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).
Author responses in *red Italic* are added at the time of the submission of the final 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.

   Update: Update: The introduction has been thoroughly rewritten. We have included many more references to relevant work, more details on the use of NWP data as forcing for snow models, and a section on the originality of our work compared to previous work. The description includes studies carried out for both Norwegian snow conditions as well as for other arctic and alpine areas.

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.

Update:

What we should also mention is that the temperature used in AROME-Crocus has a resolution of 500 m, as this is a post-processed, and terrain-adjusted variable from AROME-MetCoOp. We have now clarified this in section 2.3.1.

The following text has been added to section 2.3.1
"The AROME-MetCoOp temperature used to force SURFEX/Crocus has a grid spacing of 500m. This is a post-processed variable that uses a Kalman filter correction at observation stations, which is then interpolated horizontally using decreasing weights with increasing distance from the station. The temperature is further corrected with a height correction which also takes into account vertical temperature profiles in inversion situations in winter time. The wind speed from AROME-MetCoOp has also been statistically post-processed to represent the maximum wind speed during the last hour. In addition, correction factors are applied to the wind speed depending on wind direction and on the region."

The following text has been added to section 4.2:
"In this study, the forcing data from AROME-MetCoOp is interpolated from the original 2.5 km resolution to the 1 km resolution required by SURFEX/Crocus, apart from the temperature which is a post-processed, terrain-adjusted variable with a resolution of 500 m. The interpolation of AROME-MetCoOp data over a domain with complex topography could lead to differences in elevation (between AROME-MetCoOp and SURFEX/Crocus), which might lead to bias in for example the precipitation. We believe this might affect the AROME-Crocus experiment more than the GridObs-Crocus experiment, since the two experiments use different data sources. "

The following sentence has been added to the Conclusions chapter, in the paragraph about future work:
"When using AROME-MetCoOp as forcing data for running SURFEX/Crocus at a resolution higher than 2.5 km, terrain adjustment routines should be applied to the generation of forcing data. In this study we accounted for local terrain effects, by using post-processed AROME-MetCoOp temperature and wind, but this could be extended to other variables."

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.

   Author response update: Figure 3 has been changed in the revised manuscript, to two seperate density plots (one for AROME-Crocus and one for GridObs-Crocus). Having two plots, and adding the density of points to the plots, makes it easier to understand the data presented. Thank *you for this suggestion. The discussion of the plot has also been changed, with more discussion of what is shown in figure 3.*

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.

   Update: the text in the discussion has been changed to:

   "The SURFEX/Crocus model was originally developed for use in high alpine regions, where there is not a lot of vegetation. In those areas, the assumption of the uniform snow cover is realistic, as there is no interaction with vegetation, but for areas covered with forest and closer to sea level this could lead to an overestimation of the snow cover. When the snow cover is overestimated, the albedo will be too high and this will slow down the snow melt at the end of the season. This might explain the underestimated snow melt in both experiments. "

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)

    Update: The reference has been added to the revised manuscript and mentioned in the introduction.

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.

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).
Author responses in *red Italic* are added at the time of the submission of the final 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.

Update: An evaluation of the precipitation forcing from the two experiments show that while the amount of rainfall in both experiments is often quite close together, there are big differences in the snowfall amount. AROME-Crocus consistently has larger amounts of snow, often 1.5-2 times as much as GridObs-Crocus, and in some cases more than 3 times as much. The differences are largest on the west side of the domain compared to the east side of the domain. Altitude plays less of a role. It is interesting that the overestimation of total precipitation in AROME-MetCoOp might be limited to the snowfall amount.

The following text has been added to section 3.1:
"An evaluation of the precipitation forcing data for AROME-Crocus and GridObs-Crocus for the six stations from Fig. 6 reveals that while the rainfall amount is often quite similar in the two experiments, the largest differences are found in the snowfall amount. AROME-Crocus consistently shows a larger amount of snowfall (accumulated over a year) compared to GridObs-Crocus, often about twice as much. The differences are largest for the stations located in the western part of the domain. As described in section 2.2, this region receives the highest amounts of winter precipitation in our study area."

The following text has been added to section 3.3:
"As discussed previously, the differences between the snowfall amounts from the two precipitation forcing datasets is largest on the west side of the domain; AROME-Crocus receives about twice as much snow compared to GridObs-Crocus in this region. This explains why the differences between GridObs-Crocus and AROME-Crocus in Fig. 10 are also largest in this area. "

The following text has been added to section 4.2:
"An evaluation of the accumulated precipitation from the two forcing datasets showed that while the rainfall amounts are rather similar, the snowfall amount in AROME-Crocus is much larger compared to GridObs-Crocus, especially at the west side of the domain. This suggests that the overestimation of total precipitation in AROME-MetCoOp might be limited to the snowfall amount."

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.

Update:

What we should also mention is that the temperature used in AROME-Crocus has a resolution of 500 m, as this is a post-processed, and terrain-adjusted variable from AROME-MetCoOp. We have now clarified this in section 2.3.1.

The following text has been added to section 2.3.1
"The AROME-MetCoOp temperature used to force SURFEX/Crocus has a grid spacing of 500m. This is a post-processed variable that uses a Kalman filter correction at observation stations, which is then interpolated horizontally using decreasing weights with increasing distance from the station. The temperature is further corrected with a height correction which also takes into account vertical temperature profiles in inversion situations in winter time. The wind speed from AROME-MetCoOp has also been statistically post-processed to represent the maximum wind speed during the last hour. In addition, correction factors are applied to the wind speed depending on wind direction and on the region."

The following text has been added to section 4.2:
"In this study, the forcing data from AROME-MetCoOp is interpolated from the original 2.5 km resolution to the 1 km resolution required by SURFEX/Crocus, apart from the temperature which is a post-processed, terrain-adjusted variable with a resolution of 500 m. The interpolation of AROME-MetCoOp data over a domain with complex topography could lead to differences in elevation (between AROME-MetCoOp and SURFEX/Crocus), which might lead to bias in for example the precipitation. We believe this might affect the AROME-Crocus experiment more than the GridObs-Crocus experiment, since the two experiments use different data sources. "

The following sentence has been added to the Conclusions chapter, in the paragraph about future work:
"When using AROME-MetCoOp as forcing data for running SURFEX/Crocus at a resolution higher than 2.5 km, terrain adjustment routines should be applied to the generation of forcing data. In this study we accounted for local terrain effects, by using post-processed AROME-MetCoOp temperature and wind, but this could be extended to other variables."

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.

Update:

The topic of height differences between observations and model grid points has now been added to section 2.3.1, see below.

We have calculated statistics for only the stations with a height difference of less than 250 m. This does improve the bias and RMSE, but not by very much. The overall bias for snow depth of GridObs-Crocus remains the same (6 cm), while the RMSE changes from 28 to 25 cm. For AROME-Crocus, the bias changes from 42 to 40 cm, while the RMSE increases from 68 to 71 cm. This does not change anything for our conclusions. For the length of the snow season, the differences are larger, but still the overall conclusions would remain the same, as both experiments change in a similar way. The bias in the length of the snow season for GridObs-Crocus changes from 11 to 8 days for year 1 (RMSE 25 days vs 21 days), and from 8 to 4 days (RMSE 21 vs 18 days) for year 2. For AROME-Crocus the bias changes from 34 to 27 days for year 1 (RMSE 44 vs 37 days), and from 18 to 9 days in year 2 (RMSE: 28 vs 20 days). The changes are largest for AROME-Crocus (7-9 days improvement, versus 3-4 days for GridObs-Crocus), but GridObs-Crocus still outperforms AROME-Crocus at all times for this variable. We have added a short version of this quantification to the discussion chapter, see below.

Changes in manuscript:

The following text has been added to 2.3.1:
"In a domain with deep valleys and high mountains, it is difficult to match the elevation of the weather stations with the nearest grid point in the SURFEX/Crocus experiments. As there were only 30 stations with high quality snow depth observations in the domain, it was decided not to

filter out stations based on these height differences. This issue is discussed in more detail in section 4.1 in the Discussion chapter. "

The following text has been added to 4.1:
"It could be argued that SURFEX grid points with a large height difference compared to the corresponding station should not be used for verification purposes. However, when calculating bias and RMSE values only for the stations with a height difference below 250 m, this did not have a large impact on the results. With a few exceptions, the bias and RMSE are lower when excluding stations with a large height difference, but the differences in overall bias are only ± 3 cm, which does not change any of the conclusions in this study."

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

   Update: The introduction has been thoroughly rewritten. We have included many more references to relevant work, more details on the use of NWP data as forcing for snow models, and a section on the originality of our work compared to previous work. The description includes studies carried out for both Norwegian snow conditions as well as for other arctic and alpine areas.

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.

   Update: We have carried out 1D experiments using the GridObs-Crocus setup for two locations: Midtstova and Hemsedal II. As expected, the snow depth is decreased when the option for sublimation loss is turned on, and the snow cover duration is decreased. Where the snow depth is overestimated, this brings a slight improvement, but when the snow depth and season length is already underestimated, this does not improve the results.

   The following text has been added to the Results chapter, along with a new figure (8)::

   "SURFEX/Crocus does have an option to run with sublimation in case of snowdrift. This option has been tested for two stations from Fig. 4: Midtstova and Hemsedal II. In this experiment, SURFEX/Crocus was run twice in 1D mode for these 2 locations: one experiment with identical settings as GridObs-Crocus, and one nearly identical with the exception of the option for sublimation in case of snowdrift (GridObs-Crocus+BS). The results are shown in Fig. 8. For both locations, the snow depth in GridObs-Crocus+BS is decreased, as expected. For Hemsedal II, this reduction is an improvement compared to the GridObs-experiment, which overestimated the snow depth. The bias in GridObs-Crocus was +16 cm for Hemsedal II, which has improved to

+10 cm in GridObs-Crocus+BS. For Midtstova, GridObs-Crocus underestimates the snow depth for most of the two winter seasons (bias: -5 cm), and this underestimation is significantly larger in the GridObs+BS experiment (-28 cm). "

The following text has been added to the discussion:

"There is an option in the SURFEX/Crocus model to calculate the rate of sublimation in case of snowdrift, which results in a loss of snow. This option was tested for two stations in his study, using the GridObs-Crocus forcing dataset. As expected, this resulted in a decrease in snow depth and a decrease in season length. As GridObs-Crocus already underestimates the snow depth and snow cover, this is not an improvement."

The following sentence has been added to the conclusions:

"Using the option in SURFEX/Crocus of running with sublimation in case of snowdrift is not enough to address this issue."

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.*

Update: the introduction has been rewritten, and includes much more discussion on previous work, the originality of our work compared to these studies, and how we contribute to the field of mountain snow research. All the mentioned studies, and many more, have been discussed in the updated introduction. The growing use of high resolution NWP forecasts and their limitations has also been discussed. We believe the context of the study and our scientific questions are now more clearly described.

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).

    Update - the following text has been added to section 2.1:

    "In this study, the soil has 14 layers. A force-restore method with 3 layers for hydrology was used for soil discretization and physics within ISBA-DIF. The HSWD (Harmonized World Soil Database) 1 km resolution database for soil texture (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2012) was used for the soil properties."

    "These dates were chosen because the hydrological year starts on 1 September, and at that time there is normally no snow in the mountains. In this study, we start a new simulation on 1 September, with no snow present, and with default values for soil properties, for both 2014/2015 and 2015/2016. ."

    The following text has been added to section 2.2 (a new section about the study area):

    "The domain covers nearly 20.000 km2 (111 x 175 km), and contains 100 x 330 grid points."

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.

    The new figure 2 shows a histogram of the elevation of grid points within the SURFEX domain (as well as the elevation of the 30 weather stations). The following text has been added to 2.3.1:

    "The elevation of the grid points in the SURFEX domain is also shown in Fig. 2, which shows a typical issue with the location of weather stations, particularly in the Norwegian mountains: there are many stations located at low elevations (at the bottom of valleys), and very few stations at high elevations."

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.

Update: the following text has been added to the new section 2.2 Study Area:

"Due to the watershed and the prevailing weather patterns, there is a large gradient in precipitation amount over the domain. The far western parts of the domain receive on average around 1500 mm of precipitation during a winter season, while the eastern parts only receives 100-300 mm (Hanssen- Bauer et al, 2015). The western part of the domain has a maritime climate while the eastern part has a more inland climate, which means the average temperature during winter is higher at the western part of the domain (around or just below 0 \degree Celsius), compared to the eastern side (around -10 \degree Celsius) (Hanssen- Bauer et al, 2015). This means the gradient in average snowfall amount is not as large as the gradient in precipitation amount, but the western part of the domain still receives significantly more snow than the eastern part (Hanssen- Bauer et al, 2015). "

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.

    Update: During the snow accumulation season the temperature at Midtstova is mostly well below freezing level. There are a few episodes each winter with temperatures just above zero, where the underestimated temperature in AROME-MetCoOp means the precipitation during those episodes comes as snow instead of rain, but these do not add up to large amounts.

    The following text has been added to section 3.1:

    "In the forcing data for Midtstova we find a bias of -1.5 degrees for AROME-Crocus, compared to -0.8 degrees for GridObs-Crocus. This bias is larger than the overall bias for all nine stations measuring temperature: -0.5 degree for AROME-Crocus and -0.2 degree for GridObs-Crocus. During the snow accumulation season the temperature at Midtstova is mostly well below freezing level. There are a few episodes each winter with temperatures just above zero, where the underestimated temperature in AROME-MetCoOp means the precipitation during those episodes comes as snow instead of rain, but these do not add up to large amounts. "

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).

    Update: The case study revealed that the underestimation of snow depth at Midtstova in the beginning of the winter 2014-2015 is due to underestimation of the precipitation in GridObs-Crocus. There were no issues with the phase of precipitation, as the temperature is below zero for almost the entire period. Further investigation revealed that there is no hourly precipitation data for Midtstova between 27 October 2014 and 26 January 2015 due to a broken sensor. That means that the precipitation at Midtstova will be an interpolated value using surrounding stations during that period in the GridObs forcing, and this might explain why the precipitation is underestimated.

    The following text has been added to section 3.1:
    "GridObs-Crocus shows much more realistic results for Midtstova, although there is an underestimation of snow depth during the first part of the 2014-2015 winter. From 27 October 2014 until 26 January 2015, the precipitation sensor at Midtstova was out of order, and the forcing from GridObs-Crocus for Midtstova will therefore be represented by interpolated values from surrounding stations, which might explain the underestimation."

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.

    Update: The [-0.5 0.5] cm category has been replaced by [-1 1] cm in the new Fig. 5. With the logarithmic y axis, this is not a problem at all. Thank you for pointing this out!

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.

    Update: We have calculated the Jaccard index for the SCA images from Fig. 10 (Fig. 9 in the original manuscript). This confirms our visual conclusions that AROME-Crocus performs better in terms of snow-covered area.

    A new table has been added:

Jaccard Index	GridObs-Crocus	AROME-Crocus
15 March 2015	0.92	0.99
20 April 2015	0.82	0.94
15 May 2015	0.65	0.82
04 July 2015	0.19	0.68

Table 4. Jaccard index for the snow covered areas shown in Fig. 10. A score of 1 means the image perfectly matches the MODIS image, a score of 0 means there is no overlap between the image from the experiment compared to the MODIS image.

The following text has been added to section 3.3:
"Table 4 shows the Jaccard indices for the images from Fig. 10. The Jaccard index was also used by for example Queno et al (2016). It is a similarity index applied to the snow cover images which were remapped onto the same grid (which means that the snow cover from the MODIS images used to calculate the Jaccard index has a lower resolution than the one shown in Fig. 10). The Jaccard index is calculated as $J(X,Y) = |X\backslash \cup Y| / |X\backslash \cap Y|$, where X and Y are the simulated and observed snow cover, respectively. The number of grid points that are snow-covered in both SURFEX/Crocus and in the MODIS image is divided by the total amount of snow-covered grid points (in either SURFEX/Crocus or MODIS). When the Jaccard index equals 1, there is a perfect match between snow-covered grid points, and when the Jaccard index equals 0, there is no match at all. Table 4 shows that AROME-Crocus consistently has higher Jaccard indices compared to GridObs-Crocus. The indices decrease (for both experiments) during the melt season. "

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.

    Update: the first two references have been added to the revised manuscript and discussed in the introduction. The third reference focuses on the snow maps derived from remote sensing data, while we focus on snow modeling, which is why we have not included this reference.

Forcing the SURFEX/Crocus snow model with combined hourly meteorological forecasts and gridded observations in southern Norway

Hanneke Luijting[1], Dagrun Vikhamar-Schuler[1], Trygve Aspelien[1], and Mariken Homleid[1]

[1]The Norwegian Meteorological Institute, PO Box 43 Blindern, 0313 Oslo, Norway

Correspondence to: Hanneke Luijting (hanneke.luijting@met.no) or Dagrun Vikhamar-Schuler (dagrunvs@met.no)

Abstract. In Norway, thirty percent of the annual precipitation falls as snow. Knowledge of the snow reservoir is therefore important for energy production and water resource management. The land surface model SURFEX with the detailed snowpack scheme Crocus (SURFEX/Crocus) has been run with a grid spacing of approximately 1 km over an area in southern Norway for two years (01 September 2014 - 31 August 2016), using two different forcing data sets: 1) hourly ~~meteorological~~ forecasts from the operational weather forecast model AROME MetCoOp (2.5 km grid spacing) including post-processed temperature (500 m grid spacing) and wind variables, and 2) gridded hourly observations of temperature and precipitation (1 km grid spacing) in combination with the meteorological forecasts from AROME MetCoOp. We present an evaluation of the modeled snow depth and snow cover, as compared to 30 point observations of snow depth and to MODIS satellite images of the snow-covered area. The evaluation focuses on snow accumulation and snow melt. ~~The results are promising.~~ Both experiments are capable of simulating the snow pack over the two winter seasons, but there is an overestimation of snow depth when using ~~only~~ meteorological forecasts from AROME MetCoOp (with a bias of 42 cm and RMSE of 68 cm), although the snow-covered area throughout the melt season is better represented by this experiment. The errors, when using AROME MetCoOp as forcing, accumulate over the snow season~~, showing that assimilation of snow depth observations into SURFEX/Crocus might be necessary when using only meteorological forecasts as forcing~~. When using gridded observations, the simulation of snow depth is significantly improved (the bias for this experiment is 6 cm and RMSE 28 cm), which shows that using a combination of gridded observations and meteorological forecasts to force a snowpack model is very useful and can give better results than ~~only~~ mainly using meteorological forecasts. There is however an underestimation of snow ablation in both experiments. This is ~~mainly~~ generally due to the absence of wind-induced erosion of snow in the SURFEX/Crocus model, underestimated snow melt and biases in the forcing data.

1 Introduction

Snow is a key element in the hydrological cycle~~and in Arctic areas~~. Seasonal snow covers large areas of the Northern Hemisphere and the Arctic. In these areas the snow cover extent in spring has reduced more rapidly the past 40 years than over the past 90 years ~~(Brown and Robinson, 2011) . Generally, the size of the area covered by snow is decreasing, and the period with snow-covered ground is reducing~~(Brown and Robinson, 2011; Brown et al., 2017) . The largest declines in snow cover extent and duration are observed in Arctic coastal areas, e.g. in Scandinavia (Rasmus et al., 2015; Brown et al., 2017) . In Norway there is a general trend towards a later start and an earlier end of the snow season, although there are large annual variabilities (Hanssen-Bauer et al., 2015, 2017) . Trends in snow depth may vary with elevation, as observed for some Norwegian regions (Skaugen et al., 2012; Dyrrdal et al., 2013) . Information about changes in the seasonal variation and the amounts of snow seasonal changes in snow duration and amounts are important for many societal applications and for Arctic ecosystem changes. An overview of changes in snow and impacts due to these changes, is provided by (Bokhorst et al., 2016) . As an example, in Norway , Bokhorst et al. (2016) .

In Norway 30% of the annual precipitation falls as snow (Saloranta, 2012), and knowledge of the snow conditions is of high importance for . Observations show that changes in the winter climate the past 50 years, and particularly since 2000, give more warming events and rainfall events during winter (Vikhamar-Schuler et al., 2016; Kivinen et al., 2017) . This, in turn, affects the internal snow structure giving e.g. more wet snow conditions and ground-ice layering (Johansson et al., 2011; Vikhamar-Schuler et al., 2 Updated information on the daily local snow conditions in mountainous and lowland areas is very useful for many applications, notably local flood prediction, hydropower production planning, snow avalanche predictionfor the mountains, traffic flow and management, tourism, and mobility and food access for mammals living in the snow tourism and traffic flow management. Typical information needed for these applications are daily snow forecasts (for the next days), but depending on the application, also knowledge on the snow conditions for the past winter(s), last month(s), last week(s), past 3 days and yesterday.

For these purposes, a wide range of empirically and physically based snow models have been developed and reported in literature(, see e.g. Magnusson et al. (2015)). Several snow model intercomparison projects have also been performed, e.g. by Etchevers et al. (2004) and Essery et al. (2013). Models differ in several ways e.g. the parameterization and simplification of snow processes, the spatial and temporal resolution or the need for input data of various weather elements. Physical parameterization of the snow processes is the main focus regarding the physically based models, while the empirically based models generally include more statistics to capture describe the physical snow processes. The need for input data is therefore usually also larger and more detailed for physically based models than for the empirically based models. Empirical models also often need calibration.

In Norway, hydropower production companies and the national flood forecasting service use both the national operational flood forecasting and hydropower companies use the HBV model, which includes an empirical degree day model for snow simulations (Bergstrøm, 1976; Sælthun, 1996; Ruan and Langsholt, 2017) . Snow maps are also produced operationally on a daily basis and published at www.seNorge.no and www.xgeo.no (Saloranta, 2016) , and these maps are used by the national snow avalanche service (Barfod et al., 2013; Engeset, 2013) . Both these applications use gridded daily observations of temperature and precipitation (Mohr, 2008; Lussana et al., 2018a) . Gridded observations for hydrological applications including producing snow maps are also used in other countries, e.g. in Switzerland (Foppa et al., 2007, www.slf.ch) and Sweden (www.smhi.se).

Another type of forcing data for snow models operationally for estimating the snow distribution and runoff in the various drainage areas. Historically, the HBV model (Bergstrøm, 1976; Sælthun, 1996) is extensively models are numerical weather forecasts (NWP) from atmospheric models. NWP data provide simulations for many parameters, on an hourly scale. This type of data therefore represents different information compared to gridded daily observations of temperature and precipitation.

Using sub-daily (e.g. hourly) data in snow modeling may contribute to improved snow melt (e.g. in spring with freezing nights and melting snow during daytime) and improved snow accumulation simulations (rainfall/snowfall). Complex physically-based snow models, which simulate both the surface energy balance and the heat flow through the snowpack, usually require many weather elements as input. SnowPack (Bartelt and Lehning, 2002; Lehning et al., 2002) , SURFEX/ISBA/Crocus (Vionnet et al., 2012) , JULES (Best et al., 2011) and FSM (Essery, 2015) are examples of snow or land-surface models, which also simulate the internal structure of the snowpack. SnowPack and Crocus are used in the operational production chains. However, the snow routine in the HBV modelis an empirical degree day model , using daily values of temperature and precipitation (Sælthun, 1996) . Atmospheric models used in the operational weather forecasting snow avalanche service in Switzerland and France, respectively (Fierz et al., 2013; Lafaysse et al., 2017) . Many studies show how high resolution NWP data are very valuable in driving these snow models, see e.g. Bellaire et al. (2011, 2013); Horton et al. (2015); Vionnet et al. (2016); Quéno et al. (2016) . NWP data have also been used as driving data in hydrometeorological models (Carrera et al., 2010, e.g.) .

For a point location in the Columbia Mountains, Western Canada, Bellaire et al. (2011, 2013) used 15 km resolution weather forecasts from the NWP model GEM to force the SnowPack model. This study was later extended to a gridded area in the same region by Horton et al. (2015) who forced the SnowPack model using 2.5 km resolution NWP data from the GEM-LAM model. The use of NWP data as precipitation forcing for snow models was analysed and discussed by (Schirmer and Jamieson, 2015) . They compared two NWP datasets (GEM: 15 km and GEM-LAM: 2.5 km spatial resolution) over complex mountainous terrain during winter time, and found that the highest resolution dataset performed best in terms of precipitation forecasts. Bernier et al. (2011) used a downscaling technique to account for local terrain effects on the surface temperatures not resolved by the low-resolution NWP model. With higher spatial resolution of the NWP models, the orographic precipitation is better reproduced. In the French Alps, high-resolution forecasts (2.5 km) from the AROME NWP model were used to drive Crocus (Vionnet et al., 2016) . A similar study using the same AROME NWP model was carried out for the French and the Spanish Pyrenees by Quéno et al. (2016) . Both these studies showed that high-resolution NWP data represents a very useful and promising data source for snow models to produce snow maps of high quality. However, the authors point out some limitations of using only NWP data. Terrain effects are not well enough accounted for on a kilometric scale, whereas future development of sub-kilometric scale NWP data might improve e.g. terrain effects on the incoming solar radiation. Combining NWP data with other data sources (e.g. observations, radar) might improve the forcing data, particularly the precipitation fields. Redistribution of snow due to wind is another difficult issue in mountainous areas. Running snow models with ensemble based forecasts is a promising method to account for these uncertainties (Vernay et al., 2015b; Lafaysse et al., 2017) .

Weather forecasting models are presently evolving fast, by including more detailed models and they include more and more detailed parametrization of land surface processes such as connected to snow and soil. SURFEX (Surface Externalise??) (Masson et al., 2013) is an example of a land surface model, which can be run both inline the as part of an atmospheric weather forecast model, for example AROME MetCoOp (Müller et al., 2017), or offline as a stand-alone model. At the Norwegian Meteorological Institute (MET Norway), the AROME MetCoOp model is run operationally to provide short-term weather forecasts covering large parts of the Nordic region (Müller et al., 2017).

In this study we are evaluating evaluate the performance of the SURFEX model, using the Crocus snow scheme (?) (Vionnet et al., 2012) Norwegian snow conditions. We selected a west-east transect in a mountainous area of South Norway as the study areaSURFEX/Crocus has not previously been run in a gridded stand alone version for regions in Norway (as a 2D study). However, the model has earlier been tested for single points (1D study) with observations from weather stations and NWP data (Vikhamar-Schuler et al., 2011) . Our study is carried out as part of several research projects within hydropower and flood forecasting. The domain was chosen to cover mountains in southern Norway and to include a cross-section from west to east that crosses the watershed in this region. This domain includes catchment areas that are of high interest to hydropower companies.

The aim of our study is to test the performance and the benefit of different gridded forcing datasets as input to the SURFEX/Crocus model, and validate the simulated snow amounts and snow melt pattern in the selected domain. The originality of our work is linked to the unique combination of using both raw weather predictions, post-processed weather predictions and gridded observations. Our simulations were run without any assimilation of snow observations. Experiments were performed by applying two different data sets from the winters 2014/2015 and 2015/1016 as forcing to the SURFEX/Crocus model: 1) Predictions from the AROME MetCoOp model (Müller et al., 2017) with a grid spacing of 2.5 km (Müller et al., 2017) , of which both the temperature and the wind data were improved by post-processing algorithms, and 2) Gridded observations of precipitation and temperature (GridObs) with a grid spacing of 1 km (?) (Lussana et al., 2018b, a) . Both data sets have 1 hourly temporal resolution, and are discussed in detail in section 2.3. The aim was to compare and combine different gridded forcing data sets as input to the

Although AROME-SURFEX/Crocus has previously been used over the French Alps and the Pyrenees (Vionnet et al., 2016; Quéno et al., neither of our two datasets described above have been used as forcing for SURFEX/Crocus model and validate the computed accumulated snow amounts and snow melt pattern in both Norwegian mountains and lowlands. The simulations were run without any assimilation of snow observations. The results may be used for future for Norwegian mountains and lowland regions before. Additionally, the climate in South-Norway is different from the Alps and the Pyrenees. A west-east transect crossing the mountain chain comprises a climatic transect from maritime, alpine to more continental climate. Snow conditions and stratigraphy vary regionally as outlined by e.g. (Sturm et al., 1995) , who defined six snow classes, of which at least two classes are inside our domain (maritime and alpine). The SURFEX/Crocus snow model may therefore perform differently in individual regions. Our study contributes to a development which can produce new supplementary snow information (including snow stratigraphy) and thereby may contribute to the development of a future system for daily snow mapping. , different from the seNorge and the HBV model. The performance of the SURFEX/Crocus model is compared with three other snow models including the seNorge model in a separate study by Skaugen et al. (2017) .

2 Model setup and data sets

2.1 The SURFEX/Crocus model

The model used in this study is the detailed snowpack model Crocus (Brun et al., 1992; ?) (Brun et al., 1992; Vionnet et al., 2012) coupled with the ISBA land surface model within the SURFEX (Surface Externaliseé) interface (Masson et al., 2013). The soil scheme

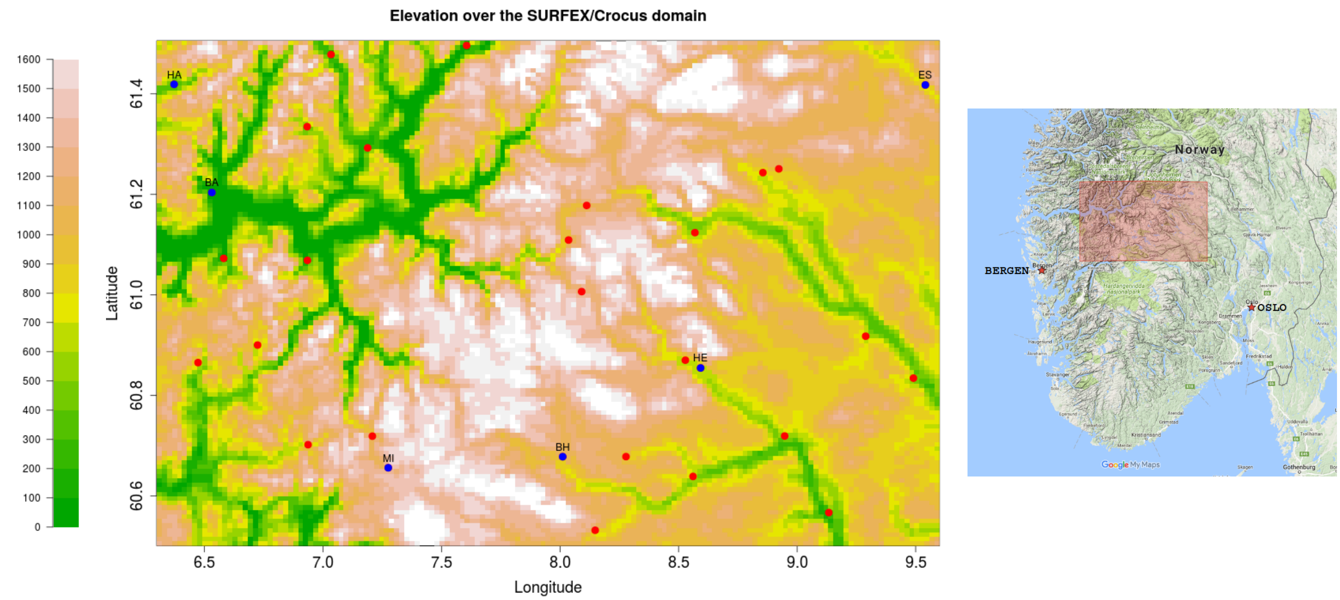

Figure 1. Map showing the domain over which the SURFEX/Crocus model was run on the ~~left~~ right (map data: Google), with on the ~~right~~ left a map showing the elevation over the SURFEX/Crocus model domain, and the locations of the 30 observations used in this paper (indicated by blue and red dots). The blue dots indicate the 6 stations used in Fig. 4: BA = Balestrand Brannstasjon, HA = Haukedal, HE = Hemsedal II, ES = Espedalen, BH = Bakko i Hol and MI = Midtstova.

used is ISBA-DIF (Boone et al., 2000; Habets et al., 2003), which uses a multi-layer approach. In this study, the soil has 14 layers. A force-restore method with 3 layers for hydrology was used for soil discretization and physics within ISBA-DIF. The HSWD (Harmonized World Soil Database) 1 km resolution database for soil texture (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2012) was used for the soil properties. The snowpack scheme Crocus models the physical properties of ~~up to~~ 50 dynamic layers within the snowpack, as well as the underlying ground. Once the snowpack reaches a threshold of 1 kg ~~/m?~~ m^{-2} SWE, the fractional snow cover fraction over a grid point is assumed to be 1. The SURFEX/Crocus model can be run in stand-alone (or offline) mode, or fully coupled to an atmospheric model.

For this study, the SURFEX/Crocus model was used in an offline mode (not fully coupled to an atmospheric model), on a 0.01°grid (approximately 1 km), with a 5 minute internal time step and output every hour. The transport of snow by wind is not simulated. The SURFEX/Crocus model was run for two winter seasons: from 1 September 2014 until 31 August 2016. These dates were chosen because the hydrological year starts on 1 September, and at that time there is normally no snow in the mountains. In this study, we start a new simulation on 1 September, with no snow present, and with default values for soil properties, for both 2014/2015 and 2015/2016.

	AROME-Crocus	**GridObs-Crocus**
Air temperature [K]	AROME-MetCoOp post-processed	Gridded observations
Specific humidity [kg ~~/kg~~kg^{-1}]	AROME-MetCoOp	AROME-MetCoOp
Wind speed [m ~~/s~~s^{-1}]	AROME-MetCoOp post-processed	AROME-MetCoOp
Wind direction [degrees]	AROME-MetCoOp post-processed	AROME-MetCoOp
Incoming direct shortwave radiation [W ~~/m?~~m^{-2}]	AROME-MetCoOp	AROME-MetCoOp
Incoming longwave radiation [W ~~/m?~~m^{-2}]	AROME-MetCoOp	AROME-MetCoOp
Surface pressure [Pa]	AROME-MetCoOp	AROME-MetCoOp
Rainfall rate [kg ~~/m??/s~~m^{-2} s^{-1}]	AROME-MetCoOp	Gridded observations
Snowfall rate [kg ~~/m??/s~~m^{-2} s^{-1}]	AROME-MetCoOp	Gridded observations

Table 1. Description of the forcing data sets used in the two experiments: 1) AROME-Crocus; and 2) GridObs-Crocus.

2.2 The study area

Figure 1 shows the domain over which the SURFEX/Crocus model was run, and the elevation over the model domain. ~~This~~ The domain covers nearly 20.000 km^{-2} (111 x 175 km), and contains 100 x 330 grid points. As mentioned in the introduction, the study area 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, as well as to include several catchment areas that are of interest to hydropower companies. The domain covers elevations from 0 masl. along fjords up to the highest mountain in Norway (2468 masl.). Therefore, the area includes different vegetation zones, ranging from high mountains above the tree line, sparsely forested and densely forested areas. This makes it a challenging area for snow modeling.

Due to the watershed and the prevailing weather patterns, there is a large gradient in precipitation amount over the domain. The far western parts of the domain receive on average around 1500 mm of precipitation during a winter season, while the eastern parts only receive 100-300 mm (Hanssen-Bauer et al., 2015) . The western part of the domain has a maritime climate while the eastern part has a more inland climate, which means the average temperature during winter is higher at the western part of the domain (around or just below 0 °Celsius), compared to the eastern side (around -10 °Celsius) (Hanssen-Bauer et al., 2015) . This means the gradient in average snowfall amount is not as large as the gradient in precipitation amount, but the western part of the domain still receives significantly more snow than the eastern part (Hanssen-Bauer et al., 2015) .

2.3 Forcing data sets

The SURFEX/Crocus model requires atmospheric forcing. For this study, we have used two different sets of forcing data. Table 1 shows an overview of which variables the SURFEX/Crocus model requires and the different sources used in the two experiments: 1) AROME-Crocus and; 2) GridObs-Crocus. AROME-Crocus uses ~~only~~ (hourly) data from the AROME MetCoOp model (described below in section 2.3.1) while GridObs-Crocus uses a combination of gridded observations of precipitation and temperature (both on a 1 hourly temporal resolution), described in section 2.3.2, and AROME MetCoOp data.

2.3.1 Numerical weather forecasts (AROME-MetCoOp)

AROME MetCoOp is a high-resolution, non-hydrostatic, convective-scale weather prediction model operated by a bilateral cooperative effort [Meteorological Cooperation on Operational Numerical Weather Prediction (MetCoOp)] between the Norwegian Meteorological Institute and the Swedish Meteorological and Hydrological Institute (Müller et al., 2017), operational since March 2014. The core of the model is based on the convection-permitting Applications of Research to Operations at Mesoscale (AROME) model developed by M??t??o-France (Seity et al., 2011). It has been modified and updated to suit advanced high-resolution weather forecasts over the Nordic regions, see Müller et al. (2017) for details. The horizontal ~~resolution~~ grid spacing is 2.5 km and the domain covers the Nordic countries. The atmosphere is divided into 65 vertical ~~layers~~levels, with the first ~~layer~~ level at approximately 12.5 m height. Atmosphere???surface interactions and surface???soil processes are described by SURFEX (Masson et al. 2013). The fluxes computed by SURFEX at the atmosphere???surface interface serve as the lower boundary conditions for the ~~atmospheric part of the model~~atmosphere within AROME MetCoOp. All surface processes are treated as one-dimensional vertical processes.

AROME MetCoOp operates with a 3-hourly update cycling, where initial fields of atmospheric and land surface variables are corrected with observations through data assimilation. At every main cycle (0000, 0600, 1200, and 1800 UTC) a 66-h forecast is produced. 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. ~~The~~ 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 (model error increases with lead time, see for example Homleid and Tveter (2016)).

The AROME-MetCoOp temperature used to force SURFEX/Crocus has a grid spacing of 500 m. This is a post-processed variable that uses a Kalman filter correction at observation stations (Homleid, 1995) , which is then interpolated horizontally using decreasing weights with increasing distance from the station. The temperature is further corrected with a height correction which also takes into account vertical temperature profiles in inversion situations in winter time. The wind speed from AROME-MetCoOp has also been statistically post-processed to represent the maximum wind speed during the last hour. In addition, correction factors are applied to the wind speed depending on wind direction and on the region. The variables from the AROME-MetCoOp 2.5 km forecasts were ~~further~~ interpolated to 1 km spatial resolution using bilinear interpolation, in order to combine the meteorological forecasts with the gridded observations (with a spatial resolution of 1 km) and to run the SURFEX/Crocus model with 1 km grid spacing.

2.3.2 Gridded observations (GridObs)

In an earlier study by Vikhamar-Schuler et al. (2011) , it was shown that snow modeling with the SURFEX/Crocus model has highest sensitivity to the temperature and precipitation input datasets. Best results were obtained when the model was forced with observations of temperature and precipitation, while replacing other input parameters with meteorological forecast data did not increase errors notably. In the study by Vikhamar-Schuler et al. (2011) the SURFEX/Crocus model was run in 1D mode for observation points, and observations from weather stations could be used directly. This is not possible when running the SURFEX/Crocus model on a 2D domain.

However, hourly gridded observations of temperature and precipitation are available on a 1 km grid over Norway~~(Lussana et al., 2016; ?)~~ This dataset uses all measurements available in MET Norway???s Climate database (eKlima, 2017). The station distribution is uneven, with more stations in the southern part of Norway and a sparser network in the north and in the mountains. The ~~gridded dataset uses a spatial interpolation procedure based on Bayesian concepts and relying on observations only. This procedure is based on statistical interpolation: the classical Optimal Interpolation (OI) scheme has been modified taking into account a scale-separation approach. Three-dimensional correlation functions are used to account for the orographic distribution of observing stations. The interpolation method is described in more detail in Uboldi et al. (2008) and Lussana et al. (2009)~~ hourly precipitation values have been obtained by using a two-step procedure. First, the spatial interpolation method described by Lussana et al. (2018a) has been applied independently to daily and hourly precipitation totals. Second, the daily precipitation totals have been disaggregated on an hourly time basis with a procedure similar to the one described by Vormoor and Skaugen (2013) . The two-step procedure has been implemented so that the final hourly product can benefit from the more accurate daily quantitative estimates that are based on a denser network of stations, if compared to the hourly ones. The hourly temperature values have been obtained by means of the method described by Lussana et al. (2018b) , while the hourly temperature dataset is described and evaluated in Lussana et al. (2016) .

The resulting gridded temperature dataset can be regarded as an unbiased estimate of the true temperature both at grid points and at station locations. Only for the most extreme negative values (temperatures below -30 °Celsius) there is a systematic warm bias of about 1 °Celsius ~~(Lussana et al., 2016; ?) . There are not as many stations measuring hourly precipitation as there are daily ones. The hourly gridded observations of precipitation have therefore been corrected by using the denser network of daily precipitationmeasurements, in such a way that the sum of hourly precipitation over a 24 hour period matches the daily observed precipitation for each grid point. ?~~ (Lussana et al., 2016, 2018b) . For precipitation, Lussana et al. (2018a) found that the precision of the estimates (at grid points) is about ??20% for precipitation, but there is a systematic underestimation of precipitation in data-sparse areas and for intense precipitation.

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.

SURFEX/Crocus requires a separate snowfall and rainfall rate, which has been derived from the hourly total precipitation by using a threshold temperature of +0.5 ° Celsius (using the gridded observations of temperature available on the same grid). This threshold temperature is commonly used for hydrological purposes in Norway (see for example Skaugen (1998)).

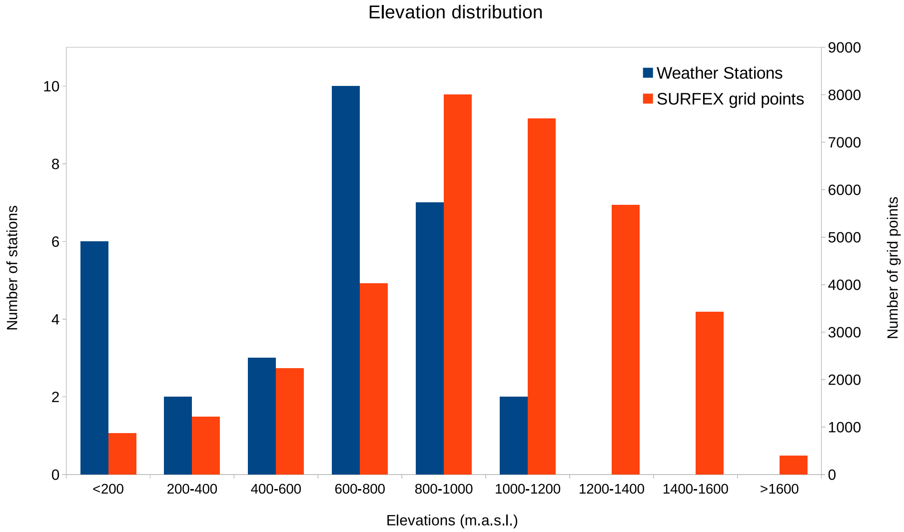

Figure 2. Distribution of elevation for all weather stations used in this study (in blue, on left axis), and of the grid points in the SURFEX domain (in red, on right axis).

2.4 Validation data set

We use two different data sets to validate the results from both experiments: point observations of snow depth and snow cover maps derived from MODIS satellite images.

2.4.1 Snow depth observations

Thirty observations of snow depth have been selected for use in verification of the model results (see Fig. 1 for their locations within the domain). Nearly all stations (25 out of 30) are official meteorological stations run by the Norwegian Meteorological Institute, while a few stations are owned by other institutions (municipalities, energy producing companies and Bane Nor, the state-owned company responsible for the Norwegian national railway infrastructure). Data from all stations is freely available from the climate database of the Norwegian Meteorological Institute (eKlima, 2017). All stations measure snow depth, nearly all (29 out of 30) measure precipitation, and 9 stations also measure air temperature. The stations were selected based on the availability of high quality snow depth observations between 1 September 2014 and 31 August 2016. The locations of the stations are reasonably well distributed over the domain (see Fig. 1) and their elevations range between 14 and 1162 meters above sea level. Figure 2 shows the elevation distribution of all stations used in this study. The elevation of the grid points in the SURFEX domain is also shown in Fig. 2, which shows a typical issue with the location of weather stations, particularly in the Norwegian mountains: there are many stations located at low elevations (at the bottom of valleys), and very few stations at high elevations.

In order to compare the observations with the results from the SURFEX/Crocus experiments, the nearest locations to the observations have been derived from the model grid. In a domain with deep valleys and high mountains, it is difficult to match the exact elevation of the weather stations with the nearest grid point in the SURFEX/Crocus experiments. As there were only 30 stations with high quality snow depth observations in the domain, it was decided not to filter out stations based on these height differences. This issue is discussed in more detail in section 4.1 in the Discussion chapter.

Daily snow depth observations taken at 06 UTC have been used for direct comparison to snow depth from the SURFEX/Crocus simulations. The observations were also used to calculate the start, length and end of the snow season, to compare against model results. The length of the snow season is defined as the number of days with more than 5 cm snow during a year. The 5 cm threshold was also used by Vionnet et al. (2016), although they used continuous snow on the ground as an additional condition. The start of the snow season is defined as the first day with more than 5 cm of snow, and the end of the snow season as the day after the last day with more than 5 cm of snow.

2.4.2 MODIS snow cover images

MODIS (Moderate Resolution Imaging Spectroradiometer; http://modis.gsfc.nasa.gov/) snow cover images (Hall and Riggs (2007); Klein and Stroeve (2002)) with a resolution of 500 m are available for the melt season of the 2014-2015 winter. The same method as described in ? Lussana et al. (2018a) was used: the MODIS images were converted to snow-covered area (SCA) on a scale from 0-100% coverage using a method based on the Norwegian linear reflectance to snow cover algorithm (NLR) (Solberg et al., 2006). The input to the NLR algorithm is the normalized difference snow index signal (NDSI- signal) (Salomonson and Appel, 2004). When cloud cover is present, there is no information on the snow-covered area.

The MODIS images were used for visual and quantitative comparison of the snow melt pattern from satellite images and from both SURFEX/Crocus experiments. For this purpose, four dates with cloud free conditions were selected throughout the melt season: 15 March 2015, 20 April 2015, 15 May 2015 and 04 July 2015.

3 Results

3.1 Snow depth

A density scatter plot of daily observed and modeled simulated snow depth for both experiments is shown in Fig. 3. The plot includes simulations for and the two winter seasons 2014/15 and 2015/16. 16 is shown in Fig. 3. Only data points where either the simulated or the observed snow depth is larger than 0 cm are taken into account. GridObs-Crocus is in reasonably good agreement with the observations (although there are cases of over- and underestimation of around 100 cm), while AROME-Crocus shows significant overestimation of snow depth . In order to show the performance of both experiments (in some cases more than 250 cm). To investigate the snow depth at individual stations over the full range of station altitudes in more detail, Fig. 4 shows snow depth plots for 6 locations: 2 located below 400 m, 2 located between 500 and 900 m, and 2 above 900

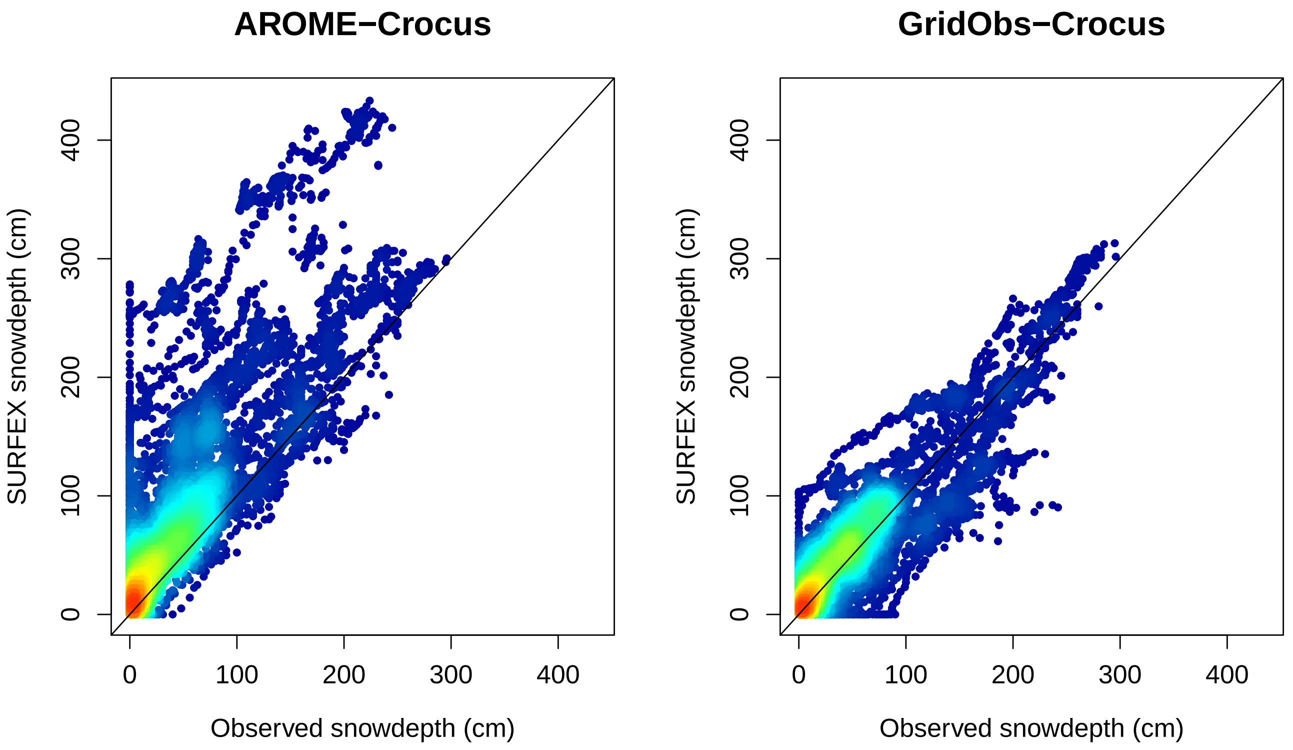

Figure 3. Scatter density plot of daily observed and simulated snow depth (cm) for ~~GridObs-Crocus~~ AROME-Crocus (~~in blue~~left) and ~~AROME-Crocus~~ for GridObs-Crocus (~~in red~~right) for all weather stations, from 01 September 2014 to 31 August 2016. The density ranges from low in blue to high in red.

m (which ~~for~~ in Norway means they are located above the tree line). For the location of these 6 stations within the domain, see Fig. 1 in which they are indicated with blue dots. Consistent with Fig. 3, Fig. 4 shows that AROME-Crocus overestimates the snow depth for all altitudes, and has more snow compared to GridObs-Crocus. This overestimation is especially strong for high altitude stations (situated above 900 m). The snow depth from GridObs-Crocus is closer to the observed snow depth, but at times underestimates the snow depth (most notably for the first winter season at Haukedal (329 masl.) and Midtstova (1162 masl.)). Episodes ~~where~~ when the snow depth decreases during the winter season (apart from snow melt in spring) are not always well captured by the SURFEX/Crocus experiments, and this issue is partly responsible for the overestimation of snow depth.

The results for Hemsedal II (604 masl., see Fig. 4) are of particular interest, as this is the only station measuring snow depth but not precipitation (and therefore not part of the gridded observation dataset used as input for GridObs-Crocus). GridObs-Crocus overestimates the snow depth at Hemsedal II, but not to the same extent as AROME-Crocus does. The bias in snow

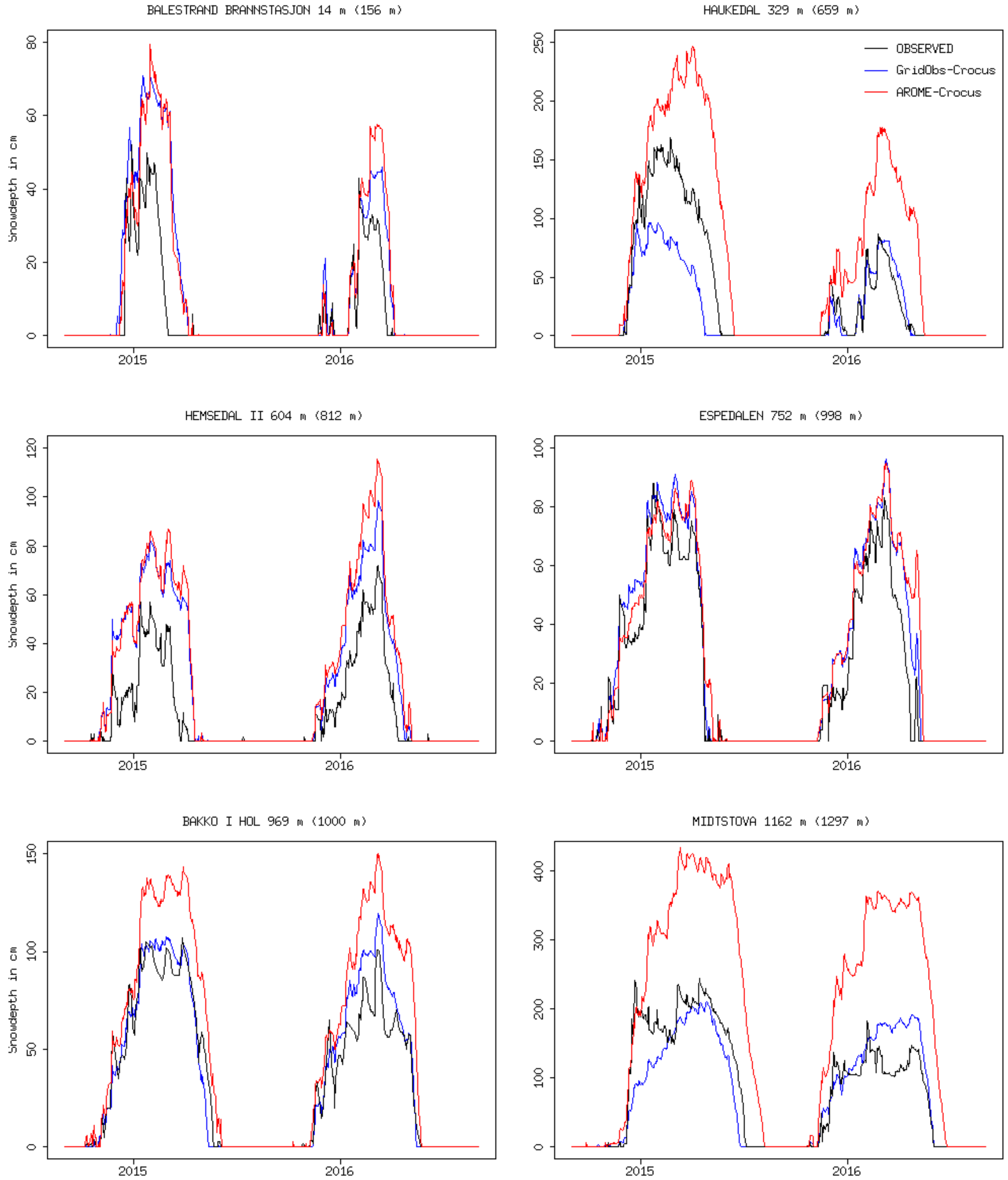

Figure 4. Observed and simulated snow depth (cm) at the location of six weather stations during the two winter seasons 2014-2016 (01 September 2014 - 31 August 2016) : 1) GridObs-Crocus (blue); 2) AROME-Crocus (red) and 3) observations (black). The elevation (in masl.) of the station is indicated above each plot, with in parentheses the elevation of the grid point in SURFEX/Crocus. The location of the 6 stations within the domain is indicated by blue dots in Fig. 1.

depth at Hemsedal II for the two seasons combined is 25 cm for GridObs-Crocus matches the observed pattern of increases and decreases more closely than (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 for most other 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.

The strong overestimation at Midtstova (see Fig. 4) by AROME-Crocus can be explained by the fact that Midtstova is located in an area with systematic and relatively large overestimation of precipitation in AROME-MetCoOp. In addition, AROME-MetCoOp underestimates the temperature by about 2 degrees in this area during winter. This can be seen in verification reports of the AROME MetCoOp model, for example in Homleid and Tveter (2016). The underestimation of In the forcing data for Midtstova we find a bias of -1.5 degrees for AROME-Crocus, compared to -0.8 degrees for GridObs-Crocus. This bias is larger than the overall bias for all nine stations measuring temperature: -0.5 degree for AROME-Crocus and -0.2 degree for GridObs-Crocus. During the snow accumulation season the temperature at Midtstova leads to even more of the already overestimated model precipitation falling is mostly well below freezing level. There are a few episodes each winter with temperatures just above zero, where the underestimated temperature in AROME-MetCoOp means the precipitation during those episodes comes as snow instead of rain, but these do not add up to large amounts. Midtstova is also a high-mountain station, which is very exposed to strong wind. Redistribution of snow due to wind is not captured in the SURFEX/Crocus model. GridObs-Crocus shows much more realistic results for Midtstova, although there is an underestimation during the first part of the 2014-2015 winter. From 27 October 2014 until 26 January 2015, the precipitation sensor at Midtstova was out of order, and the forcing from GridObs-Crocus for Midtstova will therefore be represented by interpolated values from surrounding stations, which might explain the underestimation.

An evaluation of the precipitation forcing data for AROME-Crocus and GridObs-Crocus for the six stations from Fig. 4 reveals that while the rainfall amount is often quite similar in the two experiments, the largest differences are found in the snowfall amount. AROME-Crocus consistently shows a larger amount of snowfall (accumulated over a year) compared to GridObs-Crocus, often about twice as much. The differences are largest for the stations located in the western part of the domain. As described in section 2.2, this region receives the highest amounts of winter precipitation in our study area.

Table 2 summarizes the bias over all stations for the two winter seasons (01 September 2014 - 31 August 2015 and 01 September 2015 - 31 August 2016). The bias was calculated as the mean of the differences between simulated and observed snow depth, and only for the days where there is snow present in the observations or at least one of the experiments. GridObs-Crocus shows a significantly smaller bias (4 and 9 cm) compared to AROME-Crocus (42 and 41 cm). The maximum observed snow depth is on average 112 cm for 2014-2015 and 88 cm for 2015-2016. GridObs-Crocus shows a very small bias (0 and 9 cm respectively), while AROME-Crocus overestimates the mean maximum snow depth by 43-45 cm. 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.

	2014-2015 Biassnow depth GridObs-Crocus (cm)	4 Observ
Bias snow depth AROME-Crocus Snow depth (cm)	-	+4
Mean length Length snow season (observed, days)	151	137 Bias length snow season Gr
Average date Date start of snow season (observeddays)	15 November	-2
Average date Date end of snow season (observeddays)	02 May	-3
Bias end snow season GridObs-Crocus Date max snow (days)	-3 30 January	2+13
Bias end snow season AROME-Crocus (daysMax snow (cm)	17 112	110

Table 2. Bias in snow depth, length of snow season (defined as number of days with more than 5 cm snow depth), start of snow season (defined as first day with more than 5 cm snow), end of snow season (defined as the day after the last day with more than 5 cm snow), the date for the maximum snow depth and the maximum snow depth. The two snow seasons run from 01 September 2014 to 31 August 2016. A negative bias in days means a too early date for the start/end/max snow, and a positive bias in days means a later date compared to observations. GridObs-Crocus is abbreviated to GridObs and AROME-Crocus to AROME.

RMSE	2014-2015		2015-2016	
	GridObs	AROME	GridObs	AROME
Average date max snow (observedSnow depth (cm)	30 January 29	22 February68	27	68
Bias average date max snow GridObs-Crocus Length snow season (days)	13 25	1344	21	28
Bias average date max snow AROME-Crocus Date start of snow season (days)	26 10	18	5	8
Average max snow (observed, cmDate end of snow season (days)	112 15	8826	12	17
Bias max snow GridObs-Crocus (cmDate max snow (days)	0 31	938	24	17
Bias max snow AROME-Crocus Max snow (cm)	43 30	45 68	28	70

Table 3. Bias in RMSE for snow depth, length of snow season(defined as number of days with more than 5 cm snow depth), start of snow season(defined as first day with more than 5 cm snow), end of snow season(defined as the day after the last day with more than 5 cm snow), the date for the maximum snow depth and the maximum snow depth. The two snow seasons run from 01 September 2014 to 31 August 2016. A negative bias in days means a too early date for the start/end/max snow, GridObs-Crocus is abbreviated to GridObs and a positive bias in days means a later date compared AROME-Crocus to observationsAROME.

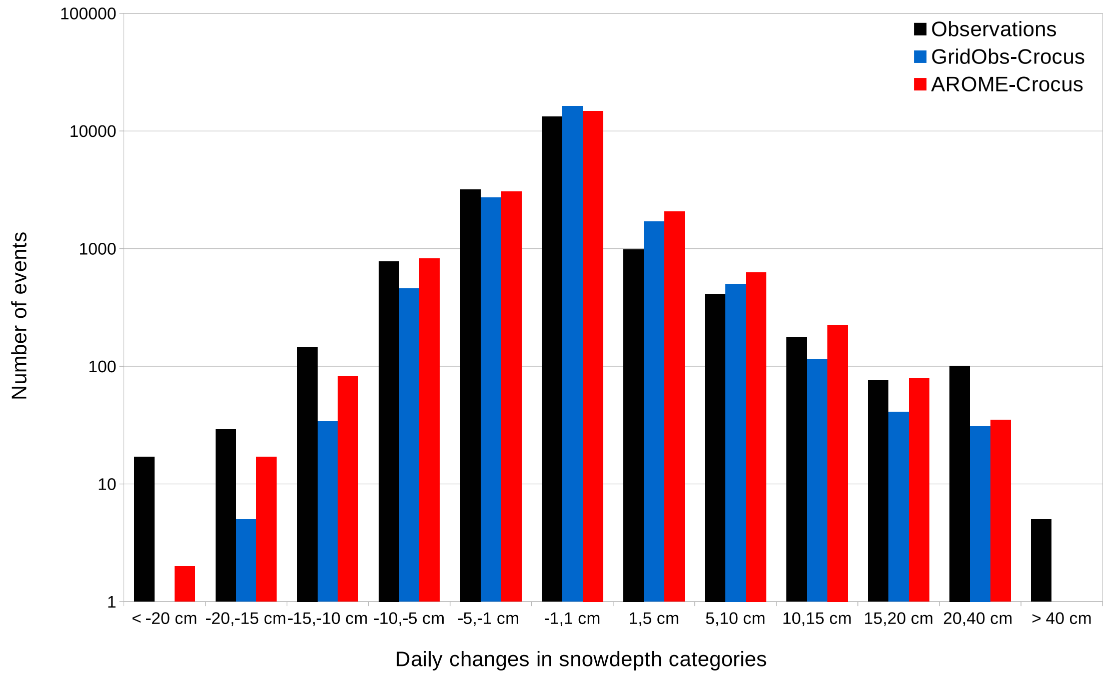

Figure 5. Categorical frequency distribution of daily changes in snow depth for observations (in black), GridObs-Crocus (in blue) and AROME-Crocus (in red), for all stations during 01 September 2014 - 31 August 2016. The y axis is on a logarithmic scale.

As snow depth accumulates over the winter season, a missed (or under/over estimated) snow event can influence the remainder of the season. It can therefore be useful to look at daily snow depth variations instead, as was also done by Quéno et al. (2016) and Schirmer and Jamieson (2015). Figure 5 shows the categorical frequency distribution of the daily change in snow depth for six accumulation categories, five decrease categories and one category centered around zero accumulation, on a logarithmic scale. The observed snow depth is measured in centimeter units, this is why there are no observations in the -0.5,0,5 cm category. The first two accumulation categories (up to 10 cm) are overestimated in both GridObs-Crocus and AROME-Crocus. The strongest observed increase category (>40 cm) as well as the strongest decrease category (< -20 cm) are strongly underestimated in both SURFEX/Crocus experiments.

SURFEX/Crocus does not allow for either the transport of blowing snow or in stand-alone mode does not account for wind-induced ablationsnow redistribution, which can be a large contributor to strong decreases in snow depth. Figure 4 showed that episodes of a decrease in snow depth (not including the snow melt at the end of the season) were not always captured well by the models, and it could be that blowing snow is the cause of this. Following Quéno et al. (2016), two diagnostics have been applied to look into this issue: blowing snow days and melting snow days. Blowing snow days are defined as days during which the wind speed (at a height of 10 m) during the past 24 hours (between 06 and 06 UTC, as this is when snow depth measurements are made) exceeds 8 m /ss^{-1}, while the snow surface temperature is below 0 °Celsius (since only dry snow can

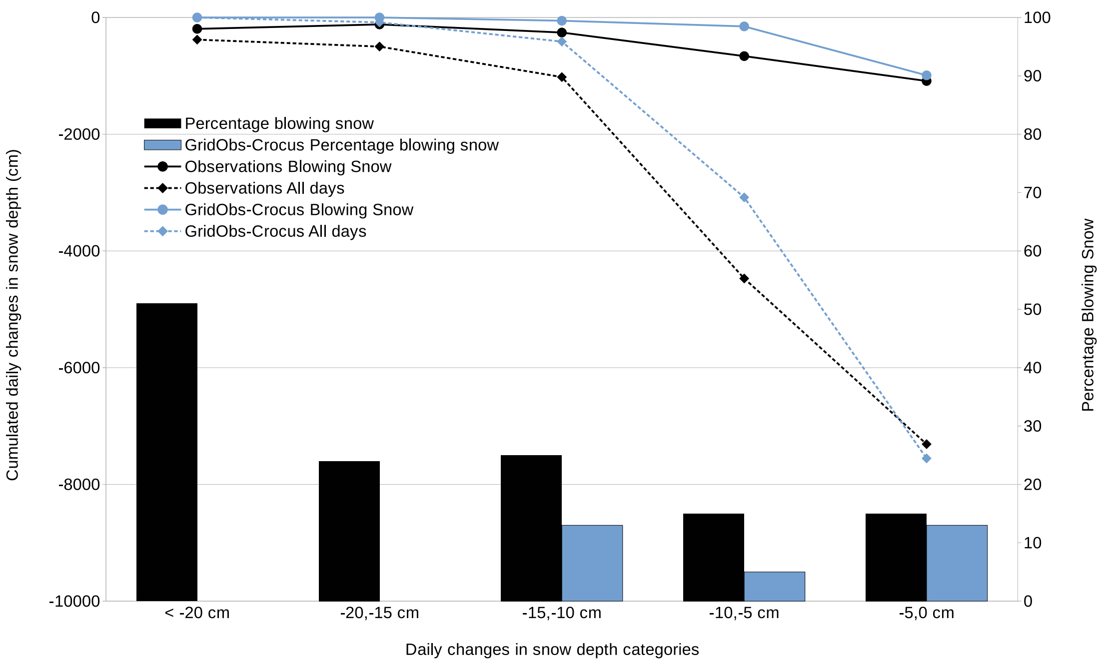

Figure 6. Cumulated daily change in snow depth for observations (in black), GridObs-Crocus (in blue) for all stations during 01 September 2014 - 31 August 2016, for blowing snow days (solid lines) and all days with decreasing snow depth (dashed lines). The columns show the percentage of snow loss that is caused by blowing snow, for observations (black) and GridObs-Crocus (blue).

be blown away). The temperature of the snow surface is taken from SURFEX the SURFEX/Crocus output for 12 UTC each day. The wind speed is taken from AROME-MetCoOp, which is used as forcing in both SURFEX/Crocus experiments. 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. This shows that the modeled wind speed can be used to determine blowing snow days. The wind threshold of 8 m /ss^{-1} for dry snow transport is taken from Li and Pomeroy (1997). Figure 6 shows the cumulated amount of the daily changes in snow depth for 5 categories of decreasing snow depth for blowing snow days and for all days where the snow depth decreases, for observations and for GridObs-Crocus, as well as the percentage of snow depth loss due to blowing snow. The cumulated amount of snow decrease is underestimated for nearly all categories. For the strongest decreasing rate (more than 20 cm), the observations indicate that 51% of the decrease in snow is caused by blowing snow. This category is not represented by GridObs-Crocus. For GridObs-Crocus, blowing snow days only contribute to the smallest decrease categories. In total (over all categories), blowing snow days contribute to 17% of the cumulated decrease in snow depth in the observations, while this amounts to 10% in GridObs-Crocus.

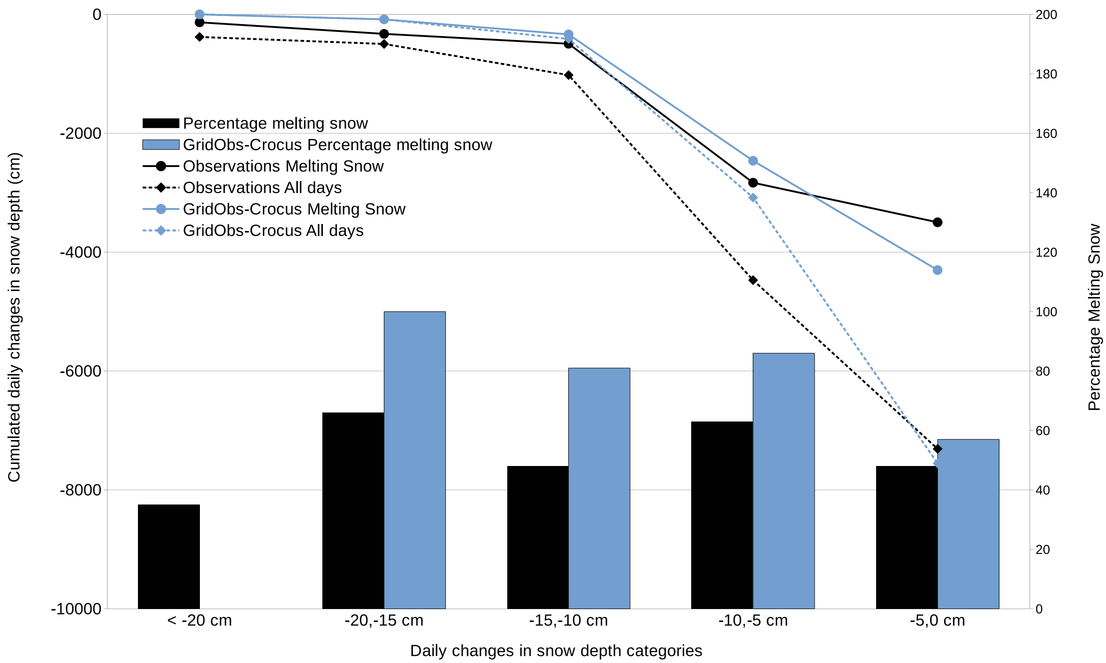

Figure 7. Cumulated daily changes in snow depth for observations (in black), GridObs-Crocus (in blue) for all stations during 01 September 2014 - 31 August 2016, for melting snow days (solid lines) and all days with decreasing snow depth (dashed lines). The columns show the percentage of snow loss that is caused by melting snow, for observations (black) and GridObs-Crocus (blue).

Melting snow days are defined as days ~~where~~ when the surface temperature of the snow is ~~at or above~~ 0 °Celsius. Figure 7 is similar to Fig. 6, but for melting snow days. For GridObs-Crocus, melting snow is the main responsible factor contributing to a decrease in snow depth. The largest decrease category is not represented by GridObs-Crocus, but for the other categories, melting snow is responsible for 57 - 100% of the decrease in snow depth. This is not surprising as SURFEX/Crocus does not represent blowing snow, so decrease in snow depth is caused by either snow melt or other processes such as snow compaction. The cumulated daily changes in snow depth for melting snow days as well as all days with a decrease in snow depth are underestimated by GridObs-Crocus for all categories except the smallest one (less than 5 cm loss in snow depth). This shows there is a general underestimation of snow ablation, as well as an underestimation of snow melt in GridObs-Crocus. The same goes for AROME-Crocus (not shown).

SURFEX/Crocus does have an option to run with sublimation in case of snowdrift. This option has been tested for two stations from Fig. 4: Midtstova and Hemsedal II. In this experiment, SURFEX/Crocus was run twice in 1D mode for these 2 locations: one experiment with identical settings as GridObs-Crocus, and one nearly identical with the exception of the option for sublimation in case of snowdrift (GridObs-Crocus+BS). The results are shown in Fig. 8. For both locations, the snow depth in GridObs-Crocus+BS is decreased, as expected. For Hemsedal II, this reduction is an improvement compared to the GridObs-experiment, which overestimated the snow depth. The bias in GridObs-Crocus was +16 cm for Hemsedal II, which

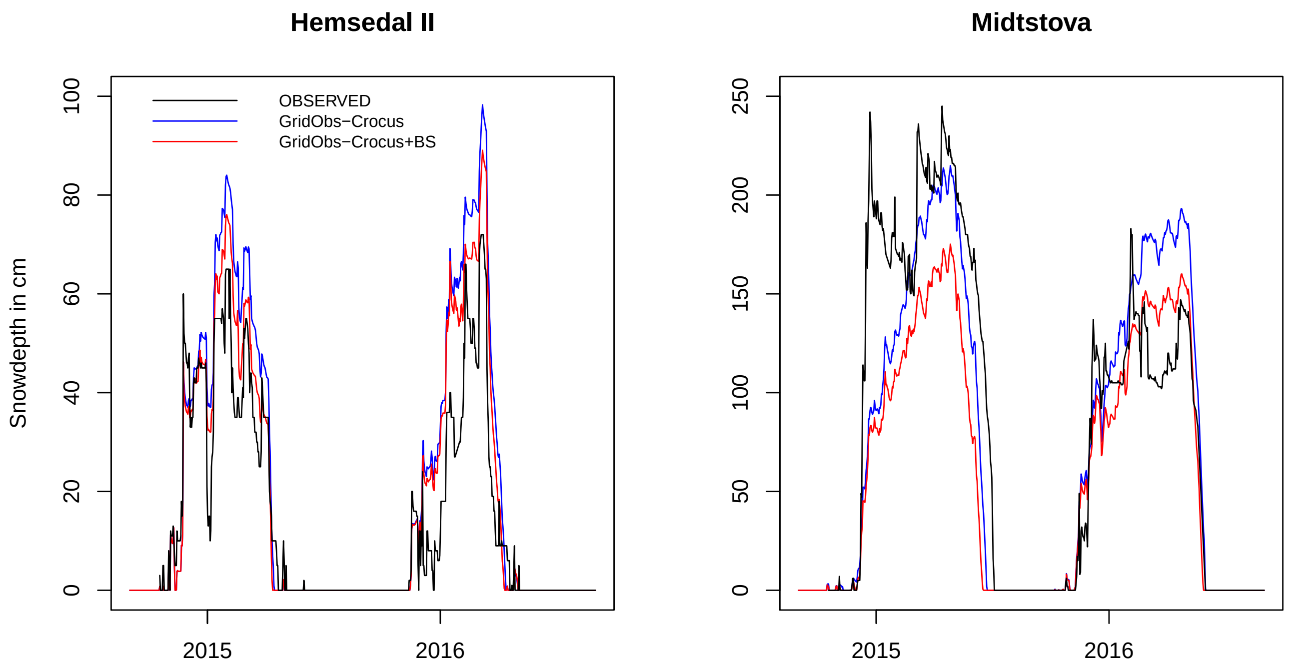

Figure 8. Observed and simulated snow depth (cm) at the location of Hemsedal II and Midtstova during the two winter seasons 2014-2016 (01 September 2014 - 31 August 2016) : 1) GridObs-Crocus 1D experiment (blue); 2) GridObs-Crocus+BS 1D experiment with sublimation loss during blowing snow events (red) and 3) observations (black).

has improved to +10 cm in GridObs-Crocus+BS. For Midtstova, GridObs-Crocus underestimates the snow depth for most of the two winter seasons (bias: -5 cm), and this underestimation is significantly larger in the GridObs+BS experiment (-28 cm).

3.2 Characteristics of the snow season

Table 2 shows statistics for the two snow seasons 2014/2015 and 2015/2016. The length of the snow season is defined as the number of days with more than 5 cm snow during a year. For GridObs-Crocus, the length of the snow season is overestimated by ~~8-10~~ 8-11 days (see table 2), while AROME-Crocus has a bias of ~~18-33~~ 18-34 days. The same positive bias was found by Vionnet et al. (2016). One possible explanation of this bias is the fact that the SURFEX/Crocus model assumes a uniform snow cover from the moment snow is present on the ground, and therefore shows less variability in snow cover compared to observations. In observations, there is often a period where the snow cover fluctuates - for example thinning to below 5 cm after the first snow has fallen and before a continuous snow cover has been established for the winter season. The start of the snow season is defined as the first day with more than 5 cm of snow, and the end of the snow season as the day after the last

Jaccard Index	GridObs-Crocus	AROME-Crocus
15 March 2015	0.92	0.99
20 April 2015	0.82	0.94
15 May 2015	0.65	0.82
04 July 2015	0.19	0.68

Table 4. Jaccard index for the snow covered areas shown in Fig. 10. A score of 1 means the image perfectly matches the MODIS image, a score of 0 means there is no overlap between the image from the experiment compared to the MODIS image.

day with more than 5 cm of snow. A negative bias in the start of the snow season means a too early start, while a positive bias means a too late start of the snow season. GridObs-Crocus has a bias of only two days (negative for the first winter and positive for the second winter) for the start of the snow season, while the snow season starts up to 10 days too early in AROME-Crocus during the first year (the second year has a bias of zero days). The season ends on average in late April or early May. In GridObs-Crocus, the season ends 3 days early during the first year, and 2 days late during the second year. In AROME-Crocus, the season ends 11-17 days too late. The observed maximum snow depth occurs on average at the end of January during the first year, and late February in the second year. Both experiments show a later date for the maximum snow depth.

Figure 9 show the distribution of the bias in the start and end of the snow season, as well as the date of maximum snow depth, for all 30 stations and for two winter seasons. Most stations show a bias near zero (between -5 and +5 days) for the start of the snow season. In AROME-Crocus, the snow season often starts too early. The bias for the end of the snow season shows more variability and an even clearer difference between the two models: GridObs-Crocus ends the snow season too early while AROME-Crocus ends the season too late. The bias for the date of the maximum snow depth of the season is mostly around zero for most stations and both models, but there are some outliers especially towards the strong positive bias. This is due to stations like for example Midtstova in Fig. 4, where the maximum observed snow depth occurs rather early in the season, while both experiments show a maximum much later in the season.

3.3 Snow-cover pattern

Figure 10 shows the spatial pattern of snow cover over the SURFEX/Crocus domain compared to MODIS data over the same area. The snow covered area is shown at different dates throughout the snow melt season: 15 March, 20 April, 15 May and 04 July 2015. On 15 March 2015, nearly the whole area is covered with snow. The only exceptions are areas right besides the fjords (white areas) in the west (well captured by both experiments), and at the bottom of valleys in the east (not captured by the SURFEX/Crocus experiments). On 20 April 2015, the snow has clearly started to melt in the valleys to the east. This is captured well by AROME-Crocus, while GridObs-Crocus shows too little snow around the valleys in the southeast of the domain. By 15 May 2015, a lot of snow had disappeared in the eastern part of the domain, while the western part has not changed much from the previous month. The average date for the end of the snow season for the 2014-2015 season was 02 May 2015 (see table 2),

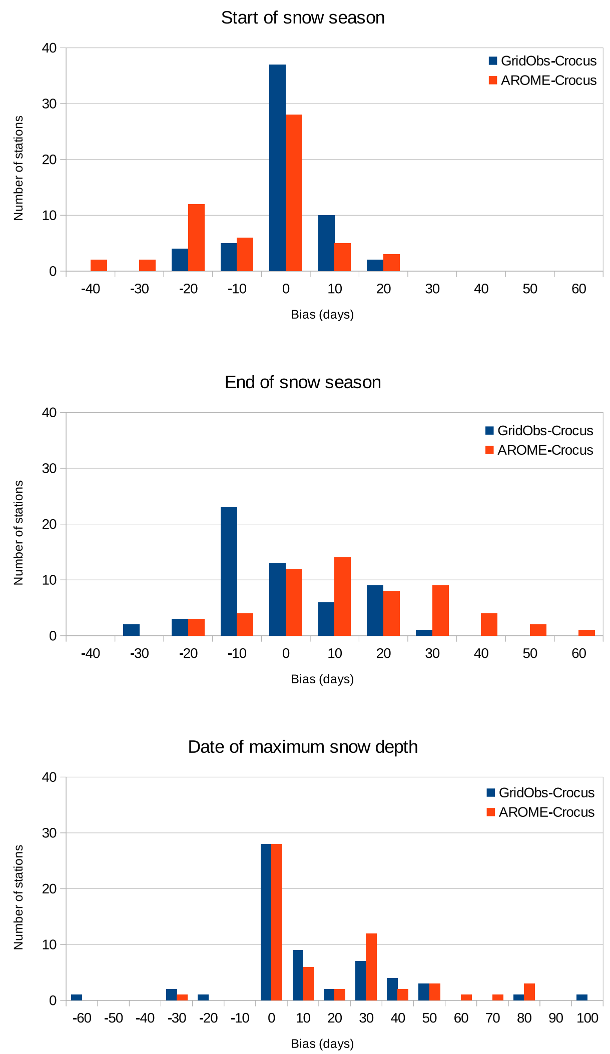

Figure 9. Distribution of the bias in start of snow season (top), end of snow season (middle) and date of maximum snow depth (bottom), for all 30 stations and for 2 winter seasons.

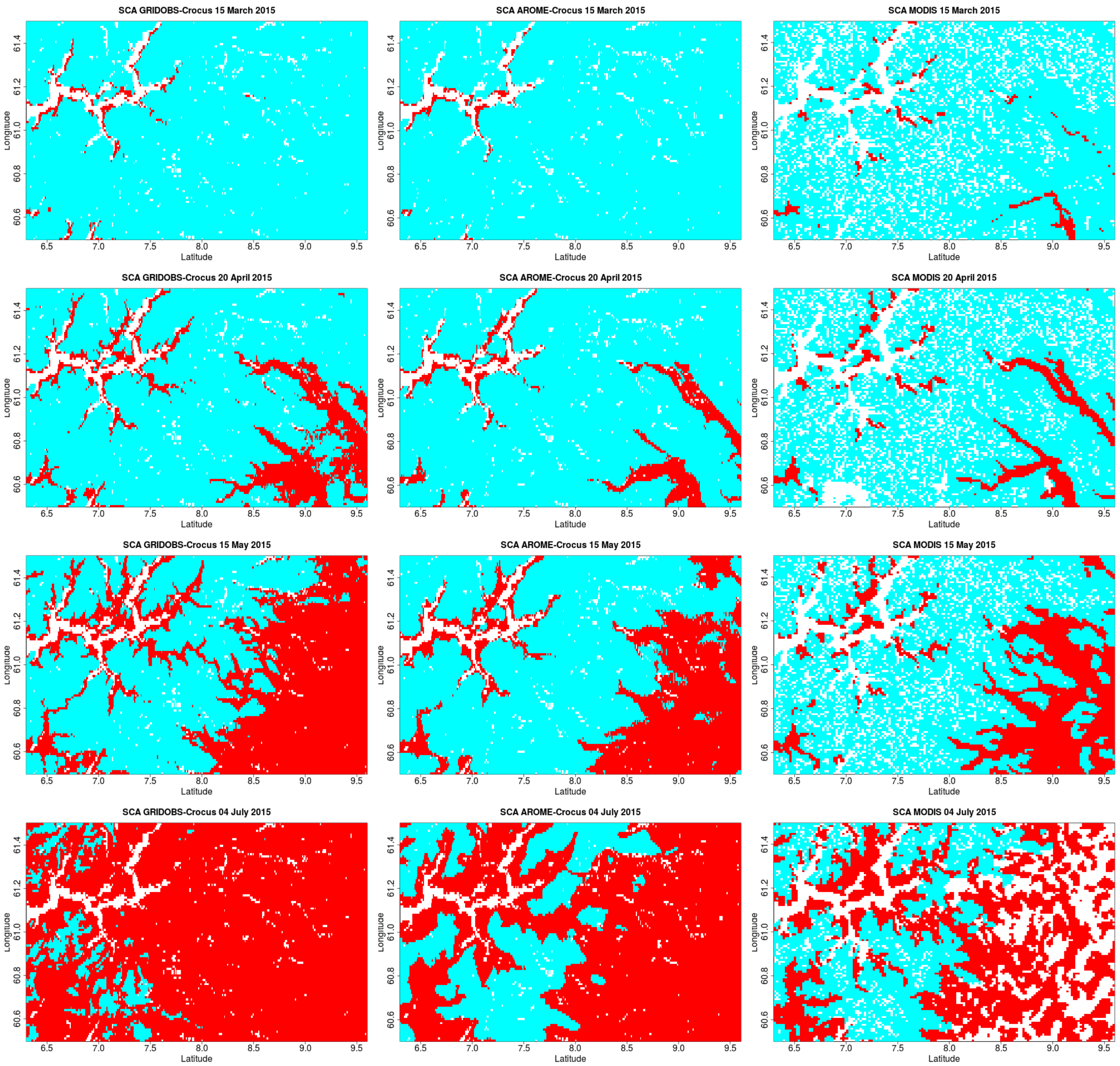

Figure 10. Snow covered area (where cyan is snow, red is no snow, and white is missing data or water surfaces) for GridObs-Crocus (left column), AROME-Crocus (middle column) and from MODIS satellite images (right column), for (rows from top to bottom): 15 March 2015, 20 April 2015, 15 May 2015 and 04 July 2015.

but the dates of the end of the snow season for individual stations range from 18 February until 05 July. Again, AROME-Crocus captures the snow cover pattern better than GridObs-Crocus. By 04 July 2015, the snow cover is limited to areas with higher elevation. AROME-Crocus captures the spatial pattern of snow cover very well. In GridObs-Crocus, nearly all snow has melted now, and the snow-covered area is underestimated. Earlier it was shown (in Fig. 9) that GridObs-Crocus has a negative bias (too late) for the end of the snow season for the 30 snow depth stations, while AROME-Crocus has a positive bias (too early). As discussed previously, the differences between the snowfall amounts from the two precipitation forcing datasets is largest on the west side of the domain; AROME-Crocus receives about twice as much snow compared to GridObs-Crocus in this region. This explains why the differences between GridObs-Crocus and AROME-Crocus in Fig. 10 are also largest in this area.

Table 4 shows the Jaccard indices for the images from Fig. 10. The Jaccard index was also used by for example Quéno et al. (2016) . It is a similarity index applied to the snow cover images which were remapped onto the same grid (which means that the snow cover from the MODIS images used to calculate the Jaccard index has a lower resolution than the one shown in Fig. 10). The Jaccard index is calculated as $J(X,Y) = |X \cap Y|/|X \cup Y|$, where X and Y are the simulated and observed snow cover, respectively. The number of grid points that are snow-covered in both SURFEX/Crocus and in the MODIS image is divided by the total amount of snow-covered grid points (in either SURFEX/Crocus or MODIS). When the Jaccard index equals 1, there is a perfect match between snow-covered grid points, and when the Jaccard index equals 0, there is no match at all. Table 4 shows that AROME-Crocus consistently has higher Jaccard indices compared to GridObs-Crocus. The indices decrease (for both experiments) during the melt season.

4 Discussion

~~The results are promising;~~ Although both experiments are capable of simulating the snow pack over the two winter seasons~~.~~ ~~There is however~~ , there is an overestimation of snow depth in the AROME-Crocus experiment~~(when using only meteorological forecasts from AROME MetCoOp), although~~ , even though the snow-covered area throughout the melt season is better represented by this experiment. The errors in AROME-Crocus accumulate over the snow season, showing that assimilation of snow depth observations into SURFEX/Crocus might be necessary when using ~~only~~ meteorological forecasts as forcing. When using gridded observations, the simulation of snow depth is significantly improved, which shows that using a combination of gridded observations and meteorological forecasts to force a snowpack model is very useful and can give better results than only using meteorological forecasts. There is an underestimation of snow ablation in both experiments, which is due to a combination of the absence of wind-induced erosion of snow and underestimation of snow melt in SURFEX/Crocus, and biases in the forcing data. The quality of the model validation, forcing data set and snowpack model are further discussed below.

4.1 Quality of the model validation

Nearly all (29) of the 30 stations that measure snow depth also measure precipitation (7 measure hourly precipitation, while 22 measure daily precipitation), which means the observed precipitation from these stations are used in the gridded observations dataset used to force GridObs-Crocus. Only 9 out of the 30 stations also measure temperature. Although precipitation is not

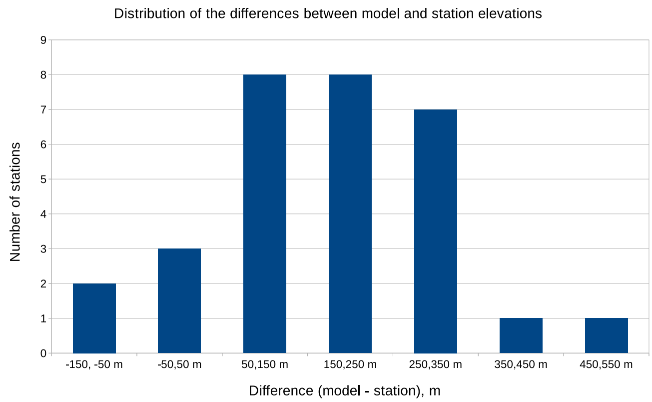

Figure 11. Differences between station elevation and the height of the station in the SURFEX/Crocus model.

directly related to snow depth, and temperature also plays an important role, it could still be argued that the GridObs-Crocus results are best in the locations of the observations that are included in the gridded dataset used to force SURFEX/~~crocus.~~ Crocus. 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.

The orography used in the SURFEX/Crocus experiments has a resolution of 1 km. Because of this, there are differences between the actual station height and the average height used for the center of the 1 km grid cell in SURFEX/Crocus. Figure 11 shows the distribution of those differences for all 30 stations. The average bias is 180 m. Most stations are placed at higher elevations in the model as compared to their actual elevation, but it should also be kept in mind that a grid point in SURFEX/Crocus describes a larger area (and range of elevation) compared to the actual observations. Especially in the mountainous region in the west of the domain (see Fig. 1), with high mountains (with steep slopes) and deep valleys, there are very large differences in height within a distance range of 1 km. An average bias of 180 m is therefore acceptable, but of course this can mean that an observation location might not be representative for the grid point it represents in SURFEX/Crocus. It could be argued that SURFEX grid points with a large height difference compared to the corresponding station should not be used for verification purposes. However, when calculating bias and RMSE values only for the stations with a height difference below 250 m, this did not have a large impact on the results. With a few exceptions, the bias and RMSE are lower when excluding stations with a large height difference, but the differences in overall bias are only $\pm$ 3 cm, which does not change any of the conclusions in this study.

4.2 Quality of the forcing data sets

Raleigh et al. (2015) showed that snow simulations are more sensitive to biases in forcing data than random errors, and that precipitation bias is the most important factor. There is a negative bias in the gridded precipitation used in GridObs-Crocus ~~(?)~~ (Lussana et al., 2018a) , especially for data-sparse areas and for intense precipitation. Missing episodes of intense snowfall would explain part of the underestimation of the snow depth in GridObs-Crocus. There are plans to improve the gridded observations of precipitation by adjusting the solid precipitation to account for the wind undercatch, and by post-processing of the predicted precipitation fields to adjust for bias ~~(?)~~ (Lussana et al., 2018a) . AROME-MetCoOp is known to overestimate the occurrence of precipitation events of less than 10 mm (Müller et al., 2017), and Fig. 5 showed that AROME-Crocus overestimated daily changes in snow depth up to 15 cm. An evaluation of the accumulated precipitation from the two forcing datasets showed that while the rainfall amounts are rather similar, the snowfall amount in AROME-Crocus is much larger compared to GridObs-Crocus, especially at the west side of the domain. This suggests that the overestimation of total precipitation in AROME-MetCoOp might be limited to the snowfall amount.

In this study, the forcing data from AROME-MetCoOp is interpolated from the original 2.5 km resolution to the 1 km resolution required by SURFEX/Crocus, apart from the temperature which is a post-processed, terrain-adjusted variable with a resolution of 500 m. The interpolation of AROME-MetCoOp data over a domain with complex topography could lead to differences in elevation (between AROME-MetCoOp and SURFEX/Crocus), which might lead to bias in for example the precipitation. We believe this might affect the AROME-Crocus experiment more than the GridObs-Crocus experiment, since the two experiments use different data sources.

Another uncertainty lies in the use of a fixed threshold temperature of 0.5 °Celsius to distinguish between rainfall and snowfall in GridObs-Crocus. This simplification could result in some actual snow events characterized as rainfall, and to a lesser extent the other way around.

Sauter and Obleitner (2015) investigated the sensitivity of SURFEX/Crocus snowpack modeling on Svalbard (Arctic Norway) to input parameters, and found that for higher elevations (in the accumulation zone), precipitation and radiation are the key factors in the evolution of the snowpack and contribute most to the model uncertainty. At lower elevations, precipitation was less important but factors such as wind speed or surface roughness increased in importance. Quéno et al. (2017) used satellite products of incoming solar and longwave radiation to force the SURFEX/Crocus model, however they concluded that improved meteorological forcing does not always lead to more accurate snowpack simulations, due to error compensations within the atmospheric forcing and the snowpack model.

4.3 Quality of the snowpack model

The SURFEX/Crocus model assumed a uniform snow cover when SWE reaches the relatively low threshold of 1 kg ~~/m?~~m^{-2}. The SURFEX/Crocus model was originally developed for use in high alpine regions, where there is not a lot of vegetation. In those areas, the assumption of the uniform snow cover is realistic, as there is no interaction with vegetation, but for areas covered with forest and closer to sea level this could lead to an overestimation of the snow cover. ~~This in turn would lead to an~~

overestimation of the snow depth, which would be expected in both experiments. AROME-Crocus does show an overestimation of snow depth. This is not the case for GridObs-Crocus, but this is partly due to the negative bias of the gridded observations of precipitation used as forcingWhen the snow cover is overestimated, the albedo will be too high and this will slow down the snow melt at the end of the season. This might explain the underestimated snow melt in both experiments.

The SURFEX/Crocus model grid is a collection of independent grid points with no transport of snow or other variables between grid points. It is therefore not possible to simulate the redistribution of snow by wind. It can be argued that with a resolution of approximately 1 km, the drifting snow would anyway be redistributed within the area of a grid point and not transported to neighboring grid points. Vionnet et al. (2014) showed that for explicit simulation of wind-induced snow transport a spatial resolution of less than 50 meters is required. This is currently not a feasible option for snowpack simulations over larger domains. There is an option in the SURFEX/Crocus model to calculate the rate of sublimation in case of snowdrift, which results in a loss of snow. In this study this option was not used. Using this option might improve the results of the AROME-Crocus experiment, but further decrease of the snow cover in This option was tested for two stations in his study, using the GridObs-Crocus would not improve the snow simulations as it would increase the underestimation of forcing dataset. As expected, this resulted in a decrease in snow depth and a decrease in season length. As GridObs-Crocus already underestimates the snow depth and snow cover, this is not an improvement.

Figure 5 showed that both SURFEX/Crocus experiments underestimate the melting of snow, further supplemented by Fig. 7 for the GridObs-Crocus experiment. Underestimated melting was also found by Quéno et al. (2016, 2017) and Vionnet et al. (2016), and complementary studies are needed to investigate the cause of this issue.

? Lafaysse et al. (2017) developed an ensemble snowpack model using SURFEX/Crocus called ESCROC (Ensemble System Crocus) to address modeling errors. They found that by using optimal members they were able to explain more than half of the simulation errors, and those ensembles have a significantly better predictive power than the classical deterministic approach. For future work, it would be interesting to use ESCROC and investigate the effect of different physical settings of SURFEX/Crocus. In addition, since November 2016, AROME-MetCoOp is run as an ensemble with 10 members, called MEPS (MetCoOp Ensemble Prediction System). This means that an ensemble of meteorological forcing is another possible direction for future work. Vernay et al. (2015a) used the 35 members of the ensemble prediction system based on the French NWP model ARPEGE as forcing to the SURFEX/Crocus model. The results indicated that accounting for the uncertainty in meteorological forecast significantly improves the skill and the usefulness of the model chain.

5 Conclusions

In this study we analyze the effect of using gridded observations of temperature and precipitation (the two most important variables for snow modeling) to force the snowpack model SURFEX/Crocus, compared to using forcing from numerical weather predictions only. Two years of model simulations were used (01 September 2014 - 31 August 2016). The two experiments (AROME-Crocus using only meteorological weather forecasts from AROME-MetCoOp, including a few post-processed variables, as forcing, and GridObs-Crocus using gridded observations of temperature and precipitation combined with meteorological forecasts from AROME-MetCoOp) were validated against snow depth observations and against MODIS images of snow-covered area. The main findings are as follows:

- Forcing SURFEX/Crocus with gridded observations of temperature and precipitation significantly improves the simulated snow depth (a bias of 6 cm and RMSE of 68 cm). This shows that using a combination of gridded observations and meteorological forecasts as forcing to SURFEX/Crocus can give better results than ~~only~~ mainly using meteorological forecasts. When using ~~purely~~ AROME-MetCoOp forcing data, the snow depth is strongly overestimated (a bias of 42 cm and RMSE of 68 cm), which was also shown by Quéno et al. (2016) and Vionnet et al. (2016).

- ~~While using only weather forecasts from~~ The AROME-MetCoOp ~~as forcing for SURFEX/Crocus resulted in an overestimating of the snow depth for individual stations, this experiment (AROME-Crocus) better represented the spatial pattern of snow cover~~ forcing data provided the best results on representing the spatial snow cover pattern throughout the melt season. Using gridded observations of temperature and ~~precipitations~~ precipitation as forcing for the SURFEX/Crocus model resulted in a snow cover that melted away too fast by the end of the season.

- Snow melt is underestimated in both experiments. This ~~appears to~~ might be an issue with the SURFEX/Crocus model, and needs to be further investigated.

- Blowing snow (which is not simulated by SURFEX/Crocus) contributes to 17% of all decreases in snow depth, and to 50% of the strongest decreases of more than 20 cm of snow depth loss in a day.

  ~~For future work, it would be interesting to run the~~ Using the option in SURFEX/Crocus ~~model with the option for~~ of running with sublimation in case of snowdrift ~~. In addition, the~~ is not enough to address this issue.

The dataset of gridded observations, and specifically the gridded precipitation is under ~~development. It~~ continuous development. For future work, it would be interesting to repeat this experiment using the new dataset which adjusts the solid precipitation to account for the wind undercatch and post-processes the predicted precipitation fields to adjust for the negative bias. ~~Using the ensemble snowpack model ESCROC~~ 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). Using the multi-physical ensemble system ESCROC (Ensemble System Crocus), and/or an ensemble of meteorological forcing would be an another interesting topic for future work. Finally, when using AROME-MetCoOp as forcing data for running SURFEX/Crocus at a resolution higher than 2.5 km, terrain adjustment routines should be applied to the generation of forcing data. In this study we accounted for local terrain effects, by using post-processed AROME-MetCoOp temperature and wind, but this could be extended to other variables.

The findings in this study have improved our understanding of regional snow modeling in Norway, which is important for not only water resource planning and flood forecasting, but also for impact studies related to climate change and winter climate. ~~The results may also~~ Running the SURFEX/Crocus model in gridded version for Norwegian conditions using a combination of data sources (raw and post-processed weather predictions and observations) is very promising. The result from this study is very valuable information which may be used for future development of a system for daily snow mapping in Norway.

Data availability. Snow depth and meteorological variables from the stations used in this study are freely available through eklima.met.no (eKlima, 2017) . AROME-MetCoOp forecasts are available through http://thredds.met.no/thredds/metno.html. The gridded dataset of temperature and precipitation is available at http://thredds.met.no/thredds/catalog/metusers/senorge2/seNorge2/archive/catalog.html. Hourly temperature and precipitation data is available from 2010 up to the present day. For daily temperature and precipitation data, the archive goes back to 1957 and can be downloaded at http://doi.org/10.5281/zenodo.845733. The data are also shown on the web-portals www.senorge.no and www.xgeo.no (both in Norwegian only). The SURFEX-Crocus simulations for both experiments can be made available for research purposes by contacting the authors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors are grateful for the funding of this study by the ~~research council~~ Research Council of Norway through the project "Better SNOW models for prediction of natural hazards and HydropOWer applications" (SNOWHOW), led by Thomas Skaugen at the Norwegian Water Resources and Energy Directorate (NVE). We would like to thank our fellow participants in the SNOWHOW project for invaluable help and discussions: Thomas Skaugen, Tuomo Saloranta (NVE), Karsten M??ller (NVE), Kjetil Melvold (NVE) and Sjur Kolberg (SINTEF). In addition, we would like to thank Tuomo Saloranta for providing the processed MODIS images, and Cristian Lussana (MET Norway) for valuable help and discussions. We are also very grateful to two anonymous reviewers, for their time, effort and very helpful suggestions, which resulted in a greatly improved paper.

References

Barfod, E., Müller, K., Saloranta, T., Andersen, J., Orthe, N., Wartianien, A., Humstad, T., Myrabø, S., and Engeset, R.: The expert tool XGEO and its applications in the Norwegian Avalanche Forecasting Service, in: International Snow Science Workshop Grenoble, October 07-11, 2013, Chamonix Mont-Blanc, France, 2013.

Bartelt, P. and Lehning, M.: A physical SNOWPACK model for the Swiss avalanche warning: Part I: numerical model, Cold Regions Science and Technology, 35, 123 – 145, http://www.sciencedirect.com/science/article/pii/S0165232X02000745, 2002.

Bellaire, S., Jamieson, J. B., and Fierz, C.: Forcing the snow-cover model SNOWPACK with forecasted weather data, The Cryosphere, 5, 1115–1125, https://doi.org/10.5194/tc-5-1115-2011, https://www.the-cryosphere.net/5/1115/2011/, 2011.

Bellaire, S., Jamieson, J. B., and Fierz, C.: Corrigendum to "Forcing the snow-cover model SNOWPACK with forecasted weather data" published in The Cryosphere, 5, 1115-1125, 2011, The Cryosphere, 7, 511–513, https://doi.org/10.5194/tc-7-511-2013, https://www.the-cryosphere.net/7/511/2013/, 2013.

Bergstrøm, S.: Development and application of a conceptual runoff model for Scandinavian catchments, SMHI report RH07, Swedish Meteorological and Hydrological Institute, Norrköping, Sweden, 1976.

Bernier, N. B., Bélair, S., Bilodeau, B., and Tong, L.: Near-surface and land surface forecast system of the Vancouver 2010 Winter Olympic and Paralympic Games, Journal of Hydrometeorology, 12, 508–530, 2011.

Best, M., Pryor, M., Clark, D., Rooney, G., Essery, R., Ménard, C., Edwards, J., Hendry, M., Porson, A., Gedney, N., et al.: The Joint UK Land Environment Simulator (JULES), model description–Part 1: energy and water fluxes, Geoscientific Model Development, 4, 677–699, 2011.

Bokhorst, S., Pedersen, S. H., Brucker, L., Anisimov, O., Bjerke, J. W., Brown, R. D., Ehrich, D., Essery, R. L. H., Heilig, A., Ingvander, S., Johansson, C., Johansson, M., Jónsdóttir, I. S., Inga, N., Luojus, K., Macelloni, G., Mariash, H., McLennan, D., Rosqvist, G. N., Sato, A., Savela, H., Schneebeli, M., Sokolov, A., Sokratov, S. A., Terzago, S., Vikhamar-Schuler, D., Williamson, S., Qiu, Y., and Callaghan, T. V.: Changing Arctic snow cover: A review of recent developments and assessment of future needs for observations, modelling, and impacts, Ambio, 45, 516–537, https://doi.org/10.1007/s13280-016-0770-0, https://doi.org/10.1007/s13280-016-0770-0, 2016.

Boone, A., Masson, V., Meyers, T., and Noilhan, J.: The Influence of the Inclusion of Soil Freezing on Simulations by a Soil–Vegetation–Atmosphere Transfer Scheme, Journal of Applied Meteorology, 39, 1544–1569, https://doi.org/10.1175/1520-0450(2000)039<1544:TIOTIO>2.0.CO;2, https://doi.org/10.1175/1520-0450(2000)039<1544:TIOTIO>2.0.CO;2, 2000.

Brown, R., Vikhamar-Schuler, D., Bulygina, O., Derksen, C., Luojus, K., Mudryk, L., Wang, L., and Yang, D.: Arctic terrestrial snow cover. Chapter 3, in: Snow, Water, Ice and Permafrost in the Arctic (SWIPA) 2017, pp. 25–64, Arctic Monitoring and Assessment Programme (AMAP), Oslo, Norway, 2017.

Brown, R. D. and Robinson, D. A.: Northern Hemisphere spring snow cover variability and change over 1922-2010 including an assessment of uncertainty, The Cryosphere, 5, 219–229, https://doi.org/10.5194/tc-5-219-2011, https://www.the-cryosphere.net/5/219/2011/, 2011.

Brun, E., David, P., Sudul, M., and Brunot, G.: A numerical model to simulate snow-cover stratigraphy for operational avalanche forecasting, Journal of Glaciology, 38, 13–22, https://doi.org/10.3189/S0022143000009552, 1992.

Carrera, M. L., Bélair, S., Fortin, V., Bilodeau, B., Charpentier, D., and Doré, I.: Evaluation of snowpack simulations over the Canadian Rockies with an experimental hydrometeorological modeling system, Journal of Hydrometeorology, 11, 1123–1140, 2010.

Dyrrdal, A. V., Saloranta, T., Skaugen, T., and Stranden, H. B.: Changes in snow depth in Norway during the period 1961-2010, Hydrology Research, 44, 169–179, 2013.

eKlima: Free access to weather- and climate data from Norwegian Meteorological Institute from historical data to real time observations, http://eklima.met.no/ (Last visited: 15 September 2017), soon to be replaced by https://frost.met.no/, 2017.

Engeset, R.: National Avalanche Warning Service for Norway. -Established 2013, in: International Snow Science Workshop Grenoble, October 07-11, 2013, Chamonix Mont-Blanc, France, 2013.

Essery, R.: A Factorial Snowpack Model (FSM 1.0), Geoscientific Model Development, 8, 3867–3876, https://doi.org/10.5194/gmd-8-3867-2015, 2015.

Essery, R., Morin, S., Lejeune, Y., and Ménard, C. B.: A comparison of 1701 snow models using observations from an alpine site, Advances in Water Resources, 55, 131 – 148, https://doi.org/http://dx.doi.org/10.1016/j.advwatres.2012.07.013, http://www.sciencedirect.com/science/article/pii/S0309170812002011, snow–Atmosphere Interactions and Hydrological Consequences, 2013.

Etchevers, P., Martin, E., Brown, R., Fierz, C., Lejeune, Y., Bazile, E., Boone, A., Dai, Y., Essery, R., Fernandez, A., Gusev, Y., Jordan, R., Koren, V., Kowalcyzk, E., Nasonova, N., Pyles, R., Schlosser, A., Shmakin, A., Smirnova, T., Strasser, U., Verseghy, D., Yamazaki, T., and Yang, Z.: Validation of the energy budget of an alpine snowpack simulated by several snow models (SnowMIP project), Annals of Glaciology, 38, 150–158, https://doi.org/10.3189/172756404781814825, International Symposium on Snow and Avalanches, Davos, Switzerland, 2-6 June, 2003, 2004.

FAO/IIASA/ISRIC/ISS-CAS/JRC: Harmonized World Soil Database (version 1.2), Tech. rep., FAO, Rome, Italy and IIASA, Laxenburg, Austria, 2012.

Fierz, C., Bavay, M., Wever, N., and Lehning, M.: SNOWPACK: where do we stand today?, in: International Snow Science Workshop, EPFL-TALK-197625, 2013.

Foppa, N., Stoffel, A., and Meister, R.: Synergy of in situ and space borne observation for snow depth mapping in the Swiss Alps, International journal of applied earth observation and geoinformation, 9, 294–310, 2007.

Habets, F., Boone, A., and Noilhan, J.: Simulation of a Scandinavian basin using the diffusion transfer version of ISBA, Global and Planetary Change, 38, 137 – 149, https://doi.org/https://doi.org/10.1016/S0921-8181(03)00016-X, http://www.sciencedirect.com/science/article/pii/S092181810300016X, project for Intercomparison of Land-surface Parameterization Schemes, Phase 2(e), 2003.

Hall, D. K. and Riggs, G. A.: Accuracy assessment of the MODIS snow products, Hydrological Processes, 21, 1534–1547, https://doi.org/10.1002/hyp.6715, http://dx.doi.org/10.1002/hyp.6715, 2007.

Hanssen-Bauer, I., Førland, E. J., Haddeland, I., Hisdal, H., Mayer, S., Nesje, A., Nilsen, J., Sandven, S., Sandø, A., Sorteberg, A., and Ådlandsvik, B.: Klima i Norge 2100. Kunnskapsgrunnlag for klimatilpasning oppdatert i 2015, Tech. Rep. 2, Norsk klimaservicesenter, https://cms.met.no/site/2/klimaservicesenteret/klima-i-norge-2100/_attachment/10990, 2015.

Hanssen-Bauer, I., Førland, E. J., Haddeland, I., Hisdal, H., Lawrence, D., Mayer, S., Nesje, A., Nilsen, J., Sandven, S., Sandø, A., Sorteberg, A., and Ådlandsvik, B.: Climate in Norway 2100. – A knowledge base for climate adaptation, Tech. Rep. 1, Norwegian Climate Service Centre, 2017.

Homleid, M.: Diurnal corrections of short-term surface temperature forecasts using the Kalman filter, Weather and Forecasting, 4, 689–707, 1995.

Homleid, M. and Tveter, F. T.: Verification of Operational Weather Prediction Models December 2015 to February 2016, Met.no report 18, Norwegian Meteorological Institute, Oslo, Norway, https://www.met.no/publikasjoner/met-info/met-info-2016/_/attachment/download/a541f975-ff95-4a38-aa4c-d47c70264ebe:11d67cb5abaffca7686fea6be5ef3b00437f9ea2/MET-info-20-2016.pdf, 2016.

Horton, S., Schirmer, M., and Jamieson, B.: Meteorological, elevation, and slope effects on surface hoar formation, The Cryosphere, 9, 1523–1533, 2015.

Johansson, C., Pohjola, V., Jonasson, C., and Callaghan, T.: Multi-decadal changes in snow characteristics in sub-Arctic Sweden, Ambio, 40, 566–574, 2011.

Kivinen, S., Rasmus, S., Jylhä, K., and Laapas, M.: Long-Term Climate Trends and Extreme Events in Northern Fennoscandia (1914–2013), 2017, https://doi.org/10.3390/cli5010016, 2017.

Klein, A. G. and Stroeve, J.: Development and validation of a snow albedo algorithm for the MODIS instrument, Annals of Glaciology, 34, 45–52, https://doi.org/10.3189/172756402781817662, 2002.

Lafaysse, M., Cluzet, B., Dumont, M., Lejeune, Y., Vionnet, V., and Morin, S.: A multiphysical ensemble system of numerical snow modelling, The Cryosphere, 11, 1173–1198, https://doi.org/10.5194/tc-11-1173-2017, https://www.the-cryosphere.net/11/1173/2017/, 2017.

Lehning, M., Bartelt, P. B., Brown, R. L., Fierz, C., and Satyawali, P.: A physical SNOWPACK model for the Swiss Avalanche Warning Services. Part II: Snow Microstructure, Cold Regions Science and Technology, 35, 147–167, 2002.

Li, L. and Pomeroy, J. W.: Estimates of Threshold Wind Speeds for Snow Transport Using Meteorological Data, Journal of Applied Meteorology, 36, 205–213, https://doi.org/10.1175/1520-0450(1997)036<0205:EOTWSF>2.0.CO;2, 1997.

Lussana, C., Salvati, M. R., Pellegrini, U., and Uboldi, F.: Efficient high-resolution 3-D interpolation of meteorological variables for operational use, Advances in Science and Research, 3, 105–112, https://doi.org/10.5194/asr-3-105-2009, https://www.adv-sci-res.net/3/105/2009/, 2009.

Lussana, C., Tveito, O. E., and Uboldi, F.: senorge v2.0: an observational gridded dataset of temperature for norway, Met.no report 14, Norwegian Meteorological Institute, Oslo, Norway, 2016.

Lussana, C., Saloranta, T., Skaugen, T., Magnusson, J., Tveito, O. E., and Andersen, J.: seNorge2 daily precipitation, an observational gridded dataset over Norway from 1957 to the present day, Earth System Science Data, 10, 235–249, https://doi.org/10.5194/essd-10-235-2018, https://www.earth-syst-sci-data.net/10/235/2018/, 2018a.

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, 2018b.

Magnusson, J., Wever, N., Essery, R., Helbig, N., Winstral, A., and Jonas, T.: Evaluating snow models with varying process representations for hydrological applications, Water Resources Research, 51, 2707–2723, https://doi.org/10.1002/2014WR016498, http://dx.doi.org/10.1002/2014WR016498, 2015.

Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari, S., Barbu, A., Boone, A., Bouyssel, F., Brousseau, P., Brun, E., Calvet, J.-C., Carrer, D., Decharme, B., Delire, C., Donier, S., Essaouini, K., Gibelin, A.-L., Giordani, H., Habets, F., Jidane, M., Kerdraon, G., Kourzeneva, E., Lafaysse, M., Lafont, S., Lebeaupin Brossier, C., Lemonsu, A., Mahfouf, J.-F., Marguinaud, P., Mokhtari, M., Morin, S., Pigeon, G., Salgado, R., Seity, Y., Taillefer, F., Tanguy, G., Tulet, P., Vincendon, B., Vionnet, V., and Voldoire, A.: The SURFEXv7.2 land and ocean surface platform for coupled or offline simulation of earth surface variables and fluxes, Geoscientific Model Development, 6, 929–960, https://doi.org/10.5194/gmd-6-929-2013, https://www.geosci-model-dev.net/6/929/2013/, 2013.

Mohr, M.: New routines for gridding of temperature and precipitation observations for “seNorge.no, Met.no Report 8, Norwegian Meteorological Institute, Oslo, Norway, 2008.

Müller, M., Homleid, M., Ivarsson, K.-I., Køltzow, M. A. Ø., Lindskog, M., Midtbø, K. H., Andrae, U., Aspelien, T., Berggren, L., Bjørge, D., Dahlgren, P., Kristiansen, J., Randriamampianina, R., Ridal, M., and Vignes, O.: AROME-MetCoOp: A Nordic Convective-Scale Operational Weather Prediction Model, Weather and Forecasting, 32, 609–627, https://doi.org/10.1175/WAF-D-16-0099.1, https://doi.org/10.1175/WAF-D-16-0099.1, 2017.

Quéno, L., Vionnet, V., Dombrowski-Etchevers, I., Lafaysse, M., Dumont, M., and Karbou, F.: Snowpack modelling in the Pyrenees driven by kilometric resolution meteorological forecasts, The Cryosphere, 10, 1571–1589, 2016.

Quéno, L., Karbou, F., Vionnet, V., and Dombrowski-Etchevers, I.: Satellite products of incoming solar and longwave radiations used for snowpack modelling in mountainous terrain, Hydrology and Earth System Sciences Discussions, 2017, 1–33, https://doi.org/10.5194/hess-2017-563, https://www.hydrol-earth-syst-sci-discuss.net/hess-2017-563/, 2017.

Raleigh, M. S., Lundquist, J. D., and Clark, M. P.: Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework, Hydrology and Earth System Sciences, 19, 3153–3179, https://doi.org/10.5194/hess-19-3153-2015, https://www.hydrol-earth-syst-sci.net/19/3153/2015/, 2015.

Rasmus, S., Boelhouwers, J., Briede, A., Brown, I., Falarz, M., Ingvander, S., Jaagus, J., Kitaev, L., Mercer, A., and Rimkus, E.: Recent change – Terrestrial cryosphere, in: Second Assessment of Climate Change for the Baltic Sea Basin, edited by Team, T. B. I. A., pp. 117–129, Springer, 2015.

Ruan, G. and Langsholt, E.: Rekalibrering av flomvarslingas HBV-modeller med inndata fra seNorge, versjon 2.0, Tech. Rep. 71, NVE Report, Oslo, Norway, 2017.

Sælthun, N. R.: The Nordic HBV model, NVE Report No. 7, Norwegian Water Resources and Energy Administration, Oslo, Norway, 1996.

Salomonson, V. V. and Appel, I.: Estimating fractional snow cover from MODIS using the normalized difference snow index, Remote Sensing of Environment, 89, 351–360, https://doi.org/10.1016/j.rse.2003.10.016, 2004.

Saloranta, T. M.: Simulating snow maps for Norway: description and statistical evaluation of the seNorge snow model, The Cryosphere, 6, 1323–1337, https://doi.org/10.5194/tc-6-1323-2012, https://www.the-cryosphere.net/6/1323/2012/, 2012.

Saloranta, T. M.: Operational snow mapping with simplified data assimilation using the seNorge snow model, Journal of Hydrology, 538, 314–325, https://doi.org/10.1016/j.jhydrol.2016.03.061, 2016.

Sauter, T. and Obleitner, F.: Assessing the uncertainty of glacier mass-balance simulations in the European Arctic based on variance decomposition, Geoscientific Model Development, 8, 3911–3928, https://doi.org/10.5194/gmd-8-3911-2015, https://www.geosci-model-dev.net/8/3911/2015/, 2015.

Schirmer, M. and Jamieson, B.: Verification of analysed and forecasted winter precipitation in complex terrain, The Cryosphere, 9, 587–601, https://doi.org/10.5194/tc-9-587-2015, https://www.the-cryosphere.net/9/587/2015/, 2015.

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C., and Masson, V.: The AROME-France convective-scale operational model, Monthly Weather Review, 139, 976–991, https://doi.org/10.1175/2010MWR3425.1, 2011.

Skaugen, T.: Studie av skilletemperatur for snø ved hjelp av samlokalisert snøpute, nedbør- og temperaturdata., 1998.

Skaugen, T., Stranden, H. B., and Saloranta, T.: Trends in snow water equivalent in Norway (1931-2009), Hydrology Research, 43, 489–499, 2012.

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

Solberg, R., Amlien, J., and Koren, H.: A review of optical snow cover algorithms. Norwegian Computing, Tech. Rep. SAMBA/40/06, Norwegian Computing Center, Oslo, Norway, 2006.

Sturm, M., Holmgren, J., and Liston, G. E.: A seasonal snow cover classification system for local to global applications, Journal of Climate, 8, 1261–1283, 1995.

Uboldi, F., Lussana, C., and Salvati, M.: Three-dimensional spatial interpolation of surface meteorological observations from high-resolution local networks, Meteorological Applications, 15, 331–345, https://doi.org/10.1002/met.76, http://dx.doi.org/10.1002/met.76, 2008.

Vernay, M., Lafaysse, M., Mérindol, L., Giraud, G., and Morin, S.: Ensemble forecasting of snowpack conditions and avalanche hazard, Cold Regions Science and Technology, 120, 251 – 262, https://doi.org/https://doi.org/10.1016/j.coldregions.2015.04.010, http://www.sciencedirect.com/science/article/pii/S0165232X15000981, 2015a.

Vernay, M., Lafaysse, M., Mérindol, L., Giraud, G., and Morin, S.: Ensemble forecasting of snowpack conditions and avalanche hazard, Cold Regions Science and Technology, 120, 251 – 262, https://doi.org/https://doi.org/10.1016/j.coldregions.2015.04.010, http://www.sciencedirect.com/science/article/pii/S0165232X15000981, 2015b.

Vikhamar-Schuler, D., Müller, K., and Engen-Skaugen, T.: Snow modeling using SURFEX with the CROCUS snow scheme, Met.no report 7, Norwegian Meteorological Institute, Oslo, Norway, 2011.

Vikhamar-Schuler, D., Hanssen-Bauer, I., Schuler, T. V., Mathiesen, S. D., and Lehning, M.: Use of a multilayer snow model to assess grazing conditions for reindeer, Annals of Glaciology, 54, 214–226, https://doi.org/10.3189/2013AoG62A306, 2013.

Vikhamar-Schuler, D., Isaksen, K., Haugen, J. E., Tømmervik, H., Luks, B., Schuler, T. V., and Bjerke, J. W.: Changes in winter warming events in the Nordic Arctic Region, Journal of Climate, https://doi.org/http://dx.doi.org/10.1175/JCLI-D-15-0763.1, 2016.

Vionnet, V., Brun, E., Morin, S., Boone, A., Faroux, S., Le Moigne, P., Martin, E., and Willemet, J.-M.: The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2, Geoscientific Model Development, 5, 773–791, https://doi.org/10.5194/gmd-5-773-2012, https://www.geosci-model-dev.net/5/773/2012/, 2012.

Vionnet, V., Martin, E., Masson, V., Guyomarc'h, G., Naaim-Bouvet, F., Prokop, A., Durand, Y., and Lac, C.: Simulation of wind-induced snow transport and sublimation in alpine terrain using a fully coupled snowpack/atmosphere model, The Cryosphere, 8, 395–415, https://doi.org/10.5194/tc-8-395-2014, https://www.the-cryosphere.net/8/395/2014/, 2014.

Vionnet, V., Dombrowski-Etchevers, I., Lafaysse, M., Quéno, L., Seity, Y., and Bazile, E.: Numerical Weather Forecasts at Kilometer Scale in the French Alps: Evaluation and Application for Snowpack Modeling, Journal of Hydrometeorology, 17, 2591–2614, 2016.

Vormoor, K. and Skaugen, T.: Temporal Disaggregation of Daily Temperature and Precipitation Grid Data for Norway, Journal of Hydrometeorology, 14, 989–999, https://doi.org/10.1175/JHM-D-12-0139.1, https://doi.org/10.1175/JHM-D-12-0139.1, 2013.