

Advancing Large-Scale Snow Modeling with Data-Driven Parameterizations

Thesis by
Andrew J. Charbonneau

In Partial Fulfillment of the Requirements for the
Degree of
Doctorate of Philosophy in Applied Physics



CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2026
Defended May 11th, 2026

© 2026

Andrew J. Charbonneau

ORCID: 0009-0004-3269-5713

All rights reserved except where otherwise noted

ACKNOWLEDGEMENTS

If I have seen further, it is by standing on the shoulders of Giants.

– Isaac Newton, 1675

As I reflect on my doctoral journey, I am reminded that you never end up where you thought you were going — not due to wandering; lost in a sea of confusion (at least, not all the time), but instead from assessing your larger goals, seizing the lucky opportunities as they appear, and riding the current wherever it takes you. I entered Caltech expecting to study optimization of metamaterials and the physics of small-scale interactions, only to find myself steered along the way towards climate modeling, and problems of physical exchange spanning much, much, larger scales. From a global pandemic, the rise of artificial intelligence, fire evacuations, and a now-precedented level of “unprecedented” events, I am thankful for the recognition that the journey I am on and how to navigate it are as important as the destination, and that *no* work is ever truly alone — it takes a village.

I would foremost like to thank Tapio Schneider and Katherine Deck, for volunteering their time in helping a graduate student in their professional development. For me to begin in the CliMA group with next to no climate background but reach where I am today is primarily due to the two of you. Your guidance, wisdom, and *immeasurable* patience has not gone unnoticed nor unappreciated. Your support and insight has shaped this piece and my overall approach to scientific investigation in ways that cannot be summarized in words, and I will carry these lessons forward wherever the currents lead me next. To both of you, I want to acknowledge in writing the innumerable informal and formal meetings to discuss physics, careers, academics, and life in general — you have been there for me from my first meeting at CliMA, and have been far more than just mentors. Something perhaps even more clear and important to me than your advice has been your enthusiasm for your own work and for assisting others; your intrigue and thoughtfulness is inspiring, and I would not be where I am today if it were not for you. You have both played an integral role as I have learned, once again, how to learn. It has been a privilege to associate with the both of you, and I look forward to further correspondence.

So too I would like to thank the rest of my committee, Andrei Faraon and Andy Thompson. It was Andrei’s choice to interview and believe in a curious undergraduate student over 6 years ago that allowed me to ultimately put forth this dissertation,

and I am thrilled to close the loop of this period with your participation in my defense. I am thankful for your recognition of my broad interests, and my excitement to apply the world of physics in new and expanding ways wherever I go. To Andy, I am thankful for your readiness to join my candidacy and my defense on a moment's notice, and our meetings before both. Your insight has also shaped my approach to these upcoming pages, and your enthusiasm for both your field and those related has been contagious. I am honored to be but the next in a long line of students defending under you and Tapio, and I hope that in the tools I present here that you might find utility in your own work; that I might begin to reciprocate what you have provided for me during my time at Caltech.

No acknowledgements would be complete without thanking the most important people for getting me to this point, and that is my mother, father, and sister. You have been there for every celebration, frustrated rant, and so much more from the (literal) first day. You have provided all the support and advice (and food) where nobody else could, as well as all the affection and fun times I could ever need. So too this shout-out goes to Grandma and Grandpa, Memere and Pepere, my faithful Godparents, and my cousins who have always been there to check in and cheer me on. There is too much for me to put into one paragraph for everyone — I love you all, and I am excited to see where the next few years take us.

I am extremely grateful for the administrators and staff of CliMA and my departments who have also encouraged and supported me during my PhD — you are the true machine that keeps research running, and your dedication is forever appreciated. Chiefly in this group lies Leilani Rivera-Dotson, Nora Oshima, and Jennifer Blankenship.

Nothing greases the gears in academia like funding, and my fellowships played a large role in making some room where there otherwise would not have been. For that, there are many to thank for their part in these programs and in my professional development. Specifically, to Eric Harper, Momo Wolapaye, and Chiara Daraio, your support in reading, evaluating, and advising me on science, on leadership, and on life will stick with me well into my career, and I cannot thank you enough.

With advisers, teachers, family, role models, and mentors out of the way, this last section is devoted to those who somehow manage to act as a combination of all five — my friends. I have no doubt the relationships I have formed in this period will last long beyond our time in Pasadena, and look forward to seeing you in the years to come. To Rohit Dilip and Tinashe Handina, you have been there through ups and

downs, late nights, late rants, hundreds of messages, and much more, since before this experience began. To Arda Seçme, Daniel Murphy, Miles Johnson, and Alex Johnson, from the beginning you have been there to help me take a break, hear me out, and everything in between. To Mikhail Hameedi, Will Lawrence, Eva Kerčmar, Kent Leslie, Adam Artymowicz, Theresa Tsaggari, and Jessica Nassimi, thank you for becoming all the above in half the time. To everyone here, you have likely taught me as much as the professors and scientists listed above. I will recall our time fondly here and am excited to see your names in journals, news, and awards, and cannot wait to catch up and continue to make more memories. As always, a shoutout to my TJ crew for keeping it going for 14 years, and my brother in Christ, Caton Yang, for all we have overcome these past few years. Last, but certainly not least, to the CliMA collaboration, and the house that made my research into a home, and a family. To all, a few sentences will never capture my gratitude, and I look forward to what lies ahead.

And, as always, to my loyal proofreaders, whose only mistake was being regarded highly enough that I asked them to check:

Steven & Tracy Charbonneau.

ABSTRACT

Snow is a key component of the terrestrial cryosphere, governing surface energy exchange, water storage, and hydrological variability across seasonal to multi-decadal timescales. Despite its importance, snow dynamics remain among the most challenging earth system components to model. This difficulty arises from the strong coupling of microphysical, radiative, and mechanical processes across scales spanning orders of magnitude, from grain-scale metamorphism to global atmospheric circulation. This dissertation develops physics-informed parameterizations for snow depth, albedo, and surface temperature for usage within large-scale snow models. For snow depth and albedo, this work develops and evaluates a hybrid physics-machine learning framework, designed to reconcile the expressive flexibility of data-driven methods with the structural constraints required for physically consistent simulation. The proposed approach embeds physical guardrails directly into a neural dynamical system, enabling small, computationally efficient models to learn nonlinear updates while remaining stable, bounded, and compatible with coupled global climate model architectures. Across both site-level and global-scale evaluations, the developed schemes match or exceed established empirical and process-based parameterizations. This performance extends from the modeled variables to downstream effects like snow season timing, and generalizes across a range of climate types. The hybrid parameterization framework substantially improves generalization compared to unconstrained approaches. Further assessment suggests observational efforts and available training ranges remain critical for further improvement, rather than deficiencies in the modeling framework itself. Overall, this work demonstrates that a hybrid framework — and the choice of simple, yet flexible structures, instead of detailed but strict formulations — offers a viable and scalable alternative to traditional snow parameterizations, achieving improved accuracy at reduced cost. Outcomes reinforce the idea that embedding even minimal physical structure into data-driven or basic schemes can yield robust improvements in complicated physical systems.

PUBLISHED CONTENT AND CONTRIBUTIONS

- [1] A. Charbonneau, K. Deck, and T. Schneider. A physics-constrained neural differential equation framework for data-driven snowpack simulation. *Artificial Intelligence for the Earth Systems*, 4(3):240040, 2025. doi: 10.1175/AIES-D-24-0040.1. URL <https://journals.ametsoc.org/view/journals/aies/4/3/AIES-D-24-0040.1.xml>. A.C. was responsible for project conception, formulating the model framework, as well as gathering data and analyzing outputs, as the lead writer of the manuscript.

Other chapters are being submitted for publication and/or are under review at the time of this thesis submission, but not yet published. Information for each work has been accordingly placed at the beginning of each chapter.

TABLE OF CONTENTS

Acknowledgements	iii
Abstract	vi
Published Content and Contributions	vii
Table of Contents	vii
List of Illustrations	x
List of Tables	xii
Chapter I: Introduction	1
1.1 Problem Definition	1
1.2 Representing Snow	2
1.3 History and Context of Snow Modeling	7
1.4 Contemporary Snow Models	9
1.5 Emerging Paradigms: Machine Learning	10
1.6 Thesis Outline	12
Bibliography	15
Chapter II: A Physics-Constrained Neural Differential Equation Framework for Data-Driven Snowpack Simulation	26
Abstract	27
2.1 Introduction	27
2.2 Methodology	30
2.3 Results	38
2.4 Discussion	48
2.5 Conclusion	50
2.6 Data Availability	51
2.A Threshold Constraint Layers	52
2.B Data Methods	55
2.C Hyperparameters and Model Benchmarking	60
Bibliography	64
Chapter III: A Next-Generation Snow Albedo Parameterization for Climate Modeling using Constrained Machine Learning	75
Abstract	76
3.1 Introduction	76
3.2 Methodology	80
3.3 Results	91
3.4 Discussion	98
3.5 Conclusion	101
3.6 Data Availability	102
3.A Data Methods	102
3.B Performance Statistics	108
3.C Hyperparameters	111

Bibliography	113
Chapter IV: Global Snow Simulations in ClimaLand: Data-Driven and Physics-Informed Parameterizations for Depth, Albedo, and Surface Temperature	132
Abstract	133
4.1 Introduction	133
4.2 Model Description	136
4.3 Methodology	141
4.4 Results	144
4.5 Discussion	169
4.6 Conclusion	172
4.A Additional Parameterization Equations	172
4.B Additional Plots	174
Bibliography	178
Chapter V: Concluding Thoughts	187

LIST OF ILLUSTRATIONS

<i>Number</i>	<i>Page</i>
1.1 Overview of snow on the land surface.	1
1.2 Diagram of snow modeling difficulty.	3
1.3 Machine Learning (ML) in snow modeling.	11
2.1 Structure of the framework predictive component.	32
2.2 Architecture of M , highlighting the constraint component attached to the predictive component (the grey pentagon) described in Fig. 2.1.	33
2.3 Distribution of SNOTEL sites used for training the network.	35
2.4 Performance of M and $\tilde{M}$ against Snow17 with (SN17O) and without (SN17) observational SWE data for generating z timeseries over the 44 validation sites.	38
2.5 Performance of the neural parameterization against Snow17 over the 14 testing sites in depth timeseries generation.	39
2.6 z and SWE timeseries over two subsets of the Alaskan testing sites.	40
2.7 Bulk density timeseries (normalized against ρ_{water}) over the same series as Fig. 2.6.	41
2.8 Predicted vs. true dz/dt and residuals against the modeled target by the M and $\tilde{M}$ models.	43
2.9 Accumulated Local Effects (ALE) plot for M 's predictive features.	44
2.10 M outputs for two snow states, with other inputs held constant.	45
2.11 Performance of M across all training and testing sites with regards to RMSE, NSE, and SPE of depth timeseries, as well as RMSE for density timeseries over the sites where density was evaluated (see Appendix B).	46
2.12 Output of M at different resolutions for a subset of the site data from Kühtai, Austria.	47
2.13 General Framework Structures.	53
3.1 Structure of M	81
3.2 Relative Root Mean Square errors of the new parameterization (NN) versus existing parameterizations.	92
3.3 Modified Taylor Diagrams of all schemes over all ground (left) and coarse (right) testing data.	93

3.4	Timeseries of albedo generated by various schemes.	94
3.5	ALE Plot for features of <i>NN</i>	95
3.6	RMSE% scores of all calibrated parameterizations.	96
3.7	Modified Taylor Diagrams of Calibrated parameterizations.	97
4.1	Single site simulation timeseries for the Col De Porte snow site. . . .	145
4.2	z performance of updated parameterizations.	147
4.3	S performance of updated parameterizations.	148
4.4	Snow Cover performance of updated parameterizations.	150
4.5	Surface Albedo performance of updated parameterizations.	152
4.6	Latent Heat Flux performance of updated parameterizations.	154
4.7	Sensible Heat Flux performance of updated parameterizations.	155
4.8	Snow cover biases by latitude.	159
4.9	Snowpack disappearance date by latitude band.	161
4.10	Date of maximum snowpack height by latitude band.	162
4.11	Snow Depth and Snow Albedo biases by elevation and average snow- pack size.	163
4.12	Snowpack Maxima variability by simulation configuration.	164
4.13	Spatial bias in interannual variability (IAV).	167
4.14	Seasonal variability in spatial SSIM score for the new snow depth (top) and snow albedo (bottom) parameterizations.	168
4.15	Single site simulation timeseries for the Swamp Angel snow site. . .	175
4.16	Single site simulation timeseries for the Senator Beck snow site. . .	175
4.17	Snow Albedo performance of updated parameterizations.	176
4.18	Upward shortwave performance of updated parameterizations.	176
4.19	Long-wave upward radiation performance of updated parameteriza- tions.	177

LIST OF TABLES

<i>Number</i>	<i>Page</i>
2.1 Results of bulk density timeseries generation for each model.	41
2.2 Performance of M at varying resolutions for z timeseries generation. .	47
2.3 Hyperparameter results.	61
2.4 Performance of the models (for SWE and z) and parameterizations (for z) in this study across testing sites.	62
2.5 Statistical significance testing of model RMSE at validation and test- ing sites.	62
2.6 Time and memory benchmarking of all models (for z and SWE) and parameterizations (for z).	63
3.1 Summary of data and variables.	86
3.2 Parameterizations compared in this study.	89
3.3 Parameterization scores over as-published parameterizations.	109
3.4 Parameterization scores over calibrated parameterizations.	110
3.5 Hyperparameter results.	111
4.1 Site Performance Statistics.	146
4.2 z and S performance statistics.	149
4.3 Snow Cover performance statistics.	151
4.4 Surface and Snow-Only albedo statistics.	153
4.5 Latent and Sensible Heat Flux statistics.	156
4.6 Upward shortwave and longwave radiation statistics.	157
4.7 Summary of simulation variable results.	158
4.8 Snowpack timing differences between model configurations.	161
4.9 Global Interannual Variability (IAV) in snow depth and snow albedo by each simulation and ERA5L.	165
4.10 Parameters for the Bulk Snowpack Anderson Parameterization. . . .	173
4.11 Parameters for HTESSSEL Albedo parameterization.	174

Chapter 1

INTRODUCTION

1.1 Problem Definition

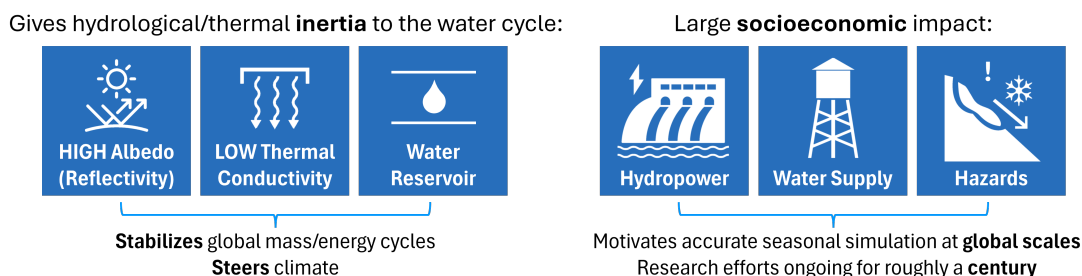


Figure 1.1: **Overview of Snow on the Land Surface.** Snow represents a critical component of the land surface, and is one of the areas most in need of improvement in land surface modeling.

Snow, or the accumulated precipitation of small ice crystals, constitutes a significant component of the terrestrial cryosphere and hydrological cycle. As a cold and reflective source of frozen water, snow acts as a regulator on the Earth’s climate system and a hydrological reservoir which influences global ecosystems and human civilization at large. At high latitudes and mountainous regions, snow accumulates and stores water during the winter months, which is either accumulated in glaciers or slowly released as it melts, which sustains and stabilizes river flow and water availability throughout the year as well as over super-annual timescales [80]. Snow is more reflective than other land cover types and has a lower thermal conductivity, which insulates the underlying ground, cools the surrounding atmosphere, and reflects solar radiation away from Earth that would otherwise warm the terrestrial surface [8]. Overall, snow conceptually provides a hydrological and thermal inertia to the land surface, which serves as a buffer and stabilizing agent to otherwise faster cycles in the global water energy and mass balance (see Fig. 1.1).

The mechanisms of snow presence further influence the biosphere, impacting soil microbial activity and plant life cycles, with downstream effects for agriculture, tourism, and other human activity [67]. Snow also serves as a hazard with impacts on traffic, housing, and larger scale natural disasters [39, 67]. Particularly in the western United States, snow serves an important role in society as a power source in addition to a source of water, through hydroelectric power generation and providing

a majority of the region’s water supply [9, 20]. These numerous socioeconomic impacts have spurred roughly a century of study in the tracking and forecasting of snow presence and its seasonal dynamics in ever-more quantitative ways, especially as its impact on the global climate has become apparent [5]. Accurate representation of snow processes is essential not only for hydrological prediction, but also for reliable simulation of climate sensitivity and future climate change.

Despite ongoing efforts, snow modeling remains an open field of research in hydrological and climate sciences, and is understood to be one of the areas most in need of improvement in land surface modeling [65]. The inertial effects of snow in the cryosphere leave the climate sensitive and susceptible to the creation and persistence of snow, yet the climate-scale circulation patterns which dictate precipitation are also the very mechanisms by which snow is created [24]. Radiative feedbacks from snow reflection (see Fig. 1.2) can propagate from local and watershed levels to entire hemisphere circulation patterns [38], or amplify the effects of cooling or warming at extreme latitudes in a phenomenon known as polar amplification [66]. Such deeply coupled feedback loops mean that errors in simulated snow will diffuse and scale into large-scale biases in surface temperature and atmospheric circulation, with downstream influence on simulations of water availability, hazard prediction, and ultimately regional climate evolution. Improving representation of snow in large-scale models is therefore a problem of both scientific and societal significance.

1.2 Representing Snow

Overview

Most of the properties relevant to snow and its impact on society and climate are bulk properties (macroscopic scales), such as the depth of the snow, the amount of water in the snow, its reflectivity, or its temperature. In the larger-scale modeling found in global climate models (GCMs), this is further abstracted to focusing on constraining the bulk energy and mass content of water stored as snow for consideration of the mass and energy budget of hydrosphere — however, the rates at which mass and energy exchange occur between the cryosphere and other model components are still dictated by these human-scale bulk properties [14].

Representing these properties constitutes a challenging combination of scales (Fig. 1.2). At the smallest levels, quantum interactions dictate the shape of the water molecule that enables the complex geometries of snowflakes, and molecular diffu-

Large-scale modeling difficulty sources from **strong feedbacks** and **scale mismatch**.

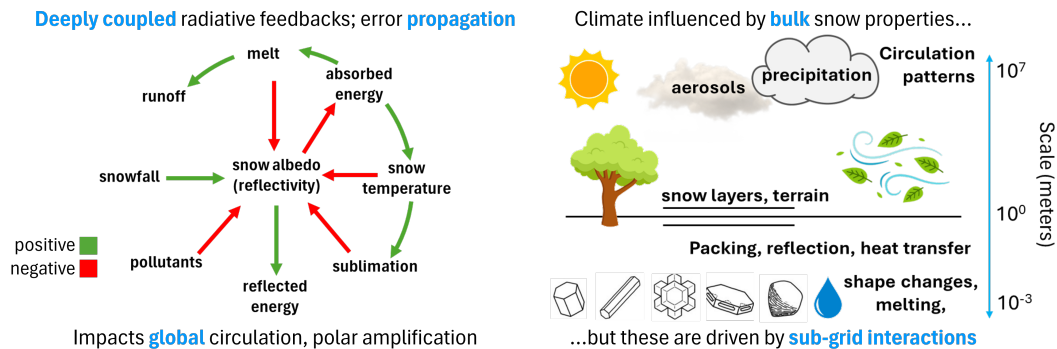


Figure 1.2: **Diagram of considerations of snow modeling.** Conceptual difficulty in representing snow at large scales stems from two particular issues, 1) its deep coupling to the radiative energy flow throughout the earth system, and 2) the multi-scale nature of snow.

sion enables the growth of snowflakes in a variety of shapes, with most snowflakes achieving a diameter on the order of a few millimeters by the time they reach the ground [53]. Their shape influences their packing properties, and thus the overall snow density (and depth, when the mass is known). Mie scattering dictates how each snowflake shape reflects radiation, which accumulates into how the bulk snow surface reflects radiation. Each shape and snow type has its own thermal properties which dictate how heat permeates the snowpack, which governs melting and liquid percolation. This drives the metamorphosis of snow grains into new shapes with new properties as the snow ages and accumulates deposited pollution, impurities, and aerosols [5, 81]. On the other hand, snow layers can form that vary over short distances and timescales which remain sensitive to local topography and wind patterns [59]. Each of these layers exhibits their own bulk thermal and optical properties, and the presence of vegetation or cloud cover can convolve with these effects, linking processes across a variety of spatial scales and land processes. Ultimately, the atmospheric properties that drive these microphysical processes within a snowpack are themselves driven by snow presence at much larger characteristic scales [8] — full physical modeling of snow over long time-scales would thus require simultaneous representation of processes spanning nearly 10 orders of magnitude.

No input data resolution nor computational bandwidth presently exists to resolve and track all these properties within their physically relevant regimes and their heterogeneity, especially when running simulations over a global domain. This intractable mismatch between the spatial and temporal scales at which snow processes occur and at which they must be represented in models constitutes a fundamental

source of uncertainty. Some parameterization, assumptions, and coarse-graining must occur for snow modeling to remain tractable, which can introduce systematic and resolution-dependent biases [62].

Within the snow modeling community, the breadth of modeling choices with respect to variable parameterization is expansive, with very few definitive standards having emerged over decades of research [54, 56]. However, all fundamentally track the dynamics and conservation of water mass and energy within the snowpack, and therefore also across the hydrological modeling domain. Water mass conservation obeys

$$\frac{dSWE}{dt} = P - R + S - E \quad (1.1)$$

where Snow Water Equivalent¹ (SWE) represents the total mass of water in the snowpack, P represents water sourced from precipitation from the atmosphere, R represents a mass sink of water to the land surface due to snowmelt and water loss, E represents mass gain (or loss) from vapor transfer like evaporation or sublimation, and S represents environmental source terms like water upwelling or blowing snow from nearby snowpacks. Each term has units of mass flux (mass per unit area of snow per unit time), though conventionally they are often represented via normalization against the density of liquid water (ρ_l , 1000 kg m^{-3}), allowing SWE to be interpreted as a single height (just like snow depth), and the fluxes as a contributed change in height per unit time. The total energy (U) conservation equation dictates

$$\frac{dU}{dt} = R_{net} + Q_P - Q_R + Q_S - Q_E + Q_{SHF} \quad (1.2)$$

where each of the sources mentioned in mass conservation can add or detract energy from the snowpack (advection or kinetic or potential energy from precipitation, latent and convective energy from evaporation/sublimation, conductive fluxes from the ground surface, water and energy lost in runoff, etc.), in addition to the net radiative energy R_{net} (usually split into shortwave and longwave) and sensible heat (SHF) contributions to the snowpack. This equation is often presented per unit volume and then integrated over the snow depth (usually denoted z) to yield an energy-per-area equation where the right-hand side components of Eq. 1.2 have energy flux units. Some implement this mass and energy system as one bulk equation for the entire snowpack at once, or implement a number of snow layers with a coupled set of conservation ordinary differential equations (ODEs) per

¹For Chapter 4, SWE will instead be denoted as S in accordance with its associated model convention, though we avoid that here due to its usage throughout the rest of the dissertation.

layer [49, 79], which allows for additional modeling of diffusive mass and energy fluxes between layers. From here, different assumptions and modeling choices enable the representation of different bulk, intrinsic, or microphysical properties to quantify these different mass and energy fluxes. Detailed models account for more complexity and processes but face notable computational scaling issues when scaled beyond single-site applications. Single bulk models instead capture aggregated statistical behavior well and can be applied globally, but often fail to exhibit sufficient variability [1].

Large-Scale Representation of Snow: Climate Models

The interaction of snow with so many other environmental processes requires snow to be simulated in tandem with other systems (or driven by representative data of these systems) to properly forecast snow effects on larger spatial or temporal scales [40]. Likewise, evaluating a snow model for its holistic efficacy requires benchmarking over a large range of possible climate types and timescales. Climate models provide a leading option for simulating the coupled dynamics of the cryosphere and atmosphere in this manner. Unlike weather prediction models, which are utilized to constrain local estimates over short time periods, climate models aim to represent the statistical behavior of a larger system over decades to centuries [13], with less sensitivity to initial conditions. This distinction emphasizes different features of model development, as climate model choices must balance achieving physical realism and generalizability against maintaining computational efficiency and (for more empirical models) long-term stability. The horizontal resolution of contemporary climate models is on the order of tens to hundreds of kilometers, which precludes explicit representation of the typical scales of snow heterogeneity [21]. This necessitates the use of parameterizations for approximating subgrid-scale variability, or the coarse-graining of the necessary physics.

When extending snow conservation equations to domains where each modeling cell dwarfs potential snow extent, a snow cover fraction² is often parameterized to quantify subgrid spatial variability in snow cover. When included, it serves to modify Eq. 1.1 and Eq. 1.2 (not shown here) by multiplying non-snow-precipitation mass and energy fluxes [21]. Models can also take advantage of the fundamental relation of bulk/effective snow density ($\overline{\rho_s}$) to z and SWE:

$$\rho_l \text{SWE} = \overline{\rho_s} z . \quad (1.3)$$

²This variable is usually denoted as f , or σ .

This means that to derive bulk density or depth, one only need pick a parameterization for the other variable (see Chapter 2). This can be applied for the whole snowpack or within a single layer, and the derived value used in other physical equations or parameterizations.

Another common approach in larger-scale modeling involves neglecting the contribution of internal air and impurities to the heat capacity of the snowpack. When this assumption is combined with the choice to model the liquid fraction of snow (i.e., what fraction of total SWE is ice versus how much is water), this enables simple diagnosis of a snow temperature ($\bar{T}$) in relation to the snow energy:

$$U = \rho_l \text{SWE} \left[(q_l c_{p,l} + (1 - q_l) c_{p,i}) (\bar{T} - T_0) - (1 - q_l) L_f \right] \quad (1.4)$$

where q_l represents the fraction of SWE that is in liquid instead of solid (ice) form, T_0 represents a reference temperature (the triple point of water, 273.16 K, or 0°C), L_f represents the latent heat of fusion of water (334 J/g at standard temperature and pressure), and $c_{p,l}, c_{p,i}$ represent the heat capacity of water and ice at constant pressure, respectively. Like bulk density or depth, this can be derived per-layer or once for the whole snowpack.

In larger scale models, constraining the *surface* energy balance also carries elevated importance, as it plays a central role in determining snow temperature and ensuing phase changes [79]. Two critical variables in the surface energy balance are the snow surface temperature (T_{surf}), which quantifies the rate of heat flux into the snowpack from surface interactions, and the snow surface albedo (α), which quantifies the reflectivity of the snow surface to incoming radiation. Both are integral to formulating the energy fluxes found in Eq. 1.2 like R_{net}, Q_{SHF}, Q_E , and must frequently be parameterized from available simulation information. Derived temperatures like $\bar{T}$ from Eq. 1.4 are often minimally necessary for a bulk model in constraining α, T_{surf} , or surface energy fluxes, though such simplistic approaches often lack the expressive capacity necessary to portray their variability across different snow environments [21, 79]. This is especially noticeable in the snow-albedo feedback [69], where small relative reductions in snow reflectivity cause larger relative increases in absorbed radiation, which amplifies further surface warming. For surface temperature, failing to capture the full range of expected variability can instead lead to slower energy accumulation and delayed melt.

Bulk formulations like the above provide a tractable way to represent snow in large-scale models, but they necessarily obscure the complexity of the underlying physical

processes. Many important behaviors emerge from smaller-scale interactions, raising challenges for parameterization and model fidelity. Much of this thesis will be devoted to efforts in permitting more faithful representation without exorbitant need of additional physical insight or information, for primarily snow depth (Chapter 2) and snow albedo (Chapter 3), as well as surface temperature (Chapter 4).

1.3 History and Context of Snow Modeling

Quantitative snow modeling as a standalone discipline is a relatively newer field, around 60 years old, though its constituent elements have existed for much longer. The Scientific Revolution brought about the invention of more direct measurements for meteorology like precipitation, snow depth and temperature in the late 1600s [86], introducing quantitative environmental science. The idea of snow albedo and its impact on the quantification of Earth's albedo was recognized as early as the conceptualization of albedo itself in the *Photometria* in 1760 by Johann Lambert [48]. Its comparatively high value compared to other materials was grasped but debated without quantification [12, 75], with explicit measurement only appearing singularly by Zöllner in 1865 [85] prior to the 1900s, mostly for quantifying albedo of astronomical bodies and not Earth. The climate definition of albedo³ was also not formalized until around this time [12].

The theories of radiative balance were developed concurrently to these concepts throughout the 1800s [35, 68]. This culminated in the application of albedo and the Stefan-Boltzmann Law in radiative balance to earth modeling by Arrhenius in 1896 [6]; recognized as the inception of energy-balance climate modeling [27]. Initial observation-based estimates of Earth's albedo occurred in 1901 and 1908 [45, 61] using Zöllner's snow measurements, after which more investigations to observe and characterize snow properties begin to appear globally [17, 44] in the ensuing decades. The equations for turbulent energy transfer and numerical weather modeling are developed during this period [47, 72], though any explicit application of these theories to snow does not appear until 1935 [73] and more theoretical snow inquiry stalls after this, due to the analytic intensity of climate calculations by hand and World War II.

During this period, the U.S. Weather Bureau and Army Corps of Engineers (ACE) began to independently organize extensive snow monitoring and characterization in

³Lambert's initial conceptualization of albedo applied well to diffuse reflection from rough surfaces, but required extension to more rigorously define the total reflectivity fraction of all incident radiation. This form of albedo is formally known as the *Bond Albedo*.

the 1920s and 1930s from the recognition of snow's importance in water storage and hydropower generation, as well as for military interests [2]. Their studies, alongside the advent of the digital computer in the 1950s, spurred a resurgence in numerical environmental prediction efforts for the purposes of hydrological forecasts. This led to the introduction of empirical snow subroutines inside hydrology models which estimated snow behavior using temperature indices, an idea first proposed (but not quantified) in 1915 [44, 55].

It is from these nascent parameterizations inside groundbreaking programs like the Stanford Watershed Model [18] and ACE-commissioned research [26] that the field of snow modeling first emerged standalone, as well as renewed interest in snow physical properties. Foundational researchers like Eric Anderson began combining these parameterizations and older theories with new [46] and self-derived [4] theoretical works to improve snow forecast accuracy, first at site and watershed scales before growing towards global scales as computational capability expanded. The recognition of the need for more rigorous physical foundations was embraced by the end of the 1960s, alongside the emergence of remotely sensed data from satellites. Formative mechanical, thermal, and optical physical research [5, 77] as well as efficient parameterization efforts expanded synergistically with earth modeling and observational network construction in the 1970s and 1980s, as the mass and energy effects of snow on the earth system became better understood. Research since the 1990s has focused more on incorporation of remotely-sensed data to constrain sub-grid snow coverage or albedo [60], characterization of snow microstructure [11], and development of more complex explicit process-based models for snow modeling [51, 74], primarily for avalanche research, hydrology studies, and climate models, as data volume and computational capability continues to expand.

Present snow modeling efforts have access to much more observational data to tune their formulations compared to their predecessors, but few large and coordinated observational efforts exist. Most high-quality data comes from a handful of small individual site measurements in the US or Europe over a number of decades [56]. Remote data can also be utilized, which has the benefit of global coverage, but consists of coarse resolution and high-uncertainty values indirectly derived from satellite radiation band ratios [76]. Few intermediaries between these extremes exist, other than the United States Snow Telemetry (SNOTEL) observation network or NASA's Airborne Snow Observatory [63], which are biased towards coverage in the western United States. Sources like remote sensing can suffer obfuscation and lack

the resolution to capture localized snow processes or internal snow structure, while ground sensors can capture local effects but face challenges like malfunctions, terrain biases, and sparse/uneven coverage, especially in remote or mountain environments [8, 22, 43, 58].

Alternatively, climate model outputs can be integrated with observations to offer an additional source of continuous reanalysis data to complement observations. However, they inherit and propagate the uncertainty of their observational basis, and combine this with the biases inherent in model assumptions. Incorporating these sources to inform large-scale modeling can result in circular bias instead of independent improvement [10]. The difficulty in explicitly measuring several snow properties as well as a lack of widespread conventions on measurement type and quality control contribute to continuing observational challenges [14, 71]. A limited set of variables exists from which to develop parameterizations [7, 71] or provide inputs to models, with varying degrees of uncertainty and inconsistency against other data products [41].

1.4 Contemporary Snow Models

Present snow models populate a spectrum of physical versus empirical parameterizations. Simple empirical models from historical hydrological works still persist in modern studies, which are computationally efficient and require minimal inputs, maximizing their usability across a variety of applications. However, they fail to resolve internal snow structure, and their underlying mechanism often fails to express sufficient output variability [1] to apply to all environmental conditions. Their singular tuning for hydrology applications (i.e., optimizing only SWE and runoff reproduction) also leads to inadequate energy tracking [14, 74].

Detailed process-based physical models have now been developed [51, 74] to improve the energy balance issues of simple schemes, representing multiple snow layers and phenomena like snow grain metamorphosis, aerosol impact on albedo, or vertical percolation of liquid within the snowpack [31]. These models provide more fine-tuned control and physical realism, but are comparatively sensitive to required inputs that are unavailable or exhibit much higher uncertainty at GCM scales, throttling their functional accuracy. Their underlying terms are often themselves parameterizations, which lack direct physical observability and must be calibrated against sparse or indirect measurements, or instead incorporate additional assumptions which impact overall model bias. This leads to systematic errors when applied

beyond the site for which they were calibrated, sometimes even being outperformed by their simpler counterparts [30]. More importantly, their increased computational load cannot always scale for global simulations [29, 42].

The range of parameterization options for a given process also varies. Almost all widespread snow models within GCMs use some form of the snow density parameterization developed decades ago by Anderson [14], with little difference between them other than fine-tuning parameter values. On the other hand, dozens of snow albedo formulations have been proposed, though none have been able to stand out among their competitors under a comprehensive range of simulation environments and conditions [29, 54, 64]. Similar variation exists for surface temperature, thermal conductivity, or other bulk properties.

Generalizability, stability, and scalability continue to remain key issues in contemporary snow modeling. Both simpler and more complex models struggle to maintain similar performance across both global scales and individual site simulations for a given calibration [52]. Persistent biases in snow cover extent, the decay of snow albedo during the melt-out period, and particularly the timing of snowpack disappearance [41] indicate that important processes remain inadequately captured. Scale-dependent uncertainties in data due to the aggregation of sub-grid variability in topography, vegetation, and meteorological input data only complicate such development and evaluation efforts. Furthermore, snow model input sensitivity is exaggerated when coupled within a GCM, as simulation error from the bias of other external systems are additionally propagated into snow model inputs.

Taken together, these core issues highlight a central theme in modern snow modeling: **the need to represent complex, small-scale processes within computationally efficient, large-scale frameworks**. This theme will only become more apparent among contemporary pushes to increase the resolution of GCMs to resolve sub-grid variability [3, 67, 78], as this further constrains available computational resources while necessitating scheme performance against the inherent variability in small-scale physical processes.

1.5 Emerging Paradigms: Machine Learning

Continued growth of data volume and computational power throughout the 2000s and 2010s across scientific industries revitalized previous investigation into artificial intelligence (AI) and other statistical prediction methods, like machine learning (ML; [50]). ML methods permit the ability to summarize and encapsulate under-

Machine Learning (ML) is an **emerging** but **nascent** technology in snow modeling.

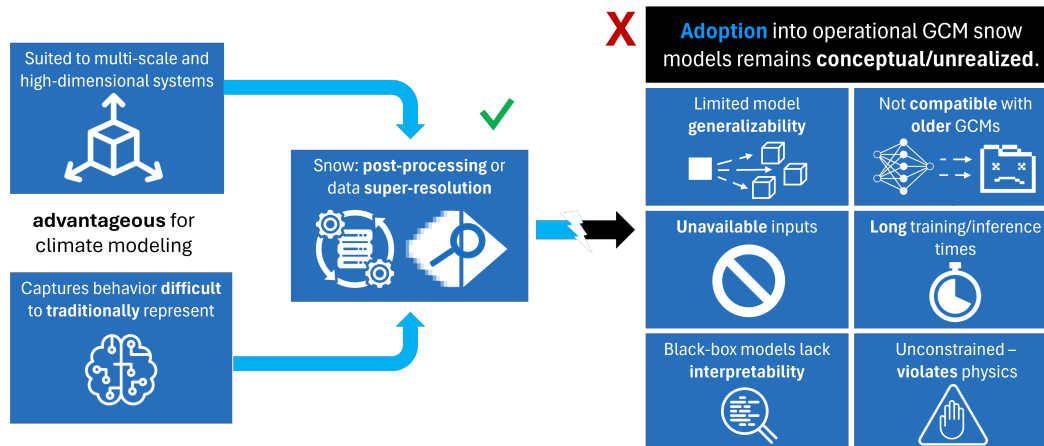


Figure 1.3: Machine Learning (ML) in snow modeling. ML is showing promise in constraining or creating new data sources in contemporary snow modeling, but has yet to be operationally implemented within a large-scale global climate model. This is due to key issues surrounding conceptual and computational issues with ML approaches, primarily in enabling physical consistency of models while preserving physical consistency, as well as in minimizing computational overhead.

lying behavior directly or indirectly from provided data, without presupposing or prescribing a functional form. They more easily represent nonlinear interactions between inputs and have been shown to be more robust to input uncertainty, as well as more flexible than other empirical methods to data nonstationarity, or the tendency of a timeseries' statistics to shift over time [19, 22, 34]. They are especially well-suited to multi-scale and high-dimensional systems with large data volume, which makes them uniquely advantageous for climate modeling applications.

Within the context of snow modeling, ML offers several potential advantages. Data-driven methods can capture complex relationships that are difficult to represent with traditional parameterizations. They can also leverage diverse or even multi-modal datasets, including remote sensing products and reanalysis outputs, to learn patterns across different spatial and temporal scales. Numerous studies now exist applying ML techniques to problems in climate sciences like data interpolation, sub-grid estimation and super-resolution, as well as enhanced masking or even short-timescale forecasting of snow bulk properties [22], though their extension to longer-term future climate domains remains limited.

Adoption of ML into operational climate applications has remained limited despite widespread consensus of their performance potential in literature [32]. A well-cited limitation to purely data-driven methods is their unconstrained black-box nature,

which can lack physical interpretability and produce outputs that violate known physical constraints [33, 36], restricting their utility within coupled models or those requiring long-term stability. Another practical limitation is that the older and less modular code frameworks (upon which many climate models are built) do not have libraries that can support most modern ML architectures [14, 37, 57]. Another common theme across ML literature in the climate sciences is long training and inference times, which can throttle prototyping and development in research applications, limit willingness to implement an ML scheme in an operational model that could potentially require future recalibration, or bottleneck computational resources in an already busy computational environment [16, 25].

These considerations have motivated investigation into hybrid physics-ML approaches, which seek to complement existing physical insight with a range of ML techniques [28, 70, 82, 83]. These approaches can enable more physically constrained modeling of conserved quantities, which can aid model generalization to future climates. Hybrid frameworks have been explored for hydrological modeling in general, but their extension to snow parameterization specifically remains a limited area of research. Many models require inputs which preclude their usage in a forecasting environment, where inputs like satellite band ratios at future times would be unavailable [15, 23]. As such, much of the existing literature in these approaches applies to data or model output post-processing, especially for remotely-sensed data [15, 23, 84]. Similar to other applications of ML in climate, concerns of large training/inference resources, system portability, generalizability, and physical consistency remain open questions regarding global-scale snow modeling.

1.6 Thesis Outline

This thesis addresses the challenge of improving snow parameterizations in large-scale climate forecasting through the development and application of a hybrid physics-ML parameterization framework. The framework guarantees compliance with prescribed physical bounds while maintaining generalizability and nonlinear expressivity for small black-box models. This approach enables the benefits of emergent ML techniques within the field while ensuring physical consistency, creating models that can run inexpensively within a GCM setting. It can demonstrably outperform other snow parameterizations both when compared directly as well as when compared within the more uncertain environment of a climate model. These improvements likewise extend from the simulated variables themselves to downstream interactions like melt timing, progressing in longstanding snow modeling

hurdles. Its performance extends not only across climate types, but across a range of spatiotemporal scales as well, matching or outperforming more complex and state-of-the art alternatives.

Chapter 2 introduces the parameterization framework, first conceptualized in 2022, which is used to construct a parameterization for snow depth and compared to the longstanding Anderson parameterization. The developed ML parameterization is benchmarked against the leading alternative at over 40 different locations across the globe. It discusses the computational benefits of the framework in detail, as well as exhibiting the benefits of the physically-bound scheme in comparison to applying the same constraints *post hoc* to a purely ML-based scheme which is trained without constraints.

Chapter 3 extends the developed framework to address present challenges in the field of snow albedo modeling within GCMs. The developed albedo parameterization is (to the author's knowledge) the first ML parameterization of snow albedo capable of operation embedded within a climate model, and is able to outperform over 20 established parameterizations within the field when assessed over 700 locations at coarse resolution and 30 site-level locations.

Chapter 4 introduces a non-ML surface temperature parameterization to complement the two neural parameterizations, within the context of the integration of the depth and albedo parameterizations into the ClimaLand snow model. The surface temperature parameterization is still built upon the same principles of incorporating a few physical insights to an otherwise expressive yet lightweight structural form. All the new parameterizations are then benchmarked within a coupled setting over global and site-level simulations, in comparison against the initial ClimaLand parameterizations as well as another simulation run using parameterizations implemented in prevalent GCMs. The new parameterizations demonstrate an ability to improve upon a number of simulation variables, as well as improve snow seasonality metrics.

Finally, Chapter 5 synthesizes the findings of the dissertation, discusses their implications, and outlines directions for future research.

Together, the presented studies provide an exhibition of a powerful parameterization technique that is adaptive to future data streams and modeling needs as snow modeling and climate modeling evolve. They highlight the potential and challenges of hybrid data-driven frameworks, underscoring the importance of continued research

and observation at the intersection of applied physics and machine learning in the earth sciences. Constrained, data-driven parameterizations can serve as effective and efficient alternatives to traditional empirical or highly complex process-based schemes, both in global-scale snow modeling and applied physical parameterization beyond.

References

- [1] R. Abolafia-Rosenzweig, C. He, S. McKenzie Skiles, F. Chen, and D. Gochis. 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), Sept. 2022. doi: 10.1029/2022ms003141. URL <http://dx.doi.org/10.1029/2022MS003141>.
- [2] ACE. Snow hydrology: Summary report of the snow investigations. published by the North Pacific division, Corps of Engineers, U.S. Army, Portland, Oregon, 1956. *Journal of Glaciology*, 3(22):148–148, 1956. doi: 10.3189/s0022143000024503. URL <http://dx.doi.org/10.3189/S0022143000024503>.
- [3] F. Aguirre, D. Bozkurt, T. Sauter, J. Carrasco, C. Schneider, R. Jaña, and G. Casassa. 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, Nov. 2023. doi: 10.3390/rs15225430. URL <http://dx.doi.org/10.3390/rs15225430>.
- [4] E. A. Anderson. Development and testing of snow pack energy balance equations. *Water Resources Research*, 4(1):19–37, Feb. 1968. ISSN 1944-7973. doi: 10.1029/wr004i001p00019. URL <http://dx.doi.org/10.1029/WR004i001p00019>.
- [5] E. A. Anderson. *A point energy and mass balance model of a snow cover*. NOAA technical report ; NWS 19. U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, Office of Hydrology, Silver Spring, Md, 1976. URL <https://repository.library.noaa.gov/view/noaa/6392>.
- [6] S. Arrhenius and E. S. Holden. On the influence of carbonic acid in the air upon the temperature of the earth. *Publications of the Astronomical Society of the Pacific*, 9(54):14–24, 1896. ISSN 00046280, 15383873. URL <http://www.jstor.org/stable/40670917>.
- [7] E. H. Bair, R. E. Davis, and J. Dozier. Hourly mass and snow energy balance measurements from Mammoth Mountain, CA USA, 2011–2017. *Earth System Science Data*, 10(1):549–563, Mar. 2018. doi: 10.5194/essd-10-549-2018. URL <http://dx.doi.org/10.5194/essd-10-549-2018>.
- [8] E. H. Bair, K. Rittger, S. M. Skiles, and J. Dozier. An examination of snow albedo estimates from MODIS and their impact on snow water equivalent reconstruction. *Water Resources Research*, 55(9):7826–7842, Sept. 2019. doi: 10.1029/2019wr024810. URL <http://dx.doi.org/10.1029/2019WR024810>.

- [9] T. P. Barnett, J. C. Adam, and D. P. Lettenmaier. Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, 438(7066): 303–309, Nov. 2005. ISSN 1476-4687. doi: 10.1038/nature04141. URL <http://dx.doi.org/10.1038/nature04141>.
- [10] L. Bengtsson, S. Hagemann, and K. I. Hodges. Can climate trends be calculated from reanalysis data? *Journal of Geophysical Research: Atmospheres*, 109 (D11), June 2004. ISSN 0148-0227. doi: 10.1029/2004jd004536. URL <http://dx.doi.org/10.1029/2004JD004536>.
- [11] A. Bernard, P. Hagemuller, M. Montagnat, and G. Chambon. Disentangling creep and isothermal metamorphism during snow settlement with x-ray tomography. *Journal of Glaciology*, 69(276):899–910, Dec. 2022. ISSN 1727-5652. doi: 10.1017/jog.2022.109. URL <http://dx.doi.org/10.1017/jog.2022.109>.
- [12] G. P. Bond. On the light of the moon and of the planet Jupiter. *Memoirs of the American Academy of Arts and Sciences*, 8(1):221–286, 1861. ISSN 00966134. URL <http://www.jstor.org/stable/25057970>.
- [13] A. Bracco, J. Brajard, H. A. Dijkstra, P. Hassanzadeh, C. Lessig, and C. Monteleoni. Machine learning for the physics of climate. *Nature Reviews Physics*, 7(1):6–20, Nov. 2024. ISSN 2522-5820. doi: 10.1038/s42254-024-00776-3. URL <http://dx.doi.org/10.1038/s42254-024-00776-3>.
- [14] A. Charbonneau, K. Deck, and T. Schneider. A physics-constrained neural differential equation framework for data-driven snowpack simulation. *Artificial Intelligence for the Earth Systems*, 4(3), July 2025. doi: 10.1175/aies-d-24-0040.1. URL <http://dx.doi.org/10.1175/AIES-D-24-0040.1>.
- [15] S. Chen, P. Xiao, X. Zhang, P. Pellikka, H. Liu, and Y. Liu. 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, Oct. 2025. doi: 10.1016/j.isprsjprs.2025.07.033. URL <http://dx.doi.org/10.1016/j.isprsjprs.2025.07.033>.
- [16] L.-A. Chevrollier, A. Wehrlé, J. M. Cook, N. Pirk, L. G. Benning, A. M. Anesio, and M. Tranter. Separating the albedo-reducing effect of different light-absorbing particles on snow using deep learning. *The Cryosphere*, 19 (4):1527–1538, 2025. doi: 10.5194/tc-19-1527-2025. URL <https://tc.copernicus.org/articles/19/1527/2025/>.
- [17] S. Colbeck. History of snow-cover research. *Journal of Glaciology*, 33(S1): 60–65, 1987. ISSN 1727-5652. doi: 10.3189/s0022143000215839. URL <http://dx.doi.org/10.3189/S0022143000215839>.
- [18] N. Crawford, R. Linsley, and S. U. D. of Civil Engineering. *Digital Simulation in Hydrology: Stanford Watershed Model IV*. Stanford University. Department

- of Civil Engineering: Technical report. Department of Civil Engineering, Stanford University, 1966. ISBN 9780598532480. URL <https://books.google.com/books?id=KOQDAAAAIAAJ>.
- [19] A. Damiani, N. N. Ishizaki, S. Feron, and R. R. Cordero. 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, Feb. 2025. doi: 10.1016/j.scitotenv.2025.178606. URL <http://dx.doi.org/10.1016/j.scitotenv.2025.178606>.
 - [20] C. De Michele, F. Avanzi, A. Ghezzi, and C. Jommi. Investigating the dynamics of bulk snow density in dry and wet conditions using a one-dimensional model. *The Cryosphere*, 7(2):433–444, 2013. URL <https://doi.org/10.5194/tc-7-433-2013>.
 - [21] K. Deck, R. K. Braghieri, A. A. Renchon, J. Sloan, G. Bozzola, E. Speer, J. Ben Mackay, T. Reddy, K. Phan, A. L. Gagné-Landmann, Y. Li, D. Yatunin, A. Charbonneau, N. Efrat-Henrici, E. Bach, S. Ma, P. Gentine, C. Frankenberg, A. A. Bloom, Y. Wang, M. Longo, and T. Schneider. ClimaLand: A land surface model designed to enable data-driven parameterizations. *Journal of Advances in Modeling Earth Systems*, 18(1), Jan. 2026. ISSN 1942-2466. doi: 10.1029/2025ms005118. URL <http://dx.doi.org/10.1029/2025MS005118>.
 - [22] G. Demil, A. T. Haghighi, B. Klöve, and M. Oussalah. Advances in image-based estimation of snow variable: A systematic literature review on recent studies. *Journal of Hydrology*, 654:132855, June 2025. doi: 10.1016/j.jhydrol.2025.132855. URL <http://dx.doi.org/10.1016/j.jhydrol.2025.132855>.
 - [23] A. Ding, S. Liang, H. Ma, T. He, A. Jia, and Q. Wang. Improved estimation of daily blue-sky snow shortwave albedo from MODIS data and reanalysis information. *Science of Remote Sensing*, 10:100163, Dec. 2024. doi: 10.1016/j.srs.2024.100163. URL <http://dx.doi.org/10.1016/j.srs.2024.100163>.
 - [24] G. T. Diro and L. Sushama. Snow–precipitation coupling and related atmospheric feedbacks over North America. *Atmospheric Science Letters*, 19(8): e831, 2018. doi: <https://doi.org/10.1002/asl.831>. URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/asl.831>.
 - [25] S. Duan, P. Ullrich, M. Risser, and A. Rhoades. Using temporal deep learning models to estimate daily snow water equivalent over the Rocky Mountains. *Water Resources Research*, 60(4):e2023WR035009, 2024. doi: <https://doi.org/10.1029/2023WR035009>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023WR035009>.
 - [26] R. V. Dunkle and J. T. Bevans. An approximate analysis of the solar reflectance and transmittance of a snow cover. *Journal of Meteorology*, 13(2):212–216,

- Apr. 1956. ISSN 0095-9634. doi: 10.1175/1520-0469(1956)013<0212:aaaots>2.0.co;2. URL [http://dx.doi.org/10.1175/1520-0469\(1956\)013<0212:AAAOTS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0469(1956)013<0212:AAAOTS>2.0.CO;2).
- [27] P. N. Edwards. History of climate modeling. *WIREs Climate Change*, 2(1):128–139, Dec. 2010. ISSN 1757-7799. doi: 10.1002/wcc.95. URL <http://dx.doi.org/10.1002/wcc.95>.
- [28] R. ElGhawi, B. Kraft, C. Reimers, M. Reichstein, M. Körner, P. Gentile, and A. J. Winkler. Hybrid modeling of evapotranspiration: Inferring stomatal and aerodynamic resistances using combined physics-based and machine learning. *Environmental Research Letters*, 18(3):034039, Mar. 2023. ISSN 1748-9326. doi: 10.1088/1748-9326/acbbe0. URL <http://dx.doi.org/10.1088/1748-9326/acbbe0>.
- [29] R. Essery, S. Morin, Y. Lejeune, and C. B. Ménard. A comparison of 1701 snow models using observations from an alpine site. *Advances in Water Resources*, 55:131–148, May 2013. doi: 10.1016/j.advwatres.2012.07.013. URL <http://dx.doi.org/10.1016/j.advwatres.2012.07.013>.
- [30] P. Etchevers, E. Martin, R. Brown, C. Fierz, Y. Lejeune, E. Bazile, A. Boone, Y.-J. Dai, R. Essery, A. Fernandez, Y. Gusev, R. Jordan, V. Koren, E. Kowalczyk, N. O. Nasonova, R. D. Pyles, A. Schlosser, A. B. Shmakin, T. G. Smirnova, U. Strasser, D. Verseghy, T. Yamazaki, and Z.-L. Yang. Validation of the energy budget of an alpine snowpack simulated by several snow models (SnowMIP project). *Annals of Glaciology*, 38:150–158, 2004. doi: 10.3189/172756404781814825. URL <http://dx.doi.org/10.3189/172756404781814825>.
- [31] M. G. Flanner, J. B. Arnheim, J. M. Cook, C. Dang, C. He, X. Huang, D. Singh, S. M. Skiles, C. A. Whicker, and C. S. Zender. SNICAR-ADv3: A community tool for modeling spectral snow albedo. *Geoscientific Model Development*, 14(12):7673–7704, Dec. 2021. doi: 10.5194/gmd-14-7673-2021. URL <http://dx.doi.org/10.5194/gmd-14-7673-2021>.
- [32] S. W. Fleming and A. G. Goodbody. A machine learning metasystem for robust probabilistic nonlinear regression-based forecasting of seasonal water availability in the US West. *IEEE Access*, 7:119943–119964, 2019. doi: 10.1109/ACCESS.2019.2936989. URL <https://ieeexplore.ieee.org/document/8809699>.
- [33] S. W. Fleming, D. C. Garen, A. G. Goodbody, C. S. McCarthy, and L. C. Landers. Assessing the new Natural Resources Conservation Service (NRCS) water supply forecast model for the American West: A challenging test of explainable, automated, ensemble artificial intelligence. *Journal of Hydrology*, 602:126782, Nov. 2021. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2021.126782. URL <http://dx.doi.org/10.1016/j.jhydrol.2021.126782>.

- [34] S. W. Fleming, K. Rittger, C. M. Oaida Taglialatela, and I. Graczyk. 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), Apr. 2024. doi: 10.1029/2023wr035785. URL <http://dx.doi.org/10.1029/2023WR035785>.
- [35] J. Fourier. *Théorie analytique de la chaleur*. Chez Firmin Didot, père et fils, 1822. URL <https://books.google.com/books?id=TDQJAAAAIAAJ>.
- [36] L. Gao, L. Zhang, Y. Shen, Y. Zhang, M. Ai, and W. Zhang. Modeling snow depth and snow water equivalent distribution and variation characteristics in the Irtysh River Basin, China. *Applied Sciences*, 11(18), 2021. ISSN 2076-3417. doi: 10.3390/app11188365. URL <https://www.mdpi.com/2076-3417/11/18/8365>.
- [37] M. García-Rodríguez, J. Rodeiro-Iglesias, and J. A. Añel. Code accessibility and code quality across phases of the models of the coupled model intercomparison project. *EGUsphere*, 2026:1–27, 2026. doi: 10.5194/egusphere-2025-6108. URL <https://egusphere.copernicus.org/preprints/2026/egusphere-2025-6108/>.
- [38] A. Hall. The role of surface albedo feedback in climate. *Journal of Climate*, 17(7):1550–1568, Apr. 2004. ISSN 1520-0442. doi: 10.1175/1520-0442(2004)017<1550:trosaf>2.0.co;2. URL [http://dx.doi.org/10.1175/1520-0442\(2004\)017<1550:TROSAF>2.0.CO;2](http://dx.doi.org/10.1175/1520-0442(2004)017<1550:TROSAF>2.0.CO;2).
- [39] J. C. Hammond, G. A. Sexstone, A. L. Putman, T. B. Barnhart, D. M. Rey, J. M. Driscoll, G. E. Liston, K. L. Rasmussen, D. McGrath, S. R. Fassnacht, and S. K. Kampf. High resolution snow model simulations reveal future elevation-dependent snow loss and earlier, flashier surface water input for the Upper Colorado River Basin. *Earth's Future*, 11(2), Jan. 2023. ISSN 2328-4277. doi: 10.1029/2022ef003092. URL <http://dx.doi.org/10.1029/2022EF003092>.
- [40] F. Hanzer, K. Helfricht, T. Marke, and U. Strasser. Multilevel spatiotemporal validation of snow/ice mass balance and runoff modeling in glacierized catchments. *The Cryosphere*, 10(4):1859–1881, Aug. 2016. ISSN 1994-0424. doi: 10.5194/tc-10-1859-2016. URL <http://dx.doi.org/10.5194/tc-10-1859-2016>.
- [41] D. Hao, G. Bisht, K. Rittger, T. Stillinger, E. Bair, Y. Gu, and L. R. Leung. Evaluation of E3SM land model snow simulations over the Western United States. *The Cryosphere*, 17(2):673–697, Feb. 2023. ISSN 1994-0424. doi: 10.5194/tc-17-673-2023. URL <http://dx.doi.org/10.5194/tc-17-673-2023>.
- [42] C. He, M. Flanner, D. M. Lawrence, and Y. Gu. 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), Feb. 2024. ISSN 1942-2466. doi: 10.1029/2023ms003861. URL <http://dx.doi.org/10.1029/2023MS003861>.
- [43] D. F. Hill, E. A. Burakowski, R. L. Crumley, J. Keon, J. M. Hu, A. A. Arendt, K. Wikstrom Jones, and G. J. Wolken. Converting snow depth to snow water equivalent using climatological variables. *The Cryosphere*, 13(7):1767–1784, 2019. doi: 10.5194/tc-13-1767-2019. URL <https://tc.copernicus.org/articles/13/1767/2019/>.
- [44] R. E. Horton. The melting of snow. *Monthly Weather Review*, 43(12):599–605, Dec. 1915. ISSN 1520-0493. doi: 10.1175/1520-0493(1915)43<599:tmos>2.0.co;2. URL [http://dx.doi.org/10.1175/1520-0493\(1915\)43<599:TMOS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1915)43<599:TMOS>2.0.CO;2).
- [45] H. H. Kimball. Some causes of the variability of earthshine. *Monthly Weather Review*, 29(5):209–211, May 1901. ISSN 1520-0493. doi: 10.1175/1520-0493(1901)29[209:scotvo]2.0.co;2. URL [http://dx.doi.org/10.1175/1520-0493\(1901\)29\[209:SCOTVO\]2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1901)29[209:SCOTVO]2.0.CO;2).
- [46] K. Kojima. Densification of seasonal snow cover. *Physics of Snow and Ice: Proceedings*, 1(2):929–952, 1967. URL <http://hdl.handle.net/2115/20351>.
- [47] T. V. Kármán. Über laminare und turbulente reibung. *Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, 1(4):233–252, Jan. 1921. ISSN 1521-4001. doi: 10.1002/zamm.19210010401. URL <http://dx.doi.org/10.1002/zamm.19210010401>.
- [48] Lambert, Johann Heinrich. *Photometria sive de mensura et gradibus luminis, colorum et umbrae*, chapter §1072, page 499. Sumptibus viduae Eberhardi Klett, 1760. doi: 10.3931/E-RARA-9488. URL <https://www.e-rara.ch/zut/doi/10.3931/e-rara-9488>.
- [49] D. M. Lawrence, R. A. Fisher, C. D. Koven, K. W. Oleson, S. C. Swenson, G. Bonan, N. Collier, B. Ghimire, L. van Kampenhout, D. Kennedy, E. Kluzek, P. J. Lawrence, F. Li, H. Li, D. Lombardozzi, W. J. Riley, W. J. Sacks, M. Shi, M. Vertenstein, W. R. Wieder, C. Xu, A. A. Ali, A. M. Badger, G. Bisht, M. van den Broeke, M. A. Brunke, S. P. Burns, J. Buzan, M. Clark, A. Craig, K. Dahlin, B. Drewniak, J. B. Fisher, M. Flanner, A. M. Fox, P. Gentine, F. Hoffman, G. Keppel-Aleks, R. Knox, S. Kumar, J. Lenaerts, L. R. Leung, W. H. Lipscomb, Y. Lu, A. Pandey, J. D. Pelletier, J. Perket, J. T. Randerson, D. M. Ricciuto, B. M. Sanderson, A. Slater, Z. M. Subin, J. Tang, R. Q. Thomas, M. Val Martin, and X. Zeng. The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, 11(12): 4245–4287, Dec. 2019. ISSN 1942-2466. doi: 10.1029/2018ms001583. URL <http://dx.doi.org/10.1029/2018MS001583>.

- [50] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553): 436–444, May 2015. ISSN 1476-4687. doi: 10.1038/nature14539. URL <http://dx.doi.org/10.1038/nature14539>.
- [51] M. Lehning, P. Bartelt, B. Brown, C. Fierz, and P. Satyawali. A physical SNOWPACK model for the Swiss avalanche warning: Part II. Snow microstructure. *Cold Regions Science and Technology*, 35(3):147–167, 2002. ISSN 0165-232X. doi: [https://doi.org/10.1016/S0165-232X\(02\)00073-3](https://doi.org/10.1016/S0165-232X(02)00073-3). URL <https://www.sciencedirect.com/science/article/pii/S0165232X02000733>.
- [52] L. R. Leung and Y. Qian. The sensitivity of precipitation and snowpack simulations to model resolution via nesting in regions of complex terrain. *Journal of Hydrometeorology*, 4(6):1025–1043, Dec. 2003. ISSN 1525-7541. doi: 10.1175/1525-7541(2003)004<1025:tsopas>2.0.co;2. URL [http://dx.doi.org/10.1175/1525-7541\(2003\)004<1025:TSOPAS>2.0.CO;2](http://dx.doi.org/10.1175/1525-7541(2003)004<1025:TSOPAS>2.0.CO;2).
- [53] J. D. Locatelli and P. V. Hobbs. Fall speeds and masses of solid precipitation particles. *Journal of Geophysical Research*, 79(15):2185–2197, May 1974. ISSN 0148-0227. doi: 10.1029/jc079i015p02185. URL <http://dx.doi.org/10.1029/JC079i015p02185>.
- [54] J. Magnusson, N. Wever, R. Essery, N. Helbig, A. Winstral, and T. Jonas. Evaluating snow models with varying process representations for hydrological applications. *Water Resources Research*, 51(4):2707–2723, Apr. 2015. doi: 10.1002/2014wr016498. URL <http://dx.doi.org/10.1002/2014WR016498>.
- [55] D. G. Marks. *Climate, energy exchange, and snowmelt in Emerald Lake watershed, Sierra Nevada*. Department of Geography and Mechanical Engineering., University of California, Santa Barbara, 1988. pdf from author available at https://www.researchgate.net/publication/258437741_Climate_Energy_Exchange_and_Snowmelt_in_Emerald_Lake_Watershed_Sierra_Nevada.
- [56] C. B. Menard, R. Essery, G. Krinner, G. Arduini, P. Bartlett, A. Boone, C. Brutel-Vuilmet, E. Burke, M. Cuntz, Y. Dai, B. Decharme, E. Dutra, X. Fang, C. Fierz, Y. Gusev, S. Hagemann, V. Haverd, H. Kim, M. Lafaysse, T. Marke, O. Nasonova, T. Nitta, M. Niwano, J. Pomeroy, G. Schädler, V. A. Semenov, T. Smirnova, U. Strasser, S. Swenson, D. Turkov, N. Wever, and H. Yuan. Scientific and human errors in a snow model intercomparison. *Bulletin of the American Meteorological Society*, 102(1):E61–E79, Jan. 2021. doi: 10.1175/bams-d-19-0329.1. URL <http://dx.doi.org/10.1175/BAMS-D-19-0329.1>.
- [57] M. Méndez, J. Overbey, and F. G. Tinetti. Legacy fortran software: Applying syntactic metrics to global climate models. In *XVIII Congreso Argentino de Ciencias de la Computación*, 2012. URL <https://sedici.unlp.edu.ar/handle/10915/23718>.

- [58] J. D. D. Meyer, J. Jin, and S.-Y. Wang. 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, 2012. doi: 10.1175/JHM-D-12-066.1. URL https://journals.ametsoc.org/view/journals/hydr/13/6/jhm-d-12-066_1.xml.
- [59] N. P. Molotch and R. C. Bales. Comparison of ground-based and airborne snow surface albedo parameterizations in an alpine watershed: Impact on snowpack mass balance. *Water Resources Research*, 42(5), May 2006. doi: 10.1029/2005wr004522. URL <http://dx.doi.org/10.1029/2005WR004522>.
- [60] A. W. Nolin. Recent advances in remote sensing of seasonal snow. *Journal of Glaciology*, 56(200):1141–1150, 2010. ISSN 1727-5652. doi: 10.3189/002214311796406077. URL <http://dx.doi.org/10.3189/002214311796406077>.
- [61] S. A. Observatory. *Annals of the Astrophysical Observatory of the Smithsonian Institution*, page 161. Number v. 2 in Harvard astronomy preservation microfilm project. U.S. Government Printing Office, 1908. URL <https://books.google.com/books?id=FcIRAAAYAAJ>.
- [62] N. Oreskes, K. Shrader-Frechette, and K. Belitz. Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, 263(5147): 641–646, Feb. 1994. ISSN 1095-9203. doi: 10.1126/science.263.5147.641. URL <http://dx.doi.org/10.1126/science.263.5147.641>.
- [63] T. H. Painter, D. F. Berisford, J. W. Boardman, K. J. Bormann, J. S. Deems, F. Gehrke, A. Hedrick, M. Joyce, R. Laidlaw, D. Marks, C. Mattmann, B. McGurk, P. Ramirez, M. Richardson, S. M. Skiles, F. C. Seidel, and A. Winstral. The Airborne Snow Observatory (ASO): 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, 2016. ISSN 0034-4257. doi: <https://doi.org/10.1016/j.rse.2016.06.018>. URL <https://www.sciencedirect.com/science/article/pii/S0034425716302577>.
- [64] C. A. Pedersen and J.-G. Winther. Intercomparison and validation of snow albedo parameterization schemes in climate models. *Climate Dynamics*, 25 (4):351–362, July 2005. ISSN 0930-7575. doi: 10.1007/s00382-005-0037-0. URL <http://dx.doi.org/10.1007/s00382-005-0037-0>.
- [65] R. Pirazzini, L. Leppänen, G. Picard, J. I. Lopez-Moreno, C. Marty, G. Macelloni, A. Kontu, A. Von Lerber, C. M. Tanis, M. Schneebeli, P. De Rosnay, and A. N. Arslan. European in-situ snow measurements: Practices and purposes. *Sensors*, 18(7):2016, June 2018. doi: 10.3390/s18072016. URL <http://dx.doi.org/10.3390/s18072016>.

- [66] F. Pithan and T. Mauritsen. Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nature Geoscience*, 7(3):181–184, Feb. 2014. ISSN 1752-0908. doi: 10.1038/ngeo2071. URL <http://dx.doi.org/10.1038/ngeo2071>.
- [67] B. Poschlod and A. S. Daloz. Snow depth in high-resolution regional climate model simulations over southern Germany – suitable for extremes and impact-related research? *The Cryosphere*, 18(4):1959–1981, Apr. 2024. ISSN 1994-0424. doi: 10.5194/tc-18-1959-2024. URL <http://dx.doi.org/10.5194/tc-18-1959-2024>.
- [68] C.-S.-M. Pouillet. *Memoir on Solar Heat: On the Radiant and Absorbing Powers of Atmospheric Air and on the Temperature of Space*. Bachelier, Paris, 1838. URL <http://ark.bnf.fr/ark:/12148/cb37284121j.public>. Bibliothèque nationale de France, ark:/12148/cb37284121j.
- [69] X. Qu and A. Hall. Assessing snow albedo feedback in simulated climate change. *Journal of Climate*, 19(11):2617 – 2630, 2006. doi: 10.1175/JCLI3750.1. URL <https://journals.ametsoc.org/view/journals/clim/19/11/jcli3750.1.xml>.
- [70] M. Raissi, P. Perdikaris, and G. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. ISSN 0021-9991. doi: <https://doi.org/10.1016/j.jcp.2018.10.045>. URL <https://www.sciencedirect.com/science/article/pii/S0021999118307125>.
- [71] M. S. Raleigh, B. Livneh, K. Lapo, and J. D. Lundquist. How does availability of meteorological forcing data impact physically based snowpack simulations? *Journal of Hydrometeorology*, 17(1):99 – 120, 2016. doi: 10.1175/JHM-D-14-0235.1. URL https://journals.ametsoc.org/view/journals/hydr/17/1/jhm-d-14-0235_1.xml.
- [72] L. F. Richardson. *Weather Prediction by Numerical Process*. Cambridge University Press, Cambridge, 1922. Reprinted by Dover Publications, 2007. Archived scan available at <https://archive.org/details/weatherpredictio00richrich/weatherpredictio00richrich/>.
- [73] H. Sverdrup. *The Eddy Conductivity of the Air Over a Smooth Snow Field: Results of the Norwegian-Swedish Spitsbergen Expedition in 1934*. Geofysiske publikationer. Brøgger, 1935. URL <https://books.google.com/books?id=-RpG0AEACAAJ>.
- [74] V. Vionnet, E. Brun, S. Morin, A. Boone, S. Faroux, P. Le Moigne, E. Martin, and J.-M. Willemet. The detailed snowpack scheme Crocus

- and its implementation in SURFEX v7.2. *Geoscientific Model Development*, 5(3):773–791, 2012. doi: 10.5194/gmd-5-773-2012. URL <https://gmd.copernicus.org/articles/5/773/2012/>.
- [75] F. von Zach. *Monatliche Correspondenz zur Beförderung der Erd- und Himmels- Kunde*, page 298. Number v. 8 in *Monatliche Correspondenz zur Beförderung der Erd- und Himmels- Kunde*. Beckerische Buchhandlung, 1803. URL <https://books.google.lk/books?id=BR84AAAAAAAJ>.
- [76] Z. Wang, C. B. Schaaf, M. J. Chopping, A. H. Strahler, J. Wang, M. O. Román, A. V. Rocha, C. E. Woodcock, and Y. Shuai. Evaluation of moderate-resolution imaging spectroradiometer (MODIS) snow albedo product (MCD43A) over tundra. *Remote Sensing of Environment*, 117:264–280, Feb. 2012. doi: 10.1016/j.rse.2011.10.002. URL <http://dx.doi.org/10.1016/j.rse.2011.10.002>.
- [77] W. J. Wiscombe and S. G. Warren. A model for the spectral albedo of snow. I: Pure snow. *Journal of the Atmospheric Sciences*, 37(12):2712–2733, Dec. 1980. doi: 10.1175/1520-0469(1980)037<2712:amftsa>2.0.co;2. URL [http://dx.doi.org/10.1175/1520-0469\(1980\)037<2712:AMFTSA>2.0.CO;2](http://dx.doi.org/10.1175/1520-0469(1980)037<2712:AMFTSA>2.0.CO;2).
- [78] F. Ye, Q. Cheng, W. Hao, D. Yu, C. Ma, D. Liang, and H. Shen. 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, Nov. 2023. doi: 10.1016/j.jag.2023.103519. URL <http://dx.doi.org/10.1016/j.jag.2023.103519>.
- [79] J. You, D. G. Tarboton, and C. H. Luce. Modeling the snow surface temperature with a one-layer energy balance snowmelt model. *Hydrology and Earth System Sciences*, 18(12):5061–5076, Dec. 2014. ISSN 1607-7938. doi: 10.5194/hess-18-5061-2014. URL <http://dx.doi.org/10.5194/hess-18-5061-2014>.
- [80] M. Zaremehrjardy, J. Victor, S. Park, B. Smerdon, D. S. Alessi, and M. Faramarzi. Assessment of snowmelt and groundwater-surface water dynamics in mountains, foothills, and plains regions in northern latitudes. *Journal of Hydrology*, 606:127449, Mar. 2022. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2022.127449. URL <http://dx.doi.org/10.1016/j.jhydrol.2022.127449>.
- [81] Y. Zhang and T. Ohata. Parameterization of albedo on snow/ice surface. A review. *Journal of the Japanese Society of Snow and Ice*, 65(1):33–51, 2003. doi: 10.5331/seppyo.65.33. URL <http://dx.doi.org/10.5331/seppyo.65.33>.
- [82] W. Zhao, J. Fang, T. Yang, X. Lian, A. J. Winkler, F. Sun, and P. Gentile. Observation-constrained physical snow water equivalent simulations using a

- physics-guided machine learning approach. *Water Resources Research*, 62(3), Mar. 2026. ISSN 1944-7973. doi: 10.1029/2025wr041406. URL <http://dx.doi.org/10.1029/2025WR041406>.
- [83] W. L. Zhao, P. Gentine, M. Reichstein, Y. Zhang, S. Zhou, Y. Wen, C. Lin, X. Li, and G. Y. Qiu. Physics-constrained machine learning of evapotranspiration. *Geophysical Research Letters*, 46(24):14496–14507, Dec. 2019. ISSN 1944-8007. doi: 10.1029/2019gl085291. URL <http://dx.doi.org/10.1029/2019GL085291>.
- [84] Z. Zhao, Y. Ding, M. Wang, and Y. Qu. 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, 2025. doi: 10.1109/tgrs.2025.3551094. URL <http://dx.doi.org/10.1109/TGRS.2025.3551094>.
- [85] J. K. F. Zöllner. *Photometrische Untersuchungen: mit besonderer Rücksicht auf die physische Beschaffenheit der Himmelskörper*, page 273. Engelmann, Leipzig, 1865. Archival scan available at <https://www.digitale-sammlungen.de/en/view/bsb10060733>.
- [86] M. Ćurić and V. Spiridonov. *Beginnings of Meteorological Measurements and Observations*, page 129–146. Springer Nature Switzerland, 2023. ISBN 9783031450327. doi: 10.1007/978-3-031-45032-7_5. URL http://dx.doi.org/10.1007/978-3-031-45032-7_5.

A PHYSICS-CONSTRAINED NEURAL DIFFERENTIAL EQUATION FRAMEWORK FOR DATA-DRIVEN SNOWPACK SIMULATION

A. Charbonneau, K. Deck, and T. Schneider. A physics-constrained neural differential equation framework for data-driven snowpack simulation. *Artificial Intelligence for the Earth Systems*, 4(3):240040, 2025. doi: 10.1175/AIES-D-24-0040.1. URL <https://journals.ametsoc.org/view/journals/aies/4/3/AIES-D-24-0040.1.xml>. A.C. was responsible for project conception, formulating the model framework, as well as gathering data and analyzing outputs, as the lead writer of the manuscript.

This work has been accepted to *Artificial Intelligence for the Earth Systems* under the American Meteorological Society (AMS). The AMS does not guarantee that the copy provided here is an accurate copy of the Version of Record (VoR). The VoR can be found at <https://doi.org/10.1175/AIES-D-24-0040.1>.

ABSTRACT

This chapter presents a physics-constrained neural differential equation framework for parameterization, and employs it to model the time evolution of seasonal snow depth given hydrometeorological forcings. When trained on data from multiple SNOTEL sites, the parameterization predicts daily snow depth with under 9% median error and Nash Sutcliffe Efficiencies over 0.94 across a wide variety of snow climates. The parameterization also generalizes to new sites not seen during training, which is not often true for calibrated snow models. Requiring the parameterization to predict snow water equivalent in addition to snow depth only increases error to $\sim 12\%$. The structure of the approach guarantees the satisfaction of physical constraints, enables these constraints during model training, and allows modeling at different temporal resolutions without additional retraining of the parameterization. These benefits hold potential in climate modeling, and could extend to other dynamical systems with physical constraints.

2.1 Introduction

Seasonal snowpacks help regulate the Earth’s energy balance, provide freshwater storage, and are crucial for understanding Earth’s climate. They hold economic and ecological significance, supplying a majority of the western United States’ (and a sixth of the world’s) water supply and influencing agriculture, flood, drought, and avalanche hazards [16, 25]. Their seasonal importance and susceptibility to climate change emphasize the need for ongoing modeling and monitoring on both seasonal and multi-decadal timescales.

Modeling the evolution of seasonal snow for regional or global climate applications offers a challenging problem of scales; it is the bulk properties of the snow (albedo, snow cover fraction, snow temperature, and snow water content) that are critical, yet microphysical and location-specific processes control these properties and must be taken into account. The most detailed models represent vertically-resolved snowpacks, including liquid percolation, phase changes, metamorphism effects, and other types of compaction; they are often calibrated and used on the site-level (e.g., [17]). Models used in global climate simulations range in complexity from single-layer/bulk models to multi-layer models with parameterizations for bulk properties that are calibrated with observational data [48]; the horizontal resolution of these models is typically $\sim 10\text{--}100$ km where microscopic processes cannot

be tractably resolved. While the laws of physics ultimately govern the evolution of snowpacks, uncertainty in how to relate essential but unresolved small-scale processes to snowpack bulk properties on the spatial scales of global models makes developing snow models a challenging task [5, 34]. This challenge is exacerbated by data availability [36, 48].

Among bulk variables in global snow models, the snow water equivalent, SWE (units of m), represents the total water storage in snow. It relates to the snow depth z (units of m) through the bulk snow density ρ_{snow} (units of kg m^{-3}) and the density of liquid water ρ_{water} (1000 kg m^{-3}) as

$$\text{SWE} = (\rho_{\text{snow}} / \rho_{\text{water}}) z; \quad (2.1)$$

it is typically used as a prognostic variable in bulk models for determining snowpack mass balance. Density and depth are critical in climate models for determining the snowpack energy balance, as they influence thermal, mechanical, and optical properties, impacting mass/energy fluxes, water retention, and spring thaw [9, 36]. However, for a given SWE, the snow depth and bulk density can vary considerably over time at a single location, or between locations under similar forcings, due to compaction, melt/refreeze cycles, and changes in the density of falling snow. These variations give rise to ongoing challenges in snow modeling.

Many prevalent snow models rely on the seminal parameterizations of Kojima [35] and Anderson [1] to derive snow density and depth from modeled SWE. These works modeled compaction and microstructure metamorphism by assuming a linear relationship between the strain rate of compaction and the weight of the overlying snow and suggested empirical parameterizations that were calibrated from a select number of observational sites and laboratory experiments and analytically extended¹. Snow evolution in models such as Snow17, iSnobal, Noah/NoahMP, CLM5, HTESSEL, and others [2, 21, 38, 46, 48, 53] employ this simplified formulation based on small-scale representations; thus, it is integral to most current US and European global snow predictions.

However, these parameterizations were primarily developed for the hydrological community instead of earth system models, and their calibration to prioritize accurate SWE and subsequent local runoff induces a trade-off in density/depth errors. Such errors impact the snowpack energy balance, which is as crucial as mass balance

¹*Post-Publishing Note:* more explicitly, they assumed $d\rho/dt = \rho * C * \text{SWE}$ for some constant C which was fit with minimal data, with later addition of temperature-dependent effects.

in global climate simulations [18, 78], and snow depth errors remain problematic in climate predictions [48]. The inadequacy for energy tracking in these formulations has been recognized for decades and spurred development of more detailed and realistic alternatives such as CROCUS or SNOWPACK, which evolve the metamorphism of snow microstructure [11, 40, 71]. While these complex models address the limitations of Anderson’s and Kojima’s original formulations, they require site-specific calibration and struggle with computational scalability for global applications. This necessitates more efficient representations for global models that can accurately predict snow depth/density. With the push toward even finer localized (1–10 km) land modeling, simple yet accurate models are increasingly essential for computational efficiency [6, 14, 59].

Advances in computing and sensing technology have led to initiatives in data assimilation and machine learning (ML) aimed at disrupting longstanding parameterizations. Today, extensive snow depth observations surpass the resolution and precision of SWE and density data/estimates [10, 24]. Improving snow models with these resources has become a dominant avenue of hydrology research, with a bias towards SWE modeling, like incorporating remotely sensed depth into iSnobal to infer SWE [27]. Several ML models (e.g., [5, 20, 47, 65]) predict SWE or depth from meteorological and topographical inputs in specific regions. Such models show satisfactory snowpack estimation but frequently yield errors over 15% when tested, especially beyond their training or calibration locations [22, 47, 70]. The ability of these empirical models to generalize to new locations or future climates and act as a universal model is limited, and their statistical or black-box nature does not inherently respect physical constraints, impeding their integration into land models [16, 25]. Combining depth observations and ML techniques to improve depth/density parameterizations has the potential to lead to improved simulation of key variables, benefiting global climate modeling. Capitalizing on this opportunity demands a representation that can generalize to many snowpacks and integrate with existing large-scale models.

This work presents a novel hybrid approach to parameterization, combining physical principles with empirical modeling that structurally guarantees compliance with prescribed bounds (for example, physical consistency or conservation conditions). We showcase its utility in designing an alternative parameterization for snow depth, with ramifications for global climate and seasonal simulations. We also create a quality-controlled dataset for snow modeling and make it publicly available.

Learning physically-informed representations from observational data across many locations enables robust performance that can generalize to new locations without recalibration. The customizability of the approach permits straightforward adaptation to different operational requirements and constraints, demonstrating additional capabilities with minimal adjustment. This offers a flexible, efficient, and scalable framework that is adaptable as the field evolves. The proposed approach exhibits a versatile means for enforcing (or learning) any function-based threshold on an optimizable model without modification of the training metrics, which can contribute to contemporary global snow modeling as well as other physics/ML hybrid models.

2.2 Methodology

Overview

Our model choices leverage ML for seasonal snow simulation in climate models, prioritizing generalizability and computational efficiency. Contemporary paradigms in hydrology research focus on parameterizing SWE from z and other data to constrain its value over global grids. However, within global climate models, SWE evolution is already well-constrained by explicitly implemented physical laws enforcing mass conservation and relatively well-understood fluxes such as sublimation, precipitation, and melt. The subsequent conversion of simulated SWE to variables like z or ρ_{snow} is typically left to longstanding parameterizations. These parameterizations can be pre-calibrated offline prior to use in a snow model, or are often further tuned “online” within a snow model. The data used for calibration can be indirect (non-SWE data), such as energy and water fluxes or land surface albedo, or be direct measurements of snow depth or density, or even gridded SWE estimates. The resulting global simulations are sometimes employed in refining gridded SWE for calibrating other models, creating circular estimation and biases.

Given the relative abundance of observations of snow depth z alongside additional snow variables, this instead justifies parameterizing z (from physics-constrained SWE) as an alternative to established formulations to address these limitations, so that tighter relationships can be determined and evaluated on the basis of direct, high-quality data. By using primary observational data instead of assimilated/reanalysis data, this approach can avoid biases and inaccuracies, ensuring more faithful representations of the underlying processes.

We model the rate of change in snowpack height (units of m s^{-1}) by an ordinary

differential equation (ODE) represented by an artificial neural network M :

$$\frac{dz}{dt} = M(z, \text{SWE}, \varphi, R, v, T_{\text{air}}, P_{\text{snow}}), \quad (2.2)$$

where SWE is the snow-water equivalent (m), φ is the relative humidity (between 0 and 1, the used data are measured with respect to liquid water), R is the broadband solar radiative energy flux (W m^{-2}), v is the wind speed (m s^{-1}), T_{air} is the air temperature ($^{\circ}\text{C}$), and P_{snow} is the liquid water-equivalent rate of snowfall (m s^{-1}). The 1D column approach permits application over any spatial grid. The chosen input variables only indirectly encode location and time dependencies through the environmental input variables, allowing the model to function in areas where topographical or temporal information is unavailable. This choice aims to enable learning about universal linear and nonlinear physical processes that apply independent of time, season, and location. Using a feed-forward neural network dependent only on the current system state aligns with land-surface models, as it matches the differential equation format used for other variables. This model is also adaptable for different applications or when SWE is unavailable (see sections 2.2 “Model Structure” and 2.2 “Training” below).

Model Structure

The model M consists of two components. The first is a “predictive” network with trainable weights to generate a dz/dt prediction (Fig. 2.1). For computational simplicity, only two hidden layers were used, which can also be interpreted as a regression on once-transformed features, with the transformational layer width set by the hyperparameter n scaled by the number of input features k (see Fig. 2.1). Inputs are easily exchangeable for alternative use cases or target predictions.

The second component consists of fixed-weight dense layers with Rectified Linear Unit (ReLU) activation, designed to impose explicit (“hard”) constraints on the predictive model. This approach allows for enforcing min/max thresholds on any predictive model, without introducing penalties into the calibration metrics. Although more advanced methods exist in literature or modern coding packages [7, 19, 33], our simplistic approach offers multiple advantages. Primarily, constraints are applied throughout training, leading to better gradient updates within prescribed limits (see Table 2.4 for comparison). This flexible framework supports most functions or specialized constructions, including “learned” constraints, and can be scalably implemented in environments with minimal to no ML support, offering potential for many fields (for more on this process, see Appendix A). By

guaranteeing physical constraints, it ensures stability during time-stepping and is conducive for integration into larger models without violating prescribed bounds such as conservation or consistency equations.

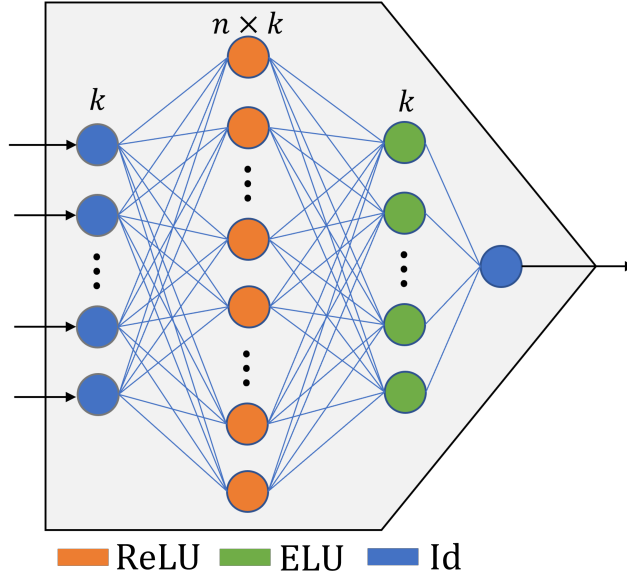


Figure 2.1: **Structure of the model’s predictive component.** Blue lines indicate a trainable linear transformation of the input (of k scalar variables), including a bias. Colors indicate the activation function used upon collection at the node (ReLU = Rectified Linear Unit, ELU = Exponential Linear Unit, Id = Identity), as noted in the legend. The hyperparameter n sets the width of the internal mixing layer.

Threshold Constraints for Snowpack Prediction

Constraints for dz/dt should keep the depth tendency within physical limits to enable generalizability and stability when M is integrated over time. This initial study selected the following basic constraints:

- Enforce depth non-negativity within a time step Δt , i.e., $M \geq -z/\Delta t$;
- Enforce depth inability to increase without snowfall, i.e., $P_{\text{snow}} = 0 \implies M \leq 0$. Processes like wind drift violate this constraint, but such effects are small in our data (see Appendix B).

These constraints can be expressed as threshold functions, the lower as $f_- = -z/\Delta t$ and the upper as $f_+ = \text{ReLU}(p) \times \mathbf{1}_{\{P_{\text{snow}} > 0\}}$, where p is the output of the predictive component and $\mathbf{1}$ is the indicator function. For these choices, f_- is nonpositive and f_+ is nonnegative, which simplifies the constraint layer structure (see Appendix A), resulting in a final structure for M as depicted in Fig. 2.2.

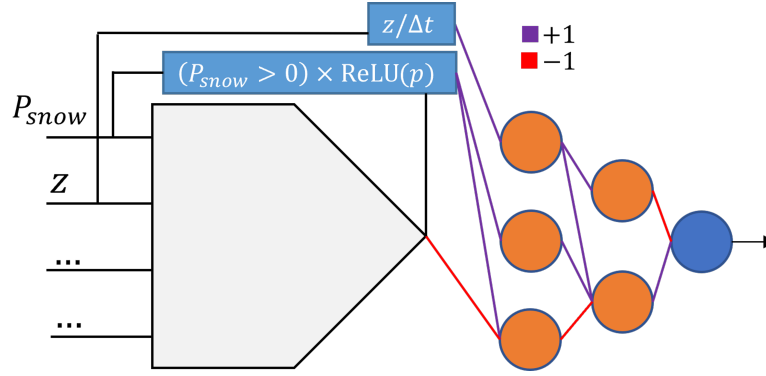


Figure 2.2: Architecture of M , highlighting the constraint component attached to the predictive component (the grey pentagon) described in Fig. 2.1. The chosen structure enforces increasing snow depth only under precipitation and non-negativity of snowpack height, and is equivalent to a max/min block on the output. Weight colors indicate the constant’s sign and activation functions follow the color scheme given in Fig. 2.1.

The first constraint includes the time step Δt , but this does not explicitly affect the time dependency or resolution of the parameterization. The predictive component contains no time nor time-step dependence. Adjusting Δt scales the constraint appropriately without altering the predictive component’s output, enabling the model’s use in adaptive time-step schemes (re-scaling the constraint per time step). This means the model in principle only requires training with data at one temporal resolution, though we would anticipate improved time-step independence if training data incorporated varied time intervals, or dz/dt values in the minimum range anticipated during usage.

Data

We required training data with simultaneous snow and meteorological measurements, preferring collocated primary ground observations over reanalysis data due to known discrepancies [49]. One primary source is the US Snow Telemetry Network (SNOTEL) by the Natural Resources Conservation Service (NRCS). Data from 44 total SNOTEL sites in the contiguous United States (CONUS) have simultaneous availability of the necessary inputs, of which the entire reporting history until February 1st, 2024 was collected. These sites span diverse climates (see Fig. 2.3), enhancing generalizability. For testing, 7 total Alaskan SNOTEL sites had the necessary inputs and were similarly collected, plus standard evaluation data from Kühtai, Austria [37], Col de Porte, France [41], and the Reynolds Mountain East catchment [58] to assess the M ’s ability to generalize to out-of-sample data. We additionally used data from Sodankylä, Finland [23], the upper Rofental [74], and

Yala Basecamp [61, 66] in the Himalayas to test performance across uncalibrated elevations and climate types.

SNOTEL data have known quality issues, such as underrepresenting complex mountainous terrain and underreporting precipitation [50, 60], along with periods of biased or unphysical values [28]. To ensure data suitability for model training, we applied established cleaning defined by Serreze et al. [60] and refined by Yan et al. [79] to SNOTEL daily snowpack data. Gauge undercatch was corrected as in Livneh et al. [43] and temperature biases addressed following the SNOTEL correction release [4]. Since no consistent quality control exists for SNOTEL meteorological or snow depth data beyond outlier tests by Hill et al. [28], we developed a custom procedure (see Appendix B). Snow fraction (to obtain snowfall from total precipitation) was estimated using the $(T_{\text{air}}, \varphi)$ bivariate logistic model from Jennings et al. [32], shown to have over 88% accuracy. From the cleaned data, we derived dz/dt , $d\text{SWE}/dt$, and P_{snow} for days with complete data, excluding all data with $\Delta t > 1$ day.

The training data were averaged (preserving start-of-window z and SWE) over a consecutive N -day moving window, with N as a hyperparameter. This enabled exploration of the tradeoff of spreading out discretized SNOTEL data (z to the inch, SWE to 0.1 inch) for smoother regression learning against preserving extreme values critical for predictions. Days with unphysical values (zero SWE and nonzero z) or no snowpack were removed to eliminate uninformative zeros in the target space. Features were scaled by their standard deviations, and the target by its absolute maximum, with scaling constants fixed in M to spare user preprocessing. This resulted in 58,484 usable sensor-days out of 105,636 for training, and 35,618 sensor-days for testing. The complete dataset and generating code are publicly available (see Data Availability).

Training

Accurate prediction of extreme values is vital in snowpack modeling. Underpredicting extreme dz/dt can prevent rapid snowpack growth or depletion, lagging snow presence early in the season or maintaining snow into the summer, which skews albedo, runoff, and energy calculations. Models like Noah, CROCUS, and SNOWPACK have struggled with this challenge [25, 44, 45, 72, 76]. Standard regression tends to under-predict extremes, so we used a custom loss function that

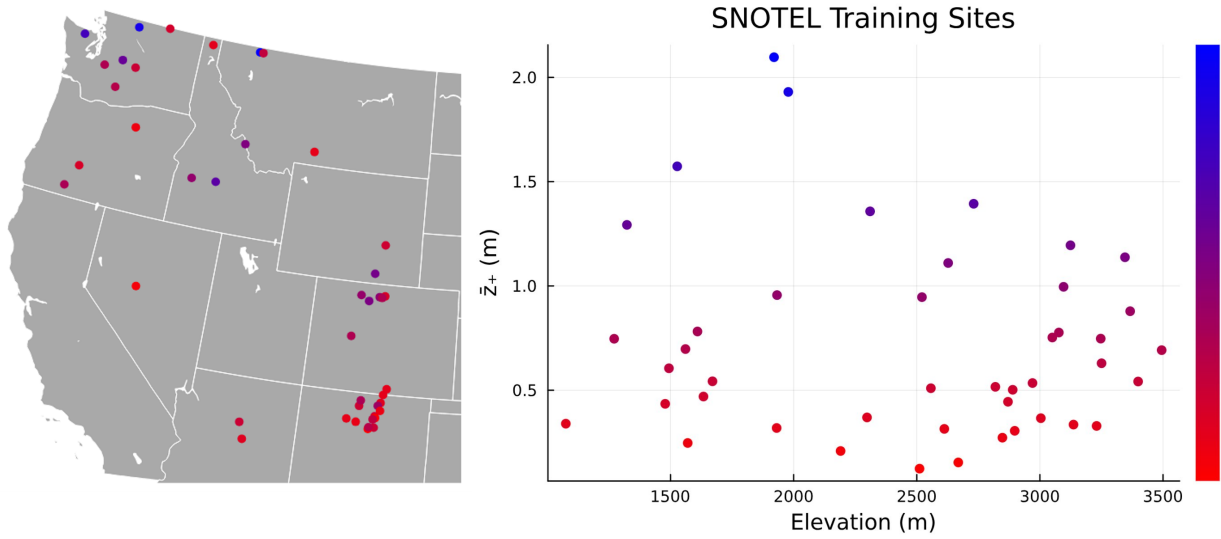


Figure 2.3: Distribution of SNOTEL sites used for training the network. (a) Training sites as visualized over the United States. (b) Training sites visualized with elevation vs their average nonzero snowpack height, $\bar{z}_+$. The color bar is a visual indicator of $\bar{z}_+$ for visualization on the spatial map.

can emphasize extreme values:

$$L = \frac{1}{N_d} \sum_{i=1}^{N_d} w_i |y_i - \hat{y}_i|^{n_1}, \quad w_i = 1 + |y_i|^{n_2}. \quad (2.3)$$

Here, N_d is the number of batched training data, $\hat{y}_i$ and y_i are the 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. Hyperparameter tuning followed a leave-one-out approach, using averaged and filtered data from 43 of 44 sites. Validation scores were generated over the remaining sites using unaveraged, unfiltered data, and averaged to guide hyperparameter selection. For timestepping (see section 2.2 “Testing”), the optimal n_1 was found with L_2 training and $n_2 > 0$, highlighting the importance of extreme points. We note that optimal hyperparameters varied between regression and timeseries tasks; see Appendix C for more details.

The model was implemented in Julia and the Flux framework [30, 31], with the RMSProp optimizer [29]. Training for 100 epochs (100 passes over all data) takes under 30 seconds on one Intel i9 CPU without GPU usage, with model storage under 3 kilobytes. Time and memory benchmarking of the network are listed in Table 2.6 in Appendix C.

Testing

Model performance was tested by timestepping the dz/dt equation with an explicit Euler method:

$$\hat{z}_{i+K} = \hat{z}_i + (K\Delta t)M(\hat{z}_i, \text{SWE}_i, \varphi_i, R_i, v_i, T_{\text{air},i}, P_{\text{snow},i}). \quad (2.4)$$

The integer K specifies sequential data transitions ($K = 1$) or “gaps” to traverse in timeseries data ($K > 1$) due to missing or cleaned data. M ’s built-in constraints ensure non-negative z values when the step size is the designated Δt (i.e., when $K = 1$; for providing other Δt see Section 2.3 “Finer Resolution Predictions”), though this choice of time-stepping procedure can create negative z when $K > 1$ (rates constrained for step-size Δt are applied over $K\Delta t > \Delta t$). In such cases, negative z values were set to zero, and similarly, timeseries were “reset” to observed values $\hat{z}_{i+K} = z_{i+K}$ whenever $K > K_{\text{max}} = 5$ days, to avoid attributing error from the method choice to M for fair evaluation. When M is used within a global model simulation, such gaps would not occur since the inputs would be available at every step, allowing all z_i to obey prescribed bounds. Selecting $K_{\text{max}} > 1$ day additionally reduced the number of resets that would otherwise beneficially skew performance metrics, with $K_{\text{max}} = 5$ ensuring resets in less than 2.3% of cases (median frequency 0.19%), with many over whole years or no-snow periods (not impacting snow simulation). Evaluation metrics included root mean square error (RMSE), mean absolute error (MAE), bias (B), and median percent error (MPE, on dz/dt for regression and z for timeseries). For timeseries, we included the Nash-Sutcliffe efficiency (NSE; [52]) and snowpack percent error (SPE), $\text{MAE}/\bar{z}_+$, where $\bar{z}_+$ is the mean nonzero height.

By swapping SWE and z features, the network can also be trained to predict $d\text{SWE}/dt$ from z , allowing SWE to be simulated from depth data. This enables standalone modeling using only weather inputs, with two networks $\tilde{M}_z$ and $\tilde{M}_{\text{SWE}}$, separately trained to run in a coupled fashion:

$$\widehat{\text{SWE}}_{i+K} = \widehat{\text{SWE}}_i + (K\Delta t)\tilde{M}_{\text{SWE}}(\hat{z}_i, \widehat{\text{SWE}}_i, \varphi_i, R_i, v_i, T_{\text{air},i}, P_{\text{snow},i}), \quad (2.5)$$

and

$$\hat{z}_{i+K} = \hat{z}_i + (K\Delta t)\tilde{M}_z(\hat{z}_i, \widehat{\text{SWE}}_i, \varphi_i, R_i, v_i, T_{\text{air},i}, P_{\text{snow},i}). \quad (2.6)$$

The only change to ensure physical consistency is to alter the lower bound of $\tilde{M}_z$ ($\tilde{M}_z = M$ otherwise) such that the z update obeys $z_{i+K} \geq \text{SWE}_{i+K}$ to enforce $z \geq \text{SWE}$, so SWE_{i+K} was calculated first before dz/dt . This permits comparison with other models without inputting observational SWE.

To compare the neural model to established parameterizations, the Snow17 temperature-index model [2] was implemented and evaluated on the same data. Snow17 (designed for modeling runoff) models SWE and infers depth through Anderson’s density parameterization, and it can also assimilate observed SWE, allowing comprehensive comparisons. We performed two comparisons: (1) M (the one-network depth parameterization) against Snow17 with both assimilating observational SWE, and (2) the two-network standalone model $\tilde{M}$ (subcomponents $\tilde{M}_z$ and $\tilde{M}_{\text{SWE}}$) against Snow17, both using only meteorological inputs. To avoid confusion, Snow17 predicting versus assimilating SWE are labeled as “SN17” and “SN17O,” respectively. We used the Snow17 parameters from the Wang et al. [73] *SN17-B-CONUS* model, calibrated over CONUS to account for regional climates. Both models were restarted with accurate depth when resets occurred, and computational benchmarking is compared in Table 2.6. Significance in differences between RMSE metrics were evaluated via a Wilcoxon signed rank test [77] with $p = 0.05$, which does not assume normal distributions or variance homogeneity.

In addition to depth, bulk density timeseries can be generated from z and SWE data and model outputs according to Eq. 2.1 and compared. The data is discrete while the model outputs are continuous, so densities were compared at sites with collocated SWE and z sensors (see Appendix B) only when both models and data yielded physical values ($0 < \rho_{\text{snow}}/\rho_{\text{water}} < 1$). Counts of “false non-snowpacks” (models show $z = 0$ while $z > 0$ in the data) and “false snowpacks” (models show $z > 0$ while $z = 0$ in the data) were recorded, along with instances of unphysical densities.

To further assess physical consistency, the estimated rates $d\text{SWE}/dt$ by $\tilde{M}_{\text{SWE}}$ were compared to the snowfall rate data P_{snow} . When the air temperature is below freezing ($T_{\text{air}} < 0$), the conservation of mass implies these rates should be roughly equivalent, limited by factors such as runoff, sublimation, snow transport, and data precision. While it is possible to directly impose conservation of mass as a threshold in this framework, forgoing this specific bound allows for an investigation of the model’s ability to represent physical constraints beyond those explicitly encoded.

Assessing physical consistency in the model directly per feature is nuanced due to strong correlations among the inputs, which confounds most interpretation methods like partial dependence, permutation, LIME, and tractable SHAP values [51]. To isolate feature effects on model output, we calculated first-order Accumulated Local Effects (ALE) plots [3]. This method shows the average change in model output

accumulated over sequential bins of feature values. For feature X at grid point x_i , data in the window $[x_i - \Delta, x_i + \Delta]$ are evaluated setting $X = x_i - \Delta$ and $X = x_i + \Delta$, storing the average of the differences $\Delta M_{x_i} = \overline{M(x_i + \Delta) - M(x_i - \Delta)}$. The final (centered) ALE value $\bar{\Delta M}$ at x_i is the sum over all ΔM_{x_k} with $x_k \leq x_i$, minus the average of all uncentered ALE values. This isolates changes solely from feature variations, and bins are defined by quantiles to ensure equal data instances in each window. The shape and slope of the ALE curve are more pertinent for interpreting physicality than the offset, especially when the feature distribution is skewed. The range of $\bar{\Delta M}$ over the feature indirectly measures feature importance in influencing model output, as the ALE value can be interpreted as a departure from the average model prediction. Further information for interpreting ALE plots can be found in Molnar [51].

2.3 Results

Depth Timeseries

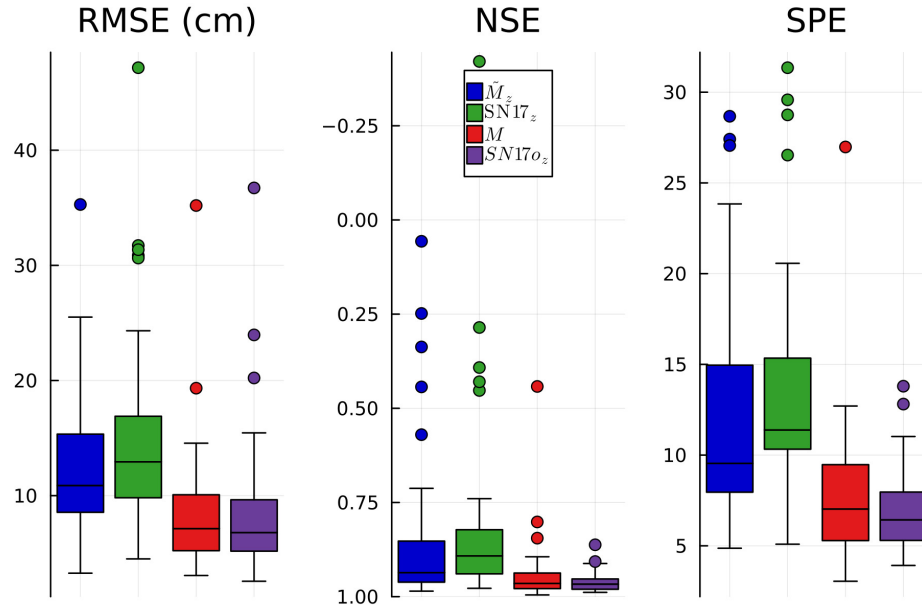


Figure 2.4: **Performance of M and $\tilde{M}$ against Snow17 with (SN17O) and without (SN17) observational SWE data for generating z timeseries over the 44 validation sites.** RMSE indicates root mean square error, NSE the Nash-Sutcliffe Efficiency, and SPE the average L_1 error normalized to the average nonzero depth, to measure percent error. Boxes outline the 25% to 75% quantiles, with the bar at the median, while whiskers mark the extremes and dots indicate outliers, which lie beyond 1.5 times the interquartile range (box width) from the box. Vertical axes limits are chosen to show all data points.

Over validation sites, neural configurations performed similarly to Snow17 inde-

pendent of including observed SWE (Fig. 2.4). Both models exhibit similarly tight spreads, though M exhibited larger spreads than SN17O when utilizing SWE data. Conversely, when simulating SWE, the spread was larger for SN17 compared to the coupled neural model $\tilde{M}$, which is more indicative of usage under SWE uncertainty or provision of SWE within a separate model. Statistically significant improvements in RMSE were found for both SWE and z modeling by $\tilde{M}$ compared to SN17 ($p = 0.029$ and $p = 0.0006$, respectively). No significant difference was observed for M against SN17O using SWE data for depth parameterization alone ($p = 0.673$), which is unsurprising as both models were calibrated for performance over the locations represented in this data.

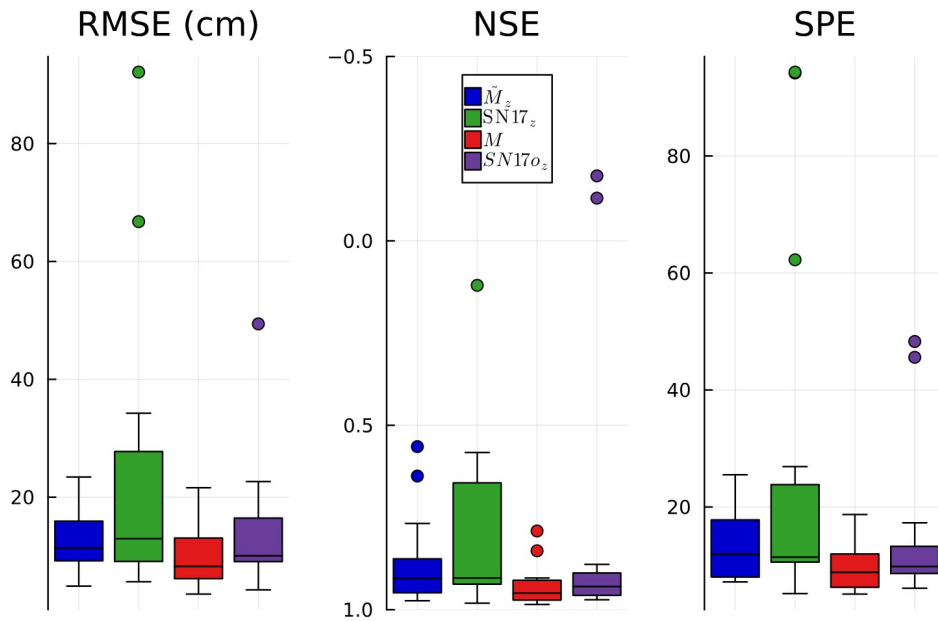


Figure 2.5: Performance of the neural parameterization against Snow17 over the 14 testing sites in depth timeseries generation. Labeling convention follows that in Fig. 2.4. One outlier for SN17 with an NSE of -2.8 is not shown on the plot to aid in the scaling of the other values.

The model performance over all testing sites is summarized in Fig. 2.5, with example timeseries in Fig. 2.6, and medians and averages reported in Appendix C (Table 2.4). Although the performance distributions are non-symmetric, both mean and median metrics are important for gauging potential (median) and consistency (mean) for generalizing to out-of-sample regimes. The neural configurations showed tighter performance spreads across testing sites and climates not included in the training data, despite no recalibration. Average neural NSE was 0.87 and 0.94 (without and with SWE data, respectively), compared to 0.35 and 0.78 for Snow17 — median differences were smaller, but still favored neural models. Snow17’s mean SPE

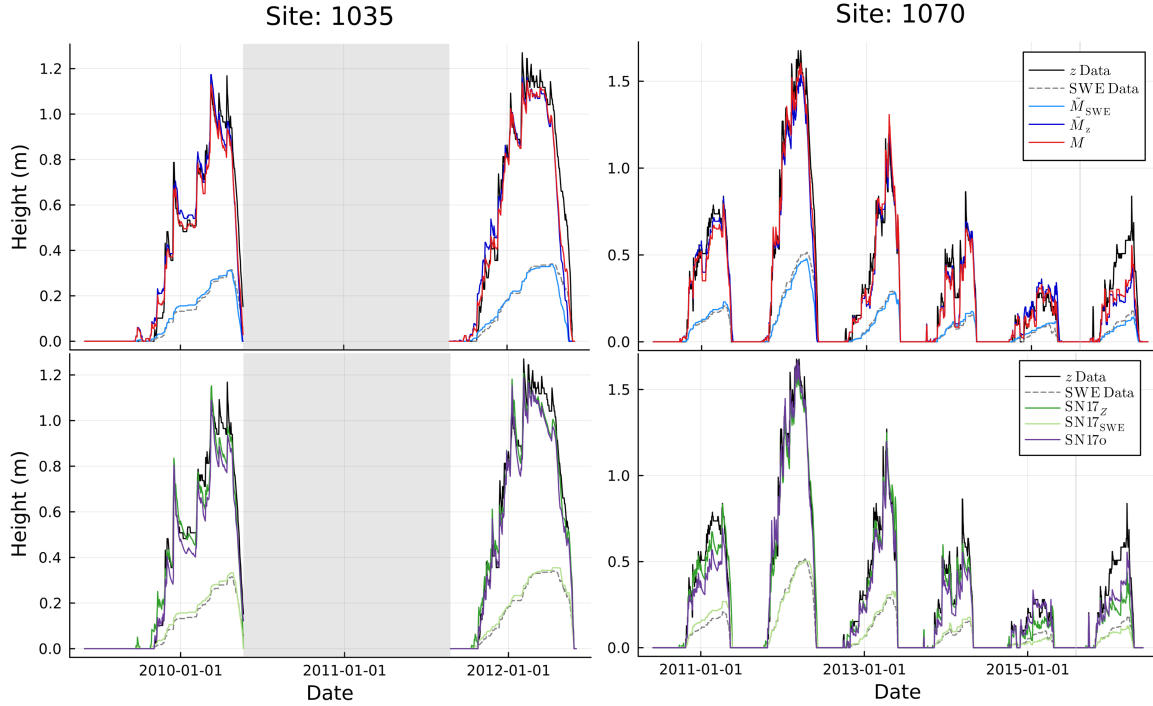


Figure 2.6: z and SWE timeseries over two subsets of the Alaskan testing sites. Data gaps (missing or cleaned data) are shaded in grey. All models perform similarly, though the neural models seem to perform poorly whenever Snow17 also performs poorly, like the winter of 2016 at Site 1070. This suggests the neural parameterization is at least an efficient surrogate for Snow17, or could insinuate a data inconsistency if both models fail similarly.

was 15% even with observational SWE, comparable to other established models without site-specific calibration [12, 16, 22, 25, 44, 47, 70, 71], and nearly doubled to 28% without SWE data. In contrast, the neural parameterization maintained mean SPE under 15%, increasing modestly from 9.5% to 13.4% without SWE data, demonstrating greater robustness and consistency out-of-sample despite the simplicity of its predictive component and lack of past snow depth state storage.

Model M demonstrates strong generalizability, with similar performance and tighter spreads across test and validation sites, without retuning. Although the neural model did not exhibit significant RMSE improvements over Snow17 on out-of-sample data at the $p = 0.05$ level ($p = 0.14$ for z , $p = 0.27$ for SWE), it did outperform Snow17 as a parameterization ($p = 0.01$). This suggests that when paired with a reliable SWE predictor within a larger model, the neural approach offers advantages over prevailing alternatives in extrapolating to uncalibrated locations.

Table 2.1: Results of bulk density timeseries generation for each model. The median score over all validation and viable testing sites is presented, and scores are derived from all predictions of physical densities during observed snowpacks, or otherwise tallied in the presented counts. The integer count gives the median number of occurrences across sites, while the percentage normalizes each count against the length of the timeseries.

Parameter	$\tilde{M}$	M	SN17	SN17O
RMSE (%)	9.1	9.5	7.4	7.6
Bias (%)	-1.9	-1.8	0.6	0.9
MPE (%)	11.3	12.9	14.1	14.9
False Non-Snowpacks	4 (0.13%)	2 (0.08%)	14 (0.68%)	0 (0%)
False Snowpacks	183 (8.6%)	136 (6.1%)	45 (3.2%)	2 (0.10%)
Unphysical Points	3 (0.17%)	3 (0.12%)	0 (0%)	0 (0%)

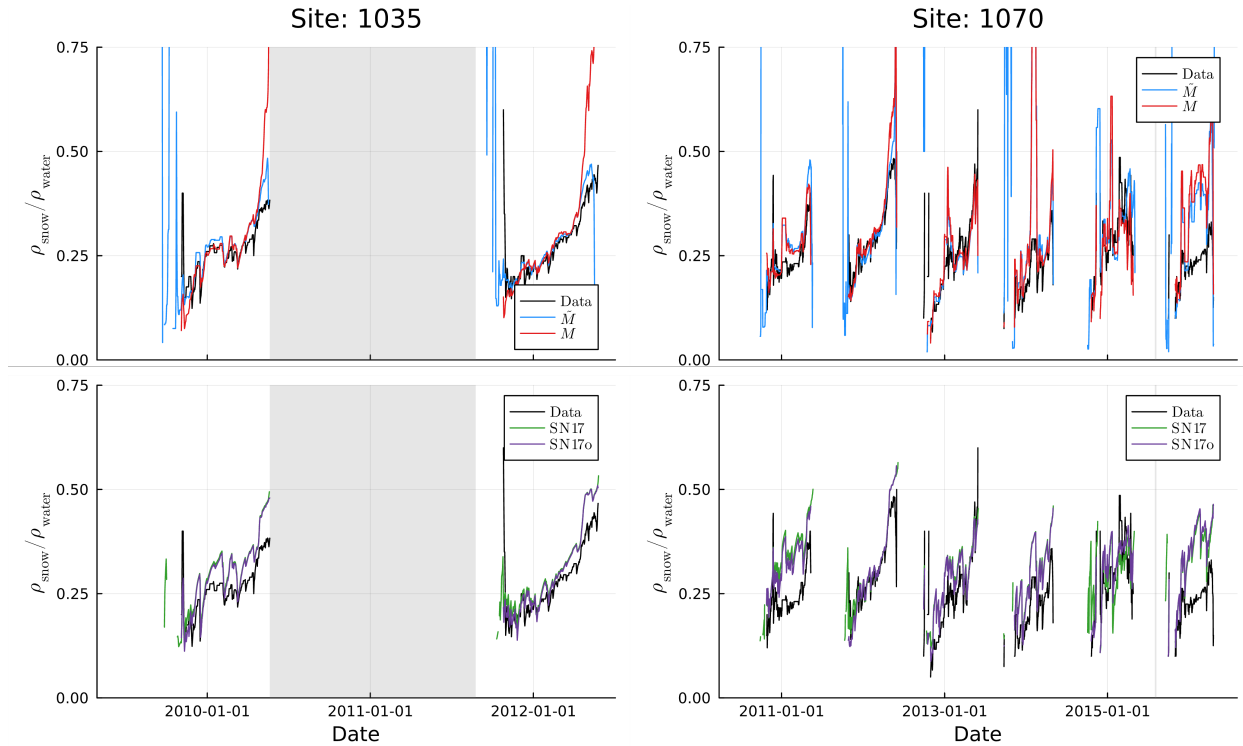


Figure 2.7: Bulk density timeseries (normalized against ρ_{water}) over the same series as Fig. 2.6. Poor representation at the start/end of the season for small snowpacks are exacerbated by data discretization and lack thereof in the models, particularly in $\tilde{M}$ and SN17, or from simpler constraints in M and $\tilde{M}$. Such fluctuations severely skew the NSE and SPE metrics and confound their measure of model performance. Timeseries are only shown where models and data show a relative density between 0 and 1.

Density Timeseries

For validation and testing sites with collocated z and SWE measurements (see Appendix B), Table 2.1 compares timeseries statistics for derived bulk density, and

Fig. 2.7 displays estimates from the same data as Fig. 2.6. RMSE values reflect relative error since density is normalized by ρ_{water} . Both SN17 and $\tilde{M}$ lack SWE discretization, inflating errors for small snowpacks compared to their assimilated counterparts. Neural unphysical densities source from a lack of constraint enforcing $\text{SWE} = 0 \implies z = 0$ (in both $\tilde{M}, M$) or $z > \text{SWE}$ (in M), though different constraint choices could eliminate this². Snow17 is roughly $3\times$ better at mitigating false snowpacks with lower RMSE (less extreme errors), but our models are roughly $3\times$ better at reducing false snow-absence and exhibit improved individual density predictions, crucial for estimating other snowpack properties. Thus, the advantage between models depends on which features are prioritized. Otherwise, the two perform similarly, despite M being designed for snow depth while the comparable parameterization within Snow17 is explicitly formulated for density. Among evaluated sites, the best and worst performances by M were better than those of SN17O.

Predicted dz/dt

Fig. 2.8 shows a histogram comparing predicted versus true dz/dt values from M (Pearson correlation $r = 0.78$) and $\tilde{M}$ ($r = 0.77$) during timeseries generation. Both still display a tendency to under-predict extreme values. This is likely from exposure to much more training data with small dz/dt , which could result in better predictions of small values at the expense of extremes without applying other methods like class balancing.

Physical Behavior of Model

For SNOTEL instances with $T_{\text{air},i} < 0$, $\tilde{M}_{\text{SWE}}$ violated the $d\text{SWE}/dt \leq P_{\text{snow}}$ condition only 0.6% of the time, with all violations under 1 mm/day and 95% under 0.75 mm/day. About half the instances showed $d\text{SWE}/dt < P_{\text{snow}}$ (prescribed bounds yielded equivalence in the rest), but the mean residual of these was -2 mm/day, with 95% above -5 mm/day, aligning with typical sublimation rates [42, 64] and effects not explicitly modeled. Given the SNOTEL data precision (2.54 mm) and negative average temperatures obscuring instantaneous positive temperatures, these results suggest strong adherence to mass conservation, even without explicitly enforcing it (as simple as choosing $f_+ = P_{\text{snow}}$ for the upper bound of $\tilde{M}_{\text{SWE}}$), highlighting the model’s ability to learn physical representations from data beyond

²Post-Publishing Note: For example, let the lower bound of the depth network be $\frac{\text{SWE}-z}{\Delta t} + M_{\text{SWE}}$, where M_{SWE} is the output of the (evaluated first) SWE network.

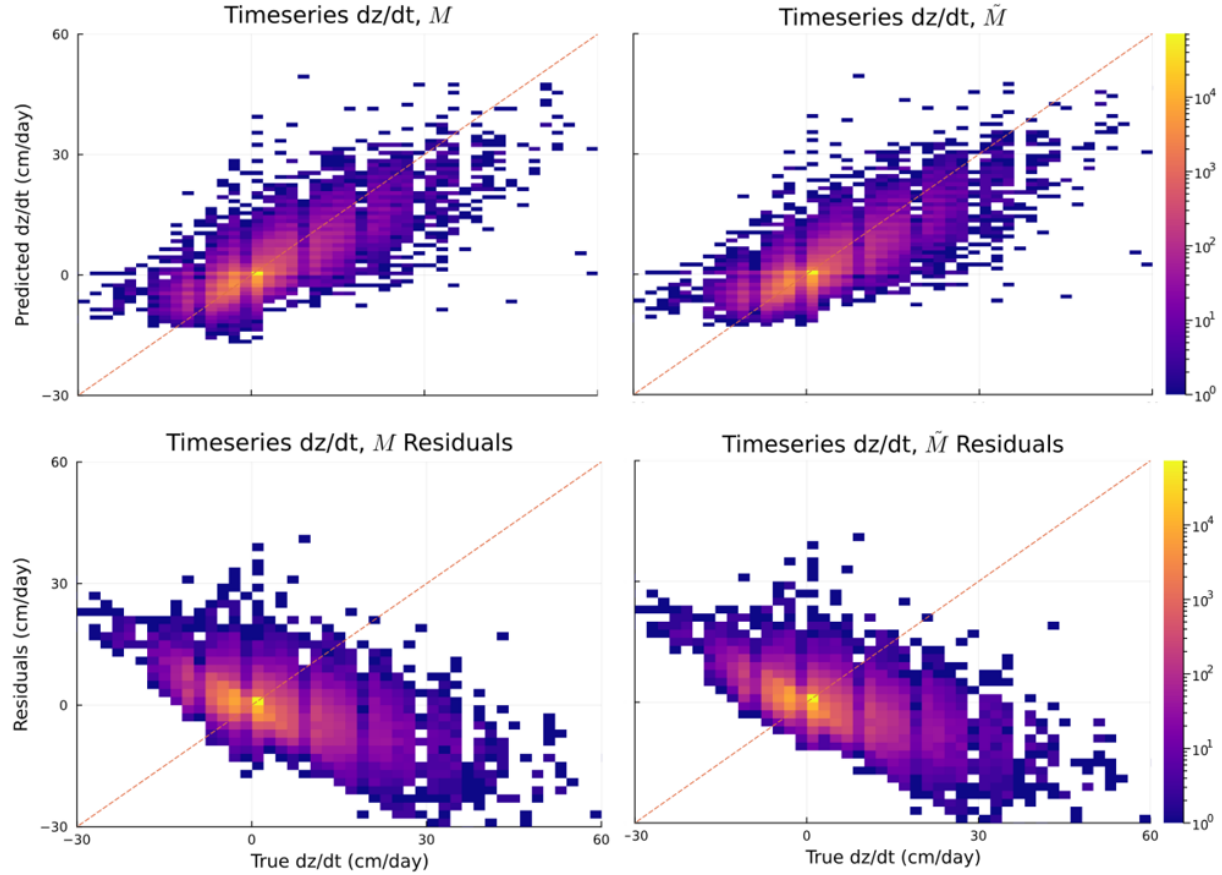


Figure 2.8: **Predicted vs. true dz/dt and residuals against the modeled target by the M and $\tilde{M}$ models.** Both models continue to under-predict extremes, despite the bias given to extremes during training. The magnitude of data in small dz/dt ranges relative to extremes could contribute to this phenomenon.

those prescribed.

The ALE plots for each feature are shown in Fig. 2.9. The average prediction of M is ≈ 0 , with negligible centering offsets, indicating a tendency to predict negative dz/dt after $T_{\text{air}} > 0$. M also exhibits a linear relationship in dz/dt with P_{snow} and stabilization at low T_{air} (where a minimum snowfall density would emerge), all aligning with physical expectations. Additionally, decreases in dz/dt are observed with increased solar radiation and wind speeds, reflecting the model’s learning of understood destructive processes. However, data availability and artifacts may influence these results; for instance, while higher relative humidity could lead to more condensation and surface melt, it also lowers estimated snowfall fraction under the data preparation used. Moreover, snow accumulation on sensors during heavy snowfall may produce saturated humidity readings linked to large dz/dt , and location biases may affect the value ranges.

ALE of Predictors

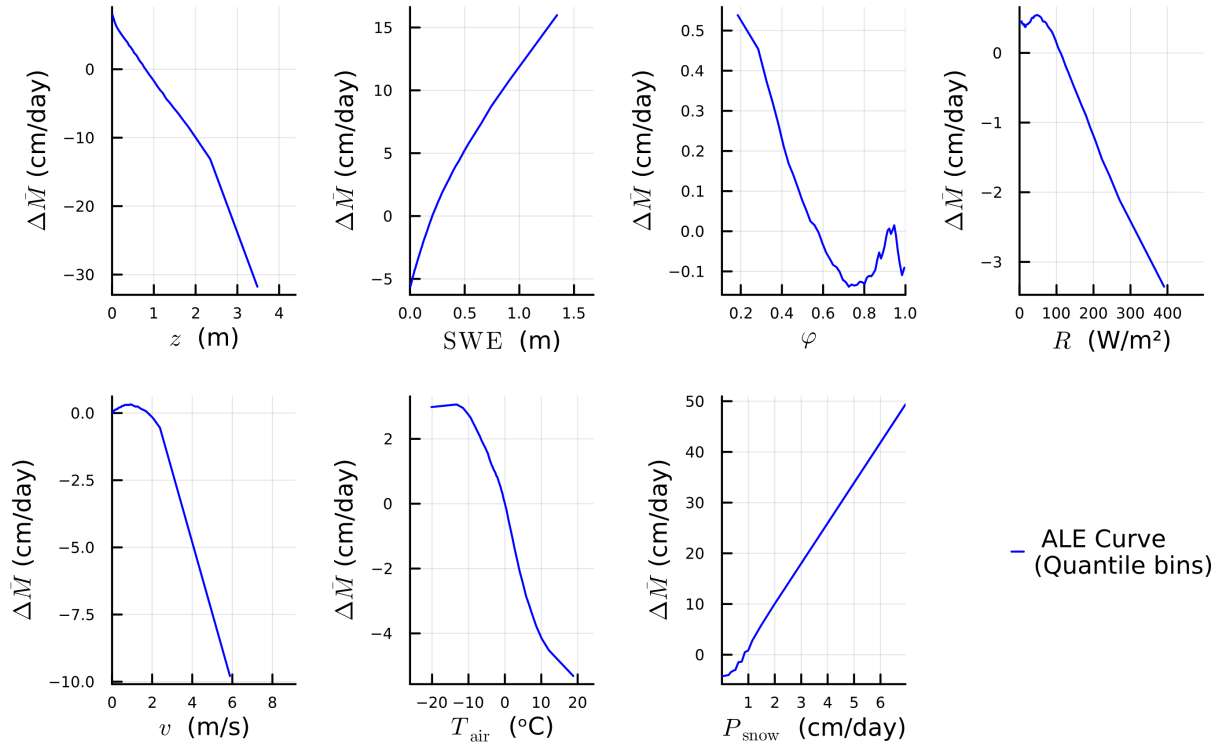


Figure 2.9: **Accumulated Local Effects (ALE) plot for M 's predictive features.** The ALE mitigates effects of feature correlations, providing a rough indication of feature importance in determining M 's output magnitude and a visual measure of physicality. x-axes mark the range of each feature, while y-axes show the change in M relative to the average prediction. Curves are binned by quantiles so each bin has at least 50 samples.

M can be directly queried over “slices” of its multidimensional input space, facilitating exploration of the model’s behavior in anticipated input regions that may not be observed in the current data. Unlike partial dependence plots, these slices are not averaged over training data ranges, allowing them to display outputs in regimes of inputs that are not physically viable. Additionally, patterns observed in one “slice” may not be conserved across others. Fig. 2.10 illustrates these slices, with areas lacking data slightly masked.

In Fig. 2.10a, contours reflect physical expectations, with even spacing indicating dz/dt is linear in P_{snow} , T_{air} becoming the dominant feature around $T_{\text{air}} = 0$, and the constraint $\Delta z \geq -z$. Fig. 2.10b examines how T_{air} and insolation impact snow depth at zero snowfall, showing depletion begins once $T_{\text{air}} > 0$ and with increasing R . All output values at zero snowfall remain nonpositive, adhering to the prescribed threshold. This slice demonstrates limited sensitivity to solar insolation at low R ,

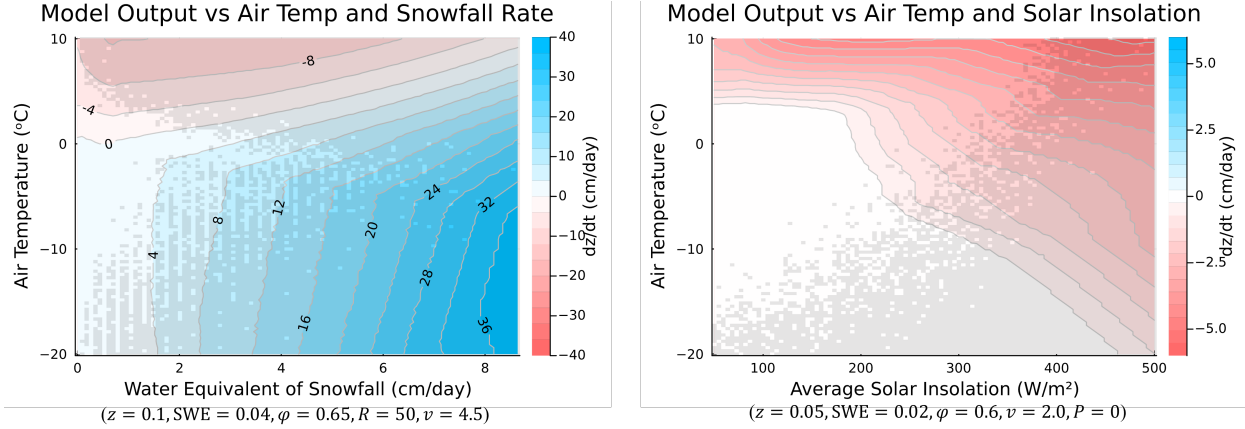


Figure 2.10: M outputs for two snow states, with other inputs held constant. In each case one threshold ($dz > -z$ or $dz/dt < 0$ for $P_{\text{snow}} = 0$) is visible. “Shaded” or “masked” pixels (partially greyed out) indicate areas of the visible parameter space with ranges not represented in the data, either from being unphysical or unobserved.

suggesting T_{air} is the dominant variable. This insensitivity may arise from high snow reflectivity at low incidence angles (and normal sensor orientation), shading effects reducing melting until higher radiation levels are exposed, or latent melting effects prevailing at low irradiance. Positive feedback loops where accumulating surface melt alters albedo could explain the transition in this slice. Overall, the model aligns with expected physical behavior over available data, though incorporating more data into extrapolated (shaded) regions could enhance universality.

Generalizability

To assess generalizability, Fig. 2.11 shows elevation vs. mean nonzero snow depth $\bar{z}_+$ for all training and testing sites, similar to Fig. 2.3. Sites are colored by the performance of M for SPE, RMSE, and NSE on z timeseries, along with density RMSE. The model succeeds comprehensively, with SPE errors under 20% for nearly all sites and most under 10%, while density errors remain predominantly below 15% RMSE³. It performs comparably on testing data from different elevations and climates, indicating robust generalizability, even for density calculations. No discernible trend with elevation appears, corroborating generalizability rather than elevation-induced effects.

³Post-Publishing Note: That is to say, we are seeing errors predominantly below the 15% threshold that is rarely broken in other literature.

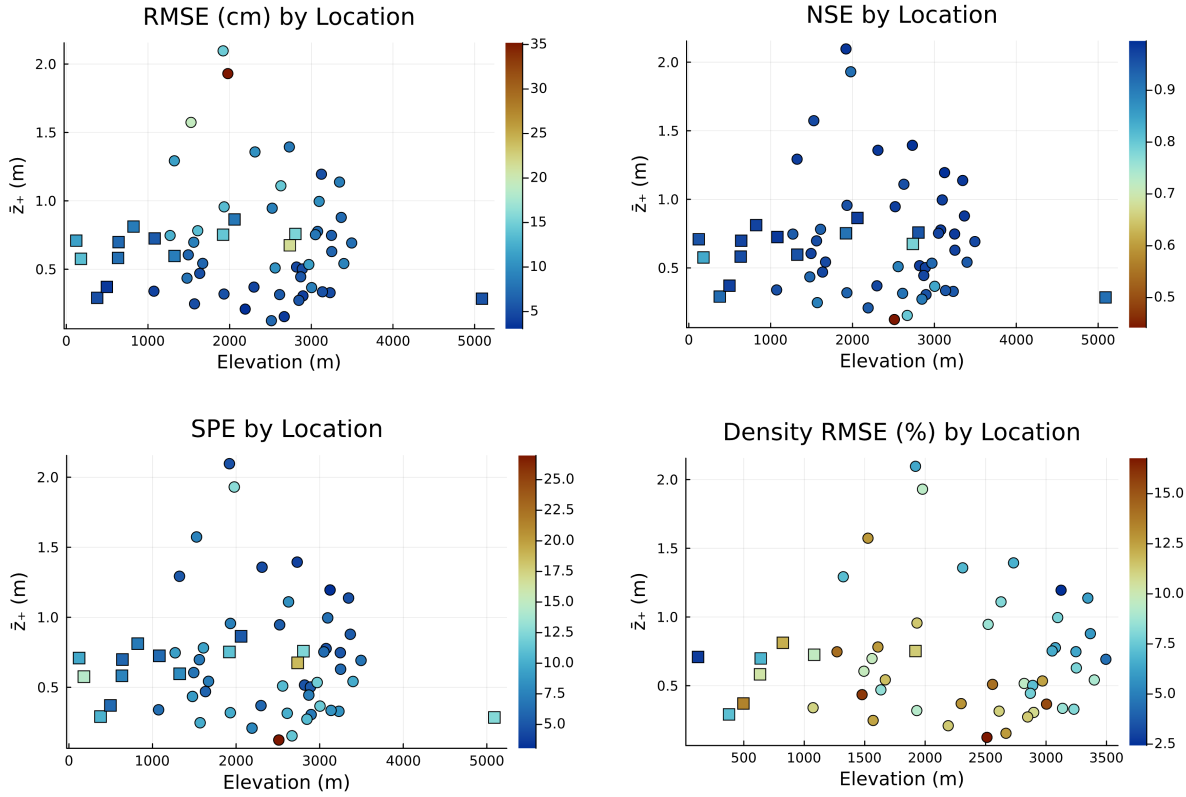


Figure 2.11: Performance of M across all training and testing sites with regards to RMSE, NSE, and SPE of depth timeseries, as well as RMSE for density timeseries over the sites where density was evaluated (see Appendix B). Testing sites are marked with squares instead of circles.

Finer Resolution Predictions

Time units only appear as rates in M via precipitation, the lower bound, and the output. By predicting rates dz/dt instead of accumulated dz , M can be evaluated at varying time-steps without retraining, by merely resetting the constraint function's scaling of $1/\Delta t$ without altering the trained predictive weights (which were trained only once in section 2.2 "Training", with $\Delta t = 1$ day). This flexibility allows testing at resolutions up to data limits, which for Kühtai is a maximum resolution of 15-minute intervals. While sub-hourly time resolutions are rarer in larger earth models, they are used in land models and will become increasingly frequent as research advances in fine-scale land-atmosphere interactions [6, 59]. Similarly, while multiday resolutions are rare, data availability can necessitate their use. Benchmarking across scales above and below the daily resolution used for training offers insights into the model's limits and potential applications. However, beyond a week, average input variables do not effectively capture critical input dynamics (e.g., monthly average

T_{air} would not reliably indicate time above freezing), leading to suppressed variations that undermine meaningful outputs, and alternative model choices would be more appropriate. Results of repeating the timeseries generation with M as described in Section 2.2 (“Testing”) for the Kühtai data with different values of Δt in the constraint but identical weights otherwise, spanning 15-minute to weekly resolutions, are shown in Fig. 2.12 and Table 2.2.

Table 2.2: **Performance of M at varying resolutions for z timeseries generation.** Error jumps beyond the daily training resolution, but performance remains near-constant between hourly and 15-minute resolution.

Statistic/Resolution	Weekly	Daily	Hourly	30-Minute	15-Minute
RMSE (m)	0.1762	0.1437	0.2008	0.1989	0.1999
NSE	0.759	0.806	0.708	0.710	0.709
SPE (%)	23.90	19.54	30.9	31.0	31.1

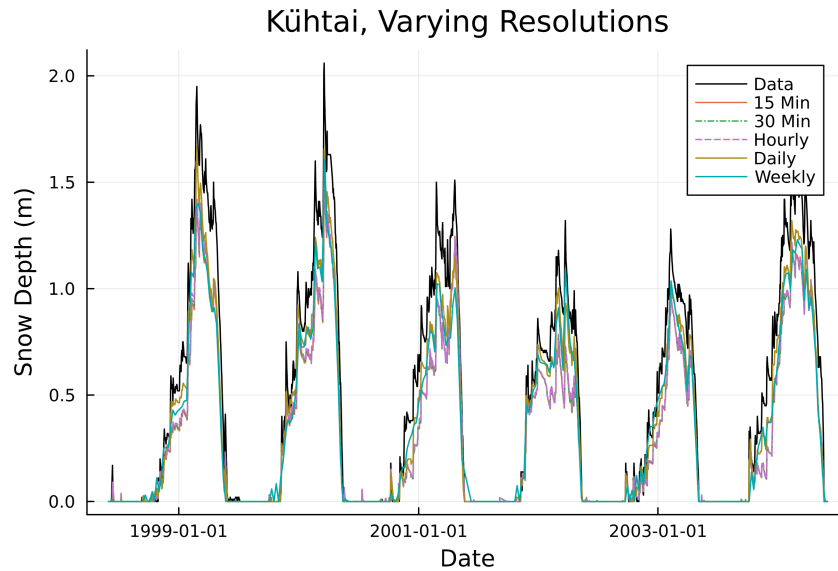


Figure 2.12: **Output of M at different resolutions for a subset of the site data from Kühtai, Austria.** The graph overlays outputs at different resolutions for direct comparison. Sub-hourly (15 min, 30 min) curves are nearly indistinguishable from the hourly curve.

The timeseries shown at this site exhibit a low bias; however, for all resolutions, M achieved over a 40% reduction in RMSE compared to Snow17 (not tabulated). RMSE increases at timesteps outside the daily interval used for training, though all sub-hourly resolutions yield nearly identical results without further trends. This performance loss may arise from the extreme values of dz/dt and precipitation observed at higher resolutions. While daily data show gradual snowpack increases,

finer resolutions might capture the same deposition over a few hours, leading to dz/dt values 10–20 times larger than those in the training data. Conversely, weekly averages smooth out extreme events, which can reduce the variance of outputs. Overall, M demonstrates an ability to transfer across temporal resolutions, though performance would likely improve with a broader range of dz/dt training values (for instance, incorporating both hourly and daily data).

2.4 Discussion

This study explored a simple, versatile data-driven framework to enhance physical parameterizations, focusing on generalizable snow parameterizations for climate modeling applications. Many choices were results-driven and informed by data availability, such as selecting widely measured variables to increase model applicability, though other choices could likely enable representation of additional processes. Data requirements limited sources to the SNOTEL network, which, while useful for local relationships, has quality issues compared to validation sites such as Col de Porte and Kühtai [49, 50], fails to represent large-scale heterogeneity effects, especially in mountain regimes [49], and lacks coverage of extreme conditions in scarcely sampled tundra and taiga biomes. The approach’s ability to extend to these unsampled terrains or perform on coarser grids with explicit large-scale effects within an Earth system model remains untested and will be the subject of a future paper⁴. Further adaptation of these models in a world of growing data volume, frequency, and quality offers an exciting opportunity for future research.

The neural model M generalizes well across locations, a crucial feature for both global consistency and local-scale modeling, particularly as climate change skews site statistics to become “new” locations. Snow17 struggles with such shifts, limiting its application on longer timescales [8, 49]. Our physics/ML hybrid approach offers a viable alternative. Generalizability is key for ML in climate models. For instance, the random forest model by Yang et al. [81]’s when applied to Chinese sites showed high out-of-sample biases and SPE compared to our model, despite similar RMSE. Their ML approach and others like Duan et al. [20] or Yang et al. [82] rely on location-specific variables (slope, aspect, topographic/vegetative indices, etc.) or

⁴*Post-Publishing Note:* This “paper” refers to Chapter 4, where the model shows reasonable ability to simulate in these regimes, though this was evaluated only at coarse resolution. Developing different networks for different biomes would be one possible approach, but this would require extra storage and dynamic identification of each grid cell’s land-type to indicate which parameter set to use. This also does not circumvent the issue of some biomes lacking sufficient observational data to train their respective networks, so augmenting the existing training data with results of future observational efforts would likely prove a more robust approach.

features like historical averages and microwave measurements [15, 63, 67, 69, 82], which hinder usage within earth system models. In contrast, Wang et al. [73] simulated SWE with recurrent neural networks using physical inputs with similarly high NSE scores and moderate generalizability, but required over 240 previous states for updates. Duan et al. [20] also used 180 days of forcing data per SWE prediction with various models at SNOTEL sites, requiring training times of 5–26 hours and hours of simulation time on a GPU. Our memoryless neural ODE model achieves comparable results to Wang et al. [73] with guaranteed consistency with physical bounds. It exhibits improved median MAE, RMSE, and NSE against all models from Duan et al. [20], with fewer inputs and significantly fewer computational resources (see Table 2.6). Steele et al. [65] used the same inputs as ours for a standalone ML model and a post-processing model for physical models. Both yielded higher SWE RMSEs (6 cm and 13 cm vs. ~ 4 cm in our model, see Table 2.4) and poorer generalization, though a slightly better derived density RMSE (implying z errors were similarly scaled to SWE errors). Both models required addition of a binary snow-presence variable to reset unphysical summer snowpacks, while our approach naturally adheres to prescribed bounds, eliminating such unphysical departures.

The best hyperparameters for timeseries generation differed from those for direct regression, underscoring better dz/dt predictions do not guarantee better accumulated seasonal timeseries. This further supports that our choices are going beyond merely matching magnitudes, and instead summarizing universal, memoryless physical processes. Notably, the parameterization faltered only when new climates introduced target magnitudes absent in the training data, rather than when locations presented different input feature magnitudes. While output magnitude extrapolation is limited as with many data-driven models, the input generalizability is a less common result, highlighting the benefits of enabling physical consistency. This underscores the need for more widespread snow sensing, particularly in extreme climates, to improve predictive power of such models.

The framework’s application of prescribed constraints around black-box models like neural networks enables easy modification of constraints or input variables. This interchangeability supported rapid experimentation and prototyping, and is synergistic for integration as a “plug-and-play” model that can adapt to available inputs and operational constraints. The approach provides linear scaling in input size with low computational overhead, which can reduce computational budgets while maintaining or improving accuracy compared to advanced models, permitting

simulations longer into the future or over finer grids. It can demonstrably act as a standalone predictive tool wherever inputs can be measured or inferred, which could be through observations, remote sensing, weather forecasts, or coupled models. This framework could be used, for example, for forecasting applications such as weekly skiing or hiking terrain predictions from weather forecasts, or tested for water supply simulation given snowpack data. The model structure is a strong candidate for simulating many types of constrained physical systems beyond snow, offering further avenues for investigation (see Appendix A).

Beyond testing the neural model M globally in a coupled setting with an entire climate model, other future directions of research could involve adapting M to continuous neural ODE structures [13] or more general timestepping schemes. While this study focused on localized relationships and comparisons with similarly-formulated standards, it did not address large-scale heterogeneity effects. Future adaptations to directly incorporate these effects might aid usage within coarser-scale (~ 100 km) models. Additional improvements could involve integrating data from NOHRSC/MADIS or detailed snow layer data to simulate temperature profiles. Alternative training strategies, such as using timeseries error as the loss function or gradient-free update rules to bypass recursive timeseries gradient issues could also be explored.

2.5 Conclusion

Using a location-agnostic and physically constrained neural ODE framework to parameterize the rate of change of snow depth, we were able to simulate seasonal snow depth with a median error of 8.8% across sites with varying climates and elevations, including some not seen during training. Though the parameterization was trained with daily data, it shows an ability to perform with moderate accuracy at other temporal resolutions without additional retraining of the model; however, retraining with higher-resolution training data may lead to further improvement. The parameterization’s structure reduces computational overhead while maintaining performance in depth simulation at the level of established, cutting-edge, or more detailed models. The design is conducive for usage in prognostic models or can be adapted to alternatively predict variables such as SWE. When driven solely by meteorological data as a standalone model, the parameterization framework can recreate seasonal timeseries with comparable error without retraining or site calibration — an improvement over other established models. In most cases, it matches or outperforms the Anderson Snow17 model in simulating seasonal snow depth,

offering an efficient formulation for use within physical models and an alternative to prevailing parameterizations.

The proposed framework demonstrates potential for a wide array of applications for both long-term climate simulations as well as short-term forecasting applications. The means of enforcing hard constraints structurally provides a simple but powerful technique for predictive modeling that can be applied beyond snowpack modeling to different climate processes or physical parameterizations.

2.6 Data Availability

The SNOTEL data utilized in this study was available via the National Water and Climate Center, which lies under the United States Department of Agriculture. Data reports of the SNOTEL data were generated using the online portal found at <https://www.nrcs.usda.gov/wps/portal/wcc/home/>. The data from Col De Porte [41] can be found at the Observatoire des Sciences de l'Univers de Grenoble DOI portal https://doi.osug.fr/public/CRYOBSCLIM_CDP/CRYOBSCLIM.CDP.2018.html, and data from Kühtai [37] can be found as the supplementary material from <https://doi.org/10.1002/2017WR020445>. Raw data from Sodankyla [23] can be found from the Intensive Observation (sensors 8, 11) data portal <https://litdb.fmi.fi/ioa.php> and automated weather station (sensor 15, portal https://litdb.fmi.fi/luo0015_data.php). Reynolds Mountain East [58] data was obtained from an ESM-SnowMIP repository <https://www.geos.ed.ac.uk/~ressery/ESM-SnowMIP.html>. Raw Yala Basecamp [61, 66] data was retrieved from <https://rds.icimod.org/Home/DataDetail?metadataId=26859> and <https://rds.icimod.org/Home/DataDetail?metadataId=1972554>. Rofental [74] data was retrieved from <https://datapub.gfz-potsdam.de/download/10.5880.FIDGEO.2023.037-MNveB/>.

The data in this study was processed from these sources. Code for scraping and cleaning SNOTEL data and tutorials for data retrieval and training/modifying the neural models, are available at https://clima.github.io/ClimaLand.jl/dev/generated/standalone/Snow/base_tutorial/. CSVs of the training/testing data are also available here, as a quality-controlled fully observational dataset of physical variables for calibration and ML applications.

2.A Threshold Constraint Layers

Defining Threshold Constraint Layers

Since $\text{ReLU}(x) = \max(x, 0)$, we can re-express the minimum and maximum functions as

$$\begin{aligned} \max(x, y) &= y + \text{ReLU}(x - y) = \\ &\text{ReLU}(y) - \text{ReLU}(-y) + \text{ReLU}(x - y) = \max(y, x), \end{aligned} \quad (2.7)$$

$$\begin{aligned} \min(x, y) &= y - \text{ReLU}(y - x) = \\ &\text{ReLU}(y) - \text{ReLU}(-y) - \text{ReLU}(y - x) = \min(y, x). \end{aligned} \quad (2.8)$$

Then for a model output p and any construction f serving to threshold p , the bounds $\max(f, p)$ or $\min(f, p)$ can be explicitly implemented with a single depth-3 fixed-weight layer with no biases acting on input $[p, f]^\top$ with ReLU activation, followed by an accumulation with no activation:

$$\begin{bmatrix} \pm 1 & 1 & -1 \end{bmatrix} \times \text{ReLU} \left(\begin{bmatrix} \pm 1 & \mp 1 \\ 0 & 1 \\ 0 & -1 \end{bmatrix} \times \begin{bmatrix} p \\ f \end{bmatrix} \right) = A_{1\pm}^\top \text{ReLU} \left(A_{2\pm}^\top \begin{bmatrix} p \\ f \end{bmatrix} \right) \quad (2.9)$$

where $+$ indicates $\max(f, p)$ and $-$ indicates $\min(f, p)$, and the ReLU acts element-wise. Eq. 2.9 offers a simple formulation, but the symmetry of the max and min functions permits resonant structures — different weights yielding the same output. If the signs of p or f are unknown, both $+f$ and $-f$ (or $+p$ and $-p$) must be passed through the ReLU, along with $p - f$, to preserve all necessary information. However, if f is always nonnegative (or nonpositive), the layer can be simplified from three layers to two, as passing $-f$ (or $+f$) becomes redundant and its ReLU always evaluates to zero. Similarly, if the threshold obeys $f \geq C$ (or $p \geq C$) for some constant C , a similar reduction is possible by including a bias term along with $A_{2\pm}$ and $A_{1\pm}$. The same reductions apply if p also exhibits similar properties. Fig. 2.13a illustrates the generalized structure for one-sided threshold constraints on a predictive component discussed in section 2.2 “Model Structure.”

Likewise, for a simultaneous upper bound f_+ and lower bound f_- on p for any constructions f_+, f_- satisfying $f_+ \geq f_-$, we have

$$\max(\min(p, f_+), f_-) = \text{ReLU}(f_-) - \text{ReLU}(-f_-) + \text{ReLU}(\alpha), \quad (2.10)$$

where

$$\alpha = \text{ReLU}(f_+) - \text{ReLU}(-f_+) - \text{ReLU}(f_-) + \text{ReLU}(-f_-) - \text{ReLU}(f_+ - p), \quad (2.11)$$

so the threshold can be explicitly implemented with a sequence of two fixed-weight layers containing no biases acting on input $[p, f_+, f_-]^\top$, followed by an accumulation with no activation:

$$\begin{aligned} & \begin{bmatrix} 1 & 1 & -1 \end{bmatrix} \times \text{ReLU} \left(\begin{bmatrix} -1 & 1 & -1 & -1 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \right. \\ & \quad \left. \times \text{ReLU} \left(\begin{bmatrix} -1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \end{bmatrix} \times \begin{bmatrix} p \\ f_+ \\ f_- \end{bmatrix} \right) \right) \end{aligned} \quad (2.12)$$

$$= \mathbf{A}_{1+}^\top \text{ReLU} \left(\mathbf{A}_3^\top \text{ReLU} \left(\mathbf{A}_4^\top \begin{bmatrix} p \\ f_+ \\ f_- \end{bmatrix} \right) \right), \quad (2.13)$$

which takes advantage of the identity $\text{ReLU}(\text{ReLU}(x)) = \text{ReLU}(x)$. Like the one-sided threshold example, many resonant structures exist according to symmetry, and bounds on the thresholds or p permit layer reductions. Fig. 2.13b shows the generalized structure for a two-sided threshold constraint function f outputting f_+, f_- on the predictive component given in section 2.2 “Model Structure.”

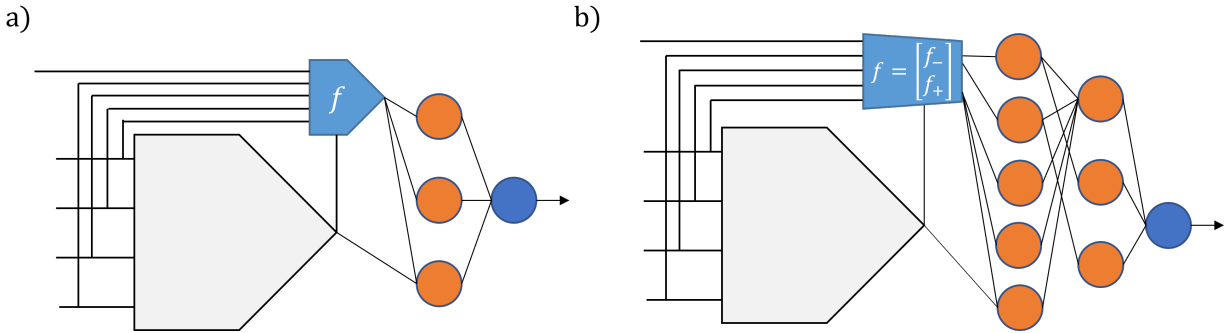


Figure 2.13: General Framework Structures. In all graphics, inputs are on the left, with predictions on the right, and the grey pentagon depicts the predictive component from Fig. 2.1. Knowledge of function f can simplify layers by removing certain orange (ReLU) nodes, with resonant structures possible from the max, min symmetry. Black weights are fixed at +1 or -1, with no biases or training. (a) General structure for a one-sided constraint (max or min) on the prediction. (b) General structure for a two-sided constraint (enforced range) on the prediction.

These constraint structures can adapt to any functional constraint f of any input (including those independent of predictive inputs) or even the predictive component

output p . This versatile framework exhibits minimal computational overhead without increasing runtime complexity. This thresholds an output rather than imposing an invariance or zeroing of a derivative, but this constraint class is relevant to many physical or nonnatural systems. This provides a simple means for emulating many systems when paired with universally approximating predictive models, or integrating data-driven parameterizations into larger models while enforcing constraints like physics or conservation laws. Absolute boundaries also enhance stability in time-stepping by keeping outputs realistic.

Many code packages and languages cannot compute gradients for logical branches or min/max functions. This structure bypasses that limitation, allowing constraints to be enforced during training, even on legacy systems with minimal functionality. This mitigates the need to penalize the loss function, which can impede learning of the main objective [55]. Enforcing constraints during training enables gradients and weight updates better-suited for predicting values within the boundaries. Unlike other penalty-free approaches like projecting outputs into constrained spaces, this method does not require constant or predefined thresholds and can explicitly predict boundary values rather than asymptotically approach them, which can inhibit learning by reducing gradient magnitudes [39]. Constraints under this construction can be analytically defined, or even parameterized and “learned” from data during training, even simultaneously with the predictive component (e.g., a network for prediction and a network for boundary values, trained simultaneously), predicting both thresholds and values for entirely data-driven modeling⁵. These functional forms can also be combined and stacked in larger networks or applied to non-network models. Figure 2.13 depicts the threshold constraint layers enveloping the predictive component from section 2.2 (“Model Structure”), but they could be placed inside larger networks, layered, or stacked as part of a larger predictive model, or applied on any non-network model.

The primary benefit of this framework is enabling analytically prescribed bounds for trainable models during training. However, by representing the “if over, then change” condition as a vectorized algebraic operation rather than logical branches, it eliminates branch misprediction in compilers and maximizes scalability for broad-

⁵*Post-Publishing Note:* Constraints can also be altered after training, without altering the learned weights of the predictive model. This preserves the predictive model without modification but changes which ranges of outputs get clamped to the supplied bounds. Such changes are useful for applications like adaptive time-stepping, like was explored in this study, but could conceivably “uncover”/expose regimes of model outputs with poorer accuracy due to their lack of emphasis during the original training.

casting and parallelization at large scale. This has performance implications for CPUs (see Table 2.6) but especially for GPUs with different (or no) protocols for branching [54]. When inputs for both p and f are similar, constraints can be implemented with one skip connection, streamlining the architecture. These constraint structures could be equivalently expressed using Maxout [26] or nested networks for a given constraint, but maintaining a single-network form with one skip connection results in faster training and sufficient variety in constraint expression.

In the bigger picture, the framework’s adaptability and guarantee of constraints makes it a strong candidate for emulating systems where complex processes defy full analytical modeling but are bound by defined limits. This versatility opens doors for use-cases even beyond snow modeling. For instance, in two-sided threshold scenarios, M can be viewed as interpolating between two boundaries, which we anticipate as offering utility in areas like predicting drag between turbulent and viscous limits or, transport in superdiffusive regimes. The approach holds potential to contribute to a comprehensive understanding of global dynamics, opening exciting avenues for future investigation.

2.B Data Methods

Snow data comes from ground stations, aerial lidar surveys, or satellite images calibrated with ground measurements [62]. Remote sensing offers broader coverage crucial for global modeling but can suffer obfuscation and lack the resolution to capture localized snow processes. Ground sensors can capture local effects but face challenges like malfunctions, terrain biases, and limited coverage, and aerial surveys track fewer variables. The SNOTEL network, with over 900 sites, uses automated sensors like snow pillows for SWE, ultrasonic sensors for depth, and weather instruments, transmitting data via ionospheric radio reflection [75].

The SNOTEL network is rare in providing simultaneous collocated hydrometeorological data across varied climates, but faces notable data uncertainties like precipitation gauge undercatch or unphysical sensor values [28, 57]. Not all sites measure all variables [56], with many in wind-shielded or flat terrain, limiting climate diversity for “universal” model calibration. All sites measure SWE, z , and precipitation, but availability of variables like T_{snow} , T_{soil} , R , φ , and v vary. Removing T_{soil} (due to its correlation with $\text{ReLU}(T_{\text{air}})$) and T_{snow} increased the usable sites to 44 in the continental U.S. and 7 in Alaska post-processing.

Established snow science evaluation sites like Col De Porte [41], Kühtai [37],

Reynolds Mountain East [58], and Sodankyla [23] were tested alongside Alaska SNOTEL data, which biases the evaluation towards lower elevations and similar $\bar{z}_+$. To address this, data from Yala Basecamp [61, 66] and two Rofental sites [74] were included, though additional daily observational timeseries for all variables are presently minimal. Non-located measurements of snow depth (z), snow water equivalent (SWE), or snow density (ρ_{snow}) can lead to inconsistencies and biases in bulk density estimates. Training on such inconsistencies could benefit accuracy in large-scale climate modeling with gridded data, as it better matches the anticipated variability from coarse-graining, but it can also introduce validation biases if observation protocols differ from the training data. Therefore, all training and testing sites were assessed for the collocation of z and SWE data collection using literature, imagery like [62], or direct communication with site representatives. This limited density evaluations to K  htai and CONUS/Alaskan SNOTEL sites.

Phenomena like wind drift can confound snow pillow data, as strong winds away from the wind sensor can push snow onto the pillow, creating positive measured dz/dt with no precipitation and insufficient wind speeds for drift, as observed by [50] in some SNOTEL sites. Our constraints in section 2.2 (“Model Structure”) can be violated by such events. However, only 1.5% of the training data and 2.2% of the testing data showed $dz/dt > 0$ without precipitation, and established models like Crocus and SNOWPACK also do not account for positive growth due to wind redistribution when run standalone. These models instead focus on the compaction or erosion of snow by wind [40, 71]; destructive effects our model captures. Ongoing work aims to integrate redistribution effects into snow models, and as this paper serves to introduce a framework compared to existing parameterizations, we find accounting for such effects beyond the scope of this study, representing an exciting area for future research. We invite interested parties to revise the variables, thresholds, and datasets in this initial formulation to the benefit of the community, as we envision broader adoption of this proposal beyond our specific formulation and calibration.

Data Cleaning Procedures

Established methods exist for cleaning SNOTEL daily SWE, T_{air} , and precipitation values, beginning with Serreze et al. [60] and extended by Yan et al. [79], but little consensus exists for depth or meteorological variables, particularly those available at hourly frequencies.

Raw hourly and daily timeseries for all input variables were retrieved from the NRCS database, covering all entries available up to 2024-02-02. Bounds were applied to each sensor based on physical limits and limits in the SNOTEL sensor handbook [68], removing any violative data. All solar, humidity, and wind speed timeseries were manually inspected, and any suspect or unphysical periods were flagged (see the function `manual_filter()` in the code repository for a list of suspect periods, listed in Data Availability). The following steps were then taken in order per site:

1. Weekly maximum hourly wind speeds were determined, and the median w_{max} and interquartile range IQR were calculated. Observed wind values w_i were flagged if $\frac{w_i - w_{max}}{IQR} > 6$. For timeseries with over 24 flagged values, “blocks” of flagged values were grown in 72-hour steps until no further flags existed. If more than 5% of values in a block were flagged, all observations in that block were flagged.
2. Unflagged hourly wind speed, solar radiation, relative humidity, and air temperature observations were binned into two-week windows and by hour, with the mean calculated for each hourly bin to generate an annual profile (24 hours per 26 biweeks). Gaps of 6 hours or less were filled using linear interpolation, while gaps of 6 to 24 hours were filled using the appropriate profile.
3. For hourly z data, for each non-missing/flagged observation z_i , the values $dz_+ = z_{i+1} - z_i$, $dz_- = z_i - z_{i-1}$, and $dt_+ = t_{z_{i+1}} - t_{z_i}$ were calculated. A threshold $Z = 20$ inches was picked and z_i having $dz_- \geq Z$ and $dz_+ \leq -Z$, or $dz_- \leq -Z$ and $dz_+ \geq Z$ were flagged. A “rut” of bad data began when $dz_- \geq Z$ and $|dz_+| \leq Z$, and continued until $|dz_+| \geq Z$, all observations in a rut were flagged. A rut lasting more than 20 observations or $dt_+ > 30$ days resulted in all observations being flagged until $z_i = 0$. From April through August, after the first time reaching $z_i = 0$, nonzero z_i were flagged. All z_i having $\frac{z_i}{SWE_i}$ less than 1, over 50, or missing SWE_i were flagged. This procedure was iterated until no more z_i were flagged, and was designed based on the structure of sensor errors in the hourly z data. This was only for the hourly data, as daily SNOTEL depth data are quality checked.
4. All hourly timeseries were then binned into three 8-hour bins per day. For solar radiation, wind speed, air temperature, and humidity, the mean of all non-missing/flagged values per bin determined (if all were missing, a missing value was given). These three averages were averaged to create the daily

values, and a daily value reported as missing if any 8-hour bin average was missing. For hourly z , the day's value was the first available observation, or missing if all observations that day were missing.

5. The annual maximums of all generated daily solar radiation observations were determined, and the median $R_{\max}$ and interquartile range IQR of these maximums determined, and all daily solar values R_i with a score $\frac{R_i - R_{\max}}{\text{IQR}} > 2$ were excised. This only removed a handful of individual irregular spikes in the rolled-up data and left most sites unaltered.
6. The converted daily timeseries was coalesced with the daily raw timeseries to form a complete daily timeseries for the site. Raw daily z and air temperature values took priority over converted-hourly values if both values existed, and converted-hourly values for solar radiation, relative humidity, wind speed took priority over daily values if both existed (only the raw daily data was used for SWE and accumulated precipitation).
7. Air temperature corrections to the air temperature data were applied in accordance with [4] and associated metadata of which sites to correct, as of May 2024.
8. Standard quality control procedures and flagging for SWE, accumulated precipitation values, and air temperature as given in [60] and extended by [79] were implemented. Inconsistent water years with maximum SWE at least 5% greater than the associated accumulated precipitation value were excised in accordance with this protocol.
9. Remaining accumulated precipitation observations were edited to account for gauge undercatch, following the procedure outlined in [43] used in [80].
10. Daily timeseries of z were compared to the quality controlled SWE, and any values showing $\frac{z_i}{\text{SWE}_i}$ less than 1, over 50, or missing SWE_i were excised.
11. For all variables, gaps of 3 days or less were then filled via linear interpolation.
12. All data was then scaled into SI units (z , SWE, and accumulated precipitation to meters from inches, relative humidity scaled from 0-1, and wind speed to meters per second from kilometers per hour).
13. Only days with complete cases (no missing values in all variables) were extracted, and sequential differences $\Delta z_i = z_{i+1} - z_i$, $\Delta \text{SWE}_i = \text{SWE}_{i+1} - \text{SWE}_i$,

$p_i = AP_{i+1} - AP_i$ were calculated, where AP is accumulated precipitation. Only values where $\Delta t_i = t_{i+1} - t_i = 1$ day were kept, and the target $\frac{dz}{dt}_i \approx \frac{\Delta z_i}{\Delta t_i}$ created, as well as analogous $\frac{dSWE}{dt}_i$ and $P_i = p_i/\Delta t_i$. Days with $P_i < 0$ due to resetting of the water year were changed to $P_i = 0$.

The data at this point was saved. Upon importing for usage in model training, precipitation P_i was split into snow P_{snow} and rain P_{rain} based off temperature and humidity values using the snow fraction equation from [32] with over an 88% success rate across the northern hemisphere:

$$f_{\text{snow}} = \frac{1}{1 + e^{\alpha + \beta T_{\text{air}} + \gamma \varphi}}, \quad (2.14)$$

with $\alpha = -10.04$, $\beta = 1.41 \text{ } ^\circ\text{C}^{-1}$, and $\gamma = 9$ (with the relative humidity $\varphi \in [0, 1]$). The correlation of P_{rain} data after quality control procedures and feature engineering to dz/dt and $dSWE/dt$ was the lowest of all variables among the training data at $r = -0.016$ and 0.005 , respectively, informing the choice to also remove it from the dataset to further increase computational simplicity and scalability of the final model. Reintroduction and retraining including the variable on the optimized model structure to verify this choice resulted in negligible changes in performance at the site level and on average.

Three exceptions apply to the above procedure specifically for this set of training data, which might not apply for other SNOTEL sites, and are as follows:

1. Air temperatures in the continental US were bounded below by -40°C and air temperatures in Alaska were bounded below by -50°C instead of the instrument limit of -60°C to remove on average 1-2 individual suspect temperature spikes per site.
2. All raw $z_i > 175$ inches were flagged (a bound solely for removing unphysical sensor spikes in this specific training data, and should be checked for alternative data).
3. For SNOTEL site 1122, the averaged air temperature hourly timeseries took priority over the daily timeseries when coalescing data.

The code to scrape SNOTEL data from the NCRS database as and apply the above processing as well as a tutorial has been made publicly available in the code repository.

The other sites were provided varying degrees of quality control beyond unit conversion and generation of targets dz/dt and $dSWE/dt$ from the resulting data. For Col De Porte, Kühtai, and Reynolds Mountain East, no action was taken beyond collecting data into daily timeseries (averaging all data except taking the first available z , SWE measurement per day, no need for intermediate 8-hour blocks). Precipitation was also split into rain and snow following the same procedure as SNOTEL data for Kühtai. In the Rofental's Bella Vista site data, a nonzero offset of precipitation data was subtracted from dates after 2022-01-01, and otherwise both sites were collected to daily data directly and treated with the same precipitation undercatch procedures as the SNOTEL data. At the Yala Basecamp site, only the 2018 year was taken due to feature availability. All negative SWE, z values were set to zero, and gaps in the SWE data up to 9 consecutive values were filled with linear interpolation. The same undercatch procedures as SNOTEL were provided, and data was aggregated to the day level directly (for z , the median was taken to ignore sensor spikes, otherwise the mean for all variables). For Sodankyla, the following measures were taken from the raw data series beyond unit conversion, target feature creation in the same manner as the SNOTEL sites, and direct aggregation to daily level:

1. 1-minute SWE data was aggregated to the 10-minute level (the same level as other variables).
2. All 10-minute variables had gaps up to 9 consecutive values filled with linear interpolation.
3. All solar radiation data less than zero was set to zero.
4. Missing z data from May to November 2016 was set to zero.
5. The same undercatch procedure was applied to precipitation data as the SNOTEL data.

2.C Hyperparameters and Model Benchmarking

Hyperparameters

Optimal hyperparameters are summarized in Table 2.3. Scores were evaluated using 44-fold leave-one-out cross-validation with a batch size of 64, tracking performance every 10 epochs over 200 epochs. Most timeseries trials performed optimally when training for approximately 100 epochs. A nonzero value of n_2 emphasized extreme points in the custom loss function, enhancing accumulated predictions,

particularly for datasets with few extreme samples. The optimal hyperparameters yielded a network size of 435 trainable parameters for the z network and 540 for the SWE network, compared to around 50 empirically tuned constants (parameters and internal code) in Snow17 for predicting SWE and z .

Table 2.3: Hyperparameter results. For timeseries, NSE and RMSE were the primary metrics, while RMSE was the main metric for regression. $n_1 = 2$ (L_2 -like metric) provided the lowest RMSE for all choices, which is unsurprising. However, it is interesting that $n_1 = 2$ also minimized SPE (an L_1 -like metric) compared to $n_1 = 1$. N represents the number of consecutive days averaged when building training data, n represents the width of the model’s mixing layer, n_1 represents the power scaling of the prediction error in the loss function, and n_2 represents the power scaling of the target magnitude in the loss function. The best value is shown for each of timeseries generation (“Series”) as well as direct regression over the data (“Regression”).

Parameter	Range	Series (z)	Series (SWE)	Regression (dz/dt)
N	1, 2, 3	1	1	1
n	3, 4, 5, 6	4	5	6
n_1	1, 2	2	2	2
n_2	0, 0.1, 1, 2, 4	4	2	2

Model Benchmarking

Table 2.4 presents the means and medians of different model configurations for timeseries generation at the testing sites. The neural parameterization is also compared to another network, M_{ifelse} , which has identical predictive structure but has no boundaries during training and only calculates and applies thresholds post-training through vectorized if/else logic. The regression RMSE scores between M and M_{ifelse} on training and testing data were within 1-2 mm/day, with M_{ifelse} performing better on training regression, despite notable superiority in M for timeseries accuracy. This trend of better training regression for M_{ifelse} but worse timeseries generation persisted across repeated training trials, reinforcing the notion that incorporating bounds during training enhances physical representation and generalizability, rather than minimizing training loss at the expense of other beneficial properties.

Table 2.5 lists the p-values from Wilcoxon Signed Rank tests used in this study. This non-parametric test assesses the significance of differences between matched samples (timeseries RMSE, in this case), comparing the performance of the presented framework over Snow17 and M_{ifelse} . The lack of significance for M at validation sites was expected, as all models were calibrated for performance on this data. However, significant improvements in out-of-sample testing data highlight the advantages of

Table 2.4: **Performance of the models (for SWE and z) and parameterizations (for z) in this study across testing sites, using labels from section 2.2 (“Testing”).** Medians are listed, with the mean in parenthesis alongside the median. Another network M_{ifelse} has been included, which trains without constraints and applies them during testing with if/else statements, to highlight the performance gains over out-of-sample data from including the framework developed in this paper. Subscripts on metrics indicate whether depth or SWE was benchmarked.

Metric	Standalone Model (SWE, z)		Parameterization (z)		
	$\tilde{M}$	SN17	M	SN17O	M_{ifelse}
MAE_z (cm)	6.9 (7.8)	8.2 (16.1)	4.6 (5.9)	6.9 (8.9)	6.4 (8.0)
RMSE_z (cm)	11.3 (12.9)	12.9 (23.7)	8.3 (9.6)	10.1 (14.3)	10.6 (13.0)
NSE_z	0.915 (0.870)	0.914 (0.347)	0.955 (0.936)	0.937 (0.782)	0.925 (0.875)
SPE_z (%)	11.9 (13.4)	11.5 (27.6)	8.8 (9.6)	9.8 (15.4)	11.5 (13.5)
MAE_{SWE} (cm)	2.1 (2.3)	2.5 (4.2)	-	-	-
RMSE_{SWE} (cm)	3.7 (4.1)	4.4 (7.4)	-	-	-
NSE_{SWE}	0.931 (0.862)	0.933 (0.663)	-	-	-
SPE_{SWE} (%)	11.3 (14.2)	11.3 (21.4)	-	-	-

the presented approach. $\tilde{M}$ ’s significant difference from Snow17 in validation sites was unexpected, potentially due to Snow17’s calibration prioritizing SWE directly (one variable), while $\tilde{M}$ benefits from optimizing z and SWE as direct inputs for SWE prediction.

Table 2.5: **Statistical significance testing of model RMSE at validation and testing sites.** p-values of the Wilcoxon Signed Rank test are shown, which compares the predictive power of two models. Labels follow those in Table 2.4. Significant ($p < 0.05$) values are highlighted in bold. M_{ifelse} has been added to underscore the significant improvement on out-of-sample data when applying the demonstrated framework.

Site RMSE	M vs. SN17O (z)	$\tilde{M}$ vs. SN17 (z)	$\tilde{M}$ vs. SN17 (SWE)	M vs. M_{ifelse} (z)
Validation Sites	0.673	0.006	0.029	0.903
Testing Sites	0.013	0.135	0.268	0.0003

Table 2.6 presents time/memory benchmarking. Testing was conducted on one Intel i9 CPU (no GPU). Models were tested in “Column” mode, processing one location’s inputs at a time (like a site simulation), and in “1.5M Grid” and “15M Grid” modes, evaluating input vectors of ~ 1.5 million and ~ 15 million locations’ inputs at once (like a global model at 10km or ~ 3 km land resolution), by stacking 14 or 141 copies of all SNOTEL inputs, respectively. Average memory/time for a single evaluation

(excluding garbage collection and compilation) were tracked and normalized by the number of inputs. Column mode was averaged over 10 SNOTEL data passes (1.06 million trials), while 1.5M and 15M Grid results were averaged over 250 trials. Snow17 can only iterate between locations in a Column-like mode. Both M and M_{ifelse} determine boundaries and adjust outputs accordingly, but M does this structurally, while M_{ifelse} uses vectorized boundary creation and broadcasts conditional if/else logic, creating the opportunity for branch misprediction effects.

While M_{ifelse} is slightly faster with less memory in Column mode (comparing and adjusting one value against two values is quicker than processing three values through matrix multiplication), M is faster by about 10% over gridded inputs (0.147 and 1.47 seconds total for all 1.5 million and 15 million inputs, respectively). Snow17 with data assimilation requires more time and memory than without, as assimilation occurs after the primary evaluation. $\tilde{M}$ is larger than standalone Snow17 (and about twice that of M , as expected) but evaluates faster, which benefits long simulations across many locations in global models. However, M is quicker and requires less memory than Snow17. Results may vary based on programming languages, libraries, compilers, or protocols used for implementation, but one could always train utilizing the framework (the main benefit) and alter boundary implementation afterwards as desired. We anticipate these findings would extend to GPUs that often lack optimized branching, suggesting further advantages for large-scale model integration. Additional comprehensive benchmarking remains an avenue for future research.

Table 2.6: Time and memory benchmarking of all models (for z and SWE) and parameterizations (for z), listing required resources per evaluation for single-instances (“Column”) or per-instance over roughly 1.5 million or 15 million instances (“1.5M Grid” and “15M Grid”) simultaneously. Snow17 only evaluates single instances, and M_{ifelse} is listed to compare constraint layers against vectorized if/else post-processing. All benchmarks were evaluated on a single Intel i9 CPU.

Metric	Standalone Model (z , SWE)		Parameterization (z)		
	$\tilde{M}$	SN17	M	SN17O	M_{ifelse}
Column, T (μs)	2.6	3.6	1.3	3.8	1.0
Column, Allocated Memory (KB)	2.5	1.6	1.2	1.7	1.1
1.5M Grid, T (μs)	0.21	-	0.099	-	0.11
1.5M Grid, Allocated Memory (KB)	0.80	-	0.37	-	0.34
15M Grid, T (μs)	0.21	-	0.099	-	0.11
15M Grid, Allocated Memory (KB)	0.80	-	0.37	-	0.34

References

- [1] E. A. Anderson. *A point energy and mass balance model of a snow cover*. NOAA technical report ; NWS 19. U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, Office of Hydrology, Silver Spring, Md, 1976. URL <https://repository.library.noaa.gov/view/noaa/6392>.
- [2] E. A. Anderson. Snow accumulation and ablation model – snow-17, Jan 2006. URL <https://www.weather.gov/media/owp/oh/hrl/docs/22snow17.pdf>.
- [3] D. W. Apley and J. Zhu. 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, 06 2020. ISSN 1369-7412. doi: 10.1111/rssb.12377. URL <https://doi.org/10.1111/rssb.12377>.
- [4] J. Atwood, B. Domonkos, K. Hill, T. Brosten, T. DeMarco, M. Hulstrand, D. Tappa, L. Austin, and A. Buckman. Evaluation of YSI temperature correction equations for bias-reducing SNOTEL network temperature data, May 2023. URL https://www.nrcs.usda.gov/sites/default/files/2023-05/Final_Temperature_Correction_Study05262023.pdf.
- [5] E. H. Bair, A. Abreu Calfa, K. Rittger, and J. Dozier. Using machine learning for real-time estimates of snow water equivalent in the watersheds of Afghanistan. *The Cryosphere*, 12(5):1579–1594, 2018. doi: 10.5194/tc-12-1579-2018. URL <https://tc.copernicus.org/articles/12/1579/2018/>.
- [6] N. Ban, C. Caillaud, E. Coppola, E. Pichelli, S. Sobolowski, M. Adinolfi, B. Ahrens, A. Alias, I. Anders, S. Bastin, D. Belušić, S. Berthou, E. Brisson, R. M. Cardoso, S. C. Chan, O. B. Christensen, J. Fernández, L. Fita, T. Frisius, G. Gašparac, F. Giorgi, K. Goergen, J. E. Haugen, O. Hodnebrog, S. Kartsios, E. Katragkou, E. J. Kendon, K. Keuler, A. Lavin-Gullon, G. Lenderink, D. Leutwyler, T. Lorenz, D. Maraun, P. Mercogliano, J. Milovac, H.-J. Panitz, M. Raffa, A. R. Remedio, C. Schär, P. M. M. Soares, L. Srnec, B. M. Steensen, P. Stocchi, M. H. Tölle, H. Truhetz, J. Vergara-Temprado, H. de Vries, K. Warrach-Sagi, V. Wulfmeyer, and M. J. Zander. The first multi-model ensemble of regional climate simulations at kilometer-scale resolution, Part I: Evaluation of precipitation. *Climate Dynamics*, 57(1-2):275–302, Apr. 2021. ISSN 1432-0894. doi: 10.1007/s00382-021-05708-w. URL <http://dx.doi.org/10.1007/s00382-021-05708-w>.
- [7] T. Beucler, M. Pritchard, S. Rasp, J. Ott, P. Baldi, and P. Gentine. Enforcing analytic constraints in neural networks emulating physical systems. *Physical Review Letters*, 126:098302, Mar 2021. doi: 10.1103/PhysRevLett.126.098302. URL <https://link.aps.org/doi/10.1103/PhysRevLett.126.098302>.

- [8] A. Boone and P. Etchevers. An intercomparison of three snow schemes of varying complexity coupled to the same land surface model: Local-scale evaluation at an alpine site. *Journal of Hydrometeorology*, 2(4):374–394, 2001. ISSN 1525755X, 15257541. URL <http://www.jstor.org/stable/24909349>.
- [9] K. J. Bormann, S. Westra, J. P. Evans, and M. F. McCabe. Spatial and temporal variability in seasonal snow density. *Journal of Hydrology*, 484:63–73, 2013. ISSN 0022-1694. doi: <https://doi.org/10.1016/j.jhydrol.2013.01.032>. URL <https://www.sciencedirect.com/science/article/pii/S0022169413000784>.
- [10] O. Bruland, A. Faerevag, I. Steinsland, G. E. Liston, and K. Sand. Weather SDM: estimating snow density with high precision using snow depth and local climate. *Hydrology Research*, 46(4):494–506, 02 2015. ISSN 0029-1277. doi: [10.2166/nh.2015.059](https://doi.org/10.2166/nh.2015.059). URL <https://doi.org/10.2166/nh.2015.059>.
- [11] E. Brun, E. Martin, V. Simon, C. Gendre, and C. Coleou. An energy and mass model of snow cover suitable for operational avalanche forecasting. *Journal of Glaciology*, 35(121):333–342, 1989. doi: [10.3189/S0022143000009254](https://doi.org/10.3189/S0022143000009254). URL <https://doi.org/10.3189/S0022143000009254>.
- [12] E. Brun, V. Vionnet, A. Boone, B. Decharme, Y. Peings, R. Valette, F. Karbou, and S. Morin. Simulation of northern Eurasian local snow depth, mass, and density using a detailed snowpack model and meteorological reanalyses. *Journal of Hydrometeorology*, 14(1):203 – 219, 2013. doi: [10.1175/JHM-D-12-012.1](https://doi.org/10.1175/JHM-D-12-012.1). URL https://journals.ametsoc.org/view/journals/hydr/14/1/jhm-d-12-012_1.xml.
- [13] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud. Neural ordinary differential equations, 2018. URL <https://arxiv.org/abs/1806.07366>.
- [14] M. P. Clark, Y. Fan, D. M. Lawrence, J. C. Adam, D. Bolster, D. J. Gochis, R. P. Hooper, M. Kumar, L. R. Leung, D. S. Mackay, R. M. Maxwell, C. Shen, S. C. Swenson, and X. Zeng. Improving the representation of hydrologic processes in earth system models. *Water Resources Research*, 51(8):5929–5956, 2015. doi: <https://doi.org/10.1002/2015WR017096>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015WR017096>.
- [15] G. Cui, M. Anderson, and R. Bales. Mapping of snow water equivalent by a deep-learning model assimilating snow observations. *Journal of Hydrology*, 616:128835, 2023. ISSN 0022-1694. doi: <https://doi.org/10.1016/j.jhydrol.2022.128835>. URL <https://www.sciencedirect.com/science/article/pii/S0022169422014056>.
- [16] C. De Michele, F. Avanzi, A. Ghezzi, and C. Jommi. Investigating the dynamics of bulk snow density in dry and wet conditions using a one-dimensional model.

- The Cryosphere*, 7(2):433–444, 2013. doi: 10.5194/tc-7-433-2013. URL <https://tc.copernicus.org/articles/7/433/2013/>.
- [17] C. De Michele, F. Avanzi, A. Ghezzi, and C. Jommi. Investigating the dynamics of bulk snow density in dry and wet conditions using a one-dimensional model. *The Cryosphere*, 7(2):433–444, 2013. URL <https://doi.org/10.5194/tc-7-433-2013>.
- [18] G. T. Diro and L. Sushama. Snow–precipitation coupling and related atmospheric feedbacks over North America. *Atmospheric Science Letters*, 19(8): e831, 2018. doi: <https://doi.org/10.1002/asl.831>. URL <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/asl.831>.
- [19] S. Dong and N. Ni. A method for representing periodic functions and enforcing exactly periodic boundary conditions with deep neural networks. *Journal of Computational Physics*, 435:110242, 2021. ISSN 0021-9991. doi: <https://doi.org/10.1016/j.jcp.2021.110242>. URL <https://www.sciencedirect.com/science/article/pii/S0021999121001376>.
- [20] S. Duan, P. Ullrich, M. Risser, and A. Rhoades. Using temporal deep learning models to estimate daily snow water equivalent over the Rocky Mountains. *Water Resources Research*, 60(4):e2023WR035009, 2024. doi: <https://doi.org/10.1029/2023WR035009>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023WR035009>. e2023WR035009 2023WR035009.
- [21] E. Dutra, G. Balsamo, P. Viterbo, P. M. A. Miranda, A. Beljaars, C. Schär, and K. Elder. An improved snow scheme for the ECMWF land surface model: Description and offline validation. *Journal of Hydrometeorology*, 11(4):899 – 916, 2010. doi: 10.1175/2010JHM1249.1. URL https://journals.ametsoc.org/view/journals/hydr/11/4/2010jhm1249_1.xml.
- [22] P. P. Ebner, F. Koch, V. Premier, C. Marin, F. Hanzer, C. M. Carmagnola, H. François, D. Günther, F. Monti, O. Hargoa, U. Strasser, S. Morin, and M. Lehning. Evaluating a prediction system for snow management. *The Cryosphere*, 15(8):3949–3973, 2021. doi: 10.5194/tc-15-3949-2021. URL <https://tc.copernicus.org/articles/15/3949/2021/>.
- [23] R. Essery, A. Kontu, J. Lemmetyinen, M. Dumont, and C. B. Ménard. 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, 2016. doi: 10.5194/gi-5-219-2016. URL <https://gi.copernicus.org/articles/5/219/2016/>.
- [24] A. Fontrodona-Bach, B. Schaeffli, R. Woods, A. J. Teuling, and J. R. Larsen. NH-SWE: Northern Hemisphere snow water equivalent dataset based on in situ snow depth time series. *Earth System Science Data*, 15(6):2577–2599, 2023.

- doi: 10.5194/essd-15-2577-2023. URL <https://essd.copernicus.org/articles/15/2577/2023/>.
- [25] L. Gao, L. Zhang, Y. Shen, Y. Zhang, M. Ai, and W. Zhang. Modeling snow depth and snow water equivalent distribution and variation characteristics in the Irtysh River Basin, China. *Applied Sciences*, 11(18), 2021. ISSN 2076-3417. doi: 10.3390/app11188365. URL <https://www.mdpi.com/2076-3417/11/18/8365>.
- [26] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio. Maxout networks, 2013. URL <https://arxiv.org/abs/1302.4389>.
- [27] A. R. Hedrick, D. Marks, S. Havens, M. Robertson, M. Johnson, M. Sandusky, H.-P. Marshall, P. R. Kormos, K. J. Bormann, and T. H. Painter. Direct insertion of NASA airborne snow observatory-derived snow depth time series into the iSnobal energy balance snow model. *Water Resources Research*, 54(10):8045–8063, 2018. doi: <https://doi.org/10.1029/2018WR023190>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR023190>.
- [28] D. F. Hill, E. A. Burakowski, R. L. Crumley, J. Keon, J. M. Hu, A. A. Arendt, K. Wikstrom Jones, and G. J. Wolken. Converting snow depth to snow water equivalent using climatological variables. *The Cryosphere*, 13(7):1767–1784, 2019. doi: 10.5194/tc-13-1767-2019. URL <https://tc.copernicus.org/articles/13/1767/2019/>.
- [29] G. Hinton, N. Srivastava, and K. Swersky. Lecture 6e: Rmsprop: Divide the gradient by a running average of its recent magnitude, February 2014. URL https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf.
- [30] M. Innes. Flux: Elegant machine learning with Julia. *Journal of Open Source Software*, 2018. doi: 10.21105/joss.00602. URL <https://joss.theoj.org/papers/10.21105/joss.00602>.
- [31] M. Innes, E. Saba, K. Fischer, D. Gandhi, M. C. Rudilosso, N. M. Joy, T. Karmali, A. Pal, and V. Shah. Fashionable modelling with Flux. *arXiv Computing Research Repository (CoRR)*, abs/1811.01457, 2018. URL <https://arxiv.org/abs/1811.01457>.
- [32] K. S. Jennings, T. S. Winchell, B. Livneh, and N. P. Molotch. Spatial variation of the rain-snow temperature threshold across the Northern Hemisphere. *Nature Communications*, 9(1):1148, Mar 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-03629-7. URL <https://doi.org/10.1038/s41467-018-03629-7>.
- [33] C. M. Jiang, K. Kashinath, Prabhat, and P. Marcus. Enforcing physical constraints in CNNs through differentiable PDE layer. In *ICLR 2020 Workshop*

- on *Integration of Deep Neural Models and Differential Equations*, 2019. URL <https://openreview.net/forum?id=q2noHUqMkK>.
- [34] S. B. Kapnick, X. Yang, G. A. Vecchi, T. L. Delworth, R. Gudgel, S. Malyshchev, P. C. D. Milly, E. Shevliakova, S. Underwood, and S. A. Margulis. Potential for western US seasonal snowpack prediction. *Proceedings of the National Academy of Sciences*, 115(6):1180–1185, 2018. doi: 10.1073/pnas.1716760115. URL <https://www.pnas.org/doi/abs/10.1073/pnas.1716760115>.
- [35] K. Kojima. Densification of seasonal snow cover. *Physics of Snow and Ice: Proceedings*, 1(2):929–952, 1967. URL <http://hdl.handle.net/2115/20351>.
- [36] K. Kouki, P. Räisänen, K. Luojus, A. Luomaranta, and A. Riihelä. Evaluation of Northern Hemisphere snow water equivalent in CMIP6 models during 1982–2014. *The Cryosphere*, 16(3):1007–1030, 2022. doi: 10.5194/tc-16-1007-2022. URL <https://tc.copernicus.org/articles/16/1007/2022/>.
- [37] P. Krajčič, R. Kirnbauer, J. Parajka, J. Schöber, and G. Blöschl. The Kühltai data set: 25 years of lysimetric, snow pillow, and meteorological measurements. *Water Resources Research*, 53(6):5158–5165, 2017. doi: <https://doi.org/10.1002/2017WR020445>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017WR020445>.
- [38] D. M. Lawrence, R. A. Fisher, C. D. Koven, K. W. Oleson, S. C. Swenson, G. Bonan, N. Collier, B. Ghimire, L. van Kampenhout, D. Kennedy, E. Kluzek, P. J. Lawrence, F. Li, H. Li, D. Lombardozzi, W. J. Riley, W. J. Sacks, M. Shi, M. Vertenstein, W. R. Wieder, C. Xu, A. A. Ali, A. M. Badger, G. Bisht, M. van den Broeke, M. A. Brunke, S. P. Burns, J. Buzan, M. Clark, A. Craig, K. Dahlin, B. Drewniak, J. B. Fisher, M. Flanner, A. M. Fox, P. Gentine, F. Hoffman, G. Keppel-Aleks, R. Knox, S. Kumar, J. Lenaerts, L. R. Leung, W. H. Lipscomb, Y. Lu, A. Pandey, J. D. Pelletier, J. Perket, J. T. Rander-son, D. M. Ricciuto, B. M. Sanderson, A. Slater, Z. M. Subin, J. Tang, R. Q. Thomas, M. Val Martin, and X. Zeng. The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, 11(12):4245–4287, 2019. doi: <https://doi.org/10.1029/2018MS001583>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001583>.
- [39] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller. *Efficient BackProp*, pages 9–48. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. ISBN 978-3-642-35289-8. doi: 10.1007/978-3-642-35289-8_3. URL https://doi.org/10.1007/978-3-642-35289-8_3.
- [40] M. Lehning, P. Bartelt, B. Brown, C. Fierz, and P. Satyawali. A physical SNOWPACK model for the swiss avalanche warning: Part II. Snow

- microstructure. *Cold Regions Science and Technology*, 35(3):147–167, 2002. ISSN 0165-232X. doi: [https://doi.org/10.1016/S0165-232X\(02\)00073-3](https://doi.org/10.1016/S0165-232X(02)00073-3). URL <https://www.sciencedirect.com/science/article/pii/S0165232X02000733>.
- [41] Y. Lejeune, M. Dumont, J.-M. Panel, M. Lafaysse, P. Lapalus, E. Le Gac, B. Lesaffre, and S. Morin. 57 years (1960–2017) of snow and meteorological observations from a mid-altitude mountain site (Col de Porte, France, 1325 m of altitude). *Earth System Science Data*, 11(1):71–88, 2019. doi: 10.5194/essd-11-71-2019. URL <https://essd.copernicus.org/articles/11/71/2019/>.
- [42] J. Liu, R. Chen, S. Ma, C. Han, Y. Ding, S. Guo, and X. Wang. The challenge of monitoring snow surface sublimation in winter could be resolved with structure-from-motion photogrammetry. *Journal of Hydrology*, 630: 130733, 2024. ISSN 0022-1694. doi: <https://doi.org/10.1016/j.jhydrol.2024.130733>. URL <https://www.sciencedirect.com/science/article/pii/S0022169424001276>.
- [43] B. Livneh, J. S. Deems, D. Schneider, J. J. Barsugli, and N. P. Molotch. Filling in the gaps: Inferring spatially distributed precipitation from gauge observations over complex terrain. *Water Resources Research*, 50(11):8589–8610, 2014. doi: <https://doi.org/10.1002/2014WR015442>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014WR015442>.
- [44] H. Luijting, D. Vikhamar-Schuler, T. Aspelien, Å. Bakketun, and M. Homleid. Forcing the SURFEX/Crocus snow model with combined hourly meteorological forecasts and gridded observations in southern Norway. *The Cryosphere*, 12(6):2123–2145, 2018. doi: 10.5194/tc-12-2123-2018. URL <https://tc.copernicus.org/articles/12/2123/2018/>.
- [45] C. C. Lundy, R. L. Brown, E. E. Adams, K. W. Birkeland, and M. Lehnig. A statistical validation of the snowpack model in a Montana climate. *Cold Regions Science and Technology*, 33(2):237–246, 2001. ISSN 0165-232X. doi: [https://doi.org/10.1016/S0165-232X\(01\)00038-6](https://doi.org/10.1016/S0165-232X(01)00038-6). URL <https://www.sciencedirect.com/science/article/pii/S0165232X01000386>. ISSW 2000: International Snow Science Workshop.
- [46] D. Marks, S. Havens, M. Johnson, and M. Sandusky. Pysnobar, v2.4.1, 2018. URL https://github.com/USDA-ARS-NWRC/pysnobar/blob/master/pysnobar/c_snobar/libsnoal/_time_compact.c.
- [47] J. Meloche, A. Langlois, N. Rutter, D. McLennan, A. Royer, P. Billecocq, and S. Ponomarenko. High-resolution snow depth prediction using random forest algorithm with topographic parameters: A case study in the Greiner watershed, Nunavut. *Hydrological Processes*, 36(3):e14546, 2022. doi: <https://doi.org/10.1002/hyp.14546>.

//doi.org/10.1002/hyp.14546. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.14546>.

- [48] C. B. Menard, R. Essery, G. Krinner, G. Arduini, P. Bartlett, A. Boone, C. Brutel-Vuilmet, E. Burke, M. Cuntz, Y. Dai, B. Decharme, E. Dutra, X. Fang, C. Fierz, Y. Gusev, S. Hagemann, V. Haverd, H. Kim, M. Lafaysse, T. Marke, O. Nasonova, T. Nitta, M. Niwano, J. Pomeroy, G. Schädler, V. A. Semenov, T. Smirnova, U. Strasser, S. Swenson, D. Turkov, N. Wever, and H. Yuan. Scientific and human errors in a snow model intercomparison. *Bulletin of the American Meteorological Society*, 102(1):E61 – E79, 2021. doi: 10.1175/BAMS-D-19-0329.1. URL <https://journals.ametsoc.org/view/journals/bams/102/1/BAMS-D-19-0329.1.xml>.
- [49] J. Meyer, J. Horel, P. Kormos, A. Hedrick, E. Trujillo, and S. M. Skiles. Operational water forecast ability of the HRRR-iSnobal combination: An evaluation to adapt into production environments. *Geoscientific Model Development*, 16(1):233–250, 2023. doi: 10.5194/gmd-16-233-2023. URL <https://gmd.copernicus.org/articles/16/233/2023/>.
- [50] J. D. D. Meyer, J. Jin, and S.-Y. Wang. 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, 2012. doi: 10.1175/JHM-D-12-066.1. URL https://journals.ametsoc.org/view/journals/hydr/13/6/jhm-d-12-066_1.xml.
- [51] C. Molnar. *Interpretable Machine Learning*. Christoph Molnar, 2 edition, 2022. URL <https://christophm.github.io/interpretable-ml-book>.
- [52] J. Nash and J. Sutcliffe. River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3):282–290, 1970. ISSN 0022-1694. doi: [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6). URL <https://www.sciencedirect.com/science/article/pii/0022169470902556>.
- [53] G.-Y. Niu, Z.-L. Yang, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, A. Kumar, K. Manning, D. Niyogi, E. Rosero, M. Tewari, and Y. Xia. The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research: Atmospheres*, 116(D12), 2011. doi: <https://doi.org/10.1029/2010JD015139>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JD015139>.
- [54] M. Pharr and R. Fernando. *GPU gems 2: Programming techniques for high-performance graphics and general-purpose computation*. Addison-Wesley Professional, 2005. URL <https://developer.nvidia.com/gpugems/gpugems2>.

- [55] I. Rahimi, A. H. Gandomi, F. Chen, and E. Mezura-Montes. 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, Dec. 2022. ISSN 1886-1784. doi: 10.1007/s11831-022-09859-9. URL <http://dx.doi.org/10.1007/s11831-022-09859-9>.
- [56] M. S. Raleigh, B. Livneh, K. Lapo, and J. D. Lundquist. How does availability of meteorological forcing data impact physically based snowpack simulations? *Journal of Hydrometeorology*, 17(1):99 – 120, 2016. doi: 10.1175/JHM-D-14-0235.1. URL https://journals.ametsoc.org/view/journals/hydr/17/1/jhm-d-14-0235_1.xml.
- [57] R. Rasmussen, B. Baker, J. Kochendorfer, T. Meyers, S. Landolt, A. P. Fischer, J. Black, J. M. Thériault, P. Kucera, D. Gochis, C. Smith, R. Nitu, M. Hall, K. Ikeda, and E. Gutmann. How well are we measuring snow: The NOAA/FAA/NCAR winter precipitation test bed. *Bulletin of the American Meteorological Society*, 93(6):811 – 829, 2012. doi: 10.1175/BAMS-D-11-00052.1. URL <https://journals.ametsoc.org/view/journals/bams/93/6/bams-d-11-00052.1.xml>.
- [58] M. L. Reba, D. Marks, M. Seyfried, A. Winstral, M. Kumar, and G. Flerchinger. A long-term data set for hydrologic modeling in a snow-dominated mountain catchment. *Water Resources Research*, 47(7), 2011. doi: <https://doi.org/10.1029/2010WR010030>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010WR010030>.
- [59] C. Schär, O. Fuhrer, A. Arteaga, N. Ban, C. Charpillot, S. D. Girolamo, L. Hentgen, T. Hoefler, X. Lapillonne, D. Leutwyler, K. Osterried, D. Panosetti, S. Rüdisühli, L. Schlemmer, T. C. Schulthess, M. Sprenger, S. Ubbiali, and H. Wernli. Kilometer-scale climate models: Prospects and challenges. *Bulletin of the American Meteorological Society*, 101(5):E567 – E587, 2020. doi: 10.1175/BAMS-D-18-0167.1. URL <https://journals.ametsoc.org/view/journals/bams/101/5/bams-d-18-0167.1.xml>.
- [60] M. C. Serreze, M. P. Clark, R. L. Armstrong, D. A. McGinnis, and R. S. Pulwarty. Characteristics of the Western United States snowpack from snowpack telemetry (SNOTEL) data. *Water Resources Research*, 35(7):2145–2160, 1999. doi: <https://doi.org/10.1029/1999WR900090>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/1999WR900090>.
- [61] J. Shea, P. Wagnon, W. Immerzeel, R. Biron, F. Brun, and F. Pellicciotti. A comparative high-altitude meteorological analysis from three catchments in the Nepalese Himalaya. *International Journal of Water Resources Development*, 31(2):174–200, 2015. doi: 10.1080/07900627.2015.1020417. URL <https://doi.org/10.1080/07900627.2015.1020417>.

- [62] E. J. Smyth, M. S. Raleigh, and E. E. Small. Improving SWE estimation with data assimilation: The influence of snow depth observation timing and uncertainty. *Water Resources Research*, 56(5):e2019WR026853, 2020. doi: <https://doi.org/10.1029/2019WR026853>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019WR026853>. e2019WR026853 2019WR026853.
- [63] Y. Song, W.-P. Tsai, J. Gluck, A. Rhoades, C. Zarzycki, R. McCrary, K. Lawson, and C. Shen. LSTM-based data integration to improve snow water equivalent prediction and diagnose error sources. *Journal of Hydrometeorology*, 25(1):223 – 237, 2024. doi: 10.1175/JHM-D-22-0220.1. URL <https://journals.ametsoc.org/view/journals/hydr/25/1/JHM-D-22-0220.1.xml>.
- [64] K. Spehlmann, E. Euskirchen, and S. Stuefer. Sublimation measurements of tundra and taiga snowpack in Alaska. *The Cryosphere Discussions*, 2023: 1–18, 2023. doi: 10.5194/tc-2023-153. URL <https://tc.copernicus.org/preprints/tc-2023-153/>.
- [65] H. Steele, E. E. Small, and M. S. Raleigh. Demonstrating a hybrid machine learning approach for snow characteristic estimation throughout the Western United States. *Water Resources Research*, 60(6):e2023WR035805, 2024. doi: <https://doi.org/10.1029/2023WR035805>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023WR035805>. e2023WR035805 2023WR035805.
- [66] E. E. Stigter, J. F. Steiner, I. Koch, T. M. Saloranta, J. D. Kirkham, and W. W. Immerzeel. Energy and mass balance dynamics of the seasonal snowpack at two high-altitude sites in the Himalaya. *Cold Regions Science and Technology*, 183:103233, 2021. ISSN 0165-232X. doi: <https://doi.org/10.1016/j.coldregions.2021.103233>. URL <https://www.sciencedirect.com/science/article/pii/S0165232X21000148>.
- [67] S. Tanniru and R. Ramsankaran. Machine learning-based estimation of high-resolution snow depth in Alaska using passive microwave remote sensing data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:6007–6025, 2023. doi: 10.1109/JSTARS.2023.3287410. URL <https://ieeexplore.ieee.org/document/10154626>.
- [68] USDA. *Title 210-National Engineering Handbook, Part 622-Snow Survey and Water Supply Forecasting*, chapter Chapter 2: Data Parameters, pages 2–2 – 2–12. USDA, Washington, D.C., 2010. URL <https://directives.nrcs.usda.gov/sites/default/files2/1720456630/Chapter%202%20-%20Data%20Parameters.pdf>.
- [69] M. Vafakhah, A. Nasiri Khiavi, S. Janizadeh, and H. Ganjkanlo. Evaluating different machine learning algorithms for snow water equivalent prediction.

- Earth Science Informatics*, 15(4):2431–2445, Aug. 2022. ISSN 1865-0481. doi: 10.1007/s12145-022-00846-z. URL <http://dx.doi.org/10.1007/s12145-022-00846-z>.
- [70] L. Viallon-Galinier, P. Hagenmuller, and M. Lafaysse. Forcing and evaluating detailed snow cover models with stratigraphy observations. *Cold Regions Science and Technology*, 180:103163, 2020. ISSN 0165-232X. doi: <https://doi.org/10.1016/j.coldregions.2020.103163>. URL <https://www.sciencedirect.com/science/article/pii/S0165232X20304109>.
- [71] V. Vionnet, E. Brun, S. Morin, A. Boone, S. Faroux, P. Le Moigne, E. Martin, and J.-M. Willemet. The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2. *Geoscientific Model Development*, 5(3):773–791, 2012. doi: 10.5194/gmd-5-773-2012. URL <https://gmd.copernicus.org/articles/5/773/2012/>.
- [72] V. Vionnet, D. Six, L. Auger, M. Dumont, M. Lafaysse, L. Quéno, M. Réveillet, I. Dombrowski-Etchevers, E. Thibert, and C. Vincent. Sub-kilometer precipitation datasets for snowpack and glacier modeling in alpine terrain. *Frontiers in Earth Science*, 7, 2019. ISSN 2296-6463. doi: 10.3389/feart.2019.00182. URL <https://www.frontiersin.org/articles/10.3389/feart.2019.00182>.
- [73] Y.-H. Wang, H. V. Gupta, X. Zeng, and G.-Y. Niu. Exploring the potential of long short-term memory networks for improving understanding of continental- and regional-scale snowpack dynamics. *Water Resources Research*, 58(3):e2021WR031033, 2022. doi: <https://doi.org/10.1029/2021WR031033>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021WR031033>. e2021WR031033 2021WR031033.
- [74] M. Warscher, T. Marke, E. Rottler, and U. Strasser. Operational and experimental snow observation systems in the upper Rofental: Data from 2017 to 2023. *Earth System Science Data Discussions*, 2024:1–26, 2024. doi: 10.5194/essd-2024-45. URL <https://essd.copernicus.org/preprints/essd-2024-45/>.
- [75] N. Water and C. Center. Telemetry and data transmission, 2023. URL <https://www.nrcs.usda.gov/wps/portal/wcc/home/aboutUs/monitoringPrograms/telemetry/>.
- [76] N. Wever, L. Schmid, A. Heilig, O. Eisen, C. Fierz, and M. Lehning. Verification of the multi-layer SNOWPACK model with different water transport schemes. *The Cryosphere*, 9(6):2271–2293, 2015. doi: 10.5194/tc-9-2271-2015. URL <https://tc.copernicus.org/articles/9/2271/2015/>.
- [77] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945. ISSN 00994987. URL <http://www.jstor.org/stable/3001968>.

- [78] L. Xu and P. Dirmeyer. Snow-atmosphere coupling strength in a global atmospheric model. *Geophysical Research Letters*, 38(13), 2011. doi: <https://doi.org/10.1029/2011GL048049>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011GL048049>.
- [79] H. Yan, N. Sun, M. Wigmosta, R. Skaggs, Z. Hou, and R. Leung. Next-generation intensity-duration-frequency curves for hydrologic design in snow-dominated environments. *Water Resources Research*, 54(2):1093–1108, 2018. doi: <https://doi.org/10.1002/2017WR021290>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017WR021290>.
- [80] H. Yan, N. Sun, M. Wigmosta, R. Skaggs, L. R. Leung, A. Coleman, and Z. Hou. Observed spatiotemporal changes in the mechanisms of extreme water available for runoff in the Western United States. *Geophysical Research Letters*, 46(2):767–775, 2019. doi: <https://doi.org/10.1029/2018GL080260>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL080260>.
- [81] J. Yang, L. Jiang, K. Luo, J. Pan, J. Lemmetyinen, M. Takala, and S. Wu. Snow depth estimation and historical data reconstruction over China based on a random forest machine learning approach. *The Cryosphere*, 14(6):1763–1778, 2020. doi: 10.5194/tc-14-1763-2020. URL <https://tc.copernicus.org/articles/14/1763/2020/>.
- [82] J. Yang, L. Jiang, J. Pan, J. Shi, S. Wu, J. Wang, and F. Pan. Comparison of machine learning-based snow depth estimates and development of a new operational retrieval algorithm over China. *Remote Sensing*, 14(12), 2022. ISSN 2072-4292. doi: 10.3390/rs14122800. URL <https://www.mdpi.com/2072-4292/14/12/2800>.

*Chapter 3***A NEXT-GENERATION SNOW ALBEDO
PARAMETERIZATION FOR CLIMATE MODELING USING
CONSTRAINED MACHINE LEARNING**

At the time of publication for this thesis, this chapter is under a second review for *Journal of Geophysical Research: Machine Learning and Computation*, from the American Geophysical Union. No digital article identifier or formal citation exists for this work at this time. The format and content of this chapter may exhibit slight differences from the version found in its affiliated version published in a journal.

This article has associated Supplementary Information (referred to within the chapter as “[Supporting Information](#).” This material will be hyperlinked where referenced, though this material can be found at <https://doi.org/10.22002/cd37q-bd803>.

References within this work to “previous work” the developed “framework,” or citation 19 in this chapter all refer to the study detailed in Chapter 2.

ABSTRACT

We demonstrate a data-driven parameterization for snow albedo using a constrained neural differential equation framework that directly predicts a range of snow albedo tendencies from standard snow and 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.

3.1 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 [33, 55, 72]. They play a significant role in terrestrial energy and water fluxes, with consequences for ecosystem and economic functions such as water supply, agriculture, supply chains, tourism, and environmental hazards [57, 63, 78]. However, snowpacks are sensitive to changing climates, and historical norms of accumulation and melt no longer reliably capture present-day patterns and timing [3, 112]. Snow dynamics remain one of the most challenging aspects to represent in simulations of the water cycle [105], 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 [9]:

$$\alpha = \frac{\int_{\lambda_1}^{\lambda_2} \alpha(\lambda) S(\lambda) d\lambda}{\int_{\lambda_1}^{\lambda_2} S(\lambda) d\lambda} \quad (3.1)$$

where $\alpha(\lambda)$ is the snow's spectrally-resolved albedo, $S(\lambda)$ is the solar spectral irra-

diance at the snow surface, and $\lambda_1 = 250$ nm and $\lambda_2 = 4000$ 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 [9] and leading to error propagation within global climate models (GCMs) [106, 128]. 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 [24, 32, 55, 74, 87]. This influence propagates to global circulation, affecting processes such as ENSO, monsoons, and poleward retreat [133], underscoring the need for accurate representation.

Reducing errors in snow albedo is difficult given its variability and multiscale drivers. Snow albedo is governed by a suite of physical processes, including the shape, size, scattering properties, and evolution of snow grains, the concentrations of light-absorbing particles (LAPs) or impurities such as dust and black carbon, snowpack layering and density, atmospheric state, precipitation history, and solar zenith angle (see [134, 142]). However, snow albedo also varies strongly across short spatial and temporal scales due to topography and microclimate [93, 137], and modification by LAPs often reflects both local and remote sources, decoupling changes from solely local snowpack processes [9]. Representing the complex interplay of these processes across scales within physical models demands input availability and process fidelity beyond what is contemporarily available at climate scales, leaving even complex schemes uncertain [25, 49]. Simpler empirical or prescribed formulations to approximate mean behaviors can be useful in mapping local geographic inputs to expected long-term trends, but the assumptions for such schemes rarely generalize beyond their calibration regime. With expanding multi-scale observations spanning diverse climates, a new alternative has become possible: to infer relationships directly from available data, within physical guardrails but without presupposing their functional form. Such a framework to represent physically consistent mappings from local snowpack, climate, and environmental properties to distributions of albedo behavior must capture the integrated effects of the snow and atmosphere while preserving generalizability to diverse climates and changing environments.

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 [35], with later extensions [138] 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 [2] 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 are mixed, with complexity showing limited influence on overall parameterization skill and no consensus on which approach best suits large-scale modeling [39, 49, 76, 86, 90]. Empirical age-based schemes reproduce mean decay rates but struggle to capture their variability under changing conditions, routinely producing albedo errors of 10–20% and comparable biases outside their calibration domains [1, 9, 25, 41]. This spurred development of detailed process models such as SNICAR, TARTES, and SNOWPACK [44, 79] to resolve radiative transfer, microstructure, and spectral effects with higher accuracy, though their usage is generally associated with a larger computational cost at regional scales or modern GCM grids (~ 10 – 100 km) [18, 25, 39, 47, 49, 59, 132]. These and intermediate-complexity models are also sensitive to input uncertainties, which are poorly constrained at GCM resolutions [1, 9, 47, 57, 90, 131] and often require site-specific calibration. As a result, they show limited transferability across climates and error ranges at large scale that frequently converge back to those of simpler schemes, or can even be outperformed by the latter under uncertain inputs [41, 77, 82, 97, 119, 122].

Therefore, many GCMs continue to rely on combinations of a few simple and older parameterizations for scalability, with wide performance spreads highly sensitive to model choice, calibration sites, and validation approaches, especially during the ablative period [39, 41, 54, 86]. Age-based schemes remain common despite evidence that more flexible alternatives are superior [9]. Snow albedo modeling choices have a larger impact on simulations than those for compaction or other processes, with most impact arising from their aptitude in representing varied albedo decay rates and their resulting influence on snowpack decay, which can vary the snow season length by 25–50+ days [1, 39, 41, 66, 78, 101, 117, 133]. Meanwhile, increasing GCM resolutions and efforts toward finer (~ 1 – 10 km) or even sub-km land scales to resolve snow albedo variability [3, 31, 47, 131, 141] will further exacerbate these issues of sensitivity, computational burden, and local errors, highlighting the need for efficient but accurate finer-scale schemes for next-generation GCMs;

capable of capturing diverse decay rates [12, 22, 120, 141].

Recent computational advances and data volume have inspired new albedo parameterization efforts emphasizing grain size, LAP concentration, and spectral resolution to replicate the success of models such as SNICAR/SNICAR-ADv3 with greater efficiency. Studies such as [9], [49], [118], and [127] report accurate performance across snow conditions, yet these schemes still require complex inputs and inherit the input sensitivity of their predecessors, limiting their utility in GCMs where necessary inputs are often unavailable [47, 49, 54, 57, 141]. Machine learning (ML) alternatives offer potential as they easily represent nonlinear interactions and are more robust to uncertainty and nonstationarity [25, 29, 46], yet 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 [20, 31, 143]. These efforts presently find limited transferability beyond calibration sites, with many errors still plateauing near $\sim 10\%$ [29, 31, 70], and many ML-based approaches do not inherently respect physical constraints [26, 48].

This work presents a novel parameterization, which uses the machine learning framework developed in 2022 and described in Charbonneau et al. [19] to predict the time derivative of the snow albedo given environmental features and snow state as input. The predictive model output, which obeys prescribed constraints, is then integrated 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 [114] 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 one minute 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.

3.2 Methodology

We utilize the framework introduced in Charbonneau et al. [19] 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.

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:

$$\frac{d\alpha}{dt} = M(\alpha, T_a, P, \mu) \quad (3.2)$$

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 Fig. 3.1, and constraint functions are picked only to limit the subsequently updated snow albedo value between an applicable range of 0.3 and 1 [60]. The constraints are enacted by fixed layers that remain during both training and inference, guaranteeing physically admissible outputs by construction (prohibiting unphysical 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 [107], 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 afterwards [19]. Unlike our previous work and other 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 the albedo absent new precipitation. Choosing initial boundaries solely to maintain

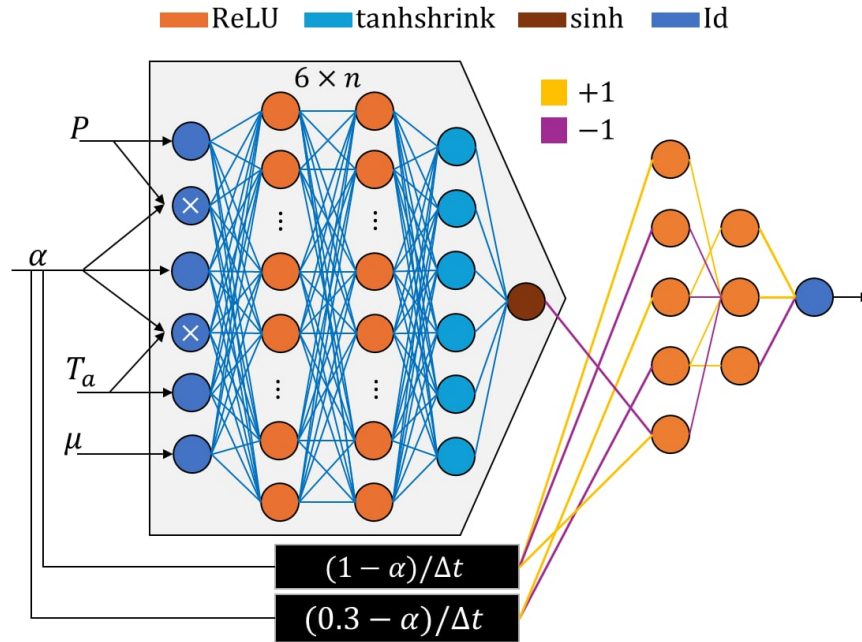


Figure 3.1: **Structure of M .** The predictive portion of the parameterization consists of a simple 3-layer dense neural network with an aggregation to a final output. Nodes are colored by their aggregation function, trainable weights (and a bias term) are shown in blue, and fixed weights (no bias term) are shown in yellow (+1) or purple (-1). One hyperparameter n determines the layer widths. The white “ $\times$ ” indicates multiplication to allow usage of the cross-term inputs αT_a and αP . A threshold constraint layer (see Appendix A in [19]) is included around the predictive component to keep the accumulated albedo between 0.3 and 1 for time-step Δt , even during training.

expected global limits did not presuppose any assumptions about albedo dynamics, and additionally allowed us to explore the extent of the model to learn physically consistent relationships absent additional structure. This prognostic form mirrors the differential equations used for other GCM variables, aiding integration, and reflects findings that prognostic parameterizations are more consistent than other forms [39]. 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 [1, 41] and improving representative capacity to observed variability. We also note that without an inclusion of snow timing, we break from conventional reliance on snow age, preferring to focus on more direct state-related snow features.

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 [51, 53], 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 [19], 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 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 [39, 86]. Like snowfall timing or LAP-affiliated features where few training samples from different climates are available¹, grain size is deliberately excluded — though often improving diagnostic accuracy, using grain size increases model sensitivity and couples performance to an uncertain or unavailable input at GCM scales, it is less directly linked to albedo derivatives than variables like temperature or precipitation [1, 41], and measurements transfer poorly across regions, risking bias and reduced generalizability. Using a 1D model over well-characterized local state variables (i.e., not directly ingesting/requiring spatiotemporal information) allows broad applicability

¹While LAP-affiliated features would be useful for capturing explicit representation of darkening processes, their relative data sparsity would leave a model more sensitive to otherwise uncertain inputs at GCM scales, and severely limit training data availability. Leaving LAP-related variables out mitigates these issues and permits more higher-quality training data from a wide variety of locations, which is critical for learning robust relationships that generalize to different climate types. As observational efforts expand and these effects become better sampled across different climate types, adapting the network to receive such inputs remains straightforward.

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 [19].

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, snow albedo timeseries 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 [3, 57, 111, 112, 119, 125], with typical validation errors of 0.05–0.08 [9, 33, 58, 100, 101, 103, 137]. Usage of SMM has yielded 10–25% gains in hydrological modeling skill, especially during melt season, underscoring its suitability for next-generation models [46]. 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 [80], Kühtai [75], and Mammoth Mountain [8], and include additional data from Sodankylä [40], the Upper Rofental [135], the Himalayas [123, 124], 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 (500m) [9, 10, 29, 112]. SMM’s underlying calibration data is not inherently representative of all climates [69] and can influence poor distinction between snow types [57], which can bias snow cover fractions and subsequently derived albedos ([78]; 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 [31, 71, 103]. For SNOTEL and other site sensors, precipitation is often underreported, and sensor records can contain biased or unphysical values [61, 91, 135]. 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 timeseries, supplemented with established methods for other variables [7, 85, 121, 140]. From cleaned records, $d\alpha/dt$ was derived for days with complete data (excluding $\Delta t > 1$), creating a table of input feature values and an associated target $d\alpha/dt$ that could be shuffled for training. Daily zenith angle cosine was computed with `Insolations.jl` [23], and snow-rain fractions were derived using the parameterization from Jennings et al. [65], shown to be over 85% accurate. Data scaling constants (standard deviation for inputs, absolute maximum for targets) were fixed in the neural model for reproducibility. This yielded roughly 1.6 million site-days with SMM albedo and 15,300 with direct observations.

Training and Testing

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 timeseries from ground sensors to assess encapsulation 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, 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 Fig. 3 from Charbonneau et al. [19], as well as Fig. S1 in the [Supporting Information](#) 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 (~ 87000 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 [64] using the root-mean-square propagation (RMSProp) optimizer [62]. Data sources and variables are summarized in Table 3.1.

Table 3.1: Summary of data and variables. 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, 1000 kg m^3 . Only variables listed in Eq. 3.2 are used within M . The spread of snowpack types and geographic locations of all sites can be found in Fig. S1 of the [Supporting Information](#).

Data					Variables			
Source	Sites	Days	Scale	Purpose	Symbol	Units	Name	Description
SNOTEL/SMM Albedo	40	86841	500m	Training	α	None	Snow Albedo	Derived, see Appendix A
SNOTEL/SMM Albedo	700	1527868	500m	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
Sodankyla, 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 ($0, T_a$)

Magnusson et al. [86] 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 3.2), representative of those used in current GCMs [78] or proposed in prior studies. Consistent with Magnusson et al. [86] 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 [39], 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](#) 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, with scheme output equal to the snow albedo at the next time-step, rather than an albedo tendency or rate relative to the present albedo. This reflects historical convention, a need for computational simplicity, and represents a leading cause for the lack of a variety of growth/decay rates in present albedo schemes, which we intend to circumvent with our continuous prognostic scheme. The distinction between these modeling forms leads to three possible equations for generating time-stepped outputs, diagnostic (Eq. 3.3), prognostic-value (Eq. 3.4), and prognostic-rate (Eq. 3.5, our neural parameterization):

$$\widehat{\alpha}_i = f(\theta_i) \quad (3.3)$$

$$\widehat{\alpha}_i = f(\widehat{\alpha}_{i-1}, \theta_i) \quad (3.4)$$

$$\widehat{\alpha_{i+K}} = \alpha_i + (K\Delta t) M(\widehat{\alpha}_i, P_i, T_{a,i}, \mu_i) \quad (3.5)$$

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. 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 [19], a maximum gap size of $K_{\max} = 5$ days was permitted before timeseries were “reset” to observed² 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 [138], though age/state dynamics after such initialization were dictated only by the scheme.

²Input data was only defined as non-missing where observational values for albedo also existed, instead of keeping data where inputs existed but not observed albedo. This mitigated the issue of needing to reset the scheme to an observed albedo value when none existed.

Table 3.2: Parameterizations compared in this study. We benchmark the proposed parameterization against an exhaustive (and fundamental, as others are similar to those below) 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 the [Supporting Information](#). Models explicitly found in CMIP6 [42] 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. Some schemes are more similar than others, though the entire range exhibits a wide spread of possible functional forms, expressive capacity, and utilized input features.

Name	Symbol	Identifiable Origins	Referenced Paper	Variables	Type	P Reset	AR	Has Been In:
ACE Curve, Polynomial	AC	1956	[93]	τ	D	T	N	
Exponential Decay (Snow Age)	ED	1969	[73]	τ	D	T	N	US ACE HEC-HMS
Exponential Decay (T_a Index)	ET	1991	[109]	P, T_a	D	T	Y	
Logistic/Exponential Decay (Ta Index)	LE	2024	[104]	P, T_a, z	D	T	Y	-
VAS	VA	2014	[87]	P, T_a, z, α_g	D	T	Y	-
Brock 2000	BR	2000	[14]	P, T_a, SWE, α_g	D	T	Y	-
Marshall (“Simple” Depth)	MD	1989	1989	τ, SWE, α_g	D	T	N	-
Oerlemans & Knap	OK	1998	[83]	τ, z, α_g	D	T	N	Noah
ESCIMO	EC	1991	[126]	τ, T_a	D	T	N	ESCIMO
Interpolation (T_a)*	IT	1992	[94]	T_a	D	-	N	ECHAM, ECMWF, JSBACH3
Interpolation (Density)	ID	1994	[52]	ρ	D	-	N	SOMARS, RACMO2, CLSM
Noah Exponential	LV	1968	[84]	τ, DOY	D	T	N	Noah, DHSVM, VIC
SNOWPACK-like	SP	2002	This	P, T_a, μ, z, ρ	D	-	N	SNOWPACK
Zhongyuan & Ohata Density	ZO	1989	[142]	ρ	D	-	N	-
NASA GISS*	NG	1983	[56]	P	D	C	Y	GISS Model II, ORCHIDEE, CLSM
CLASS*	CL	1991	[81]	P, α	PV	C	Y	Noah-MP, ISBA, CLASS
UTOPIA	UT	1991, 1990	[17]	P, α, T_s, μ	PV	C	Y	UTOPIA
HTESSEL*	HT	1991, 1990	[36]	α, P, SWE, T_s	PV	C	Y	HTESSEL/ECMWF
Giddings and La Chapelle	GC	1961	[99]	z, α_g	D	-	N	-
JULES*	JL	2001	[13]	P, T_s	D	C	Y	JULES
Marks 1988	MS	1988	[88]	τ, μ	D	T	N	SNOBAL
BATS*	BA	1986	[1]	P, T_s, μ	D	C	Y	Many (Noah-MP, WRF CoLM...)
ISBA*	IS	1992	[27]	z, T_s, T_a, P	D	C	Y	Crocus, ISBA MAR
Neural Network	NN	Prev. Work, 2025	This	α, P, T_a, μ	PR	-	N	-

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 four 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 Eq. (3.3-3.5), 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 $\times$) 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 “AR” in Table 3.2.

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 timeseries. Statistical significance of RMSE and RMSE% differences across validation and test sites was assessed using the Wilcoxon signed-rank test [136] 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 confounds more traditional interpretability methods like Shapley additive explanations (SHAP) analysis, permutation importance, or local interpretable model-agnostic explanations (LIME) [92]. Therefore, to qualitatively assess physical consistency, we derived a first-order accumulated local effects (ALE) plot of the neural parameterization, which isolates the influence of individual features even under correlated inputs. Additional discussion of ALE construction and interpretation is available in Charbonneau et al. [19], Apley and Zhu [6], and Molnar [92].

3.3 Results

Model Benchmarking

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. This is reflected by our scheme exhibiting the lowest spread in RMSE% for daily timeseries generation compared to all other schemes as-published. The distribution of RMSE% scores from all schemes as-published (various symbols, see Table 3.2) and our approach over both coarser (500-m scale) and site-level (point observations) daily data not seen during training can be seen in Fig. 3.2. All generated albedos by our scheme (NN) 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 each testing site can be further compared against other schemes in the modified Taylor Diagrams [37], which show both standard deviations and correlations of simulations and observations (Fig. 3.3). Table 3.3 in Appendix B displays numerical RMSE%, RMSE, B, B%, and Wilcoxon Signed Rank [136] 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 (Fig. 3.2). The two halves of the violin plot are relatively symmetric in addition to exhibiting narrow range, 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 Fig. S1 in the Supporting Information), 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 compared to the explicit calibration to coarse data in our approach. The

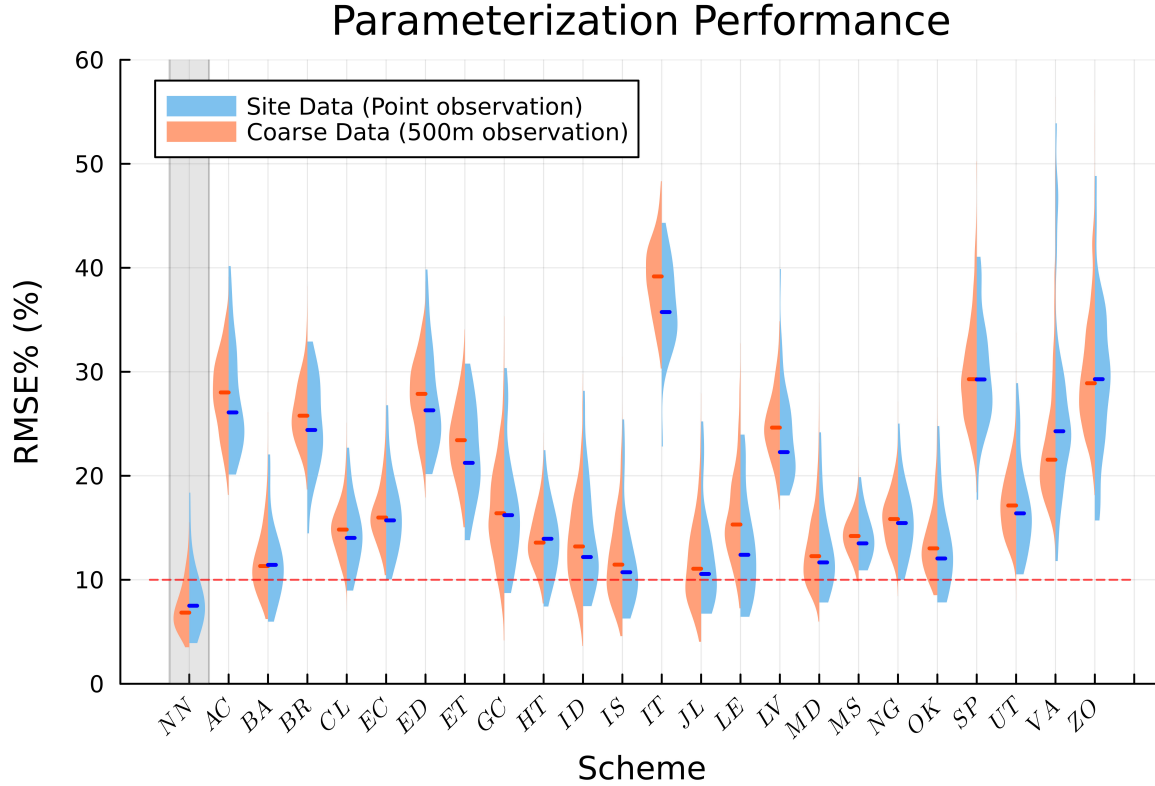


Figure 3.2: **Relative Root Mean Square errors of the new parameterization (NN) versus existing parameterizations.** This split-violin plot shows the distribution of RMSE% scores for each scheme in daily timeseries generation across all evaluated sites (700 500m scale sites on the left in orange, 31 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 3.2, 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.

lack of schemes near the unity normalized standard deviation contour in Fig. 3.3 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 3.3.

Model Consistency

Timeseries 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

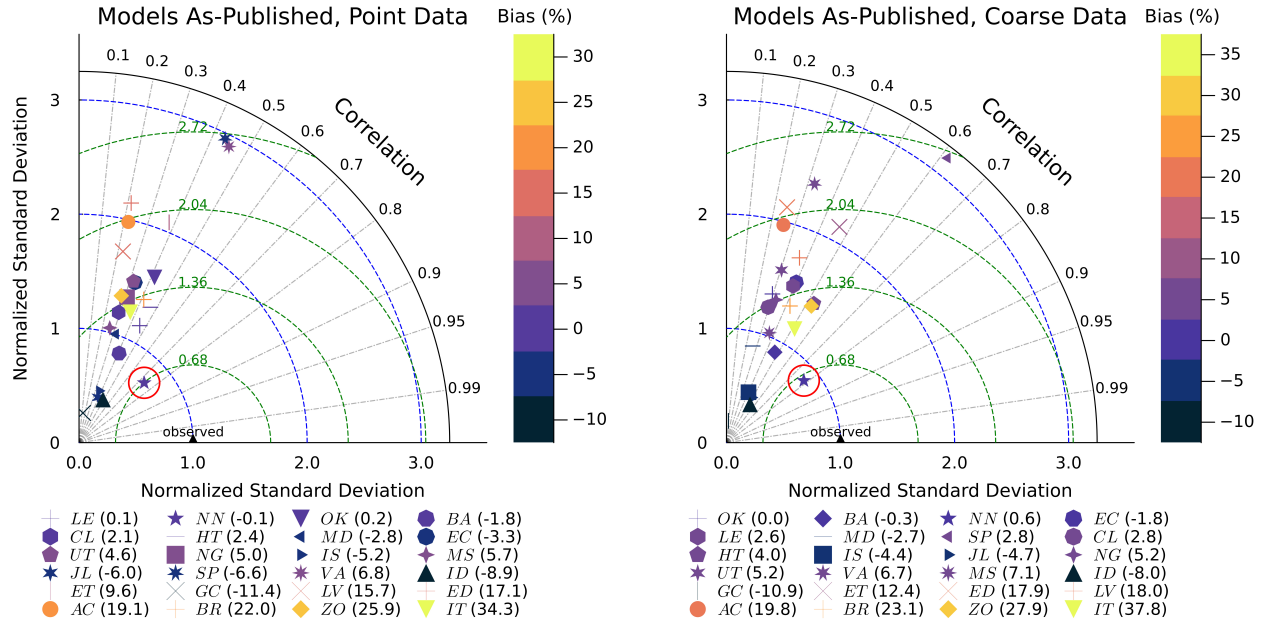


Figure 3.3: Modified Taylor Diagrams of all schemes over all ground (left) and coarse (right) testing data. The polar coordinates of each scheme mark the standard deviation of daily timeseries outputs (normalized to that of the observations; the radial coordinate) and the Pearson correlation of daily timeseries 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.

in Fig. 3.4. 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. While tracking overall seasonal dynamics via a wider variety of output ranges 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

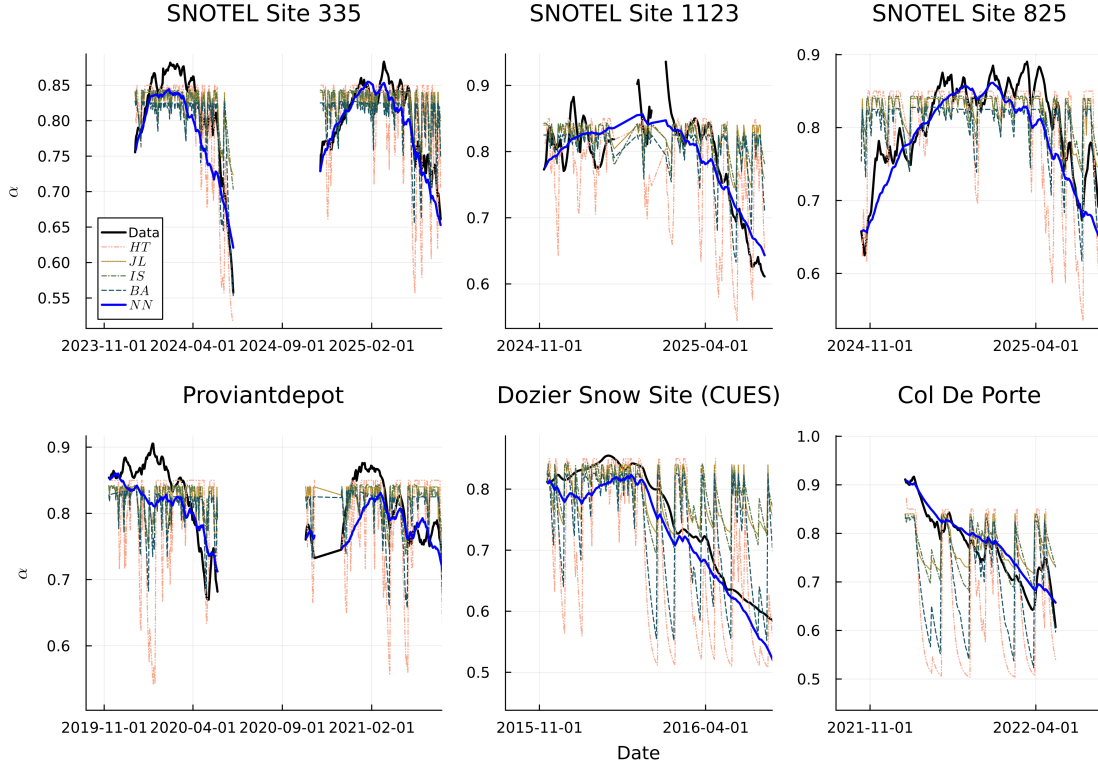


Figure 3.4: Timeseries of albedo generated by various schemes. This figure details generated timeseries over a selection of site-level testing sites compared to the processed timeseries data, using symbols provided in Table 3.2. While other schemes show a better capability at creating fast fluctuations 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 vs the fluctuations of other parameterizations.

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 [95, 96].

An Accumulated Local Effects (ALE) plot of our scheme is shown in Fig. 3.5 below. The ALE plot displays the centered changes in model output (the change in output $d\alpha/dt$ compared to the average output $d\alpha/dt$), accumulated over sequential bins of feature values. The accumulation of centered changes instead of displaying averaged outputs over data subsets isolates the model changes independent of feature correlations, allowing interpretation of curve shape (but less so scale) to gauge physical consistency.

The mean prediction of our model over all data is $d\alpha/dt \approx 0$, meaning the average changes in model output by feature value align well with physical expectations. Albedo tends to decrease unless there is a minimal amount of snowfall (~ 5 mm

ALE of Predictors

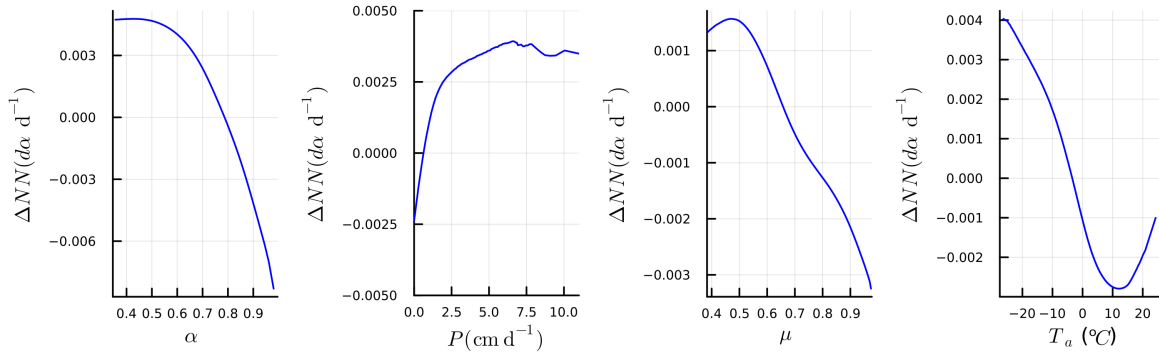


Figure 3.5: ALE Plot for features of NN. The horizontal axes give the range of each of the four input variables (present albedo α , water-equivalent snowfall P , cosine of solar noon zenith angle μ , and air temperature T_a) while the vertical axes show the average change in the scheme relative to the average prediction of albedo change per day (which is $d\alpha/dt \approx 0$). Curves are binned so that each bin has at least 50 samples to average. The rough patch for P after $\sim 5.5 \text{ cm d}^{-1}$ is an artifact of very limited sample sizes in this range, towards the extremes of the available data.

liquid-equivalent, see Fig. 3.5), which then increases but levels off as the amount of snowfall increases after a certain threshold. Similarly, the model begins to tend towards predicting decreases in albedo as temperatures approach and increase past the melting temperature, with a slower rate of change in output near colder temperatures where less melting would occur. While intraday albedo changes decrease and then increase as the sun enters and exits at low zenith angles, the daily albedo trained here uses the zenith angle cosine at local noon, which increases during warmer seasons. The plot thus shows the tendency for albedo to decrease during warmer days of the year, again aligning with physical expectations. However, data limitations can influence these results — the concavity at higher T_a could indicate 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 $T_a > 15 \text{ }^\circ\text{C}$ 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).

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

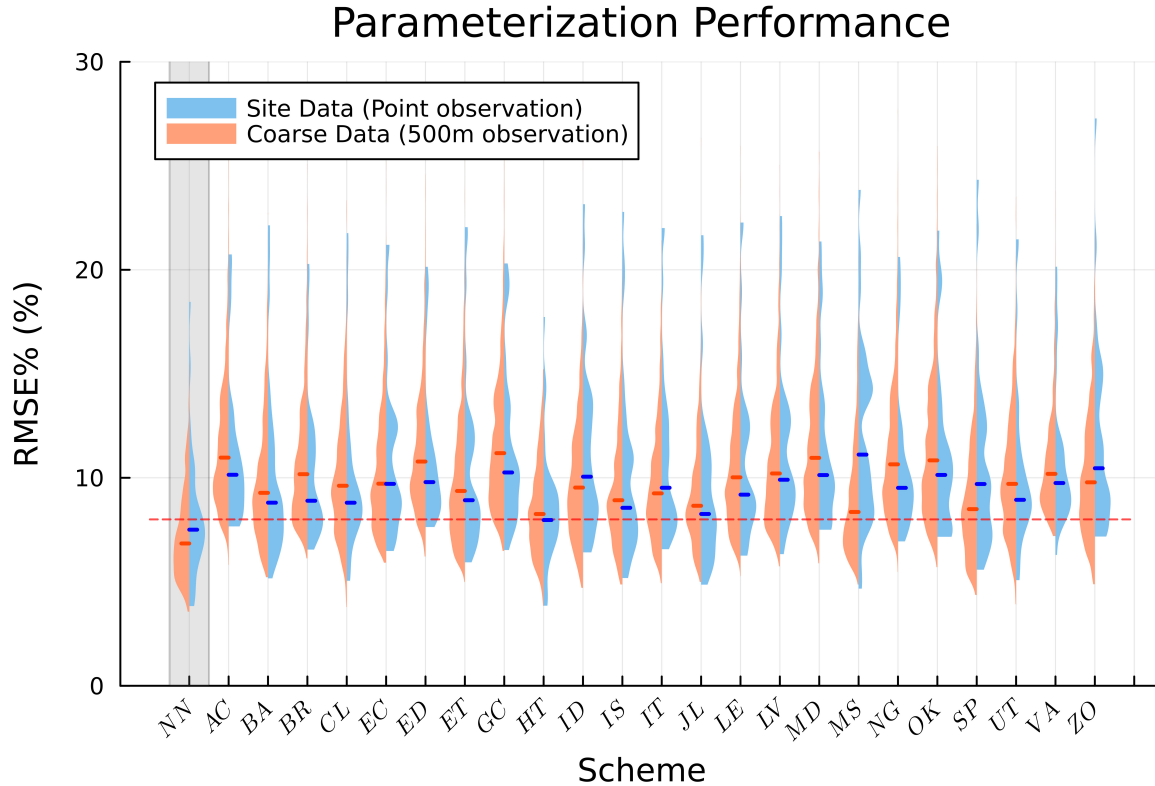


Figure 3.6: **RMSE% scores of all calibrated parameterizations.** Format follows that seen in Fig. 3.2 for daily timeseries generation, though this time the red dotted line is placed at an RMSE% of 8% instead of 10% to help better distinguish scheme performance. Our scheme remains clearly separated from other schemes.

form could achieve (see Overview in the [Supporting Information](#)). Calibrating specifically to the same data did not change the overall conclusions and rough scheme ranking, in alignment with findings from Magnusson et al. [86] and Essery et al. [39]. Many other schemes still show a large number of sites with errors in the 10-20% range after calibration (see Fig. 3.6), 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 daily timeseries generation at 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 HTESSSEL scheme (HT), for only the point-level data — our model still showed significant improvement over HTESSSEL at the coarse level. Despite this, our scheme still showed roughly 0.5 RMSE% (RMSE 0.007) better performance over point-level sites than HTESSSEL, and reduced bias by nearly an order of magnitude (see Table 3.4 in Appendix B). The BATS (BA)

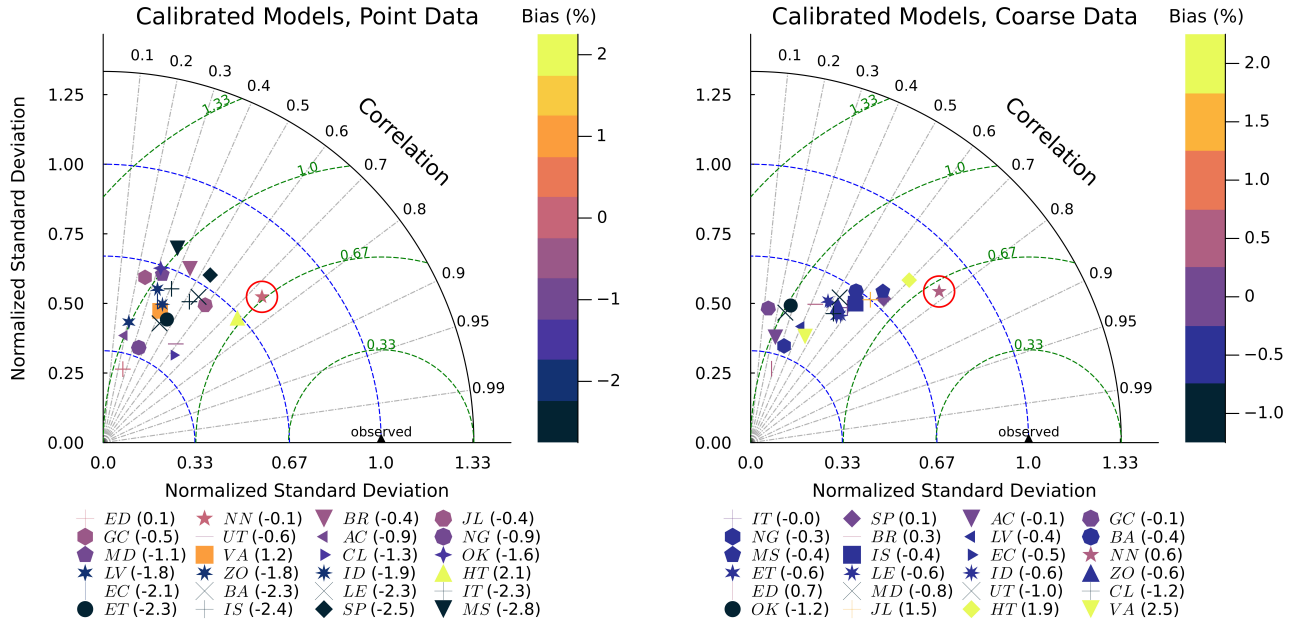


Figure 3.7: Modified Taylor Diagrams of Calibrated parameterizations. Format follows that seen in Fig. 3.3 for daily timeseries generation. Our model still exhibits the highest average correlation and is the closest to the observed values, though the spread is unsurprisingly tighter than in the uncalibrated 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.

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. [39], 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. Fig. 3.7 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 the [Supporting Information](#).

3.4 Discussion

Our parameterization improves accuracy by 10–30% relative to standard schemes while maintaining generalizability — a single quick calibration transfers across diverse locations and scales, from site-level ground observations to coarser 500 m grid-point data. This consistency suggests the approach is extracting physically meaningful processes from the input environmental and snow state data that can translate across scales. The results mirror our earlier work developing a similar neural model for snow thickness [19], reinforcing the promise of the utilized framework for usage in ever-finer-resolution GCMs. Despite its simplicity, the network maintains accuracy across diverse climate types, which is beneficial both for consistency at global scales and accuracy for local applications. This could aid robustness against nonstationarity and long-term forecasts as climate regimes shift over time, compared to other parameterizations more sensitive to their initial calibration regime. The minimal variable set also decouples the output albedo from compounding 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 [122] or full physical models such as Crocus (using the IS scheme) [4, 47] 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 [57, 82, 97, 101], 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 physical inputs, which can be applied within a GCM for future albedo simulation. Previous work like that of Ding et al. [31] 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. [143] 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 $\sim 0.055 - 0.075$ (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.,

[20, 68, 70]) and statistical approaches (e.g., [67, 141]) 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 physical inputs, ensuring interpretability and long-term applicability. Chevrollier et al. [21] 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 1000x smaller, enabling online tuning and rapid prototyping, and achieves our results with 100x less data, requiring under $2 \mu\text{s}$ 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, uncertainty, and limitations of the available 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 [61, 91]. 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 [66]. 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. The tradeoff between smoothing noisy timeseries and preserving 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 [58], grain size, snow chemistry [119], or atmospheric state. Aggregating these influences into one broadband albedo, rather than spectral or direct/diffuse components, compounds this challenge, especially near more extreme $d\alpha/dt$. While additional variables could represent more processes, direct measurement of such effects is rare, uncertain, or infeasible at GCM scales [8, 54, 102]. Our framework, however, can readily adapt if such data becomes available. Our errors are already reaching the 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.

The model captures a diverse range of $d\alpha/dt$ well but struggles with sharp increases from new snowfall at the site level. This limitation likely reflects data separability issues for extreme $d\alpha/dt$ from both the input variable choice (the feature values associated with extreme $d\alpha/dt$ 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 [1], 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 [95, 106]. 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 [9]. The proposed scheme adapts better to inherent uncertainty, enables rapid iteration, and can readily incorporate new inputs, motivating broader adoption as observational capacity expands. We note the physically consistent behavior with precipitation was learned even without choosing boundary functions to explicitly enforce it, illustrating the potential for this approach. 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 timeseries from 500 m resolution. Still, these resolutions are finer than the resolutions of current GCMs, where even more coarse-graining 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 addressed in a companion study³ where the prognostic model is integrated into a new land surface model [28]. Other additional avenues for research include sub-daily time-step evaluation, 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.

3.5 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 physical inputs at low computational burden. This reinforces the potential of a physics-informed, 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

³See Chapter 4.

(where snowlines generally move to higher altitudes and latitudes), the generalizability we demonstrated will likely translate to generalizability to different climate states. This first-time application of physics-based 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.

3.6 Data Availability

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://www.nrcs.usda.gov/wps/portal/wcc/home/>. Links to raw and processed data for Col De Porte [80], Kühtai [75], Sodankyla [40], Yala [123, 124], and Upper Rofental [135] data can be found in our previous work [19]. 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 [114]. El Tapado 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> [8]. The Insolations.jl package is developed by the Climate Modeling Alliance (CliMA), able to be downloaded at <https://clima.github.io/Insolation.jl/dev/> [23], 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 datasets 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 the [Supporting Information](#).

3.A Data Methods

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 (here-

after “SMM”) product [66, 100, 101, 112, 114] 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 [46]. SMM has been recommended for next-generation model calibration [46], though many studies still rely on outdated products, impairing parameterization accuracy and performance interpretability [3]. 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 [9, 29, 57, 66, 69, 112]. Sensor drift in the Terra/Aqua MODIS sensors also introduces minor bias [43], 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 [113], but we used SMM due to broader validation, minor differences in accuracy [57, 125], and better availability during much of this study.

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 [108]. 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 [45, 61]. Some QC standards exist in some networks, such as SNOTEL (e.g., [121]), though many data streams remain raw.

In prior work establishing our parameterization framework [19], 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 [71]. While site slope/topography data are more rare, most SNOTEL stations are built on cleared flat ground, reducing the need for additional correction.

Data Cleaning Methods

For each of the 804 SNOTEL sites inside SMM bounds, from Oct 1 2000 to Oct 1 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 [129] were set to missing, and CQC steps 7–13 were applied. Zenith angle (ZA) at solar noon was calculated using `Insolations.jl` [23] 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 [65] (>85% accuracy):

$$f = \frac{1}{1 + e^{\gamma + \beta T_a}} \quad (3.6)$$

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 June 1 2023 to June 1 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} timeseries.
7. Excise $\alpha_{\text{snow},i}$ where z_i or SWE_i are 0, missing, or $\text{SWE}_i > z_i$.
8. Excise $\alpha_{\text{snow},i}$ where $P_i = 0$ but $d\text{SWE}/dt_i > 0$ (avoid $d\alpha/dt$ 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. [114].
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 towards possible cloud albedos, but more inclined to higher values than an arithmetic mean), and create a more complete α_{snow} timeseries 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 $z_i = \text{SWE}_i = P_i = P_{i-1} = \text{days_without_observation}_i = 0$. 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 timeseries by 100 (to [0-1]).

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

Pre-Training Procedures

Remaining series exhibited unrealistic high-frequency oscillations, especially in-situ or during peak season, exceeding expected daily changes [8, 119]. 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 [10, 41]. To address these issues, we applied a two-step correction:

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:

$$EMA_i = w\alpha_i + (1 - w) EMA_{i-1}. \quad (3.7)$$

Other smoothing approaches exist (e.g., [16, 141], 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 $d\alpha/dt$ 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.

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 :

$$\alpha_{obs} = (1 - f_{snow}) \alpha_{bg} + f_{snow} \alpha_{snow}. \quad (3.8)$$

If persistent non-snow contamination prevents full snow coverage ($f_{\max} < 1$) 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:

$$\alpha_{\text{obs},\max} = (1 - f_{\max}) \alpha_{bg} + f_{\max} \alpha_{\text{theory},\max}. \quad (3.9)$$

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

$$f_{\text{max}} = \frac{\alpha_{\text{obs,max}} - \alpha_{bg}}{\alpha_{\text{theory,max}} - \alpha_{bg}}, \quad \alpha_i = \frac{\alpha_{\text{obs},i} - \alpha_{bg}(1 - f_{\text{max}})}{f_{\text{max}}}. \quad (3.10)$$

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 [34]. 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 $d\alpha/dt$ was then computed as $\frac{\alpha_{i+1} - \alpha_i}{t_{i+1} - t_i}$, retaining only consecutive days (when $t_{i+1} - t_i = 1$ 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 5,7,8,9 only. For CUES, relative humidity was capped at 1, only albedo steps 4,6,8,9 were used, and snowfall was estimated using the Zwart parameterization [145] 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. [131].

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 [139], including P_{rain} had negligible impact on outcome (as in our prior study) and was omitted for computational simplicity. A 12-h 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.

3.B Performance Statistics

Table 3.3: Parameterization scores over as-published parameterizations. Performance of the different schemes in this study across both validation and testing sites, denoting models using their labels provided in Table 3.2 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 (NN). 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.

Model	Coarse Testing Data (500 m)				Significance		Site Testing Data (Point Observation)					Significance	
	RMSE%	RMSE	Bias%	Bias	RMSE%	RMSE	RMSE%	RMSE	Bias%	Bias	RMSE%	RMSE	
NN	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)	-	-	
CL	14.83 (14.9)	0.1095 (0.1089)	2.93 (2.85)	0.0282 (0.0288)	2.87E-116	2.87E-116	14.03 (14.43)	0.1050 (0.1061)	2.15 (2.13)	0.0199 (0.0223)	9.31E-10	1.86E-09	
OK	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)	-0.46 (0.17)	-0.0004 (0.0067)	6.52E-09	1.3E-8	
LV	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)	15.06 (15.73)	0.1192 (0.1259)	9.31E-10	9.31E-10	
MD	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)	-2.11 (-2.84)	-0.0104 (-0.0144)	9.31E-10	1.86E-09	
JL	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)	-4.77 (-5.99)	-0.0311 (-0.0376)	6.52E-08	4.0E-6	
SP	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)	-8.52 (-6.59)	-0.0761 (-0.0509)	9.31E-10	9.31E-10	
MS	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)	6.61 (5.68)	0.0564 (0.0503)	9.31E-10	9.31E-10	
UT	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)	3.28 (4.62)	0.0306 (0.0398)	9.31E-10	9.31E-10	
HT	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)	2.27 (2.39)	0.0205 (0.022)	9.31E-10	2.79E-09	
EC	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)	-3.2 (-3.33)	-0.0227 (-0.0204)	9.31E-10	1.86E-09	
ZO	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)	26.75 (25.9)	0.2113 (0.2045)	9.31E-10	9.31E-10	
IT	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)	33.02 (34.28)	0.2582 (0.2673)	9.31E-10	9.31E-10	
ID	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)	-8.33 (-8.87)	-0.06 (-0.0594)	1.86E-09	1.77E-08	
BA	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)	-1.42 (-1.77)	-0.0058 (-0.007)	1.3E-8	2.86E-7	
VA	21.55 (23.13)	0.1588 (0.1740)	4.46 (6.62)	0.0371 (0.0555)	2.87E-116	2.87E-116	24.28 (26.5)	0.1776 (0.1900)	4.1 (6.83)	0.0321 (0.0525)	9.31E-10	9.31E-10	
LE	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)	0.32 (0.06)	0.0073 (0.0054)	6.52E-08	2.36E-7	
AC	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)	17.16 (19.07)	0.1357 (0.1507)	9.31E-10	9.31E-10	
GC	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)	-10.97 (-11.43)	-0.0746 (-0.0776)	1.86E-09	2.79E-09	
NG	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)	4.25 (4.96)	0.0375 (0.0434)	9.31E-10	9.31E-10	
ED	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)	15.35 (17.08)	0.1201 (0.1353)	9.31E-10	9.31E-10	
ET	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)	7.66 (9.57)	0.0643 (0.0759)	9.31E-10	9.31E-10	
BR	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)	21.86 (21.97)	0.1741 (0.1724)	1.86E-09	9.31E-10	
IS	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)	-4.38 (-5.24)	-0.028 (-0.0322)	1.57E-07	7.01E-6	

Table 3.4: Parameterization scores over calibrated parameterizations. Labeling and formatting follows that as described in Table 3.3. 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.

Model	Coarse Testing Data (500 m)				Significance		Site Testing Data (Point Observation)				Significance	
	RMSE%	RMSE	Bias%	Bias	RMSE%	RMSE	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>AC</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

3.C Hyperparameters

Similar to our previous work [19], we made use of a custom loss-function:

$$L = \frac{1}{N_d} \sum_{i=1}^{N_d} w_i |y_i - \hat{y}_i|^{n_1}, \quad w_i = 1 + |y_i|^{n_2}. \quad (3.11)$$

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 as undershooting one extreme $d\alpha/dt$ 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 timeseries RMSE% and bias (B) 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 timeseries trials performed optimally when training for approximately 100 epochs with the small network, similar to our previous findings.

Table 3.5: Hyperparameter results. Timeseries RMSE% and B% were used as the primary metrics. $n_1 = 2$ (L_2 -like metric) provided the lowest RMSE% for all choices.

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

The configuration of $n = 2$, $n_1 = 2$, and $n_2 = 0$ showed optimal performance for both timeseries RMSE% and 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 timeseries that better replicated sub-weekly variation⁴, but at the expense of whole-timeseries offsets that skewed overall performance statistics (which would similarly

⁴This is unsurprising as $n_2 \neq 0$ and/or $n_1 = 1$ choices do not suppress variance (which can lead to larger abrupt changes in model outputs) as much as training with mean squared error. Squared error for an autoregressive model is more akin to learning the mean output conditioned on the given input, while absolute error is more akin to learning the median output conditioned on the given input. For input data with smaller representation at its extremes (distribution tails), the difference between the conditional median and conditional mean in these regimes can suppress more extreme outputs when using $n_1 = 2$ compared to $n_1 = 1$.

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.

References

- [1] R. Abolafia-Rosenzweig, C. He, S. McKenzie Skiles, F. Chen, and D. Gochis. 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), Sept. 2022. doi: 10.1029/2022ms003141. URL <http://dx.doi.org/10.1029/2022MS003141>.
- [2] ACE. Snow hydrology: Summary report of the snow investigations. Published by the north pacific division, Corps of Engineers, U.S. Army, Portland, Oregon, 1956. *Journal of Glaciology*, 3(22):148–148, 1956. doi: 10.3189/s0022143000024503. URL <http://dx.doi.org/10.3189/S0022143000024503>.
- [3] F. Aguirre, D. Bozkurt, T. Sauter, J. Carrasco, C. Schneider, R. Jaña, and G. Casassa. 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, Nov. 2023. doi: 10.3390/rs15225430. URL <http://dx.doi.org/10.3390/rs15225430>.
- [4] P. M. Alexander, M. Tedesco, X. Fettweis, R. S. W. van de Wal, C. J. P. P. Smeets, and M. R. van den Broeke. Assessing spatio-temporal variability and trends in modelled and measured Greenland ice sheet albedo (2000–2013). *The Cryosphere*, 8(6):2293–2312, Dec. 2014. doi: 10.5194/tc-8-2293-2014. URL <http://dx.doi.org/10.5194/tc-8-2293-2014>.
- [5] E. A. Anderson. Development and testing of snow pack energy balance equations. *Water Resources Research*, 4(1):19–37, Feb. 1968. doi: 10.1029/wr004i001p00019. URL <http://dx.doi.org/10.1029/WR004i001p00019>.
- [6] D. W. Apley and J. Zhu. 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, June 2020. doi: 10.1111/rssb.12377. URL <http://dx.doi.org/10.1111/rssb.12377>.
- [7] J. Atwood, B. Domonkos, K. Hill, T. Brosten, T. DeMarco, M. Hultstrand, D. Tappa, L. Austin, and A. Buckman. Evaluation of YSI temperature correction equations for bias-reducing SNOTEL network temperature data. In *2023 Western Snow Conference*, 2023. URL https://www.nrcs.usda.gov/sites/default/files/2023-05/Final_Temperature_Correction_Study05262023.pdf.
- [8] E. H. Bair, R. E. Davis, and J. Dozier. Hourly mass and snow energy balance measurements from Mammoth Mountain, CA USA, 2011–2017. *Earth System Science Data*, 10(1):549–563, Mar. 2018. doi: 10.5194/essd-10-549-2018. URL <http://dx.doi.org/10.5194/essd-10-549-2018>.

- [9] E. H. Bair, K. Rittger, S. M. Skiles, and J. Dozier. An examination of snow albedo estimates from MODIS and their impact on snow water equivalent reconstruction. *Water Resources Research*, 55(9):7826–7842, Sept. 2019. doi: 10.1029/2019wr024810. URL <http://dx.doi.org/10.1029/2019WR024810>.
- [10] E. H. Bair, J. Dozier, C. Stern, A. LeWinter, K. Rittger, A. Savagian, T. Stilling, and R. E. Davis. 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, May 2022. doi: 10.5194/tc-16-1765-2022. URL <http://dx.doi.org/10.5194/tc-16-1765-2022>.
- [11] D. G. Baker, D. L. Ruschy, and D. B. Wall. The albedo decay of prairie snows. *Journal of Applied Meteorology*, 29(2):179–187, Feb. 1990. doi: 10.1175/1520-0450(1990)029<0179:tadops>2.0.co;2. URL [http://dx.doi.org/10.1175/1520-0450\(1990\)029<0179:TADOPS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0450(1990)029<0179:TADOPS>2.0.CO;2).
- [12] N. Ban, C. Caillaud, E. Coppola, E. Pichelli, S. Sobolowski, M. Adinolfi, B. Ahrens, A. Alias, I. Anders, S. Bastin, D. Belušić, S. Berthou, E. Brisson, R. M. Cardoso, S. C. Chan, O. B. Christensen, J. Fernández, L. Fita, T. Frisius, G. Gašparac, F. Giorgi, K. Goergen, J. E. Haugen, O. Hodnebrog, S. Kartsios, E. Katragkou, E. J. Kendon, K. Keuler, A. Lavin-Gullon, G. Lenderink, D. Leutwyler, T. Lorenz, D. Maraun, P. Mercogliano, J. Milovac, H.-J. Panitz, M. Raffa, A. R. Remedio, C. Schär, P. M. Soares, L. Srnec, B. M. Steensen, P. Stocchi, M. H. Tölle, H. Truhetz, J. Vergara-Temprado, H. de Vries, K. Warrach-Sagi, V. Wulfmeyer, and M. J. Zander. The first multi-model ensemble of regional climate simulations at kilometer-scale resolution, Part I: Evaluation of precipitation. *Climate Dynamics*, 57(1–2):275–302, Apr. 2021. doi: 10.1007/s00382-021-05708-w. URL <http://dx.doi.org/10.1007/s00382-021-05708-w>.
- [13] M. J. Best, M. Pryor, D. B. Clark, G. G. Rooney, R. L. H. Essery, C. B. Ménard, J. M. Edwards, M. A. Hendry, A. Porson, N. Gedney, L. M. Mercado, S. Sitch, E. Blyth, O. Boucher, P. M. Cox, C. S. B. Grimmond, and R. J. Harding. The joint UK land environment simulator (JULES), model description – Part 1: Energy and water fluxes. *Geoscientific Model Development*, 4(3):677–699, Sept. 2011. doi: 10.5194/gmd-4-677-2011. URL <http://dx.doi.org/10.5194/gmd-4-677-2011>.
- [14] B. W. Brock, I. C. Willis, and M. J. Sharp. Measurement and parameterization of albedo variations at Haut Glacier d’Arolla, Switzerland. *Journal of Glaciology*, 46(155):675–688, 2000. doi: 10.3189/172756500781832675. URL <http://dx.doi.org/10.3189/172756500781832675>.
- [15] E. Brun, P. David, M. Sudul, and G. Brunot. A numerical model to simulate snow-cover stratigraphy for operational avalanche forecasting. *Journal of*

- Glaciology*, 38(128):13–22, 1992. doi: 10.3189/s0022143000009552. URL <http://dx.doi.org/10.3189/S0022143000009552>.
- [16] J. F. Calleja, A. Corbea-Perez, S. Fernandez, C. Recondo, J. Peon, and M. A. de Pablo. Snow albedo seasonality and trend from MODIS sensor and ground data at Johnsons Glacier, Livingston Island, Maritime Antarctica. *Sensors*, 19(16):3569, Aug. 2019. doi: 10.3390/s19163569. URL <http://dx.doi.org/10.3390/s19163569>.
 - [17] C. Cassardo. UTOPIA: The manual of version 2015. 8594, 2015. doi: 10.13140/RG.2.2.29664.38404. URL <https://www.researchgate.net/doi/10.13140/RG.2.2.29664.38404>.
 - [18] H. G. Chan. Parameterization of snowpack albedo reduction by light-absorbing impurities. *Geophysical Fluid Dynamics Laboratory (U.S.)*, 2022. doi: 10.25923/527J-0A46. URL <https://repository.library.noaa.gov/view/noaa/45392>.
 - [19] A. Charbonneau, K. Deck, and T. Schneider. A physics-constrained neural differential equation framework for data-driven snowpack simulation. *Artificial Intelligence for the Earth Systems*, 4(3), July 2025. doi: 10.1175/aies-d-24-0040.1. URL <http://dx.doi.org/10.1175/AIES-D-24-0040.1>.
 - [20] S. Chen, P. Xiao, X. Zhang, P. Pellikka, H. Liu, and Y. Liu. 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, Oct. 2025. doi: 10.1016/j.isprsjprs.2025.07.033. URL <http://dx.doi.org/10.1016/j.isprsjprs.2025.07.033>.
 - [21] L.-A. Chevrollier, A. Wehrlé, J. M. Cook, N. Pirk, L. G. Benning, A. M. Anesio, and M. Tranter. Separating the albedo-reducing effect of different light-absorbing particles on snow using deep learning. *The Cryosphere*, 19(4):1527–1538, 2025. doi: 10.5194/tc-19-1527-2025. URL <https://tc.copernicus.org/articles/19/1527/2025/>.
 - [22] M. P. Clark, Y. Fan, D. M. Lawrence, J. C. Adam, D. Bolster, D. J. Gochis, R. P. Hooper, M. Kumar, L. R. Leung, D. S. Mackay, R. M. Maxwell, C. Shen, S. C. Swenson, and X. Zeng. Improving the representation of hydrologic processes in earth system models. *Water Resources Research*, 51(8):5929–5956, Aug. 2015. doi: 10.1002/2015wr017096. URL <http://dx.doi.org/10.1002/2015WR017096>.
 - [23] C. Contributors. Insolation.jl, 2020. URL <https://github.com/CliMA/Insolation.jl>. GitHub repository. Apache-2.0 license.
 - [24] R. R. Cordero, E. Sepúlveda, S. Feron, C. Wang, A. Damiani, F. Fernandoy, S. Neshyba, P. M. Rowe, V. Asencio, J. Carrasco, J. A. Alfonso, S. MacDonell, G. Seckmeyer, J. M. Carrera, J. Jorquera, P. Llanillo, J. Dana, A. L. Khan, and

- G. Casassa. Black carbon in the Southern Andean snowpack. *Environmental Research Letters*, 17(4):044042, Mar. 2022. doi: 10.1088/1748-9326/ac5df0. URL <http://dx.doi.org/10.1088/1748-9326/ac5df0>.
- [25] A. Damiani, N. N. Ishizaki, S. Feron, and R. R. Cordero. 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, Feb. 2025. doi: 10.1016/j.scitotenv.2025.178606. URL <http://dx.doi.org/10.1016/j.scitotenv.2025.178606>.
- [26] C. De Michele, F. Avanzi, A. Ghezzi, and C. Jommi. Investigating the dynamics of bulk snow density in dry and wet conditions using a one-dimensional model. *The Cryosphere*, 7(2):433–444, Mar. 2013. doi: 10.5194/tc-7-433-2013. URL <http://dx.doi.org/10.5194/tc-7-433-2013>.
- [27] B. Decharme, E. Brun, A. Boone, C. Delire, P. Le Moigne, and S. Morin. 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, Apr. 2016. doi: 10.5194/tc-10-853-2016. URL <http://dx.doi.org/10.5194/tc-10-853-2016>.
- [28] K. Deck, R. K. Braghiere, A. A. Renchon, J. Sloan, G. Bozzola, E. Speer, J. Ben Mackay, T. Reddy, K. Phan, A. L. Gagné-Landmann, Y. Li, D. Yatunin, A. Charbonneau, N. Efrat-Henrici, E. Bach, S. Ma, P. Gentine, C. Frankenberg, A. A. Bloom, Y. Wang, M. Longo, and T. Schneider. Climaland: A land surface model designed to enable data-driven parameterizations. *Journal of Advances in Modeling Earth Systems*, 18(1):e2025MS005118, 2026. doi: <https://doi.org/10.1029/2025MS005118>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2025MS005118>. e2025MS005118 2025MS005118.
- [29] G. Demil, A. T. Haghighi, B. Klöve, and M. Oussalah. Advances in image-based estimation of snow variable: A systematic literature review on recent studies. *Journal of Hydrology*, 654:132855, June 2025. doi: 10.1016/j.jhydrol.2025.132855. URL <http://dx.doi.org/10.1016/j.jhydrol.2025.132855>.
- [30] R. Dickinson, A. Henderson-Sellers, P. Kennedy, and M. Wilson. Biosphere-atmosphere transfer scheme (BATS) for the NCAR community climate model. Technical report, NATIONAL CENTER FOR: ATMOSPHERIC RESEARCH, 1986. URL <https://doi.org/10.5065/d6668b58>.
- [31] A. Ding, S. Liang, H. Ma, T. He, A. Jia, and Q. Wang. Improved estimation of daily blue-sky snow shortwave albedo from MODIS data and reanalysis information. *Science of Remote Sensing*, 10:100163, Dec. 2024. doi: 10.1016/j.srs.2024.100163. URL <http://dx.doi.org/10.1016/j.srs.2024.100163>.

- [32] G. T. Diro and L. Sushama. Snow-precipitation coupling and related atmospheric feedbacks over North America. *Atmospheric Science Letters*, 19(8), Aug. 2018. doi: 10.1002/asl.831. URL <http://dx.doi.org/10.1002/asl.831>.
- [33] J. Dong, M. Ek, D. Hall, C. Peters-Lidard, B. Cosgrove, J. Miller, G. Riggs, and Y. Xia. 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, Apr. 2014. doi: 10.1175/jhm-d-13-060.1. URL <http://dx.doi.org/10.1175/JHM-D-13-060.1>.
- [34] J. Dozier, R. O. Green, A. W. Nolin, and T. H. Painter. Interpretation of snow properties from imaging spectrometry. *Remote Sensing of Environment*, 113:S25–S37, Sept. 2009. doi: 10.1016/j.rse.2007.07.029. URL <http://dx.doi.org/10.1016/j.rse.2007.07.029>.
- [35] R. V. Dunkle and J. T. Bevens. An approximate analysis of the solar reflectance and transmittance of a snow cover. *Journal of Meteorology*, 13(2):212–216, Apr. 1956. doi: 10.1175/1520-0469(1956)013<0212:aaaots>2.0.co;2. URL [http://dx.doi.org/10.1175/1520-0469\(1956\)013<0212:AAAOTS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0469(1956)013<0212:AAAOTS>2.0.CO;2).
- [36] E. Dutra, G. Balsamo, P. Viterbo, P. M. A. Miranda, A. Beljaars, C. Schär, and K. Elder. An improved snow scheme for the ECMWF land surface model: Description and offline validation. *Journal of Hydrometeorology*, 11(4):899–916, Aug. 2010. doi: 10.1175/2010jhm1249.1. URL <http://dx.doi.org/10.1175/2010JHM1249.1>.
- [37] S. Elvidge, M. J. Angling, and B. Nava. On the use of modified Taylor diagrams to compare ionospheric assimilation models. *Radio Science*, 49(9): 737–745, Sept. 2014. doi: 10.1002/2014rs005435. URL <http://dx.doi.org/10.1002/2014RS005435>.
- [38] R. Essery, M. Best, and P. Cox. *MOSES 2.2 technical documentation*, volume 30. Hadley Centre for Climate Prediction and Research, 2001. URL <https://jules.jchmr.org/sites/default/files/2023-06/JULES-HCTN-30.pdf>.
- [39] R. Essery, S. Morin, Y. Lejeune, and C. B. Ménard. A comparison of 1701 snow models using observations from an alpine site. *Advances in Water Resources*, 55:131–148, May 2013. doi: 10.1016/j.advwatres.2012.07.013. URL <http://dx.doi.org/10.1016/j.advwatres.2012.07.013>.
- [40] R. Essery, A. Kontu, J. Lemmetyinen, M. Dumont, and C. B. Ménard. 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, June 2016. doi: 10.5194/gi-5-219-2016. URL <http://dx.doi.org/10.5194/gi-5-219-2016>.

- [41] P. Etchevers, E. Martin, R. Brown, C. Fierz, Y. Lejeune, E. Bazile, A. Boone, Y.-J. Dai, R. Essery, A. Fernandez, Y. Gusev, R. Jordan, V. Koren, E. Kowalczyk, N. O. Nasonova, R. D. Pyles, A. Schlosser, A. B. Shmakin, T. G. Smirnova, U. Strasser, D. Versegny, T. Yamazaki, and Z.-L. Yang. Validation of the energy budget of an alpine snow-pack simulated by several snow models (Snow MIP project). *Annals of Glaciology*, 38:150–158, 2004. doi: 10.3189/172756404781814825. URL <http://dx.doi.org/10.3189/172756404781814825>.
- [42] V. Eyring, S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor. Overview of the coupled model intercomparison project phase 6 (CMIP6): Experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958, 2016. doi: 10.5194/gmd-9-1937-2016. URL <https://gmd.copernicus.org/articles/9/1937/2016/>.
- [43] S. Feng, A. Wehrlé, J. M. Cook, A. M. Anesio, J. E. Box, L. G. Benning, and M. Tranter. The apparent effect of orbital drift on time series of MODIS MOD10A1 albedo on the Greenland ice sheet. *Science of Remote Sensing*, 9:100116, June 2024. doi: 10.1016/j.srs.2023.100116. URL <http://dx.doi.org/10.1016/j.srs.2023.100116>.
- [44] M. G. Flanner, J. B. Arnheim, J. M. Cook, C. Dang, C. He, X. Huang, D. Singh, S. M. Skiles, C. A. Whicker, and C. S. Zender. SNICAR-ADv3: A community tool for modeling spectral snow albedo. *Geoscientific Model Development*, 14(12):7673–7704, Dec. 2021. doi: 10.5194/gmd-14-7673-2021. URL <http://dx.doi.org/10.5194/gmd-14-7673-2021>.
- [45] S. W. Fleming, L. Zukiewicz, M. L. Strobel, H. Hofman, and A. G. Goodbody. SNOTEL, the soil climate analysis network, and water supply forecasting at the Natural Resources Conservation Service (NRCS): Past, present, and future. *JAWRA Journal of the American Water Resources Association*, 59(4):585–599, Feb. 2023. doi: 10.1111/1752-1688.13104. URL <http://dx.doi.org/10.1111/1752-1688.13104>.
- [46] S. W. Fleming, K. Rittger, C. M. Oaida Taglialatela, and I. Graczyk. 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), Apr. 2024. doi: 10.1029/2023wr035785. URL <http://dx.doi.org/10.1029/2023WR035785>.
- [47] M. Gaillard, V. Vionnet, M. Lafaysse, M. Dumont, and P. Ginoux. Improving large-scale snow albedo modeling using a climatology of light-absorbing particle deposition. *The Cryosphere*, 19(2):769–792, Feb. 2025. doi: 10.5194/tc-19-769-2025. URL <http://dx.doi.org/10.5194/tc-19-769-2025>.

- [48] L. Gao, L. Zhang, Y. Shen, Y. Zhang, M. Ai, and W. Zhang. Modeling snow depth and snow water equivalent distribution and variation characteristics in the Irtysh River Basin, China. *Applied Sciences*, 11(18):8365, Sept. 2021. doi: 10.3390/app11188365. URL <http://dx.doi.org/10.3390/app11188365>.
- [49] A. S. Gardner and M. J. Sharp. A review of snow and ice albedo and the development of a new physically based broadband albedo parameterization. *Journal of Geophysical Research: Earth Surface*, 115(F1), Mar. 2010. doi: 10.1029/2009jf001444. URL <http://dx.doi.org/10.1029/2009JF001444>.
- [50] J. C. Giddings and E. LaChapelle. Diffusion theory applied to radiant energy distribution and albedo of snow. *Journal of Geophysical Research*, 66(1): 181–189, Jan. 1961. doi: 10.1029/jz066i001p00181. URL <http://dx.doi.org/10.1029/JZ066i001p00181>.
- [51] Y. Gorishniy, I. Rubachev, V. Khrulkov, and A. Babenko. Revisiting deep learning models for tabular data, 2023. URL <https://arxiv.org/abs/2106.11959>.
- [52] W. Greuell and T. Konzelmann. 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, Jan. 1994. doi: 10.1016/0921-8181(94)90010-8. URL [http://dx.doi.org/10.1016/0921-8181\(94\)90010-8](http://dx.doi.org/10.1016/0921-8181(94)90010-8).
- [53] L. Grinsztajn, E. Oyallon, and G. Varoquaux. Why do tree-based models still outperform deep learning on tabular data?, 2022. URL <https://arxiv.org/abs/2207.08815>.
- [54] D. Günther, T. Marke, R. Essery, and U. Strasser. 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, Apr. 2019. doi: 10.1029/2018wr023403. URL <http://dx.doi.org/10.1029/2018WR023403>.
- [55] D. K. Hall and G. A. Riggs. Accuracy assessment of the MODIS snow products. *Hydrological Processes*, 21(12):1534–1547, May 2007. doi: 10.1002/hyp.6715. URL <http://dx.doi.org/10.1002/hyp.6715>.
- [56] J. Hansen, G. Russell, D. Rind, P. Stone, A. Lacis, S. Lebedeff, R. Ruedy, and L. Travis. Efficient three-dimensional global models for climate studies: Models I and II. *Monthly Weather Review*, 111(4):609–662, Apr. 1983. doi: 10.1175/1520-0493(1983)111<0609:etdgmf>2.0.co;2. URL [http://dx.doi.org/10.1175/1520-0493\(1983\)111<0609:ETDGMF>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1983)111<0609:ETDGMF>2.0.CO;2).

- [57] D. Hao, G. Bisht, K. Rittger, T. Stillinger, E. Bair, Y. Gu, and L. R. Leung. Evaluation of E3SM land model snow simulations over the Western United States. *The Cryosphere*, 17(2):673–697, Feb. 2023. doi: 10.5194/tc-17-673-2023. URL <http://dx.doi.org/10.5194/tc-17-673-2023>.
- [58] B. J. Hatchett, A. L. Koshkin, K. Guirguis, K. Rittger, A. W. Nolin, A. Heggli, A. M. Rhoades, A. E. East, E. R. Siirila-Woodburn, W. T. Brandt, A. Gershunov, and K. Haleakala. Midwinter dry spells amplify post-fire snowpack decline. *Geophysical Research Letters*, 50(3), Feb. 2023. doi: 10.1029/2022gl101235. URL <http://dx.doi.org/10.1029/2022GL101235>.
- [59] C. He, M. Flanner, D. M. Lawrence, and Y. Gu. 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), Feb. 2024. doi: 10.1029/2023ms003861. URL <http://dx.doi.org/10.1029/2023MS003861>.
- [60] H. E. Henneman and H. G. Stefan. Albedo models for snow and ice on a freshwater lake. *Cold Regions Science and Technology*, 29(1):31–48, Apr. 1999. doi: 10.1016/S0165-232X(99)00002-6. URL [http://dx.doi.org/10.1016/S0165-232X\(99\)00002-6](http://dx.doi.org/10.1016/S0165-232X(99)00002-6).
- [61] D. F. Hill, E. A. Burakowski, R. L. Crumley, J. Keon, J. M. Hu, A. A. Arendt, K. Wikstrom Jones, and G. J. Wolken. Converting snow depth to snow water equivalent using climatological variables. *The Cryosphere*, 13(7):1767–1784, July 2019. doi: 10.5194/tc-13-1767-2019. URL <http://dx.doi.org/10.5194/tc-13-1767-2019>.
- [62] G. Hinton, N. Srivastava, and K. Swersky. Lecture 6e: Rmsprop: Divide the gradient by a running average of its recent magnitude, 2012. URL https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf.
- [63] Y. Huang, Z. Song, H. Yang, B. Yu, H. Liu, T. Che, J. Chen, J. Wu, S. Shu, X. Peng, Z. Zheng, and J. Xu. Snow cover detection in mid-latitude mountainous and polar regions using nighttime light data. *Remote Sensing of Environment*, 268:112766, Jan. 2022. doi: 10.1016/j.rse.2021.112766. URL <http://dx.doi.org/10.1016/j.rse.2021.112766>.
- [64] M. Innes. Flux: Elegant machine learning with Julia. *Journal of Open Source Software*, 3(25):602, May 2018. doi: 10.21105/joss.00602. URL <http://dx.doi.org/10.21105/joss.00602>.
- [65] K. S. Jennings, T. S. Winchell, B. Livneh, and N. P. Molotch. Spatial variation of the rain–snow temperature threshold across the Northern Hemisphere. *Nature Communications*, 9(1), Mar. 2018. doi: 10.1038/s41467-018-03629-7. URL <http://dx.doi.org/10.1038/s41467-018-03629-7>.

- [66] A. S. Jensen, K. Rittger, and M. S. Raleigh. 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, Mar. 2024. doi: 10.1088/2752-5295/ad285a. URL <http://dx.doi.org/10.1088/2752-5295/ad285a>.
- [67] A. Jia, D. Wang, S. Liang, J. Peng, and Y. Yu. Improved cloudy-sky snow albedo estimates using passive microwave and VIIRS data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 196:340–355, Feb. 2023. doi: 10.1016/j.isprsjprs.2023.01.004. URL <http://dx.doi.org/10.1016/j.isprsjprs.2023.01.004>.
- [68] J. Jia, Z. Zeng, B. Tian, S. Chen, L. Chen, Y. Wang, X. Wang, and M. Ding. The construction of 20-year daily surface albedo along PANDA Transect, Antarctica. *Journal of Geophysical Research: Atmospheres*, 130(5), Mar. 2025. doi: 10.1029/2024jd042863. URL <http://dx.doi.org/10.1029/2024JD042863>.
- [69] L. Keuris, M. Hetzenecker, T. Nagler, N. Mölg, and G. Schwaizer. 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, Feb. 2023. doi: 10.3390/rs15051231. URL <http://dx.doi.org/10.3390/rs15051231>.
- [70] J. Kiem, A. Hammerle, L. Montagnani, and G. Wohlfahrt. Future scenarios of albedo and radiative forcing resulting from changes in snow depth in Austria. *EGUsphere*, 2024:1–22, Apr. 2024. doi: 10.5194/egusphere-2024-881. URL <https://egusphere.copernicus.org/preprints/2024/egusphere-2024-881/>.
- [71] A. Klein. 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, July 2003. doi: 10.1016/s0034-4257(03)00097-x. URL [http://dx.doi.org/10.1016/S0034-4257\(03\)00097-X](http://dx.doi.org/10.1016/S0034-4257(03)00097-X).
- [72] A. G. Klein and J. Stroeve. Development and validation of a snow albedo algorithm for the MODIS instrument. *Annals of Glaciology*, 34:45–52, 2002. doi: 10.3189/172756402781817662. URL <http://dx.doi.org/10.3189/172756402781817662>.
- [73] J. Kondo and T. Yamazaki. A prediction model for snowmelt, snow surface temperature and freezing depth using a heat balance method. *Journal of Applied Meteorology*, 29(5):375–384, May 1990. doi: 10.1175/1520-0450(1990)029<0375:apmfss>2.0.co;2. URL [http://dx.doi.org/10.1175/1520-0450\(1990\)029<0375:APMFSS>2.0.CO;2](http://dx.doi.org/10.1175/1520-0450(1990)029<0375:APMFSS>2.0.CO;2).
- [74] K. Kouki, P. Räisänen, K. Luojus, A. Luomaranta, and A. Riihelä. Evaluation of Northern Hemisphere snow water equivalent in CMIP6 models during

- 1982–2014. *The Cryosphere*, 16(3):1007–1030, Mar. 2022. doi: 10.5194/tc-16-1007-2022. URL <http://dx.doi.org/10.5194/tc-16-1007-2022>.
- [75] P. Krajčič, R. Kirnbauer, J. Parajka, J. Schöber, and G. Blöschl. The Kühtai data set: 25 years of lysimetric, snow pillow, and meteorological measurements. *Water Resources Research*, 53(6):5158–5165, June 2017. doi: 10.1002/2017wr020445. URL <http://dx.doi.org/10.1002/2017WR020445>.
- [76] G. Krinner, C. Derksen, R. Essery, M. Flanner, S. Hagemann, M. Clark, A. Hall, H. Rott, C. Brutel-Vuilmet, H. Kim, C. B. Ménard, L. Mudryk, C. Thackeray, L. Wang, G. Arduini, G. Balsamo, P. Bartlett, J. Boike, A. Boone, F. Chéruiy, J. Colin, M. Cuntz, Y. Dai, B. Decharme, J. Derry, A. Ducharne, E. Dutra, X. Fang, C. Fierz, J. Ghattas, Y. Gusev, V. Haverd, A. Kontu, M. Lafaysse, R. Law, D. Lawrence, W. Li, T. Marke, D. Marks, M. Ménégos, O. Nasonova, T. Nitta, M. Niwano, J. Pomeroy, M. S. Raleigh, G. Schaedler, V. Semenov, T. G. Smirnova, T. Stacke, U. Strasser, S. Svensson, D. Turkov, T. Wang, N. Wever, H. Yuan, W. Zhou, and D. Zhu. ESM-SnowMIP: Assessing snow models and quantifying snow-related climate feedbacks. *Geoscientific Model Development*, 11(12):5027–5049, Dec. 2018. doi: 10.5194/gmd-11-5027-2018. URL <http://dx.doi.org/10.5194/gmd-11-5027-2018>.
- [77] D. M. Lawrence, K. W. Oleson, M. G. Flanner, P. E. Thornton, S. C. Swenson, P. J. Lawrence, X. Zeng, Z.-L. Yang, S. Levis, K. Sakaguchi, G. B. Bonan, and A. G. Slater. Parameterization improvements and functional and structural advances in version 4 of the Community Land Model (CLM). *Journal of Advances in Modeling Earth Systems*, 3(1), Jan. 2011. doi: 10.1029/2011ms00045. URL <http://dx.doi.org/10.1029/2011MS00045>.
- [78] W. Y. Lee, H.-J. Gim, and S. K. Park. 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, Nov. 2023. doi: 10.1007/s13143-023-00344-2. URL <http://dx.doi.org/10.1007/s13143-023-00344-2>.
- [79] M. Lehning, P. Bartelt, B. Brown, and C. Fierz. A physical SNOWPACK model for the Swiss avalanche warning. *Cold Regions Science and Technology*, 35(3):169–184, Nov. 2002. doi: 10.1016/S0165-232X(02)00072-1. URL [http://dx.doi.org/10.1016/S0165-232X\(02\)00072-1](http://dx.doi.org/10.1016/S0165-232X(02)00072-1).
- [80] Y. Lejeune, M. Dumont, J.-M. Panel, M. Lafaysse, P. Lapalus, E. Le Gac, B. Lesaffre, and S. Morin. 57years (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, Jan. 2019. doi: 10.5194/essd-11-71-2019. URL <http://dx.doi.org/10.5194/essd-11-71-2019>.

- [81] H. Li, G. Zhang, C. Wang, Z. Liu, C. Ju, and A. Mamtimin. Improving snow albedo parameterization scheme based on remote sensing data. *Atmospheric Research*, 284:106602, Mar. 2023. doi: 10.1016/j.atmosres.2022.106602. URL <http://dx.doi.org/10.1016/j.atmosres.2022.106602>.
- [82] T.-S. Lin, C. He, R. Abolafia-Rosenzweig, F. Chen, W. Wang, M. Barlage, and D. Gochis. 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, Feb. 2025. doi: 10.1175/jhm-d-24-0082.1. URL <http://dx.doi.org/10.1175/JHM-D-24-0082.1>.
- [83] L. Liu, Y. Ma, M. Menenti, R. Su, N. Yao, and W. Ma. Improved parameterization of snow albedo in Noah coupled with Weather Research and Forecasting (WRF): applicability to snow estimates for the Tibetan Plateau. *Hydrology and Earth System Sciences*, 25(9):4967–4981, Sept. 2021. doi: 10.5194/hess-25-4967-2021. URL <http://dx.doi.org/10.5194/hess-25-4967-2021>.
- [84] B. Livneh, Y. Xia, K. E. Mitchell, M. B. Ek, and D. P. Lettenmaier. Noah LSM snow model diagnostics and enhancements. *Journal of Hydrometeorology*, 11(3):721–738, June 2010. doi: 10.1175/2009jhm1174.1. URL <http://dx.doi.org/10.1175/2009JHM1174.1>.
- [85] B. Livneh, J. S. Deems, D. Schneider, J. J. Barsugli, and N. P. Molotch. Filling in the gaps: Inferring spatially distributed precipitation from gauge observations over complex terrain. *Water Resources Research*, 50(11):8589–8610, Nov. 2014. doi: 10.1002/2014wr015442. URL <http://dx.doi.org/10.1002/2014WR015442>.
- [86] J. Magnusson, N. Wever, R. Essery, N. Helbig, A. Winstral, and T. Jonas. Evaluating snow models with varying process representations for hydrological applications. *Water Resources Research*, 51(4):2707–2723, Apr. 2015. doi: 10.1002/2014wr016498. URL <http://dx.doi.org/10.1002/2014WR016498>.
- [87] M. J. Malik, R. van der Velde, Z. Vekerdy, and Z. Su. Improving modeled snow albedo estimates during the spring melt season. *Journal of Geophysical Research: Atmospheres*, 119(12):7311–7331, June 2014. doi: 10.1002/2013jd021344. URL <http://dx.doi.org/10.1002/2013JD021344>.
- [88] D. G. Marks. *Climate, energy exchange, and snowmelt in Emerald Lake watershed, Sierra Nevada*. Department of Geography and Mechanical Engineering., University of California, Santa Barbara, 1988. pdf from author available at https://www.researchgate.net/publication/258437741_Climate_Energy_Exchange_and_Snowmelt_in_Emerald_Lake_Watershed_Sierra_Nevada.

- [89] S. E. Marshall. A physical parameterization of snow albedo for use in climate models, NCAR cooperative thesis 123. *Ph. D. Thesis*, 1989. URL https://atmos.uw.edu/~sgw/PAPERS/1989_Marshall.pdf.
- [90] C. B. Menard, R. Essery, G. Krinner, G. Arduini, P. Bartlett, A. Boone, C. Brutel-Vuilmet, E. Burke, M. Cuntz, Y. Dai, B. Decharme, E. Dutra, X. Fang, C. Fierz, Y. Gusev, S. Hagemann, V. Haverd, H. Kim, M. Lafaysse, T. Marke, O. Nasonova, T. Nitta, M. Niwano, J. Pomeroy, G. Schädler, V. A. Semenov, T. Smirnova, U. Strasser, S. Swenson, D. Turkov, N. Wever, and H. Yuan. Scientific and human errors in a snow model intercomparison. *Bulletin of the American Meteorological Society*, 102(1):E61–E79, Jan. 2021. doi: 10.1175/bams-d-19-0329.1. URL <http://dx.doi.org/10.1175/BAMS-D-19-0329.1>.
- [91] J. D. D. Meyer, J. Jin, and S.-Y. Wang. 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, Dec. 2012. doi: 10.1175/jhm-d-12-066.1. URL <http://dx.doi.org/10.1175/JHM-D-12-066.1>.
- [92] C. Molnar. *Interpretable Machine Learning*. Christoph Molnar, 3 edition, 2025. ISBN 978-3-911578-03-5. URL <https://christophm.github.io/interpretable-ml-book>.
- [93] N. P. Molotch and R. C. Bales. Comparison of ground-based and airborne snow surface albedo parameterizations in an alpine watershed: Impact on snowpack mass balance. *Water Resources Research*, 42(5), May 2006. doi: 10.1029/2005wr004522. URL <http://dx.doi.org/10.1029/2005WR004522>.
- [94] N. Mölders, H. Luijting, and K. Sassen. 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, Oct. 2007. doi: 10.1007/s00703-007-0271-6. URL <http://dx.doi.org/10.1007/s00703-007-0271-6>.
- [95] National Research Council. *Making Climate Forecasts Matter*, chapter 2, pages 18–37. The National Academies Press, Washington, DC, 1999. ISBN 978-0-309-06475-0. doi: 10.17226/6370. URL <https://nap.nationalacademies.org/catalog/6370/making-climate-forecasts-matter>.
- [96] H. Nguyen, R. Mehrotra, and A. Sharma. Correcting for systematic biases in GCM simulations in the frequency domain. *Journal of Hydrology*, 538:117–126, 2016. ISSN 0022-1694. doi: <https://doi.org/10.1016/j.jhydrol.2016.04.018>. URL <https://www.sciencedirect.com/science/article/pii/S0022169416301998>.

- [97] C. M. Oaida, Y. Xue, M. G. Flanner, S. M. Skiles, F. De Sales, and T. H. Painter. 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, Apr. 2015. doi: 10.1002/2014jd022444. URL <http://dx.doi.org/10.1002/2014JD022444>.
- [98] J. Oerlemans and W. H. Knap. A 1 year record of global radiation and albedo in the ablation zone of Morteratschgletscher, Switzerland. *Journal of Glaciology*, 44(147):231–238, 1998. doi: 10.3189/s0022143000002574. URL <http://dx.doi.org/10.3189/S0022143000002574>.
- [99] A. D. J. O’Neill and D. M. Gray. Solar radiation penetration through snow. In *The Role of Snow and Ice in Hydrology, Proceedings of the Banff Symposium, Int. Assoc. Hydrol. Sci*, volume 107, pages 227–240, 1972. URL https://research-groups.usask.ca/hydrology/documents/pubs/papers/oneill_gray_1973_2.pdf.
- [100] T. H. Painter, K. Rittger, C. McKenzie, P. Slaughter, R. E. Davis, and J. Dozier. Retrieval of subpixel snow covered area, grain size, and albedo from MODIS. *Remote Sensing of Environment*, 113(4):868–879, Apr. 2009. doi: 10.1016/j.rse.2009.01.001. URL <http://dx.doi.org/10.1016/j.rse.2009.01.001>.
- [101] T. H. Painter, A. C. Bryant, and S. M. Skiles. Radiative forcing by light absorbing impurities in snow from MODIS surface reflectance data. *Geophysical Research Letters*, 39(17), Sept. 2012. doi: 10.1029/2012gl052457. URL <http://dx.doi.org/10.1029/2012GL052457>.
- [102] T. H. Painter, D. F. Berisford, J. W. Boardman, K. J. Bormann, J. S. Deems, F. Gehrke, A. Hedrick, M. Joyce, R. Laidlaw, D. Marks, C. Mattmann, B. McGurk, P. Ramirez, M. Richardson, S. M. Skiles, F. C. Seidel, and A. Winstral. The Airborne Snow Observatory (ASO): 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, Oct. 2016. doi: 10.1016/j.rse.2016.06.018. URL <http://dx.doi.org/10.1016/j.rse.2016.06.018>.
- [103] R. T. Palomaki, K. Rittger, S. J. Lenard, E. Bair, J. Dozier, S. M. Skiles, and T. H. Painter. Assessment of methods for mapping snow albedo from MODIS. *Remote Sensing of Environment*, 326:114742, Aug. 2025. doi: 10.1016/j.rse.2025.114742. URL <http://dx.doi.org/10.1016/j.rse.2025.114742>.
- [104] C. Pike, D. Riley, H. Toal, and L. Burnham. Presenting a model to predict changing snow albedo for improving photovoltaic performance simulation. *Solar*, 4(3):422–439, Aug. 2024. doi: 10.3390/solar4030019. URL <http://dx.doi.org/10.3390/solar4030019>.

- [105] R. Pirazzini, L. Leppänen, G. Picard, J. I. Lopez-Moreno, C. Marty, G. Macelloni, A. Kontu, A. Von Lerber, C. M. Tanis, M. Schneebeli, P. De Rosnay, and A. N. Arslan. European in-situ snow measurements: Practices and purposes. *Sensors*, 18(7):2016, June 2018. doi: 10.3390/s18072016. URL <http://dx.doi.org/10.3390/s18072016>.
- [106] X. Qu and A. Hall. Assessing snow albedo feedback in simulated climate change. *Journal of Climate*, 19(11):2617 – 2630, 2006. doi: 10.1175/JCLI3750.1. URL <https://journals.ametsoc.org/view/journals/clim/19/11/jcli3750.1.xml>.
- [107] I. Rahimi, A. H. Gandomi, F. Chen, and E. Mezura-Montes. A review on constraint handling techniques for population-based algorithms: From single-objective to multi-objective optimization. *Archives of Computational Methods in Engineering*, 30:2181–2209, 2023. URL <https://doi.org/10.1007/s11831-022-09859-9>.
- [108] M. S. Raleigh, B. Livneh, K. Lapo, and J. D. Lundquist. How does availability of meteorological forcing data impact physically based snowpack simulations? *Journal of Hydrometeorology*, 17(1):99–120, Dec. 2015. doi: 10.1175/jhm-d-14-0235.1. URL <http://dx.doi.org/10.1175/JHM-D-14-0235.1>.
- [109] R. Ranzi and R. Rosso. *A physically based approach to modelling distributed snowmelt in a small alpine catchment*, volume 205, pages 141–150. IAHS, 07 1991. ISBN 0947571183. pdf available from author at https://www.researchgate.net/publication/255659250_A_physically_based_approach_to_modelling_distributed_snowmelt_in_a_small_alpine_catchment.
- [110] J. P. Riley, D. G. Chadwick, and K. O. Eggleston. Snowmelt simulation. *Reports, paper 122*, 1969. URL https://digitalcommons.usu.edu/water_rep/122/.
- [111] K. Rittger, T. H. Painter, and J. Dozier. Assessment of methods for mapping snow cover from MODIS. *Advances in Water Resources*, 51:367–380, Jan. 2013. doi: 10.1016/j.advwatres.2012.03.002. URL <http://dx.doi.org/10.1016/j.advwatres.2012.03.002>.
- [112] K. Rittger, M. S. Raleigh, J. Dozier, A. F. Hill, J. A. Lutz, and T. H. Painter. Canopy adjustment and improved cloud detection for remotely sensed snow cover mapping. *Water Resources Research*, 56(6), June 2020. doi: 10.1029/2019wr024914. URL <http://dx.doi.org/10.1029/2019WR024914>.
- [113] K. Rittger, M. Serreze, M. J. Brodzik, M. S. Raleigh, M. Skiles, K. Musselman, S. Lenard, T. H. Painter, J. Dozier, N. Bair, T. Stillinger, and “NSIDC Team”. 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*, College Park, MD, USA, May 2023. NASA Goddard Space Flight Center. URL https://modis.gsfc.nasa.gov/sci_team/meetings/202305/presentations/rittger.pdf. Held 1–4 May 2023 at The Hotel at The University of Maryland. Presentation slides available at: https://modis.gsfc.nasa.gov/sci_team/meetings/202305/presentations/rittger.pdf.
- [114] K. Rittger, S. J. P. Lenard, R. Palomaki, M. Brodzik, T. Stillinger, E. Bair, J. Dozier, and T. Painter. Historical MODIS/Terra L3 global daily 500m SIN grid snow cover, snow albedo, and snow physical properties, 2024. URL https://nsidc.org/data/STC_MODSCGDRF_HIST/versions/1.
 - [115] E. Roeckner, K. Arpe, L. Bengtsson, S. Brinkop, L. Dümenil, M. Esch, E. Kirk, F. Lunkeit, M. Ponater, B. Rockel, R. Sausen, U. Schleese, S. Schubert, and M. Windelband. Simulation of the present-day climate with the ECHAM model: Impact of model physics and resolution. *Max-Planck-Institut für Meteorologie*, 1992. doi: 10.17617/2.1852612. URL https://pure.mpg.de/pubman/item/item_1852612.
 - [116] M. B. Rohrer. *Die Schneedecke im schweizerischen Alpenraum und ihre Modellierung: Analyse langjähriger Schneemessungen und Modellierung des Wasseräquivalentes aufgrund zeitlich hoch aufgelöster operationell gemessener meteorologischer Daten*. PhD thesis, ETH Zurich, 1991. URL <http://hdl.handle.net/20.500.11850/140438>.
 - [117] A. Ryken, L. A. Bearup, J. L. Jefferson, P. Constantine, and R. M. Maxwell. Sensitivity and model reduction of simulated snow processes: Contrasting observational and parameter uncertainty to improve prediction. *Advances in Water Resources*, 135:103473, Jan. 2020. doi: 10.1016/j.advwatres.2019.103473. URL <http://dx.doi.org/10.1016/j.advwatres.2019.103473>.
 - [118] M. Saito, P. Yang, N. G. Loeb, and S. Kato. 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, May 2019. doi: 10.1175/jas-d-18-0308.1. URL <http://dx.doi.org/10.1175/JAS-D-18-0308.1>.
 - [119] C. Sarangi, Y. Qian, K. Rittger, K. J. Bormann, Y. Liu, H. Wang, H. Wan, G. Lin, and T. H. Painter. 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, May 2019. doi: 10.5194/acp-19-7105-2019. URL <http://dx.doi.org/10.5194/acp-19-7105-2019>.
 - [120] C. Schär, O. Fuhrer, A. Arteaga, N. Ban, C. Charpillot, S. Di Girolamo, L. Hentgen, T. Hoeffler, X. Lapillonne, D. Leutwyler, K. Osterried,

- D. Panosetti, S. Rüdüsühli, L. Schlemmer, T. C. Schulthess, M. Sprenger, S. Ubbiali, and H. Wernli. Kilometer-scale climate models: Prospects and challenges. *Bulletin of the American Meteorological Society*, 101(5): E567–E587, May 2020. doi: 10.1175/bams-d-18-0167.1. URL <http://dx.doi.org/10.1175/BAMS-D-18-0167.1>.
- [121] M. C. Serreze, M. P. Clark, R. L. Armstrong, D. A. McGinnis, and R. S. Pulwarty. Characteristics of the Western United States snowpack from snowpack telemetry (SNOTEL) data. *Water Resources Research*, 35(7):2145–2160, July 1999. doi: 10.1029/1999wr900090. URL <http://dx.doi.org/10.1029/1999WR900090>.
- [122] D. Shao, W. Xu, H. Li, J. Wang, and X. Hao. Modeling snow surface spectral reflectance in a land surface model targeting satellite remote sensing observations. *Remote Sensing*, 12(18):3101, Sept. 2020. doi: 10.3390/rs12183101. URL <http://dx.doi.org/10.3390/rs12183101>.
- [123] J. Shea, P. Wagon, W. Immerzeel, R. Biron, F. Brun, and F. Pellicciotti. A comparative high-altitude meteorological analysis from three catchments in the Nepalese Himalaya. *International Journal of Water Resources Development*, 31(2):174–200, Apr. 2015. doi: 10.1080/07900627.2015.1020417. URL <http://dx.doi.org/10.1080/07900627.2015.1020417>.
- [124] E. E. Stigter, J. F. Steiner, I. Koch, T. M. Saloranta, J. D. Kirkham, and W. W. Immerzeel. Energy and mass balance dynamics of the seasonal snowpack at two high-altitude sites in the Himalaya. *Cold Regions Science and Technology*, 183:103233, Mar. 2021. doi: 10.1016/j.coldregions.2021.103233. URL <http://dx.doi.org/10.1016/j.coldregions.2021.103233>.
- [125] T. Stillinger, K. Rittger, M. S. Raleigh, A. Michell, R. E. Davis, and E. H. Bair. Landsat, MODIS, and VIIRS snow cover mapping algorithm performance as validated by airborne lidar datasets. *The Cryosphere*, 17(2):567–590, Feb. 2023. doi: 10.5194/tc-17-567-2023. URL <http://dx.doi.org/10.5194/tc-17-567-2023>.
- [126] U. Strasser and T. Marke. 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, Nov. 2010. doi: 10.5194/gmd-3-643-2010. URL <http://dx.doi.org/10.5194/gmd-3-643-2010>.
- [127] J. Ström, J. Svensson, H. Honkanen, E. Asmi, N. B. Dkhar, S. Tayal, V. P. Sharma, R. Hooda, O. Meinander, M. Leppäranta, H.-W. Jacobi, H. Li-havainen, and A. Hyvärinen. 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), 2022. doi: 10.1525/elementa.2021.00118. URL <http://dx.doi.org/10.1525/elementa.2021.00118>.

- [128] C. W. Thackeray, X. Qu, and A. Hall. Why do models produce spread in snow albedo feedback? *Geophysical Research Letters*, 45(12):6223–6231, 2018. doi: <https://doi.org/10.1029/2018GL078493>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL078493>.
- [129] USDA. *Title 210-National Engineering Handbook, Part 622-Snow Survey and Water Supply Forecasting*, chapter Chapter 2: Data Parameters, pages 2–2 – 2–12. USDA, Washington, D.C., 2010. URL <https://directives.nrcs.usda.gov//sites/default/files2/1720456630/Chapter%202%20-%20Data%20Parameters.pdf>.
- [130] D. L. Verseghy. Class — a Canadian land surface scheme for GCMS. I. Soil model. *International Journal of Climatology*, 11(2):111–133, Mar. 1991. doi: 10.1002/joc.3370110202. URL <http://dx.doi.org/10.1002/joc.3370110202>.
- [131] A. Voordendag, M. Réveillet, S. MacDonell, and S. Lhermitte. 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, Sept. 2021. doi: 10.5194/tc-15-4241-2021. URL <http://dx.doi.org/10.5194/tc-15-4241-2021>.
- [132] W. Wang, K. Yang, L. Zhao, Z. Zheng, H. Lu, A. Mamtimin, B. Ding, X. Li, L. Zhao, H. Li, T. Che, and J. C. Moore. 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, Apr. 2020. doi: 10.1175/jhm-d-19-0193.1. URL <http://dx.doi.org/10.1175/JHM-D-19-0193.1>.
- [133] Z. Wang, C. B. Schaaf, M. J. Chopping, A. H. Strahler, J. Wang, M. O. Román, A. V. Rocha, C. E. Woodcock, and Y. Shuai. Evaluation of moderate-resolution imaging spectroradiometer (MODIS) snow albedo product (MCD43A) over tundra. *Remote Sensing of Environment*, 117:264–280, Feb. 2012. doi: 10.1016/j.rse.2011.10.002. URL <http://dx.doi.org/10.1016/j.rse.2011.10.002>.
- [134] S. G. Warren. Optical properties of snow. *Reviews of Geophysics*, 20(1): 67–89, Feb. 1982. doi: 10.1029/rg020i001p00067. URL <http://dx.doi.org/10.1029/RG020i001p00067>.
- [135] M. Warscher, T. Marke, E. Rottler, and U. Strasser. Operational and experimental snow observation systems in the upper Rofental: Data from 2017 to 2023. *Earth System Science Data*, 16(8):3579–3599, Aug. 2024. doi: 10.5194/essd-16-3579-2024. URL <http://dx.doi.org/10.5194/essd-16-3579-2024>.

- [136] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80, Dec. 1945. doi: 10.2307/3001968. URL <http://dx.doi.org/10.2307/3001968>.
- [137] S. N. Williamson, L. Copland, and D. S. Hik. The accuracy of satellite-derived albedo for northern alpine and glaciated land covers. *Polar Science*, 10(3):262–269, Sept. 2016. doi: 10.1016/j.polar.2016.06.006. URL <http://dx.doi.org/10.1016/j.polar.2016.06.006>.
- [138] W. J. Wiscombe and S. G. Warren. A model for the spectral albedo of snow. I: Pure snow. *Journal of the Atmospheric Sciences*, 37(12):2712–2733, Dec. 1980. doi: 10.1175/1520-0469(1980)037<2712:amftsa>2.0.co;2. URL [http://dx.doi.org/10.1175/1520-0469\(1980\)037<2712:AMFTSA>2.0.CO;2](http://dx.doi.org/10.1175/1520-0469(1980)037<2712:AMFTSA>2.0.CO;2).
- [139] Y. Xie, C. Pettersen, M. Flanner, and J. Shates. Ground-observed snow albedo changes during rain-on-snow events in Northern Alaska. *Journal of Geophysical Research: Atmospheres*, 129(16), Aug. 2024. doi: 10.1029/2024jd040975. URL <http://dx.doi.org/10.1029/2024JD040975>.
- [140] H. Yan, N. Sun, M. Wigmosta, R. Skaggs, Z. Hou, and R. Leung. Next-generation intensity-duration-frequency curves for hydrologic design in snow-dominated environments. *Water Resources Research*, 54(2):1093–1108, Feb. 2018. doi: 10.1002/2017wr021290. URL <http://dx.doi.org/10.1002/2017WR021290>.
- [141] F. Ye, Q. Cheng, W. Hao, D. Yu, C. Ma, D. Liang, and H. Shen. 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, Nov. 2023. doi: 10.1016/j.jag.2023.103519. URL <http://dx.doi.org/10.1016/j.jag.2023.103519>.
- [142] Y. Zhang and T. Ohata. Parameterization of albedo on snow/ice surface. A review. *Journal of the Japanese Society of Snow and Ice*, 65(1):33–51, 2003. doi: 10.5331/seppyo.65.33. URL <http://dx.doi.org/10.5331/seppyo.65.33>.
- [143] Z. Zhao, Y. Ding, M. Wang, and Y. Qu. 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, 2025. doi: 10.1109/tgrs.2025.3551094. URL <http://dx.doi.org/10.1109/TGRS.2025.3551094>.
- [144] B. Zhongyuan and T. Ohata. Variations of albedo on the glacier no. 1 at the headwater of Urumqi River, Tianshan Mountains, during the summer ablation period. *Journal of Glaciology and Geocryology*, 11(4):311–324, Nov. 1989. URL <https://doi.org/10.7522/j.issn.1000-0240.1989.0042>.

- [145] C. Zwart and C. Fierz. Significance of new snow properties for snow-cover development. In *Geophysical Research Abstracts*, volume 9, page 10287, 2007. pdf available from author at https://www.researchgate.net/publication/267513945_Significance_of_new_snow_properties_for_snow-cover_development. Formula available in SNOWPACK source code.

*Chapter 4***GLOBAL SNOW SIMULATIONS IN CLIMALAND:
DATA-DRIVEN AND PHYSICS-INFORMED
PARAMETERIZATIONS FOR DEPTH, ALBEDO, AND
SURFACE TEMPERATURE**

At the time of publication for this thesis, this chapter is being submitted to *Journal of Advances in Modeling Earth Systems*, from the American Geophysical Union. No digital article identifier or formal citation exists for this work at this time. The format and content of this article may exhibit slight differences from the version found in its affiliated version published in a journal.

References within this work to the developed “framework,” or citation 13 in this chapter all refer to the study detailed in Chapter 2. References to the neural albedo parameterization refer to the model developed in Chapter 3, and are indicated as such.

ABSTRACT

Snow exerts a strong control on land-atmosphere energy exchange and freshwater availability, yet it remains a major source of uncertainty in land and climate models due to challenges in modeling snow processes across multiple temporal and spatial scales. Here we present updates to the snow model in ClimaLand, a land surface model used in climate simulations and for weather prediction. Specifically, we incorporate physics-constrained neural network parameterizations for snow depth and albedo, with an equilibrium-gradient formulation for snow surface temperature that improves surface energy exchange. At point scales, the updated model reduces errors in snow depth and snow water equivalent and improves melt timing. At global scales forced with reanalysis data, the new parameterizations reduce widespread discrepancies in bulk snow properties while maintaining comparable skill across spatial and temporal metrics and evaluation scale. Snow albedo shows the largest improvement, with global bias reduced by nearly a factor of two compared to leading alternatives, leading to improved surface energy budgets. Similar to the site-level results, these changes translate into more realistic snow phenology, including earlier and more accurate timing of peak snowpack and melt-out. Performance remains comparable or improved to previous and established empirical schemes, while remaining spatial discrepancies are associated with environments underrepresented in the training data. The results demonstrate that lightweight and flexible physics-informed frameworks can provide accurate, efficient, and scalable alternatives to traditional approaches for large-scale snow modeling.

4.1 Introduction

Snow is a critical component of Earth’s climate system. Snow serves as an energy insulator and as a natural capacitor for water, modifying and delaying the mass transport between the atmosphere and underlying land surface. These effects drive complex feedbacks that impact climate at local and global scales [15]. Snow processes also govern many socioeconomic and ecological phenomena, including agricultural irrigation, hydropower generation, and natural disasters [31, 55]. Therefore, characterization and prediction of seasonal snow processes remain important for various natural and societal concerns, especially since snow cover changes with the evolving climate state [2, 31, 32].

Accurately modeling snow is not a straightforward task, as snow evolution encom-

passes various processes that span a range of spatiotemporal scales and underlying physical interactions. Snow density and depth influence the insulation properties of the snowpack [9, 36], while its high reflectivity drives strong radiative feedbacks that are sensitive to warming [11, 15, 37, 59]. Local modification of surface air temperatures by snow can propagate to the large-scale precipitation patterns that dictate snow cover extent, or couple large-scale dynamics to the local topographic complexity and aerosol/pollutant transport that affect snow radiative properties [8, 20]. Global energy balance and water mass distribution is thus susceptible to variability in snow accumulation and distribution, particularly in mid-latitude regions, where snow cover and local processes exhibit the highest variation [10, 37]. Uncertain inputs and imprecise formulations can propagate errors through simulations [60].

Limiting these errors in land models at global scales requires faithful representation of numerous processes where inputs might be unresolved, uncertain, or unavailable. The best choice of snow parameterizations for global climate models (GCMs) remains an open area of research, with myriad contenders and comparisons as data volume and computational access expand [38, 46, 54]. Model outputs vary considerably with parameterization selection [23, 55], yet the tradeoffs of computational complexity, input sensitivity, expressivity, and generalizability continue to limit the emergence of a clear winner for any given snow process [23, 29, 38].

Snow parameterization research for land modeling exhibits a wide spectrum of techniques, physical realism, and complexity, though several common structural and practical bottlenecks have emerged. Intermodel variability remains high, with albedo or density schemes driving large discrepancies in snow cover duration and radiative forcing, especially during the ablation period where observational agreement is weakest [15, 23]. These errors are intertwined with the spatial scales represented by the model. Parameterizations are often calibrated/evaluated over localized or more homogeneous regions, which then fail to generalize across resolutions, elevations, or locations, as unresolved heterogeneity in topography, precipitation, or pollutant concentration introduces scale-dependent biases [32, 33, 55]. Approaches calibrated at large scales, however, often fail to exhibit sufficient variability for higher-resolution accuracy or generalizability to evolving climate statistics [1, 24].

At the same time, schemes are often monolithic within specific land model implementations, or embedded within legacy languages/infrastructure [28, 48]. This can obfuscate the ability to add, isolate, or compare alternative formulations within a consistent framework. Rigid implementation can limit applicability in other models

run with different time-steps, hinder incorporation of newer approaches such as machine learning (ML), or limit their usage on parallelizable hardware like GPUs. These weaknesses throttle a model’s ability to take full advantage of contemporary computational advances [13, 18]. For both modular and more rigid land models, high dimensionality and correlation of overall parameter spaces can limit computational efficiency and hinder more comprehensive online calibration or uncertainty quantification, beyond the computational cost of already-expensive forward simulations and spin-up times [17, 40]. There is growing recognition that fine-scale processes, in both space and time, affect coarse-grained behavior, driving ever-higher resolution modeling [2, 15, 31, 37, 55, 65], yet many of these processes are not resolved in global climate simulations because both the data required to constrain them and the input variables required to drive them are not available. These challenges motivate the development of modeling frameworks designed for flexibility, scalability, and systematic experimentation.

ClimaLand, the land component of the Climate Modeling Alliance (“CliMA”) Earth system model, was released (version 1.0) in late 2025 [18]. It is built in a high-level language (Julia) to be modular and abstract, with ML library support for the easy development, incorporation, and experimentation of different subprocess parameterizations. ClimaLand is also designed to run natively on GPUs. This yields performance gains permitting global simulations without the need for supercomputers (e.g., on single machines), greatly facilitating online calibration. These design principles reflect the potential of an emerging hybrid paradigm where ML is used not as a replacement for physics, but as a complement, embedding flexible data-driven parameterizations directly from growing observational data volumes within the guardrails of a physical model infrastructure. This allows the nonlinear expressiveness of ML methods [16, 19, 26] to be leveraged while avoiding the inconsistencies of completely unconstrained approaches or the rigidity of formulaic assumptions. Parameterizations which maintain physical consistency yet also are able to model a range of outputs will be crucial for forecasting an evolving climate at ever-higher resolutions, as regional statistics and extremes shift over time, and it becomes necessary to generalize beyond fixed singular calibration regimes.

This article introduces updates to the ClimaLand Snow model, to catalog new parameterization developments (Section 4.3) and benchmark them (Section 4.4) over point-level site simulations and global spatial scales. We first introduce ML parameterizations for snow depth (4.2 “Snow Depth Parameterization”) and snow

albedo (4.2 “Snow Albedo Parameterization”) based upon a physics-constrained neural network framework first developed in 2022 and described in [13] that exhibit lightweight performance and can adapt to growing model needs as resolution increases. We then also introduce a lightweight surface temperature scheme (4.2 “Snow Surface Temperature Parameterization”) to enable stronger surface temperature variance. Simulation discrepancies are then evaluated at point and global scales (Section 4.4) for the updated variables as well as additional bulk and radiative properties. Snow phenology effects as a result of the parameter updates are explicitly investigated. Findings suggest that the choice of simple, yet flexible schemes instead of detailed formulations offer lightweight and competitive alternatives to traditional approaches. Outcomes are contextualized against those of the previous schemes and widely used alternatives [3, 21].

4.2 Model Description

The ClimaLand snow model (hereafter abbreviated as “CLS”) is a 1D bulk model tracking overall internal energy and water balance per column. CLS is modular and allows users to implement and compare custom parameterizations within the scope of the snow model API, without the need to touch model internal code. A more detailed description of the original model can be found in the original technical documentation (Section 2.2. of [18]).

Snow Depth Parameterization

The snow depth and density are required to determine heat fluxes through the snowpack, as well as the rate meltwater leaves the snowpack [18]. CLS always tracks snow water equivalent (S), which is related to snow depth (z) and bulk snow density ($\rho_s \leq \rho_w$) via liquid water density (ρ_w):

$$S = \left(\frac{\rho_s}{\rho_w} \right) z \quad (4.1)$$

which allows tracking of all three variables via a choice of a depth or a density parameterization.

Most prevalent snow models choose to parameterize ρ_s and derive z , with many employing the empirical formulation originating in Kojima [34] that was analytically extended by Anderson [3]. Small deviations exist amongst implementations, but most stay close to the original calibrated parameters, with downstream deviation arising from differing estimations of new-snowfall density [23, 55]. Despite its

central role in U.S. and European snow prediction systems, the parameterization is sensitive to the climate used for its calibration (with limited transferability to other climate types) and not robust to uncertainty in snow liquid content, a required input [6, 13]. CLS also tracks snow liquid content (S_l), enabling density diagnosis from the liquid fraction without additional state variables. CLS originally reduced calibration sensitivity to a single minimum-density parameter:

$$\rho_s = (1 - q_l) \rho_{\min} + q_l \rho_w, \quad q_l = \frac{S_l}{S}. \quad (4.2)$$

This parameterization has been replaced by an ML scheme, employing a parameterization framework developed in 2022 by Charbonneau et al. [13] that is applicable to various physical systems. The approach uses fixed matrices to enforce supplied constraints (analytical functions, other trainable structures, etc.) upon a model without the need to penalize the loss function during gradient-based training. This focuses parameter updates solely on conformant performance during constrained learning and guarantees compliant outputs in operation. The framework was shown to enable consistent and generalizable schemes from simple model structures, capable of competing with more complex models at minimal computational cost. It also exhibited faster and more accurate inference compared to clamping the outputs of the same structures trained without constraints, with better GPU scalability [13]. This parameterization evolves state variable z , according to

$$\frac{dz}{dt} = M_z \left(z, S, \bar{\varphi}, \bar{R}, \bar{v}, \bar{T}_{air}, \bar{P}_{snow} \right) \quad (4.3)$$

where M_z is a small constrained feed-forward neural network, and φ (0-1) is relative humidity, R (W/m^2) is received downward shortwave radiation, v (m/s) is surface wind speed, T_{air} ($^{\circ}\text{C}$) is the air temperature, and P_{snow} (m/s) is the liquid-equivalent rate of snowfall. The variable overbars reflect the scheme being trained offline with SNOTEL station daily averages (but instantaneous z, S), so a 24-hour moving average input for these features is approximated within CLS by an exponential moving average (EMA). One additional constraint is imposed to maintain $S_{new} \leq z + \Delta t \cdot dz/dt \leq S_{new} \rho_w / \rho_{min}$, with ρ_{min} picked as 50 kg m^{-3} and $S_{new} = S + \Delta t \cdot dS/dt$ once the prognostic update for S has been determined. Training data was compiled from different site locations so as to learn one set of coefficients that should apply independent of local climate. It was shown to exhibit lower error than the Anderson parameterization across numerous site simulations

both when prescribing observational S and when simulating input S (an S network for this scheme versus Anderson’s Snow17 model), and due to being 1D can be used in all CLS domain types [13].

This scheme was briefly specified in the initial ClimaLand report. However, the framework has been abstracted into a more performant module since its direct implementation, enabling further parameterization development within CLS (see 2.3) or adoption by interested researchers (see Data Availability). The scheme has been evaluated standalone at the site-level to date, though never within the land model and not at global scales, allowing first-time investigation of spatial and coupled performance in this work.

Snow Albedo Parameterization

Snow shortwave albedo (α) governs the fraction of incoming shortwave radiation reflected to the atmosphere, and together with soil and canopy albedo dictates the grid-cell reflectivity. Acting as a direct land-atmosphere coupling, its high values (70–90%) amplify sensitivity to small perturbations or uncertainty [8]. Biases in snow albedo have been linked to overestimates of snow depth and snow cover duration [55], and studies identify albedo parameterizations as among the most influential, yet under-characterized, controls on radiative fluxes during the snow ablative period [38, 56].

Despite its importance, a leading approach for snow albedo modeling in GCMs remains elusive. α depends on wavelength, pollutant concentration, snow age, and morphology [67], encompassing processes from many spatiotemporal scales. Complex models such as SNICAR/SNICAR-ADv3 within the Community Land Model (CLM; [40]) can provide precise spectrally-resolved estimates, but require inputs such as grain size/shape or pollutant concentration that are uncertain or unavailable at computationally tractable scales. This can yield errors comparable to simpler schemes at higher computational cost [24, 43]. Meanwhile, simple schemes rely on precise knowledge of snowfall timing and struggle with generalizability, especially in representing diverse decay rates with a single set of parameters [1, 8]. Consequently, modest differences in parameterization can produce large divergences in simulated snowpack properties [23, 30, 46, 55].

CLS initially offered a constant albedo scheme, or a diagnostic albedo dependent on solar zenith angle cosine (μ) and snow density as a fraction of ρ_w ($\tilde{\rho} = \rho_s/\rho_w$):

$$\alpha(\mu, \tilde{\rho}) = \beta(\tilde{\rho}) \cdot (a_0 + a_1 e^{-k\mu}), \quad \beta(\tilde{\rho}) = \min(1 - b_0(\tilde{\rho} - \tilde{\rho}_0), 1) \quad (4.4)$$

with calibrated parameters $a_0, a_1, k, b_0, \tilde{\rho}_0$, and scheme output being clamped between 0 and 1. Like the depth model, the snow albedo has been added as a prognostic state variable, and evolves according to a small constrained neural network based upon the aforementioned constrained model framework:

$$\frac{d\alpha}{dt} = M_\alpha(\alpha, T_{air}, P_{snow}, \mu_{LN}) \quad (4.5)$$

where all inputs are instantaneous save μ_{LN} , the zenith angle cosine at that day's local solar noon for the column (calculated via a broadcasted function over the domain, the same as μ). Whenever a snowpack disappears and a new snowpack appears ($S < 1\text{mm}$, $P_{snow} > 3\text{mm/d}$), α is reinitialized as a constant parameter that we set as 0.8. These inputs reflect training with daily albedo products [57] and deliberate choice of widely available inputs. Similar to the depth parameterization, the lack of explicit time or location dependence enables one set of weights to be learned that can be applied at any time/location. This developed ML scheme was shown to outperform over 20 existing or prevalent snow albedo schemes across hundreds of site evaluations at point-site and coarser 500m evaluation while maintaining physical consistency (Chapter 3), though like the depth parameterization, this study serves as an initial analysis of spatial performance at contemporary GCM resolutions when coupled to a larger model environment.

Snow Surface Temperature Parameterization

Heat exchanges between the snowpack and the atmosphere are quantified by the temperature gradient at the snow surface. While the whole snowpack has a high thermal inertia, the surface temperature can fluctuate on minute timescales [5], necessitating distinct modeling from the bulk for accurate accumulated energy flux into the snowpack over time.

This need is often invoked as justification for implementing multilayer snow models, as their additional degrees of freedom provide more range to express diurnal surface variability, clear distinction from the bulk and soil surface temperatures, and better representation of nonlinear gradients [55, 66]. However, multiple layers add complexity and computational load, requiring construction of a full vertical temperature profile even when only the surface slope governs heat flux. Implementation

also requires additional assumptions or inputs about the internal snowpack, many poorly constrained or unavailable in GCMs (nor widely observed), limiting their robustness and evaluability across climates [66]. A direct estimation of only the surface slope for a bulk model like CLS can bypass the computational bottleneck of a matrix system solve at each grid point, while improving surface flux resolution that ultimately governs melt timing and snow season duration.

The original CLS release lacked a surface temperature parameterization, using only the bulk temperature (T_s) in a rough approximation of surface heat flux. This was linked to a high bias in S and noticeably delayed melting compared to observations [18], an issue observed across many snow models [27, 45, 61, 62]. The updated version implements an “equilibrium-gradient” approach inspired by the Utah Energy Balance model (UEB; [66]), identifying the gradient that equilibrates the net surface fluxes. A gradient between the surface and bulk temperatures is quantified in accordance with Fourier’s Law via numerically solving a nonlinear equation for T^* :

$$SW_{\text{net}} + LW_{\text{net}}(T^*) + LHF(T^*) + SHF(T^*) = \frac{-\kappa (T^* - T_s)}{d} \quad (4.6)$$

where SW_{net} , LW_{net} , LHF , SHF represent the net shortwave flux, net longwave flux, latent heat flux, and sensible heat flux at the snow surface, respectively (all are positive if upwards). All except the shortwave radiative flux are nonlinear functions of T^* , calculated with the `Thermodynamics.jl` module [64]. If T^* is less than the freezing temperature (T_f), T^* is set as the surface temperature. If not, the surface temperature is set to the freezing temperature, and the leftover energy flux

$$F_{\text{surf}, \text{residual}} = \frac{-\kappa (T^* - T_f)}{d} \quad (4.7)$$

is added to the total energy flux for the bulk snow energy. The reference distance d is selected as

$$d = d_0 \left(1 - \exp \left(-\frac{z}{d_0} \right) \right), \quad d_0 = \sqrt{\frac{\kappa D}{\pi c_{p,i} \rho_s}} \quad (4.8)$$

where κ is snow thermal conductivity and $c_{p,i}$ is the specific heat capacity of ice. The right-hand-side of Eq. 4.6 and d_0 match the surface slope found from solving the diffusive heat transfer equation using periodic boundary conditions with a period ($D = 86400$ s) of 1 day [66]. This means d_0 represents the e-folding length of the

expected temperature-deviation quenching scale in a diurnally-varying snowpack. The exponential modification to d from d_0 ensures the reference depth remains inside the snowpack as the snowpack becomes small. Eq. 4.6 is numerically solved via the secant method, using a maximum of 8 iterations and a tolerance of 0.1 K, which was found to be sufficient in nearly all cases. In the few cases where the algorithm fails to converge, the bulk temperature is instead used as the surface temperature.

4.3 Methodology

Data

Point-location snow observations are necessary to constrain the timing and magnitude of accumulation and melt [32], but in-situ networks remain sparse, limiting their representativeness at GCM scales. Consequently, spatially resolved continental-scale snow cover must instead be evaluated using satellite retrievals and/or reanalysis products [37]. To permit comprehensive evaluation under both considerations, model evaluation was conducted across point-scale and global data to assess simulated snowpack evolution, spatial structure, and variability.

Atmospheric forcing data for global simulations were obtained from ERA5 reanalysis data at 1-degree resolution [12]; the forcing data was loaded at 3-hour intervals and then linearly interpolated to the simulation time. The simulation also prescribes leaf area index from MODIS [53](see section 4.3 of the initial release for a more comprehensive description of all model inputs, including static surface parameters). Model outputs were compared to those of ERA5-Land (“ERA5L”; [52]), which downscales the land surface model outputs of ERA5 (forced with ERA5 atmospheric reanalysis) to 9km resolution, plus altitudinal corrections (lapse-rate based) for air temperature, humidity, and pressure [22]. ERA5 and ERA5L have been found to perform well compared to other reanalyses, particularly in albedo consistency [37]. Grid cell albedo was derived from the ratio of net and downward shortwave radiation, as in Kouki et al. [37]. We note that ERA5 and ERA5-Land products are themselves model-constrained reanalyses and therefore do not constitute independent ground truth as a means of absolute valuation, with numerous studies outlining the structural biases in these sources (e.g., [33, 37]). However, this allows scheme comparison against a widely-used baseline to permit structural evaluation of discrepancies over more climate types than those used in model development.

At point scales, site-level data from Col de Porte (“CDP”; [42]) from 2008 to 2012

was used. In addition, the same analysis was carried out using data from the Senator Beck ("SNB") and Swamp Angel ("SWA") snow data collection sites [39], which were used in the same intercomparison project as Col de Porte [38, 47]. The same data preparation for CDP was used for SWA and SNB, converting precipitation mass fluxes into water-equivalent volume fluxes via division by ρ_w . For these simulations, in addition to z , S , and surface temperature (T_{surf}), ground albedo α_g was compared to the sensor-derived albedo, combining effects of snow cover fraction, snow albedo, and underlying soil albedo.

Experimental Design

Global simulations were run at $1^\circ \times 1^\circ$ spatial resolution, with 15 soil layers in the vertical direction. An implicit-explicit euler scheme (forward-backward Euler, or "ARS111"; [7]) with a time-step of 7.5 minutes was used to step the model, with three iterations per implicit solve. Global simulations were run from summer 2000 through summer 2020 to generate 20 total snowpack cycles in the northern hemisphere, with outputs saved at daily resolution. Single-site simulations were run from 2008-2012 with only the snow model standalone (only explicit time-stepping required) and a time-step of 1 hour, saving all outputs. For additional details on global simulations, we refer the reader to Sections 4.2.2, 4.3 in [18]).

Spin-up procedure for model initial conditions involved loading the same spin-up state as was generated for the initial ClimaLand report, which has zero flux at the bottom of the soil domain. Any existing snow is initialized with a liquid fraction of 0. This loaded state does not specify initial snow depth, density, or albedo due to the form of the original CLS parameterizations (see Section 4.3), so any initial snow in the spin-up file was assigned a density of 250 kg/m^3 ($z = 4S$) in accordance with observed settled snow densities [41]. The spin-up state initializes on March 1st 2000, with the model stepped until data collection starting summer 2000 when most snow had melted. For single-site simulations, all initial conditions were provided by the first observational values.

One simulation was run with the original CLS parameterizations, and another was run with all updated parameterizations. To further compare CLS in context with other snow modeling initiatives, an additional simulation was run using the prevalent Anderson density parameterization found in the Snow17 model [4], and the albedo scheme from HTESSEL [21] (parameterization equations are provided in Appendix 4.A). This albedo scheme is used in ECMWF and was the leading non-ML GCM

albedo scheme during albedo parameterization development (Chapter 3). This third simulation configuration also utilized the new snow temperature parameterization. All simulations otherwise used ClimaLand v1.7 defaults to isolate differences from only snow parameterizations. For site simulations, snow water content results are further contextualized against the summary statistics of the 27 snow models used in the Snow Model Intercomparison Project (ESM-SNOWMIP; [47]).

Previous studies acknowledge a dearth of evaluations explicitly assessing snow phenology and interannual variability within snow models, or correlations in bias [33]. Furthermore, equifinality oversight from assessing only the updated variables can mask underlying compensating biases (e.g., high snow albedo offset by low snow cover fraction) or internal deviations from otherwise similar outputs. Direct comparison of only updated variables is more acceptable in point-scale and short-term applications, but can amplify uncertainty over large domains and timescales [32]. Therefore, additional intermediaries such as radiative fluxes, snow cover, S , snow season duration, and annual snow maxima were also evaluated to mitigate these limitations. As improvements in individual state variables may reflect compensating structural errors, emphasis was placed on assessing larger patterns and consistency across snow mass, albedo, and phenology rather than isolated metric optimization.

Analysis / Metrics

For global outputs, daily timeseries per column were aggregated (mean) to monthly values, assessing root mean square error (RMSE), bias, and R^2 of monthly timeseries to analyze long-period climatology and statistical behavior rather than instantaneous variation. For site-simulation timeseries, a snowpack percentage error (SPE; the mean absolute error relative to the average nonzero value) was also assessed. For global-simulation albedo, root mean square percentage error (RMSE%) and bias percentage error (B%) were assessed alongside their absolute counterparts. For spatial analysis, output maps were compared to reanalysis with the Structural Similarity Index Measure (SSIM) metric and R^2 over space instead of time. We distinguish between grid- R^2 for one given time step with R_s^2 and timeseries- R^2 for one column with R_t^2 . For snow cover fraction, in addition to previously mentioned metrics, outputs per column were next denoted as snow/no-snow at each time-step with a threshold of 0.5, enabling assessment of binary classification metrics (i.e., scores derived from the confusion matrix of true positives, true negatives, false positives, and false negatives) in addition to regression scores, such as accuracy and critical success index (CSI). The significance of the difference between paired outputs for all

metrics (among different simulations) were assessed with a Wilcoxon signed-rank test [63] at a $p = 0.05$ significance level, a non-parametric method that avoids assumptions of normality or equal variance. All model metric differences were found to be statistically significant beyond the $p = 0.05$ significance level, so p -values are not reported here.

For z , S , snow cover, and snow phenology, outputs were only compared over columns where snow existed, excluding pixels with less than 1 mm of snow throughout the whole simulation. Columns identified as glaciers by the ERA5L glacial mask were also excluded, since CLS does not yet incorporate the conversion of condensed snow to glacial ice. Snow albedos were compared only where snow cover exceeded 95% to minimize contamination from non-snow signals. To assess phenology of seasonal snow, outputs were analyzed only over columns where at least 2 months of continuous no-snow cover (< 5 cm snow depth) could be identified. Snow presence in viable columns was counted only if the snow achieved a depth of at least 5 cm for at least two weeks. All other outputs are compared over all time over the entire land surface. For radiative fluxes, metrics are reported as integrated over the entire land area (area-weighted), while all other point properties are reported as a direct aggregation over the modeling grid.

4.4 Results

Site Simulations

Output statistics for the three site simulations are provided in Table 4.1, and outputs for CDP are displayed in Fig. 4.1. Outputs for SWA and SNB are provided in 4.B. Overall, the new parameterizations provide a competitive alternative to established parameterizations and represent an improvement over the old parameterizations. The delayed melting due to the older scheme and high bias in simulated S is visible for CDP and even more apparent for SWA and SNB. While the bulk temperature more closely estimates the observed snow surface temperature for SWA and SNB, the new scheme mitigates above-freezing surface temperatures, which were a notable issue in the initial release [18].

For site simulations, the HTESSSEL albedo scheme exhibits stronger high-frequency variability than the new albedo scheme, which instead tracks the seasonal statistics of albedo more than the sub-weekly or hourly variability. This corroborates the findings from testing in a standalone configuration (Chapter 3). Across all site-level variables, the new parameterizations exhibit slight advantage over the established

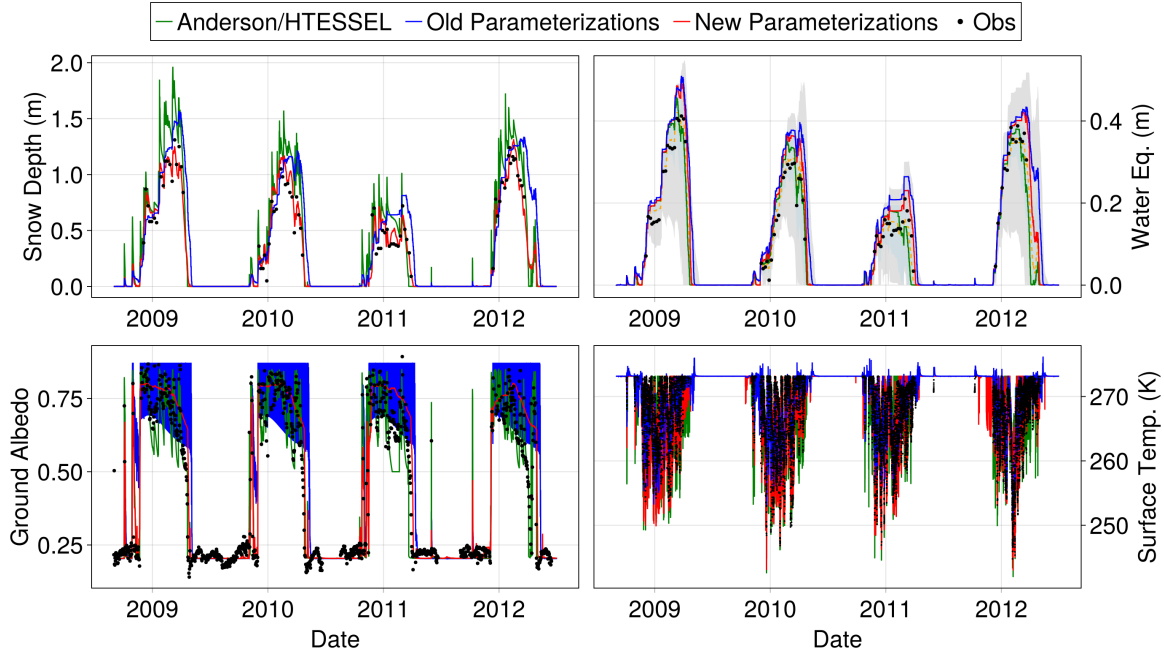


Figure 4.1: Single site simulation timeseries for the Col de Porte snow site. The above plot shows the results of the tested simulation configurations over site data. Each simulation corresponds to the curve color indicated in the legend, while observations are shown as black points. In the curve for Snow Water Equivalent (S), gray bands indicate the maximum/minimum simulation values by the 27 models in ESM-SNOWMIP, light blue bands indicate the 25th to 75th percentiles, while a dotted orange line indicates the median. The delayed melting and positive surface temperature phenomena from the old CLS parameterizations are apparent, while the new parameterizations mitigate these issues. The old albedo parameterization appears as a block due to the fast-varying nature of the zenith angle within a day in addition to its variance throughout the year.

Anderson/HTESSSEL (henceforth referred to as “A/H”) parameterizations in their ability to reduce relative and absolute error, though with slightly higher bias. These results also mimic previous analyses finding the new depth parameterization yields correlated z and S biases while the Anderson parameterization tends to more holistically under-predict S and over-predict z [13].

The Anderson parameterization more closely matches the median estimation of S among the ensemble of simulations in the ESM-SNOWMIP comparison, but this is less surprising as most of the models in this ensemble utilize a form of that parameterization. For these site simulations, the new neural parameterization is better at tracking the ensemble prediction during meltout, while the combination of the A/H parameterizations tends to melt a few days too early. At the site level, the largest benefit is sourced from the melt-timing improvement due to the improved surface temperature parameterization, and the neural snow depth parameterization offers an effective option in snow depth modeling.

Table 4.1: Site Performance Statistics. The RMSE, Bias, and SPE (the MAE divided by its mean nonzero value) are displayed for each site and each simulation (with the new parameterizations, old parameterizations, and established Anderson/HTESSSEL parameterizations labeled as “New,” “Old,” and “A/H,” respectively, with the best for each site highlighted in bold.

Var	Site	RMSE			Bias			SPE		
		New	Old	A/H	New	Old	A/H	New	Old	A/H
z	cdp	0.103	0.26	0.254	0.029	0.154	0.207	0.121	0.293	0.323
	swa	0.26	1.08	0.277	-0.152	0.874	0.119	0.133	0.597	0.154
	snb	0.605	0.601	0.529	-0.49	0.355	0.273	0.388	0.341	0.343
α_g	cdp	0.092	0.124	0.076	0.03	0.003	-0.014	0.136	0.181	0.112
	swa	0.131	0.444	0.101	0.034	0.341	-0.013	0.154	0.695	0.129
	snb	0.11	0.289	0.114	-0.003	0.169	-0.042	0.112	0.302	0.132
T_{surf}	cdp	3.303	4.259	3.599	-0.718	0.387	-0.68	0.009	0.012	0.01
	swa	7.141	6.402	8.635	5.835	-3.713	7.478	0.024	0.019	0.029
	snb	5.034	4.146	6.908	3.939	-1.078	5.947	0.017	0.01	0.023
S	cdp	0.052	0.072	0.04	0.044	0.062	0.01	0.21	0.296	0.152
	swa	0.102	0.359	0.135	-0.014	0.314	-0.11	0.133	0.604	0.211
	snb	0.13	0.17	0.168	0.039	0.106	-0.061	0.24	0.299	0.296

Global Simulations

Snow Depth and Water Equivalent

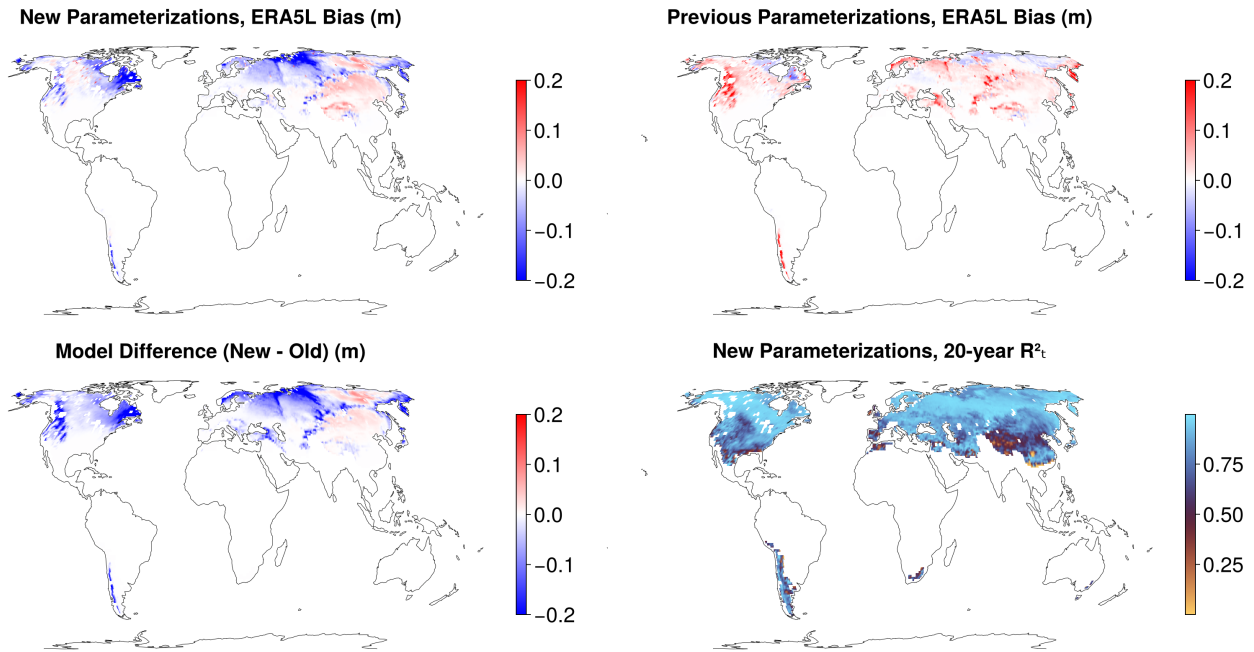


Figure 4.2: **z performance of updated parameterizations.** These plots show the bias against ERA5-Land data in snow depth for the new (top-left) versus old (top-right) parameterizations, as well as the difference between the new and old models (bottom-left), and the R^2 value per grid point over the entire 20-year simulation (bottom-right). Glacial pixels with perennial snow as well as no-snow pixels were excluded from this evaluation. The new parameterizations serve to reduce the one-sided bias of the original configuration, though this reduction over-corrects in more northern latitudes.

Fig. 4.2 and 4.3 outline model differences and performance from the new parameterizations compared to the original schemes, while Table 4.2 summarizes global simulation statistics. Spatial differences in bias arise from the difference in simulation forcing data at each location on account of varying climate types across the globe, as well as the trained parameterizations' abilities to generalize to these climates and their snowpack scenarios. The original CLS configuration exhibits global high positive bias in snow depth compared to ERA5L that is mitigated by the new parameterizations. This correction is stronger in the northern latitudes, resulting in a smaller but overall negative snow depth bias that is emphasized near the Arctic circle. However, the new configuration mitigates most bias in the Middle East and western North America. This is unsurprising, as despite the snow depth neural network exhibiting capable generalizability, the locations with more (less) bias align with the climate and snowpack types with less (more) representation in

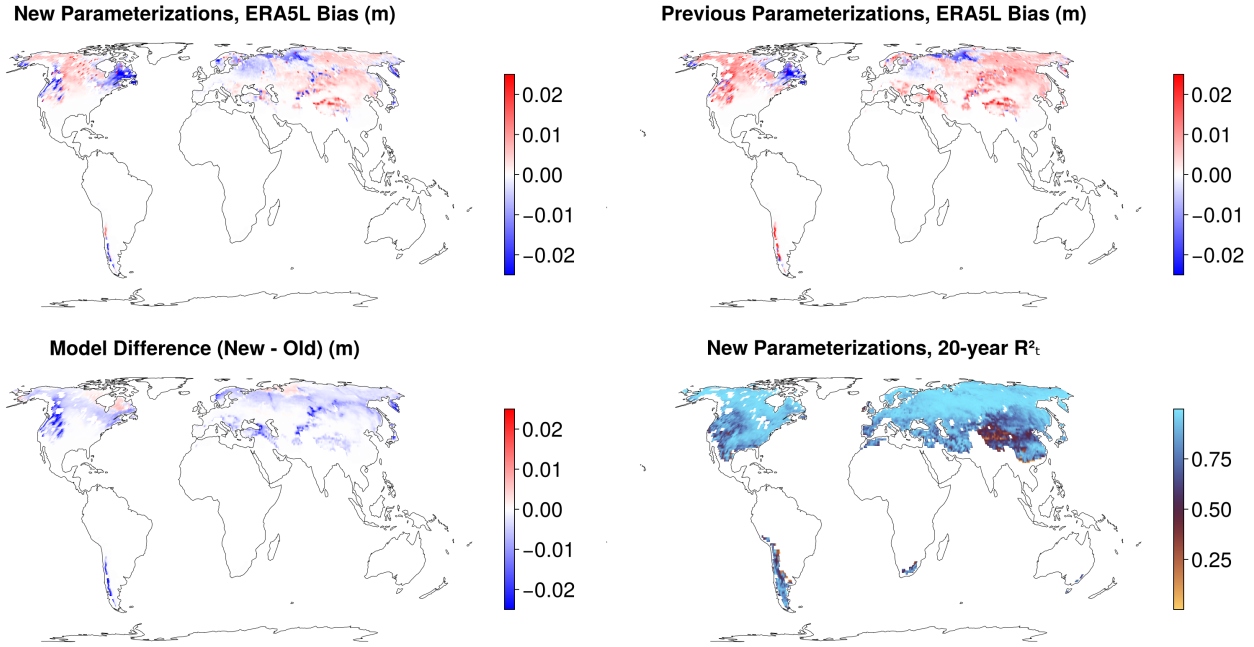


Figure 4.3: **S performance of updated parameterizations.** These plots follow the format outlined in Fig. 4.2, using the same pixel mask. For snow water equivalent, the bias reduction is more effective globally, reducing the absolute magnitude of the bias. Like snow depth and results, the temporal R^2_t values for snowy columns remain holistically high.

the developed training data [13]. For S , the correction is more effective at mitigating bias overall regardless of original direction, reducing the bias magnitude globally.

The new parameterizations show overall improved snow depth RMSE and R^2_t over each column, but slight regression in spatial aggregation metrics. For snow water equivalent, the absolute difference between new and old parameterizations is smaller but favors the new parameterizations more holistically. The medians for the new parameterizations are better than their means, suggesting the neural parameterizations are performing well globally but struggling to fully generalize to more extreme environments where less training data was available. However, they show a dramatic improvement in reducing the bias of total water mass accounting in the global snow system.

While all configurations exhibit capable performance, the A/H configuration performs slightly better in snow depth in most metrics except for bias, in a reversal from its ranking at the site level. However, the same configuration also shows worse performance with regards to representing the water mass budget of the snow climate system, with some of the worst bias and spatial representation metrics. Similar to the neural parameterizations, this could be an artifact of the limited climate regimes

Table 4.2: z and S performance statistics. The mean and median performance scores (see section 4.3 “Analysis”) over the global surface are reported for the three simulation configurations. For each metric, the best median and mean score are highlighted. The new parameterizations improve temporal scores such as RMSE and R_t^2 in snow depth compared to the old parameterizations, but exhibit slightly worse spatial (SSIM, R_s^2) scores. The new parameterization configuration also performs best with regard to snow water equivalent, though all configurations remain close and high-capability. The configuration utilizing the Anderson depth parameterization exhibits slightly better snow depth performance, yet exhibits worse overall S performance, with the worst performance in spatial reconstruction of S and worst bias in S .

		z				
Setup		RMSE	Bias	R_t^2	SSIM	R_s^2
New	mean	0.077	-0.033	0.838	0.831	0.854
	median	0.058	-6E-03	0.894	0.797	0.847
Old	mean	0.084	0.015	0.834	0.883	0.875
	median	0.072	0.002	0.87	0.891	0.891
A/H	mean	0.071	-0.04	0.904	0.898	0.917
	median	0.058	-0.026	0.933	0.874	0.931
		S				
Setup		RMSE	Bias	R_t^2	SSIM	R_s^2
New	mean	0.01	-0.001	0.875	0.955	0.903
	median	0.008	1E-05	0.945	0.963	0.948
Old	mean	0.011	0.002	0.874	0.955	0.928
	median	0.01	0.001	0.944	0.951	0.944
A/H	mean	0.011	-0.003	0.901	0.95	0.899
	median	0.007	-0.001	0.944	0.963	0.941

present in the original calibration data for this scheme.

Snow Coverage

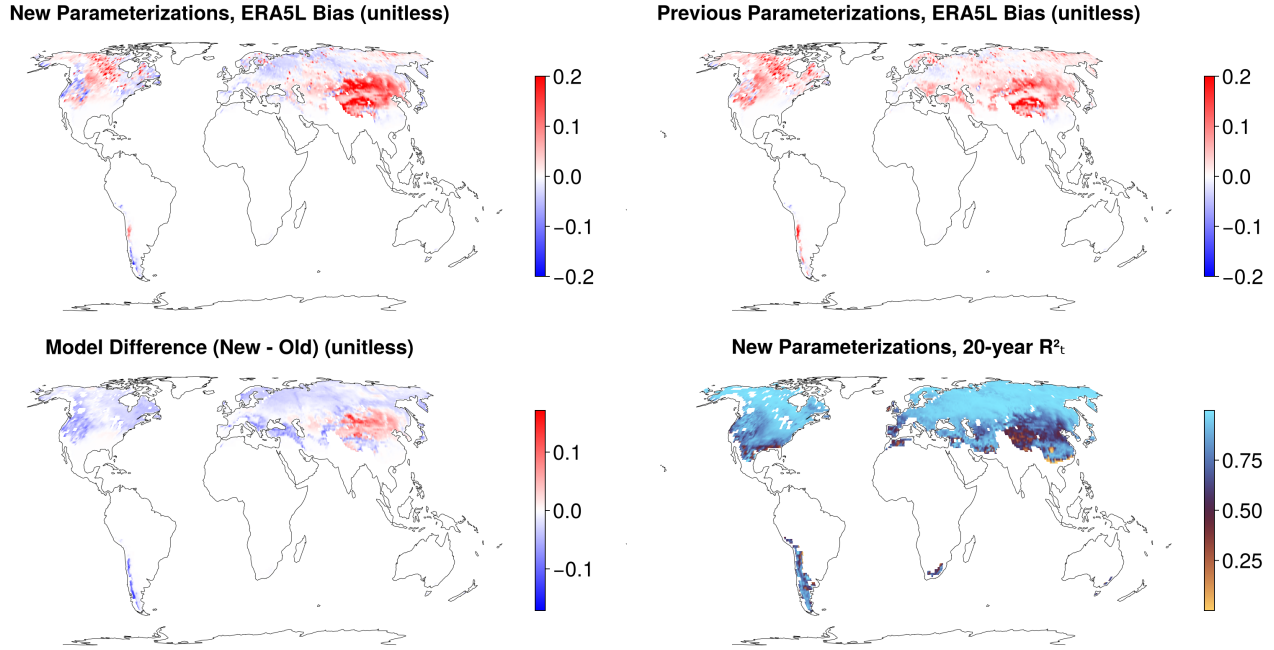


Figure 4.4: **Snow Cover performance of updated parameterizations.** These plots follow the format outlined in Fig. 4.2, using the same pixel mask. Like the snow depth results, one-sided biases are reduced by the new parameterizations.

Snow coverage results are analogous to those for snow depth, with overall reduction of positive bias across both hemispheres compared to the old parameterizations. Results of snow coverage simulations are displayed in Fig. 4.4 and summarized in Table 4.3. The original CLS configuration exhibited a positive bias in snow cover fraction against ERA5L over the northern hemisphere and Chilean Matorral which has been reduced globally by the new parameterizations, except for similar or increased bias over the Tibetan/Mongolian Plateaus. This overall reduction in bias leads to improved snow cover bias scores and precision in the new configuration, particularly in Canada and southern Europe, though it now exhibits slightly higher RMSE and other binary comparison metrics. These changes propagate to accumulated energy fluxes and dynamics in the climate model due to the relative gap between normal surface albedos and snow albedos.

For this variable, the established parameterizations outperform the CLS parameterizations. However, performance difference between all configurations for a given metric and aggregation is small, with nearly all within ~ 0.05 and many within ~ 0.02 . Even for the worst-performing configuration, the R^2_T scores for each column in the snow regimes of the northern hemisphere remain high. We note that the snow

Table 4.3: Snow Cover performance statistics. Table formatting follows that in Table 4.2, showing scores for snow depth and snow water equivalent. For all schemes, most performance is high and similar within ~ 0.05 , suggesting all would be effective within a GCM setting. Compared to the old parameterizations, the new parameterizations reduce the bias, though are slightly lower for other scores.

Setup		RMSE	Bias	R_t^2	SSIM	R_s^2	Accuracy	Precision	CSI
New	mean	0.095	0.011	0.868	0.74	0.869	0.955	0.904	0.793
	median	0.075	-2E-4	0.945	0.727	0.875	0.948	0.916	0.853
Old	mean	0.083	0.024	0.909	0.8	0.891	0.963	0.852	0.815
	median	0.064	0.01	0.965	0.805	0.901	0.958	0.91	0.888
A/H	mean	0.072	-0.007	0.934	0.811	0.925	0.972	0.945	0.841
	median	0.062	-0.009	0.963	0.779	0.944	0.971	0.972	0.881

cover fraction parameterization scheme among the simulations is identical (a simple asymptotic fractional threshold, $\propto z/(z + z_0)$), so these performance differences are contextualized not only against the depth schema differences within each simulation, but also largely against the sensitivity of the snow cover fraction scheme's calibration parameters, which have not yet been optimized holistically in CLS nor for each simulation's scheme choice. Work in ClimaLand remains ongoing to develop a more complex scheme for this subsystem to accompany the new parameterizations and reduce present alternative metrics.

Albedo

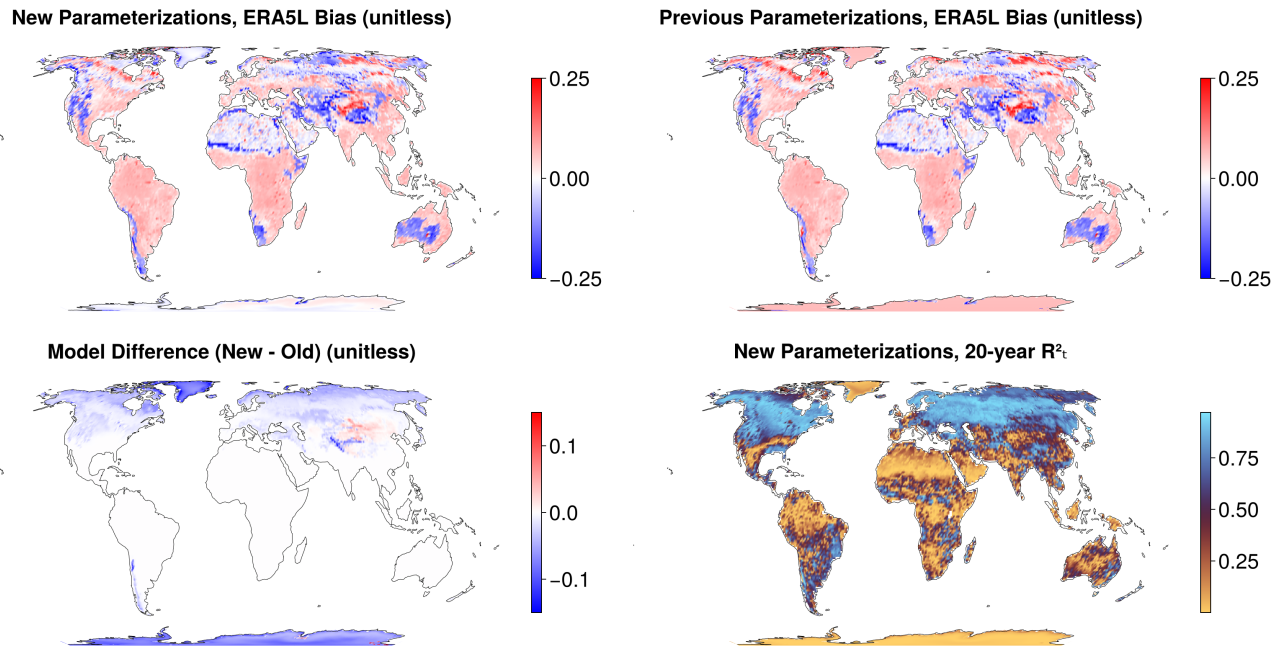


Figure 4.5: Surface Albedo performance of updated parameterizations. These plots follow the format outlined in Fig. 4.2, with the entire global surface analyzed. With only the snow parameterizations changed, it is unsurprising there is no difference in radiative fluxes over the columns without snow presence. The new parameterizations improve the overall model biases over Greenland and Antarctica, with some reduction at snowy latitudes due to the snow albedo reduction (see below). Other biases remain about the same globally.

Albedo outputs for the global surface albedo are displayed in Fig. 4.5, and results for surface and snow-only albedo are summarized in Table 4.4. Visualization of snow albedo changes are also provided in 4.B. The new parameterizations this time demonstrate a notable improvement in simulation fidelity compared to the alternative simulation configurations, dominating improvement or closely matching the top performer in most metrics. Compared to performance in the site-level hourly simulations, the tendency of the new albedo scheme to track expected seasonal tendencies reduces simulation bias across the globe by nearly a factor of two compared to other schemes, implying improved radiative energy flux simulation. The snow albedo RMSE and RMSE% scores of the new parameterization are also similar to those found during standalone development and benchmarking at site-level and 500m coarse data from a variety of locations, highlighting the strong extrapolative capability of the new scheme (Chapter 3).

There is a notable drop in temporal correlation of surface albedo over locations without snow. This could relate to prescribing static non-ERA5L soil parameters,

Table 4.4: **Surface and Snow-Only albedo statistics.** Table formatting follows that in Table 4.2. The new parameterizations show a dramatic improvement compared to the alternative simulation configurations for these variables, though the temporal correlation with ERA5L reanalysis data is low.

		α_{surf}						
Setup		RMSE	RMSE%	Bias	Bias%	R_t^2	SSIM	R_s^2
New	mean	0.078	0.281	-0.003	0.088	0.352	0.532	0.892
	median	0.055	0.245	-0.002	0.007	0.217	0.543	0.91
Old	mean	0.102	0.312	0.032	0.138	0.357	0.559	0.889
	median	0.1	0.246	0.062	0.079	0.229	0.566	0.916
A/H	mean	0.145	0.353	-0.082	-0.014	0.364	0.438	0.775
	median	0.119	0.317	-0.046	-0.062	0.228	0.438	0.819

		α_{snow}						
Setup		RMSE	RMSE%	Bias	Bias%	R_t^2	SSIM	R_s^2
New	mean	0.07	0.103	0.035	0.056	0.239	0.641	0.226
	median	0.05	0.062	0.011	0.014	0.155	0.65	0.203
Old	mean	0.152	0.21	0.129	0.178	0.286	0.658	0.36
	median	0.108	0.127	0.067	0.079	0.244	0.629	0.412
A/H	mean	0.148	0.182	-0.097	-0.109	0.553	0.471	0.152
	median	0.133	0.18	-0.082	-0.097	0.607	0.496	0.116

instead of dynamically modeling or prescribing ERA5L soil parameters (though soil moisture, energy, and subsequent impact on soil albedo are dynamically modeled). Compared to the A/H configuration, snow albedo simulation is greatly improved but exhibits worse temporal correlation scores among just-snow albedo values. This is likely an artifact of the mean-tracking behavior of the neural parameterization compared to other schemes that better attempt to represent higher-frequency variance in snow albedo timeseries.

Radiative Fluxes

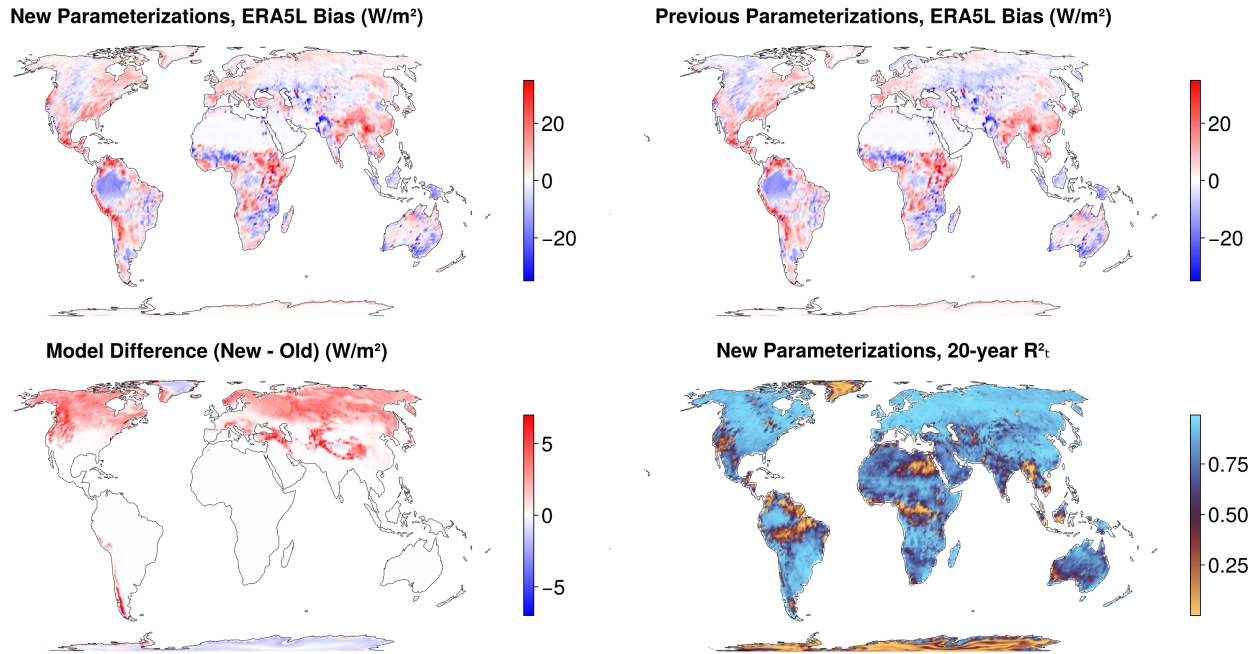


Figure 4.6: Latent Heat Flux performance of updated parameterizations. These plots follow the format outlined in Fig. 4.2, with the entire global surface analyzed. Like the surface albedo, changes are only discernible over snowy areas. The new parameterizations serve to reduce negative bias in the upper latitudes, though they exacerbate already-positive bias within non-glacial arctic columns.

Similar to the albedo results, the radiative latent and sensible heat fluxes over the global surface (Fig. 4.6 and Fig. 4.7, summarized in Table 4.5) show improvement in RMSE and bias between the new and the old parameterizations, with both outperforming the established A/H parameterizations. The opposite sign of latent and heat flux biases mimics findings in the original ClimaLand release [18] which can be attributed to remaining model structural error and yet-uncalibrated parameters. However, the new CLS updates lessen the magnitude of these biases, especially over glacial columns in Greenland and Antarctica. In ClimaLand, negative fluxes are downward (into the snow), so the new parameterizations show overall averaged increase in absorbed surface energy (sensible heat flux) and additional evaporation (upward latent heat flux). These are consistent with the overall average decrease in surface albedo, though we note temporally aggregating the configuration differences and overall biases can mask more nuanced interactions.

The different configurations overall show little difference regarding shortwave and longwave upward radiation (see Table 4.6, also visualized in 4.B) beyond worse

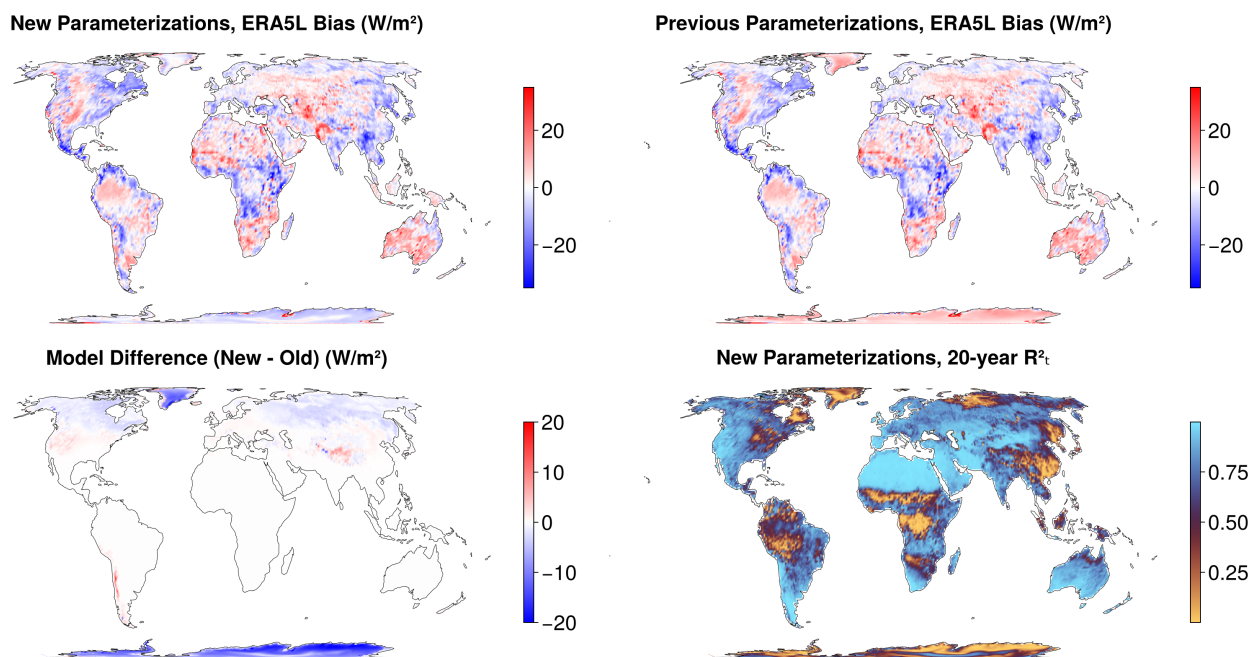


Figure 4.7: Sensible Heat Flux performance of updated parameterizations. These plots follow the format outlined in Fig. 4.2, with the entire global surface analyzed. Like Fig. 4.5, model changes are distinguishable only in columns with snow presence. Changes are mostly localized to Greenland and Antarctica, which serve to help reduce global bias statistics.

performance by the A/H configuration for shortwave, though the new parameterizations serve to reduce model bias over Greenland. Like the original ClimaLand release, the land surface is still too bright, though the older parameterizations remain a stronger predictor of longwave radiation. Radiative changes should not be qualitatively compared in a direct manner to snow changes due to different temporal aggregation (all times vs. periods with sufficient snow coverage) and the combined effects of the canopy, soil, or snow season duration changes.

Summary

The below table (Table 4.7) summarizes the leading performer over simulated variables with regards to RMSE, Bias, and SSIM. In most cases, the new parameterizations lead to the best improvement compared to ERA5-L or site observations, providing the leading or second-best score.

Table 4.5: Latent and Sensible Heat Flux statistics. Table formatting follows that in Table 4.2. Similar to albedo improvements, the new parameterizations tend to perform best with regard to the evaluated metrics and show holistic improvement compared to the old parameterizations, except for temporal correlation scores aggregated over columns. In these tables, the mean and median refer to direct aggregations over the modeling grid, while the “ $\int g$ ” statistic refers to the integrated average (area-weighted) over the grid. As SSIM and R_s^2 are one scalar per modeling grid and instead aggregated directly across time (while RMSE, bias, and R_t^2 are assessed per pixel and then aggregated across space), no area-weighted score exists for these values.

		LHF				
Setup		RMSE	Bias	R_t^2	SSIM	R_s^2
New	mean	9.49	0.935	0.583	0.434	0.884
	median	8.492	0.053	0.728	0.431	0.888
	$\int g$	12.614	0.98	0.697	-	-
Old	mean	9.773	0.35	0.672	0.413	0.88
	median	8.711	0.394	0.788	0.42	0.883
	$\int g$	12.716	0.282	0.722	-	-
A/H	mean	11.468	2.412	0.69	0.405	0.868
	median	9.642	2.836	0.784	0.413	0.866
	$\int g$	13.157	1.486	0.734	-	-
		SHF				
Setup		RMSE	Bias	R_t^2	SSIM	R_s^2
New	mean	13.867	-1.432	0.551	0.351	0.815
	median	12.164	-2.064	0.627	0.369	0.819
	$\int g$	14.286	-1.306	0.658	-	-
Old	mean	15.782	1.45	0.619	0.289	0.742
	median	13.194	2.54	0.708	0.307	0.75
	$\int g$	15.284	-0.11	0.671	-	-
A/H	mean	22.633	8.282	0.714	0.317	0.61
	median	15.757	4.301	0.797	0.328	0.77
	$\int g$	16.488	2.032	0.71	-	-

Table 4.6: **Upward shortwave and longwave radiation statistics.** Both CLS parameterizations improve over the A/H parameterization with regards to shortwave radiation, with slight improvements using the new parameterizations, however, the differences are minimal regarding longwave radiation and mostly similar.

		SW_{up}				
Setup		RMSE	Bias	R_t^2	SSIM	R_s^2
New	mean	11.044	2.642	0.877	0.653	0.934
	median	9.089	3.533	0.959	0.661	0.942
	$\int g$	10.38	1.72	0.822	-	-
Old	mean	12.133	-0.804	0.879	0.645	0.932
	median	10.827	-1.301	0.956	0.651	0.934
	$\int g$	10.689	0.429	0.826	-	-
A/H	mean	26.261	-11.785	0.892	0.581	0.891
	median	12.089	-2.311	0.955	0.575	0.905
	$\int g$	13.879	-3.525	0.85	-	-
		LW_{up}				
Setup		RMSE	Bias	R_t^2	SSIM	R_s^2
New	mean	7.166	-1.058	0.971	0.903	0.996
	median	6.064	-1.089	0.991	0.904	0.996
	$\int g$	6.864	0.539	0.966	-	-
Old	mean	6.27	-0.228	0.97	0.901	0.996
	median	5.073	-0.321	0.993	0.902	0.996
	$\int g$	6.345	1.02	0.965	-	-
A/H	mean	6.611	1.778	0.979	0.908	0.996
	median	6.166	2.193	0.994	0.908	0.997
	$\int g$	6.215	1.756	0.969	-	-

Table 4.7: Summary of simulation variable results. This table displays the best simulating variable with regards to RMSE, Bias, and SSIM for the discussed variables at the site or global level. The leading scheme is denoted by “N” for the new parameterizations, “O” for the old parameterizations, or “A/H” for the Anderson/HTESEL configuration. When the new parameterizations were not the leading configuration, the cell is highlighted if the new parameterizations were the second-best performer. In nearly all cases, the newly developed parameterizations represent an improvement over previous or prevalent configurations.

Variable	Site			Global		
	RMSE	Bias	SPE	RMSE	Bias	SSIM
z	N	N	N	A/H	O	A/H
SWE	N	N	N	N	N	N
Albedo	-	-	-	N	N	O
Snow Albedo	A/H	N	A/H	N	N	O
SW_{up}	-	-	-	N	O	N
LW_{up}	-	-	-	O	O	A/H
LHF	-	-	-	N	O	N
SHF	-	-	-	N	N	N

Phenology

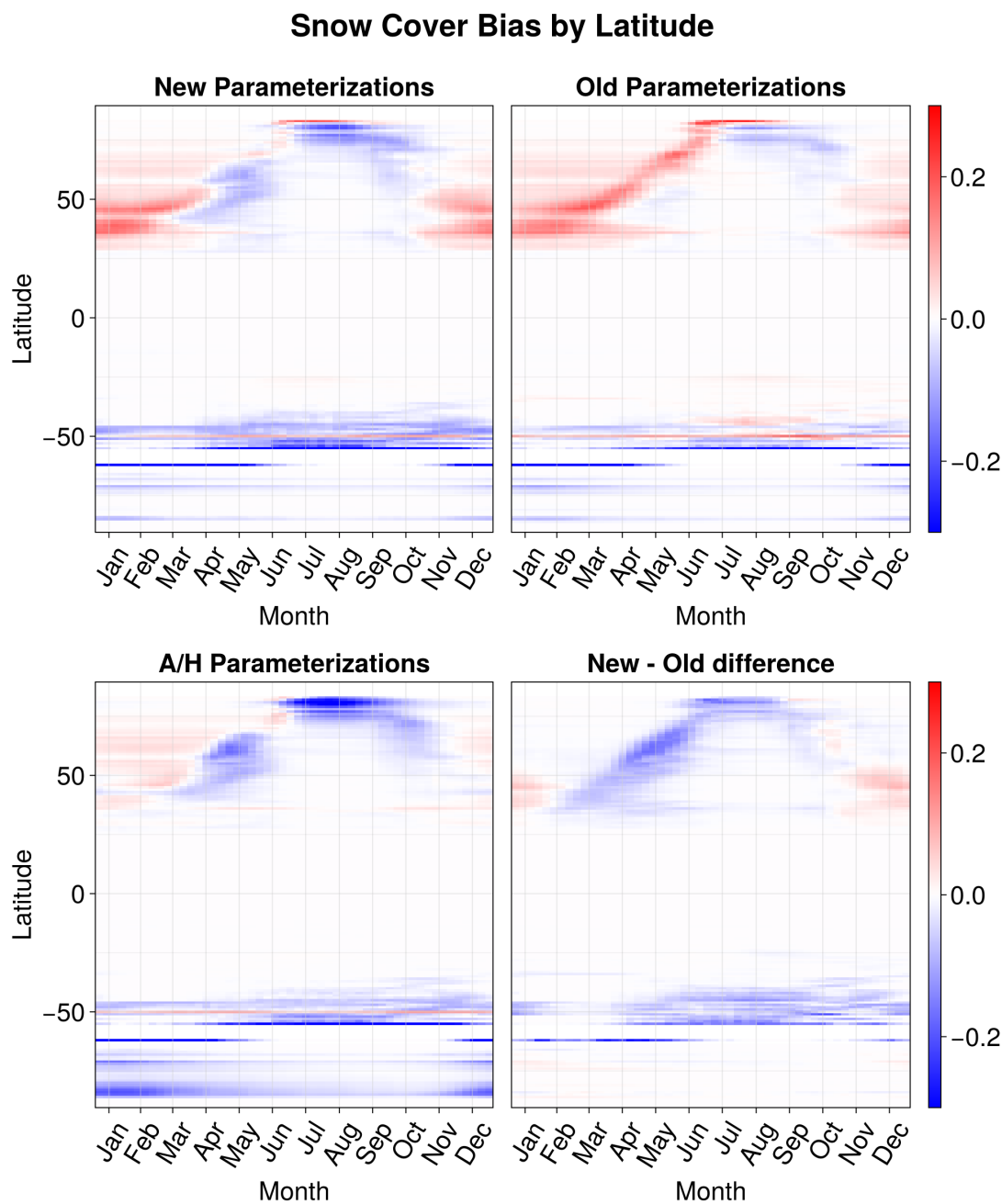


Figure 4.8: **Snow cover biases by latitude.** These plots show nonzero snow cover fractions averaged by month of year and across longitudes. The biases against ERA5L are shown, as well as the difference between the new and old parameterizations (bottom right). The negative difference at the start of the year in the Northern hemisphere and the later months in the Southern hemisphere indicates the new parameterizations are reducing the snow season globally, compared to the lag in the old parameterizations. Like global plots, glacial and no-snow columns were excluded from this analysis.

The timing of the snow season for different model configurations is summarized in Fig. 4.8-4.10 and Table 4.8. The onset of the snow season is similar in all simulation configurations (later months for northern latitudes in Fig. 4.8, or earlier months for southern latitudes), with minimal spread in snowpack appearance dates. In Fig. 4.8, snow cover fraction is unilaterally reduced in the southern hemisphere compared to the previous parameterizations, which does serve to over-correct snow coverage bias relative to ERA5L at the beginning and end of the snow season in the mid-southern latitudes, or at the end of the snow season in the northern hemisphere. However, the magnitude of the new negative bias is still less than that visible in the A/H configuration, especially at more extreme latitudes in both hemispheres. This reflects a shorter snow season duration unilaterally under the new parameterizations, mimicking the results identified for the site-level simulations. The earlier end to the snow season also corroborates the average improvements to radiative fluxes, since the discrepancy between snow and surface albedo in these transient phases is large.

Downstream differences in snow phenology primarily arise from mid- or late-season differences, rather than the beginning of the snow season. All simulations exhibited median snowpack appearance dates within 3 days of the data and from each other. However, simulation distinguishability increases for the season melt-out date (Fig. 4.9) as well as maximum snowpack date (Fig. 4.10), especially at northern latitudes. In particular, the dates of maximum snowpack height were significantly delayed with the old CLS parameterizations, which have been mitigated through the addition of the new parameterizations. Much of this improvement can be attributed to the added energy flux into the snow throughout the melting season due to the new surface temperature parameterization, speeding up the melting process.

The new parameterizations serve to reduce the median date of maximum snow depth by over 4 weeks compared to the old parameterizations, and simulate the melt-out date with reduced bias compared to the old parameterizations and to the A/H parameterization when aggregating over all latitude bands (Table 4.8). These are crucial for mitigating energy flux biases due to the land surface sensitivity to snow presence. We note the date of maximum snowpack height does not necessarily indicate the onset date of the melting season for a given latitude, particularly at higher latitudes where snowpacks can persist without significant depletion or compaction until later in the season.

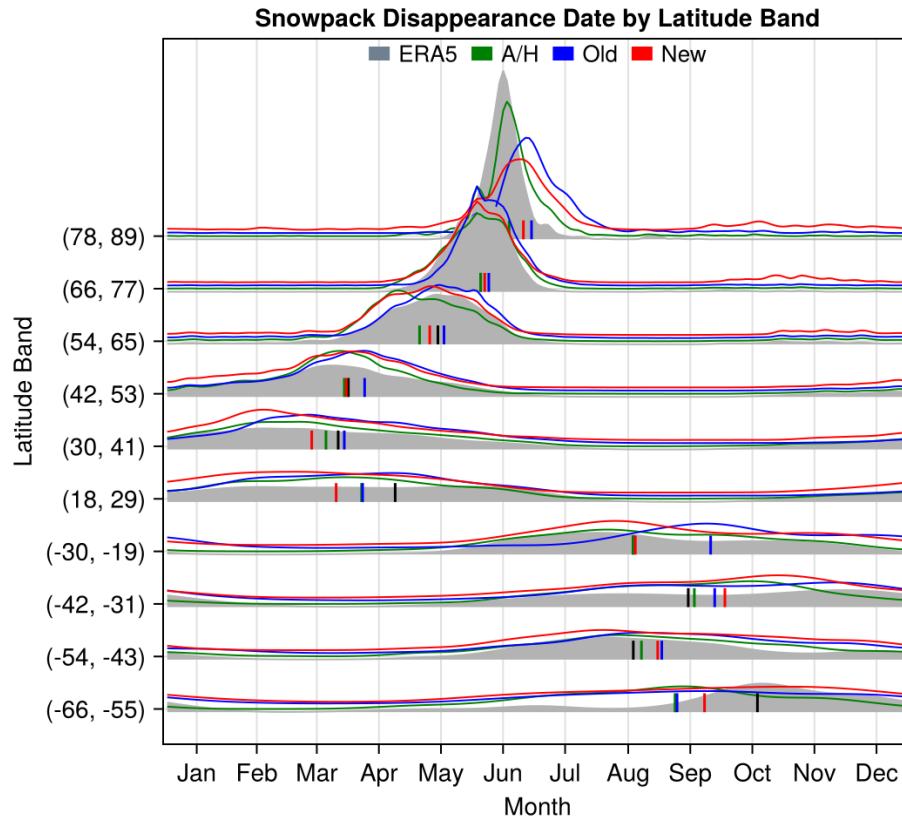


Figure 4.9: **Snowpack disappearance date by latitude band.** Within applicable latitude bands where seasonal snow exists, the distribution of snowpack disappearance dates in that band are visualized on the plot for all simulation configurations and ERA5L data. Distribution medians are also marked for each band in the same color as their corresponding curve as indicated by the legend. A small vertical shift from the 0-mass line for each distribution has been added to aid in distribution visualization. Configuration distinguishability increases at more northern latitudes, with some configurations performing better than others for any particular latitude band.

Table 4.8: **Snowpack timing differences between model configurations.** This table displays the average (in days) across latitude bands for the bias against ERA5L in snowpack onset date, disappearance date, and maximum depth date, as well as the difference between the old and new parameterization simulations. Biases are determined from the distribution medians seen in Fig. 4.9 and 4.10.

Date	New	Old	A/H	(New-Old) Diff
Onset	2.5	2	0.5	1
Disappearance	-0.5	6	-7	-5
Maximum	-7.5	23.25	4	-23.5

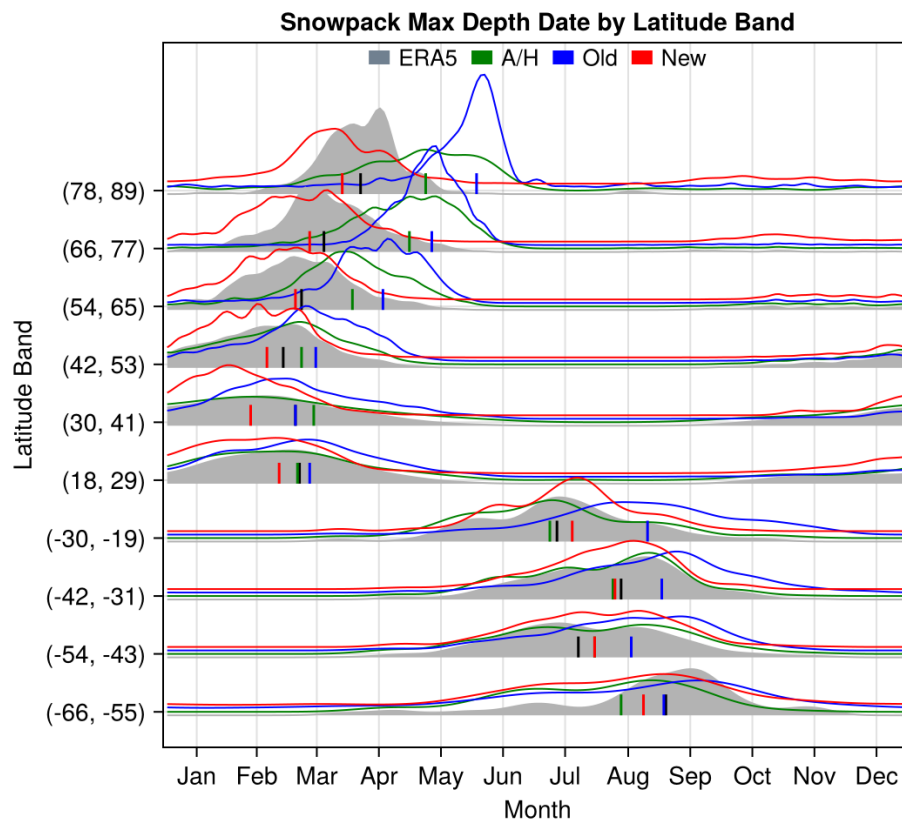


Figure 4.10: Date of maximum snowpack height by latitude band. Plot format follows that seen in Fig. 4.9, though now for the date of maximum snowpack height, which can influence melt-out date as well as the timing of snow-stored energy and mass in the global climate system. The delay of the old parameterizations compared to data and new parameterizations is apparent in this visualization.

Output Structure and Variability

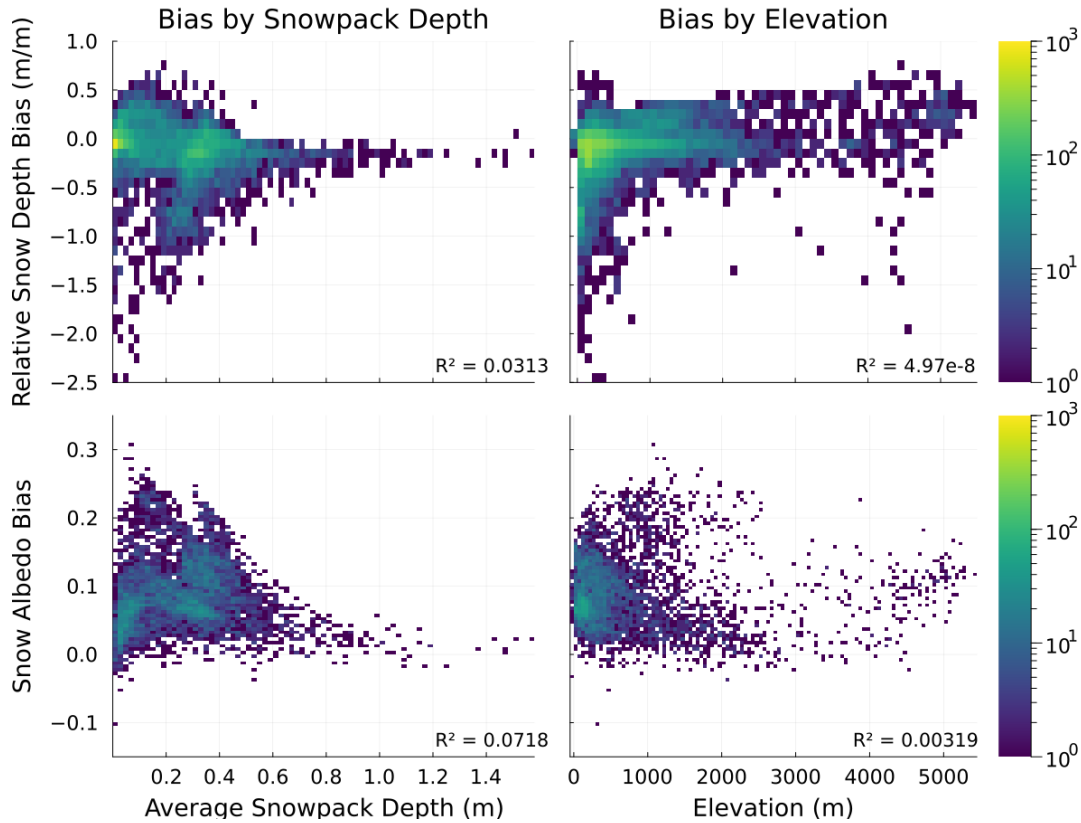


Figure 4.11: Snow Depth and Snow Albedo biases by elevation and average snowpack size. This plot shows the 20-year bias scores for each global column against that column’s elevation and average nonzero snowpack depth. Snow albedo bias (already unitless) is presented directly, while snow depth bias is presented relative to the average nonzero snowpack height. Glacial columns or those without snow are excluded to assess the parameterization capability in recreating seasonal snow.

Biases in the new snow depth and albedo parameterizations are contextualized against elevation and snowpack type (the average nonzero snow depth per column, across all years) in Fig. 4.11. Overall, both parameterizations show little correlation between output bias and elevation or snowpack type, suggesting the neural parameterizations extrapolate well to a range of elevations and snowpacks not present in their training data. The spread in biases is wider for low elevations and small snowpacks than high elevations and large snowpacks, though this is influenced by the relative abundance of low-elevation columns over the globe. Biases per snowpack type exhibit a slight positive tendency for small snowpacks and negative bias for small snowpacks — this regression to the mean is to be expected for small neural parameterizations that, through training, aim to best replicate the statistics of the underlying training data.

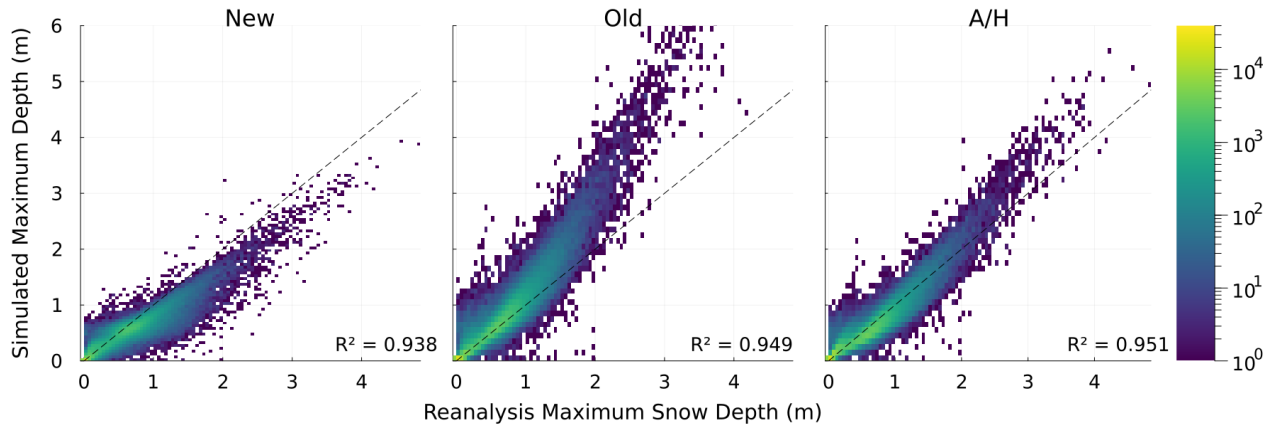


Figure 4.12: **Snowpack Maxima variability by simulation configuration.** The maximum snowpack height per year for each column are visualized against ERA5L maxima for each of the three assessed model configurations. All three configurations show good agreement with the ERA5L reanalysis data. Though showing a slightly lower R^2 score and more under-prediction, the new parameterizations do a better job at handling the largest snowpack maxima compared to other parameterizations that tend to overpredict maxima.

Parameterization ability to recreate maximal snowpack depth (and its inherent inter-annual variability) is compared by simulation configuration in Fig. 4.12. This plot shows all annual snowpack maxima (per year per column), so a wider spread around the 1:1 line (dashed) and a lower correlation indicates a configuration is less capable in recreating seasonal variability in snowpack maxima, in addition to portraying an over- or under-estimate of snowpack maxima. Again, all three schemes show a strong and similar ability to replicate maximal snowpack depth, though the new scheme tends to over-correct the positive snowpack bias exhibited by the old scheme and Anderson scheme. However, the new scheme does demonstrate a slightly stronger capability in recreating annual ERA5L maxima in larger snowpacks compared to the other two schemes, whose extremes bend further away from the 1:1 line.

Interannual variability (IAV) is compared more directly in Fig. 4.13. The standard deviation in the annual maxima of snow depth and snow albedo (aggregated and reported in Table 4.9) can be compared for each simulation configuration against those found in ERA5L, and the biases portrayed globally, indicating where each configuration exhibits too much or too little variance. The mean-tracking behavior of the neural network parameterizations manifests as an underperformance in northern latitude IAV and too much variability over more stable glacial regimes for snow albedo, though similar behavior is visible in all schemes. The new snow depth parameterization reverses and reduces the positive IAV bias from the older configuration, offering a competitive alternative to the A/H configuration with bet-

ter southern hemisphere performance. The single-sided positive bias in snow depth maxima IAV for the old configuration likely sources from insufficient melting due to the lack of a surface temperature parameterization, as well as poor representation of the range of new-snow densities with its density parameterization, which couples depth maxima IAV to the higher-IAV precipitation forcing data. Meanwhile, the single-sided negative bias in snow depth maxima IAV from the new parameterizations is more likely from structural under-representation in the neural model of other input features that could lead to better separability in different snowpack behavior, instead of regression to the mean within the training data, particularly in high latitudes where effects like lateral snow transfer and wind loading become more important. Sufficient snow albedo IAV over eastern Asia in areas like the Tibetan plateau remains a problem for all configurations, though this is a common issue across the snow modeling community due to the area’s anomalous environment. Aggregated globally (Table 4.9), the new snow albedo scheme best replicates the global variability.

Table 4.9: Global Interannual Variability (IAV) in snow depth and snow albedo by each simulation and ERA5L. This table summarizes the average variability (standard deviation) in maximum snowpack depth and snow albedo across all years over the modeling grid, compared to ERA5. Spatial assessment of these scores can be visualized in Fig. 4.13.

IAV	New	Old	A/H	ERA5L
z	0.0823	0.143	0.1212	0.1138
α_{snow}	0.0106	0.0055	0.0202	0.0114

In addition to spatial performance of temporal metrics, temporal performance of spatial metrics can be seen in Fig. 4.14, which visualizes the variation in SSIM scores by month for the new parameterizations, binned over the 20-year simulation. The simulation shows consistent performance within each month and larger spread throughout the year. The higher and narrower range in summer scores is likely due to the relative consistency of the land surface during the summer months, where remaining snow cover is more perpetual and homogeneous. For both variables, the performance at the onset of the Northern hemisphere snow season is higher than the performance at the end of the season, corroborating the phenology results and common trends in the snow modeling community. Overall, the inter-annual variability is well-matched by the new parameterizations, and the improvements in bias and performance by the new schemes do not show structural error, indicating

competitive long-term simulation capability by the new parameterizations in global climate modeling.

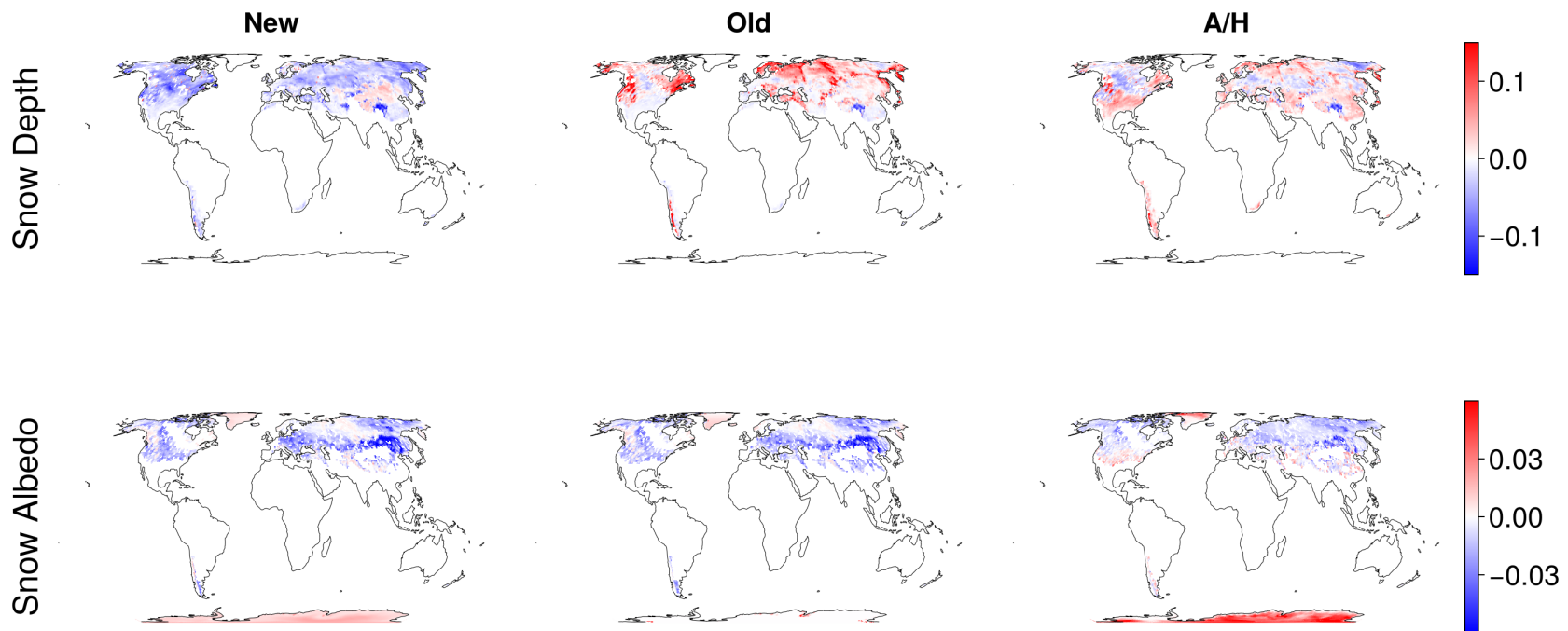


Figure 4.13: **Spatial bias in interannual variability (IAV)**. These plots visualize the difference in the standard deviation in the annual maximum values of snow depth and snow albedo between each configuration and the ERA5L data, as an indication of each configuration's capability to replicate the IAV in each variable across different climates and locations. A negative (blue) bias indicates a model exhibits less variability between years than ERA5L (more similar predictions each year), while a positive (red) bias indicates too much variability. Snow depth is analyzed over non-glacial columns with annual snow cover, while snow albedo is analyzed over all spatiotemporal points with > 95% snow cover.

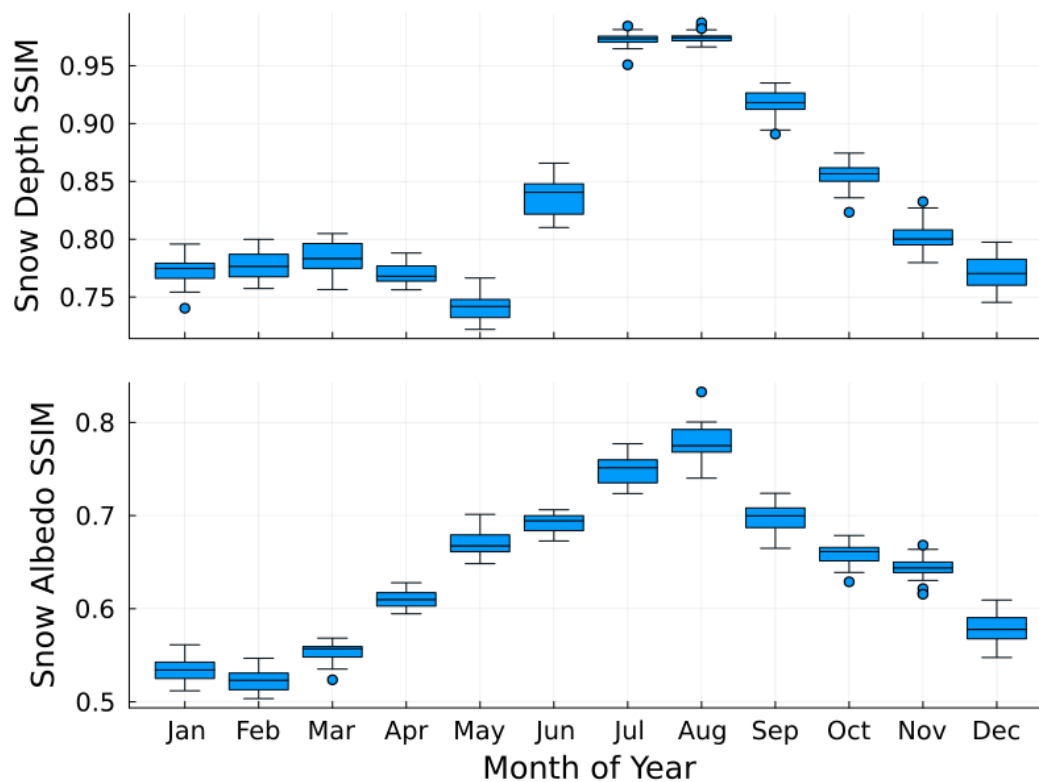


Figure 4.14: **Seasonal variability in spatial SSIM score for the new snow depth (top) and snow albedo (bottom) parameterizations.** As in Fig. 4.11 and Fig. 4.12, only non-glacial columns with annual snow cover were analyzed, which can over-inflate summer scores (where depth and albedo are more consistent globally). However, the new parameterizations show consistent performance within each month/season, with slightly better performance in the early snow season compared to the end of the snow season. Barplot central bars indicate the median, while boxes span the first and third quartiles. Points outside the whiskers are at least 1.5x the interquartile range from the median.

4.5 Discussion

The updated parameterizations constitute a notable advancement over their previous versions and demonstrate overall competitive performance in simulating snow properties, providing more relevant representations of the physics that govern terrestrial fluxes and snowpack state. These updates provide new defaults and optional functionality for future ClimaLand experimentation. We note the improvement in snow season duration from the surface temperature scheme only required a basic amount of physical intuition applied to a sufficiently versatile but computationally inexpensive expression, instead of the complexity of multiple snow layers, decoupling this property from conventional perceptions. The computational footprint of the neural parameterizations stays comparable to simple empirical schemes, enabling their use in large-scale simulations. This is in contrast to more detailed schemes or conventional ML model sizes. Both neural parameterizations showed similar accuracy at global scales when coupled to a model to that in their standalone local runs. This highlights the potential of the developed parameterization framework in creating schemes that can generalize in scale as well as climate diversity.

The error ranges seen here with the new parameterizations are on par with or competitive with those seen in other global models. Models such as CLM or the Exascale Land Model (ELM) frequently find snowpack depth errors on the order of 7-10+ cm and S bias near 1-2 cm, with snow albedo errors near and above 0.1 [33, 40]. Both models make use of more complex parameterizations such as SNICAR or additional input variables like aerosol concentration that can dramatically increase the computational cost. The new CLS schemes can replicate or reduce these errors, particularly in S and snow albedo, at reduced computational burden. The schemes are even competitive with the outputs of regional models such as WRF or AMUNDSEN run with higher-resolution and lower-uncertainty forcing data [32, 55]. Their architectural flexibility and their ability to perform at global scales when training only on site data bodes well for future research into the extension of ClimaLand/CLS to other modeling applications.

Notably, larger and more complex models (or our older schemes) struggle with melt-out timing with errors on the order of weeks [33, 58], while our new temperature scheme demonstrates statistical bias of snowpack phenology on the order of days. The sensitivity of climate models and the seasonal cryosphere to snow presence suggests our improvement in snow phenology will benefit the accumulated energy budget of long climate simulations. We further emphasize that a primary benefit of

the updated parameterizations lies in improved cross-scale consistency and reduced tuning fragility, rather than fixation on any single performance metric.

In complex coupled model systems, clear attribution of improvement or impairment remains challenging, as performance in any single variable or metric in isolation does not holistically determine scheme suitability. Apparent biases could reflect internal representation limitations, but could also arise from uncertainties in the forcing or evaluation data. Despite their roles as state-of-the-art reanalysis sources, ERA5 and ERA5-Land data are known to exhibit biases in temperature, precipitation, and subsequent snow properties that can vary with elevation [51, 55]. They are also known to overestimate S and melt timing, especially in mountainous regimes [37]. All of these could recontextualize interpretation of the reported CLS biases. At the same time, true observations are themselves limited. Sensor networks are sparse and often unrepresentative of complex terrain, and SNOTEL stations are typically placed on cleared, flat ground that might not capture substantial subgrid variability in snow-pack properties [33, 50]. Existing monitoring systems undersample the full range of elevations, land covers, and terrain classes present in snow-covered environments. Furthermore, comparisons between gridded data and point observations are inherently affected by strong spatial heterogeneity in snow properties [14, 25, 31]. These scale mismatches complicate model evaluation, highlighting the need for additional observational efforts or motivating efforts toward ever-higher-resolution modeling. Modern studies suggest that high-resolution simulations can better resolve meteorological and snow processes in mountainous regions than observation-constrained gridded products, which remain limited by measurement difficulty [44, 51]. With these considerations, the strong site-level performance demonstrated by the updated parameterizations suggests that these schemes are well-positioned to differentiate their performance and leverage increased GCM resolution, where improved process representation can translate more directly into large-scale skill.

Opportunities remain for improving snow process representation in CLS. Formal assessment of parameter uncertainty or online calibration was not pursued in this work in order to isolate the changes from behavioral updates alone. These parameterization updates demonstrate notable improvement, but leave room for additional complexity. The characteristic length scale for the surface temperature gradient is derived assuming homogeneous snow properties. This is acceptable for a bulk framework, though refinements (e.g., calibrated or parameterized scaling) could further improve flux magnitudes. Likewise, the present snow cover fraction parameter-

ization (not discussed here) or new-snowpack albedo adopt simplified formulations that remain areas of active development. In particular, equifinality considerations of the more basic snow cover fraction scheme could be impacting results (tightly coupling coverage error to depth error), and a stronger parameterization could further distinguish presented scheme performance. Calibrating this scheme for a particular configuration choice while investigating alternative formulations is a leading priority for the ClimaLand team. Additional processes such as wind-driven redistribution or canopy interception are not yet represented, reflecting our focus on computational efficiency and parsimony at this stage. More holistic benchmarking via ILAMB protocols also remains an ongoing effort.

Slight performance reduction for a subset of comparisons here likely reflects limited representation of extreme or sparsely-observed climates in the training data, rather than an instability in the scheme designs. Similarly, the presented approaches can embody accurate relationships in accordance with the expressive capacity of the chosen model class, but cannot eliminate structural errors arising from unavailable variables. Both ideas underscore the need for increased observation of a variety of snow properties in undersampled environments, and at scales unresolvable by present remote sensing capabilities [37, 49].

Our results highlight the potential of ML-based and physically-informed lightweight parameterizations as a complementary pathway for improving land surface modeling. By construction, these approaches offer increased flexibility and robustness to uncertainty and nonstationarity [16, 19, 26], while avoiding the rigidity of symbolic functional forms with even fewer parameters. They also provide favorable computational scaling relative to more complex process-based schemes, enabling efficient integration into large-scale models. Within ClimaLand, the underlying model engine is designed to support modularity and extensibility, allowing new parameterizations to be incorporated, adapted, and evaluated as observational constraints evolve. The constrained neural model framework also facilitates broader applicability beyond snow processes, opening the door to alternative representations across land surface components and related physical systems. As these tools continue to mature, we look forward to the opportunities that arise from community engagement. Users and developers are encouraged to contribute feedback, report bugs, or build additional parameterizations to help advance the next generation of land surface modeling.

4.6 Conclusion

This study introduced and evaluated updated snow parameterizations within the ClimaLand snow model, including first-time spatial benchmarking of ML-based schemes for snow depth and albedo as well as an equilibrium-gradient surface temperature formulation. Across site-level and global simulations, the updated configurations demonstrated improved representation of snowpack evolution while maintaining competitive computational performance. Notably, improvements in melt timing were achieved without the added complexity of multilayer snow schemes, and the ML-based parameterizations exhibited capable generalization across climates and spatial scales relative to established approaches.

The modular yet constrained ML framework enables stable cross-scale snow parameterizations with competitive physical behavior while also offering reduced structural rigidity. Within the architecture of ClimaLand, these updates demonstrate that data-driven parameterizations can be robustly integrated into physically based land models and can yield measurable improvements in cryospheric state variables and coupled surface energy fluxes. Our results imply that the choice of simple, constrained, yet flexible structures, instead of detailed but strict formulations, can serve as effective and efficient alternatives to traditional empirical or highly complex process-based schemes in large-scale land models. The modular structure of ClimaLand further enables systematic testing and integration of such approaches, supporting continued model development as new data and higher-resolution applications become available.

4.A Additional Parameterization Equations

Anderson Density Parameterization

This scheme is often formulated as a prognostic equation for (unitless; fraction of ρ_w) snow density, incorporating effects for temperature-dependent compaction and microstructure metamorphism

$$\frac{1}{\rho} \frac{d\rho}{dt} = W_{above} c_1 \exp(c_2 (T_s - T_f) - c_3 \rho) + c_4 \exp(c_5 (T_s - T_f) - c_6 \max(0, \rho - \rho_d)) \quad (4.9)$$

with calibrated coefficients c_i and ρ_d , T_s is the bulk snow temperature, and W_{above} the snow water equivalent or weight of *overlying* snow. However, this formulation is mostly associated with multilayer schemes, where W_{above} is readily determined and density evolves independently in each layer. This is less applicable to a bulk

snow model where the single snow layer constitutes its own overlying snow, and the average density over the whole snowpack is desired. As is discussed in Koren et al. [35], the right-hand side of Eq. 4.9 changes sufficiently slowly within a model time-step (usually on the order of minutes) that it can be treated as a constant, permitting a solution that can be integrated over the snowpack height to get a parameterization for the bulk density evolution. This bulk form of the parameterization is the type implemented in Snow17 [4], with the formulation:

$$\rho(t + \Delta t) = \rho(t) \left(\frac{e^A (e^{B \cdot S_i} - 1)}{B \cdot S_i} \right) \quad (4.10)$$

$$B = c_1 \Delta t \exp(c_2 (T_s - T_f) - c_3 \rho) \quad (4.11)$$

$$A = c_4 \Delta t \exp(c_5 (T_s - T_f) - c_6 \max(0, \rho - \rho_d)) \quad (4.12)$$

with updated densities clamped between a provided minimum (ρ_{min}) and maximum (ρ_{max}) density. $S_i = (1 - q_l)S$ is the ice content of the snowpack. This compaction scheme is utilized twice; once to determine the depth contribution from new snowfall once it falls and compacts itself on top of the snowpack (this requires another parameterization for the density of new snowfall, which is given in Snow17 as $\rho_{ns} = \rho_{min} + a(T_{air} - T_0)^b$), and once to determine the depth change when the whole snowpack compacts. These two are combined into one depth change that can be divided by the time-step to yield a prognostic update for snow depth. Table 4.10 lists typical values for the scheme parameters.

Table 4.10: Parameters for the Bulk Snowpack Anderson Parameterization. Units for temperature difference are given as K^{-1} since celsius difference and kelvin difference are identical. These parameters are utilized in many GCMs where the Fortran implementation of this scheme is found.

Parameter	Value	Parameter	Value
c_1	$0.01 \text{ hr}^{-1} \text{ cm}^{-1}$	ρ_d	0.2
c_2	0.08 K^{-1}	ρ_{min}	0.05
c_3	$21 \text{ cm}^3/\text{g}$	ρ_{max}	0.45
c_4	0.01 hr^{-1} if $q_l = 0$, otherwise 0.02 hr^{-1}	a	$1.7\text{e-}3$
c_5	0.04 K^{-1}	T_0	$-15 \text{ }^\circ\text{C}$
c_6	46	b	1.5

HTESSEL Snow Albedo Parameterization

This simplistic snow albedo scheme uses a continuous precipitation-based albedo reset as well as either a linear or exponential albedo decay, dictated by the bulk snow temperature or the presence of melt (negative dS/dt). The update rules are as follows:

- $\alpha(t+1) = \alpha(t) + \min\left(1, \frac{P_{snow}\Delta t}{Q_0}\right) (\alpha_{\max} - \alpha(t))$, when $P_{snow} > 0$
- $\alpha(t+1) = \alpha_{\min} + (\alpha(t) - \alpha_{\min}) e^{-k_1 \Delta t}$, when $P_{snow} = 0$ and $dS/dt < 0$ or $T_s > -2^\circ\text{C}$
- $\alpha(t+1) = \alpha(t) - k_2 \Delta t$, otherwise.

Adaptation of this scheme to a prognostic format is straightforward. Table 4.11 lists typical values for the scheme parameters.

Table 4.11: Parameters for HTESSEL Albedo parameterization.

Parameter	Value
$\alpha_{\min}$	0.5
$\alpha_{\max}$	0.85
k_1	0.24 d^{-1}
k_2	0.008 d^{-1}
Q_0	1 cm

4.B Additional Plots

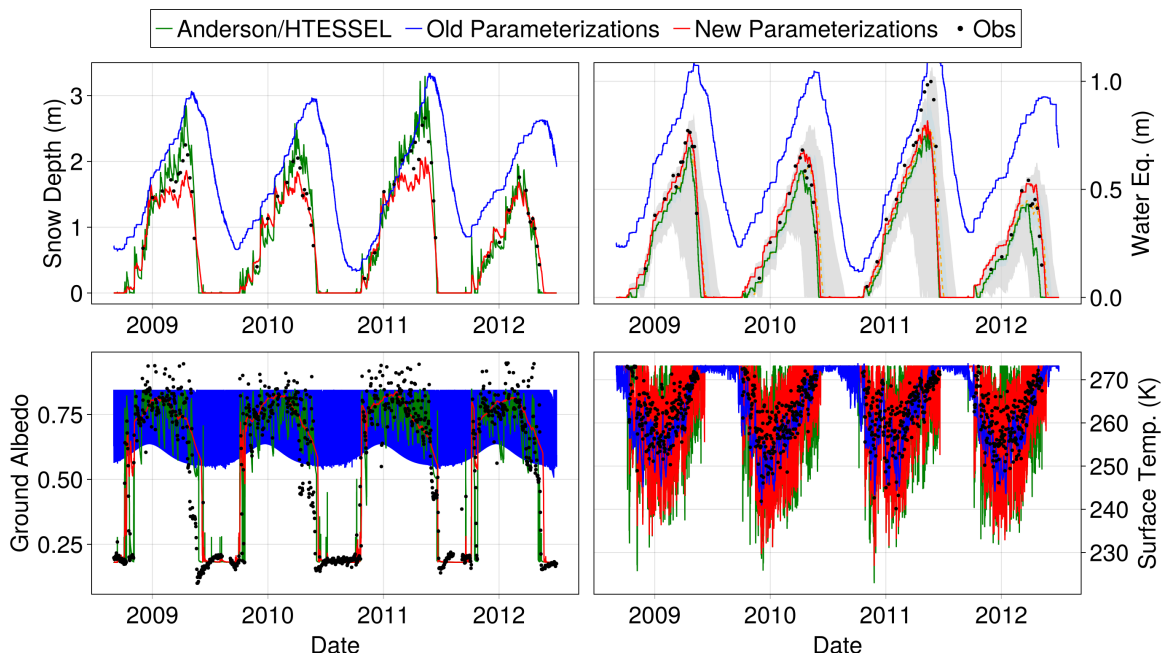


Figure 4.15: **Single site simulation timeseries for the Swamp Angel snow site.** Formatting follows that in Fig. 4.1. In this simulation, the melting is sufficiently delayed that the snowpack never disappears for the old parameterizations.

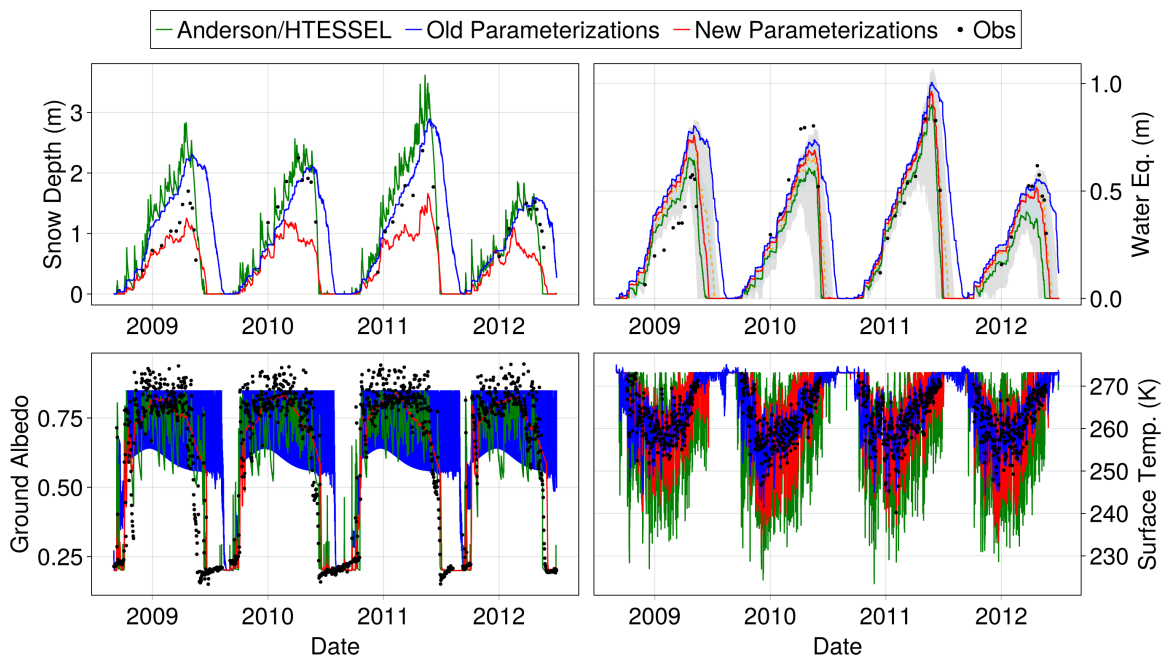


Figure 4.16: **Single site simulation timeseries for the Senator Beck snow site.** Formatting follows that in Fig. 4.1. For this site, all simulations do not perform as well with regards to snow depth. The new parameterizations tend to under-predict depth, while the A/H parameterization tends to over-predict depth

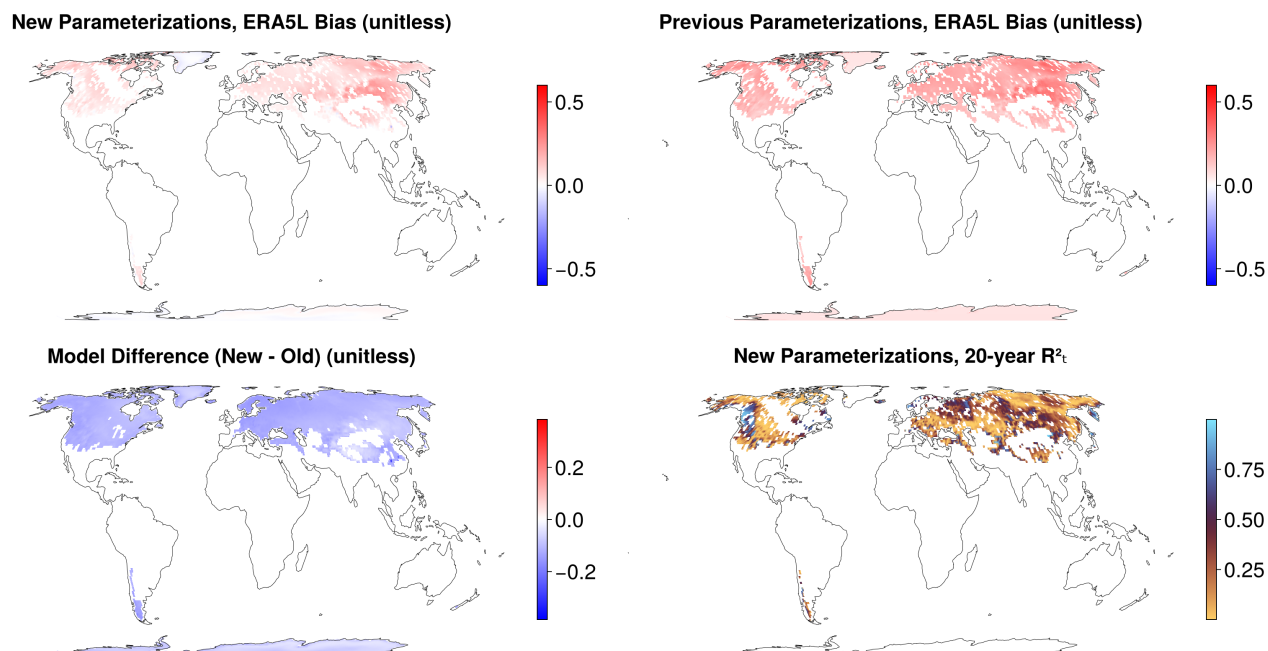


Figure 4.17: **Snow Albedo performance of updated parameterizations.** These plots follow the format outlined in Fig. 4.2, with masking over all columns and times where snow cover fraction is less than 95%. For specifically snow albedo, the reduction in bias magnitude is obvious with little over-correction.

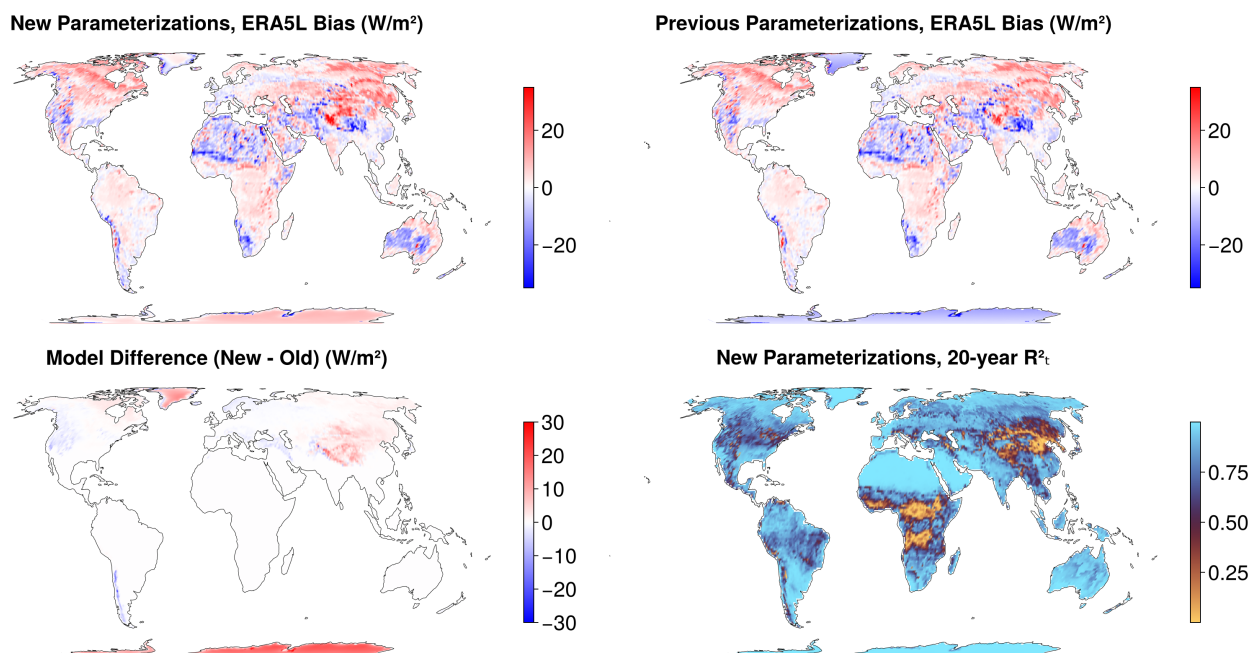


Figure 4.18: **Upward shortwave performance of updated parameterizations.** These plots follow the format outlined in Fig. 4.2, with the entire global surface analyzed. The new parameterizations mitigated bias in Greenland and the western United States, but also over-corrected the bias over glacial areas such as Antarctica, and increased bias over Eastern Asia.

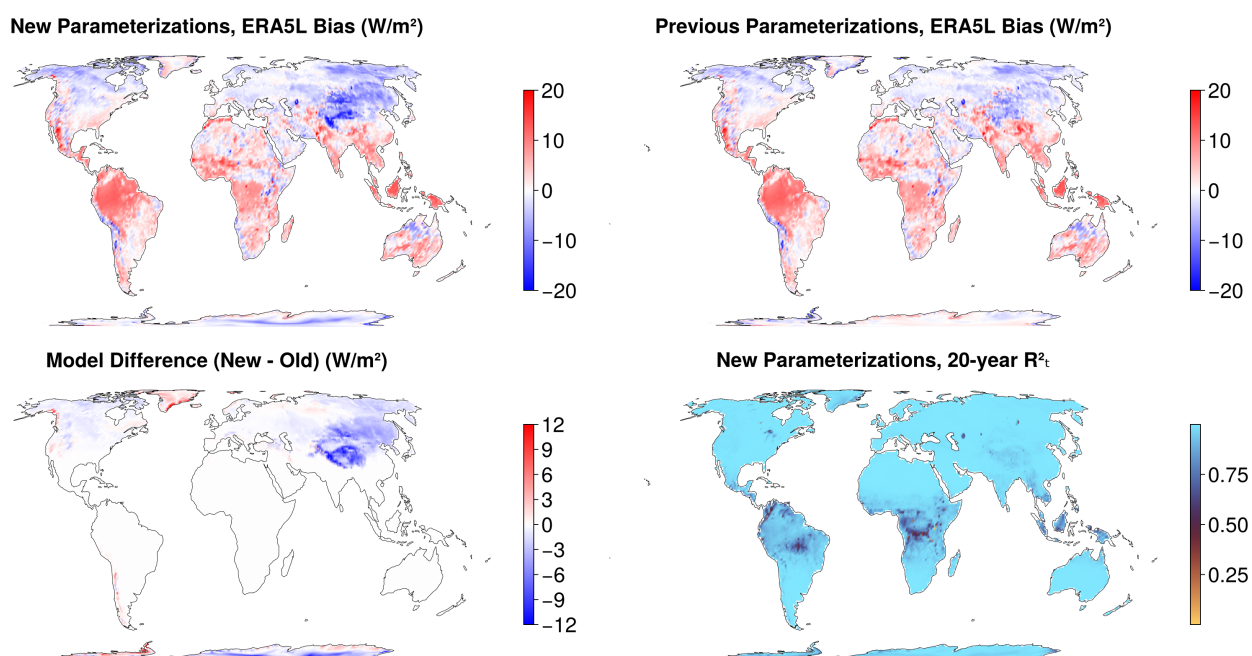


Figure 4.19: Long-wave upward radiation performance of updated parameterizations. These plots follow the format outlined in Fig. 4.2, with the entire global surface analyzed. Like the shortwave radiation results, the new parameterizations over-corrected bias over glacial areas such as Antarctica, though also mitigating some negative bias in Greenland.

References

- [1] R. Abolafia-Rosenzweig, C. He, S. McKenzie Skiles, F. Chen, and D. Gochis. 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), Sept. 2022. doi: 10.1029/2022ms003141. URL <http://dx.doi.org/10.1029/2022MS003141>.
- [2] F. Aguirre, D. Bozkurt, T. Sauter, J. Carrasco, C. Schneider, R. Jaña, and G. Casassa. 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, Nov. 2023. doi: 10.3390/rs15225430. URL <http://dx.doi.org/10.3390/rs15225430>.
- [3] E. A. Anderson. *A point energy and mass balance model of a snow cover*. NOAA technical report ; NWS 19. U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, Office of Hydrology, Silver Spring, Md, 1976. URL <https://repository.library.noaa.gov/view/noaa/6392>.
- [4] E. A. Anderson. Snow accumulation and ablation model – snow-17, Jan 2006. URL <https://www.weather.gov/media/owp/oh/hrl/docs/22snow17.pdf>.
- [5] S. Arioli, G. Picard, L. Arnaud, S. Gascoin, E. Alonso-González, M. Poizat, and M. Irvine. Time series of alpine snow surface radiative-temperature maps from high-precision thermal-infrared imaging. *Earth System Science Data*, 16(9):3913–3934, Sept. 2024. ISSN 1866-3516. doi: 10.5194/essd-16-3913-2024. URL <http://dx.doi.org/10.5194/essd-16-3913-2024>.
- [6] J. Aschauer, A. Michel, T. Jonas, and C. Marty. An empirical model to calculate snow depth from daily snow water equivalent: SWE2HS 1.0. *Geoscientific Model Development*, 16(14):4063–4081, July 2023. ISSN 1991-9603. doi: 10.5194/gmd-16-4063-2023. URL <http://dx.doi.org/10.5194/gmd-16-4063-2023>.
- [7] U. M. Ascher, S. J. Ruuth, and R. J. Spiteri. Implicit-explicit runge-kutta methods for time-dependent partial differential equations. *Applied Numerical Mathematics*, 25(2–3):151–167, Nov. 1997. ISSN 0168-9274. doi: 10.1016/s0168-9274(97)00056-1. URL [http://dx.doi.org/10.1016/S0168-9274\(97\)00056-1](http://dx.doi.org/10.1016/S0168-9274(97)00056-1).
- [8] E. H. Bair, K. Rittger, S. M. Skiles, and J. Dozier. An examination of snow albedo estimates from MODIS and their impact on snow water equivalent reconstruction. *Water Resources Research*, 55(9):7826–7842, Sept. 2019. doi: 10.1029/2019wr024810. URL <http://dx.doi.org/10.1029/2019WR024810>.

- [9] K. J. Bormann, S. Westra, J. P. Evans, and M. F. McCabe. Spatial and temporal variability in seasonal snow density. *Journal of Hydrology*, 484:63–73, 2013. ISSN 0022-1694. doi: <https://doi.org/10.1016/j.jhydrol.2013.01.032>. URL <https://www.sciencedirect.com/science/article/pii/S0022169413000784>.
- [10] R. D. Brown and P. W. Mote. The response of Northern Hemisphere snow cover to a changing climate. *Journal of Climate*, 22(8):2124–2145, Apr. 2009. ISSN 0894-8755. doi: 10.1175/2008jcli2665.1. URL <http://dx.doi.org/10.1175/2008JCLI2665.1>.
- [11] M. I. Budyko. The effect of solar radiation variations on the climate of the Earth. *Tellus*, 21(5):611–619, 1969. doi: <https://doi.org/10.1111/j.2153-3490.1969.tb00466.x>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2153-3490.1969.tb00466.x>.
- [12] C3S. ERA5 hourly data on single levels from 1940 to present, 2018. URL <https://cds.climate.copernicus.eu/doi/10.24381/cds.adbb2d47>.
- [13] A. Charbonneau, K. Deck, and T. Schneider. A physics-constrained neural differential equation framework for data-driven snowpack simulation. *Artificial Intelligence for the Earth Systems*, 4(3), July 2025. doi: 10.1175/aies-d-24-0040.1. URL <http://dx.doi.org/10.1175/AIES-D-24-0040.1>.
- [14] M. P. Clark, J. Hendrikx, A. G. Slater, D. Kavetski, B. Anderson, N. J. Cullen, T. Kerr, E. Örn Hreinsson, and R. A. Woods. Representing spatial variability of snow water equivalent in hydrologic and land-surface models: A review. *Water Resources Research*, 47(7), July 2011. ISSN 1944-7973. doi: 10.1029/2011wr010745. URL <http://dx.doi.org/10.1029/2011WR010745>.
- [15] A. S. Daloz, C. Schwingshackl, P. Mooney, S. Strada, D. Rechid, E. L. Davin, E. Katragkou, N. de Noblet-Ducoudré, M. Belda, T. Halenka, M. Breil, R. M. Cardoso, P. Hoffmann, D. C. A. Lima, R. Meier, P. M. M. Soares, G. Sofiadis, G. Strandberg, M. H. Toelle, and M. T. Lund. Land–atmosphere interactions in sub-polar and alpine climates in the CORDEX flagship pilot study Land Use and Climate Across Scales (LUCAS) models – Part 1: Evaluation of the snow-albedo effect. *The Cryosphere*, 16(6):2403–2419, June 2022. ISSN 1994-0424. doi: 10.5194/tc-16-2403-2022. URL <http://dx.doi.org/10.5194/tc-16-2403-2022>.
- [16] A. Damiani, N. N. Ishizaki, S. Feron, and R. R. Cordero. 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, Feb. 2025. doi: 10.1016/j.scitotenv.2025.178606. URL <http://dx.doi.org/10.1016/j.scitotenv.2025.178606>.

- [17] G. Danabasoglu, J. Lamarque, J. Bacmeister, D. A. Bailey, A. K. DuVivier, J. Edwards, L. K. Emmons, J. Fasullo, R. Garcia, A. Gettelman, C. Hannay, M. M. Holland, W. G. Large, P. H. Lauritzen, D. M. Lawrence, J. T. M. Lenaerts, K. Lindsay, W. H. Lipscomb, M. J. Mills, R. Neale, K. W. Oleson, B. Otto-Bliesner, A. S. Phillips, W. Sacks, S. Tilmes, L. van Kampenhout, M. Vertenstein, A. Bertini, J. Dennis, C. Deser, C. Fischer, B. Fox-Kemper, J. E. Kay, D. Kinnison, P. J. Kushner, V. E. Larson, M. C. Long, S. Mickelson, J. K. Moore, E. Nienhouse, L. Polvani, P. J. Rasch, and W. G. Strand. The Community Earth System Model version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, 12(2), Feb. 2020. ISSN 1942-2466. doi: 10.1029/2019ms001916. URL <http://dx.doi.org/10.1029/2019MS001916>.
- [18] K. Deck, R. K. Braghiere, A. A. Renchon, J. Sloan, G. Bozzola, E. Speer, J. Ben Mackay, T. Reddy, K. Phan, A. L. Gagné-Landmann, Y. Li, D. Yatunin, A. Charbonneau, N. Efrat-Henrici, E. Bach, S. Ma, P. Gentine, C. Frankenberg, A. A. Bloom, Y. Wang, M. Longo, and T. Schneider. ClimaLand: A land surface model designed to enable data-driven parameterizations. *Journal of Advances in Modeling Earth Systems*, 18(1), Jan. 2026. ISSN 1942-2466. doi: 10.1029/2025ms005118. URL <http://dx.doi.org/10.1029/2025MS005118>.
- [19] G. Demil, A. T. Haghighi, B. Klöve, and M. Oussalah. Advances in image-based estimation of snow variable: A systematic literature review on recent studies. *Journal of Hydrology*, 654:132855, June 2025. doi: 10.1016/j.jhydrol.2025.132855. URL <http://dx.doi.org/10.1016/j.jhydrol.2025.132855>.
- [20] G. T. Diro and L. Sushama. Snow–precipitation coupling and related atmospheric feedbacks over North America. *Atmospheric Science Letters*, 19(8), Aug. 2018. doi: 10.1002/asl.831. URL <http://dx.doi.org/10.1002/asl.831>.
- [21] E. Dutra, G. Balsamo, P. Viterbo, P. Miranda, A. Beljaars, C. Schär, and K. Elder. *New snow scheme in HTESSEL: Description and offline validation*. ECMWF Reading, UK, 2009. doi: 10.21957/98X9MRV1Y. URL <https://www.ecmwf.int/node/9167>.
- [22] E. Dutra, J. Muñoz-Sabater, S. Boussetta, T. Komori, S. Hirahara, and G. Balsamo. Environmental lapse rate for high-resolution land surface downscaling: An application to ERA5. *Earth and Space Science*, 7(5), May 2020. ISSN 2333-5084. doi: 10.1029/2019ea000984. URL <http://dx.doi.org/10.1029/2019EA000984>.
- [23] R. Essery, S. Morin, Y. Lejeune, and C. B. Ménard. A comparison of 1701 snow models using observations from an alpine site. *Advances in Water Resources*, 55:131–148, May 2013. doi: 10.1016/j.advwatres.2012.07.013. URL <http://dx.doi.org/10.1016/j.advwatres.2012.07.013>.

- [24] P. Etchevers, E. Martin, R. Brown, C. Fierz, Y. Lejeune, E. Bazile, A. Boone, Y.-J. Dai, R. Essery, A. Fernandez, Y. Gusev, R. Jordan, V. Koren, E. Kowalczyk, N. O. Nasonova, R. D. Pyles, A. Schlosser, A. B. Shmakin, T. G. Smirnova, U. Strasser, D. Versegny, T. Yamazaki, and Z.-L. Yang. Validation of the energy budget of an alpine snowpack simulated by several snow models (Snow MIP project). *Annals of Glaciology*, 38:150–158, 2004. doi: 10.3189/172756404781814825. URL <http://dx.doi.org/10.3189/172756404781814825>.
- [25] S. R. Fassnacht, K. A. Dressler, D. M. Hultstrand, R. C. Bales, and G. Patterson. Inconsistencias temporales en los patrones espaciales del equivalente de agua en nieve: regresiones entre telemetría de nieve y topografía en la cuenca del Río Colorado. *Pirineos*, 167:165–185, May 2012. ISSN 0373-2568. doi: 10.3989/pirineos.2012.167008. URL <http://dx.doi.org/10.3989/pirineos.2012.167008>.
- [26] S. W. Fleming, K. Rittger, C. M. Oaida Taglialatela, and I. Graczyk. 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), Apr. 2024. doi: 10.1029/2023wr035785. URL <http://dx.doi.org/10.1029/2023WR035785>.
- [27] L. Gao, L. Zhang, Y. Shen, Y. Zhang, M. Ai, and W. Zhang. Modeling snow depth and snow water equivalent distribution and variation characteristics in the Irtysh River Basin, China. *Applied Sciences*, 11(18), 2021. ISSN 2076-3417. doi: 10.3390/app11188365. URL <https://www.mdpi.com/2076-3417/11/18/8365>.
- [28] M. García-Rodríguez, J. Rodeiro-Iglesias, and J. A. Añel. Code accessibility and code quality across phases of the models of the Coupled Model Intercomparison Project. *EGUsphere*, 2026:1–27, 2026. doi: 10.5194/egusphere-2025-6108. URL <https://egusphere.copernicus.org/preprints/2026/egusphere-2025-6108/>.
- [29] A. S. Gardner and M. J. Sharp. A review of snow and ice albedo and the development of a new physically based broadband albedo parameterization. *Journal of Geophysical Research: Earth Surface*, 115(F1), Mar. 2010. doi: 10.1029/2009jf001444. URL <http://dx.doi.org/10.1029/2009JF001444>.
- [30] D. Günther, T. Marke, R. Essery, and U. Strasser. 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, Apr. 2019. doi: 10.1029/2018wr023403. URL <http://dx.doi.org/10.1029/2018WR023403>.
- [31] J. C. Hammond, G. A. Sexstone, A. L. Putman, T. B. Barnhart, D. M. Rey, J. M. Driscoll, G. E. Liston, K. L. Rasmussen, D. McGrath, S. R. Fassnacht,

- and S. K. Kampf. High resolution SnowModel simulations reveal future elevation-dependent snow loss and earlier, flashier surface water input for the Upper Colorado River Basin. *Earth's Future*, 11(2), Jan. 2023. ISSN 2328-4277. doi: 10.1029/2022ef003092. URL <http://dx.doi.org/10.1029/2022EF003092>.
- [32] F. Hanzer, K. Helfricht, T. Marke, and U. Strasser. Multilevel spatiotemporal validation of snow/ice mass balance and runoff modeling in glacierized catchments. *The Cryosphere*, 10(4):1859–1881, Aug. 2016. ISSN 1994-0424. doi: 10.5194/tc-10-1859-2016. URL <http://dx.doi.org/10.5194/tc-10-1859-2016>.
- [33] D. Hao, G. Bisht, K. Rittger, T. Stillinger, E. Bair, Y. Gu, and L. R. Leung. Evaluation of E3SM land model snow simulations over the western united states. *The Cryosphere*, 17(2):673–697, Feb. 2023. ISSN 1994-0424. doi: 10.5194/tc-17-673-2023. URL <http://dx.doi.org/10.5194/tc-17-673-2023>.
- [34] K. Kojima. Densification of seasonal snow cover. *Physics of Snow and Ice: Proceedings*, 1(2):929–952, 1967. URL <http://hdl.handle.net/2115/20351>.
- [35] V. Koren, J. Schaake, K. Mitchell, Q. Duan, F. Chen, and J. M. Baker. A parameterization of snowpack and frozen ground intended for NCEP weather and climate models. *Journal of Geophysical Research: Atmospheres*, 104(D16):19569–19585, Aug. 1999. ISSN 0148-0227. doi: 10.1029/1999jd900232. URL <http://dx.doi.org/10.1029/1999JD900232>.
- [36] K. Kouki, P. Räisänen, K. Luojus, A. Luomaranta, and A. Riihelä. Evaluation of Northern Hemisphere snow water equivalent in CMIP6 models during 1982–2014. *The Cryosphere*, 16(3):1007–1030, 2022. doi: 10.5194/tc-16-1007-2022. URL <https://tc.copernicus.org/articles/16/1007/2022/>.
- [37] K. Kouki, K. Luojus, and A. Riihelä. Evaluation of snow cover properties in ERA5 and ERA5-Land with several satellite-based datasets in the Northern Hemisphere in spring 1982–2018. *The Cryosphere*, 17(12):5007–5026, Nov. 2023. ISSN 1994-0424. doi: 10.5194/tc-17-5007-2023. URL <http://dx.doi.org/10.5194/tc-17-5007-2023>.
- [38] G. Krinner, C. Derksen, R. Essery, M. Flanner, S. Hagemann, M. Clark, A. Hall, H. Rott, C. Brutel-Vuilmet, H. Kim, C. B. Ménard, L. Mudryk, C. Thackeray, L. Wang, G. Arduini, G. Balsamo, P. Bartlett, J. Boike, A. Boone, F. Chérut, J. Colin, M. Cuntz, Y. Dai, B. Decharme, J. Derry, A. Ducharne, E. Dutra, X. Fang, C. Fierz, J. Ghattas, Y. Gusev, V. Haverd, A. Kontu, M. Lafaysse, R. Law, D. Lawrence, W. Li, T. Marke, D. Marks, M. Ménégoz, O. Nasonova, T. Nitta, M. Niwano, J. Pomeroy, M. S. Raleigh, G. Schaedler, V. Semenov, T. G. Smirnova, T. Stacke, U. Strasser, S. Svenson, D. Turkov,

- T. Wang, N. Wever, H. Yuan, W. Zhou, and D. Zhu. ESM-SnowMIP: Assessing snow models and quantifying snow-related climate feedbacks. *Geoscientific Model Development*, 11(12):5027–5049, Dec. 2018. doi: 10.5194/gmd-11-5027-2018. URL <http://dx.doi.org/10.5194/gmd-11-5027-2018>.
- [39] C. C. Landry, K. A. Buck, M. S. Raleigh, and M. P. Clark. Mountain system monitoring at Senator Beck Basin, San Juan Mountains, Colorado: A new integrative data source to develop and evaluate models of snow and hydrologic processes. *Water Resources Research*, 50(2):1773–1788, Feb. 2014. ISSN 0043-1397. doi: 10.1002/2013wr013711. URL <http://dx.doi.org/10.1002/2013WR013711>.
- [40] D. M. Lawrence, R. A. Fisher, C. D. Koven, K. W. Oleson, S. C. Swenson, G. Bonan, N. Collier, B. Ghimire, L. van Kampenhout, D. Kennedy, E. Kluzek, P. J. Lawrence, F. Li, H. Li, D. Lombardozzi, W. J. Riley, W. J. Sacks, M. Shi, M. Vertenstein, W. R. Wieder, C. Xu, A. A. Ali, A. M. Badger, G. Bisht, M. van den Broeke, M. A. Brunke, S. P. Burns, J. Buzan, M. Clark, A. Craig, K. Dahlin, B. Drewniak, J. B. Fisher, M. Flanner, A. M. Fox, P. Gentine, F. Hoffman, G. Keppel-Aleks, R. Knox, S. Kumar, J. Lenaerts, L. R. Leung, W. H. Lipscomb, Y. Lu, A. Pandey, J. D. Pelletier, J. Perket, J. T. Randerson, D. M. Ricciuto, B. M. Sanderson, A. Slater, Z. M. Subin, J. Tang, R. Q. Thomas, M. Val Martin, and X. Zeng. The Community Land Model (CLM) version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, 11(12): 4245–4287, Dec. 2019. ISSN 1942-2466. doi: 10.1029/2018ms001583. URL <http://dx.doi.org/10.1029/2018MS001583>.
- [41] W. Y. Lee, H.-J. Gim, and S. K. Park. 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, Nov. 2023. doi: 10.1007/s13143-023-00344-2. URL <http://dx.doi.org/10.1007/s13143-023-00344-2>.
- [42] Y. Lejeune, M. Dumont, J.-M. Panel, M. Lafaysse, P. Lapalus, E. Le Gac, B. Lesaffre, and S. Morin. 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, Jan. 2019. doi: 10.5194/essd-11-71-2019. URL <http://dx.doi.org/10.5194/essd-11-71-2019>.
- [43] T.-S. Lin, C. He, R. Abolafia-Rosenzweig, F. Chen, W. Wang, M. Barlage, and D. Gochis. 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, Feb. 2025. doi: 10.1175/jhm-d-24-0082.1. URL <http://dx.doi.org/10.1175/JHM-D-24-0082.1>.

- [44] J. Lundquist, M. R. Abel, E. Gutmann, and S. Kapnick. Our skill in modeling mountain rain and snow is bypassing the skill of our observational networks. *Bulletin of the American Meteorological Society*, 100(12):2473–2490, Dec. 2019. ISSN 1520-0477. doi: 10.1175/bams-d-19-0001.1. URL <http://dx.doi.org/10.1175/BAMS-D-19-0001.1>.
- [45] C. C. Lundy, R. L. Brown, E. E. Adams, K. W. Birkeland, and M. Lehnig. A statistical validation of the snowpack model in a Montana climate. *Cold Regions Science and Technology*, 33(2):237–246, 2001. ISSN 0165-232X. doi: [https://doi.org/10.1016/S0165-232X\(01\)00038-6](https://doi.org/10.1016/S0165-232X(01)00038-6). URL <https://www.sciencedirect.com/science/article/pii/S0165232X01000386>. ISSN 2000:International Snow Science Workshop.
- [46] J. Magnusson, N. Wever, R. Essery, N. Helbig, A. Winstral, and T. Jonas. Evaluating snow models with varying process representations for hydrological applications. *Water Resources Research*, 51(4):2707–2723, Apr. 2015. doi: 10.1002/2014wr016498. URL <http://dx.doi.org/10.1002/2014WR016498>.
- [47] C. B. Menard, R. Essery, G. Krinner, G. Arduini, P. Bartlett, A. Boone, C. Brutel-Vuilmet, E. Burke, M. Cuntz, Y. Dai, B. Decharme, E. Dutra, X. Fang, C. Fierz, Y. Gusev, S. Hagemann, V. Haverd, H. Kim, M. Lafaysse, T. Marke, O. Nasonova, T. Nitta, M. Niwano, J. Pomeroy, G. Schädler, V. A. Semenov, T. Smirnova, U. Strasser, S. Swenson, D. Turkov, N. Wever, and H. Yuan. Scientific and human errors in a snow model intercomparison. *Bulletin of the American Meteorological Society*, 102(1):E61–E79, Jan. 2021. doi: 10.1175/bams-d-19-0329.1. URL <http://dx.doi.org/10.1175/BAMS-D-19-0329.1>.
- [48] M. Méndez, J. Overbey, and F. G. Tinetti. Legacy fortran software: Applying syntactic metrics to global climate models. In *XVIII Congreso Argentino de Ciencias de la Computación*, 2012. URL <https://sedici.unlp.edu.ar/handle/10915/23718>.
- [49] J. Meyer, J. Horel, P. Kormos, A. Hedrick, E. Trujillo, and S. M. Skiles. Operational water forecast ability of the HRRR-iSnoB combination: An evaluation to adapt into production environments. *Geoscientific Model Development*, 16(1):233–250, 2023. doi: 10.5194/gmd-16-233-2023. URL <https://gmd.copernicus.org/articles/16/233/2023/>.
- [50] J. D. D. Meyer, J. Jin, and S.-Y. Wang. 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, 2012. doi: 10.1175/JHM-D-12-066.1. URL https://journals.ametsoc.org/view/journals/hydr/13/6/jhm-d-12-066_1.xml.
- [51] D. Monteiro and S. Morin. Multi-decadal analysis of past winter temperature, precipitation and snow cover data in the European Alps from reanalyses, climate models and observational datasets. *The Cryosphere*, 17(8):

- 3617–3660, Aug. 2023. ISSN 1994-0424. doi: 10.5194/tc-17-3617-2023. URL <http://dx.doi.org/10.5194/tc-17-3617-2023>.
- [52] J. Muñoz-Sabater, E. Dutra, A. Agustí-Panareda, C. Albergel, G. Arduini, G. Balsamo, S. Boussetta, M. Choulga, S. Harrigan, H. Hersbach, B. Martens, D. G. Miralles, M. Piles, N. J. Rodríguez-Fernández, E. Zsoter, C. Buontempo, and J.-N. Thépaut. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13(9):4349–4383, Sept. 2021. ISSN 1866-3516. doi: 10.5194/essd-13-4349-2021. URL <http://dx.doi.org/10.5194/essd-13-4349-2021>.
- [53] R. Myneni, Y. Knyazikhin, and T. Park. MCD15A2H MODIS/Terra+Aqua leaf area index/FPAR 8-day LZ4 global 500m SIN grid V006, 2015. URL <https://www.earthdata.nasa.gov/data/catalog/lpcloud-mcd15a2h-006>.
- [54] C. A. Pedersen and J.-G. Winther. Intercomparison and validation of snow albedo parameterization schemes in climate models. *Climate Dynamics*, 25(4):351–362, July 2005. ISSN 0930-7575. doi: 10.1007/s00382-005-0037-0. URL <http://dx.doi.org/10.1007/s00382-005-0037-0>.
- [55] B. Poschlod and A. S. Daloz. Snow depth in high-resolution regional climate model simulations over southern Germany – suitable for extremes and impact-related research? *The Cryosphere*, 18(4):1959–1981, Apr. 2024. ISSN 1994-0424. doi: 10.5194/tc-18-1959-2024. URL <http://dx.doi.org/10.5194/tc-18-1959-2024>.
- [56] X. Qu and A. Hall. Assessing snow albedo feedback in simulated climate change. *Journal of Climate*, 19(11):2617 – 2630, 2006. doi: 10.1175/JCLI3750.1. URL <https://journals.ametsoc.org/view/journals/clim/19/11/jcli3750.1.xml>.
- [57] K. Rittger, S. J. P. Lenard, R. Palomaki, M. Brodzik, T. Stillinger, E. Bair, J. Dozier, and T. Painter. Historical MODIS/Terra L3 global daily 500m SIN grid snow cover, snow albedo, and snow physical properties, 2024. URL https://nsidc.org/data/STC_MODSCGDRF_HIST/versions/1.
- [58] A. Ryken, L. A. Bearup, J. L. Jefferson, P. Constantine, and R. M. Maxwell. Sensitivity and model reduction of simulated snow processes: Contrasting observational and parameter uncertainty to improve prediction. *Advances in Water Resources*, 135:103473, Jan. 2020. doi: 10.1016/j.advwatres.2019.103473. URL <http://dx.doi.org/10.1016/j.advwatres.2019.103473>.
- [59] W. D. Sellers. A global climatic model based on the energy balance of the Earth-atmosphere system. *Journal of Applied Meteorology and Climatology*, 8:392–400, 1969. URL [https://doi.org/10.1175/1520-0450\(1969\)008%3C0392:AGCMBO%3E2.0.CO;2](https://doi.org/10.1175/1520-0450(1969)008%3C0392:AGCMBO%3E2.0.CO;2).

- [60] C. W. Thackeray, X. Qu, and A. Hall. Why do models produce spread in snow albedo feedback? *Geophysical Research Letters*, 45(12):6223–6231, 2018. doi: <https://doi.org/10.1029/2018GL078493>. URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL078493>.
- [61] V. Vionnet, D. Six, L. Auger, M. Dumont, M. Lafaysse, L. Quéno, M. Réveillet, I. Dombrowski-Etchevers, E. Thibert, and C. Vincent. Sub-kilometer precipitation datasets for snowpack and glacier modeling in alpine terrain. *Frontiers in Earth Science*, 7, 2019. ISSN 2296-6463. doi: 10.3389/feart.2019.00182. URL <https://www.frontiersin.org/articles/10.3389/feart.2019.00182>.
- [62] N. Wever, L. Schmid, A. Heilig, O. Eisen, C. Fierz, and M. Lehning. Verification of the multi-layer SNOWPACK model with different water transport schemes. *The Cryosphere*, 9(6):2271–2293, 2015. doi: 10.5194/tc-9-2271-2015. URL <https://tc.copernicus.org/articles/9/2271/2015/>.
- [63] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80, Dec. 1945. doi: 10.2307/3001968. URL <http://dx.doi.org/10.2307/3001968>.
- [64] D. Yatunin, S. Byrne, C. Kawczynski, S. Kandala, G. Bozzola, A. Sridhar, Z. Shen, A. Jaruga, J. Sloan, J. He, D. Z. Huang, V. Barra, R. Chew, A. Boral, Y. Chen, O. Knoth, P. Ullrich, C. Mbengue, and T. Schneider. The Climate Modeling Alliance atmosphere dynamical core: Concepts, numerics, and scaling. *Journal of Advances in Modeling Earth Systems*, 18(3), Mar. 2026. ISSN 1942-2466. doi: 10.1029/2025ms005014. URL <http://dx.doi.org/10.1029/2025MS005014>.
- [65] F. Ye, Q. Cheng, W. Hao, D. Yu, C. Ma, D. Liang, and H. Shen. 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, Nov. 2023. doi: 10.1016/j.jag.2023.103519. URL <http://dx.doi.org/10.1016/j.jag.2023.103519>.
- [66] J. You, D. G. Tarboton, and C. H. Luce. Modeling the snow surface temperature with a one-layer energy balance snowmelt model. *Hydrology and Earth System Sciences*, 18(12):5061–5076, Dec. 2014. ISSN 1607-7938. doi: 10.5194/hess-18-5061-2014. URL <http://dx.doi.org/10.5194/hess-18-5061-2014>.
- [67] Y. Zhang and T. Ohata. Parameterization of albedo on snow/ice surface. A review. *Journal of the Japanese Society of Snow and Ice*, 65(1):33–51, 2003. doi: 10.5331/seppyo.65.33. URL <http://dx.doi.org/10.5331/seppyo.65.33>.

CONCLUDING THOUGHTS

In this dissertation, a neural differential equation framework for hybrid physics-ML parameterizations was developed and applied to snow depth and snow albedo, to demonstrate the efficacy of ML-informed modeling capabilities in large-scale snow simulations like global climate modeling. The central challenge in this endeavor is not an absence of physical understanding of snow processes, but instead the persistent mismatch between the multiscale coupled nature of snow dynamics and the limitations of traditional parameterizations in capturing these processes consistently, both when generalizing to a range of climate regimes but also across a range of spatial and temporal scales. By embedding physical constraints directly into a neural dynamical system, the presented framework enables data-driven parameterizations that remain numerically stable and consistent with conservation laws while retaining the flexibility to represent nonlinear process interactions across a variety of evaluation environments.

In three main areas — snow depth evolution, snow albedo seasonality, and coupled climate model integration — the framework was shown to outperform multiple established parameterizations at point-scale assessment with quality-controlled observational data as well as remaining competitive at coarser-scale or even global surface modeling. These applications of the framework demonstrated an ability to generalize to physically distinct regions of the cryosphere, benefitting from core principles of embedding physical behavior and learning nonlinear forcing-response relationships. Their shared success reveal the advantage of the framework lies not in process-specific tuning, but in its ability to encode physically consistent state evolution within a small and flexible functional form, where widespread data is available. This emphasizes that constraints are more than just regularization. They are structural priors that explicitly enable cross-domain transfer when combined with a sufficiently flexible underlying predictive component, an idea on which the developed framework effectively exploits.

These overall results constitute two key findings: first, that one singular framework holds the potential for improvement across a range of physical systems, and second, that such improvements are not confined to solely the variable the framework is em-

ployed to represent. Improvements in hydrological bulk properties were propagated (with the help of the new temperature parameterization as well; built upon similar principles) to other hydrological and even radiative components of the new system. The act of propagation itself is a straightforward conclusion given the fundamental nature of the coupling within GCMs, but, more abstractly, this also suggests the proposed approach serves an important role in encapsulating a more general structure of snow evolution as a constrained dynamical system, compared to serving only as a flexible functional approximation for individual variables. Synthesized, these outcomes reinforce the idea that improvements at the process level can translate into meaningful system-level gains when embedded within a coupled modeling architecture. However, not every variable showed substantial improvement at global scales, emphasizing that coupling introduces additional interactions that are not fully captured in point-scale evaluation.

Compared to the very large ML/AI models prevalent in research and industry, the schemes discussed in this dissertation are unexpectedly diminutive; yielding architectures that could have been utilized for hydrological modeling on a mainframe from the 1970s, or that are certainly reproducible on fortran or legacy systems still used in climate modeling today. It is notable for all three new parameterizations that adding only limited physical insight and basic boundary functions allows us to replicate or outclass the performance achievable by comparable physical/empirical schemes or with pure ML alone. This further highlights the importance of identifying and applying physical insight when it can be employed, even when moving towards more black-box techniques. A reduced but effective computational footprint permits more advanced functionality for the same cost in global simulations; offering a competitive alternative to existing detailed schemes. This reduces computational slowdown that would otherwise arise from the usage of larger models, or implies a ready path forward for prototyping and substitution into identified bottlenecks, even beyond snow modeling specifically. In a similar vein, the usage of the presented methodology compared to more traditional broadcasted conditional logic permits a hardware speed-up that is conducive to modern and future computational languages and hardware. This scalability will be crucial as growing model resolution continues to push the limits of what is possible in global climate simulations.

Despite the demonstrated performance improvements, several limitations remain. The framework is fundamentally constrained by the availability and representativeness of observational training data, which remains spatially uneven and biased

toward well-instrumented snow regimes. As a result, performance degradation is observed here in environments that are underrepresented, particularly for complex, remote, or more extreme terrain, as well as transitional snow climates. Spread in data temporal and aggregation scale is also as important as climate and spatial diversity. Improvements over point-scale simulations do not guarantee global improvement, and the range of extremes in variables like precipitation rates do not readily transfer between different temporal aggregations, which demonstrably decreased the snow depth parameterization performance when evaluated at different time-steps. Holistic coverage of data efforts will always remain an issue in global climate modeling, so incorporating insight from reanalysis or gridded remote sensing data can likely help in extracting leading order information from under-sampled regimes. However, when using these products, proper care must be taken to limit the effects of structural or footprint-dependent biases in their preparation to permit proper generalization. Benchmarking or pruning remote samples in accordance with existing observational samples (as was done with the snow albedo gridded data and SNOTEL observations) plays a critical role in quality assurance to create a pool of physically consistent, trusted, and informative training samples to enable generalizable learning.

The developed hybrid ML framework demonstrates strong capability to address many conventional challenges in model generalization, particularly when training data and feature representations are thoughtfully constructed. Its ability to adapt to input uncertainty, incorporate new variables, and perform effectively under limited representative data highlights its potential relative to traditional approaches. However, as with all data-informed parameterizations, its performance can be tempered by the availability and representativeness of observational data. In particular, processes that remain poorly observed, such as the effects of aerosols or grain-scale variability in snow albedo, introduce unavoidable sources of uncertainty. Importantly, this constraint is not unique to the present framework, but reflects a broader limitation across empirical and hybrid modeling approaches. Indeed, within this context, the results of this thesis suggest that the proposed methodology is comparatively robust to such gaps, while also providing a clear pathway for continued improvement as new data become available. Rather than representing a limitation of the framework itself, these challenges identify a critical opportunity for targeted observational efforts and data collection to further advance snow modeling capabilities. It is important to reiterate that the theoretical foundations of many of these processes are already well-characterized and fundamentally understood; theoretical advances in the process-based understanding of snow are presently less impor-

tant than determining how to faithfully represent these processes at-scale when the necessary inputs are unavailable or uncertain¹. In a few of these cases, further observational effort can provide much-needed data volume to permit a richer and better-partitioned learning space for the proposed framework to harness, though for many processes their closure at coarser scales will necessarily require more indirect and/or statistical handling of anticipated biases. Spatial assessment of model bias patterns with regards to land/climate type beyond those seen here could influence the development of additional climate/location-aware features to introduce more separability, or identify critical climates for which to gather more observations, in order to begin mitigating the errors associated with these types of processes.

The ideas of this thesis were exciting to develop, but are far from absolute — there will always be more ideas to explore. The proof-of-concept boundary functions selected for the presented parameterizations served as a foundational baseline. This choice was deliberate in these expositional works and debut of the framework, particularly for preliminary assessment of the created methodology and for exploration of each model’s capacity to learn other physical properties absent more specific guidelines. With the general behavior now better qualified, the introductory approach remains flexible to incorporating additional functional forms to curate further physical consistency. Explicit conservation of mass could be built into the depth parameterization, or more physical limits on the snow depth upper bound based on the Zwart parameterization² could be explored. A more detailed investigation of the pareto boundary from further physically-informed restriction against maintaining sufficient expressive flexibility could aid in translating more qualitative rules-of-thumb into explicit design principles for future application of the framework. The models presented here were also trained on primarily mixed-location point-scale data, but not mixed-frequency or global-resolution data. Data extremes vary with temporal and spatial aggregation, so training on a mixture of hourly-resolution and daily-resolution data or incorporating reanalysis data could enhance a scheme’s ability to be run at different time-steps, or over more extreme climates. Online training (fine-tuning) of model parameters once coupled inside a global model could also improve downstream simulation fidelity. Another direction

¹To put it another way, the problem is not so much of a lack of understanding of snow processes themselves, but how to manufacture *separability* of the target variable *only* from the available observations/features, in order to create a better one-to-one mapping between the expected process behavior and the available environmental state (which might necessarily never include that system’s governing inputs). Until GCMs reach sub-km scales, many inputs will remain too coarse-grained to explicitly represent their understood behavior to model their column-scale impact to bulk energy/mass.

²See section 3.A in Chapter 3.

could involve non-local modeling, by incorporating convolutional neural networks or inputs from other columns to permit representation of lateral fluxes and other transfer effects. While such approaches are within the scope of the framework, many lateral processes are inherently tied to specific spatial scales, which throttles a trained parameterizations ability to generalize over different modeling grids, and was thus not investigated here. Including transfer effects from other columns is more representative of many under-represented processes in many GCMs or the present form of the ClimaLand Snow model, but aggregating non-local information also exacerbates issues from missing or uncertain inputs, which can further reduce the available pool for training data. Furthermore, lateral transfer effects are less important when the horizontal resolution exceeds 10–100km. Further qualification of the tradeoffs of these considerations remains an avenue for future work.

In a broader sense, the developments of this thesis are not restricted to just deep neural networks, nor gradient-based training, nor even snow modeling. The architectures here are adaptable to many classical empirical schemes, or alternative ML methods like tree-based models, transformers, or even non-parametric schemes. Probabilistic schemes like Bayesian networks could be utilized to create uncertainty-aware climate parameterizations. The possibilities for representable constraints are even more vast. Black-box constraints or specific model types like Kolmogorov Arnold Networks could be learned simultaneously with the data as a form of equation discovery for multi-scale systems with unknown limiting behavior. On modern code systems, automatic differentiation packages allow the gradient of a neural network output with respect to its inputs to be passed to a constraint function along with the prediction, enabling additional transformations on the output that represent a broader class of physical constraints than bounds alone. This technique could be combined with the loss-function augmentation techniques seen in other physics-informed network techniques to enable an exciting regime of hybrid physical modeling, in snow modeling and beyond. I am optimistic for wider interest about the tools presented here, and have abstracted them into a Julia module to use with your own constraint functions and predictive models, with tutorials available [here](#). Taken together, I hope these ideas invite further curiosity and iteration, contributing in a small way to the continued evolution of physically grounded, data-driven modeling approaches.

The works of this thesis ultimately yield a unified framework for cryosphere parameterizations, and perhaps the first successful integration of ML-based snow parameterizations into a GCM. With careful forethought, deliberate construction of

schemes that are computationally efficient yet hold the same expressive capability as more complex approaches provide improvements in snow modeling that extend from local to global scales, particularly in snow albedo and their combined effects on snow seasonality. This demonstration of cross-process generalization is evidence that ML-physics models can improve climate-relevant outputs without destabilizing coupled dynamical systems while remaining anchored in physical law, and even go so far as reduce longstanding biases in snow schemes. This research lays the foundation for next-generation parameterizations in effective and scalable large-scale climate modeling, or other physical modeling systems.