



RESEARCH ARTICLE

10.1029/2026JH001312

Key Points:

- Our data-driven albedo scheme achieves improved and generalizable predictions versus existing models at both fine and coarse scales
- A simple scheme using only standard meteorological inputs provides competitive output at low computational burden
- The utilized framework offers potential for efficient and scalable physical parameterizations in next-generation modeling regimes

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. Charbonneau,
aacharbon@caltech.edu

Citation:

Charbonneau, A., Deck, K., & Schneider, T. (2026). A next-generation snow albedo parameterization for climate modeling using constrained machine learning. *Journal of Geophysical Research: Machine Learning and Computation*, 3, e2026JH001312. <https://doi.org/10.1029/2026JH001312>

Received 24 FEB 2026

Accepted 17 JUL 2026

Author Contributions:

Conceptualization: A. Charbonneau**Investigation:** A. Charbonneau**Methodology:** A. Charbonneau, K. Deck, T. Schneider**Supervision:** K. Deck, T. Schneider**Visualization:** A. Charbonneau**Writing – original draft:** A. Charbonneau**Writing – review & editing:** K. Deck, T. Schneider

© 2026 The Author(s). *Journal of Geophysical Research: Machine Learning and Computation* published by Wiley Periodicals LLC on behalf of American Geophysical Union.

This is an open access article under the terms of the [Creative Commons Attribution License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

A Next-Generation Snow Albedo Parameterization for Climate Modeling Using Constrained Machine Learning

A. Charbonneau¹ , K. Deck¹ , and T. Schneider¹ ¹California Institute of Technology, Pasadena, CA, USA

Abstract We demonstrate a data-driven parameterization framework for snow albedo evolution using a constrained auto-regressive neural differential equation that directly predicts snow albedo change from standard meteorological inputs. After training with multi-year in situ and satellite observations from a wide variety of locations, the scheme effectively reproduces daily albedo evolution across diverse climate zones, with median error under 7.5% (RMSE $\approx$ 0.05), a 10%–30% improvement over established models. Furthermore, the model generalizes to sites not seen during training and scales from coarser grids to point locations. The scheme can easily incorporate new features as observational networks expand, offering an adaptive and computationally lightweight framework for next-generation land and climate models.

Plain Language Summary Snow reflects sunlight back to space, and small changes in its reflectivity (albedo) can influence our climate. Accurately modeling how snow albedo changes over time remains a challenge in climate models. We develop a data-driven approach that learns the rate of change of albedo using commonly available weather and snow measurements, instead of variables that would be unavailable or more uncertain within a climate model. The model is trained using several years of satellite and ground-based observations from a wide range of environments. Our model reproduces daily changes in snow albedo across many locations with lower error than widely used alternatives, performing well for both ground measurements and satellite-scale data. Our approach can easily incorporate new observations and is computationally inexpensive, providing a flexible means for aiding next-generation earth science models and beyond.

1. Introduction

Seasonal snowpacks play a crucial role in the cryosphere, influencing climate at small and large scales. They constitute an integrated response to a region's weather, covering nearly half the Northern Hemisphere land surface in winter while also ranking among the fastest-changing land covers by season (Dong et al., 2014; Hall & Riggs, 2007; Klein & Stroeve, 2002). They play a significant role in terrestrial energy and water fluxes, with consequences for ecosystem and economic functions such as water supply, agriculture, and environmental hazards (Hao et al., 2023; Huang et al., 2022; Lee et al., 2023). However, snowpacks are sensitive to changing climates, and historical norms of accumulation and melt no longer reliably capture present-day patterns and timing (Aguirre et al., 2023; Rittger et al., 2020). Snow dynamics remain one of the most challenging aspects to represent in simulations of the water cycle (Pirazzini et al., 2018), emphasizing the need for improved snow models that can adapt and generalize to a range of possible futures at seasonal and multidecadal timescales.

Within snow models, one of the most critical parameters affecting the climate system is the snowpack albedo (α), the ratio of reflected to incident shortwave radiation from the snow surface. Broadband snow albedo integrates wavelength-dependent snow albedo over the spectrum of incoming radiation (Bair et al., 2019):

$$\alpha = \frac{\int_{\lambda_1}^{\lambda_2} \alpha(\lambda) S(\lambda) d\lambda}{\int_{\lambda_1}^{\lambda_2} S(\lambda) d\lambda} \quad (1)$$

where $\alpha(\lambda)$ is the snow's spectrally resolved albedo, $S(\lambda)$ is the solar spectral irradiance at the snow surface, and $\lambda_1 = 250$ nm and $\lambda_2 = 4,000$ nm are conventional cutoffs for the solar shortwave spectrum. Snow exhibits some of the highest natural albedos on Earth (0.7–0.9), making snow albedo and subsequent energy fluxes uniquely sensitive to perturbations or small errors. A small relative reduction in α creates a large relative increase in absorbed shortwave energy ($1-\alpha$) which accelerates melt and lowers albedo further, amplifying a positive feedback on warming (Bair et al., 2019) and leading to error propagation within global climate models (GCMs)

(Qu & Hall, 2006; Thackeray et al., 2018). Through this mechanism, snow albedo exerts a strong control on the surface energy balance and overlying air circulation, and thus regional climate patterns and precipitation (Cordero et al., 2022; Diro & Sushama, 2018; Hall & Riggs, 2007; Kouki et al., 2022; Malik et al., 2014). This influence propagates to global circulation, affecting processes such as ENSO, monsoons, and poleward retreat (Wang et al., 2012), underscoring the need for accurate representation.

Reducing errors in modeled snow albedo is difficult given its variability and multiscale drivers. Snow albedo is governed by a suite of physical processes, including snow metamorphism, deposition of new snow or light-absorbing particles (LAPs) such as dust and black carbon, snow compaction and layer formation, and interaction with the evolving solar/atmospheric state (see Warren, 1982; Zhang & Ohata, 2003). However, snow albedo also varies strongly across short spatial and temporal scales due to topography and microclimate (Molotch & Bales, 2006; Williamson et al., 2016), and modification by LAPs often reflects both local and remote sources, decoupling changes from solely local snowpack processes (Bair et al., 2019). A main challenge in modeling snow albedo lies in representing the interplay of these processes across scales.

Most prevalent snow albedo models are built upon two formative ideas from the 1950s. The first was the application of basic scattering theory to model snow albedo analytically (Dunkle & Bevens, 1956), with later extensions (Wiscombe & Warren, 1980) to incorporate multiple scattering, solar zenith angle, LAPs, and other snowpack properties. The second was the quantification of empirical surface-age and positive-temperature indices in an Army Corps of Engineers report (ACE, 1956) to estimate snow albedo decay. These ideas established the spectrum of theoretical versus empirical approaches to albedo modeling that are still in use today.

Results of these studies remain inconclusive, with complexity showing limited influence on overall parameterization skill and no consensus on which approach best suits large-scale modeling (Essery et al., 2013; Gardner & Sharp, 2010; Krinner et al., 2018; Magnusson et al., 2015; Menard et al., 2021). Simpler age-based schemes can reproduce global mean decay rates, but struggle to capture variability under changing conditions, routinely producing errors of 10%–20% and comparable biases outside their calibration domains (Abolafia-Rosenzweig et al., 2022; Bair et al., 2019; Damiani et al., 2025; Etchevers et al., 2004). Complex or process-based models developed to address these limitations such as SNICAR, TARTES, and SNOWPACK (Flanner et al., 2021; Lehning et al., 2002; Picard & Libois, 2024) demand input availability and process fidelity beyond what is contemporarily available at GCM scales (~10–100 km), leaving them uncertain or intractable even before accounting for increased computational cost (Bair et al., 2019; Chan, 2022; Damiani et al., 2025; Essery et al., 2013; Gaillard et al., 2025; Gardner & Sharp, 2010; Hao et al., 2023; He et al., 2024; Menard et al., 2021; W. Wang et al., 2020). Their sensitivity to poorly constrained inputs and site-specific calibration leave these and intermediate-complexity models with limited climate transferability and error ranges that frequently converge back to those of simpler schemes (Etchevers et al., 2004; Lawrence et al., 2011; Lin et al., 2025; Oaida et al., 2015; Sarangi et al., 2019; Shao et al., 2020).

Therefore, many GCMs continue to rely on combinations of simpler age-based parameterizations for tractability, despite wide performance spreads highly sensitive to model choice, calibration sites, and validation approaches, particularly during the ablative period (Essery et al., 2013; Etchevers et al., 2004; Günther et al., 2019; Magnusson et al., 2015). Snow albedo parameterizations exert larger influence on snow simulations than those for other snow processes due to variability in albedo decay rates affecting snowpack decay, which can alter snow season length by 25–50+ days (Lee et al., 2023; Z. Wang et al., 2012; Etchevers et al., 2004; Essery et al., 2013; Abolafia-Rosenzweig et al., 2022). As GCM resolutions continue to increase toward kilometer or sub-kilometer scales (Aguirre et al., 2023; Ding et al., 2024; Gaillard et al., 2025; Voordendag et al., 2021; Ye et al., 2023), issues of generalization, input sensitivity, and computational burden will only become more acute. This highlights the need for computationally efficient but accurate and generalizable schemes for the next generation of GCMs, which are more data-informed than previous generations in an effort to reduce recalcitrant errors (e.g., in surface energy fluxes) yet still face similar input uncertainty of desired variables like snow grain size or LAP concentration. This necessitates a framework that can represent diverse albedo decay behaviors across climates while still remaining computationally efficient, transferable, and dependent only on widely available inputs.

Recent computational advances and data volume have inspired new albedo parameterization efforts emphasizing grain size, LAP concentration, and spectral resolution such as Bair et al. (2019), Gardner and Sharp (2010), Saito et al. (2019), and Ström et al. (2022). These studies report accurate performance across snow conditions but still inherit the input sensitivity of their predecessors, limiting their utility in GCMs. Machine Learning (ML) instead

offers potential, as ML approaches can capture nonlinear interactions directly from data while avoiding the systematic bias inherent to presupposing explicit functional forms (Damiani et al., 2025; Demil et al., 2025; Fleming et al., 2024). However, efforts specific to snow albedo modeling remain scarce. Many focus on snow coverage to improve data fidelity, or largely post-process input reflectance products or predetermined albedo, lacking the physical basis for incorporation into GCMs where such inputs are unavailable (Chen et al., 2025; Ding et al., 2024; Zhao et al., 2025). These efforts likewise find limited transferability beyond calibration sites, with many errors still plateauing near $\sim 10\%$ (Demil et al., 2025; Ding et al., 2024; Kiem et al., 2024), and many ML-based approaches do not inherently respect physical constraints (De Michele et al., 2013; Gao et al., 2021).

This work presents a novel parameterization, which uses the machine learning framework developed in Charbonneau et al. (2025) to predict the time derivative of the snow albedo given only meteorological inputs and the present snow albedo as inputs. The predictive model output, which obeys prescribed constraints, is then integrated over time to obtain the broadband snow albedo. It is able to outperform a suite of established models over a large variety of sites from only a single calibration on coarse-scale data, with results that extend from the coarser scales of present data products to fine-scale point validation without recalibration. We use the U.S. Snow Telemetry (SNOTEL) network to gather site-level snow state data, in combination with the STC/MODSCAG-MODDRFS 500-m albedo product (Rittger et al., 2024) to create coarse-level data, along with established validation site data and automated weather station data to allow testing over a variety of climates. The approach is also accessible to many computational systems and fast to train and prototype (under 1 min for training on a single CPU, and able to process millions of inputs per second). It is adaptable as new variables (e.g., characterizing aerosol deposition) become available, demonstrating versatile capability for advancing climate modeling.

2. Methodology

We utilize a ML framework developed in 2022 and introduced in Charbonneau et al. (2025) to construct our parameterization, which enforces prescribed constraints (supplied analytical functions, additional parameterizations, etc.) using fixed layers during both training and operation, enabling constrained learning without penalizing the loss function. This scheme was shown to enable efficient, physically consistent, and generalizable parameterizations from simple predictive structures, applicable to a variety of systems and capable of performing on par with more complex models. We combine remotely sensed data in addition to point observations to support benchmarking across scales and a variety of climates. By prioritizing observational data rather than age-based or assimilated features, we avoid circular bias from reanalysis-driven inputs, break from continued reliance on age indices, and achieve more faithful representations of underlying snow processes for usage within GCMs.

2.1. Model

We model the rate of change in snow albedo with a simple neural network, M , driving an ordinary differential equation for the evolution of snow albedo:

$$\frac{d\alpha}{dt} = M(\alpha, T_a, P, \mu) \quad (2)$$

where T_a is air temperature ($^{\circ}\text{C}$), P is liquid water-equivalent snowfall rate (m s^{-1}), μ is the cosine of the solar zenith angle (at solar noon for daily data), and α is current albedo (0–1). The network structure is shown in Figure 1, and constraint functions are picked only to limit the subsequently updated snow albedo value between an applicable range of 0.3 and 1 (Henneman & Stefan, 1999), which restricts reflectivities to the physical range [0, 1]. The constraints are enacted by fixed layers that remain during both training and inference, guaranteeing compliant outputs by construction (prohibiting violative outputs that would arise from their removal) and improving gradient updates within valid bounds during training. This eliminates the need to add loss function penalties, which can impede optimization of the primary objective (Rahimi et al., 2023), and was shown to protect against overfitting and enable learning of more extrapolative final models, compared to identical models trained without constraints that were only applied afterward (Charbonneau et al., 2025).

Unlike our previous work and other snow albedo schemes, we did not want to a priori restrict albedo increases to occur only under precipitation with the selected constraints, since phenomena like the smoothing of the snow surface during melt or overnight freezing can increase snow albedo absent new precipitation. Our utilized

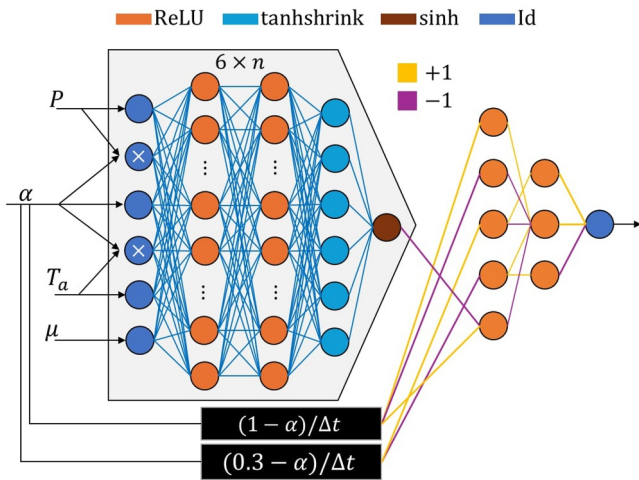


Figure 1. Structure of M . The predictive portion of the parameterization consists of a simple 3-layer dense neural network with an aggregation to a final output. Nodes are colored by their aggregation function, trainable weights (and a bias term) are shown in blue, and fixed weights (no bias term) are shown in yellow (+1) or purple (−1). One hyperparameter n determines the layer widths. The white “x” indicates multiplication to allow usage of the cross-term inputs αT_a and αP . A threshold constraint layer (see Appendix A in Charbonneau et al., 2025) is included around the predictive component to keep the accumulated albedo between 0.3 and 1 for time-step Δt , even during training.

boundaries relax this prevalent but narrow assumption. We also chose not to impose additional structure in this initial study—such as preventing albedo increases under higher air temperatures, or preventing albedo decreases under sufficient precipitation—instead favoring less restrictive bounds which maximally bound the network (not guide the network internal behavior), though the framework can readily represent such choices. We note that while prohibiting albedos greater than 1 or negative albedos imposes physical limits on the generated dynamics, the actual lower value of 0.3 is empirically motivated. This could introduce systematic biases for more extreme snow environments where darker albedos are needed but are architecturally prohibited by the model. However, alteration of this bound is straightforward and does not inherently require retraining of the model weights.

Choosing initial boundaries solely to maintain expected global limits and physical bounds on reflectivity did not presuppose any assumptions about albedo dynamics. This choice allowed us to investigate the extent to which physically consistent relationships could be learned from the data without imposing additional model structure, or whether specifying additional behavior might be necessary. The prognostic format mirrors the differential equations used for other GCM variables, aiding integration, and reflects findings that prognostic parameterizations are more consistent than other forms (Essery et al., 2013). Learning the rate of change directly via a neural network permits a nonlinear expression of a diverse range of rates from an otherwise simplistic input space, addressing one of the main limitations in previous simple or prognostic schemes (Abolafia-Rosenzweig et al., 2022; Etchevers et al., 2004) and improving representative capacity to observed variability.

We deliberately employ a small black-box neural network (NN) as the internal predictive model within the constraint structure, rather than linear regression, tree-based methods, or other ML approaches. For training volumes available in climate parameterizations and this study, reported tabular-data advantages of tree ensembles diminish (Gorishniy et al., 2023; Grinsztajn et al., 2022), while NNs exhibit stronger expressive capacity for the continuous nonlinear interactions governing albedo. While white-box models like single trees are interpretable, ensemble methods lose most structural transparency, offering little practical interpretability over an intentionally compact NN for GCM integration. Explicit cross-terms and ReLU in the first layer reduce the required depth/width to otherwise learn multiplication and enable explicit representation of “controlled” behaviors (e.g., $(\alpha - \alpha_0) * H(P - P_0)$; H is the Heaviside function) with fewer parameters and operations. Unlike trees, inference with a NN is continuous, fully differentiable, more GPU/cache-scalable without conditional branching, and more compatible with Kalman/gradient-based calibration workflows utilized in GCM sub-model tuning. Our framework with NNs was also shown to permit physically generalizable extrapolation beyond the convex hull of the training targets (Charbonneau et al., 2025), whereas tree models cannot, which is important in evolving climate systems where no data is wholly representative. Convolutional or spatiotemporal networks were not pursued, as the governing processes are effectively localized in the selected input features at GCM-relevant scales, and spatial aggregation amplifies missing-data and uncertainty propagation. We emphasize that these choices showcase an informed baseline rather than a final architecture, as the framework remains extensible to alternative model classes and other constraint choices.

Since model complexity shows little correlation with model skill, using a simple network aids computational efficiency, robustness to input uncertainty, and protection against overfitting; all necessary for high-resolution GCMs (Essery et al., 2013; Magnusson et al., 2015). Grain size is deliberately excluded; although it often improves diagnostic accuracy, it increases sensitivity and couples performance to an uncertain or unavailable input at GCM scales, and measurements transfer poorly across regions, risking bias and reduced generalizability. The utilized variable set also breaks from a continued reliance on snow age. Using a 1D model over well-characterized local state variables (i.e., not directly ingesting/requiring spatiotemporal information) allows broad applicability over any spatial grid and better representation of inherently universal processes, while the lack of explicit time

dependence within the predictive component (only time-step dependence in the constraints) ensures the scheme can be run at any GCM time step without retraining (Charbonneau et al., 2025).

The proposed architecture should be able to emulate physically plausible relationships present in the training data, but this chosen construction does not explicitly enforce conservation of energy, radiative transfer equations, snow microstructure evolution, or other mechanistic constraints that can shape the dynamics between the bounds. Physically realistic behavior only arises from the combination of bounds and training data relationships rather than from guaranteed adherence to snow radiative transfer constraints. This model should augment and interact with but not replace radiative-transfer schemes, which resolves physics unavailable to the present parameterization and serves different purposes within a larger model. Rather, this scheme provides an empirical representation of the snow albedo emerging from processes that would otherwise require substantially more complex, uncertain, or unavailable intermediate state variables, and can be used within a larger model to characterize the reflectivity which is needed in determining radiative transfer.

2.2. Data

We combined in situ observations from the U.S. Snow Telemetry (SNOTEL) network with the remotely sensed STC/MODSCAG-MODDRFS 500-m albedo product (hereafter “SMM”). SNOTEL provides collocated ground measurements of snow and meteorological variables. In recent years, over 20 SNOTEL sites have also begun to collect surface radiation data, enabling direct snow albedo derivation for site-level validation. For earlier periods or other sites, snow albedo time series were extracted from the corresponding SMM grid point. SMM improves on earlier MODIS-based products through enhanced corrections for clouds, canopy viewing angle, and ground slope, reducing errors by nearly threefold against widely used products (e.g., MOD10A1), especially in forested regions where most SNOTEL sensors are located. It more accurately represents snow albedo instead of mixed surface albedo as well as snow onset, disappearance, and melt season dynamics down to lower snow cover fractions (Aguirre et al., 2023; Hao et al., 2023; Rittger et al., 2013, 2020; Sarangi et al., 2019; Stillinger et al., 2023), with typical validation errors of 0.05–0.08 (Bair et al., 2019; Dong et al., 2014; Hatchett et al., 2023; Painter et al., 2009, 2012; Palomaki et al., 2025; Williamson et al., 2016). Usage of SMM has yielded 10%–25% gains in hydrological modeling skill, especially during melt season, underscoring its suitability for next-generation models (Fleming et al., 2024). We are additionally not aware of prior studies using the new SNOTEL albedo data for model validation, offering a novel benchmark. To further test generalization and provide comparable benchmarks, we incorporated standard evaluation sites from Col de Porte (Lejeune et al., 2019), Kūhtai (Krajčič et al., 2017), and Mammoth Mountain (Bair et al., 2018), and include additional data from Sodankylä (Essery et al., 2016), the Upper Rofental (Warscher et al., 2024), the Himalayas (Shea et al., 2015; Stigter et al., 2021), and one site in Chile from the Centro de Estudios Avanzados en Zonas Áridas (CEAZA) meteorological network, to test performance across out-of-sample elevations and climates.

Both SMM and site sensor data have known quality issues. For SMM, accuracy varies by location and with vegetation cover, and it can exhibit geolocation errors of up to one pixel (Bair et al., 2019, 2022; Demil et al., 2025; Rittger et al., 2020). SMM’s underlying calibration data is not inherently representative of all climates (Keuris et al., 2023) and can influence poor distinction between snow types (Hao et al., 2023), which can bias snow cover fractions and subsequently derived albedos (Lee et al., 2023; see Appendix A). Systematic underestimation of snow-covered areas, summertime values, and misrepresentation of shallow snowpacks (<5 cm) suggest non-snow albedo continues to contaminate reported values (Ding et al., 2024; Palomaki et al., 2025; A. Klein, 2003). For SNOTEL and other site sensors, precipitation is often underreported, and sensor records can contain biased or unphysical values (Hill et al., 2019; Meyer et al., 2012; Warscher et al., 2024). Because no universal cleaning methods exist—particularly for the new SNOTEL albedo records—we developed a custom procedure (see Appendix A) to quality-control, de-noise, smooth, and scale albedo time series, supplemented with established methods for other variables (Atwood et al., 2023; Livneh et al., 2014; Serreze et al., 1999; Yan et al., 2018). From cleaned records, da/dt was derived for days with complete data (excluding $\Delta t > 1$ day), creating a table of input feature values and an associated target da/dt that could be shuffled for training. Daily zenith angle cosine was computed with `Insolations.jl` (Contributors, 2020), and snow–rain fractions were derived using the parameterization from Jennings et al. (2018), shown to be over 85% accurate. Data were scaled but not normalized (no shift) to preserve the physical meaning of input zeroes. Data scaling constants (standard deviations for inputs, absolute maxima for targets) were embedded within the neural network to ensure consistent preprocessing and postprocessing, eliminating the need for users to manually apply input and output scaling

Table 1
Summary of Data and Variables

Data					Variables			
Source	Sites	Days	Scale	Purpose	Symbol	Units	Name	Description
SNOTEL/SMM Albedo	40	86841	500 m	Training	α	None	Snow Albedo	Derived, see Appendix A
SNOTEL/SMM Albedo	700	1527868	500 m	Testing	P	ms^{-1}	Snowfall (w.eq.)	Direct or derived (Appendix A)
SNOTEL/SNOTEL Radiation	23	5318	Point	Testing	T_a	$^{\circ}\text{C}$	Air Temp.	Averaged to daily level
Col De Porte, France	1	2340	Point	Testing	μ	None	Zenith Angle	Cosine at solar noon
Kühtai, Austria	1	4532	Point	Testing	z	m	Snow depth	Direct data
Sodankylä, Finland	1	317	Point	Testing	SWE	m	Snow w.eq.	Direct data
Bella Vista, Upper Rofental	1	457	Point	Testing	ρ	None	Bulk Density	Derived as SWE/z
Proviantdepot, Upper Rofental	1	741	Point	Testing	τ	days	Snow Age	See below & Appendix A
Yala Basecamp, Himalayas	1	126	Point	Testing	α_g	None	Ground Albedo	Derived, see Appendix A
El Tapado, Chile (CEAZA Network)	1	268	Point	Testing	DOY	None	Day of Year	Derived from Date
Dozier Snow Study Site (CUES)	1	1490	Point	Testing	Date	Date	Date	Direct data
TOTAL	771	1630298			T_s	$^{\circ}\text{C}$	Surface Temp.	Estimated, $\min(0, T_a)$

Note. This table denotes the names of utilized data sources and their purpose, as well as the variables mentioned throughout the paper. The Days column refers to the number of sensor days with complete records of at least α , P , and T_a . “Water equivalent” is abbreviated as “w. eq.” Bulk Density is represented as a unitless fraction of the density of liquid water, $1,000 \text{ kg m}^{-3}$. Only variables listed in Equation 2 are used within M . The spread of snowpack types and geographic locations of all sites can be found in Figure S1 of Supporting Information S1.

during model evaluation. This yielded roughly 1.6 million site-days with SMM albedo and 15,300 with direct observations.

2.3. Model Training

The combination of SNOTEL sites with the SMM values along with the other validation sites provides two albedo spatial scales for model evaluation: point-scale time series from ground sensors to assess emulation of underlying albedo processes, and extracted 500 m grid-point timeseries to gauge representation of coarse-grained dynamics more relevant for climate models and evolving GCM needs. To emphasize generalizability, the model was trained on a small subset of the coarser SMM data (40 locations out of 740; see below), saving the rest for coarse-scale testing. We also used the same resulting calibration to test over out-of-sample ground-level records from varying climate types, forcing a more stringent test of the parameterization's capacity to extract and learn relationships that should transfer across scales. We used only the daily data from the 40 SNOTEL sites overlapping with our previous study for training, as these were shown to exhibit sufficient diversity in elevation and local climate to enable robust parameterization (see spread in altitude and snowpack types in Figure 3 from Charbonneau et al. (2025), as well as Figure S1 in Supporting Information S1 for all sites in this study). This left 700 unused SNOTEL sites for coarser testing and 31 site-level test locations, providing extensive performance statistics while highlighting the framework's ability to generalize from a relatively smaller training sample ($\sim 87,000$ batched and shuffled site-day samples from SMM). The model was trained with L_2 loss, using $n = 2$ for the width hyperparameter (see Appendix C for hyperparameter selection), creating a total of 331 weights. Training for 100 epochs (100 passes over all training data) with batch sizes of 256 requires <30 s on a single Intel i9 CPU, with a storage footprint under 2 kB. This likewise competes with the usual training time advantage associated with tree ensembles and other non-adjoint machine learning approaches. Training was implemented in Julia's Flux framework (Innes, 2018) using the root-mean-square propagation (RMSProp) optimizer (Hinton et al., 2012). Data sources and variables are summarized in Table 1.

2.4. Model Testing and Comparison

Magnusson et al. (2015) notes new parameterizations are often benchmarked without comparison against existing models, limiting assessment of their true value. To address this, we compare our parameterization against a comprehensive set of alternatives (Table 2), representative of those used in current GCMs (Lee et al., 2023) or

Table 2
Parameterizations Compared in This Study

Name	Symbol	Identifiable origins	Referenced paper	Variables	Type	P reset	AR	Has been in:
ACE Curve, Polynomial	AC	ACE (1956)	Molotch and Bales (2006)	τ	D	T	N	–
Exponential Decay (Snow Age)	ED	Riley et al. (1969)	Kondo and Yamazaki (1990)	τ	D	T	N	US ACE HEC-HMS
Exponential Decay (T_a Index)	ET	Ranzi and Rosso (1991)	Ranzi and Rosso (1991)	P, T_a	D	T	Y	–
Logistic/Exponential Decay (T_a Index)	LE	Pike et al. (2024)	Pike et al. (2024)	P, T_a, z	D	T	Y	–
VAS	VA	Malik et al. (2014)	Malik et al. (2014)	P, T_a, z, α_g	D	T	Y	–
Brock et al. (2000)	BR	Brock et al. (2000)	Brock et al. (2000)	P, T_a, SWE, α_g	D	T	Y	–
Marshall (“Simple” Depth)	MD	Marshall (1989)	Marshall (1989)	τ, SWE, α_g	D	T	N	–
Oerlemans & Knap	OK	Oerlemans and Knap (1998)	Liu et al. (2021)	τ, z, α_g	D	T	N	Noah
ESCIMO	EC	Rohrer (1991)	Strasser and Marke (2010)	τ, T_a	D	T	N	ESCIMO
Interpolation (T_a)*	IT	Roeckner et al. (1992)	Mölders et al. (2007)	T_a	D	–	N	ECHAM, ECMWF, JSBACH3
Interpolation (Density)	ID	Greuell and Konzelmann (1994)	Greuell and Konzelmann (1994)	ρ	D	–	N	SOMARS, RACMO2, CLSM
Noah Exponential	LV	Anderson (1968)	Livneh et al. (2010)	τ, DOY	D	T	N	Noah, DHSVM, VIC
SNOWPACK-like	SP	Lehning et al. (2002)	This	P, T_a, μ, z, ρ	D	–	N	SNOWPACK
Zhongyuan & Ohata Density	ZO	Zhongyuan and Ohata (1989)	Zhang and Ohata (2003)	ρ	D	–	N	–
NASA GISS*	NG	Hansen et al. (1983)	Hansen et al. (1983)	P	D	C	Y	GISS Model II, ORCHIDEE, CLSM
CLASS*	CL	Versegny (1991)	Li et al. (2023)	P, α	PV	C	Y	Noah-MP, ISBA, CLASS
UTOPIA	UT	Versegny (1991), Baker et al. (1990)	Cassardo (2015)	P, α, T_s, μ	PV	C	Y	UTOPIA
HTESSEL*	HT	Versegny (1991), Baker et al. (1990)	Dutra et al. (2010)	α, P, SWE, T_s	PV	C	Y	HTESSEL/ECMWF
Giddings and La Chapelle	GC	Giddings and LaChapelle (1961)	O’Neill and Gray (1972)	z, α_g	D	–	N	–
JULES*	JL	Essery et al. (2001)	Best et al. (2011)	P, T_s	D	C	Y	JULES
Marks (1988)	MS	Marks (1988)	Marks (1988)	τ, μ	D	T	N	SNOBAL
BATS*	BA	Dickinson et al. (1986)	Abolafia-Rosenzweig et al. (2022)	P, T_s, μ	D	C	Y	Many (Noah-MP, WRF CoLM...)
ISBA*	IS	Brun et al. (1992)	Decharme et al. (2016)	z, T_s, T_a, P	D	C	Y	Crocus, ISBA MAR
Neural Network	NN	Prev. Work (Charbonneau et al., 2025)	This	α, P, T_a, μ	PR	–	N	–

Note. We benchmark the proposed parameterization against an exhaustive (and fundamental, as others are similar to those above) list of other simple forms (no grain size, only available macroscopic state variables) used in past and present GCMs or proposed in literature, noting their original appearance and reference consulted for parameter values. The table denotes each scheme’s symbol for usage throughout the paper, as well as the model type (prognostic-rate, prognostic-value, or diagnostic, as PR, PV, and D, respectively), whether the scheme uses a continually- (C) or threshold- (T) based reset of internal snow age/indices, and whether the scheme was calibrated autoregressively (AR) or via regression (yes[Y]/no[N], see below). Equations are provided in Supporting Information S1. Models explicitly found in CMIP6 (Eyring et al., 2016) are denoted with an asterisk. Some other CMIP6 parameterizations are very similar to listed table members (e.g., UTOPIA, BATS, T_a Interpolation), or constant values. Other listed schemes have also been used in non-CMIP6 global or regional modeling.

proposed in prior studies. Consistent with Magnusson et al. (2015) and similar efforts, alternative schemes are primarily compared as-published beyond adjusting time-step-dependent parameters to daily resolution, since many have already been tuned for global use. Additional recalibration may reduce absolute errors but rarely alters the rough ranking or performance categories of schemes (Essery et al., 2013), making such adjustments

unnecessary. However, since our neural parameterization must be calibrated to be compared, for fair and holistic comparison a supplemental benchmarking was included for all schemes upon recalibrating them to the same data (see the Supporting Information S1 for equations and parameter values).

The benchmarked schemes fall into diagnostic or prognostic classes, with some incorporating internal snow-age or other “latent-state” formulations which rendered them unable to be trained via regression like the neural scheme (see below). All schemes except ours predict albedo directly rather than its derivative. This distinction leads to three equations for generating time-stepped outputs, diagnostic (Equation 3), prognostic-value (Equation 4), and prognostic-rate (Equation 5, our neural parameterization):

$$\hat{\alpha}_i = f(\theta_i) \quad (3)$$

$$\hat{\alpha}_i = f(\widehat{\alpha_{i-1}}, \theta_i) \quad (4)$$

$$\widehat{\alpha_{i+K}} = \alpha_i + (K\Delta t) M(\hat{\alpha}_i, P_i, T_{a,i}, \mu_i) \quad (5)$$

where a hat indicates the model prediction (and no hat for the observation) and θ_i indicate the scheme's (denoted f) relevant set of non-albedo input variables at step i . All schemes were initialized to observed values $\hat{\alpha}_1 = \alpha_1$. For the neural scheme, the integer K denotes the step size across sequential days ($K = 1$) or “gaps” from missing/cleaned/no-snow periods ($K > 1$). We note that when traversing data gaps via an Euler step with $K > 1$, the resulting albedo could violate the bounds dictated by the predictive model (designed for step-size $K = 1$), though such error should be attributed to the time-stepping choice, not the parameterization itself. To mitigate this, following our previous work (Charbonneau et al., 2025), a maximum gap size of $K_{\max} = 5$ days was permitted before time series were “reset” to observed values $\widehat{\alpha_{i+K}} = \alpha_{i+K}$ if $K > K_{\max}$, and unphysical albedos from $1 < K < K_{\max}$ traversals were likewise clamped to $[0.3, 1]$ to avoid attributing error from the time-stepping choice to the parameterization for fair evaluation, as such issues would not occur within a global model setting. Other schemes were reset for any gaps, giving them a “best-case” comparison. Latent-state schemes were not resumed upon reaching a gap until receiving at least $Q_0 = 3$ mm of liquid-equivalent new snowfall to reinitialize internal snow “age”/state from an otherwise indeterminate value (unless Q_0 was specified differently in their documentation), consistent with theoretical analysis from Wiscombe and Warren (1980), though age/state dynamics after such initialization were dictated only by the scheme.

2.4.1. Calibration of Existing Schemes

Most schemes with an internal snow age (τ) utilized a simplistic threshold formulation equal to the number of days since the last snowfall event ($P\Delta t > Q_0$), allowing a priori external engineering of the τ feature and training of such schemes via regression over the same batches as the neural scheme. However, some latent-state schemes used continuous or adaptive expressions, requiring autoregressive sequence-based training. In such cases, “batches” were constructed from contiguous cleaned data segments of at least 4 days. To avoid indeterminate initial latent states, new sequences were only started at days with $P\Delta t > Q_0$ liquid-equivalent snowfall. Each sequence was initialized with a new latent state and stepped forward by the parameterization using the appropriate form from Equations 3–5, scoring sequences by comparing simulated and observed albedos using the L_2 norm. All schemes could be trained on their respective batches with the same L_2 optimization framework. The neural scheme (and others) could also be trained in this autoregressive manner if desired, though such training was slower (roughly 5×) without necessarily leading to better results compared to regression on single-day inputs and targets, leading to autoregressive training being reserved only for the models labeled “Y” for the “AR” column in Table 2.

2.5. Statistical Analysis

Evaluation metrics included root-mean-square error (RMSE), root-mean-square percentage error (RMSE%), bias (B), and bias-percentage (B%) between simulated and observed albedo time series. Statistical significance of RMSE and RMSE% differences across validation and test sites was assessed using the Wilcoxon signed-rank test (Wilcoxon, 1945) at a $p = 0.05$ significance level, a nonparametric method over matched samples that avoids assumptions of normality or equal variance. The input features of the model exhibit non-negligible correlation, which can distort more traditional interpretability methods like Shapley additive explanations (SHAP) analysis,

Table 3
Quantified Scheme Behavior

Values	Coarse	Site
Mean RMSE%	4.98%	5.79%
Mean RMSE% , Snowing	4.88%	5.57%
Mean RMSE% , No Snow	5.05%	5.85%
Mean RMSE% , No Snow, $T_a > 0^\circ\text{C}$	5.86%	6.38%
Mean RMSE% , No Snow, $T_a < 0^\circ\text{C}$	4.55%	5.49%
Mean $d\alpha/dt$, Snowing	0.00085 d^{-1}	0.0012 d^{-1}
Mean $d\alpha/dt$, No Snow	−0.0021 d^{-1}	−0.0016 d^{-1}
Mean $d\alpha/dt$, No Snow, $T_a > 0^\circ\text{C}$	−0.0036 d^{-1}	−0.0031 d^{-1}
Mean $d\alpha/dt$, No Snow, $T_a < 0^\circ\text{C}$	−0.0013 d^{-1}	−0.00067 d^{-1}

Note. This table displays outputs and errors from the scheme over both the coarse and site level testing data across a range of environmental conditions.

permutation importance, or local interpretable model-agnostic explanations (LIME) (Molnar, 2025). Therefore, to qualitatively assess physical consistency, we derived a first-order accumulated local effects (ALE) plot of the neural parameterization, which isolates the marginal influence of individual features even under correlated inputs. Additional discussion of ALE construction and interpretation is available in Charbonneau et al. (2025), Apley and Zhu (2020), and Molnar (2025). Feature interaction strength was quantified with pairwise Friedman's H-statistics (see Table C1), which were sufficiently small to focus analysis on primary variable behavior. Emphasis was placed on assessing larger generalization patterns and error robustness rather than isolated metric optimization or detailed mechanistic decomposition, as the study serves to introduce the potential of an adaptable parameterization framework for near-term modeling needs rather than provide an exhaustive technical analysis of a single finalized model.

3. Results

3.1. Model Benchmarking

The distribution of RMSE% scores from all schemes as-published (various symbols, see Table 2) and our approach (*NN*) over both coarser (500-m scale) and site-level (point observations) data not seen during training are shown in Figure 2. All generated albedos by our scheme show a correlation against observations of $R^2 = 0.786$ over the coarse data and $R^2 = 0.677$ over the site-level data. The average correlation scores over testing sites can be further compared against other schemes in the modified Taylor Diagrams (Elvidge et al., 2014), which show both standard deviations and correlations of simulations and observations (Figure 3). Table B1 in Appendix B displays numerical RMSE%, RMSE, B, B%, and Wilcoxon Signed Rank (Wilcoxon, 1945) scores comparing our scheme to other schemes.

Our model exhibits a median snow albedo RMSE of 0.049 (RMSE% of 6.8%) across all coarse sites, increasing by less than 1% (RMSE 0.056) for point observations, and with less than 0.01 bias at both scales, improving over 30% from other schemes and surpassing the routine 10% error found in other studies. The distribution of RMSE% scores of our scheme is narrower than that of other schemes, showing a consistent range of performance, with fewer outliers (Figure 2). The two halves of the violin plot are relatively symmetric, suggesting good generalizability across different climates (consistency across sites) as well as generalizability across spatial scales (500 m and site-level sites). This is further corroborated by a lack of distinct patterns in performance with regards to location or snowpack type (see Figure S1 in Supporting Information S1), suggesting broad applicability. With only one calibration at coarser scales, our scheme shows an ability to readily compete against other models that are specifically calibrated for global usage across a range of locations and resolutions.

Most other schemes show less of an ability to maintain consistent performance when transferring across spatial scales—while *NN* shows slightly better performance on coarse data than site data, other schemes tend to show slightly better performance on site data, which could be influenced by the calibration data used for these models

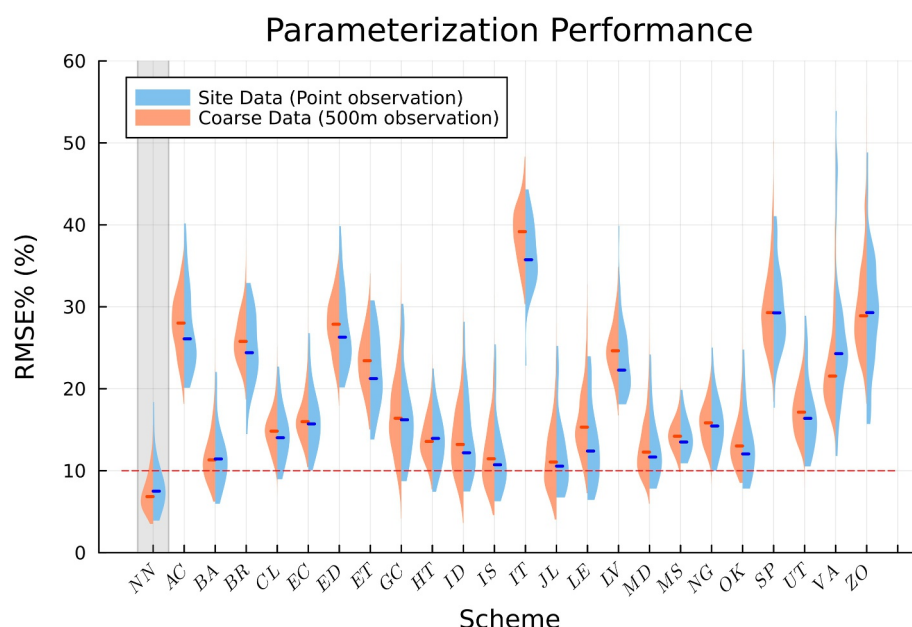


Figure 2. Relative Root Mean Square errors of the new parameterization (*NN*) versus existing parameterizations. This split-violin plot shows the distribution of RMSE% scores for each scheme across all evaluated sites (700 500 m scale sites on the left in orange, 31 point observational sites on the right in blue), with the median of each set marked by the horizontal bars. Parameterization symbols match those described in Table 2, and our scheme is highlighted in gray to the left of the other schemes (organized alphabetically; scheme parameters taken as published). Our proposed scheme outperforms all other compared schemes, with the bulk of sites below the marked 10% error threshold (dotted red line) common in snow albedo parameterizations. It also shows good generalizability; with a short tail (a tight spread of scores) and relatively symmetric halves between validation and testing sites (consistent performance across varied out-of-sample regimes). All scores are visible on the plot.

compared to the explicit calibration to coarse data in our approach. The lack of schemes near the unity normalized standard deviation contour in Figure 3 also indicates our model is more capable at replicating the variability of snow albedo observations. Our approach also shows the highest average correlation to observations. All comparisons were significant at the 5% level ($p < 0.05$ for all Wilcoxon tests, with 700 samples for coarse sites and 31 for point-observation sites), detailed in Table B1.

3.2. Model Consistency

Time series of our scheme output versus that of well-known top-quintile alternative schemes like BATS (BA), JULES (JL), ISBA (IS), and HTESSEL (HT) are given in Figure 4. Compared to the other schemes, our approach is more able to trace the trends of the albedo within the growth and melt season and match the range of slopes created by the observations across different conditions. By expressing a variety of growth rates, the neural model can better track the seasonal dynamics compared with other models with more rigid evolution structures or a finite set of growth and depletion rates. While this is useful over accumulated seasonal timescales (and thus useful for gauging the accuracy of accumulated energy fluxes within a GCM), we do note this approach does not capture the abrupt increase in albedo due to new snowfall or fast decreases as well as the other schemes, preferring to instead more smoothly climb and fall and transition between different growth and depletion rates, which is likely influenced by our use of training data at coarser scales, L_2 loss, and data cleaning approaches. At higher temporal resolutions where high-frequency albedo changes can occur from snowfall or intraday zenith angle dependencies, our approach will likely track fluctuations less but track the average well. This emphasis is more conducive to the needs of climate models and long-term studies, where low systematic bias in trends and integrated fluxes is critical over sub-weekly variation or transient fluctuations (National Research Council, 1999; Nguyen et al., 2016).

If the spread of scores aggregated across snow seasons was primarily dominated only by strong seasonal signals (i.e., merely reproducing the average increase-decrease cycle from the training distribution), strong degradation

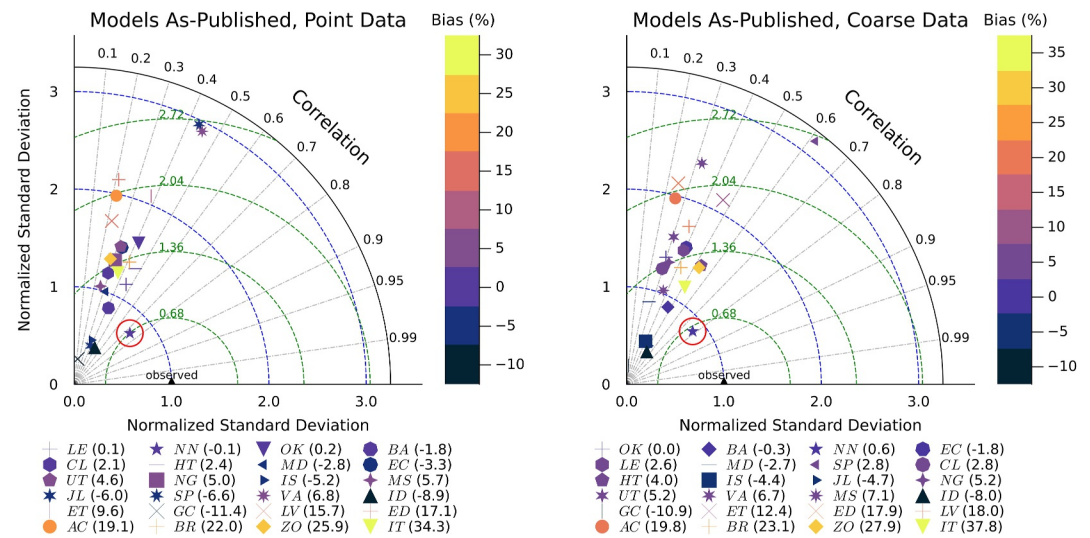


Figure 3. Modified Taylor Diagrams of all schemes over all ground (left) and coarse (right) testing data. The polar coordinates of each scheme mark the standard deviation of time series outputs (normalized to that of the observations; the radial coordinate) and the Pearson correlation of time series outputs against the observations (the azimuth), averaged over all testing sites. The geometric relationship between standard deviation, correlation, and error means the Euclidean distance from each scheme's marker to the observations is equal to the centered root-mean-squared (RMS) error against the observations (also equal to the standard deviation in model error), for which contours of centered RMS error are marked in green, contours of normalized standard deviation are marked in blue, and contours of correlation in black. Scheme markers are colored in accordance with their B% score, and the legend is ordered in terms of increasing |B%|, with the value listed next to the symbol. For both types of data, our scheme (circled in red) is best able to replicate the observations with one of the best B% scores.

and patterns should be expected when transferring the scheme between regimes of different climatic and snow conditions. To investigate this, time series errors decomposed by snow surface age are displayed in Figure 5, which additionally allows exploration of snow errors within a snow season and by snow type. A similar breakdown of error spread by snow year, snow season average temperature, and season total snow accumulation were also assessed (see Figures S2–S4 in Supporting Information S1), exploring the scheme robustness against more extreme or out-of-distribution conditions.

For error decomposition by snow age, the difference between coarse and site-level performance is more separated for older snow surfaces. Error increases slightly with snow age for site-level evaluations, but the majority of errors still remain under the 10% level, meaning the spread in errors even with snow age effects remains small. For coarse-level data, the distribution of errors remains the same regardless of snow age, and errors between coarse and site-level data remain symmetric until the snow surface is roughly 2 weeks old. These similarities again suggest the developed simple scheme exhibits capable generalization between snow types and across spatial scales. Results of the other error decompositions by snow year, average temperature, or total snowfall exhibited even weaker patterns (Figures S2–S4 in Supporting Information S1), with little pattern or variance in error spread across a variety of snow age or environmental conditions, particularly for the coarse-level data. Although seasonal evolution contributes substantially to snow albedo variability, the model is trained only on daily albedo tendencies independent of the bigger snow season, and exhibits relatively uniform performance across geographically distinct sites (Figure S1 in Supporting Information S1) in addition to diverse snow and environmental regimes. This consistency implies that model skill is not solely derived from reproducing a generic seasonal cycle and exhibits capable generalization across a wide variety of evaluation modes from a single training.

To assess the scheme's behavior by-feature, an Accumulated Local Effects (ALE) plot of our scheme is shown in Figure 6 below, with average outputs and errors under different environmental conditions tabulated in Table 3. The ALE plot displays the centered changes in model output (the change in output da/dt compared to the average output da/dt), accumulated over sequential bins of feature values. This mitigates some of the distortions that can arise in marginal effect analyses when predictors are correlated, making it more suitable than partial dependence methods for interpreting learned responses in this setting where predictors are not fully independent.

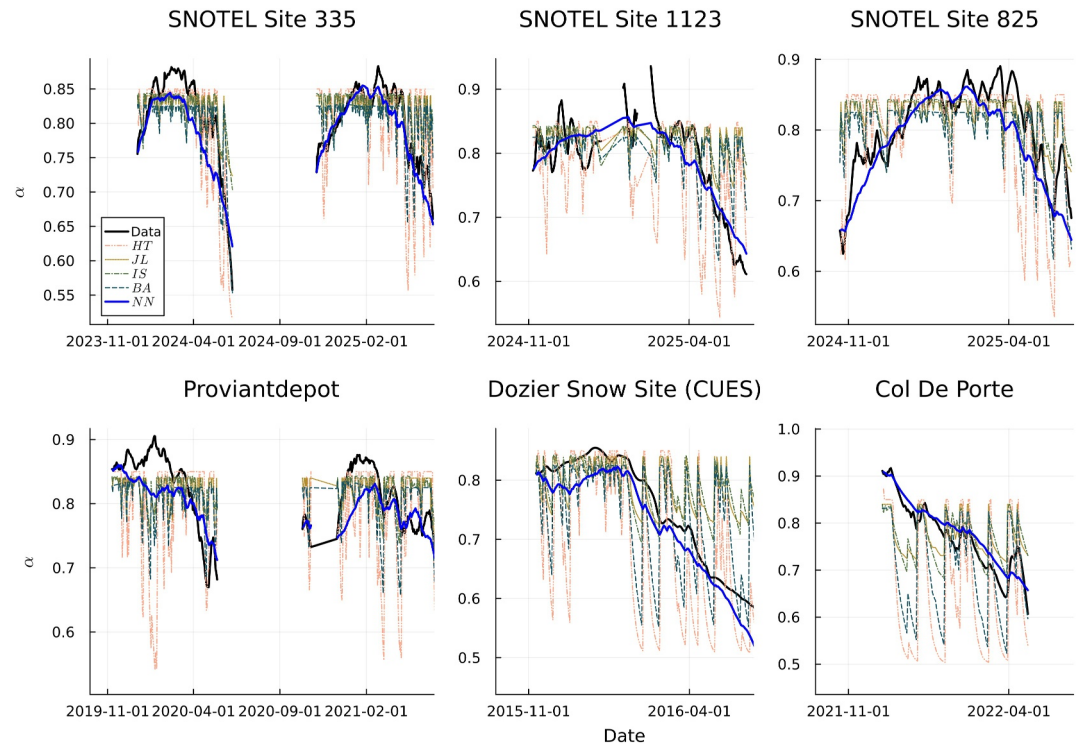


Figure 4. Time series of albedo generated by various schemes. This figure details generated time series over a selection of site-level testing sites compared to the processed time series data, using symbols provided in Table 2. While other schemes show a better capability at creating fast fluctuations or jumps from snowfall, our scheme shows a better ability to follow the overall trend of the albedo throughout the snow season. The shown years were selected to highlight the difference between the trend-tracking behavior of our model versus the fluctuations of other parameterizations.

Interpretation of curve shape (but less so scale) can then be used to gauge behavioral alignment with physically informed expectations. In addition to the curve for the presented scheme, the distribution of curves from another 50 trained schemes (identical structure, different initial weights) is also displayed to visualize scheme sensitivity.

The RMSE% errors in Table 3 for both coarse and site-level data vary by only $\sim 1\%$ across a range of environmental conditions, without accumulation in a seasonal manner. Despite the model failing to capture abrupt increases from snowfall compared to other schemes and a low average output derivative during snowfall events (~ 0.001 per day), the average error between snowfall and no-snowfall events remains similar, though it is slightly higher for days without snow that are above the melting temperature than those below the melting temperature. The output derivatives by the network are also positive under snow conditions and negative under no-snow conditions, with larger decreases for above-freezing T_a than below-freezing T_a , aligning with expected physical behavior. The output derivatives without precipitation average roughly -0.002 d^{-1} , which is similar order to (and about half the magnitude, though the means are averaged over a range of values) those found in analogous linear albedo decay equations like that of Baker et al. (1990) or the schemes in UTOPIA (UT) and HTESSEL (HT). The slightly higher error under warmer temperatures and no precipitation corroborates the findings with snow age in Figure 5, suggesting the model performs slightly better in early season than in late season like many established albedo models.

The mean prediction of our model over all data is $da/dt \approx 0$, so positive/negative y-axis values in each ALE plot (the average change in model output relative to the mean prediction) can be interpreted as the model outputting positive/negative derivatives, as opposed to merely more positive/negative values than some finite average. The ALE curves suggest that the learned statistical relationships are broadly consistent with established physical expectations for snow albedo evolution (see Figure 6). Albedo tends to decrease unless there is a minimal amount of snowfall (2.1 mm liquid-equivalent), which then increases but levels off as the amount of snowfall increases past a $\sim 2 \text{ cm/day}$ threshold. The minimal snowfall amount is consistent with the utilized data resolution

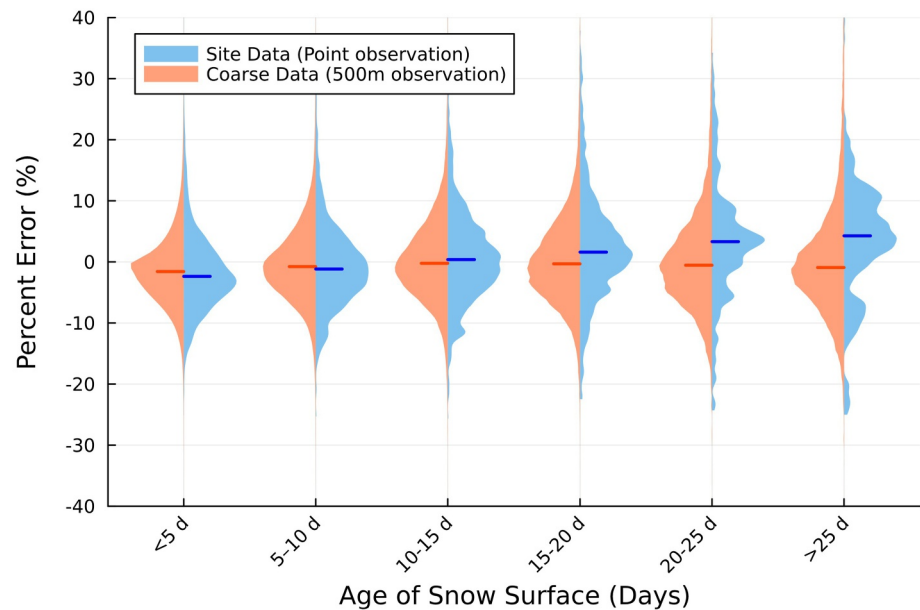


Figure 5. Time series error decomposition by snow surface age. This figure details the breakdown of RMSE% scores in generated time series by the age of the snow surface for the developed ML scheme, following the split-violin format introduced in Figure 2. The symmetry between coarse and site-level results breaks when snow age is more than 2 weeks old, with higher error for older snow surfaces, though all errors still remain small compared to those found in the other schemes.

(2.54 mm), and the stabilization threshold is consistent with measured limits of optically thick snowpacks for both old and new grain sizes (Dozier, 1991; Wiscombe & Warren, 1980). Similarly, the model begins to tend toward predicting decreases in albedo as temperatures approach and increase past -3.6°C , with a slower rate of change in output near colder temperatures where less melting would occur. This value is sufficiently close to the melting temperature where we would expect melt-based albedo decay to manifest, and is consistent with the temperature threshold found in HTESSEL (-2°C) to dictate albedo decay. The albedo curve exhibits regression to the mean, with an intercept of 0.78, the mean albedo in the training data.

While intraday albedo changes decrease and then increase (an intercept of $\mu = 0.66$, about 48° , which is also the mean in the training data) as the sun enters and exits at low zenith angles, the daily albedo trained here uses the

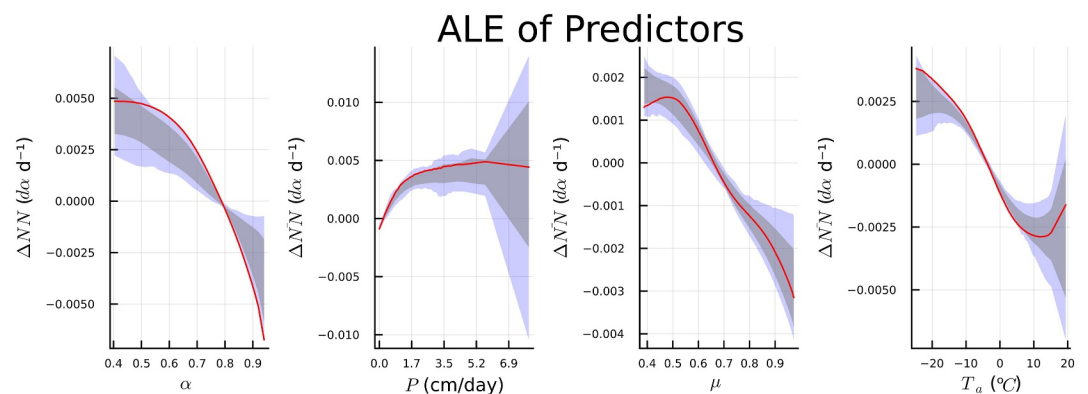


Figure 6. ALE Plot for features of NN . The horizontal axes give the range of each of the four input variables (present albedo α , water-equivalent snowfall P , cosine of solar noon zenith angle μ , and air temperature T_a) while the vertical axes show the average change in the scheme relative to the average prediction of albedo change per day (which is $da/dt \approx 0$). Curves are binned so that each bin has at least 50 samples to average. Gray bands represent the mean and one standard deviation of the ALEs from an ensemble of trained models, while the blue bands fill the 5%–95% quantiles. Scheme behavior becomes more variable in regions where there are limited samples to inform the training process, toward the extremes of the available data (e.g., P after 5.5 cm d^{-1}).

zenith angle cosine at local noon, which increases during warmer seasons. The ALE plot thus shows the tendency for albedo to decrease during warmer days of the year, again consistent with physical expectations. However, observational or structural limitations can influence these results—the concavity at higher T_a could indicate an artifact of the training, the slowing rate of albedo decrease as albedo approaches its minimal values due to low/no snowpack at the temperature ranges shown, or could alternatively be influenced by the number of samples in the bins at the extreme ranges of the data (there are orders of magnitude less samples after $T_a > 15^\circ\text{C}$ or after $P > 6\text{ cm/day}$ compared to elsewhere along the curve, so statistics can become slightly skewed or the curve become less smooth compared to other feature ranges). The estimation uncertainty of each ALE curve in the ensemble (the standard errors of the means that build the curve, accumulated in quadrature when the local effects are accumulated) is negligible (less than the width of the line on the plot) compared to the range of curves across different trained schemes, implying the structural/parameter uncertainty of the presented model form is dominant. This means the spread seen in Figure 6 is indicative of a high sensitivity to the scheme initial state, though this is unsurprising for a small neural network. The stability of the learned relationships, however, is comparatively robust within most of the ranges found in the data, with the ALE ranges varying more at extreme values (which are poorly represented in the data and thus carry comparatively little influence on the training process) but the overall shapes and scales of the curves remain similar. We note that the reported interpretability results should be viewed as diagnostic rather than mechanistic or implicative of causality. The ALE analysis characterizes the statistical sensitivities learned by the network and does not establish causal physical relationships, though it permits exploration of their alignment with expected behavior.

3.3. Calibrated Results

A “best-case” ancillary consideration for holistic comparison against our scheme was also examined, by additionally calibrating all other schemes (already designed for global usage) directly to the data to examine the best performance their functional form could achieve (see Overview in Supporting Information S1). Calibrating specifically to the same data did not change the overall conclusions and rough scheme ranking, in alignment with findings from Magnusson et al. (2015) and Essery et al. (2013). Many other schemes still show a large number of sites with errors in the 10%–20% range after calibration (see Figure 7), while the bulk of sites at both scales by our model are in the <10% error range. Even after calibrating the other schemes (maintaining our scheme's original calibration), our model exhibited significant improvement in RMSE and RMSE% at the 5% level for both coarse and point resolution, except for one insignificant RMSE% score ($p \approx 0.21$) yet a near-significant improvement in the corresponding RMSE score ($p = 0.0695$) against the HTESSEL scheme (HT), for only the point-level data—our model still showed significant improvement over HTESSEL at the coarse level. Despite this, our scheme still showed roughly 0.5 RMSE% (RMSE 0.007) better performance over point-level sites than HTESSEL, and reduced bias by nearly an order of magnitude (see Table B2 in Appendix B). The BATS (BA) and JULES (JL) schemes were still among the better schemes after calibration, though the HTESSEL (HT) scheme increased in competitive ranking. That our model and other prognostic schemes are among the top performers corroborates earlier findings from Essery et al. (2013), implying prognostic schemes fare better than diagnostic schemes. Even after calibration to the same data, most other schemes still exhibit better performance at the site level than the coarse level - this could be due to their tendency to fluctuate more than the NN, impacting their ability to emphasize the lower-frequency trends that dominate at coarse-grained scales. Figure 8 shows that calibration aided all schemes in reproducing the variability seen in the data, though our proposed scheme continues to be the best at reproducing the observations, having the second lowest mean B% and top quintile of median B% for point observations. All scheme equations and both their published and calibrated parameters are provided in Supporting Information S1.

4. Discussion

This study provides a proof-of-concept demonstration and initial benchmark of a constrained machine-learning parameterization framework for snow albedo evolution, which focuses on deterministic prediction skill and generalization. With only one-time training and a simple model structure, our parameterization improves simulation accuracy by 10%–30% relative to standard schemes. The approach demonstrates transferability across a broad set of contemporary climates, spatial scales, and snow/environmental conditions, at levels unattained by existing analogous schemes. While the different evaluation data do not cover all possible snow environments nor provide inputs to explicitly benchmark across varying snow types, we observe broad generalization across a

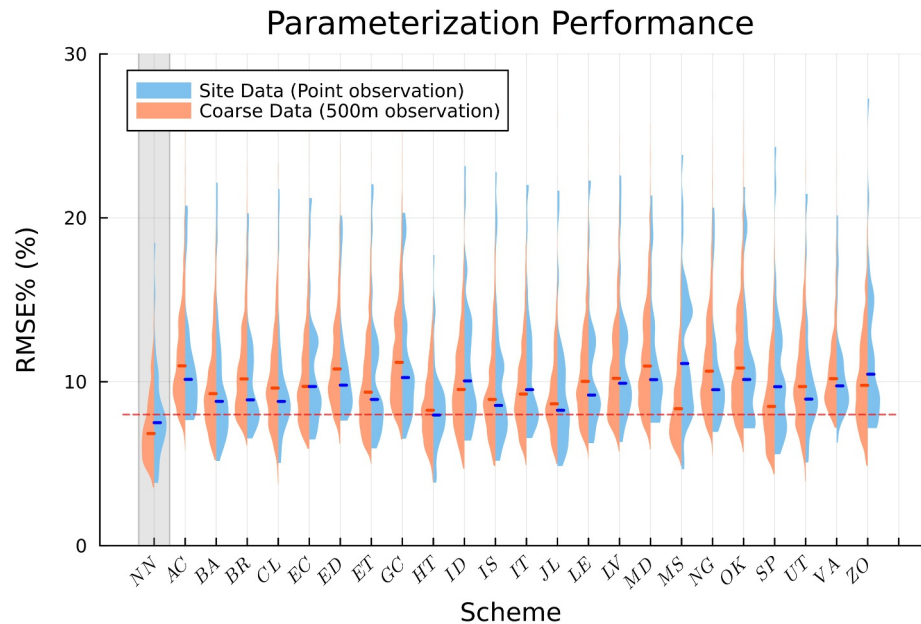


Figure 7. RMSE% scores of all calibrated parameterizations. Format follows that seen in Figure 2, though this time the red dotted line is placed at an RMSE% of 8% instead of 10% to help better distinguish scheme performance. Our scheme remains clearly separated from other schemes.

spread of geographically and climatically distinct regimes and snow states. This performance consistency indicates the approach is extracting relationships broadly consistent with physical expectations from the input environmental and snow state data that can translate across scales and environmental regimes. The results mirror our earlier work developing a similar neural model for snow thickness (Charbonneau et al., 2025), reinforcing the promise of the utilized framework for usage in ever-finer-resolution GCMs.

Despite its simplicity, the network exhibits accuracy and transferability which is beneficial for consistency at global scales and fidelity in local applications. The established generalization by this scheme suggests a robustness against nonstationarity and long-term forecasts as climate regimes shift over time, compared to other

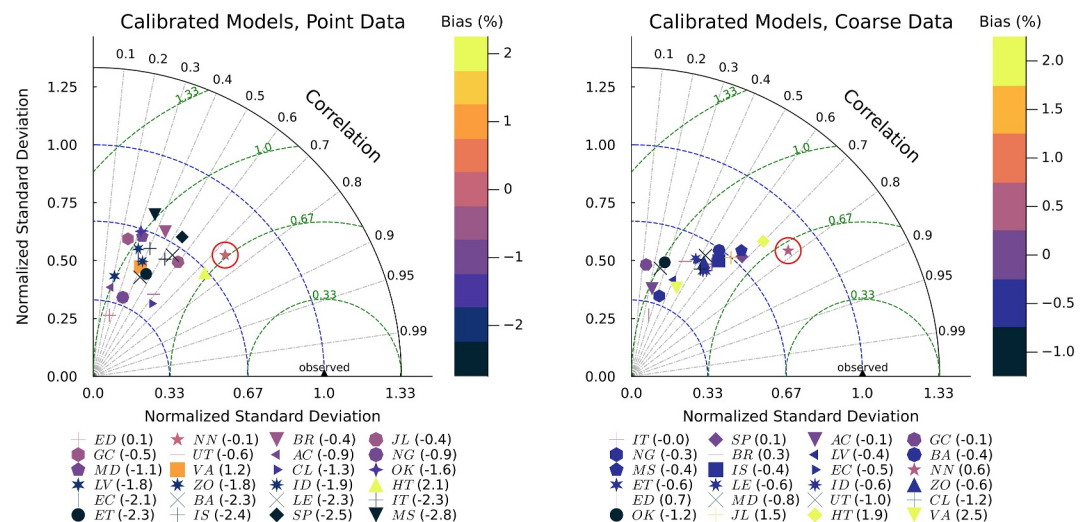


Figure 8. Modified Taylor Diagrams of Calibrated parameterizations. Format follows that seen in Figure 3. Our model still exhibits the highest average correlation and is the closest to the observed values, though the spread is unsurprisingly tighter than in the uncalibrated case. Our scheme also shows a stronger ability than the other models to match the observational variation in albedo, and is still among the better performers with regards to bias, especially at point scales.

parameterizations more sensitive to their calibration regime, though explicit evaluation of extrapolation under future or strongly out-of-distribution conditions remains an important direction for future work. The minimal variable set also decouples the output albedo from compounding or systematic errors in other simulated or uncertain inputs that are necessary in alternative schemes, further aiding robustness. Performance levels are within the 0.05–0.08 RMSE ranges of more complex schemes from radiative-transfer theory like SNICAR (Shao et al., 2020) or full physical models such as Crocus (using the IS scheme) (Alexander et al., 2014; Gaillard et al., 2025) when optimized within a simulation. In real simulation contexts, where coarser scales and uncertain inputs degrade accuracy, these advanced models typically exceed 0.1 RMSE (Hao et al., 2023; Lin et al., 2025; Oaida et al., 2015; Painter et al., 2012), whereas our scheme will not require such intermediate modeling inputs.

The model also demonstrates competitive skill within the ML space. To our knowledge, this is the first application of ML specifically for the parameterization of snow albedo from meteorological inputs, which can be applied within a GCM for future albedo simulation. Previous work like that of Ding et al. (2024) retrieved snow albedo using non-state inputs such as day of year, top-of-atmosphere reflectance values, and satellite band ratios, but also reanalysis-derived albedo, or that of Zhao et al. (2025) which used gradient-boosted trees and other ML models to predict snow surface albedo from satellite surface reflectance values, trained using simulated data from an atmospheric radiative transfer model. Both studies found their designs improved prediction consistency against site-level observations, achieving mean RMSE ~ 0.055 – 0.075 (about 7.5%–10.3%), over 101 worldwide sites or 26 sites in Greenland. However, by relying on reanalysis or satellite inputs, it cannot forecast albedo forward in time from the present within coupled systems, limiting the approach to post-processing instead of prognostic use. Comparable ML efforts (e.g., Chen et al., 2025; Jia et al., 2025; Kiem et al., 2024) and statistical approaches (e.g., Jia et al., 2023; Ye et al., 2023) achieve similar accuracy but rely on geographic indices or predictors such as satellite band ratios, which constrain resolution and applicability and reduce physical interpretability. Our smaller, simpler scheme achieves similar results to these state-of-the-art studies and other ML techniques, while trained on only $\sim 5\%$ of available samples and only meteorological inputs, improving interpretability and long-term applicability. Chevrollier et al. (2025) trained a neural network to emulate SNICAR along with a network inversion technique to additionally infer surface properties from input albedo spectra. Their approach achieved substantial speed-up relative to the full radiative transfer model while diagnosing spectrally resolved albedo (212 bands) with albedo errors below 0.01. However, the emulator needed over 5 million training samples to train 12 layers and 266k parameters, orders of magnitude more parameters than entire GCMs; limiting feasibility for online fine-tuning within GCM ensembles. As with the SNICAR model they emulate, their performance depends on precise input availability (snow grain size, liquid water content, pollutant concentrations, or full 212-band spectra for inversion), requirements unlikely to be met in near-term climate modeling. The reported cost of inversion to obtain such surface properties (~ 0.1 s per grid point) likewise severely constrains scalability for global models. In contrast, our standalone albedo dynamics model is nearly $1000\times$ smaller, enabling online tuning and rapid prototyping, and achieves our results with $100\times$ less data, requiring under 2 μ s per grid point. More notably, it does not inherit the assumptions in an underlying model (e.g., constant snow density and semi-infinite snow depth for the SNICAR emulator), thereby avoiding associated downstream biases.

The model outcomes are constrained by the context, observational uncertainty, and limitations of the available data, as any data-driven model will inherit characteristics of its training data. The SNOTEL network and site-level observations capture localized relationships but are noisy, prone to quality issues, and do not represent large-scale heterogeneity, particularly in complex terrain (Hill et al., 2019; Meyer et al., 2012). The STC/MODSCAG-MODDRFS product likewise suffers terrain as well as cloud effects and is less accurate in forests, partly because its reference spectra exclude burned vegetation. This omission is noteworthy as soot- or fire-derived LAPs increasingly drive abrupt albedo darkening in forested regions, where most SNOTEL sensors reside (Jensen et al., 2024). Both data sources are affected by persistent obfuscation from debris, vegetation, or shadows, and lack coverage in extreme snow environments like tundra and taiga biomes, leaving our model's extension to these domains untested. The utilized SNOTEL sites express a range of possible LAP environments, but are not wholly encompassing, so it is possible that areas with more extreme sand or soot content could yield biased model estimates. Our simulation errors are already reaching the observational uncertainty range of modern data efforts

themselves, underscoring the effectiveness of the modeling paradigm but implying further model gains are unlikely without significant advances in observational quality and availability. Future retraining with improved products is straightforward within the proposed framework.

The tradeoff between smoothing noisy time series and preserving authentic short-lived fluctuations is difficult to avoid. We favor smoothing in our methodology (see Appendix A for more detail), which provides a stable representation of the trends relevant for climate modeling, though this can blur the true input-output relationships; potentially limiting model extension to sub-daily scale variability without retraining. Separability of learnable albedo responses is also restricted by the chosen inputs, which cannot capture LAP concentrations, wildfires (Hatchett et al., 2023), grain size, snow chemistry (Sarangi et al., 2019), or atmospheric state. Aggregating these influences into one broadband albedo, rather than spectral or direct/diffuse components, compounds this challenge, especially near more extreme da/dt . While additional variables could represent more processes, direct measurement of such effects is rare, uncertain, or infeasible at GCM scales (Bair et al., 2018; Günther et al., 2019; Painter et al., 2016). Our framework, however, can readily adapt if such data becomes available.

This work was designed to evaluate predictive performance and transferability across scales for the proposed framework, but does not aim to provide one operational design or formal uncertainty quantification of a finalized model. Several sources of uncertainty are therefore intertwined in the reported results. In addition to the aforementioned observational uncertainty arising from measurement error, quality-control procedure, and retrieval availability, representativeness differences between point and gridded observations can impact the derived results. Structural uncertainty arises from the parameterization itself, including feature selection and omission, network architecture, and temporal resolution. Additional uncertainty may arise from the imposed bounds, which were selected to improve stability and realism while minimizing assumptions, but may contribute to representation errors under extreme conditions and do not inherently prohibit inconsistent behavior between the limits. The distributions of model errors across sites provide an indirect indication of performance variability, but should not be interpreted as formal prediction intervals. Future work could include ensemble, Bayesian, or probabilistic formulations to estimate uncertainty and propagate observational uncertainty through the parameterization, as the framework remains adaptable to such techniques.

The model captures a diverse range of da/dt well but struggles with sharp increases from new snowfall at the site level. This limitation likely reflects data separability issues for extreme da/dt from both the input variable choice (the feature values associated with extreme da/dt are also associated with other non-extreme targets within the spread of data) and the cleaning techniques, as well as the selected L_2 loss, which emphasizes long-term statistics over individual predictions. Prognostic schemes also require an initial albedo value at the onset of a new snowpack, which can vary by environment (Abolafia-Rosenzweig et al., 2022), leading to spin-up effects. However, high-frequency albedo variation at point-scales will be suppressed in the coarse-graining over grid cells within a GCM. While replicating extreme and sub-weekly variability are important for event-level and point-scale studies in the earth sciences, seasonal energy balance and time-integrated radiative effects dominate evolution within an earth system model and long-term simulations, as these govern large-scale feedbacks and equilibrium responses (National Research Council, 1999; Qu & Hall, 2006). This leaves our computationally efficient approach well-suited for GCM integration. A small network structure can approximate a variety of nonlinear functions while maintaining linear computational scalability. This also lowers overfitting risk while evading the rigidity of symbolic functions with even fewer parameters. This contrasts with competing age-based parameterizations, which rely on precise knowledge of snowfall timing that are uncertain at large scales (Bair et al., 2019). The proposed scheme adapts better to inherent observational uncertainty, enables rapid iteration, and can readily incorporate new inputs, motivating broader adoption as observational capacity expands.

We note that behavior consistent with physical expectations of precipitation effects was learned even without choosing boundary functions to explicitly enforce it, illustrating the potential for this approach. However, the absence of explicit physical constraints may increase predictive uncertainty when extrapolating beyond the environmental conditions found in the training and testing data. While the developed scheme demonstrates strong transferability across locations and scales, explicit evaluation under extreme out-of-distribution or future climate states and among features that can quantify radiative transfer effects remains an important avenue for future work. Further experimentation with alternative analytical boundaries could also enforce other desirable behaviors or further increase consistency beyond this initial investigation.

The parameterization performs well at point scales and daily time series from 500 m resolution. Still, these spatial resolutions are finer than the resolutions of current GCMs, where even more coarse-graining or shorter time-steps may introduce additional variability. However, the continuing push for finer-resolution highlights our approach as a candidate for the growing needs of next-generation climate models. Benchmarking accuracy and spatial structure within a full model over contemporaneous spatial grids at global scale or by land type was not assessed here, and is being addressed in an upcoming paper where the prognostic model has been integrated into a new land surface model (Deck et al., 2026) at sub-daily resolution. Other additional avenues for research include other temporal dependencies, validation across global flux tower networks, or adapting the framework for spectrally resolved albedo using alternative MODIS products (e.g., MCD43D) to compare against more complex parameterizations.

5. Conclusion

By developing a simple prognostic parameterization for snow-albedo change that breaks from a continued reliance on snow-age, we simulated seasonal daily snow albedo with a median relative error of under 7.5% across diverse locations (median RMSE of ~ 0.05 at both coarse and point scales). Training over a wide data set creates more generalizable results. Our performance extends from point-scale to coarse-grained data and multiple climate types; outperforming or matching advanced data products, complex physical models, and contemporary ML approaches using only standard meteorological inputs at low computational burden. This reinforces the potential of a constrained data-driven framework that preserves physical consistency while offering accuracy, scalability, and adaptability for next-generation Earth system models. The approach can be used within larger models and can accommodate evolving observational availability. Because snow cover at different altitudes and locations samples situations that may arise in a future, warmer climate (where snowlines generally move to higher altitudes and latitudes), the generalizability shown in this study suggests potential robustness under changing climate conditions, which can be quantified more rigorously in future study. This application of constrained ML for snow albedo modeling shows potential for developing more accurate forecasts and effective mitigation strategies in an evolving climate, and offers a flexible path for integrating ML into physical modeling without compromising interpretability or physical consistency.

Appendix A: Data Appendix

A1. Remote Data

Many remotely sensed snow albedo products exist, with MODIS-derived (~ 500 m, daily) products remaining standard despite the emergence of finer but less frequent (< 250 m) imagery. The STC/MODSCAG-MODDRFS 500-m albedo product (hereafter “SMM”) product (Jensen et al., 2024; Painter et al., 2009, 2012; Rittger et al., 2020, 2024) applies spectral unmixing of MODIS reflectances against simulated radiative-transfer and observational libraries to isolate snow albedo, outperforming empirical products like MOD10A1 that more readily represent the mixed total land surface albedo (Fleming et al., 2024). SMM has been recommended for next-generation model calibration (Fleming et al., 2024), though many studies still rely on outdated products, impairing parameterization accuracy and performance interpretability (Aguirre et al., 2023). SMM ownership recently transferred to Snow Today via the National Snow and Ice Data Center (NSIDC).

Work continues on improving known limitations in SMM data, including pixel geolocation errors (~ 1 pixel), high solar-zenith-angle albedos, vegetation/cloud/topographic interference, spectral library representativeness, and inherent one-to-many ambiguity in the unmixing algorithm (Bair et al., 2019; Demil et al., 2025; Hao et al., 2023; Jensen et al., 2024; Keuris et al., 2023; Rittger et al., 2020). Sensor drift in the Terra/Aqua MODIS sensors also introduces minor bias (Feng et al., 2024), though effects are minimal among the temporal ranges and quality measures present in SMM and this study. NSIDC recently prioritized SPIReS, a newer unmixing algorithm (Rittger et al., 2023), but we used SMM due to broader validation, minor differences in accuracy (Hao et al., 2023; Stillinger et al., 2023), and better availability during much of this study.

A2. In Situ Data

In situ data measure collocated snow and meteorological variables, though coverage and variable selection vary widely, with few stations reporting full GCM-relevant parameters (Raleigh et al., 2015). Example instruments for

automated measurements include ultrasonic depth sensors, snow pillows, and tipping-bucket gauges, which often are noisy, downtime-prone, or yield unphysical values (Fleming et al., 2023; Hill et al., 2019). Some QC standards exist in some networks, such as SNOTEL (e.g., Serreze et al., 1999), though many data streams remain raw.

In prior work establishing our parameterization framework (Charbonneau et al., 2025), established quality control protocols were extended across additional variables and sites. Subsets of these procedures (denoted “CQC” henceforth) were reused/modified in this work for snow albedo and SMM integration. MODIS and SNOTEL products generally agree once snow depth exceeds 5 cm (A. Klein, 2003). While site slope/topography data are more rare, most SNOTEL stations are built on cleared flat ground, reducing the need for additional correction.

A3. Data Cleaning Methods

For each of the 804 SNOTEL sites inside SMM bounds, from 1 Oct 2000 to 1 Oct 2023, daily `albedo_dirty_terrain_corrected`, `snow_fraction`, and `days_without_observation` fields from the corresponding SMM pixel (and 8 surrounding pixels) were paired with SNOTEL snow depth (z), snow water equivalent (SWE), air temperatures (daily average), and accumulated precipitation (AP). Site-years beginning collection after the start of the water-year (October 1st) or sites having under 60 SWE/precipitation values were excluded. SNOTEL values beyond instrument limits (USDA, 2010) were set to missing, and CQC steps 7–13 were applied. Zenith angle (ZA) at solar noon was calculated using `Insolations.jl` (Contributors, 2020) with UTC offsets derived from the free TimeZoneDB API (<https://timezonedb.com/api>), and precipitation was split into rain and snow using the bivariate logistics model from (Jennings et al., 2018) (>85% accuracy):

$$f = \frac{1}{1 + e^{\gamma + \beta T_a}} \quad (\text{A1})$$

where $\beta = 1.24^\circ\text{C}^{-1}$ (temperature in Celsius) and $\gamma = -1.54$. SNOTEL sites with radiation data used the same steps from 1 June 2023 to 1 June 2025, with the additional steps for hourly incoming (SW_{INC}) and outgoing (SW_{OUT}) shortwave radiation instead of the three SMM fields:

1. Set negative radiation values to zero
2. Derive hourly albedo α_i as $SW_{\text{OUT},i}/SW_{\text{INC},i}$
3. Excise all α_i outside [0–1]
4. Excise α_i with timestamp $ZA > 90^\circ$ (using `Insolations.jl`)
5. Aggregate all $\alpha_i \leq 0.4$ to daily level (mean), and estimate α_g as the minimum of 0.1 or the median of all aggregated values having $z = 0$ and $P = 0$
6. Aggregate all $\alpha_i > 0.35$ to daily level (mean) to make α_{snow} time series
7. Excise $\alpha_{\text{snow},i}$ where z_i or SWE_i are 0, missing, or $SWE_i > z_i$
8. Excise $\alpha_{\text{snow},i}$ where $P_i = 0$ but $dSWE/dt_i > 0$ (avoid da/dt samples under indeterminate snowfall)
9. Fill holes of 5 days or less in remaining α_{snow} via linear interpolation
10. Exclude sites with <60 remaining albedo values

For the SMM albedos (0–100), we used the following steps per site:

1. Excise `albedo_dirty_terrain_correctedi` having `days_without_observationi` > 10, according to Rittger et al. (2024).
2. Apply a power mean ($p = 3$) over the remaining 3×3 pixels of `albedo_dirty_terrain_corrected` to mitigate geolocation error, extract the most representative “snowy” estimate (mitigate non-snow or erroneous snow-fraction contamination, with a smoother non-max norm to avoid amplifying high-frequency noise or biasing toward possible cloud albedos, but more inclined to higher values than an arithmetic mean), and create a more complete α_{snow} time series estimate. The count and standard deviation of utilized pixels was tracked.
3. Similarly apply a power mean ($p = -3$, now more inclined to lower values) over the same 3×3 pixels of `albedo_dirty_terrain_corrected`, and estimate α_g as the median of all means ≤ 40 where $z_i = SWE_i = P_i = P_{i-1} = \text{days_without_observation}_i = 0$. Exclude sites lacking any samples for this estimate.

4. Repeat steps 7–10 from above, excising $\alpha_{\text{snow},i} < 35$ as ground-contaminated signal (this choice was corroborated by the standard deviation over utilized 3×3 pixels at such values being < 8 , while snowy pixels exhibited abruptly higher variation)
5. Scale resulting albedo time series by 100 (to [0–1])

Finally, all albedos with $z < 5$ cm were excised to avoid small snowpack discrepancies.

A4. Pre-Training Procedures

Remaining series exhibited unrealistic high-frequency oscillations, especially in situ or during peak season, exceeding expected daily changes (Bair et al., 2018; Sarangi et al., 2019). These could arise from transient shadows or wind-driven debris, local heterogeneity contamination from the above cleaning procedures, or systematic sensor issues requiring further consideration. At both coarse and site scales, persistent non-snow surfaces (rocks, vegetation, shadows, sensor arrays) can also skew true snow fraction even when reported snow cover fraction is 100% at large snowpack, biasing reported albedo downward (Bair et al., 2022; Etchevers et al., 2004). To address these issues, we applied a two-step correction:

A4.1. Exponential Smoothing

An exponential moving average (EMA) was applied as a low-pass filter to dampen noise, extracting the trends relevant for climate modeling scales:

$$EMA_i = w\alpha_i + (1 - w)EMA_{i-1} \quad (\text{A2})$$

Other smoothing approaches exist (e.g., Calleja et al., 2019; Ye et al., 2023), though we prefer the EMA due to simpler implementation, true low-pass filtering compared to techniques like max-filtering which can amplify noise, and better preservation of precipitation-driven albedo changes via weighting bias to recent samples. This approach will smooth out some more abrupt authentic transitions and influence the smoothness of the learned dynamics, but also aids in filling out the otherwise more discrete space of raw da/dt targets due to data precision, which can throttle a neural network's learning ability. We chose $w = 0.33$ for SMM and $w = 0.1$ for noisier in situ data to balance noise suppression against preserving physically relevant trends. Each time the EMA was restarted due to missing/cleaned values, the first 5 values were discarded for sufficient burn-in.

A4.2. Rescaling

A maximum-scaling was applied to correct for persistent non-snow contamination. Measured albedo is a mix of background (α_{bg}) and snow signals weighted by snow fraction, f :

$$\alpha_{obs} = (1 - f_{snow})\alpha_{bg} + f_{snow}\alpha_{snow} \quad (\text{A3})$$

If persistent non-snow contamination prevents full snow coverage ($f_{\text{max}} < 1$) even for peak snowpack in a state of maximum possible albedo ($\alpha_{\text{theory,max}}$), the measured maximum albedo ($\alpha_{\text{obs,max}}$) will be reduced:

$$\alpha_{\text{obs,max}} = (1 - f_{\text{max}})\alpha_{bg} + f_{\text{max}}\alpha_{\text{theory,max}} \quad (\text{A4})$$

Suitable choice of $\alpha_{\text{theory,max}}$ and α_{bg} knowledge thus allows estimation of f_{max} and a scaling transformation to rescale albedos:

$$f_{\text{max}} = \frac{\alpha_{\text{obs,max}} - \alpha_{bg}}{\alpha_{\text{theory,max}} - \alpha_{bg}}, \quad \alpha_i = \frac{\alpha_{\text{obs},i} - \alpha_{bg}(1 - f_{\text{max}})}{f_{\text{max}}} \quad (\text{A5})$$

Although the best $\alpha_{\text{theory,max}}$ (i.e., fresh pure snowfall on large snowpack) choice could fluctuate by site, it should exhibit tight range across climates and be well-represented within multi-year records, enabling estimation of f_{max} . The top 5% of all in situ sites' cleaned albedos were collected, using the mean of these values to estimate

$\alpha_{\text{theory,max}} = 0.88$, which aligns with theoretical maximums of broadband albedo (Dozier et al., 2009). This rescaling serves to reduce non-snow contamination and potential unmeasured slope effects, and recorrects EMA-induced downward bias while also leaving series with $\alpha_{\text{obs,max}} \approx \alpha_{\text{theory,max}}$ largely unchanged. $\alpha_{\text{obs,max}}$ was estimated with the 95th percentile of each site's cleaned and smoothed albedo measurements, removing and interpolating any $\alpha_{\text{snow}} > 1$ after the transformation.

Daily da/dt was then computed as $\frac{\alpha_{t+1} - \alpha_t}{t_{t+1} - t_t}$, retaining only consecutive days (when $t_{t+1} - t_t = 1$ day). The final training set excluded cases with `snow_fraction` < 0.05 (the mean over the available 3×3 pixels) and $z/\text{SWE} = 1$, though these were kept in testing data. Sites offering less than 30 training samples were excluded from the study.

For non-SNOTEL sites (excluding El Tapado and CUES), non-albedo cleaning is described in CQC, and albedo cleaning followed the steps for SNOTEL above. Col de Porte already gives daily albedo, requiring steps 5, 7, 8, and 9 only. For CUES, relative humidity was capped at 1, only albedo steps 4, 6, 8, and 9 were used, and snowfall was estimated using the Zwart parameterization (Zwart & Fierz, 2007) from positive changes in z , in accordance with a site representative. No ground albedo was determined but $\alpha_{\text{obs,max}}$ was sufficiently high to only apply the EMA. For El Tapado, the following steps were used:

1. Excise negative z , SWE, and AP ; $AP > 1$ m, $T_a > 20^\circ\text{C}$ (sensor noise)
2. Use only 2017 and 2019 water years from availability and quality concerns (per site correspondence)
3. Fill T_a in 2017 via lapse-rate interpolation, in accordance with Voordendag et al. (2021)
4. Calculate changes in AP , setting values < 0.5 mm or negative to zero, then reconstruct cumulative AP from these differences.
5. Apply all SNOTEL albedo steps and CQC steps 9–13

The feature set was largely driven by near-term availability, computational simplicity, and quality considerations. Although rain can affect albedo (Xie et al., 2024), including P_{rain} had negligible impact on outcome (as in our prior study) and was omitted for computational simplicity. A 12-hr offset exists between SNOTEL precipitation (local midnight) and SMM albedo (roughly local noon) measurements, which can blur interpretation of albedo changes and relationship “separability.” However, the EMA already coarse-grains changes and additional methods would likely further blur these relationships, so no further adjustment was applied.

Appendix B: Performance Statistics

Table B1
Parameterization Scores Over As-Published Parameterizations

Model	Coarse testing data (500 m)				Significance				Site testing data (Point observation)				Significance	
	RMSE%	RMSE	Bias%	Bias	RMSE%	RMSE	Bias%	Bias	RMSE%	RMSE	Bias%	Bias	RMSE%	RMSE
<i>NN</i>	6.84 (7.29)	0.0494 (0.0509)	0.46 (0.57)	0.0059 (0.0079)	—	—	—	—	7.51 (8.05)	0.0555 (0.0584)	0.48 (−0.12)	0.0076 (0.0038)	—	—
<i>CL</i>	14.83 (14.9)	0.1095 (0.1089)	2.93 (2.85)	0.0282 (0.0288)	2.87E−116	2.87E−116	2.15 (2.13)	0.0199 (0.0223)	14.03 (14.43)	0.1050 (0.1061)	2.15 (2.13)	0.0199 (0.0223)	9.31E−10	1.86E−09
<i>OK</i>	13.02 (13.56)	0.0929 (0.0959)	−0.26 (−0.02)	0.0042 (0.0075)	2.88E−116	2.96E−116	12.05 (13.3)	0.0872 (0.0945)	12.05 (13.3)	0.0872 (0.0945)	−0.46 (0.17)	−0.0004 (0.0067)	6.52E−09	1.3E−8
<i>LV</i>	24.64 (24.84)	0.1905 (0.1918)	18.27 (18.04)	0.1433 (0.1429)	2.87E−116	2.87E−116	22.28 (23.82)	0.1706 (0.1843)	22.28 (23.82)	0.1706 (0.1843)	15.06 (15.73)	0.1192 (0.1259)	9.31E−10	9.31E−10
<i>MD</i>	12.27 (12.88)	0.0858 (0.0870)	−2.4 (−2.71)	−0.0109 (−0.0126)	6.06E−116	1.02E−115	11.67 (12.91)	0.0855 (0.0887)	11.67 (12.91)	0.0855 (0.0887)	−2.11 (−2.84)	−0.0104 (−0.0144)	9.31E−10	1.86E−09
<i>JL</i>	11.06 (11.59)	0.0732 (0.0741)	−4.44 (−4.68)	−0.0276 (−0.0283)	1.3E−111	1.51E−104	10.56 (12.15)	0.0723 (0.0800)	10.56 (12.15)	0.0723 (0.0800)	−4.77 (−5.99)	−0.0311 (−0.0376)	6.52E−08	4.0E−6
<i>SP</i>	29.29 (29.91)	0.2082 (0.2157)	1.68 (2.97)	0.0046 (0.015)	2.87E−116	2.87E−116	29.26 (29.85)	0.2123 (0.2167)	29.26 (29.85)	0.2123 (0.2167)	−8.52 (−6.59)	−0.0761 (−0.0509)	9.31E−10	9.31E−10
<i>MS</i>	14.21 (14.25)	0.1097 (0.1093)	7.39 (7.11)	0.0634 (0.0612)	2.87E−116	2.87E−116	13.5 (14.11)	0.1041 (0.1081)	13.5 (14.11)	0.1041 (0.1081)	6.61 (5.68)	0.0564 (0.0503)	9.31E−10	9.31E−10
<i>UT</i>	17.14 (17.35)	0.1261 (0.1294)	4.95 (5.24)	0.0435 (0.0462)	2.87E−116	2.87E−116	16.39 (17.02)	0.1183 (0.1256)	16.39 (17.02)	0.1183 (0.1256)	3.28 (4.62)	0.0306 (0.0398)	9.31E−10	9.31E−10
<i>HT</i>	13.57 (13.92)	0.0986 (0.1011)	3.7 (3.96)	0.0309 (0.0339)	2.87E−116	2.87E−116	13.94 (13.76)	0.0985 (0.1011)	13.94 (13.76)	0.0985 (0.1011)	2.27 (2.39)	0.0205 (0.022)	9.31E−10	2.79E−09
<i>EC</i>	15.98 (16.32)	0.1147 (0.1157)	−1.71 (−1.74)	−0.0105 (−0.0089)	2.87E−116	2.87E−116	15.71 (16.4)	0.1154 (0.1163)	15.71 (16.4)	0.1154 (0.1163)	−3.2 (−3.33)	−0.0227 (−0.0204)	9.31E−10	1.86E−09
<i>ZO</i>	28.9 (29.91)	0.2267 (0.2329)	27.18 (27.91)	0.2138 (0.2178)	2.87E−116	2.87E−116	29.29 (29.45)	0.2348 (0.2301)	29.29 (29.45)	0.2348 (0.2301)	26.75 (25.9)	0.2113 (0.2045)	9.31E−10	9.31E−10
<i>IT</i>	39.17 (39.09)	0.3059 (0.3063)	37.87 (37.83)	0.2946 (0.2948)	2.87E−116	2.87E−116	35.74 (36.26)	0.2816 (0.2851)	35.74 (36.26)	0.2816 (0.2851)	33.02 (34.28)	0.2582 (0.2673)	9.31E−10	9.31E−10
<i>ID</i>	13.2 (13.56)	0.0859 (0.0865)	−7.99 (−8.0)	−0.0545 (−0.0534)	4.46E−115	1.02E−112	12.19 (14.22)	0.0874 (0.0939)	12.19 (14.22)	0.0874 (0.0939)	−8.33 (−8.87)	−0.06 (−0.0594)	1.86E−09	1.77E−08
<i>BA</i>	11.32 (11.69)	0.0778 (0.0795)	−0.24 (−0.29)	0.0029 (0.0037)	2.87E−116	2.88E−116	11.43 (11.72)	0.0813 (0.0810)	11.43 (11.72)	0.0813 (0.0810)	−1.42 (−1.77)	−0.0058 (−0.007)	1.3E−8	2.86E−7
<i>VA</i>	21.55 (23.13)	0.1588 (0.1740)	4.46 (6.62)	0.0371 (0.0355)	2.87E−116	2.87E−116	24.28 (26.5)	0.1776 (0.1900)	24.28 (26.5)	0.1776 (0.1900)	4.1 (6.83)	0.0321 (0.0525)	9.31E−10	9.31E−10
<i>LE</i>	15.31 (16.07)	0.1047 (0.1101)	2.44 (2.62)	0.0228 (0.0245)	2.87E−116	2.87E−116	12.4 (13.27)	0.0846 (0.0898)	12.4 (13.27)	0.0846 (0.0898)	0.32 (0.06)	0.0073 (0.0054)	6.52E−08	2.36E−7
<i>AC</i>	28.02 (27.89)	0.2192 (0.2175)	20.22 (19.78)	0.1612 (0.1577)	2.87E−116	2.87E−116	26.09 (27.44)	0.2040 (0.2124)	26.09 (27.44)	0.2040 (0.2124)	17.16 (19.07)	0.1357 (0.1507)	9.31E−10	9.31E−10
<i>GC</i>	16.4 (16.67)	0.1065 (0.1065)	−10.73 (−10.88)	−0.0734 (−0.0737)	3.88E−116	7.07E−116	16.22 (16.68)	0.1062 (0.1090)	16.22 (16.68)	0.1062 (0.1090)	−10.97 (−11.43)	−0.0746 (−0.0776)	1.86E−09	2.79E−09
<i>NG</i>	15.84 (15.8)	0.1178 (0.1181)	5.31 (5.24)	0.0464 (0.0464)	2.87E−116	2.87E−116	15.46 (15.59)	0.1130 (0.1169)	15.46 (15.59)	0.1130 (0.1169)	4.25 (4.96)	0.0375 (0.0434)	9.31E−10	9.31E−10
<i>ED</i>	27.88 (27.77)	0.2167 (0.2159)	18.42 (17.93)	0.1464 (0.1433)	2.87E−116	2.87E−116	26.3 (27.38)	0.2022 (0.2111)	26.3 (27.38)	0.2022 (0.2111)	15.35 (17.08)	0.1201 (0.1353)	9.31E−10	9.31E−10
<i>ET</i>	23.42 (23.43)	0.1722 (0.1748)	12.04 (12.47)	0.0935 (0.0972)	2.87E−116	2.87E−116	21.24 (22.2)	0.1625 (0.1656)	21.24 (22.2)	0.1625 (0.1656)	7.66 (9.57)	0.0643 (0.0759)	9.31E−10	9.31E−10
<i>BR</i>	25.78 (26.08)	0.2028 (0.2054)	22.91 (23.06)	0.1808 (0.1825)	2.87E−116	2.87E−116	24.4 (25.07)	0.1939 (0.1961)	24.4 (25.07)	0.1939 (0.1961)	21.86 (21.97)	0.1741 (0.1724)	1.86E−09	9.31E−10
<i>IS</i>	11.46 (11.91)	0.0756 (0.0773)	−4.24 (−4.35)	−0.0255 (−0.0255)	4.02E−113	4.2E−110	10.73 (11.99)	0.0751 (0.0796)	10.73 (11.99)	0.0751 (0.0796)	−4.38 (−5.24)	−0.028 (−0.0322)	1.57E−07	7.01E−6

Note. Performance of the different schemes in this study across both validation and testing sites, denoting models using their labels provided in Table 2 in the main text. Medians are listed, with the mean in parentheses alongside the median, for each of the performance metrics discussed in the text. For both coarse and point-level sites, each scheme also lists the p -value from the Wilcoxon signed-rank tests in this study, comparing each scheme to our proposed scheme (VN). Significant p -values (<0.05) are listed in bold, as well as the top scheme within each metric, unless no clear winner was determinable. Our scheme is one of the top performers for most metrics, and all statistical tests are significant ($p > 0.05$) comparing the proposed scheme with regards to RMSE and RMSE%. Scores are also similar between coarse and point-level sites for our model, indicating generalizability across spatial scales.

Table B2
Parameterization Scores Over Calibrated Parameterizations

Model	Coarse testing data (500 m)				Significance				Site testing data (Point observation)				Significance	
	RMSE%	RMSE	Bias%	Bias	RMSE%	RMSE	Bias%	Bias	RMSE%	RMSE	Bias%	Bias	RMSE%	RMSE
<i>NN</i>	6.84 (7.29)	0.0494 (0.0509)	0.46 (0.57)	0.0059 (0.0079)	—	—	—	—	7.51 (8.05)	0.0555 (0.0584)	0.48 (−0.12)	0.0076 (0.0038)	—	—
<i>CL</i>	9.62 (10.1)	0.0674 (0.0680)	−1.14 (−1.23)	−0.0029 (−0.0024)	8.03E−109	4.46E−104	—	—	8.8 (9.83)	0.0646 (0.0677)	−0.7 (−1.34)	−0.0017 (−0.0031)	3.15E−04	4.44E−03
<i>OK</i>	10.84 (11.52)	0.0750 (0.0770)	−0.94 (−1.24)	0.0003 (−0.0005)	1.46E−114	7.17E−113	—	—	10.14 (11.17)	0.0723 (0.0770)	−0.77 (−1.62)	0.001 (−0.0039)	1.93E−07	4.99E−07
<i>LV</i>	10.21 (10.85)	0.0720 (0.0737)	−0.28 (−0.41)	0.0045 (0.0049)	1.51E−114	1.26E−112	—	—	9.91 (11.47)	0.0758 (0.0792)	−0.68 (−1.77)	0.0001 (−0.0049)	4.00E−08	1.57E−07
<i>MD</i>	10.96 (11.55)	0.0760 (0.0777)	−0.46 (−0.79)	0.0037 (0.0031)	4.68E−115	6.86E−114	—	—	10.13 (11.43)	0.0722 (0.0790)	−0.2 (−1.11)	0.005 (−0.0001)	5.12E−08	1.02E−07
<i>JL</i>	8.65 (9.26)	0.0639 (0.0653)	1.65 (1.57)	0.0171 (0.0179)	2.32E−109	2.39E−112	—	—	8.26 (9.43)	0.0639 (0.0670)	−0.11 (−0.44)	0.0064 (0.0032)	1.04E−02	9.15E−03
<i>SP</i>	8.49 (9.01)	0.0596 (0.0609)	−0.02 (0.08)	0.0047 (0.0062)	3.66E−107	1.43E−96	—	—	9.7 (10.56)	0.0673 (0.0732)	−1.99 (−2.49)	−0.012 (−0.0122)	1.19E−05	2.82E−05
<i>MS</i>	8.36 (9.11)	0.0591 (0.0612)	−0.29 (−0.44)	0.0023 (0.0022)	6.45E−107	1.63E−92	—	—	11.11 (11.89)	0.0815 (0.0831)	−3.22 (−2.77)	−0.0167 (−0.0138)	1.28E−07	4.16E−07
<i>UT</i>	9.71 (10.16)	0.0682 (0.0691)	−0.99 (−1.01)	−0.0013 (−0.001)	1.38E−111	4.58E−109	—	—	8.94 (10.11)	0.0635 (0.0707)	0.05 (−0.58)	0.0034 (0.0025)	1.18E−04	1.72E−03
<i>HT</i>	8.26 (8.54)	0.0599 (0.0617)	1.63 (1.89)	0.016 (0.0192)	1.18E−87	1.79E−98	—	—	7.97 (8.5)	0.0633 (0.0638)	3.36 (2.06)	0.0283 (0.0209)	2.39E−01	6.64E−02
<i>EC</i>	9.72 (10.43)	0.0680 (0.0703)	−0.35 (−0.46)	0.0038 (0.0036)	2.34E−115	1.25E−114	—	—	9.71 (10.9)	0.0692 (0.0750)	−1.18 (−2.14)	−0.0043 (−0.0086)	1.57E−07	2.57E−06
<i>ZO</i>	9.78 (10.26)	0.0666 (0.0686)	−0.77 (−0.62)	0.0008 (0.0025)	8.1E−114	6.06E−110	—	—	10.46 (11.7)	0.0762 (0.0810)	−0.73 (−1.79)	−0.0003 (−0.0054)	4.99E−07	3.46E−06
<i>IT</i>	9.25 (9.81)	0.0657 (0.0666)	0.03 (−0.01)	0.0058 (0.0066)	7.01E−115	2.82E−113	—	—	9.52 (10.96)	0.0693 (0.0762)	−2.06 (−2.33)	−0.0096 (−0.0102)	2.86E−07	2.57E−06
<i>ID</i>	9.53 (9.99)	0.0648 (0.0668)	−0.7 (−0.57)	0.0003 (0.0024)	2.99E−112	1.09E−106	—	—	10.05 (11.26)	0.0734 (0.0783)	−1.12 (−1.93)	−0.0012 (−0.0067)	1.38E−06	1.35E−05
<i>BA</i>	9.28 (9.92)	0.0656 (0.0674)	−0.28 (−0.41)	0.0029 (0.0032)	2.53E−112	8.42E−113	—	—	8.8 (9.95)	0.0617 (0.0684)	−1.29 (−2.31)	−0.0041 (−0.011)	3.15E−04	1.47E−03
<i>VA</i>	10.19 (10.72)	0.0754 (0.0768)	2.67 (2.47)	0.0276 (0.027)	4.32E−116	4.33E−116	—	—	9.75 (10.83)	0.0725 (0.0782)	1.92 (1.19)	0.0214 (0.0173)	2.21E−06	3.46E−06
<i>LE</i>	10.02 (10.62)	0.0693 (0.0714)	−0.47 (−0.57)	0.0028 (0.0028)	1.36E−115	6.25E−115	—	—	9.19 (10.83)	0.0655 (0.0735)	−1.47 (−2.33)	−0.0056 (−0.01)	1.17E−06	3.57E−05
<i>CA</i>	10.97 (11.64)	0.0782 (0.0796)	0.2 (−0.12)	0.0093 (0.0081)	1.37E−115	5.74E−115	—	—	10.14 (11.21)	0.0732 (0.0781)	0.13 (−0.88)	0.0081 (0.002)	1.93E−07	5.96E−07
<i>GC</i>	11.18 (11.75)	0.0778 (0.0801)	0.01 (−0.13)	0.0091 (0.0086)	7.01E−116	2.09E−115	—	—	10.26 (11.33)	0.0737 (0.0795)	0.54 (−0.48)	0.008 (0.0054)	8.42E−07	1.90E−06
<i>NG</i>	10.65 (11.27)	0.0755 (0.0768)	0.09 (−0.27)	0.0084 (0.0067)	1.12E−114	1.11E−113	—	—	9.52 (10.66)	0.0693 (0.0738)	0.29 (−0.89)	0.008 (0.0016)	2.99E−06	8.02E−06
<i>ED</i>	10.78 (11.39)	0.0776 (0.0789)	1.03 (0.73)	0.0156 (0.0147)	2.16E−115	5.09E−115	—	—	9.79 (10.92)	0.0722 (0.0771)	1.02 (0.09)	0.0139 (0.0095)	7.10E−07	1.17E−06
<i>ET</i>	9.37 (10.0)	0.0648 (0.0669)	−0.46 (−0.55)	0.003 (0.0027)	8.56E−114	5.1E−111	—	—	8.92 (10.5)	0.0639 (0.0716)	−1.55 (−2.34)	−0.0057 (−0.0103)	9.16E−06	2.15E−04
<i>BR</i>	10.17 (10.81)	0.0730 (0.0741)	0.58 (0.33)	0.0115 (0.0107)	1.77E−115	5.55E−115	—	—	8.9 (10.27)	0.0662 (0.0727)	0.97 (−0.35)	0.0134 (0.0048)	4.00E−06	6.11E−06
<i>IS</i>	8.92 (9.62)	0.0634 (0.0649)	−0.33 (−0.4)	0.0026 (0.0032)	4.54E−109	1.95E−108	—	—	8.56 (10.07)	0.0641 (0.0694)	−1.86 (−2.39)	−0.0077 (−0.0114)	2.15E−04	1.59E−03

Note. Labeling and formatting follows that as described in Table B1. In this case, the proposed scheme is still dominant in many metrics. Results of statistical tests are again all significant ($p > 0.05$) with the exceptions of *HT* over point-level data.

Table C1
Feature Interaction Strength

H-statistic	
P, μ	0.13
P, α	0.11
P, T_a	0.12
μ, T_a	0.18
α, T_a	0.04
α, μ	0.09

Note. This table displays the pairwise Friedman H-statistics of each feature pair in the neural network. No value exceeds 0.2, implying minimal higher-order effects compared to the additive effects diagnosed from Figure 6.

Appendix C: Interaction Strength

The pairwise Friedman H-statistic (Molnar, 2025) quantifies the strength of an interaction between two model features. The square of the statistic can be interpreted as quantifying what fraction of the prediction's variance comes from solely the interaction effect between two features rather than their additive main effects. For this network, all values are less than 0.2, implying weak pairwise interactions (less than 5% of the prediction variance stems from interaction effects). This means the overall prediction effects can be summarized as primarily being additive in the main single-variable effects (those assessed in the ALE in Figure 6), instead of also including strong higher-order effects. The reported statistics remain stable between different binning sizes as well as under consideration of the feature correlations, implying the results are faithful for the developed scheme and robust against numerical artifacts that can arise for a particular calculation approach.

Appendix D: Hyperparameters

Similar to our previous work (Charbonneau et al., 2025), we made use of a custom loss-function:

$$L = \frac{1}{N_d} \sum_{i=1}^{N_d} w_i |y_i - \hat{y}_i|^{n_1}, \quad w_i = 1 + |y_i|^{n_2}. \quad (\text{D1})$$

Table D1
Hyperparameter Results

Parameter	Description	Range	Best value
n	Width of Mixing Layer	2, 3, 4, 5	2
n_1	Power Scaling of Prediction Error	1, 2, 3	2
n_2	Power Scaling of Target Magnitude	0, 0.1, 1, 2	0

Note. Timeseries RMSE% and B% were used as the primary metrics. $n_1 = 2$ (L_2 -like metric) provided the lowest RMSE% for all choices.

where N_d is the number of batched training data, $\hat{y}_i$ and y_i are the (scaled) prediction and target data, and n_1 and n_2 are constant positive numbers. Using ($n_1 = 1, n_2 = 0$) or ($n_1 = 2, n_2 = 0$) is equivalent to optimizing the average L_1 or L_2 losses, respectively. This function enables investigation of emphasizing extremes, which can be important in prognostic (accumulated) modeling since undershooting one extreme da/dt can propagate the offset to future time steps. Optimal hyperparameters and the assessed values are summarized in Table C1, for n_1, n_2 , and layer width hyperparameter n . Scores for each configuration were evaluated using leave-one-out validation (training on 39 of the 40 sites, tracking the time series RMSE% and bias of the remaining site, and averaging over the 40 possible combinations) with a batch size of 256, tracking performance every 10 epochs over 200 epochs. Most time series trials performed optimally when training for approximately 100 epochs with the small network, similar to our previous findings. Table D1.

The configuration of $n = 2, n_1 = 2$, and $n_2 = 0$ showed optimal performance for both time series RMSE% and bias (B). As $n_2 = 0$ matches the L_2 loss function, true L_2 was used for final model training. We note $n_2 \neq 0$ and/or $n_1 = 1$ did often exhibit time series that better replicated sub-weekly variation, but at the expense of whole timeseries offsets that skewed overall performance statistics (which would similarly skew the energy balance within a GCM) to favor $n_2 = 0$. As L_2 loss is associated with RMSE, it is less surprising that $n_1 = 2$ exhibited optimal RMSE%, though optimal $n_2 = 0$ differs from our previous study. This optimal configuration is likely influenced by the relative lack of variation in the coarse-scale (and cleaned) albedo data used for training, compared to the testing site-level data with more variation or the site-level snow depth data in our previous work.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

All utilized data are publicly available. The SNOTEL data utilized in this study are available via the National Water and Climate Center. SNOTEL Data reports were generated using the online portal found at <https://wcc.sc.egov.usda.gov/reportGenerator/>. Links to raw and processed data for Col De Porte (Lejeune et al., 2019), Kühtai (Krajčič et al., 2017), Sodankylä (Essery et al., 2016), Yala (Shea et al., 2015; Stigter et al., 2021), and Upper Rofental (Warscher et al., 2024) data can be found in our previous work (Charbonneau et al., 2025). SMM data can be found at the Snow Today website, hosted by the National Snow and Ice Data Center (NSIDC) https://nsidc.org/data/stc_modscgdrf_hist/versions/1 (Rittger et al., 2024). El Tapado (station code “TPF”) raw data was retrieved from the CEAZA meteorological network data portal at https://www.ceazamet.cl/index.php?e_cod=TPF&pag=mod_estacion, and CUES data can be found for download at <https://snow.ucsb.edu/index.php/level-2-model-ready> (Bair et al., 2018). The *Insolations.jl* package is developed by the Climate Modeling Alliance (CliMA), able to be downloaded at <https://clima.github.io/Insolation.jl/dev/> (Contributors, 2020), while the API for time-zone retrieval can be found at <https://timezonedb.com/api>. As in our previous work, code for scraping and cleaning data, CSVs of data, and tutorials for data retrieval and modifying our (or building custom) constrained neural networks are available at https://clima.github.io/ClimaLand.jl/dev/generated/standalone/Snow/base_tutorial/, to provide quality-controlled data sets for calibration and ML applications. All data cleaning procedures are described above in Appendix A. Time-stepping is described in the main text. Descriptions and equations of the compared models are provided in Supporting Information S1.

Acknowledgments

We thank Edward Bair and Annelies Voordendag for their correspondence regarding CUES and El Tapado data, respectively, and for Cristian Nelson for correspondence about El Tapado data. We thank the Col de Porte, Kühtai, Sodankylä, Upper Rofental, CUES, CEAZA, and Yala Basecamp teams for their data. A. C. was supported by the AI4Science initiative at the California Institute for Technology and a Department of Defense National Defense Science and Engineering Graduate (NDSEG) Fellowship. All authors were generously supported by Schmidt Sciences, L.L.C. and by the Resnick Sustainability Institute.

References

- Abolafia-Rosenzweig, R., He, C., McKenzie Skiles, S., Chen, F., & Gochis, D. (2022). Evaluation and optimization of snow albedo scheme in noah-mp land surface model using in situ spectral observations in the Colorado rockies. *Journal of Advances in Modeling Earth Systems*, 14(10), e2022MS003141. <https://doi.org/10.1029/2022ms003141>
- ACE. (1956). Snow hydrology: Summary report of the snow investigations. *Journal of Glaciology*, 3(22), 148. <https://doi.org/10.3189/s002214300024503>
- Aguirre, F., Bozkurt, D., Sauter, T., Carrasco, J., Schneider, C., Jaña, R., & Casassa, G. (2023). Snow cover reconstruction in the brunswick peninsula, Patagonia, derived from a combination of the spectral fusion, mixture analysis, and temporal interpolation of modis data. *Remote Sensing*, 15(22), 5430. <https://doi.org/10.3390/rs15225430>
- Alexander, P. M., Tedesco, M., Fettweis, X., van de Wal, R. S. W., Smeets, C. J. P. P., & van den Broeke, M. R. (2014). Assessing spatio-temporal variability and trends in modelled and measured Greenland ice sheet albedo (2000–2013). *The Cryosphere*, 8(6), 2293–2312. <https://doi.org/10.5194/tc-8-2293-2014>
- Anderson, E. A. (1968). Development and testing of snow pack energy balance equations. *Water Resources Research*, 4(1), 19–37. <https://doi.org/10.1029/wr004001p00019>
- Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(4), 1059–1086. <https://doi.org/10.1111/rssb.12377>
- Atwood, J., Domonkos, B., Hill, K., Brosten, T., DeMarco, T., Hultstrand, M., et al. (2023). Evaluation of ysi temperature correction equations for bias-reducing snotel network temperature data. In *2023 western snow conference*. Retrieved from https://www.nrcs.usda.gov/sites/default/files/2023-05/Final_Temperature_Correction_Study05262023.pdf
- Bair, E. H., Davis, R. E., & Dozier, J. (2018). Hourly mass and snow energy balance measurements from Mammoth mountain, ca usa, 2011–2017. *Earth System Science Data*, 10(1), 549–563. <https://doi.org/10.5194/essd-10-549-2018>
- Bair, E. H., Dozier, J., Stern, C., LeWinter, A., Rittger, K., Savagian, A., et al. (2022). Divergence of apparent and intrinsic snow albedo over a season at a sub-alpine site with implications for remote sensing. *The Cryosphere*, 16(5), 1765–1778. <https://doi.org/10.5194/tc-16-1765-2022>
- Bair, E. H., Rittger, K., Skiles, S. M., & Dozier, J. (2019). An examination of snow albedo estimates from modis and their impact on snow water equivalent reconstruction. *Water Resources Research*, 55(9), 7826–7842. <https://doi.org/10.1029/2019wr024810>
- Baker, D. G., Ruschy, D. L., & Wall, D. B. (1990). The albedo decay of prairie snows. *Journal of Applied Meteorology*, 29(2), 179–187. [https://doi.org/10.1175/1520-0450\(1990\)029<0179:tadops>2.0.co;2](https://doi.org/10.1175/1520-0450(1990)029<0179:tadops>2.0.co;2)
- Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., et al. (2011). The joint UK land environment simulator (jules), model description – Part 1: Energy and water fluxes. *Geoscientific Model Development*, 4(3), 677–699. <https://doi.org/10.5194/gmd-4-677-2011>
- Brock, B. W., Willis, I. C., & Sharp, M. J. (2000). Measurement and parameterization of albedo variations at haut glacier d’arolla, Switzerland. *Journal of Glaciology*, 46(155), 675–688. <https://doi.org/10.3189/172756500781832675>
- Brun, E., David, P., Sudul, M., & Brunot, G. (1992). A numerical model to simulate snow-cover stratigraphy for operational avalanche forecasting. *Journal of Glaciology*, 38(128), 13–22. <https://doi.org/10.3189/s0022143000009552>
- Calleja, J. F., Corbea-Perez, A., Fernandez, S., Recondo, C., Peon, J., & de Pablo, M. A. (2019). Snow albedo seasonality and trend from modis sensor and ground data at johnsons glacier, livingston island, maritime Antarctica. *Sensors*, 19(16), 3569. <https://doi.org/10.3390/s19163569>
- Cassardo, C. (2015). Utopia: The manual of version 2015. 8594. <https://doi.org/10.13140/RG.2.2.29664.38404>
- Chan, H. G. (2022). Parameterization of snowpack albedo reduction by light-absorbing impurities. Geophysical Fluid Dynamics Laboratory. <https://doi.org/10.25923/527J-0A46>

- Charbonneau, A., Deck, K., & Schneider, T. (2025). A physics-constrained neural differential equation framework for data-driven snowpack simulation. *Artificial Intelligence for the Earth Systems*, 4(3). <https://doi.org/10.1175/aies-d-24-0040.1>
- Chen, S., Xiao, P., Zhang, X., Pellikka, P., Liu, H., & Liu, Y. (2025). Mapping snow-covered forest albedo via hybrid radiative transfer and machine learning across northern hemisphere. *ISPRS Journal of Photogrammetry and Remote Sensing*, 228, 489–504. <https://doi.org/10.1016/j.isprsjprs.2025.07.033>
- Chevrollier, L.-A., Wehrli, A., Cook, J. M., Pirk, N., Benning, L. G., Anesio, A. M., & Tranter, M. (2025). Separating the albedo-reducing effect of different light-absorbing particles on snow using deep learning. *The Cryosphere*, 19(4), 1527–1538. <https://doi.org/10.5194/tc-19-1527-2025>
- Contributors, C. (2020). Insolation.jl. Retrieved from <https://github.com/CliMA/Insolation.jlGitHubrepository.Apache-2.0license>
- Cordero, R. R., Sepúlveda, E., Feron, S., Wang, C., Damiani, A., Fernandez, F., et al. (2022). Black carbon in the southern andean snowpack. *Environmental Research Letters*, 17(4), 044042. <https://doi.org/10.1088/1748-9326/ac5df0>
- Damiani, A., Ishizaki, N. N., Feron, S., & Cordero, R. R. (2025). Projecting future snow changes at kilometer scale for adaptation using machine learning and a cmip6 multi-model ensemble. *Science of The Total Environment*, 964, 178606. <https://doi.org/10.1016/j.scitotenv.2025.178606>
- Decharme, B., Brun, E., Boone, A., Delire, C., Le Moigne, P., & Morin, S. (2016). Impacts of snow and organic soils parameterization on northern Eurasian soil temperature profiles simulated by the isba land surface model. *The Cryosphere*, 10(2), 853–877. <https://doi.org/10.5194/tc-10-853-2016>
- Deck, K., Braghieri, R. K., Renchon, A. A., Sloan, J., Bozzola, G., Speer, E., et al. (2026). Climland: A land surface model designed to enable data-driven parameterizations. *Journal of Advances in Modeling Earth Systems*, 18(1), e2025MS005118. <https://doi.org/10.1029/2025MS005118>
- De Michele, C., Avanzi, F., Ghezzi, A., & Jommi, C. (2013). Investigating the dynamics of bulk snow density in dry and wet conditions using a one-dimensional model. *The Cryosphere*, 7(2), 433–444. <https://doi.org/10.5194/tc-7-433-2013>
- Demil, G., Haghighi, A. T., Klöve, B., & Oussalah, M. (2025). Advances in image-based estimation of snow variable: A systematic literature review on recent studies. *Journal of Hydrology*, 654, 132855. <https://doi.org/10.1016/j.jhydrol.2025.132855>
- Dickinson, R., Henderson-Sellers, A., Kennedy, P., & Wilson, M. (1986). *Biosphere-atmosphere transfer scheme (bats) for the near community climate model* (Tech. Rep.). National Center For: Atmospheric Research. <https://doi.org/10.5065/D6668B58>
- Ding, A., Liang, S., Ma, H., He, T., Jia, A., & Wang, Q. (2024). Improved estimation of daily blue-sky snow shortwave albedo from modis data and reanalysis information. *Science of Remote Sensing*, 10, 100163. <https://doi.org/10.1016/j.srs.2024.100163>
- Diro, G. T., & Sushama, L. (2018). Snow–precipitation coupling and related atmospheric feedbacks over North America. *Atmospheric Science Letters*, 19(8), e831. <https://doi.org/10.1002/asl.831>
- Dong, J., Ek, M., Hall, D., Peters-Lidard, C., Cosgrove, B., Miller, J., et al. (2014). Using air temperature to quantitatively predict the modis fractional snow cover retrieval errors over the continental United States. *Journal of Hydrometeorology*, 15(2), 551–562. <https://doi.org/10.1175/jhm-d-13-060.1>
- Dozier, J. (1991). Retrieval of snow properties from aviris data. In *Third airborne science workshop, jpl report*.
- Dozier, J., Green, R. O., Nolin, A. W., & Painter, T. H. (2009). Interpretation of snow properties from imaging spectrometry. *Remote Sensing of Environment*, 113, S25–S37. <https://doi.org/10.1016/j.rse.2007.07.029>
- Dunkle, R. V., & Bevens, J. T. (1956). An approximate analysis of the solar reflectance and transmittance of a snow cover. *Journal of Meteorology*, 13(2), 212–216. [https://doi.org/10.1175/1520-0469\(1956\)013<0212:aaaots>2.0.co;2](https://doi.org/10.1175/1520-0469(1956)013<0212:aaaots>2.0.co;2)
- Dutra, J., Balsamo, G., Viterbo, P., Miranda, P. M. A., Beljaars, A., Schär, C., & Elder, K. (2010). An improved snow scheme for the ecmwf land surface model: Description and offline validation. *Journal of Hydrometeorology*, 11(4), 899–916. <https://doi.org/10.1175/2010jhm1249.1>
- Elvidge, S., Angling, M. J., & Nava, B. (2014). On the use of modified taylor diagrams to compare ionospheric assimilation models. *Radio Science*, 49(9), 737–745. <https://doi.org/10.1002/2014rs005435>
- Essery, R., Best, M., & Cox, P. (2001). *Moses 2.2 technical documentation* (Vol. 30). Hadley Centre for Climate Prediction and Research.
- Essery, R., Kontu, A., Lemmetyinen, J., Dumont, M., & Ménard, C. B. (2016). A 7-year dataset for driving and evaluating snow models at an arctic site (sodankylä, Finland). *Geoscientific Instrumentation, Methods and Data Systems*, 5(1), 219–227. <https://doi.org/10.5194/gi-5-219-2016>
- Essery, R., Morin, S., Lejeune, Y., & B Ménard, C. (2013). A comparison of 1701 snow models using observations from an alpine site. *Advances in Water Resources*, 55, 131–148. <https://doi.org/10.1016/j.advwatres.2012.07.013>
- Etchevers, P., Martin, E., Brown, R., Fierz, C., Lejeune, Y., Bazile, E., et al. (2004). Validation of the energy budget of an alpine snowpack simulated by several snow models (snow mip project). *Annals of Glaciology*, 38, 150–158. <https://doi.org/10.3189/172756404781814825>
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model inter-comparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>
- Feng, S., Wehrli, A., Cook, J. M., Anesio, A. M., Box, J. E., Benning, L. G., & Tranter, M. (2024). The apparent effect of orbital drift on time series of modis mod10a1 albedo on the Greenland ice sheet. *Science of Remote Sensing*, 9, 100116. <https://doi.org/10.1016/j.srs.2023.100116>
- Flanner, M. G., Arnheim, J. B., Cook, J. M., Dang, C., He, C., Huang, X., et al. (2021). Snicar-adv3: A community tool for modeling spectral snow albedo. *Geoscientific Model Development*, 14(12), 7673–7704. <https://doi.org/10.5194/gmd-14-7673-2021>
- Fleming, S. W., Rittger, K., Oaida Taghialatela, C. M., & Graczyk, I. (2024). Leveraging next-generation satellite remote sensing-based snow data to improve seasonal water supply predictions in a practical machine learning-driven river forecast system. *Water Resources Research*, 60(4), e2023WR035785. <https://doi.org/10.1029/2023wr035785>
- Fleming, S. W., Zukiewicz, L., Strobel, M. L., Hofman, H., & Goodbody, A. G. (2023). Snotel, the soil climate analysis network, and water supply forecasting at the natural resources conservation service: Past, present, and future. *JAWRA Journal of the American Water Resources Association*, 59(4), 585–599. <https://doi.org/10.1111/1752-1688.13104>
- Gaillard, M., Vionnet, V., Lafaysse, M., Dumont, M., & Ginoux, P. (2025). Improving large-scale snow albedo modeling using a climatology of light-absorbing particle deposition. *The Cryosphere*, 19(2), 769–792. <https://doi.org/10.5194/tc-19-769-2025>
- Gao, L., Zhang, L., Shen, Y., Zhang, Y., Ai, M., & Zhang, W. (2021). Modeling snow depth and snow water equivalent distribution and variation characteristics in the Irtysh River basin, China. *Applied Sciences*, 11(18), 8365. <https://doi.org/10.3390/app11188365>
- Gardner, A. S., & Sharp, M. J. (2010). A review of snow and ice albedo and the development of a new physically based broadband albedo parameterization. *Journal of Geophysical Research*, 115(F1). <https://doi.org/10.1029/2009j001444>
- Giddings, J. C., & LaChapelle, E. (1961). Diffusion theory applied to radiant energy distribution and albedo of snow. *Journal of Geophysical Research*, 66(1), 181–189. <https://doi.org/10.1029/jz066i001p00181>
- Gorishniy, Y., Rubachev, I., Khrulkov, V., & Babenko, A. (2023). Revisiting deep learning models for tabular data. Retrieved from <https://arxiv.org/abs/2106.11959>

- Greuell, W., & Konzmann, T. (1994). Numerical modelling of the energy balance and the englacial temperature of the Greenland ice sheet. Calculations for the eth-camp location (west Greenland, 1155 m a.s.l.). *Global and Planetary Change*, 9(1–2), 91–114. [https://doi.org/10.1016/0921-8181\(94\)90010-8](https://doi.org/10.1016/0921-8181(94)90010-8)
- Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on tabular data? Retrieved from <https://arxiv.org/abs/2207.08815>
- Günther, D., Marke, T., Essery, R., & Strasser, U. (2019). Uncertainties in snowpack simulations—assessing the impact of model structure, parameter choice, and forcing data error on point-scale energy balance snow model performance. *Water Resources Research*, 55(4), 2779–2800. <https://doi.org/10.1029/2018wr023403>
- Hall, D. K., & Riggs, G. A. (2007). Accuracy assessment of the modis snow products. *Hydrological Processes*, 21(12), 1534–1547. <https://doi.org/10.1002/hyp.6715>
- Hansen, J., Russell, G., Rind, D., Stone, P., Lacis, A., Lebedeff, S., et al. (1983). Efficient three-dimensional global models for climate studies: Models i and ii. *Monthly Weather Review*, 111(4), 609–662. [https://doi.org/10.1175/1520-0493\(1983\)111<0609:etdgmf>2.0.co;2](https://doi.org/10.1175/1520-0493(1983)111<0609:etdgmf>2.0.co;2)
- Hao, D., Bisht, G., Rittger, K., Stillinger, T., Bair, E., Gu, Y., & Leung, L. R. (2023). Evaluation of e3sm land model snow simulations over the western United States. *The Cryosphere*, 17(2), 673–697. <https://doi.org/10.5194/tc-17-673-2023>
- Hatchett, B. J., Koshkin, A. L., Guirguis, K., Rittger, K., Nolin, A. W., Heggli, A., et al. (2023). Midwinter dry spells amplify post-fire snowpack decline. *Geophysical Research Letters*, 50(3), e2022GL101235. <https://doi.org/10.1029/2022gl101235>
- He, C., Flanner, M., Lawrence, D. M., & Gu, Y. (2024). New features and enhancements in community land model (clm5) snow albedo modeling: Description, sensitivity, and evaluation. *Journal of Advances in Modeling Earth Systems*, 16(2), e2023MS003861. <https://doi.org/10.1029/2023ms003861>
- Henneman, H. E., & Stefan, H. G. (1999). Albedo models for snow and ice on a freshwater lake. *Cold Regions Science and Technology*, 29(1), 31–48. [https://doi.org/10.1016/s0165-232x\(99\)00002-6](https://doi.org/10.1016/s0165-232x(99)00002-6)
- Hill, D. F., Burakowski, E. A., Crumley, R. L., Keon, J., Hu, J. M., Arendt, A. A., et al. (2019). Converting snow depth to snow water equivalent using climatological variables. *The Cryosphere*, 13(7), 1767–1784. <https://doi.org/10.5194/tc-13-1767-2019>
- Hinton, G., Srivastava, N., & Swersky, K. (2012). *Lecture 6e: Rmsprop: Divide the gradient by a running average of its recent magnitude*. University of Toronto.
- Huang, Y., Song, Z., Yang, H., Yu, B., Liu, H., Che, T., et al. (2022). Snow cover detection in mid-latitude mountainous and polar regions using nighttime light data. *Remote Sensing of Environment*, 268, 112766. <https://doi.org/10.1016/j.rse.2021.112766>
- Innes, M. (2018). Flux: Elegant machine learning with julia. *Journal of Open Source Software*, 3(25), 602. <https://doi.org/10.21105/joss.00602>
- Jennings, K. S., Winchell, T. S., Livneh, B., & Molotch, N. P. (2018). Spatial variation of the rain–snow temperature threshold across the northern hemisphere. *Nature Communications*, 9(1), 1148. <https://doi.org/10.1038/s41467-018-03629-7>
- Jensen, A. S., Rittger, K., & Raleigh, M. S. (2024). Spatio-temporal patterns and trends in modis-retrieved radiative forcing by snow impurities over the western us from 2001 to 2022. *Environmental Research: Climate*, 3(2), 025001. <https://doi.org/10.1088/2752-5295/ad285a>
- Jia, A., Wang, D., Liang, S., Peng, J., & Yu, Y. (2023). Improved cloudy-sky snow albedo estimates using passive microwave and viirs data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196, 340–355. <https://doi.org/10.1016/j.isprsjprs.2023.01.004>
- Jia, J., Zeng, Z., Tian, B., Chen, S., Chen, L., Wang, Y., et al. (2025). The construction of 20-year daily surface albedo along panda transect, Antarctica. *Journal of Geophysical Research: Atmospheres*, 130(5), e2024JD042863. <https://doi.org/10.1029/2024jd042863>
- Keuris, L., Hetzenecker, M., Nagler, T., Mölg, N., & Schwaizer, G. (2023). An adaptive method for the estimation of snow-covered fraction with error propagation for applications from local to global scales. *Remote Sensing*, 15(5), 1231. <https://doi.org/10.3390/rs15051231>
- Kiem, J., Hammerle, A., Montagnani, L., & Wohlfahrt, G. (2024). Future scenarios of albedo and radiative forcing resulting from changes in snow depth in Austria. *Copernicus GmbH*. <https://doi.org/10.5194/egusphere-2024-881>
- Klein, A. (2003). Validation of daily modis snow cover maps of the upper rio Grande River basin for the 2000–2001 snow year. *Remote Sensing of Environment*, 86(2), 162–176. [https://doi.org/10.1016/s0034-4257\(03\)00097-x](https://doi.org/10.1016/s0034-4257(03)00097-x)
- Klein, A. G., & Stroeve, J. (2002). Development and validation of a snow albedo algorithm for the modis instrument. *Annals of Glaciology*, 34, 45–52. <https://doi.org/10.3189/172756402781817662>
- Kondo, J., & Yamazaki, T. (1990). A prediction model for snowmelt, snow surface temperature and freezing depth using a heat balance method. *Journal of Applied Meteorology*, 29(5), 375–384. [https://doi.org/10.1175/1520-0450\(1990\)029<0375:apmfss>2.0.co;2](https://doi.org/10.1175/1520-0450(1990)029<0375:apmfss>2.0.co;2)
- Kouki, K., Räisänen, P., Luojus, K., Luomaranta, A., & Riihelä, A. (2022). Evaluation of northern hemisphere snow water equivalent in cmip6 models during 1982–2014. *The Cryosphere*, 16(3), 1007–1030. <https://doi.org/10.5194/tc-16-1007-2022>
- Krajčič, P., Kirnbauer, R., Parajka, J., Schöber, J., & Blöschl, G. (2017). The kühnau data set: 25 years of lysimetric, snow pillow, and meteorological measurements. *Water Resources Research*, 53(6), 5158–5165. <https://doi.org/10.1029/2017wr020445>
- Krinner, G., Derksen, C., Essery, R., Flanner, M., Hagemann, S., Clark, M., et al. (2018). Esm-snowmip: Assessing snow models and quantifying snow-related climate feedbacks. *Geoscientific Model Development*, 11(12), 5027–5049. <https://doi.org/10.5194/gmd-11-5027-2018>
- Lawrence, D. M., Oleson, K. W., Flanner, M. G., Thornton, P. E., Swenson, S. C., Lawrence, P. J., et al. (2011). Parameterization improvements and functional and structural advances in version 4 of the community land model. *Journal of Advances in Modeling Earth Systems*, 3(1). <https://doi.org/10.1029/2011ms00045>
- Lee, W. Y., Gim, H.-J., & Park, S. K. (2023). Parameterizations of snow cover, snow albedo and snow density in land surface models: A comparative review. *Asia-Pacific Journal of Atmospheric Sciences*, 60(2), 185–210. <https://doi.org/10.1007/s13143-023-00344-2>
- Lehning, M., Bartelt, P., Brown, B., & Fierz, C. (2002). A physical snowpack model for the swiss avalanche warning. *Cold Regions Science and Technology*, 35(3), 169–184. [https://doi.org/10.1016/s0165-232x\(02\)00072-1](https://doi.org/10.1016/s0165-232x(02)00072-1)
- Lejeune, Y., Dumont, M., Panel, J.-M., Lafaysse, M., Lapalus, P., Le Gac, E., et al. (2019). 57 years (1960–2017) of snow and meteorological observations from a mid-altitude mountain site (col de porte, france, 1325m of altitude). *Earth System Science Data*, 11(1), 71–88. <https://doi.org/10.5194/essd-11-71-2019>
- Li, H., Zhang, G., Wang, C., Liu, Z., Ju, C., & Mamtimin, A. (2023). Improving snow albedo parameterization scheme based on remote sensing data. *Atmospheric Research*, 284, 106602. <https://doi.org/10.1016/j.atmosres.2022.106602>
- Lin, T.-S., He, C., Abolafia-Rosenzweig, R., Chen, F., Wang, W., Barlage, M., & Gochis, D. (2025). Improved snow albedo evolution in noah-mp land surface model coupled with a physical snowpack radiative transfer scheme. *Journal of Hydrometeorology*, 26(2), 185–200. <https://doi.org/10.1175/jhm-d-24-0082.1>
- Liu, L., Ma, Y., Menenti, M., Su, R., Yao, N., & Ma, W. (2021). Improved parameterization of snow albedo in noah coupled with weather research and forecasting: Applicability to snow estimates for the Tibetan Plateau. *Hydrology and Earth System Sciences*, 25(9), 4967–4981. <https://doi.org/10.5194/hess-25-4967-2021>
- Livneh, B., Deems, J. S., Schneider, D., Barsugli, J. J., & Molotch, N. P. (2014). Filling in the gaps: Inferring spatially distributed precipitation from gauge observations over complex terrain. *Water Resources Research*, 50(11), 8589–8610. <https://doi.org/10.1002/2014wr015442>

- Livneh, B., Xia, Y., Mitchell, K. E., Ek, M. B., & Lettenmaier, D. P. (2010). Noah lsm snow model diagnostics and enhancements. *Journal of Hydrometeorology*, 11(3), 721–738. <https://doi.org/10.1175/2009jhm1174.1>
- Magnusson, J., Wever, N., Essery, R., Helbig, N., Winstral, A., & Jonas, T. (2015). Evaluating snow models with varying process representations for hydrological applications. *Water Resources Research*, 51(4), 2707–2723. <https://doi.org/10.1002/2014wr016498>
- Malik, M. J., van der Velde, R., Vekedy, Z., & Su, Z. (2014). Improving modeled snow albedo estimates during the spring melt season. *Journal of Geophysical Research: Atmospheres*, 119(12), 7311–7331. <https://doi.org/10.1002/2013jd021344>
- Marks, D. G. (1988). *Climate, energy exchange, and snowmelt in emerald lake watershed, sierra Nevada*. Department of Geography and Mechanical Engineering., University of California.
- Marshall, S. E. (1989). A physical parameterization of snow albedo for use in climate models, ncar cooperative thesis 123 [Ph. D. Thesis].
- Menard, C. B., Essery, R., Krinner, G., Arduini, G., Bartlett, P., Boone, A., et al. (2021). Scientific and human errors in a snow model inter-comparison. *Bulletin of the American Meteorological Society*, 102(1), E61–E79. <https://doi.org/10.1175/bams-d-19-0329.1>
- Meyer, J. D. D., Jin, J., & Wang, S.-Y. (2012). Systematic patterns of the inconsistency between snow water equivalent and accumulated precipitation as reported by the snowpack telemetry network. *Journal of Hydrometeorology*, 13(6), 1970–1976. <https://doi.org/10.1175/jhm-d-12-066.1>
- Mölders, N., Luijting, H., & Sassen, K. (2007). Use of atmospheric radiation measurement program data from barrow, Alaska, for evaluation and development of snow-albedo parameterizations. *Meteorology and Atmospheric Physics*, 99(3–4), 199–219. <https://doi.org/10.1007/s00703-007-0271-6>
- Molnar, C. (2025). *Interpretable machine learning* (3rd ed.). Author. Retrieved from <https://christophm.github.io/interpretable-ml-book>
- Molotch, N. P., & Bales, R. C. (2006). Comparison of ground-based and airborne snow surface albedo parameterizations in an alpine watershed: Impact on snowpack mass balance. *Water Resources Research*, 42(5). <https://doi.org/10.1029/2005wr004522>
- National Research Council. (1999). In P. C. Stern & W. E. Easterling (Eds.), *Making climate forecasts matter* (pp. 18–37). The National Academies Press. <https://doi.org/10.17226/6370>
- Nguyen, H., Mehrotra, R., & Sharma, A. (2016). Correcting for systematic biases in gcm simulations in the frequency domain. *Journal of Hydrology*, 538, 117–126. <https://doi.org/10.1016/j.jhydrol.2016.04.018>
- Oaida, C. M., Xue, Y., Flanner, M. G., Skiles, S. M., De Sales, F., & Painter, T. H. (2015). Improving snow albedo processes in wrf/ssib regional climate model to assess impact of dust and black carbon in snow on surface energy balance and hydrology over western u.s. *Journal of Geophysical Research: Atmospheres*, 120(8), 3228–3248. <https://doi.org/10.1002/2014jd022444>
- Oerlemans, J., & Knap, W. H. (1998). A 1 year record of global radiation and albedo in the ablation zone of morteratschgletscher, Switzerland. *Journal of Glaciology*, 44(147), 231–238. <https://doi.org/10.3189/s0022143000002574>
- O'Neill, A., & Gray, D. M. (1972). Solar radiation penetration through snow. In *The role of snow and ice in hydrology, proceedings of the banff symposium, int. Assoc. Hydrol. Sci* (Vol. 107, pp. 227–240).
- Painter, T. H., Berisford, D. F., Boardman, J. W., Bormann, K. J., Deems, J. S., Gehrke, F., et al. (2016). The airborne snow observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo. *Remote Sensing of Environment*, 184, 139–152. <https://doi.org/10.1016/j.rse.2016.06.018>
- Painter, T. H., Bryant, A. C., & Skiles, S. M. (2012). Radiative forcing by light absorbing impurities in snow from modis surface reflectance data. *Geophysical Research Letters*, 39(17). <https://doi.org/10.1029/2012gl052457>
- Painter, T. H., Rittger, K., McKenzie, C., Slaughter, P., Davis, R. E., & Dozier, J. (2009). Retrieval of subpixel snow covered area, grain size, and albedo from modis. *Remote Sensing of Environment*, 113(4), 868–879. <https://doi.org/10.1016/j.rse.2009.01.001>
- Palomaki, R. T., Rittger, K., Lenard, S. J., Bair, E., Dozier, J., Skiles, S. M., & Painter, T. H. (2025). Assessment of methods for mapping snow albedo from modis. *Remote Sensing of Environment*, 326, 114742. <https://doi.org/10.1016/j.rse.2025.114742>
- Picard, G., & Libois, Q. (2024). Simulation of snow albedo and solar irradiance profile with the two-stream radiative transfer in snow (tartes) v2.0 model. *Geoscientific Model Development*, 17(24), 8927–8953. <https://doi.org/10.5194/gmd-17-8927-2024>
- Pike, C., Riley, D., Toal, H., & Burnham, L. (2024). Presenting a model to predict changing snow albedo for improving photovoltaic performance simulation. *Inside Solaris*, 4(3), 422–439. <https://doi.org/10.3390/solar4030019>
- Pirazzini, R., Leppänen, L., Picard, G., Lopez-Moreno, J. I., Marty, C., Macelloni, G., et al. (2018). European in-situ snow measurements: Practices and purposes. *Sensors*, 18(7), 2016. <https://doi.org/10.3390/s18072016>
- Qu, X., & Hall, A. (2006). Assessing snow albedo feedback in simulated climate change. *Journal of Climate*, 19(11), 2617–2630. <https://doi.org/10.1175/JCLI3750.1>
- Rahimi, I., Gandomi, A. H., Chen, F., & Mezura-Montes, E. (2023). A review on constraint handling techniques for population-based algorithms: From single-objective to multi-objective optimization. *Archives of Computational Methods in Engineering*, 30(3), 2181–2209. <https://doi.org/10.1007/s11831-022-09859-9>
- Raleigh, M. S., Livneh, B., Lapo, K., & Lundquist, J. D. (2015). How does availability of meteorological forcing data impact physically based snowpack simulations? *Journal of Hydrometeorology*, 17(1), 99–120. <https://doi.org/10.1175/jhm-d-14-0235.1>
- Ranzi, R., & Rosso, R. (1991). A physically based approach to modelling distributed snowmelt in a small alpine catchment, 205, p. 141–150).
- Riley, J. P., Chadwick, D. G., & Eggleston, K. O. (1969). Snowmelt simulation. *Report*.
- Rittger, K., Lenard, S. J. P., Palomaki, R., Brodzik, M., Stilling, T., Bair, E., et al. (2024). *Historical modis/terra l3 global daily 500m sin grid snow cover, snow albedo, and snow physical properties*. National Snow and Ice Data Center. <https://doi.org/10.7265/F6J3-F387>
- Rittger, K., Painter, T. H., & Dozier, J. (2013). Assessment of methods for mapping snow cover from modis. *Advances in Water Resources*, 51, 367–380. <https://doi.org/10.1016/j.advwatres.2012.03.002>
- Rittger, K., Raleigh, M. S., Dozier, J., Hill, A. F., Lutz, J. A., & Painter, T. H. (2020). Canopy adjustment and improved cloud detection for remotely sensed snow cover mapping. *Water Resources Research*, 56(6), e2019WR024914. <https://doi.org/10.1029/2019wr024914>
- Rittger, K., Serreze, M., Brodzik, M. J., Raleigh, M. S., Skiles, M., Musselman, K., et al. (2023). Co-production of snow cover and snow albedo at sub-pixel scale with analysis, uncertainty, and data distribution in near real-time. In *Proceedings of the 2023 modis/viirs science team meeting, panel 8: Snow cover and albedo*. Retrieved from https://modis.gsfc.nasa.gov/sci_team/meetings/202305/presentations/rittger.pdf
- Roekner, E., Arpe, K., Bengtsson, L., Brinkop, S., Dümenil, L., Esch, M., et al. (1992). Simulation of the present-day climate with the echam model: Impact of model physics and resolution. <https://doi.org/10.17617/2.1852612>
- Rohrer, M. B. (1991). Die schneedecke im schweizerischen alpenraum und ihre modellierung: Analyse langjähriger schneemessungen und modellierung des wasseräquivalentes aufgrund zeitlich hoch aufgelöster operationell gemessener meteorologischer daten (Doctoral dissertation). <https://doi.org/10.3929/ETHZ-A-000601662>
- Saito, M., Yang, P., Loeb, N. G., & Kato, S. (2019). A novel parameterization of snow albedo based on a two-layer snow model with a mixture of grain habits. *Journal of the Atmospheric Sciences*, 76(5), 1419–1436. <https://doi.org/10.1175/jas-d-18-0308.1>

- Sarangi, C., Qian, Y., Rittger, K., Bormann, K. J., Liu, Y., Wang, H., et al. (2019). Impact of light-absorbing particles on snow albedo darkening and associated radiative forcing over high-mountain Asia: High-resolution wrf-chem modeling and new satellite observations. *Atmospheric Chemistry and Physics*, 19(10), 7105–7128. <https://doi.org/10.5194/acp-19-7105-2019>
- Serreze, M. C., Clark, M. P., Armstrong, R. L., McGinnis, D. A., & Pulwarty, R. S. (1999). Characteristics of the western United States snowpack from snowpack telemetry (snotel) data. *Water Resources Research*, 35(7), 2145–2160. <https://doi.org/10.1029/1999wr900090>
- Shao, D., Xu, W., Li, H., Wang, J., & Hao, X. (2020). Modeling snow surface spectral reflectance in a land surface model targeting satellite remote sensing observations. *Remote Sensing*, 12(18), 3101. <https://doi.org/10.3390/rs12183101>
- Shea, J., Wagnon, P., Immerzeel, W., Biron, R., Brun, F., & Pellicciotti, F. (2015). A comparative high-altitude meteorological analysis from three catchments in the Nepalese himalaya. *International Journal of Water Resources Development*, 31(2), 174–200. <https://doi.org/10.1080/07900627.2015.1020417>
- Stigter, E. E., Steiner, J. F., Koch, I., Saloranta, T. M., Kirkham, J. D., & Immerzeel, W. W. (2021). Energy and mass balance dynamics of the seasonal snowpack at two high-altitude sites in the himalaya. *Cold Regions Science and Technology*, 183, 103233. <https://doi.org/10.1016/j.coldregions.2021.103233>
- Stillinger, T., Rittger, K., Raleigh, M. S., Michell, A., Davis, R. E., & Bair, E. H. (2023). Landsat, modis, and viirs snow cover mapping algorithm performance as validated by airborne LiDAR datasets. *The Cryosphere*, 17(2), 567–590. <https://doi.org/10.5194/tc-17-567-2023>
- Strasser, U., & Marke, T. (2010). Escimo.spread – A spreadsheet-based point snow surface energy balance model to calculate hourly snow water equivalent and melt rates for historical and changing climate conditions. *Geoscientific Model Development*, 3(2), 643–652. <https://doi.org/10.5194/gmd-3-643-2010>
- Ström, J., Svensson, J., Honkanen, H., Asmi, E., Dkhar, N. B., Tayal, S., et al. (2022). Snow albedo and its sensitivity to changes in deposited light-absorbing particles estimated from ambient temperature and snow depth observations at a high-altitude site in the himalaya. *Elementa: Science of the Anthropocene*, 10(1), 00118. <https://doi.org/10.1525/elementa.2021.00118>
- Thackeray, C. W., Qu, X., & Hall, A. (2018). Why do models produce spread in snow albedo feedback? *Geophysical Research Letters*, 45(12), 6223–6231. <https://doi.org/10.1029/2018GL078493>
- USDA. (2010). *Title 210-national engineering handbook, part 622-snow survey and water supply forecasting* (pp. 2–2–12). Author.
- Verseghy, D. L. (1991). Class—A Canadian land surface scheme for gems. I. Soil model. *International Journal of Climatology*, 11(2), 111–133. <https://doi.org/10.1002/joc.3370110202>
- Voordendag, A., Réveillet, M., MacDonell, S., & Lhermitte, S. (2021). Snow model comparison to simulate snow depth evolution and sublimation at point scale in the semi-arid andes of Chile. *The Cryosphere*, 15(9), 4241–4259. <https://doi.org/10.5194/tc-15-4241-2021>
- Wang, W., Yang, K., Zhao, L., Zheng, Z., Lu, H., Mantimin, A., et al. (2020). Characterizing surface albedo of shallow fresh snow and its importance for snow ablation on the interior of the Tibetan Plateau. *Journal of Hydrometeorology*, 21(4), 815–827. <https://doi.org/10.1175/jhm-d-19-0193.1>
- Wang, Z., Schaaf, C. B., Chopping, M. J., Strahler, A. H., Wang, J., Román, M. O., et al. (2012). Evaluation of moderate-resolution imaging spectroradiometer (modis) snow albedo product (mcd43a) over tundra. *Remote Sensing of Environment*, 117, 264–280. <https://doi.org/10.1016/j.rse.2011.10.002>
- Warren, S. G. (1982). Optical properties of snow. *Reviews of Geophysics*, 20(1), 67–89. <https://doi.org/10.1029/rg020i001p00067>
- Warscher, M., Marke, T., Rottler, E., & Strasser, U. (2024). Operational and experimental snow observation systems in the upper rofental: Data from 2017 to 2023. *Earth System Science Data*, 16(8), 3579–3599. <https://doi.org/10.5194/essd-16-3579-2024>
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6), 80. <https://doi.org/10.2307/3001968>
- Williamson, S. N., Copland, L., & Hik, D. S. (2016). The accuracy of satellite-derived albedo for northern alpine and glaciated land covers. *Polar Science*, 10(3), 262–269. <https://doi.org/10.1016/j.polar.2016.06.006>
- Wiscombe, W. J., & Warren, S. G. (1980). A model for the spectral albedo of snow. I: Pure snow. *Journal of the Atmospheric Sciences*, 37(12), 2712–2733. [https://doi.org/10.1175/1520-0469\(1980\)037<2712:amftsa>2.0.co;2](https://doi.org/10.1175/1520-0469(1980)037<2712:amftsa>2.0.co;2)
- Xie, Y., Pettersen, C., Flanner, M., & Shates, J. (2024). Ground-observed snow albedo changes during rain-on-snow events in northern Alaska. *Journal of Geophysical Research: Atmospheres*, 129(16), e2024JD040975. <https://doi.org/10.1029/2024jd040975>
- Yan, H., Sun, N., Wigmosta, M., Skaggs, R., Hou, Z., & Leung, R. (2018). Next-generation intensity-duration-frequency curves for hydrologic design in snow-dominated environments. *Water Resources Research*, 54(2), 1093–1108. <https://doi.org/10.1002/2017wr021290>
- Ye, F., Cheng, Q., Hao, W., Yu, D., Ma, C., Liang, D., & Shen, H. (2023). Reconstructing daily snow and ice albedo series for Greenland by coupling spatiotemporal and physics-informed models. *International Journal of Applied Earth Observation and Geoinformation*, 124, 103519. <https://doi.org/10.1016/j.jag.2023.103519>
- Zhang, Y., & Ohata, T. (2003). Parameterization of albedo on snow/ice surface. A review. *Journal of the Japanese Society of Snow and Ice*, 65(1), 33–51. <https://doi.org/10.5331/seppyo.65.33>
- Zhao, Z., Ding, Y., Wang, M., & Qu, Y. (2025). A radiative transfer model-driven machine learning approach to estimate the snow surface albedo over the Greenland ice sheet. *IEEE Transactions on Geoscience and Remote Sensing*, 63, 1–18. <https://doi.org/10.1109/tgrs.2025.3551094>
- Zhongyuan, B., & Ohata, T. (1989). Changes in reflectance of glacier no. 1 at the source of the urumqi river in the tianshan Mountains during the summer ablation period. *Journal of Glaciology and Geocryology*, 11(4), 311–324.
- Zwart, C., & Fierz, C. (2007). Significance of new snow properties for snow-cover development. In *Geophysical research abstracts* (Vol. 9), 10287