SNOWPACK MODEL SENSITIVITY TO OBSERVATIONAL AND WEATHER MODEL INPUTS

Ryan Szczerbinski[1], Travis J. Morrison[1]*, McKinley Talty[1], and Jim Steenburgh[2]

[1] *Utah Avalanche Center, Salt Lake City, UT, USA*
[2] *University of Utah, Salt Lake City, UT, USA*

ABSTRACT: The SNOWPACK model has shown skill at predicting the formation and evolution of various layers within the seasonal snowpack, including instances of avalanche-conducive properties which have supported human forecasters. While ground truth atmospheric and snowpack observations are ideal for SNOWPACK inputs, weather stations typically experience some intermittent failure, leading to gaps in the observational records. Additionally, they are expensive to install and maintain, and only provide point information, which has limited relevance to broad mountainous regions. In response, some agencies are turning to Numerical Weather Prediction (NWP) model data to entirely drive SNOWPACK over broader spatial scales. While this method improves spatial coverage, it comes at the cost of accuracy. Little Cottonwood Canyon's Atwater (ATH20) snow study site provides a continuous observational record, which enables an opportunity to study the sensitivity of SNOWPACK results to observed versus NWP inputs. ATH20 site observations and raw and downscaled values from NOAA's 3-km resolution High-Resolution Rapid Refresh (HRRR) model were each used as inputs to SNOWPACK over the 2025 snow year, as well as various observation-HRRR combinations. Results were validated against manual snow profile observations by quantifying profile similarities with dynamic time warping (DTW). HRRR downscaling improved the rate of SNOWPACK convergence and had a better DTW result than the raw HRRR data compared to ATH20-driven SNOWPACK. All SNOWPACK runs including the fully ATH20-driven performed similarly against the manual snow profiles, indicating a need for manual profile data collection tailored to numerical model validation. We conclude by discussing the organization's future model validation and operational forecasting intentions.

KEYWORDS: Snowpack and Weather Modeling, Observation-Model Comparisons, Snowpack Properties, Weather Model Downscaling

1. INTRODUCTION

The SNOWPACK numerical snow cover model is used in snow science and avalanche forecasting operations to simulate the formation and evolution of seasonal snowpack structure, including weak layers associated with avalanche activity (Bartelt and Lehning, 2002). Geophysical model inputs including air temperature, humidity, wind, radiation budget, snow height, and precipitation are ideally supplied by an automated weather station at or near the site of interest. Ground stations provide the most direct measurement of local conditions, but they are expensive to install and maintain, and are prone to intermittent failure from power loss, icing, and sensor drift, producing gaps in the observational record ranging from hours to months. Complex terrain variations and sparse spatial coverage relative to the areas that forecasters must assess further compound point observation challenges. Numerical weather prediction (NWP) models provide continuous, spatially complete coverage that does not depend on any single station's uptime.

This spatial improvement comes at a cost to accuracy; even a 3-km convection-allowing model can misrepresent local terrain effects, and model output is fundamentally a forecast rather than a measurement. However, some high-resolution NWP model estimates can outperform networks of precipitation gauges for total annual rain and snowfall in mountain terrain, showing potential to improve hydrometeorological datasets (Lundquist el at., 2019).

Some avalanche forecasting operations have begun using NWP output to drive SNOWPACK over broad spatial scales in place of point observations, but model prediction value is limited by a lack of meaningful verification data (Horton et al., 2023). Further, observation assimilation was noted as a method for improvement. Herla et al. (2024) used raw NWP data to drive SNOWPACK over an extensive 1,004 grid points spanning three snow climates and 10 snow years, and found skillful representation of weak layers, but also demonstrated substantial differences between model results and human assessments, which are both imperfect. Palomaki and Miller

* Corresponding author address:
Travis J. Morrison, Utah Avalanche Center, Salt Lake City, UT 84109;
email: travis@utahavalanchecenter.org

https://doi.org/10.15788/1790099265

(2023) ran SNOWPACK with weather station observations and bias-corrected NWP data and compared the results to manual snow profiles for a single site over a single season, and found closer alignment between observationally derived profiles and manual profiles than NWP derived profiles.

For this study we ran SNOWPACK using observations from the Atwater snow study site in Little Cottonwood Canyon, UT, which maintains a continuous, high-quality observational record. This gave us the opportunity to substitute portions of a known-good station record with HRRR-derived values under controlled conditions and measure the resulting effect on SNOWPACK output: artificial gaps were created in the observational record to simulate missing data, which was then filled with raw and downscaled NWP variables. Simulated snowpacks were compared against manual profile observations using a tailored snow profile alignment tool following Herla et al. (2021), establishing a framework for evaluating how the amount, pattern, and type of gap-filled weather data affect SNOWPACK's representation of observed snowpack properties.

2. METHODS

2.1 Study site

The Atwater (ATH20) snow study site (40.5913, -111.6375) sits at 2,665 m in the town of Alta, UT. The site is a flat and sheltered ~25 x 40 m field surrounded by steep hillsides and subalpine forest (Figure 1). It has an intermountain snow climate, with an average annual snowfall of 13.8 m measured at the adjacent Alta Ski Area. ATH20 (Figure 2) has a long history of snow and avalanche study, dating back to its establishment in 1939 (Skiles et al., 2022).

ATH20 is equipped to measure a broad array of meteorological parameters, including 2 m air temperature, wind speed and direction, radiation budget, relative humidity, and snow depth at hourly or higher time resolution.

Data from 11 manual snow profile observations during the 2025 snow year were used to validate model results and were collected by the Utah Avalanche Center and the University of Utah's Environmental Fluid Dynamics and Snow Hydro Labs.

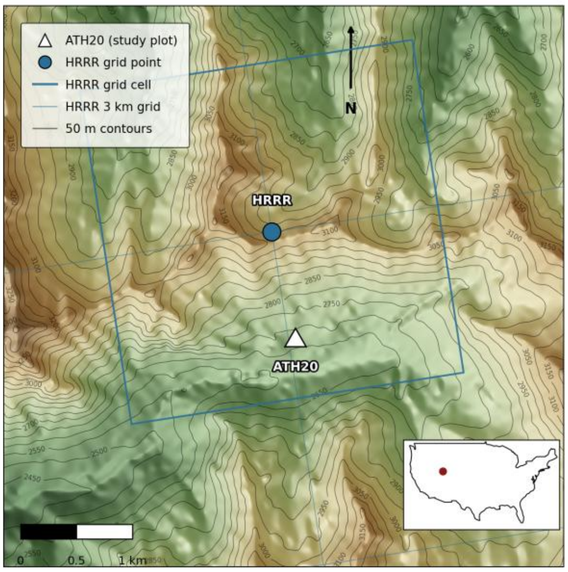

Figure 1: ATH20 field site in Little Cottonwood Canyon, UT.

Figure 2: ATH20 field site where snow profiles and weather station data were collected.

2.2 SNOWPACK inputs

To simulate weather station downtime, two methods were used to create artificial gaps in the ATH20 observational record for the 2025 test snow year. The "block" method created three to four continuous gaps for the entire snow year, simulating a station with intermittent yet persistent missing data. The "scatter" method created discontinuous gaps, simulating shorter duration but more frequent missing data. Real weather station downtime is a combination of these two methods, but for simplicity we tested them independently. The gaps were then filled with raw and downscaled HRRR data in various ratios, thus creating four gap-fill methods to test: block and scatter, each independently filled with raw and downscaled HRRR data. Additionally, each gap-fill method was tested with various gap percentages between 5-100%, creating a continuous test platform between pure observational data with 0% fill, all the way up to pure model data at 100% fill.

2.3 Numerical weather prediction data

We used NOAA's High Resolution Rapid Refresh (HRRR) v4 atmospheric model, an hourly-updated mesoscale NWP model with 3-km horizontal grid spacing over the conterminous United States. The HRRR produces 18 hour forecasts every hour, and 48 hour forecasts every 6 hours (Dowell et al., 2022). To significantly reduce the computational load of retrieving two snow years' worth of NWP data, rather than downloading single-hour HRRR analysis data files, also called the "nowcast", we retained the data from the full forecast horizon from a given model run, and then repeated this process at the run's expiration for the full season. As a result, the season-long NWP files used for the training and test data sets are composed of a series of typically 18 but occasionally 48-hour HRRR forecasts. We acknowledge that this may not necessarily represent how an operationally NWP gap-filled observational data set would be constructed, but determined it was adequate given the time and computational resources available for this study. The nearest HRRR grid point (40.60, -111.64) at 3,027 m is located ~1 km NNW of ATH20 in an adjacent watershed and was used for this study (Figure 1). Outputs from the HRRR at this point were used in raw and downscaled versions for gap filling as described in section 2.2.

2.4 Downscaling

Data from ATH20 from the 2026 snow year between October 1, 2025 and April 15, 2026 were used to develop downscaling models for each meteorological variable that was tested. These parameters were then applied to the 2025 snow year between October 1, 2024 and June 1, 2025, and this test data set was used as input to SNOWPACK with the methods described in Section 2.2. The downscaled variables were 2 m air temperature, surface temperature, incoming and outgoing shortwave radiation, incoming longwave radiation, relative humidity, and 10 m wind speed.

A two-parameter Weibull distribution was used to map HRRR 10 m wind speed to ATH 20 wind speed (Wais, 2017). For each of the other variables, several correction methods were fit against the ATH20 data: additive bias correction, multiplicative ratio correction, ordinary least-squares linear regression, and empirical quantile mapping. The method minimizing RMSE against the ATH20 observations was retained for that variable: bias correction for 2 m air temperature and incoming longwave radiation, linear regression for surface temperature, ratio correction for both shortwave radiation terms, and quantile mapping for relative humidity.

Similar to a related study by Palomaki and Miller (2023), we found a poor relationship between observed and modeled snow heights. Instead of downscaling, a simple linear interpolation between the last known observed snow heights was used to fill the artificial data gaps. While discarding NWP snow height provides no insight into its performance in the SNOWPACK model and thus a fully NWP-driven virtual "site" without physically measured observations, operational avalanche forecaster trust in these methods is paramount, and SNOWPACK simulations driven by measured snow height outperform simulations driven by measured precipitation since snow height driving is self-correcting against precipitation gauge errors (Wever et al., 2015; Morin et al., 2020).

In the future we will explore switching to precipitation driven simulations at certain thresholds due to its robustness for NWP-chain and remote applications lacking a reliable height record (Bellaire et al., 2011; Keenan et al., 2021). Downscaling approaches for precipitation and snow to liquid ratio will also be assessed (Veals et al., 2025).

2.5 SNOWPACK model simulations

We used the SNOWPACK (v3.7) one-dimensional snowpack model (Bartelt and Lehning, 2002). The model was initialized at the start of the snow year, October 1, 2024, with predicted persistent snow cover beginning mid-November. The model timestep was 15 minutes, and profile outputs were written every 3 hours. The Zwart new snow density scheme and measured snow height enforcement were used in the model's initialization.

Each gap-fill method percentage was run through SNOWPACK until the full season was complete, changing the location of the gaps for each new run if the previous run failed for up to five attempts. Once the full season ran successfully, the inverse of the number of attempts became its success rate, and these values were plotted in section 3.2. At large gap percentages, the success rate became unstable due to the large amount of forecast data.

2.6 SNOWPACK postprocessing

We used an internally developed snowpack-validator, a Python programming language-translation of *sarp_snowprofile.R* and *sarp.snowprofile-*

alignment.R, which were built to account for variations in snow layer depth, thickness, hardness, grain type, and other properties between numerical and physical snow profiles with a dynamic time warping (DTW) method (Herla et al., 2021). DTW rescales a query profile to a reference profile, and assigns a cost between 0-1, with 0 being low cost and high resemblance, and 1 being high cost and poor resemblance. DTW was used to make two types of snow profile comparisons: HRRR-filled simulations to observation-based simulations, and HRRR-filled and observation based snowpack simulations to manual snow profiles. Similarity subscores are calculated by assigning distance functions for the weak layer, crust, precipitation particle, and bulk grain subtypes on a normalized scale. The simple similarity score is the mean of the four subscores, with 0 indicating poor similarity and 1 indicating perfect similarity.

3. RESULTS

3.1 Downscaling

The HRRR 2 m air temperature distribution had a mean bias error of -3.11 K, which can be explained by comparing the grid point locations of the HRRR and ATH20 station (an elevation difference of ~360 m) corresponding to a physically realistic adiabatic lapse rate of 8.64 K/km.

The HRRR surface temperature distribution had a mean bias error of -1.41 K, composed of two percentile regimes: a cold bias below 274 K and a warm bias above 274 K. This was best corrected by a linear regression with a slope of 1.10 and an intercept of -27.32 K.

The incoming shortwave radiation distribution had a mean bias error of +39.0 W/m^2 and was corrected with a ratio of 0.75. The outgoing shortwave radiation distribution had a mean bias error of –19.9 W/m^2 and was corrected with a ratio of 1.46. Incoming longwave radiation had a mean bias error of –42.6 W/m^2 and was corrected by the same amount.

Relative humidity had a mean bias error of 0.00, and RMSE of 0.19. None of the correction methods altered the RMSE, so empirical quantile mapping was used to correct a slight model quantile underestimate of ~0.05 between observed values of 0.7 and 0.9.

The wind speed Weibull distribution at ATH20 had shape parameter k = 1.09 and scale parameter λ = 1.13, and the HRRR had k = 2.21 and λ = 6.59. After mapping the HRRR wind speed distribution parameters to the ATH20 values, the mean bias error dropped from 4.7 m/s to 0 m/s, and the RMSE dropped from 5.3 m/s to <1 m/s (Figure 3). Additionally, a linear regression fit through the origin, a simpler but less physics-based method, was tested. With a slope = 0.18 it produced very similar bulk statistical results to the Weibull distribution mapping, with mean bias error of 0 m/s and RMSE <1 m/s.

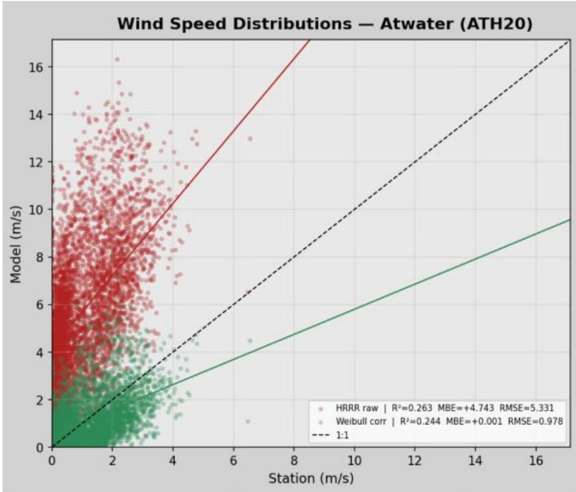

Figure 3: The improved correlation between ATH20 and Weibull-corrected HRRR values (green data) compared to ATH20 and raw HRRR values (red data).

3.2 SNOWPACK model

Raw NWP data inputs caused model run failure at lower fill percentages than downscaled NWP data inputs across both the block and scatter methods. Model run failure typically occurred mid-season with a developed snowpack, when the surface temperature fell out of range below 210 K. We suspect this numerical divergence was encouraged by the NWP season download method described in Section 2.3, but further investigation is required. Figure 4 shows the higher success rates of downscaled compared to raw HRRR fills as a function of season fill percentage for the block and scatter methods.

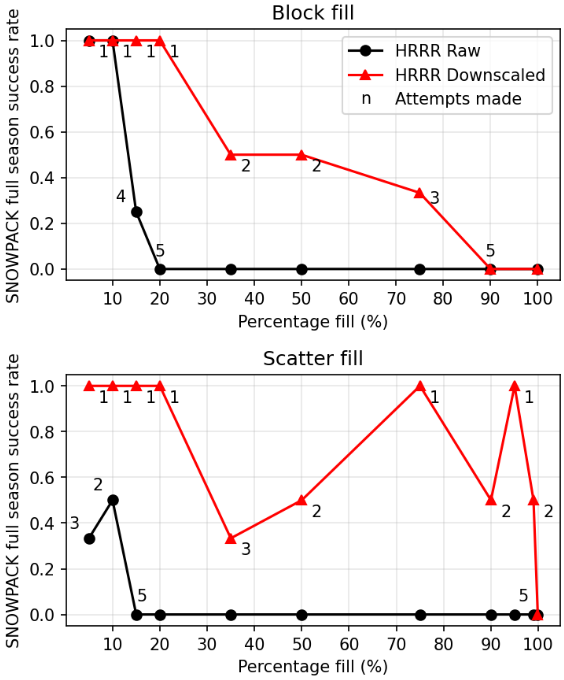

Figure 4: Higher success rates for downscaled (red triangles) compared to raw (black dots) HRRR as a function of fill percentage.

We suspect the return to 1.0 at 75% and 95% scatter fill was due to the low number of simulations tested and that those rates would decrease with further attempts at those fill percentages, following a similar trend in the block fill. However, at higher percent fills the meteorological inputs become more self-similar (i.e. mostly forecast data).

Figure 5 shows two example profile time series, driven with ATH20 data (top) and 50% downscaled scatter fill (bottom). There are clearly visible similarities including the early and mid-winter distribution of faceted crystals (light blue), rounded grains (light pink), melt-freeze crust (black and red) and depth hoar (dark blue) layers, and the onset of springtime melt forms (solid red).

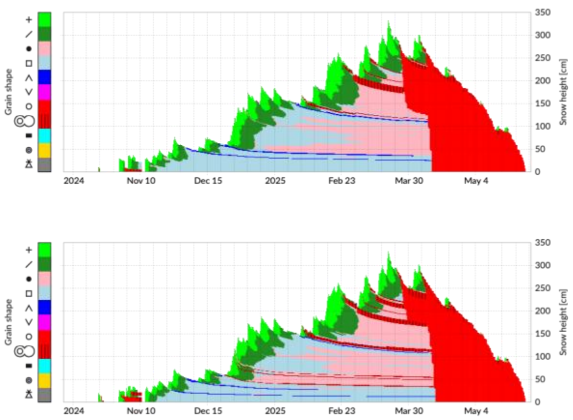

Figure 5: ATH20 observational weather station (top) and 50% downscale scatter fill (bottom) simulated snow profiles.

For the scatter method, the median simple similarity scores between the various raw and downscaled model runs against the ATH20 run ranged from 0.79 (75% downscaled) to 0.98 (10% downscaled) (Figure 6). The 5% and 10% downscaled fills had the highest medians of >0.95 and the smallest interquartile range. Not only did the raw HRRR input data fail to run past 10% fill, but the median score for every downscaled fill percentage except 75% was higher than the 10% raw fill. This presents a strong case for downscaling the HRRR at a station when filling missing data to maintain similarity with existing observation-driven model outputs and to have a reliable modeling chain which prioritizes weather station observations.

Figure 7 shows the simple similarity scores for the same model runs including ATH20 against 11 manual snow profiles. Model run median scores ranged from 0.38 (75% downscaled) to 0.56 (5% raw). Amongst the models tested there is no clear decrease in score with higher fill percentage against the manual profiles. Further, the only medians >0.5 are from the 5% raw and 20% downscaled runs. This in combination with the results from Figure 6 showing a decreasing correlation to the observationally-driven SNOWPACK model with higher fill percentages suggests that there is uncertainty in both the model and the manual profile data collection techniques. Equivalently, replacing observational model inputs with more NWP data does not notedly impact comparisons to the manual profiles.

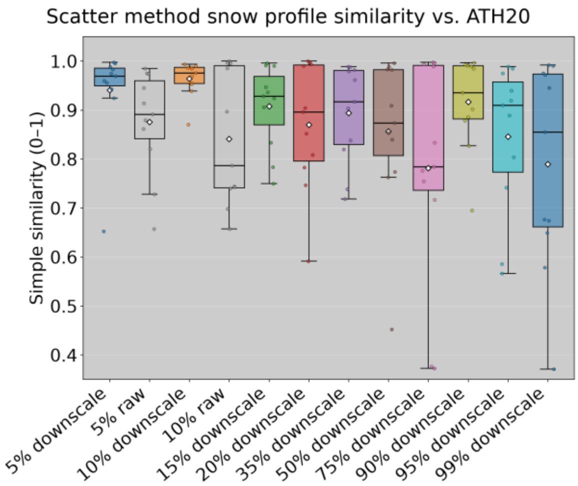

Figure 6: Simple similarity of the scatter runs against ATH20, showing the higher scores and smaller range for downscaled data.

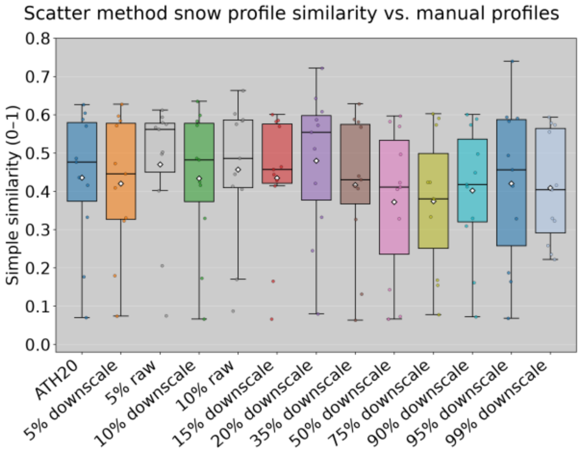

Figure 7: Simple similarity scores of the scatter runs against manual profiles, showing consistency across the runs.

The block method median similarity scores for the various fill percentages resembled the scatter method, where 5% and 10% downscaled fills had the combined highest scores and smallest interquartile ranges against ATH20. Median similarity scores against ATH20 ranged from 0.77 (35% downscaled) to 0.97 (5% downscaled). Similarity scores among all runs against manual profiles ranged from 0.42 (15% raw) to 0.52 (35% downscaled). These charts are available upon request.

4. DISCUSSION

Downscaled HRRR data used for input to the SNOWPACK numerical snow cover model led to improved numerical convergence in SNOWPACK, and resulted in higher similarity to fully observationally driven profiles compared to raw HRRR data. Minimal similarity differences were shown between the various successful model runs and observed snow profiles. The median simple similarity scores among all completed runs ranged from 0.77 to 0.98 against ATH20, and from 0.38 to 0.56 against manual profiles. Palomaki and Miller (2023) found comparable median simple similarity scores of ~0.4 between their nearest HRRR point-averaged and weather station SNOWPACK runs and manual profiles. These differences in scores between model to model comparisons and model to observation comparisons are likely due to imperfections in both human and numerical model-derived profiles (Herla et al., 2024). Horton et al. (2023) similarly determined that a lack of meaningful verification data limits model prediction value. However, this study represents the best data currently available and further emphasizes the importance of high-quality snow profile data and collection standardization. Future validation in alpine and polar regions throughout the world will serve to refine and increase confidence in the model.

Other validation approaches and technologies may provide useful insights, such as rapidly deployable snow penetrometers which offer repeatable measurements of snow hardness profiles. Additionally, during the 2026 snow year, the Utah Avalanche Center (UAC) deployed a vertical thermocouple array at ATH20 to measure snowpack temperature in-situ to provide another validation method (Talty et al., 2026).

5. CONCLUSION

For the UAC's avalanche forecasting operations, missing or poor weather station data will be filled with HRRR analysis "nowcast" data, downscaled to the respective stations. Results here provide confidence that the approach will improve weather station-centric SNOWPACK modeling when coupled with downscaled NWP data. The SNOWPACK models ran at ATH20 maintained high fidelity against the observational record when filled with downscaled data up to 10%, and maintained fair convergence with up to 50% missing data. These methods will need to be further tested at other stations in different climates and for more seasons.

Downscaled NWP data also benefits forward looking forecasts, which can be readily used to pad SNOWPACK inputs into the future. Leveraging downscaled forecasts for this will also provide a more accurate forward looking numerical snow cover profile.

While the initial cost of installing an automated weather station is on the order of tens of thousands of dollars (Palomaki and Miller, 2023), many regional avalanche forecast centers and highway departments already operate their own stations, or leverage existing mesonets. For these use cases, large language model-assisted computational tool development to run and validate SNOWPACK with downscaled NWP-supplemented station data is a viable option.

ACKNOWLEDGEMENTS

Most of the in-situ snow profiles used in this study were performed and generously shared by the University of Utah's Environmental Fluid Dynamics and Snow Hydro Labs.

We thank the UAC forecasters, Alta Ski Area, Deer Valley Ski Area, Park City Powder Cats, Utah Department of Transportation, Avalanche Risk Consultants, and the University of Utah Snow HydRO Lab for their support of this project through weather station data and continual feedback throughout the development process. This project was made possible through funding from the Utah Outdoor Adventure Commission and other interested donors.

REFERENCES

Bartelt, P. and M. Lehning, 2002. A physical SNOWPACK model for the Swiss avalanche warning: Part I, numerical model, Cold Reg. Sci. Technol., 35, 123-145.

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

Dowell, D. C., Alexander, C. R., James, E. P., Weygandt, S. S., Benjamin, S. G., Manikin, G. S., Blake, B. T., Brown, J. M., Olson, J. B., Hu, M., Smirnova, T. G., Ladwig, T., Kenyon, J. S., Ahmadov, R., Turner, D. D., Duda, J. D., and Alcott, T. I.: The High-Resolution Rapid Refresh (HRRR): An Hourly Updating Convection-Allowing Forecast Model. Part I: Motivation and System Description, Weather and Forecasting, 37, 1371–1395, https://doi.org/10.1175/WAF-D-21-0151.1, 2022.

Herla, F., Haegeli, P., Horton, S., and Mair, P., 2024. A large-scale validation of snowpack simulations in support of avalanche forecasting focusing on critical layers, Nat. Hazards Earth Syst. Sci., 24, 2727-2756, https://doi.org/10.5194/nhess-24-2727-2024.

Herla, F., Horton, S., Mair, P., and Haegeli, P.: Snow profile alignment and similarity assessment for aggregating, clustering, and evaluating snowpack model output for avalanche forecasting, Geoscientific Model Development, 14, 239–258, https://doi.org/10.5194/gmd-14-239-2021, 2021.

Horton, S., P. Haegeli, K. Klassen, J. Floyer, and G. Helgeson, 2023. Adopting snowpack models into an operational forecasting program: successes, challenges, and future outlook, in: Proceedings of the International Snow Science Workshop, Bend, OR, 8-13 October 2023, 1544-1549.

Keenan, E., N. Wever, M. Dattler, J.T.M. Lenaerts, B. Medley, P. Kuipers Munneke, and C. Reijmer, 2021: Physics-based SNOWPACK model improves representation of near-surface Antarctic snow and firn density. The Cryosphere, 15, 1065–1085, doi:10.5194/tc-15-1065-2021.

Lundquist, J., M. Hughes, E. Gutmann, and S. Kapnick, 2019. Our skill in modeling mountain rain and snow is bypassing the skill of our observational networks, Bull. Am. Meteorol. Soc., 100, 2473-2490, https://doi.org/10.1175/BAMS-D-19-0001.1.

Morin, S., S. Horton, F. Techel, M. Bavay, C. Coléou, C. Fierz, A. Gobiet, P. Hagenmuller, M. Lafaysse, M. Ližar, C. Mitterer, F. Monti, K. Müller, M. Olefs, J. S. Snook, A. van Herwijnen, and V. Vionnet, 2020. Application of physical snowpack models in support of operational avalanche hazard forecasting: a status report on current implementations and prospects for the future, Cold Reg. Sci.Technol.,170,102910, https://doi.org/10.1016/j.coldregions.2019.102910.

Palomaki, R. T. and Z. S. Miller, 2023. Assessing snowpack stratigraphy accuracy based on different input data: insights for operational avalanche forecasting, in: Proceedings of the International Snow Science Workshop, Bend, OR, 8-13 October 2023, 287 - 294.

Skiles, S. M., Clark, S., and Ragar, D., 2022. Hourly snow energy and mass balance at Atwater study plot, Alta, Utah, Hydrol. Process., 36, e14558, https://doi.org/10.1002/hyp.14558, 2022.

Talty, M., Morrison, T. J., and Eskeland, J., 2026. An applied validation and verification framework for computer-assisted public avalanche forecasting, in: Proceedings of the International Snow Science Workshop, Whistler, BC, 28 September-2 October 2026.

Veals, P. G., Pletcher, M., Schwartz, A. J., Chase, R. J., Harnos, K., Correia, J., Wessler, M. E., and Steenburgh, W. J., 2025. Predicting snow-to-liquid ratio in the mountains of the western United States, Weather Forecast., 40, 1825–1838, https://doi.org/10.1175/WAF-D-24-0233.1.

Wais, P., 2017. A review of Weibull functions in wind sector, Renew. Sust. Energ. Rev., 70, 1099-1107, https://doi.org/10.1016/j.rser.2016.12.014.

Wever, N., L. Schmid, A. Heilig, O. Eisen, C. Fierz, and M. Lehning, 2015. Verification of the multi-layer SNOWPACK model with different water transport schemes, The Cryosphere, 9, 2271-2293, https://doi.org/10.5194/tc-9-2271-2015.