Anonymous Referee #1

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 article, entitled "Forcing the SURFACE/Crocus snow model with continued hourly meteorological forecasts and gridded observations in southern Norway", has for its main objective to improve numerical prediction of terrestrial snow in Norway. The approach that is presented and examined is based on the used of an external land surface system, SURFEX, that includes the sophisticated snow model CROCUS, and which is driven by a combination of surface analyses (for precipitation and air temperature) and of short-range numerical atmospheric prediction.

This subject is certainly of interest. Better analysis and prediction of terrestrial snow remains one of the great challenges that national environmental prediction centers face. But unfortunately the scope and ambitions of this study are insufficient, in the sense that this kind of work has often been presented in previous papers (which are not cited), and are actually used Operations in several centers. While the results analysis is very well done, with interesting metrics and with interesting discussion, the overall impact (positive) is not Earth shattering.

Unless the authors come up with a major overhaul of the article, maybe based on some of the more specific comments below, this paper should be rejected.

Author response: Thank you for taking the time to review our paper, it is very much appreciated. The manuscript will be greatly improved thanks to the comments and suggestions from the 2 reviewers.

We will address the issues you mention and improve the introduction to discuss previous papers on this topic. 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 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 the use of gridded observations with a temporal resolution of 1 hour and a spatial resolution of 1 km to force SURFEX/Crocus is both interesting and original.

We will answer to the specific comments below (which we have numbered for easier reference), and 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

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

Specific comments: ————————

1. Page 2, line 25: The main objective of the paper is stated as: "The aim was to compare and combine different forcing data sets as input to the SURFEX/Crocus model and validate the computed accumulated snow amounts and snow melt pattern in both Norwegian mountains and lowlands." The authors have to realize that this type of work has been done often. Practically every snow scheme that has been developed in the last few decades have been tested at observational surface stations using atmospheric forcing that are not unlike what is used in this study. Many land data assimilation systems also use snow models forced with a mixture of observations and model predictions to produce terrestrial snow analyses. In order to make this article acceptable for publication, the authors need to describe and include these previous studies in their Introduction, and explain how exactly what they have done contributes in an original manner to advancing this body of research.

   Author response: Thank you for your suggestion. The introduction will be rewritten and improved in the revised manuscript, following your suggestions. We will include a section in the introduction that better explains the originality of our work compared to previous work, which lies in the evaluation of the two datasets against each other, and the use of gridded observations with a temporal resolution of 1 hour. See also our answer above, and to comments 4 and 8 of review #2.

2. Page 4, Table 1: Was there any adaptation applied to the atmospheric forcing, e.g., for air temperature, or precipitation phase? My understanding that the atmospheric model that provided some of the forcings was at lower horizontal resolution than the external land surface model. There might be some inconsistencies in terrain height (between model and reality) that could lead to biases.

   Author response: Thank you for your comment. It is correct that the atmospheric model has a lower resolution (2.5 km) than SURFEX/Crocus (1 km). The forecasts from the atmospheric model were interpolated to a 1 km grid using bilinear interpolation. No downscaling or corrections have been performed to account for the difference in terrain height. 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 carry out the simulations as described here.

   We will add this topic to the discussion (and suggestions for future work) in the revised manuscript. See also our answer to review #2, comment 2.

3. Page 7, line 8: "... is in good agreement with the observations, ..." I don't know how we could say that from Fig. 3. More generally, more discussion is needed to correctly present and understand what is shown in that figure. For instance, it seems there is a significant bias for snow depth

below 100 cm. Also, it might be better to show plot the density of points rather than just the cloud of points.

Author response: Thank you for your helpful comment. We will add more discussion about what is shown in figure 3 to the revised manuscript. We also agree that figure 3 could be improved, and we will investigate if a density scatter plot(s) would be an improvement.

4. Page 9, line 9: "... matches the observed pattern of increases and decreases more closely than AROME-Crocus." I disagree with that statement... or I would say instead that this improvement is quite marginal. Is it what we should expect in terms of impact for points that do not benefit of having a surface observational station?

   Author response: To support this statement we have calculated the bias and the RMSE for Hemsedal II: the bias is 25 cm for GridObs-Crocus (RMSE 27 cm) and 30 cm for AROME-Crocus (RMSE: 33 cm). This shows that GridObs-Crocus does perform better than AROME-Crocus at this location, but indeed not by a very large margin. However, Hemsedal II is performing much better than most stations in AROME-Crocus (overall bias for all stations for AROME-Crocus is 42 cm, and RMSE 68 cm). The fact that GridObs-Crocus still outperforms AROME-Crocus also at points that do not benefit of having precipitation measurements (and moreover, a point that performs better than most in AROME-Crocus) is interesting. See also our response to comment 15 from reviewer 2.

   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."

5. Page 16, line 1: "The results are promising." Too vague.

   Author response: We have removed "The results are promising" from the first sentence, which now reads "Although both experiments are capable of simulating the snow pack over the two winter seasons, there is an overestimation of snow depth in the AROME-Crocus experiment…" (see also our reply to the next question). The same sentence ("The results are promising") has also been removed from the abstract, see question nr 12 under "minor comments".

6. Page 16, line 1: "Both experiments are capable of simulating the snow pack over the two winter seasons." Based on current and recent scientific and technological achievements in this research area, should we consider this an achievement worth being presented as a conclusion?

   Author response: Thank you for your comment. This sentence is not meant to be presented as a conclusion (it does not return in the Conclusions chapter either), merely as an introductory sentence for discussing the results in more detail. We first state the obvious, before going into more details about the positive and negative results from both experiments.

   To make it clearer that it is not meant as a conclusion, we have changed the sentence as follows: "Although both experiments are capable of simulating the snow pack over the two winter seasons, there is an overestimation of snow depth in the AROME-Crocus experiment…"

7. Page 16, line 16: "... it could still be argued that the GridObs-Crocus are best in locations of the observations that are included in the gridded dataset used to force SURFACE/Crocus." This statement by the authors is not substantiated by evidence in this paper.

   Author response: Thank you for your comment. This sentence was not meant as a statement that we have evidence for, it was meant to show that we are aware that we might be criticized for evaluating GridObs-Crocus using stations that were part of the gridded precipitation observations used to force GridObs-Crocus. One way to investigate this issue is to have a closer look at the results for Hemsedal II, the only station not included in the gridded precipitation dataset - see our answer to comment 4 (as well as to comment 15 from reviewer 2).

   Changes in the manuscript: We have added the following text to the statement: "The only station that was not part of the gridded precipitation dataset is Hemsedal II. The RMSE for GridObs-Crocus at Hemsedal II is 27 cm, about the same as the overall RMSE for all stations for GridObs-Crocus (28 cm). Although this shows that the performance of a station not included in the gridded precipitation dataset is about the same as the performance of stations that are part of this dataset, one station is not enough to draw conclusions about the entire domain."

8. Page 17, line 28: "... this could lead to an overestimation of the snow cover. This in turn would lead to an overestimation of the snow depth." How does that work? What's the physical link here, to explain this cause and effect?

   Author response: Thank you for bringing this up. We agree that this was not explained very well. We will improve the discussion of this issue in the revised manuscript.

General comments:

9. The impact of precipitation and air temperature observations on the simulation could be better highlighted with "leave-one-out" experiments.

   Author response: Thank you for your good suggestion. It would have indeed been interesting to investigate the impact of precipitation and air temperature separately. This was not the focus of our paper however, and it would be too comprehensive to start a new set of experiments and discuss the results in this paper. Instead we will add this topic as a very interesting direction for

future work. We will also include more in-depth evaluation of the temperature and precipitation forcing datasets in the revised manuscript.

Changes in manuscript: the following sentence has been added to the conclusions: "To investigate the impact of using gridded observations of temperature and precipitation separately, "leave-one-out" experiments could be carried out (two extra experiments where one uses only gridded observations of temperature, and one uses only gridded observations of precipitation, while all other variables come from AROME-MetCoOp)."

10. **General comment:** The article would gain in quality if a comparison of the results presented in this article would be compared with what is currently available (operationally) in Norway.

    Author response: Thank you for this suggestion. A paper has been submitted by Skaugen et al last month, titled "In search of operational snow model structures for the future - comparing four snow models for 17 catchments in Norway" (full reference below). In this paper, GridObs-Crocus is compared to 3 other models. One of the models is seNorge, which is used operationally by NVE (The Norwegian Water Resources and Energy Directorate). In this study by Skaugen et al, all models use the same gridded dataset of precipitation and temperature as in our study. We have therefore decided to focus on evaluating the use of this gridded dataset as forcing versus using only meteorological forecasts, so that the two papers compliment each other.

    In the revised manuscript, we will mention this paper by Skaugen et al.

    Reference: T. Skaugen , H. Luijting, T. Saloranta, D. Vikhamar-Schuler and K. Müller: "In search of operational snow model structures for the future - comparing four snow models for 17 catchments in Norway", Hydrology Research (Submitted December 2017)

11. **General comment:** Has there been any tests to evaluate the impact of air temperature and precipitation seperately?

    Author response: Thank you for your comment. We are not entirely sure of the difference (if any) between comment 9 (above) and this comment. In the revised manuscript we will have a stronger focus on the evaluation of the forcing data sets, and we will add more discussion on the temperature and precipitation on the SURFEX/Crocus results in the revised manuscript. We have not looked at the impact of air temperature and precipitation separately, but this would be a very interesting experiment for future work. It has been added to the section about future work, see our answer to comment 9 above.

Minor comments: ——————

12. Page 1, line 8: "The results are promising." This is too vague.

    Author response: Thank you for your comment. We have removed this sentence from the abstract, as we agree that this is too vague. It was meant as an introduction to the following sentences that describe the results in more detail, but this is not really necessary.

13. Page 2, line 8: "... include more statistics to capture the physical snow processes". The word "capture" is not appropriate in that sentence.

Author response: We have replaced the word “capture” with “describe”.

14. Page 2, line 10: "grid spacing" instead of "resolution".

    Author response: We cannot find the word “resolution” on page 2, line 10. The word is used two times on page 2, both in the context of temporal resolution, where we would not want to use grid spacing. On page 4, line 10, we changed horizontal resolution to horizontal grid spacing. We hope this is what was meant here.

15. Page 2, line 11: "levels" instead of "layers".

    Author response: We again have to assume that the reviewer means page 4, line 11 here. This was changed to “The atmosphere is divided into 65 vertical levels, with the first level at…”

16. Page 2, line 14: "... the atmospheric part..." this is too vague.

    Author response: This should again be page 4, line 14. The sentence has been rewritten to: “The fluxes computed by SURFEX at the atmosphere–surface interface serve as the lower boundary conditions for the atmosphere within AROME MetCoOp.”

17. Page 7, Fig. 3: The text on this figure is too small, unreadable.

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

18. Page 10, Fig. 6: The two colors chosen for this figure are too alike (difficult to distinguish for old eyes like mine...)

    Author response: Thank you for pointing this out. We have changed the color to a lighter blue to make it easier to distinguish the two colors. We wanted to keep blue as this color is used for the GridObs-Crocus results throughout the paper. We have also increased the font size for all text, as well as the size of the markers, in response to a comment (nr 33) from reviewer #2. We hope this makes the plot easier to read.

19. Page 11, Table 2: The statistics presented in that table are quite interesting, but reading it is a bit tedious. I wonder if there could be another way of arranging the table.

    Author response: Thank you for your comment. The table has been changed: the results for GridObs-Crocus and AROME-Crocus are now placed in columns instead of multiple rows, which makes the table more compact, and hopefully easier to read.

20. Page 11, line 3: "... exceeds 8 m/s..." this is for the winds at what height, 10 m? This should be mentioned.

    Author response: Yes the wind speed is at 10 m height. The sentence has been changed to clarify this: “Blowing snow days are defined as days during which the wind speed (at a height of 10 m)...”

21. Page 12, Fig. 7: same comment as Fig. 6 concerning the colors.

    Author response: This plot has been changed in the same way as figure 6, see our reply to comment nr 18 above.

22. Page 15, Fig. 9: the text is too small, unreadable.

    Author response: This figure has been changed so that all the text (on the axes, axes labels and titles) is larger, and readable.