8h lead time for the 0000 UTC cycle, and 3-5h lead time for the 1800 UTC cycle) forecasts combined into a forcing file for each day."

26. P 9 L 4: "Episodes when" instead of "episodes where".

    Author response: Changed to "Episodes when"

27. P 9 L 33: " the transport of blowing snow or wind-induced ablation" is not clear and should be rewritten. Maybe : "SURFEX/Crocus in stand-alone mode does not account for wind-induced snow redistribution".

Author response: Thank you for your suggestion. The sentence has been changed to: "SURFEX/Crocus in stand-alone mode does not account for wind-induced snow redistribution, which can be a large contributor to strong decreases in snow depth."

28. P 12 L2: when snow is present on the ground, maximal surface temperature is 0◦C. Please remove "or above".

    Author response: Changed to "Melting snow days are defined as days when the surface temperature of the snow is 0 °Celsius."

29. P 18 L 21: incoming longwave and shortwave radiations are also a key component of the snowpack evolution. Therefore, I recommend the authors to remove the sentence "the two most important variables for snow modeling".

    Author response: Changed to "important variables for snow modeling"

Figure

30. Figure 1: the name of all the cities on the snapshots from Google Maps are not easy to read and may be removed. Google Maps may not be the most relevant background map.

    Author response: Thank you for your comment. We have added the names of Oslo and Bergen in bigger font on the map. The elevation map is more relevant than the Google overview map, and we have changed figure 1 to better reflect this: the elevation map is now much larger than the Google map. As a quick overview of where the domain is located in Norway we believe the small Google overview map is sufficient.

31. Figure 3: the axis labels and the legend are too small and hard to read.

    Author response: The figure has been changed so that the axis labels and legend are much easier to read.

32. Figure 4: it would be very interesting to know the elevation of the model grid point corresponding to the station location. Such information is really relevant to analyse model results (see the general comment on this particular point).

    Author response: The elevation of the model grid point has been added to the figure, in parentheses after the station elevation. The caption of the figure has been modified to reflect this: "The altitude of the station is indicated above each plot, with in parentheses the elevation of the grid point in SURFEX/Crocus."

33. Figure 6 and 7: the size of the markers (squares and diamonds) and of text (legend, axis labels, . . .) is too small on these figures.

Author response: We have changed the size of the markers, and increased the font size of all text for both figures. The color of the dark blue has also been changed to a lighter blue, in response to a comment by reviewer #1 that the two colors (black and dark blue) were hard to distinguish.

References (not included in the initial manuscript):

34. Carrera, M. L., Bélair, S., Fortin, V., Bilodeau, B., Charpentier, D., & Doré, I. (2010). Evaluation of snowpack simulations over the Canadian Rockies with an experimental hydrometeorological modeling system. Journal of Hydrometeorology, 11(5), 1123- 1140.

Bernier, N.B., S. Bélair, B. Bilodeau, and L. Tong, 2011: Near-Surface and Land Surface Forecast System of the Vancouver 2010 Winter Olympic and Paralympic Games. J. Hydrometeor., 12, 508–530, https://doi.org/10.1175/2011JHM1250.1

Gascoin, S., Hagolle, O., Huc, M., Jarlan, L., Dejoux, J.-F., Szczypta, C., Marti, R., and Sánchez, R.: A snow cover climatology for the Pyrenees from MODIS snow products, Hydrol. Earth Syst. Sci., 19, 2337-2351, https://doi.org/10.5194/hess-19-2337-2015, 2015.

Author response: Thank you for your suggestions. We will add these references to the revised manuscript.