

Response to Reviewer 2

Dear Reviewers and Editor,

We want to thank you for putting the time and effort into reviewing our manuscript and for your valuable comments and questions. We are certain that the revised manuscript will benefit from your input.

The main points we took away from both reviews is that the manuscript needs (1) a more thorough evaluation of the model uncertainties, (2) more contextualization of the model regarding existing modeling approaches and (3) a clearer structure and explanations.

To address these points, we propose the following main revisions:

- We will add a quantitative point-wise evaluation of wind direction and snow redistribution for the case study similar to the evaluation of wind speed. We will also expand the analysis with a contextualization of the SNOWstorm errors with regard to the errors of the benchmark HEF-LES, errors in the coarse-scale input data, as well as errors reported in comparable modeling approaches. We also want to point out that a direct comparison of reported error metrics in other modeling approaches is problematic, due to the differences in settings (e.g., temporal resolutions) and validation strategies. This last point leaves room for a coordinated model intercomparison experiment in the future (for blowing snow models on the one hand, and near-surface wind models on the other hand).
- We will add a new section on model limitations, which explicitly summarizes the model uncertainties as presented in the ideal-case experiments and the case study. Additionally, this section will summarize which settings the model is designed for, and which settings are not represented in the model. These will include the atmospheric boundary condition, topographic limitations, and limitations in the surface structure (e.g. focus on non-vegetated areas).
- We will add additional explanations on the model structure and specific choices, where the reviewers pointed out questions in the text. We will also re-make Fig.1 to facilitate understanding of the topographic analysis.
- We agree with the reviewer comment that the case study is rather short, and that conclusions drawn from this are only valid for this one case. However, we believe that evaluation against the HEF-LES (which introduces the short-period constraint) can provide valuable insights on a process level. We will therefore put an even stronger focus on the process perspective in our evaluation. This will include an analysis of which flow features in the HEF-LES are, or are not, represented in the SNOWstorm predictions.

In the following, we address all specific points; please find our answers in blue. References to equations, figures, and tables refer to the original numbering in the manuscript and might change in the revised version.

Yours sincerely,
Manuel Saigger on behalf of all co-authors

This is a review of SNOWstorm (v1.0). Overall, I really like the idea of going after blowing snow with deep learning. Blowing snow is critical to include, and it's very tricky. However, I think this manuscript needs more work and contextualization to be a strong contribution.

I found that the manuscript layout to be difficult to follow, with the mix of results and discussion together making for a difficult read to really figure out what was tested and where. For example, the conclusion notes that direction and speed were tested against observations, but I'm struggling to find this in the text. The heavy use of model variables, e.g., SUBL_VI, SNOW_VI, makes parsing the paragraphs difficult and makes it easy to get lost. I would encourage the authors to have clear numerical model methodology, the tests, and the case study results, with a synthesis discussion section to tie them together. And then use this to tighten the text as much as possible.

Broadly, I am left with an uncertainty as to how good this model really is. I find both the wind velocity and blowing snow analysis underwhelming. Specifically, I would really like to see figures showing model speed and direction, compared to both a model but also observations. I see figure 9, but I would prefer this as a wind rose and to see >24h of observation analysis. How are the ridges and lees /exactly/ compared to the more complex model or obs? The lee is where the snow deposition will happen and if this is wrong, the deposition is wrong (which seems like the case). Can the model perturb the wind direction very much? It looks like it barely does, but this might just be a visualisation limitation. How large of a domain can this scale up to? How does this approach compare to other blowing snow approaches like those of Snowmodel, ICAR, CHM, SnowPappus, and Alpine3D? For example, Figure 5 d-e shows a ~4x under estimation in transport rate. Figure 6 is completely missing deposition zones. This seems like a lot. I would also like to see side by side plots of the lidar snowdepth observations and the model outputs. Is the spatial variability we observe, that all models struggle to get 100% right, at least qualitatively present in the model outputs? Without these contextualizations, it is difficult to evaluate the manuscript.

I think having a core set of research questions that are answered throughout the manuscript would help tighten and refine the story. As is, this is a statement of what was done, and the clear common thread tying it all together, including the advancement versus existing approaches is lacking. Getting the wrong answer faster is not particularly useful unless the error in that approximation is well bounded, understood, and justified.

We want to thank the reviewer for their valuable feedback. As stated in the general response above, our revision will concentrate on a more thorough presentation of model uncertainties and contextualization in the framework of existing models, as well as on a modified structure that aims to improve readability of the manuscript. This will include a new section on model limitations.

Please find our answers below to your specific points raised in the upper paragraphs :

First, we will change the abbreviations for the model output to more intuitive ones (NSW \rightarrow wind_{surf}; DSW \rightarrow ΔM_{snow} ; SUBL_VI \rightarrow ΔM_{subl} ; SNOW_VI \rightarrow flux_{snow}). For the abbreviations of the individual experiments in the case study, we will use the description of the experiment setup wherever possible, and repeat the explanation of the abbreviation otherwise.

We agree that a real-world validation for longer than 24 h would be desirable. Sadly, the HEF-LES as one part of the benchmark data only exists for this period and this data basis is not easily extendable due to the large computational demands. We therefore concentrate on this rather short

case study, in which we can still learn a lot about the model's behavior in various flow scenarios as we can validate against the highly resolved LES as benchmark.

We want to argue against the change to wind roses in Figure 9, as we would lose the temporal variability information, which is one important aspect in our analysis of the case study. As stated in the introduction of our response and in the specific comment below on the evaluation of wind direction, we agree that a more thorough analysis of the wind direction is necessary, which will be added in the revised manuscript.

We agree, that a direct comparison of the simulated snow redistribution with the observed snow height change could strengthen the validation. We can unfortunately not do this here, as the observed snow height change is a difficult signal to interpret. It is not only due to redistribution but confounded by compaction and avalanching, and additional restrictions in the scanning geometry. Disentangling these effects would be out of scope of this work, and has been the focus of dedicated studies like Voordendag et al. (2024). However, the HEF-LES data were validated against the laser scans in the original publication and we use these validated HEF-LES as benchmark in our study. In the section 4.1 (case study overview) we will emphasize this validation approach more clearly. Apart from that we will add a quantitative validation of snow redistribution with respect to the HEF-LES.

"How large of a domain can this scale up to?" The output of SNOWstorm is fixed to tiles of 256x256 points with spatial resolution of 50 m. For larger domains, multiple tiles have to be predicted, and in this way any domain sizes are possible. Note that the model is not transferable to other spatial resolutions (see also response to Reviewer 1). We will point this out more clearly in the revised manuscript.

We could not follow some remarks of the reviewer, in particular we do not see the missing deposition zones in Fig. 6 (blue shaded areas), which the reviewer is referring to.

Specific comments:

L1 "and (the) resulting" missing the

We will add that.

L22 e.g., add the comma after e.g. throughout

We will add that.

L 91, Figure1 these need scale bars, north/south arrows, and lat/lon reticules. As is, it's not at all clear what we are looking at.

We agree, and we will re-do this figure into a conceptual figure to showcase the topographic analysis procedure.

L95 The synthetic topographies are neat, but why go through the effort when global 30m DEM exist? I suspect it's for the periodic boundary conditions? If so, this needs to be stated up front and not as a buried lede

As described in L170-172, the periodic nature of these is a useful feature for our application, though not the main reason for working with this approach. It is correct that high-resolution DEMs are globally available and that building a training data set based on real terrain would be possible as well, however, then the problem would be how to choose individual regions and how to avoid sampling bias.

The main idea is for our model to learn the fundamental behavior of terrain-atmosphere interaction. For this reason, we build the training data set on synthetic topographies that reflect fundamental terrain characteristics.

L 114 "to fit the requirements of the later model" where in the MS? what section?

This refers to the design choices in model resolution and domain extend of SNOWstorm. We will make this more clear in the revised manuscript.

L120 `a` term is what?

Thank you for pointing this out, "a" is the intercept of the spectral slope, we will add this in the revised manuscript. Additionally, it will be shown in the re-worked Fig.1.

L145 "is used" -> "are used"

Will be corrected.

L155 Why are the particle sizes not a distribution as commonly noted in the literature?

In the current version of the snow drift WRF module the particle size is described for simplicity as the mean particle size at ground level with decreasing size with height, the size at ground level is either prescribed or calculated as a function of friction velocity.

L155 Is this domain masked for vegetation? how is that handled? I realize that the authors are interested in non-vegetated areas, but some description of how this is handled and impact on flow fields is needed

Yes, the entire model domain is set up without vegetation. As we run the numerical simulation in ideal-case setup with no land-surface model, this is reflected in the prescribed roughness length. It is correct that this is not a problem in our application, as we are interested in non-vegetated areas, but it is a limitation for other applications. We will state this more clearly here and in the additional section on model limitations.

L 156 table 1, how are these sampled while maintaining physical coherence? Please cite ranges for these values

As the values of Tab.1 are chosen to reflect the variability seen in nature, we did not identify any parameter combination as incoherent. Nevertheless, some parameter combinations (especially with very strong static stability) might occur less frequent. We still want to include these settings in the training data, as we want to expose the model to a training data set that is as variable as possible.

L174 "model run time" are these wall clocks or internal model hours?

We refer here to internal model hours and will make this clearer in the revised manuscript.

L229 how is T changed but RH is not updated? this will dry/wet the atmosphere out and impact the sublimation estimates

We are convinced that this should not be a problem. In this step, we extrapolate the temperature from the coarse-scale input to the height difference between the coarse-scale model topography and the fine-scale DEM, under the assumption of a moist-adiabatically stratified atmosphere of constant relative humidity. You are correct, that this will change the absolute humidity, however, the sublimation calculation in SNOWstorm is only dependent on the relative humidity.

L236 I would like to see clear wind direction eval here to

As stated in the general introduction, we will add a point-wise evaluation of the wind direction error compared to the AWS and HEF-LES. Additionally, we will add a more process-based evaluation of how flow features are spatially captured. This will include the flow deflection and flow splitting on the windward side of Weißkugel summit, the flow deflection and acceleration at the ridges north and south of the glacier, and the channeling inside the valley.

As you pointed out in your general comments, it is crucial for the model to represent the wind direction at ridges to correctly to identify wind-ward and lee-ward slopes correctly. This is happening largely throughout the domain, except for the area around station IHE on the ridge south of the glacier. Here, the ridge is almost aligned with the large-scale flow direction and therefore both HEF-LES and SNOWstorm fail to predict the observed southerly wind direction. We will discuss this more critically in the revised manuscript.

L308 has results in the overview

This is a result of the original publication of Voordendag et al. (2024), for which we are giving the overview here. In the revised manuscript we will state more clearly, that this paragraph only gives a brief overview on the analysis of the original publication.

L320 how sensitive are the results to this light-ish density? Windslab layers are of course also missing here

We ran a sensitivity experiment with higher density of 400 kg m⁻³. In this experiment the average erosion is reduced by roughly a factor 3, the maximum erosion by a factor 1.4. With the fresh snowfall immediately before the drifting snow phase, we consider the 200 kg m⁻³ realistic. You are correct, that snow surface hardening and densification is not included in our approach, and that this is an important aspect to include. In our case, we wanted to be consistent with the HEF-LES, where it is not included either. However, for future application with SNOWstorm coupled to a snow model, this aspect will be included.

L330-> would love to have S_WDI_W and other variables described or reiterated for the reader. By the time I got here, I had no idea what these were anymore. There is some description on L 325 but just they are inputs, of what?

Thank you for pointing out these difficulties, we will re-introduce variable names for better readability.

The structure of the abbreviation in your example is:

S → SNOWstorm experiment

WD1 → atmospheric input for the SNOWstorm prediction taken from WRF domain D1

W → SNOWstorm run on smoothed topography of HEF-LES