REVIEWER 1

This is a review for the second revision of "Snow cover prediction in the Italian Central Apennines using weather forecast and snowpack numerical models."

We thank the reviewer for the time spent evaluating our work and we thank her for the useful and constructive comments. We have answered all comments. Note that our answers are in **BLUE** in the following text.

The reviewers have addressed the majority of the review, however my core concern remains. Specifically, this is the concern that there is a fundamental scale mismatch between the forcing data, the grid square, the numerical model, and the evaluation dataset. To reiterate: the snowmodel is an O(mm) to O(cm) vertical resolution; the soil model is O(mm) to O(cm) parameterization (Richard's equation) that generally only holds at a point scale and is not representative of a single hillslope let alone 3km; the forcing is at 3km (missing key variability); the models are applied as homogenous for a 3km x 3km grid square; and finally are evaluated against point observations that they-themselves are often not representative of the landscape without careful transect application.
We are aware that there is a strong scale mismatch between the point observations and kilometer resolution of the modeled data. And even if upscaling techniques have been proposed (Hou et at. (2022), Horton and Haegeli (2022)), we believe point measurements of snowpack properties are usually preferable, since they are directly taken on the field, and their interpolation on a kilometer grid may even increase the degree of uncertainty in the evaluation of a model performances in reproducing the observations. Indeed, the interpolation of point snow measurements on a grid needs several assumptions on wind transport of snow, vegetation interception, slope aspect and elevation which may introduce uncertainty also on the values used for validation. For these reasons, point data are widely used in the snow science community to validate model forecasts. For example, Lute et al. (2022) used point observations of snow water equivalent from the SNOTEL sites to validate the SnowClim model at 210 m resolution, Avanzi et al. (2021) validated the S3M model using different sources of point measurements at 240 m resolution and Revuelto et al. (2018) validated Crocus model at 250 meters resolution using point observations of snow height. However, also at kilometric resolution the site measurements are still used to validate snow cover models, like in Chen et al. (2014), Bellaire et al. (2011), Bellaire et al. (2013), Schirmer and Jamieson (2015), Vionnet et al. (2012), Quéno et al. (2016) and Luijting et al. (2018). The majority of the papers cited above are already included in the manuscript introduction and are also briefly summarized. Concerning atmospheric forcing, we are aware that regional atmospheric models (albeit at km resolution) need to be parameterized for some subgrid processes: it is now and always will be due to the inherent nature of atmospheric turbulent motions. Therefore, we have used state-of-the-art parameterizations to simulate subgrid processes in our atmospheric model. Moreover, we made a great effort in validating our results with main meteorological measurements, in order to make sure that WRF simulations do not have a bias so large as to make any snow cover evaluation impossible. However, Terzago et al. (2020) showed that a coarse resolution atmospheric forcing for SNOWPACK produces only slightly worse results than higher quality atmospheric forcing data, despite the fact that they are only using one site for validation. Indeed, as suggested by Ikeda et al. (2010), Barlage er al. (2010) and Pavelsky et al. (2011), the scale mismatch between simulations at kilometer resolution and point observations can have a

large impact if only one site is used for validation, but if a large number of validation sites is used, the mean model bias tends to minimize, as the error at measurement sites tends to be randomly distributed. We also want to highlight that the measurement sites chosen in our manuscript are located in zones representative of large areas, in wide fields far from the canopy, thus particularly suitable to validate our simulations. However, in order to reduce the possible mismatch between the model resolution and the point observations, we divided the study domain in 200 meters elevation bands, from 800 m a.s.l. to 1800 m a.s.l. (Table 2) and for each elevation band we averaged the snow heights and snow water equivalents simulated and observed. We used the elevation-averaged values to produce the scatter plots and calculate the statistics.

- Lute, A. C., Abatzoglou, J., and Link, T.: SnowClim v1.0: high-resolution snow model and data for the western United States, Geosci. Model Dev., 15, 5045–5071, https://doi.org/10.5194/gmd-15-5045-2022, 2022.
- Avanzi, F., Gabellani, S., Delogu, F., Silvestro, F., Cremonese, E., Morra di Cella, U., Ratto, S., and Stevenin, H.: Snow Multidata Mapping and Modeling (S3M) 5.1: a distributed cryospheric model with dry and wet snow, data assimilation, glacier mass balance, and debris-driven melt, Geosci. Model Dev., 15, 4853–4879, https://doi.org/10.5194/gmd-15-4853-2022, 2022.
- Revuelto J, Lecourt G, Lafaysse M, Zin I, Charrois L, Vionnet V, Dumont M, Rabatel A, Six D, Condom T, Morin S, Viani A, Sirguey P. Multi-Criteria Evaluation of Snowpack Simulations in Complex Alpine Terrain Using Satellite and In Situ Observations. *Remote Sensing*. 2018; 10(8):1171. https://doi.org/10.3390/rs10081171
- Chen, F., et al. (2014), Modeling seasonal snowpack evolution in the complex terrain and forested Colorado Headwaters region: A model intercomparison study, *J. Geophys. Res. Atmos.*, 119, 13,795– 13,819, doi:10.1002/2014JD022167.
- 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://tc.copernicus.org/articles/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://tc.copernicus.org/articles/7/511/2013/, 2013.
- 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://tc.copernicus.org/articles/9/587/2015/, 2015.
- 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://gmd.copernicus.org/articles/5/773/2012/, 2012.
- 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,

https://doi.org/10.5194/tc-10-1571-2016,https://tc.copernicus.org/articles/10/1571/2016/, 2016.
- Luijting, H., Vikhamar-Schuler, D., Aspelien, T., Bakketun, Å., and Homleid, M.: Forcing the SURFEX/Crocus snow model with combined hourly meteorological forecasts and gridded observations in southern Norway, The Cryosphere, 12, 2123–2145, https://doi.org/10.5194/tc-12-2123-2018, https://tc.copernicus.org/articles/12/2123/2018/, 2018.
- Terzago, S., Andreoli, V., Arduini, G., Balsamo, G., Campo, L., Cassardo, C., Cremonese, E., Dolia, D., Gabellani, S., von Hardenberg, J., Morra di Cella, U., Palazzi, E., Piazzi, G., Pogliotti, P., and Provenzale, A.: Sensitivity of snow models to the accuracy of meteorological forcings in mountain environments, Hydrol. Earth Syst. Sci., 24, 4061–4090, https://doi.org/10.5194/hess-24-4061-2020, 2020.
- Pavelsky, T. M., Kapnick, S., and Hall, A.: Accumulation and melt dynamics of snowpack from a multiresolution regional climate model in the central Sierra Nevada, California, Journal of Geophysical Research: Atmospheres, 116, https://doi.org/https://doi.org/10.1029/2010JD015479,https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JD015479, 2011.
- Ikeda K, Rasmussen R, Liu C, Gochis D, Yates D, Chen F, Tewari M, Barlage M, Dudhia J, Miller K, Arsenault K. Simulation of seasonal snowfall over Colorado. Atmospheric Research. 2010 Sep 1;97(4):462-77.
- Barlage,M.,F. Chen,M. Tewari,K.Ikeda,D.Gochis,J.Dudhia,R. Rasmussen, B. Livneah, M. Ek, and K. Mitchell (2010), Noah land surface model modifications to improve snowpack prediction in the Colorado Rocky Mounains,
- Horton, S. and Haegeli, P.: Using snow depth observations to provide insight into the quality of snowpack simulations for regional-scale avalanche forecasting, The Cryosphere, 16, 3393–3411, https://doi.org/10.5194/tc-16-3393-2022, 2022.
- Hou, Yingxu, Xiaodong Huang, and Lin Zhao. 2022. "Point-to-Surface Upscaling Algorithms for Snow Depth Ground Observations" *Remote Sensing* 14, no. 19: 4840. https://doi.org/10.3390/rs14194840

Worrying about the microstructure (L466) seems to be really putting the cart before the horse and avoiding key processes improvements such as met downscaling.

We never pretended to compare observed snow microstructure with modeled microstructure at 3 km resolution, indeed the word "microstructure" appears only once at the end of the manuscript, when we suggest future possible developments of this work. We are actually working on higher resolution Alpine3D simulations (500 m) forced with WRF forecasts and auxiliary inputs, in order to evaluate the ability of the model in reproducing the snow microstructure observed at some particular sites. However we point out that Bellaire and Jamieson (2013),Horton et al. (2015) and Horton and Jamieson (2016) even evaluated the possibility of estimating some snowpack microstructure properties forcing snow cover models with forecasted weather data (these papers are already cited in the manuscript)

- Bellaire, S. and Jamieson, B.: Forecasting the formation of critical snow layers using a coupled snow cover and weather model, Cold Regions Science and Technology, 94, 37–44, https://doi.org/https://doi.org/10.1016/j.coldregions.2013.06.007, https://www.sciencedirect.com/science/article/pii/S0165232X13000840, 2013.
- Horton, S. and Jamieson, B.: Modelling hazardous surface hoar layers across western Canada with a coupled weather and snow cover model, Cold Regions

Science and Technology, 128, 22–31, https://doi.org/https://doi.org/10.1016/j.coldregions.2016.05.002, https://www.sciencedirect.com/science/article/pii/S0165232X16300854, 2016.

- Horton, S., Schirmer, M., and Jamieson, B.: Meteorological, elevation, and slope effects on surface hoar formation, The Cryosphere, 9, 1523–1533, https://doi.org/10.5194/tc-9-1523-2015, https://tc.copernicus.org/articles/9/1523/2015/, 2015

This also exists with the two (one unknown in Noah-LSM) of the fractional snowcover area parameterization used to diagnose the snow covered area. For example, how sensitive are the results to the fixed constants in this parameterization? Are there implicit scale assumptions in the fSCA parameterization? This is not discussed, so the appropriatness is not clear.

The parametrization for snow cover fraction used in Noah LSM is known (Koren et al. (1999)) and we also reported it in our manuscript (section 3.3). This parametrization is already implemented in Noah, while it is applied a-posteriori to Alpine3D, in order to compare modeled and retrieved fractional snow cover. We didn't tune the parametrization on our specific domain, since it was far from the objectives of this study, so we didn't do any sensitivity test on it. However, we are aware that other FSC parametrizations or other thresholds for deriving SCA maps may lead to different results. Further comments about it have been added to the manuscript in sections 4.2.3 and 5.

- Koren, V., Schaake, J., Mitchell, K., Duan, Q.-Y., Chen, F., and Baker, J. M.: A parameterization of snowpack and frozen ground intended for NCEP weather and climate models, Journal of Geophysical Research: Atmospheres, 104, 19 569–19 585, https://doi.org/https://doi.org/10.1029/1999JD900232, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/1999JD900232, 1999

In summary, the lack of clear explanation on the scale mismatches present in this manuscript remain my concern and I would strongly suggest this be incorporated into a subsequent revision to help contextualize the results.

L265 Same scale mismatch as Snowpack. How are you parameterizing the soils at a 3km resolution? What is the reason it does better in this context?

The soil has the same resolution of the atmospheric forcing. Our "alpha" test was with the soil simulation not active in Alpine3D, using the default "bucket" scheme for water transport. We then activated soil simulation, initialized as described in the manuscript, and we tested again the "bucket" scheme and Richards water transport scheme. We obtained better simulations in terms of snow height, snow water equivalent and snow cover extent with the Richards scheme, thus we decided to use it in our Alpine3D setup.

Since it was clear from the preliminary tests that the Richards scheme was more performing compared to the "bucket" scheme, we didn't include them in the sensitivity test shown in the supplementary materials.

L283 "hitting" word choice

The word has been corrected in the manuscript

L355 I'm surprised the SNOWPACK density estimation isn't better than Noah. This suggests to me a scale mismatch between the WRF forcing gridsquare, and the point model used to represent this 3km domain
At L355 we are not talking about density or snow water equivalent, but we are comparing simulated and measured snow height variations, which can be caused by densification and/or melting in our model setup (we don't consider snow erosion in the model). As we already discussed in the manuscript, WRF-Alpine3D presents slightly better performances compared to WRF-Noah in the estimations of the negative daily snow height variations. We also showed that WRF-Alpine3D has better performances than WRF-Noah in reproducing the snow height and snow water equivalent observed at the measurement sites. This suggests that the snowpack shrinking is mainly due to densification in WRF-Alpine3D and to melting in WRF-Noah, highlighting a strong difference between the models.

L367 w/c Totally
The word has been corrected in the manuscript

L377 “anticipated”?
The word has been corrected in the manuscript

L389 Reproduces what?
The sentence has been modified in the manuscript

L396 Note what month period this is
The sentence has been modified in the manuscript

L415 “Where already melted” awkward — where it has already melted?
Melted on the trees. The sentence has been modified in the manuscript.

L460 Could this be due to the FCA equation (eq2)? Or whatever Noah is using? Indeed this might not be a SD/SWE estimation mismatch but rather a limitation in the fractional snow cover param used.
Of course the parametrization has a large impact on the estimation of the snow cover fraction. We added a discussion on this aspect in the manuscript in sections 4.2.3 and 5.

L465 How representative would you expect the microstructure to be of a 3km domain
We agree with the reviewer that the microstructure of the snow at 3 km resolution can be hardly compared with the microstructure measured at a single point. Thus, as we also stated in the manuscript, we will (and we are already doing it actually) increase the resolution of WRF simulations (this time with Noah-MP LSM) to 1km, we will apply a subsequent downscaling to 500 m resolution and we will use these data, together with MODIS fractional snow cover observations and RADAR observed precipitation field to force Alpine3D at 500 m resolution. We believe that at this resolution and with the auxiliary inputs we will be able to make a more robust comparison of observed and simulated snow vertical profiles.

Fig3 No validation of longwave?

Unfortunately, in the study region there are no longwave sensors and we know that this is a large limitation for the model validation. Thus in the past year we installed two measurement sites in the study domain with several weather and snow sensors, and we also planned to install longwave sensors. We are going to use the measurements collected at our sites in the future.

Finally, we have to inform the Reviewer that in the past months, discussing with people of the Meteomont service, we discovered that the daily snow density data which they provide refer only to the density of the snowpack top layer, thus they are not mean values along the entire snowpack profile. So, using an independent dataset of almost weekly vertical snow density profiles, still provided by the Meteomont service, we made a linear model for the estimation of mean snowpack density starting from the density of the snowpack top layer. In this way we could still use the daily density measurements to calculate the snow water equivalent. This argument is treated in more detail in the supplementary materials.

EDITOR

Comments to the author:
Dear Authors,

Thank you for providing a revised version of your manuscript. After receiving one reviewer's comments as well as assessing the revised manuscript myself, I would like to invite you to respond and to prepare a further revision. In addition to considering the reviewer's second assessment, I would kindly request that you address the following points.

We thank the Editor for the time spent evaluating our work and we thank her for the useful and constructive comments. We have answered all comments. Note that our answers are in **BLUE** in the following text.

Technical comments:

1. Please reference the supplementary figures and texts directly in the main manuscript where appropriate.
As the Editor suggested, we explicitly referred to the supplementary materials in the manuscript when necessary

2. L30-33: Please combine into one sentence.
The sentences have been modified in the manuscript

3. L102: Please explicitly define "Cfb" in the text.
Cfb has been explicitly defined in the manuscript

4. Please explicitly define what is meant by "winter", as the synoptic discussion extends to February and March and some simulations seem to extend into April.
In the new manuscript we explicitly defined the selected periods of interest and we referred to them using the words “snow season”, replacing the word “winter”

5. I suggest replacing references to "cloud-resolving" at 3-km grid spacing to "convection-permitting"
We modified the text in the manuscript as the Editor suggested

6. L291: I suggest "By applying an arbitrary threshold of 51%" to clarify that the appropriateness of this choice was not investigated
We modified the text in the manuscript as the Editor suggested

7. The Noah model is a generalised land-surface model that is not intended to represent snow processes in great physical detail as in Alpine3D. Therefore, I suggest not referring to it as a "snow-cover model" (e.g. L335)
We replaced the therms “snow cover model” with “land surface model”, as this definition is more general and also preferable for Alpine3D (it is defined as “surface model” also in the model description page: https://alpine3d.slf.ch)

8. Please clarify where appropriate (e.g. Table 4) where statistics represent an average over all three winters

All the statistics are an average over the three snow seasons. It has been better specified in the manuscript

9. L386-391: Please indicate that this text refers to 2018-2019

It has been clarified in the manuscript

10. L415 "trees"

The typo has been corrected in the manuscript

11. Please indicate if/where the offline coupling scripts can be accessed by the research community

Actually all the interfacing scripts are a mix of Bash and Python. We are now developing a better code based only on Python language, which we will release in the future. Instead the datasets presented in this study can be obtained upon request

Scientific comments:

1. L79: it would be useful to indicate why the Noah LSM was selected instead of the updated Noah-MP LSM, as the latter has largely superseded the former in literature on high-resolution applications of WRF

We are aware that Noah-MP is a more sophisticated land surface model compared to Noah, however the use of Noah-MP takes us away from the goal of the study, which is to compare the abilities of a single-layer (Noah) and a multi-layer (Alpine3D) snowpack model to reproduce the observed values of snow height, snow water equivalent and snow extent using for both models the same meteorological forcing.

The sentence has been modified as follows:

"The aim of this study is to investigate the ability of a simple single-layer and a more sophisticated multi-layer snow cover numerical model to reproduce the observed snow height, snow water equivalent and snow extent in Central Apennines, using for both models the same forecast weather data as meteorological forcing. To this purpose, we use two well-known snow and soil models: i) Noah LSM, an Eulerian model which simulates the snowpack as a bulk single layer (Chen and Dudhia, 2001); ii) Alpine3D, a multi-layer Lagrangian model which simulates the snowpack stratification (Lehning et al., 2006)."

However in future works we will compare Noah-MP at 1km resolution and Alpine3D at 500m resolution, both driven by WRF at 1km resolution, using also auxiliary input data in order to evaluate the ability of the models to reproduce the observed snow vertical profiles.

2. L230: Please clarify further the simulation strategy in the methods section, as requested by Reviewer (R) 1

We clarified the simulation strategy for WRF, Noah and Alpine3D in the sections 3.2.1, 3.2.2 and 3.2.3, respectively.

3. L240: Please indicate the exact starting and end dates of all three simulations (and why they differ, if they do), as well as address R1's comment on how realistic the initial snow condition is. It would also be useful to provide a justification for using a different initial snow condition for the two simulations (as I understood it to be) and to address R2's original comment on how the simulation strategy impacts soil temperature/conditions.
The initial dates have been added to the manuscript. We didn't use different simulations setup for the three simulated snow seasons, and in order to avoid misunderstanding we better explained how the atmosphere and the soil have been initialized in the method section. We confirm also here that for each snow season the simulation started when there was no snow observed at all the measurement sites. Unfortunately we didn't investigate the impact of the soil temperature on the simulations, however we believe that the initial soil conditions in the models are representative of the state of the real soil because we started the simulations using NCEP analysis several days before the first snowfall occurred at the measurement sites. These considerations have been added to the manuscript in section 3.2.1.

4. L344, L377: How can the results be interpreted in terms of differential settlement as opposed to differential melting?
Unfortunately we couldn't calculate the differential melting because of the not constant measurement frequency of the snow water equivalent. However we discussed more in detail the snow water equivalent timeseries, showing that the snow water equivalent simulated with WRF-Alpine3D is well in line with the observations, instead WRF-Noah tends to simulate a snow water equivalent decrease rate faster than the observations. This causes also a faster decrease in the snow height in WRF-Noah compared to WRF-Alpine3D, suggesting that the main driver of snow compaction in Noah is the snow melt, while in Alpine3D is the snow densification. These comments have been added to the manuscript.

5. L349: Why is there a cluster of observed snow heights of ~150 cm?
We looked even more carefully than how we did before at all the timeseries of snow height and snow water equivalent in our dataset and we found some bad data coming from measurement errors. We removed that data and this caused a slight reduction of points in the scatterplots but resulted in better accordance of both models with the observations and no cluster of observed snow heights at ~150 cm. However this didn't change our conclusions.

6. L424-425: Where can the agreement vs. elevation be seen by the reader?
The agreement of the models with observations as a function of elevation can be qualitatively observed comparing Fig. 8 and the right panel of Fig. 1, which shows the topography of the study domain. We added this comment to the manuscript.

7. L454-455: On the other hand, WRF-A3D seems to overestimate max snow height towards the end of the 2021 simulation
We agree with the Editor comment and we added it in the manuscript.

8. For Section 2 of the Supplementary Information, I suggest presenting a figure or some statistics from the sensitivity simulations so that the reader can visualise the best agreement described

As the Editor suggests, we added to the supplementary materials a table showing the scores obtained in the snow height estimation for the combinations that presented MAE smaller than 18 cm and MBE between -1 cm and 1 cm.

Finally, we have to inform the Editor that in the past months, discussing with people of the Meteomont service, we discovered that the daily snow density data which they provide refer only to the density of the snowpack top layer, thus they are not mean values along the entire snowpack profile. So, using an independent dataset of almost weekly vertical snow density profiles, still provided by the Meteomont service, we made a linear model for the estimation of mean snowpack density starting from the density of the snowpack top layer. In this way we could still use the daily density measurements to calculate the snow water equivalent. This argument is treated in more detail in the supplementary materials.