

Review of "SNOWstorm (v1.0) - a deep-learning based model for near-surface winds and drifting snow in mountain environments", by Saigger et al.

Summary

This manuscript presents a new machine-learning-based approach to simulate near-surface winds, snow drift sublimation and transport over complex topography at a high spatial resolution of 50 m. The model is trained on controlled large-eddy simulations (LES) conducted over synthetic topographies, which supports its broad applicability. An independent validation on a glacier in the Austrian Alps demonstrates comparable wind fields and snow redistribution patterns to those obtained from numerical LES results.

This novel downscaling model for wind and snow redistribution opens the door to long-term, high-resolution glacier mass-balance applications. However, I have several major and minor concerns in its current form. Most importantly, the proposed approach is not yet sufficiently situated within the context of existing models. The manuscript would benefit from clearer benchmarking against established wind and snow redistribution models, a more explicit discussion of how specific modeling choices relate to existing approaches (and the motivation for differing choices), and a clearer definition of the applicability range and limitations of the SNOWstorm framework and its individual components. In particular, additional detail on the training dataset (e.g., the representation of atmospheric conditions and terrain types) would help assess whether model mismatches may stem from limited coverage of certain flow regimes or synthetic topographic characteristics. In this context, the rationale for adopting power-law spectral scaling based on parameters derived from real terrain should be more explicitly justified, especially given the potential uncertainties when extrapolating the model beyond the training range in terms of topography or spatial resolution. Clearer methodological explanations (e.g., regarding the sequential execution and interaction of model components) and a more detailed quantitative evaluation would aid the manuscript. This could include benchmarking modeled snow redistribution rate errors against errors from previous studies and incorporating available observational datasets, such as the mentioned terrestrial laser scans. Finally, the manuscript would benefit from a more transparent discussion of model capabilities and limitations, ensuring that conclusions remain closely aligned with the evidence presented. If these aspects are addressed, the study would make a valuable contribution to the field.

Major comments

Please try to be more specific to increase understanding and avoid overgeneralizing results. Below, I refer to my concerns at the corresponding text positions.

Abstract:

Line 7: "[...] to be applicable over a wide range of atmospheric conditions and for a wide

range of mountain regions.”: Be more clear: Aren’t the atmospheric conditions only for winter time conditions of glaciated mountain regions in mid- to high latitudes and the mountain regions for rather moderate slopes of maximum 40°?

Line 11: ”In a first real-world application study in the European Alps, SNOWstorm predicts wind fields [...]”. : Be more specific: ”on a glacier located in the European Alps”.

Introduction:

Line 71-72: ”The model of Le Toumelin et al. (2023) used the data set of idealized numerical simulations of Helbig et al. (2017) as training data.”: Be more specific: [...] idealized numerical simulations across diverse synthetic topographies [...]

Line 72-73: ”Despite the successful implementation, this model has the shortcomings of assuming a neutral stratification of the atmosphere, neglecting turbulent motions and assuming a linear dependence in the wind velocity.” : Please be more specific. While Le Toumelin et al’s wind downscaling model assumes linear scaling with respect to coarse wind velocity, consistent with linear flow theory, the model itself is also a nonlinear convolutional neural network and does not assume linear terrain-flow interactions.

Line 76-84: ”[...] we specifically aim for these characteristics:” Are these goals fully achieved with SNOWstorm or not? Be more specific, several goals seem really vague, e.g. ”-large speed up rate compared to conventional numerical simulations of several orders of magnitude,” or ”-representation of turbulent motions in the atmosphere,”.

Data and methods:

Line 105-110: While it is briefly mentioned that the approach for the SNOWstorm development strategy was based on work from Helbig and Löwe, 2012, Helbig et al, 2017 and Le Toumelin et al, 2023, this should be clearly indicated in other sections such as the introduction in the manuscript. Le Toumelin et al, 2023 developed a Convolutional Neural Network model for near-surface winds based on Helbig et al, 2017’s database consisting of thousands of atmospheric model simulations on Gaussian Random Fields as topographies. Please be also more concrete in ”Atmospheric simulations run on such topographies can inform downscaling tools for, e.g., wind (Helbig et al., 2017; Le Toumelin et al., 2023).”

Line 105-110: The motivation and reasoning for choosing topographies with power law spectral scaling are not made clear. Particularly, given that your approach is built on that from Helbig et al, 2017, who used Gaussian over power law topography models based on a better statistical agreement with real terrain slope characteristics, and given that power laws do not hold across spatial scales.

Eq. 1: If I am not mistaken you used the 9 regions shown in Fig. 1 and cut them in 256x256 tiles with 50 m resolution. Based on the spectral parameters a, b from these topographies

you derived the 72 power-law spectrally scaled topographies. Please indicate how many 256x256 tiles were used for the a, b parameter pool. Please also clearly state that the 72 resulting synthetic topographies only represent the observed topographic scaling range. Thus, training a statistical model on processes computed on these topographies does not guarantee similar statistical model accuracy on topographies that fall outside this observed scaling range. Also, isn't the wavenumber k bound by domain and grid cell size? I believe this also introduces quite some uncertainty in model applications at different domain and grid cell sizes then used for the training data, i.e. the model may not be that widely transferable. Please clarify and discuss. Please also better describe a, b , set them in context and provide the formula for k .

Fig. 1: Please indicate the reasoning for choosing exactly these nine regions as well as their terrain characteristics.

Fig. 2 and Line 130-132: How can the Fourier Land b parameter range be broader than the b range extracted from the real topographies from which they were randomly drawn?

Line 145: "With the ideal-case setup of our simulations, only [...]": What do you mean by "ideal-case setup" here? And how does this compare to the mentioned "semi-idealized" simulations in abstract and conclusions?

Line 16-163: Please detail your model setup with 10 simulations per topography:

1. One simulation describes one atmospheric condition with randomly drawn pressure, temperature and relative humidity. How you can draw random values of pressure, temperature and humidity with the ranges given in Table 1 and ensuring physical meaningfulness from a random combination at the same time?
2. When you are shifting the wind direction by 1 degree, does this mean you have 360 additional simulations for each topography or does this mean you have each wind direction twice in the 720 simulations?
3. Were the same 10 atmospheric conditions run for all 72 topographies, or did every topography had their own conditions? I think it is essential to detail what atmospheric conditions are in the training data sets. Can you provide a table for the conditions if the same 10 conditions were used for each topography and if not, can you provide a histogram describing the atmospheric conditions of the data set, e.g. a barplot showing how often neutral, stable were drawn, which wind directions and also which snow density etc.

Line 174-176: What is the reasoning and goal to average and accumulate modeled fields over 2 hours for training?

Line 203-205: Why is this application of a square-root transformation on terrain height done? Doesn't this mean resulting slopes are smoothed and the resulting topographies are not comparable anymore to the ones generated with Eq. 1? Please clarify if any implications for the topography characteristics result.

Line 212-216: Following the data augmentation with one shift per topography-wind field, please indicate the total of wind fields that were used for training. If I understand correctly, the data augmentation was only applied for the snow U-Nets not for the NSW U-Net? This means the models have different amount of training data? Please clarify.

Line 212-216: Please consider showing a few spatial examples for shifted wind fields on topography, perhaps in the Supplement?

Line 229-223: How are the low-resolution fields interpolated to the fine-scale grid?

Line 230-231: Are predictions of DSM, SUBL_VI and SNOW_VI sequential? How are interactions between sublimation and snow transport considered, as it is the case in the real-world?

Line 252-253: To strengthen this hypothesis an overview of actual atmospheric conditions used in the training data set would be beneficial. See my comment above.

Line 255: Compared to studies with a similar spatial resolution to that achieved here (30 m, 50 m), the phrase "fall in line" may be somewhat optimistic, as the MAE of 0.8 m/s is notably higher than the value reported by Le Toumelin et al. (2023), and the mean bias is higher than the ones reported by Dujardin and Lehning as well as Le Toumelin. In addition, given the substantially coarser target resolution of 1 km in Dupuy et al., a direct comparison may not be entirely appropriate.

Line 263-271: Can you set these errors in context or, if possible, compare to errors from previous studies modeling or measuring these rates, as the magnitudes of these errors alone are not intuitive.

Line 308-309: Is this validation referring to what we see in Fig. 9, i.e. green versus black lines?

Section 3.2: Could you provide terrain characteristics for these three cases, similar to the background conditions described in the figures? Could it be informative to investigate three different atmospheric background conditions over the same topography? I am unsure about the origin of the mismatches: do they arise from an insufficient representation of flat terrain, or from a lack of neutral and low-wind conditions in the training dataset? This is why I suggested above providing more detailed information about the training dataset as

to understand the model behavior.

Line 325-327: "are in agreement" and "slightly smoother" seem a bit overly optimistic for the spatial patterns shown in Fig. 8b,c versus Fig. 8d. Perhaps change to "agree overall well" or similar.

Line 329-330: 1-4 m/s is a rather large absolute error range. Please be more specific in saying which experiment led to better performances and please also set the errors in context to errors from previous studies. From what I see in Table 3, the correct error range is 2-4 m/s not 1-4 m/s.

Line 330: Can you give an explanation for this overestimation?

Table 3: Please provide the errors for ERA5 and WRF low-resolution input as well as for HEF-LES at the stations. Set in context, this would allow to better understand the errors. Please also provide the errors for wind direction, as both wind speed and wind direction, are used as input for redistribution and drifting snow sublimation in SNOWstorm.

Line 337-340: Can you provide error measures to demonstrate this improvement? Please consider showing the error statistics similarly to Table 3 for all "*_G" experiments presented in Table 2 (e.g., in supplementary).

Line 347-448: While the overall spatial patterns seem to agree well, the agreement in magnitude appears less strong. It would be helpful to be a bit more specific.

Line 370-373: Same comment as above in the abstract: "wide range of regions world wide", please be more clear.

Line 373: "European Alps", please be more specific.

Line 375: Please rephrase as this does not fully agree with the results shown. Particularly, strengths and amounts compared less well. What do you mean with "ground truth"? Why do you not show any evaluations with laser scan observations given that they were available for you case study (Line 306)?

Line 375: What do you mean with "sharp gradients"? Where do you show that SNOWstorm predicts turbulent flow well and can particularly well predict atmospheric conditions that haven't been included in existing models?

Line 380: Please consider to be more specific here. See my comment to errors in the results section.

Line 380: "SNOWstorm generally succeeds to capture the overall flow structure and re-distribution patterns by the LES both with the smoothed topography from the LES and an un-smoothed high-resolution DEM.": Again, this is general statement. Please be more specific and open. Also, I don't see error statistics demonstrating this for the unsmoothed topography, only the spatial fields are shown in S4 and S5.

Line 385: "Validated against automatic weather stations, errors in the wind speed are slightly higher than in the crossvalidation experiments with MAEs between about 1.5 and 4 m/s": Please rephrase. I don't agree that 1.6 - 4 m/s absolute errors are only "slightly higher" than 0.8 m/s. Please also set these errors in context to errors obtained from previous ML-based models.

Line 390: "Overall amounts of snow mass change as well as the placement of zones of erosion and deposition agree between SNOWstorm and the LES." : Please be more specific. See my other comments to this regards.

Line 405-408: "semi-idealized". Why do you call it semi-idealized? Compared to previous models that used model experiments, I would say the approach here is idealized as well, although you explicitly cover broader atmospheric stability and turbulent conditions? Please clarify and, if applicable, adapt in the manuscript.

Minor comments

Line 58-61: Please verify referencing and be specific about developments and applications of wind downscaling models. As far as I know the studies, Schirmer et al used the Winstral approach, Vionnet et al and Marsh et al used wind libraries from WindNinja. WindNinja is the diagnostic model described in Wagenbrenner et al. Dacic et al 2010 used atmospheric model wind data, Helbig et al 2017 presented a statistical wind downscaling, Helbig et al, 2024 a statistical wind and snowfall downscaling approach.

Line 70-71: "For this, Dujardin and Lehning (2022) trained their model on data from weather stations and high-resolution digital elevation models in Switzerland.": Dujardin and Lehning trained on COSMO, weather station and terrain data.

Line 196: What is meant with "above-crest height"? Please indicate which level or height above ground should be used for the low-resolution wind input.

Technical comments

Line 273: Do you mean "than" instead of "that"?

Line 355-358: I think it should be S4 and S5 here.