ClimaLand: a land surface model designed to enable data-driven parameterizations

Deck, K., Braghiere, R. K., Renchon, A. A., Sloan, J., Bozzola, G., Speer, E., Mackay, J. B., Reddy, T., Phan, K., Gagné-Landmann, A. L., Li, Y., Yatunin, D., Charbonneau, A., Efrat-Henrici, N., Bach, E. ORCID: https://orcid.org/0000-0002-9725-0203, Ma, S., Gentine, P., Frankenberg, C., Bloom, A. A., Wang, Y., Longo, M. and Schneider, T. (2026) ClimaLand: a land surface model designed to enable data-driven parameterizations. Journal of Advances in Modeling Earth Systems, 18 (1). e2025MS005118. ISSN 1942-2466 doi: 10.1029/2025MS005118 Available at https://centaur.reading.ac.uk/127790/

JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2025MS005118

Special Collection:
The CliMA Earth System Model

ClimaLand: A Land Surface Model Designed to Enable Data-Driven Parameterizations

Katherine Deck[1], Renato K. Braghiere[1,2], Alexandre A. Renchon[1], Julia Sloan[1], Gabriele Bozzola[1], Edward Speer[1], J. Ben Mackay[3], Teja Reddy[1], Kevin Phan[1], Anna L. Gagné-Landmann[4], Yuchen Li[5], Dennis Yatunin[1], Andrew Charbonneau[1], Nat Efrat-Henrici[1], Eviatar Bach[1,6,7], Shuang Ma[2,8], Pierre Gentine[9], Christian Frankenberg[1], A. Anthony Bloom[2], Yujie Wang[1], Marcos Longo[10], and Tapio Schneider[1]

[1]California Institute of Technology, Pasadena, CA, USA, [2]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, [3]Scripps Institution of Oceanography, La Jolla, CA, USA, [4]Northern Arizona University, Flagstaff, AZ, USA, [5]Department of Physics, Stanford University, Stanford, CA, USA, [6]University of Reading, Reading, UK, [7]National Centre for Earth Observation, Reading, UK, [8]Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, CA, USA, [9]Department of Earth and Environmental Engineering, Columbia University, New York, NY, USA, [10]Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Key Points:

- ClimaLand, a land surface model, simulates energy, water, and carbon fluxes within and across soil, canopy and snow components
- The soil model simulates flow and phase changes of water in both saturated and unsaturated zones
- The model runs natively on graphics processing units and is designed to enable the inclusion of data-driven parameterizations

Correspondence to:
K. Deck,
kdeck@caltech.edu

Citation:
Deck, K., Braghiere, R. K., Renchon, A. A., Sloan, J., Bozzola, G., Speer, E., et al. (2026). ClimaLand: A land surface model designed to enable data-driven parameterizations. *Journal of Advances in Modeling Earth Systems*, *18*, e2025MS005118. https://doi.org/10.1029/2025MS005118

Received 28 MAR 2025
Accepted 19 NOV 2025

Abstract Land surface models (LSMs) are essential tools for simulating the coupled climate system, representing the dynamics of water, energy, and carbon fluxes on land and their interaction with the atmosphere. However, parameterizing sub-grid processes at the scales relevant to climate models (∼10–100 km) remains a considerable challenge. The parameterizations typically have a large number of unknown and often correlated parameters, making calibration and uncertainty quantification difficult. Moreover, many existing LSMs are not readily adaptable to the incorporation of modern machine learning (ML) parameterizations trained with in situ and satellite data. This article presents the first version of ClimaLand, a new LSM designed for overcoming these limitations, including a description of the core equations underlying the model, the results of an extensive set of validation exercises, and an assessment of the computational performance of the model. We show that ClimaLand can leverage graphics processing units for computational efficiency, and that its modular architecture and high-level programming language, Julia, allows for integration with ML libraries. In the future, this will enable efficient simulation, calibration, and uncertainty quantification with ClimaLand.

Plain Language Summary Simulating the Earth's atmosphere, ocean, and land surface is an important method that scientists use for understanding the Earth's climate, including its response to climate change. Due the complexity of the processes involved, approximations are made when representing certain aspects of the land surface, such as vegetation heterogeneity or topographical variation. These approximations can be improved by using data ("calibration"), but doing so has a large computational cost. They can also be improved using machine learning (ML), but this requires models to be easily integrated with ML packages. ClimaLand is a new land surface model which has been designed from the start to incorporate ML parameterizations and to more efficiently calibrate parameterizations with data. This article presents the ClimaLand model, benchmarks its computational performance, and compares model output against data in a variety of regimes. Follow-on studies will improve the core model using ML parameterizations and by calibrating the model.

1. Introduction

A land surface model (LSM) is an integral part of an Earth system model (ESM), a key tool used to understand and predict Earth's climate. Fluxes exchanged between the land, atmosphere and ocean close the water, energy, and carbon cycles and provide sources and sinks driving other ESM components. LSMs are in turn driven by the atmosphere and can be used to understand and predict how a changing climate will affect hydrology, the cryosphere, and the terrestrial biosphere, including their coupled feedbacks on other Earth system components. These land dynamics are regulated not only by physical principles but also by the complex biological and ecological processes that drive biosphere–atmosphere coupling (G. B. Bonan et al., 2024). Differences in the representation of these land processes in models lead to varying predictions in future climate state. For example, the land surface

has absorbed nearly 30% of anthropogenic CO_2 emissions, but the magnitude and sign of the future carbon uptake rate by land is uncertain (Friedlingstein et al., 2023).

There are limitations in how current LSMs represent certain physical behaviors. Most LSMs do not explicitly include processes such as lateral groundwater flow and water vapor diffusion in soil, requiring parameterizations based on known prognostic variables (M. P. Clark et al., 2015). Heterogeneity in the land surface, and the fact that different nonlinear physical processes are important at different scales, also necessitate parameterization, which describe how the modeled variables, which typically represent horizontal scales of $\mathcal{O}$(10–100 km), are affected by unresolved subgrid scales. Examples include parameterizations for the role of topography in runoff generation (e.g., Niu et al. (2005)) and for partially snow-covered grid cells. Disturbances, including wildfires, permafrost thaw, and land-use change, introduce further variability that influences carbon storage, surface albedo, and aerosol emissions (Braghiere, Fisher, et al., 2023; Schädel et al., 2024). Addressing these uncertainties requires integrating ecological and geophysical processes and their multi-scale interactions (G. B. Bonan et al., 2024).

Calibration and uncertainty quantification are required for all parameterizations included in LSMs but remain challenging due to the complexity of terrestrial processes and their combined impacts on the climate system. Current land models incorporate a vast array of correlated parameters spanning hydrology, carbon cycling, and plant physiology, many of which are poorly constrained by observational data (Dagon et al., 2020; Zaehle et al., 2005). Calibration of computationally expensive models with many free parameters is difficult, and quantifying the uncertainty that these parameters induce in downstream predictions is even more challenging because it requires many model runs. Perturbed parameter ensemble (PPE) studies highlight that land parameter uncertainty propagates through land–atmosphere fluxes, influencing global temperature and precipitation distributions (Fischer et al., 2011), and could drive a 2.2°C variation in global-mean land surface temperature and significant regional precipitation variability (Zarakas et al., 2024). The computational demand of running a model many times with different parameter values is a key bottleneck in both inference and uncertainty quantification (Danabasoglu et al., 2020).

Machine learning (ML) is transforming climate modeling in several areas, including model emulation and replacing process parameterizations (Eyring et al., 2024; Lai et al., 2025; Reichstein et al., 2019; Schneider, Teixeira, et al., 2017). In land surface modeling, for example, ML has excelled in modeling runoff, river discharge, and flooding (Kratzert et al., 2018; Kratzert, Klotz, Herrnegger, et al., 2019; Kratzert, Klotz, Shalev, et al., 2019; Nearing et al., 2024). It also has great promise for replacing highly uncertain parameterizations within a process-based LSM framework that can, for example, encode conservation laws for energy, water, and carbon (Fang & Gentine, 2024; Kraft et al., 2021; Schneider, Lan, et al., 2017). These hybrid models have reduced biases in subgrid-scale processes such as cloud cover and ocean mesoscale eddies, outperforming traditional parameterizations (e.g., Beucler et al., 2021; Christopoulos et al., 2024; Lopez-Gomez et al., 2022; Perezhogin et al., 2024). Furthermore, ML-enhanced approaches to modeling terrestrial processes such as evapotranspiration have shown improved generalization, especially during extreme events, compared to purely physical models (Zhao et al., 2019). ML can address longstanding challenges in land surface modeling when the necessary training data are available. Such data are available for many applications, including rain and river discharge gauges (e.g., Färber et al., 2024), spectrometers that measure the land albedo (e.g., Schaaf & Wang, 2015), gravimeters that measure the terrestrial storage of water (e.g., Landerer & Swenson, 2012), or flux towers that measure atmosphere-land fluxes (e.g., Ameriflux, 2015).

This article details version 1.0 of the Climate Modeling Alliance (CliMA) land model, ClimaLand. The model is developed from scratch in Julia, a high-level language that provides ML libraries that make it easy to incorporate ML parameterizations. To demonstrate this, a prognostic snow depth model based on a neural differential equation (Charbonneau et al., 2025) is added to ClimaLand; since ClimaLand is designed to be modular, experimenting with other machine-learned parameterizations is similarly straight forward. ClimaLand has also been designed to run natively on GPUs. This leads to a computational performance that enables global simulations without requiring access to a supercomputer, and it greatly reduces the computational cost of calibration and uncertainty quantification. It also enables kilometer-scale simulations on multiple GPUs. This article presents the model equations (Section 2), quantifies the performance of the model on GPUs (Section 3), and validates the model output using laboratory, site-level, and global data (Section 4). Further contextualization of the model, including plans for calibration and a discussion of model equations compared with other LSMs, is provided in

Section 5. This work demonstrates the potential of ClimaLand as a tool for physics/ML hybrid modeling. The code itself with additional documentation and tutorials is publicly available on GitHub (Deck et al., 2025).

2. ClimaLand Model Description

The land model is composed of three components: soil, snow, and canopy. The soil model, described in Section 2.1, encodes water and energy balances, including freezing and thawing, in both the water table and the vadose zone. Advancing these equations in time requires solving partial differential equations for soil moisture and energy. An option to model heterotrophic respiration and the vertical profile of CO_2 in the pore space is available. The snow model, described in Section 2.2, also encodes energy and water balances, but because we do not vertically resolve the snowpack structure, only a set of ordinary differential equations is solved. As described, the snow depth is a variable predicted with an ML parametrization. The canopy model, described in Section 2.3, models photosynthesis, radiative transfer in the canopy, transpiration, and the flow of water from the soil and through the plant. As with the snow model, a set of ordinary differential equations is solved. These components interact via fluxes or sources and sinks in their equations, as described in detail in the following sections.

2.1. Soil Model

The soil model consists of a set of equations describing the evolution of the soil volumetric internal energy ρe $\left(\text{J m}^{-3}\right)$, liquid water content θ_l $\left(\text{m}^3\ \text{m}^{-3}\right)$, ice content θ_i $\left(\text{m}^3\ \text{m}^{-3}\right)$, and, optionally, concentrations of tracers such as CO_2. We use the notation $\boldsymbol{\theta} = (\theta_l, \theta_i)$ for the water content, $\boldsymbol{x}$ for spatial location, and $\boldsymbol{\nu}$ for the dry-soil composition, which is a prescribed, spatially varying but temporally constant, parameter field (see Appendix A1). The bold notation indicates vector-valued variables. Table 1 lists the key variables and parameters in the soil model, including prognostic variables, functions of the prognostic state, global physical constants, and empirical —and often spatially varying—properties of the soil.

2.1.1. Water Balance

2.1.1.1. Liquid Water Flux

The central quantity around which the soil water balance revolves is the volume flux of water $\boldsymbol{d}_l$ $\left(\text{m s}^{-1}\right)$, given by Darcy's law (e.g., Dingman, 2015):

$$\boldsymbol{d}_l = -K\nabla h. \tag{1}$$

Here, $K = K(\boldsymbol{\theta};\boldsymbol{\nu})$ is the hydraulic conductivity $\left(\text{m s}^{-1}\right)$ and $h = h(\boldsymbol{x},\boldsymbol{\theta};\boldsymbol{\nu})$ is the hydraulic head or water potential (m). The hydraulic conductivity K is assumed to be a scalar, implying an isotropic conductivity. Horizontal layering of soil may induce important anisotropy, which can be represented by an anisotropic hydraulic conductivity tensor, if needed.

The hydraulic head is the sum of the elevation head, given by the height z above a reference elevation, and the pressure head ψ:

$$h(\boldsymbol{x},\boldsymbol{\theta};\boldsymbol{\nu}) = z + \psi(\boldsymbol{x},\boldsymbol{\theta};\boldsymbol{\nu}). \tag{2}$$

The pressure head represents the water pressure within the soil matrix. In unsaturated soils, it is the (negative) matric potential $\psi_m(\boldsymbol{\theta};\boldsymbol{\nu})$, which is determined from the capillary pressure across the pore liquid/gas interface by assuming that the gas pressure within the pores is constant across the soil matrix (Bear, 2018). In the saturated zone, below the water table, there are no capillary effects, and the pressure head $\psi = \psi(\boldsymbol{x};\boldsymbol{\nu})$ represents the (positive) water pressure in the soil matrix.

2.1.1.2. Conservation Law for Liquid Water

Following Woodward and Dawson (2000) and Endrizzi et al. (2014), the Richards' equation suitable for the vadose (unsaturated) zone is combined with a form of Richards' equation suitable for the saturated zone by defining the augmented liquid fraction

Table 1
Key Variables and Parameters in Soil Model

Prognostic variables	Description	Units	Definition	Range/Typical value
ρe	Internal energy per unit volume	J m^{-3}	Equation 6	4×10^6 J m^{-3}
ϑ_l	Augmented liquid fraction	m^3 m^{-3}	Equation 3	$0 \le \vartheta_l$
θ_i	Volume fraction of ice	m^3 m^{-3}		$0 \le \theta_i \le \nu_p$
Functions of prognostic state	**Description**	**Units**	**Definition**	**Range/Typical value**
θ_l	Volume fraction of liquid water	m^3 m^{-3}	Equations A13a–A13c	$0 \le \theta_l \le \nu_{\text{eff}}$
$\theta_w = \theta_l + \theta_i$	Volume fraction of total water	m^3 m^{-3}		$0 \le \theta_w \le \nu_p$
T	Soil Temperature	K	Equation 11	288 K
$\nu_{\text{eff}} = \nu_p - \theta_i$	Effective porosity		Equation A12	$0 \le \nu_{\text{eff}} \le \nu_p$
ρ	Bulk density of (wet) soil	kg m^{-3}	Equation 7	1.5×10^3 kg m^{-3}
$\rho_s c_s$	Volumetric heat capacity	J m^{-3} K^{-1}	Equation 8	2×10^6 J m^{-3} K^{-1}
κ	Thermal conductivity	W m^{-1} K^{-1}	Equation A2	0.85 W m^{-1} K^{-1}
κ_{sat}	Saturated thermal conductivity	W m^{-1} K^{-1}	Equation A11	3 W m^{-1} K^{-1}
K_e	Kersten number		Equation A3	$0 \le K_e \le 1$
$S_l, S_{l,\text{eff}}$	Liquid saturation, Effective liquid saturation		Equations A15 and A16	$S_l, S_{l,\text{eff}} > 0$
h	Hydraulic head	m	Equation 2	
ψ	Pressure head	m	Equation A14	
ψ_m	Matric potential (van Genuchten)	m	Equation A17	$\psi_m \le 0$
K	Hydraulic conductivity (van Genuchten)	m s^{-1}	Equation A18	10^{-6} m s^{-1}
Global constants	**Description**	**Units**		**Value**
g	Standard gravitational acceleration at the surface of Earth	m s^{-2}		9.81 m s^{-2}
T_f	Freezing point temperature at 10^5 Pa	K		273.15 K
T_0	Reference temperature	K		273.16 K
$L_{f,0}$	Latent heat of fusion at T_0	J kg^{-1}		333.6×10^3 J kg^{-1}
c_l	Specific heat capacity of liquid water	J kg^{-1} K^{-1}		4181 J kg^{-1} K^{-1}
c_i	Specific heat capacity of ice	J kg^{-1} K^{-1}		2100 J kg^{-1} K^{-1}
ρ_l	Density of liquid water	kg m^{-3}		1000 kg m^{-3}
ρ_i	Density of ice	kg m^{-3}		916.7 kg m^{-3}
Empirical properties	**Description**	**Units**	**Obtained from**	**Typical value**
ν_p	Porosity		Input data	$0 \le \nu_p \le 1$
θ_{res}	Residual fraction		Input data	$0 \le \nu_p$
ν_j	Volume fraction of χ_j, relative to bulk soil		Input data	$0 \le \chi_j \le 1$
$\nu_{j,ss}$	Volume fraction of χ_j, relative to soil solids		Input data	$0 \le \chi_j \le 1$
$\rho_j c_j$	Volumetric heat capacity of χ_j	J m^{-3} K^{-1}	Input data	$O(1$ J m^{-3} K^{-1}
ρ_{ds}	Particle density dry soil	kg m^{-3}	Input data	$O(10^3$ kg m$^{-3})$
ρ_b	Bulk density of dry soil	kg m^{-3}	Equation A7	$O(10^3$ kg m$^{-3})$
κ_{dry}	Dry thermal conductivity	W m^{-1} K^{-1}	Input data	1.5 W m^{-1} K^{-1}
S_s	Aquifer specific storage	m^{-1}	Input data	10^{-4} m^{-1}
K_{sat}	Saturated hydraulic conductivity	m s^{-1}	Input data	10^{-5} m s^{-1}
n	Van Genuchten shape parameter		Input data	2
α	Van Genuchten inverse ref. potential	m^{-1}	Input data	2 m^{-1}
τ_{FT}	Phase transition time	s	Equation 14	

Note. Some variables are defined in Appendix A.

$$\vartheta_l(\psi) = \theta_l(\psi) + \begin{cases} 0 & \text{when unsaturated } (\psi \le 0), \\ S_s\psi & \text{when saturated } (\psi > 0), \end{cases} \tag{3}$$

where S_s $\left(\mathrm{m}^{-1}\right)$ is the specific storage (or specific storativity) of an aquifer, a coefficient that measures the increase in water volume in a soil volume per increase in head h (Dingman, 2015, Chapter 7). The specific storage S_s can vary with location and is proportional to the compressibility of water and of the soil matrix; it vanishes if both compressibilities are zero (Bear, 2018, Chapter 5). The form of ψ in the saturated zone is given implicitly by Equation 3, as $\psi = (\vartheta_l - \theta_l)/S_s$.

The augmented liquid fraction ϑ_l and ice fraction θ_i satisfy the conservation laws

$$\frac{\partial \vartheta_l}{\partial t} = -\nabla \cdot (-K\nabla h) - \frac{F_T}{\rho_l} + \mathcal{S}_{\vartheta_l}, \tag{4a}$$

$$\frac{\partial \theta_i}{\partial t} = \frac{F_T}{\rho_i} + \mathcal{S}_{\theta_i}. \tag{4b}$$

The conservation laws assume constant densities ρ_l and ρ_i of liquid and ice, so that the volume fractions are conserved. The first equation, for liquid water, is a generalized form of Richards' equation in terms of the augmented liquid fraction ϑ_l, which holds in saturated and unsaturated soils (Woodward & Dawson, 2000). The second equation, for ice, assumes that water frozen into the soil matrix does not flow, so no flux term appears on the right-hand side. The term F_T $\left(\mathrm{kg\ m^{-3}\ s^{-1}}\right)$ on the right-hand side of both equations represents the mass conversion rate between liquid and ice in phase changes of water, formulated so that total water mass is conserved. The terms $\mathcal{S}_{\vartheta_l}, \mathcal{S}_{\theta_i}$ represents additional source/sink terms.

ClimaLand supports the van Genuchten (1980) and Brooks and Corey (1964) parameterizations for hydraulic conductivity K and matric potential ψ_m as functions of volumetric liquid water content in the vadose zone, along with factors that adjust K for the presence of ice and the temperature dependence of the viscosity of water (Hansson et al., 2004; Lundin, 1990). These equations are presented for completeness in Appendices A3 and A4.

2.1.2. Energy Balance

We characterize the thermodynamic state of soil by its internal energy. The specific internal energy (energy per unit mass, J kg^{-1}) is the mass-weighted mean of the constituent specific internal energies of dry soil e_{ds}, liquid water e_l, and ice e_i; the small contribution of gases to the mass and specific internal energy of soil is neglected. The specific internal energies of the constituents depend on the temperature T of the constituents (which are assumed to be in local thermal equilibrium, so they have the same temperature in any one location) and the specific heat capacities c_{ds}, c_l, and c_i:

$$e_{ds}(T;\boldsymbol{\nu}) = c_{ds}(\boldsymbol{\nu})(T - T_0), \tag{5a}$$

$$e_l(T) = c_l(T - T_0), \tag{5b}$$

$$e_i(T) = c_i(T - T_0) - L_{f,0}. \tag{5c}$$

The temperature T_0 is a reference temperature at which the specific internal energy of dry soil and liquid water are zero, and $L_{f,0}$ is the specific latent heat of fusion at T_0; for consistency with the CliMA atmosphere model, we choose T_0 as the triple-point temperature of water (Yatunin et al., 2025). As is common for liquids and solids, we have equated the specific internal energies and enthalpies of liquid water and ice (consistent with taking their densities to be constant), so that the specific latent heat of fusion appears in the internal energy definition in place of the specific internal energy difference between the phases.

For the soil model, it is convenient to use internal energies per unit volume $\left(\mathrm{J\ m^{-3}}\right)$ rather than per unit mass. The internal energy per unit volume can be written as the mass-weighted mean

$$\begin{aligned}\rho e(T,\boldsymbol{\theta};\boldsymbol{\nu}) &= \theta_{ds}\rho_{ds}e_{ds}(T;\boldsymbol{\nu}) + \theta_l\rho_l e_l(T) + \theta_i\rho_i e_i(T)\\ &= \rho_s c_s(\boldsymbol{\theta};\boldsymbol{\nu})(T - T_0) - \theta_i\rho_i L_{f,0}.\end{aligned} \tag{6}$$

Here, ρ_s is the soil density, the mass-weighted mean of the constituent densities neglecting gases,

$$\rho_s(\boldsymbol{\theta};\boldsymbol{\nu}) = (1-\nu_p)\rho_{ds}(\boldsymbol{\nu}) + \theta_l\rho_l + \theta_i\rho_i, \tag{7}$$

with porosity ν_p and with the dry soil particle density ρ_{ds} in turn being the mass-weighted mean of the dry-soil constituents. Similarly, the specific heat capacity $c_s(\boldsymbol{\theta};\boldsymbol{\nu})$ of soil is the mass-weighted mean of those of the constituents,

$$c_s(\boldsymbol{\theta};\boldsymbol{\nu}) = \frac{(1-\nu_p)\rho_{ds}c_{ds}(\boldsymbol{\nu}) + \theta_l\rho_l c_l + \theta_i\rho_i c_i}{\rho_s}, \tag{8}$$

where the dry soil specific heat capacity c_{ds} likewise is a mass-weighted mean of the dry-soil constituents. Note that the volumetric internal energy is equal to zero when $T = T_0$ and all water is in the liquid phase.

The conservation law for internal energy is defined by the first law of thermodynamics assuming an open system (e.g., Longo et al., 2019; Walko et al., 2000):

$$\frac{\partial \rho e}{\partial t} = -\nabla\cdot(-\kappa\nabla T + \boldsymbol{D}) + S_{\rho e}. \tag{9}$$

On the right-hand side are the divergences of fluxes and source/sink terms. These include the conductive heat fluxes $-\kappa\nabla T\ (\mathrm{W\ m^{-2}})$, where $\kappa = \kappa(\boldsymbol{\theta};\boldsymbol{\nu})\ (\mathrm{W\ m^{-1}\ K^{-1}})$ is the thermal conductivity of soil, which we take to be isotropic (scalar) and parametrized as in Appendix A2; and energy fluxes $\boldsymbol{D}\ (\mathrm{W\ m^{-2}})$ carried by moving water,

$$\boldsymbol{D} = \rho_l e_l \boldsymbol{d}_l, \tag{10}$$

with $\boldsymbol{d}_l$ given by Darcy's law (1). The source/sink term $S_{\rho e}$ includes changes in energy, for example, due to root extraction of liquid water, subsurface runoff, metabolic heat from microbes, etc. Surface runoff of liquid water contributes to an energy flux at the boundary, as does radiation, which we assume does not penetrate into the soil and hence is described via a boundary flux.

2.1.2.1. Temperature

Given the internal energy ρe and composition $(\boldsymbol{\theta};\boldsymbol{\nu})$ of soil, the definition (6) of the internal energy can be inverted to give the soil temperature:

$$T = T_0 + \frac{\rho e + \theta_i\rho_i L_{f,0}}{\rho_s c_s(\boldsymbol{\theta};\boldsymbol{\nu})}. \tag{11}$$

Using internal energy rather than temperature as a prognostic variable is advantageous because internal energy, unlike temperature, is conserved when water freezes or thaws. Thus, internal energy is unaffected by phase transitions of water and remains continuous at freezing/thawing fronts.

2.1.3. Freezing and Thawing

We represent the liquid-ice conversion rate F_T as a relaxation of liquid and ice fractions toward their respective thermodynamic equilibrium values θ_l^* and θ_i^*:

$$F_T = \frac{\rho_l(\theta_l - \theta_l^*)}{\tau_{\mathrm{FT}}} = -\frac{\rho_i(\theta_i - \theta_i^*)}{\tau_{\mathrm{FT}}}. \tag{12}$$

Here, τ_{FT} is a timescale for a model grid cell to reach thermodynamic equilibrium.

For the thermodynamic equilibrium fractions $\theta_l^*(T,\rho\Theta)$ and $\theta_i^*(T,\rho\Theta)$, which depend on temperature T and the total water content $\rho\Theta \equiv \rho_l\theta_l + \rho_i\theta_i$, we use the approximations derived by Painter et al. (2016):

$$\theta_l^* = \psi^{-1}\left[\psi\left(\frac{\rho\Theta}{\rho_l}\right) + \frac{L_{f,0}}{g}\log\left(\frac{T}{T_f^*}\right)\mathcal{H}\left(T_f^* - T\right)\right], \tag{13a}$$

$$\theta_i^* = \left(\rho\Theta - \rho_l\theta_l^*\right)/\rho_i, \tag{13b}$$

where

$$T_f^* = T_f \exp\left(\frac{g\psi\left(\rho\Theta/\rho_l\right)}{L_{f,0}}\right) \tag{13c}$$

represents a freezing-point temperature shifted by pressure effects and $\mathcal{H}(\cdot)$ is the Heaviside function. These shifts are typically less than a degree but can be larger for fine-textured soils.

To derive this, one applies the Gibbs-Duhem relation to each phase, yielding a more general Clausius-Clapeyron relation that accounts for the matric potential of the unsaturated soil, which accounts for capillary forces that do not act on the ice and air; therefore, ice and liquid water do not occupy the same point in the temperature-pressure phase diagram. The derivation also assumes that the liquid pressure of freezing soils has the same functional form as the soil water retention curve, relating capillary pressure to water content in unfrozen soils (Dall'Amico, 2011; Kurylyk & Watanabe, 2013; Painter & Karra, 2014). This implies that the residual water fraction is always in the liquid form, even for very cold soil. Finally, it also approximates the specific latent heat of fusion near the freezing point by the value $L_{f,0}$ at the triple-point reference temperature T_0, which lies within 0.01 K of the freezing point at sea level pressure. Experiments by Watanabe et al. (2011) confirm that the equilibrium moisture content is close to that given by Equation 13. Above the freezing point T_f^*, the partitioning reduces to $\theta_i^* = 0$ and $\theta_l^* = \rho\Theta/\rho_l = \theta_l$.

The relaxation timescale τ_{FT} is finite because even when the freezing point is reached on the grid scale, additional thermal diffusion or other energy fluxes are required to add or remove the latent heat of fusion. Moreover, unresolved subgrid-scale temperature variations across a cell lead to different portions undergoing the phase transition at different times, effectively smearing out the transition in time for the grid cell as a whole. The timescale of thermal diffusion across a grid cell of vertical extent Δz, assuming negligible horizontal temperature gradients relative to vertical ones, is

$$\tau_{\mathrm{diff}} = \frac{\rho_s c_s \Delta z^2}{\kappa}. \tag{14}$$

We approximate $\tau_{\mathrm{FT}} = \tau_{\mathrm{diff}}$.

2.1.4. Heterotrophic Respiration

Optionally, the soil model includes a balance equation for carbon dioxide in the soil pore space to model heterotrophic respiration and to predict the flux of CO_2 at the soil surface. Expressing the concentration of CO_2 as the carbon mass per volume of soil ρ_{C} $\left(\mathrm{kg\ m^{-3}}\right)$, the evolution equation is

$$\frac{\partial \rho_{\mathrm{C}}}{\partial t} = \nabla \cdot \left(D_{\mathrm{C}}(T,\boldsymbol{\theta};\boldsymbol{\nu})\nabla\rho_{\mathrm{C}}\right) + \mathcal{S}_{\mathrm{C}}, \tag{15}$$

where $D_{\mathrm{C}}(T,\boldsymbol{\theta};\boldsymbol{\nu})$ is the diffusivity $\left(\mathrm{m^2\ s^{-1}}\right)$ of CO_2 gas, which depends on the soil temperature, water content, and type, and $\mathcal{S}_{\mathrm{C}}$ represents volumetric sources/sinks of CO_2 $\left(\mathrm{kg\ m^{-3}\ s^{-1}}\right)$ due to microbes, which we represent with the dual Arrhenius and Michaelis-Menton approach (Davidson et al., 2012); this term depends on soil moisture and oxygen levels (see Appendix A6 for details). The boundary flux at the surface of the soil $F_{\mathrm{C,sfc}}$

depends on the values of D_C, ρ_C at the top of the soil as well as the atmospheric value of ρ_C just above the soil. This, along with the soil carbon mass per volume in the pore space ρ_C, can be compared with observations and used as an indirect measure of heterotrophic respiration.

Additional gas constituents (water vapor, oxygen) could be modeled in a similar fashion but are not included in the current version of ClimaLand. As noted, oxygen in the pore spaces affects respiration rates, but it is currently prescribed. Autotrophic respiration is also modeled, but treated as part of the canopy model (Section 2.3).

2.1.5. Surface Boundary Conditions and Sources

At the soil surface, the boundary condition for the volumetric liquid water flux is given by

$$\boldsymbol{F}_{\mathrm{w}} = I\hat{\boldsymbol{z}}, \tag{16}$$

where $\hat{\boldsymbol{z}}$ is the local vertical unit vector and

$$I = F_{w,\mathrm{sfc}} - R_s \tag{17}$$

is the infiltration flux $\left(\mathrm{m\ s^{-1}}\right)$, defined as the available liquid water volume flux at the surface $F_{w,\mathrm{sfc}}$ less the surface runoff R_s (Appendix A5). The available water flux is defined in terms of water mass fluxes $\left(\mathrm{kg\ m^{-2}\ s^{-1}}\right)$ as

$$\rho_l F_{w,\mathrm{sfc}} = (1-\sigma)(E_l - P_{\mathrm{liq}}) - \sigma\rho_l R_{\mathrm{snow}}. \tag{18}$$

Here, $P_{\mathrm{liq}} > 0$ is the liquid precipitation rate represented as a mass flux of water, E_l is the evaporation rate of liquid water mass (Appendix E2), $\rho_l R_{\mathrm{snow}} > 0$ is the meltwater flux from snow, and σ is the snowcover fraction. All snow-related fluxes are discussed in Section 2.2. Note that if snow melts or sublimates within a timestep, the excess snow liquid water volume flux is included in the soil boundary condition to conserve mass. Precipitation is either prescribed from meteorological forcings or supplied by an atmosphere model.

For volumetric internal energy, the flux boundary condition vector is given by

$$\boldsymbol{F}_{\mathrm{energy}} = \left[(1-\sigma)\left(-e_{l,\mathrm{rain}}\left(P_{\mathrm{liq}} - \rho_l R_s\right) + H + LE - R_n\right) + \sigma G\right]\hat{\boldsymbol{z}}, \tag{19}$$

where $e_{l,\mathrm{rain}}$ is the specific internal energy of rain (assumed to be the same as for the surface runoff), H is the sensible heat flux, LE is the latent heat flux (Appendix E2), R_n is the net radiation (Appendix D1; positive if warming), G is the soil-snow heat flux, and we have ignored transmission of radiation through the snowpack to the soil surface. If snow melts or sublimates within a timestep, the excess snow energy flux is included in the soil boundary condition to conserve energy. All energy fluxes are in $\mathrm{W\ m^{-2}}$.

The volumetric source terms for water $\left(\mathrm{s^{-1}}\right)$ are sublimation of the surface layer of ice $\mathcal{S}_{\mathrm{subl}}$ (Section 2.1.5.1), subsurface runoff $\mathcal{S}_{ss,w}$ (SA5), and root extraction $\mathcal{S}_{re,w}$ (Section 2.3.1). We also include energy sources $\left(\mathrm{W\ m^{-3}}\right)$ due to root extraction $\mathcal{S}_{re,e}$ and subsurface runoff $\mathcal{S}_{ss,e}$. Therefore, ClimaLand currently has the source/sink terms

$$\mathcal{S}_{\vartheta_l} = \mathcal{S}_{re,w} + \mathcal{S}_{ss,w}, \tag{20}$$

$$\mathcal{S}_{\theta_i} = \mathcal{S}_{\mathrm{subl}}, \tag{21}$$

$$\mathcal{S}_{\rho e} = \mathcal{S}_{re,e} + \mathcal{S}_{ss,e}, \tag{22}$$

in the conservation equations for water, Equation 4, and for energy, Equation 9.

2.1.5.1. Ice Sublimation

Sublimation from soil ice is computed as a mass flux E_i in a manner analogous to how we compute the evaporation of soil liquid water E_l. We define the sublimation source $\mathcal{S}_{\mathrm{subl}}$ as

$$S_{\text{subl}} = -\frac{E_i}{\rho_i \Delta z_{\text{sfc}}}(1-\sigma)\mathcal{H}(z - z_{\text{sfc}} + \Delta z_{\text{sfc}}), \tag{23}$$

where $z_{\text{sfc}}(\boldsymbol{x})$ (m) is the surface elevation, defined by the topography, and, as a reminder, $z - z_{\text{sfc}}$ is negative by definition (z increases upward).

2.1.6. Bottom Boundary Conditions

We support multiple options for bottom boundary conditions of the soil domain, including zero-flux and free-drainage boundary conditions. The former applies for water at bedrock. However, in global simulations, ClimaLand does not make use of bedrock depth data yet—the domain has the same depth everywhere—nor does the model account for lateral flow yet (but the numerical core already supports lateral flow). Consequently, in the global simulations presented here, we set a free-drainage boundary condition for water with a corresponding flux boundary condition for energy (equal to the energy flux created by water leaving the domain).

2.2. Snow Model

Snow is modeled using a bulk approach with conservation equations for water mass and internal energy, similar to the Utah-Energy-Balance (UEB) model (Tarboton et al., 1996), and with a diagnostic equation for surface temperature. The snow depth, which affects the surface temperature and the rate at which meltwater leaves the snowpack, is modeled using a neural ordinary differential equation as described in Charbonneau et al. (2025). Variables and parameters of the snow model are listed in Table 2.

2.2.1. Conservation Laws

The conservation law for snow water mass per unit snow area arises from vertically integrating the partial differential equation for snow density and ignoring the time derivatives of snow depth. The resulting conservation law is usually expressed in terms of the snow water equivalent S, the equivalent volume of liquid water per unit ground area. The snow water mass per unit snow cover area is then $\rho_l S/\sigma$, where σ is the snow cover fraction. The snow water equivalent satisfies, at each location on the land surface,

$$\rho_l \frac{\partial S}{\partial t} = P_{\text{snow}} + \sigma\left(P_{\text{liq}} - E - \rho_l R_{\text{snow}}\right). \tag{24}$$

Here, P_{snow} and P_{liq} are precipitation in solid and liquid forms (positive by convention), E is the vapor flux (positive if toward the atmosphere) and $\rho_l R_{\text{snow}}$ is a flux due to percolation of water through and subsequent loss at the bottom of the snowpack (positive by convention). In this equation, all fluxes are mass fluxes.

The snow internal energy per unit ground area, U, is

$$U = \rho_l S\left[c_{\text{snow}}(T - T_0) - \left(1 - q_l\right)L_{f,0}\right], \tag{25}$$

where $c_{\text{snow}} = c_i\left(1 - q_l\right) + c_l q_l$ is the specific heat capacity of snow, with q_l the bulk liquid water mass fraction and $1 - q_l$ the frozen water mass fraction, and T the bulk snow temperature.

The snow energy per ground area satisfies, at each horizontal location,

$$\frac{\partial U}{\partial t} = e_{i,\text{snow}} P_{\text{snow}} + \left(-\rho_l e_l R_{\text{snow}} + e_{l,\text{rain}} P_{\text{liq}} + G - H - LE + R_n\right)\sigma. \tag{26}$$

This can be derived by vertically integrating the partial differential equation for snow volumetric internal energy, ignoring the time derivatives in snow depth, and $e_{i,l}$ is the internal energy per mass for water or ice, as defined in Equation 5. In standalone runs, we estimate the internal energy per unit mass of the solid precipitation as $e_{i,\text{snow}} = -L_{f,0}$, and approximate $e_{l,\text{rain}} = 0$. In coupled runs, the energy fluxes from precipitation are provided by the atmosphere model in order to conserve energy. We also have the sensible and latent heat fluxes H and LE

Table 2
Key Variables and Parameters in the Snow Model

Prognostic variables	Description	Units	Definition	Range/typical value
U	Internal energy per unit ground area	J m^{-2}	Equation 25	$U \sim \mathcal{O}(-10^8 \text{J m}^{-2})$
S	Snow water equivalent	m^3 m^{-2}		$0 \le S$
S_l	Snow liquid water per ground area	m^3 m^{-2}		$0 \le S_l \le S$
z_{snow}	Snow depth	m		$S \le z_{\text{snow}}$
Functions of prognostic state	**Description**	**Units**	**Definition**	**Range/Typical value**
T	Snow bulk temperature	K	Equation 29	$T \le 273.15$ K
T_{sfc}	Snow surface temperature	K	Equation E6	$T \le 273.15$ K
q_l	Liquid water mass fraction per snow mass		Equation 28	$0 \le q_l \le 1$
c_{snow}	Specific heat capacity	J kg^{-1} K^{-1}	Equation in text after Equation 25	$c_i \le c_{\text{snow}} \le c_l$
ρ_{snow}	Snow density	kg m^{-3}	Equation 30	$0 < \rho_{\text{snow}} \le \rho_l$
κ_{snow}	Snow thermal conductivity	W m^{-1} K^{-1}	Equation B6	$\kappa_{\text{air}} < \kappa_{\text{snow}} \le \kappa_{\text{ice}}$
Empirical properties	**Description**	**Units**	**Obtained from**	**Typical value**
$K_{\text{sat, snow}}$	Saturated hydraulic conductivity of snow	m s^{-1}	Free parameter	$\mathcal{O}(10^{-4})$
α_{snow}	Albedo of snow in shortwave		Free parameter	$0 < \alpha_{\text{snow}} \le 1$
ϵ_{snow}	Emissivity of snow in longwave		Free parameter	0.97
$\theta_{l,c}$	Residual volumetric liquid water content		Free parameter	$0 \le \theta_{l,c} \le 1$

(positive if toward the atmosphere), net radiation R_n (positive if warming snow), the energy loss due to liquid water loss $(\rho_l e_l R_{\text{snow}})$ and G, the soil/snow energy flux, which is positive if warming the snow.

The fraction of the snow water equivalent in liquid form $S_l = q_l S$ satisfies, at each horizontal location,

$$\rho_l \frac{\partial S_l}{\partial t} = \left(P_{\text{liq}} - E_l + \rho_l M - \rho_l R_{\text{snow}}\right)\sigma, \tag{27}$$

where $\rho_l M$ is the mass flux for liquid water arising from phase change.

If, in a given timestep, the snow mass is predicted to become negative due to the melt or sublimation rates, we set S and U to zero and we take the remainder of the fluxes and add them to the soil boundary conditions to conserve mass and energy.

The parameterizations for the mass flux resulting from phase change $\rho_l M$, the runoff of liquid water from the snowpack $\rho_l R_{\text{snow}}$, the ground heat flux G, and the snow cover fraction σ are described in Appendix B. The radiation calculation is described in Appendix D1. The expressions for latent heat flux, sensible heat flux, and vapor fluxes are provided in Appendix E1.

2.2.2. Temperature and Liquid Mass Fraction

The temperature T is a diagnostic variable, obtained from the three prognostic variables S, S_l, and U. First, the liquid mass fraction q_l is solved for as

$$q_l = \frac{S_l}{S + \delta S}, \tag{28}$$

where the small constant δS is included for numerical stability when S approaches zero. By solving the definition of internal energy (Equation 25) for temperature, we find

$$T = T_0 + \left[\frac{U + \rho_l S(1 - q_l) L_{f,0}}{\rho_l c_{\text{snow}}(S + \delta S)}\right]. \tag{29}$$

Note that this expression alone does not guarantee that the snow temperature is strictly less than or equal to the freezing temperature; however, the melt term $\rho_l M$ approximately achieves that.

2.2.3. Snow Depth

The snow depth is required to compute heat fluxes throughout the snowpack. These are driven by gradients in temperature; in a bulk model, they arise due to differences in temperature between the surface of the snow and the bulk, or between the bulk and the underlying soil. Density is required to compute the thermal conductivity of the snow, which also affects the heat fluxes within the snowpack. The depth also appears in the liquid water runoff term, $\rho_l R_{\text{snow}}$ (Appendix B). While the snow model needs both depth and density, only one (in addition to S) is required to compute the other, as

$$\rho_{\text{snow}} z_{\text{snow}} = \rho_l S. \tag{30}$$

To model the evolution of snow depth, we use the neural network based snow depth model of Charbonneau et al. (2025). In this model, the snow depth evolves according to the differential equation

$$\frac{dz_{\text{snow}}}{dt} = g(\ldots, S, z_{\text{snow}}), \tag{31}$$

where g is a feed-forward neural network. The function g was trained offline using SNOTEL observations of local environmental conditions, as well as the instantaneous S and z_{snow}. Though trained with site-level data, the model learns a single set of coefficients which, in principle, are location independent. In offline simulations, it was shown to outperform the Snow17 model (Anderson, 1976) at depth predictions when provided with the observed S (Charbonneau et al., 2025).

ClimaLand also supports a simpler single parameter density model, where $\rho_{\text{snow}} = \rho_{\min} q_l + (1 - q_l)\rho_{\text{liq}}$, and $\rho_{\min} < \rho_l$ is the minimum density of the snow, a free parameter. In this case, snow depth is derived from density and snow water equivalent and not modeled using Equation 31.

2.3. Canopy Model

The canopy model consists of eight sub-modules: radiative transfer, photosynthesis, autotrophic respiration, stomatal conductance, canopy energy, plant hydraulics, water stress, and solar induced fluorescence. Within these sub-modules, we currently support the following:

- Radiative transfer: Beer-Lambert law (for a bulk canopy) or the vertically resolved two-stream scheme of Sellers (1985), as modified by Braghiere et al. (2021);
- Photosynthesis: Farquhar model (Farquhar et al., 1980) or the optimality-based P-model (Mengoli et al., 2022; H. Wang et al., 2017), which is based on the Farquhar model but posits that plants find an optimal balance between transpiration (water loss) and carbon assimilation. In both cases, rates are modified by a moisture stress factor;
- Autotrophic respiration: D. Clark et al. (2011);
- Stomatal conductance: Medlyn conductance model (Medlyn et al., 2001) or the conductance as predicted by the P-model (must be paired with P-model photosynthesis);
- Solar induced fluorescence: Lee et al. (2015);
- Canopy temperature: prescribed and equal to the air temperature, or prognostic;
- Plant hydraulics: prognostic water content, following D. M. Lawrence et al. (2019);
- Moisture stress: computed as a function of leaf water potential (Duursma & Medlyn, 2012; Tuzet et al., 2003), or as a function of soil moisture (Egea et al., 2011).

At the top of the canopy, we compute exchanges of water, energy, and carbon with the atmosphere. At the bottom of the canopy, we compute exchanges of water and energy with a prescribed or prognostic ground model.

Therefore, the canopy can be run entirely in standalone mode (prescribed ground model) or as part of a land model integrated with a prognostic soil and/or snow model. Radiative transfer is described in Appendix D2; the P-model is defined in Appendix C, and other parameterizations are detailed in the works cited above. Note that several previous studies employed an early hyperspectral prototype of CliMA Land (v0.1) that integrated shortwave radiative transfer, trait-based optical properties, solar-induced chlorophyll fluorescence, and photosynthesis and energy fluxes (Braghiere et al., 2021; Y. Wang et al., 2021, 2023, 2024, 2025) within a standalone multi-layer canopy model. The hyperspectral modules are maintained in a separate repository (https://github.com/CliMA/Land).

2.3.1. Plant Hydraulics

The plant hydraulics equations govern the volumetric water content along the water flow path within the canopy, including the flux of water from the soil to the roots, which is computed dynamically. The equations are similar to those of CLM5 (D. M. Lawrence et al., 2019) but without the assumption of steady state.

The volumetric liquid water content is denoted by ϑ, and the scalar volume flux of water per plant cross-sectional area by q. The water flow path is assumed to be aligned with the local Cartesian $\hat{z}$ coordinate. We also assume that the plant has an area basis $\mathcal{A}(z)$ (areal cross section per unit ground area) which varies along the flow path, and possibly in time. Then the conservation law for water volume is

$$\frac{\partial \mathcal{A}\vartheta}{\partial t} = -\frac{\partial (q\mathcal{A})}{\partial z}, \tag{32}$$

where we approximate q with Darcy's law as

$$q = -K(\psi)\frac{\partial h}{\partial z}, \tag{33}$$

where K is the conductivity $\left(\mathrm{m\ s^{-1}}\right)$ and $h = \psi + z\,(\mathrm{m})$ is the hydraulic head. We use a Weibull relationship for conductivity with constant parameters along the flow path,

$$K(\psi) = \begin{cases} K_{\mathrm{sat}} \exp\left[-\left(\dfrac{\psi}{\psi_{63}}\right)^{c}\right] & \vartheta < \nu_p \\ K_{\mathrm{sat}} & \vartheta > \nu_p, \end{cases} \tag{34}$$

and a linear retention curve, likewise with constant parameters along the flow path,

$$\psi(\vartheta) = \begin{cases} \dfrac{1}{a_1}(\vartheta/\nu_p - 1) & \vartheta < \nu_p \\ \dfrac{1}{a_2}(\vartheta/\nu_p - 1) & \vartheta > \nu_p. \end{cases} \tag{35}$$

Here, $K_{\mathrm{sat}}\left(\mathrm{m\ s^{-1}}\right)$ is the vegetation saturated conductivity, $\psi_{63}\,(\mathrm{m})$ and c are parameters of the Weibull function, ν_p is the porosity of the vegetation, and a_1 and $a_2\left(\mathrm{m^{-1}}\right)$ are the inverse slopes of the retention curve. The modeling of $\vartheta > \nu_p$ is primarily used for numerical purposes.

At the top of the canopy, the boundary condition is the canopy transpiration (per ground area). The formula for this is given in Appendix E3. At the bottom of the domain, we compute the flux from root extraction per cross sectional area $q = q_{\mathrm{re}}$ as

$$q_{\mathrm{re}} = -\int_{z_{\mathrm{min}}}^{z_{\mathrm{sfc}}} P(z) \int_{z}^{z_{\mathrm{sfc}}} \frac{\partial q}{\partial z'} dz' dz \tag{36}$$

where $P(z)$ is the root distribution function (satisfying $\int_{-\infty}^{z_{\text{sfc}}} P(z)dz = 1$) and where z_{min} is the bottom soil layer of the simulation. The inner integral over z' computes the flux between the soil at z and the vegetation at the surface z_{sfc}, along the root path, and this is then weighted by the root density at that z. We then sum these contributions over all soil depths.

The sink term of the soil is in terms of a volumetric fraction change, that is, a volume of water per volume of soil per second. This is obtained as

$$S_{re,w}(z) = \mathcal{A}(z_{\text{sfc}})\, P(z) \int_{z}^{z_{\text{sfc}}} \frac{\partial q}{\partial z'} dz'. \tag{37}$$

The sign change occurs in the expression for $S_{re,w}$ because a positive value of q indicates flow from the soil to the plant. This is a sink term for the soil.

The corresponding energy sink term is

$$S_{re,e}(z) = e_l(z) S_{re,w}(z), \tag{38}$$

where $e_l(z)$ is the specific internal energy of liquid water in the soil at z.

In practice, we discretize this following a finite difference scheme into two layers, a stem and a leaf. We thus require a root area index $\mathcal{A}_{\text{roots}}$ (for computing $\mathcal{A}_{\text{sfc}}$), stem area index $\mathcal{A}_{\text{stem}}$, and leaf area index $\mathcal{A}_{\text{leaf}}$. We use harmonic averages for interpolating between the cell center and the cell face. We also ignore the time derivative of $\mathcal{A}(z)$ appearing in Equation 32.

2.3.2. Canopy Energy

The equation for a bulk canopy temperature T can be derived by starting from a multi-layer canopy model and integrating in the vertical. Ignoring time derivatives in the canopy height and the area indices, and ignoring phase changes in canopy water, this reduces to

$$(\rho ch)\mathcal{A}_{\text{canopy}} \frac{dT}{dt} = R_n - LE - H + \int_{z_{\text{min}}}^{z_{\text{sfc}}} S_{re,e}(z)dz, \tag{39}$$

where ρch is the bulk specific heat of the canopy per unit area, $\mathcal{A}_{\text{canopy}} = \mathcal{A}_{\text{leaf}} + \mathcal{A}_{\text{stem}}$ is an above ground bulk area index, LE is the latent heat flux between the canopy and the atmosphere (Appendix E3), H is the sensible heat flux between the canopy and the atmosphere (Appendix E3), R_n is the net absorbed radiation in the canopy (Appendix D2), and the last term is the energy flux due to extraction of water from soil by roots (required to conserve energy).

3. Numerical Implementation

ClimaLand is written in the Julia programming language (Bezanson et al., 2017) and hosted in a public GitHub repository (Deck et al., 2025). The Julia programming language was chosen for being high-level, yet performant and flexible. One of the goals of the CliMA project is to expand access and ease of use of climate models and facilitate the incorporation of ML components; the streamlined installation and use provided by Julia contribute to that goal.

3.1. Modularity

The ClimaLand model has been designed from the start for modularity. ClimaLand can be run as a single column, suitable for modeling at a flux tower site; in a defined region bounded by latitude and longitude, suitable for regional simulations; or at a global scale, suitable for integration into a full ESM. Beyond spatial flexibility, ClimaLand supports various sets of equations. For example, one can simulate only soil, only snow, or only the canopy by running a component model in standalone mode. One can also choose to simulate different combinations of these components, including combining all of them in a fully integrated LSM. Examples of some of

these various configurations will be demonstrated in Section 4. Furthermore, ClimaLand supports modularity in terms of parameterizations, allowing for multiple options for soil hydraulics, photosynthesis, and snow physics.

This scientific flexibility is enabled by the software's design, which leverages Julia's multiple dispatch capability —a feature where the behavior of a function is determined by the type of its input arguments. This approach is central to how the model constructs its prognostic state vector (the set of variables to be predicted). A function, `prognostic_vars`, is called on each component model (e.g., snow, soil), which is identified by a unique type. Dispatch automatically selects the correct method to list that component's prognostic variables. This process is hierarchical, allowing an integrated model to query its sub-components, which in turn query their specific parameterizations. For example, a neural network snow model, with its own type, adds snow depth (z_{snow}) to the state vector and provides a method for computing its time derivative (dz_{snow}/dt) by evaluating the network. Once the complete state and its derivative function are assembled, the core time-stepping algorithm handles the time evolution, remaining entirely agnostic to the scientific components included in the model.

3.2. Performance Portability

ClimaLand is also designed for performance portability, allowing the same code to run efficiently on both CPUs and GPUs, including in parallel across multiple cores. For the user, switching between these hardware architectures requires minimal configuration. This is possible because all architecture-specific code, including data structures and communication protocols, is abstracted away from the science model and managed by two backend CliMA modules: `ClimaComms.jl` and `ClimaCore.jl`. This separation allows the core ClimaLand algorithms to remain hardware-agnostic, ensuring both flexibility and performance without requiring specialized code from the model developer.

3.3. Temporal and Spatial Discretization

ClimaLand is able to access multiple timestepping algorithms, as implemented in the ClimaTimeSteppers.jl package. Soil liquid water content, soil internal energy, and canopy temperature are treated implicitly in time, while all other variables (soil CO_2, soil ice, snow prognostic variables, and canopy water content) are stepped explicitly. Therefore, when running the full integrated land model, or a standalone model with some implicitly-stepped prognostic variables, it is necessary to use an implicit-explicit (IMEX) method to be able to solve both the implicit and explicit components of the model's tendencies (see Appendix F for details). For most such runs, the ARS111 method is used by default, which has 1 implicit stage, 1 explicit stage, and first-order accuracy (also known as forward–backward Euler). When running a standalone model that has only explicitly-stepped prognostic variables, the RK4 algorithm is typically used, which is a Runge–Kutta method with 4 stages and fourth-order accuracy.

In the vertical direction, a finite difference scheme is used to discretize the space. In both global and regional simulations (3D domains), a spectral element spatial discretization is used in the horizontal directions. Note that the land model does not currently include any lateral fluxes, and so the horizontal connectivity of the domain is not used: columns are independent of each other.

3.4. Computational Performance

High computational performance is a requirement for LSMs, because repeated multi-year simulations at one degree resolution (or higher) are required for calibration of models, and because longer simulations are required for scenario-based modeling efforts. This section focuses on the computational performance of ClimaLand on GPUs. The benchmark simulations carried out below do not include diagnostic output or forcing updates. We did not include diagnostics primarily because the time taken for diagnostics depends on the number and frequency of diagnostic output one needs. We did not include forcing because we do not currently have forcing at higher resolutions than 1°, and so we cannot accurately convey how reading in the larger data would impact the performance at the highest resolutions. All tests were performed on Nvidia A100 GPU with 32-bit floating point precision; all multi-GPU simulations were carried out on the Derecho supercomputer (Computational and Information Systems Laboratory, 2023). It is important to note that, despite GPU support being a key feature of ClimaLand, we have not yet focused on extensive code optimization for GPU performance.

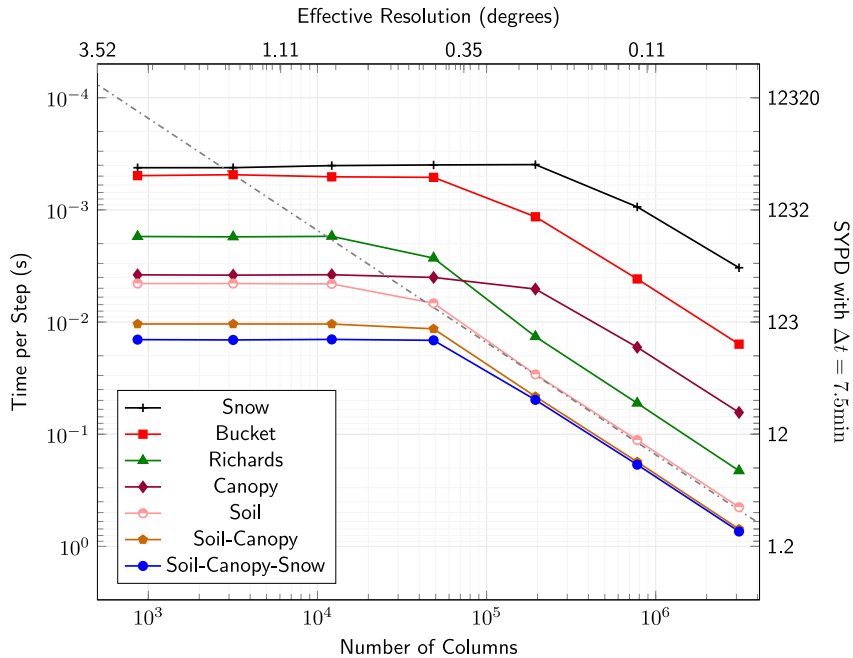

Figure 1. Performance and scaling of the ClimaLand models as a function of the resolution, measured in terms of time to advance the model by one timestep. For a fair comparison, all runs used the same time-stepper (ARS 111, with 3 N iterations) and a fixed timestep. The top x axis shows the equivalent spatial resolution assuming all columns are on land, and the right y axis indicates the number of simulated years per day assuming a timestep of $\Delta t = 7.5$ min. The dash-dotted line shows linear scaling with the number of columns. The performance plateau at low resolution arises when the number of columns is insufficient to fully utilize the graphics processing unit (GPU). With enough columns, the scaling turns to linear. These results demonstrate the feasibility of running high-resolution land-surface simulations on a single modern GPU.

3.4.1. Simulations on a Single GPU

When running the full land model, different components have computational costs that vary by orders of magnitude, and one of the goals of this section is to highlight the relative costs. Given modularity of ClimaLand, we assess performance in terms of the number of columns simulated and the wall time required to complete a timestep. Using the number of columns as a metric abstracts the spatial configuration, as these columns can be part of a global simulation or a focused regional simulation. Furthermore, expressing performance in terms of wall time per timestep allows us to generalize across models that might have different maximum allowed timesteps.

We present benchmark results in Figure 1, illustrating the time required to complete a timestep as a function of the number of simulated (independent) columns for various model combinations. In this evaluation, we began with the fully integrated land model (soil, canopy, and snow) and then executed each component (soil, canopy, or snow) in standalone mode. Additionally, we included two other models: a bucket model and a simpler soil model solving only Richards' equation for a single prognostic variable, ϑ_l. To enhance the interpretability of the plot, we have added two supplementary axes: the top x-axis displays the effective resolution, calculated assuming all columns are located on continents (each column covers $0.3 \times 4\pi/N_{\text{columns}}$ square radians, where 0.3 represents the global land fraction), and the right y-axis indicates the equivalent number of simulated years per day, assuming a timestep of $\Delta t = 7.5$min.

The curves in Figure 1 exhibit two distinct regimes: an initial plateau followed by a linear decrease, as visually highlighted by the dash-dotted linear function. The plateau at lower resolutions reflects the parallel nature of GPUs, where all columns are processed concurrently, and the GPU is not yet fully utilized. However, as the number of columns increases, the scaling transitions to linear. For simpler models like the bucket, canopy, or snow model, this transition from constant to linear scaling occurs at a higher number of columns than for the full model. Profiling analysis reveals that these models are largely latency-bound, meaning the GPU cannot be kept fully utilized with the available workload. This is partly attributed to known inefficiencies in ClimaLand's GPU

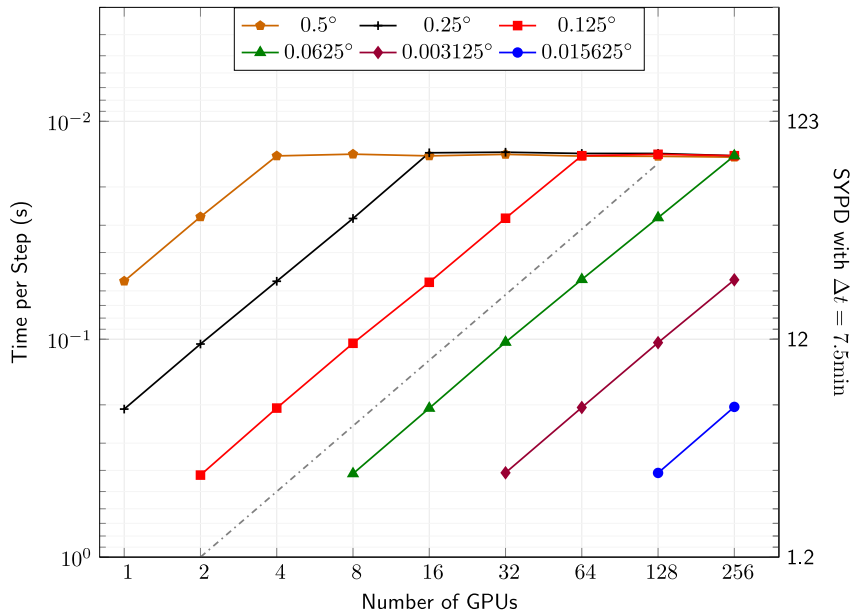

Figure 2. Parallelization performance and strong scaling of the ClimaLand integrated model as a function of the resolution and the number of graphics processing units (GPUs). The *x* axis shows the number of GPUs used, the right *y* axis indicates the number of simulated years per day assuming a timestep of $\Delta t = 7.5$ min, and the left *y* axis shows the wall time per step. For a given number of GPUs, with high enough resolution, the scaling is linear, as indicated by the gray dashed line. These results demonstrate the feasibility of running kilometer-scale land-surface simulations on modern GPUs with ClimaLand. The behavior shown is expected since there is no communication between columns.

memory management, which forces the GPU to wait for new instructions. While this effect is more pronounced in simpler models, it is also present in integrated models.

Figure 1 demonstrates that ClimaLand's full model can process 1–2 million columns per second, highlighting the feasibility of simulating high-resolution land-surface models on a single modern GPU. At these high resolutions, the soil component is the most computationally demanding, consuming over 90% of the total execution time. This is expected, as the soil model involves solving partial differential equations, unlike other components that solve ordinary differential equations. Conversely, the addition of the snow model has a negligible impact on runtime, especially at higher resolution. We also note that the primary constraint on achieving even higher resolutions is ClimaLand's memory footprint, as simulating more than 3 million columns exhausts the GPU memory. While memory optimization has not been a primary focus thus far, there are clear avenues for improvement in this area should one aim to accommodate even more columns on a single GPU. Finally, we tested the full land model on a laptop GPU (Nvidia RTX A2000 Mobile) for a configuration with 10,584 columns, which resulted in a time per step of 0.5 s, demonstrating that it is possible to perform medium-scale simulations on readily available hardware. To provide additional context about the meaning of the report metrics, one step in uELM (Ultrahigh-Resolution E3SM Land Model), with 36,000 grid cells, takes 1 s on one CPU node on Summit and one third of a second on 6 NVidia V100 GPUs (D. Wang et al., 2022).

3.4.2. Simulations on Multiple GPUs

The performance portability of ClimaLand enables simulations on multiple GPUs, which is essential for modeling the land surface at high resolutions. To demonstrate this capability, we present the model's strong scaling performance across various horizontal resolutions in Figure 2.

As shown in Figure 2, for a given resolution, the simulation speedup is linear with the number of GPUs. This implies that doubling the horizontal resolution in each direction (a four-fold increase in columns) requires four times the number of GPUs to maintain the same time-to-solution. The performance plateaus when the number of columns per GPU, and hence the arithmetic intensity, is too low for the hardware to be fully utilized (around 65,000 columns per GPU in this case). This linear scaling, an expected result given the independence of model columns, demonstrates that ClimaLand is capable of kilometer-scale simulations.

4. Model Validation

In this section, we show the results of ClimaLand simulations, compared with observations, in various configurations: in columns and on the globe, with the standalone soil and snow models or the fully integrated ClimaLand model. The primary goal is to demonstrate that the model, when run with realistic parameters, returns realistic results. Tutorials running the following simulations are available (Deck et al., 2025).

4.1. Phase Changes in Soil

4.1.1. Freezing Front: Comparison to Laboratory Data

Mizoguchi (1990) conducted an experiment in which columns of Kanagawa sandy loam were frozen by removing heat from above. The resulting total water content $\theta_i + \theta_l$ was measured as a function of depth at three instances after the experiment began. The data show the presence of a freezing front propagating into the soil, and the phenomenon of cryosuction is observed, where (liquid) water below the front flows above the front, where it then freezes. Many soil models are tested against this data set (Dall'Amico, 2011; Hansson et al., 2004; Painter, 2011).

The columns were 20 cm deep, and the initial volumetric water content was $\theta_l(t = 0, z) = 0.33$, with no water flux at the boundaries. The initial temperature was uniform at 279.85 K. Although the surface temperature was held at 267.15 K via a circulating fluid and the bottom was insulated, other groups have modified these boundaries conditions in order to obtain a better fit to the data (Hansson et al., 2004; Dall'Amico, 2011; Painter, 2011), hypothesizing that some heat loss occurred at the bottom and there was resistance between the fluid and the soil surface. We set a Neumann condition of zero flux at the bottom, and a time-varying surface flux of

$$F(z = 0, t) = 28\ \text{W m}^{-2}\ \text{K}^{-1} \times (T_{\text{soil}}(t) - 267.15\ \text{K})$$

where T_{soil} is value of the soil temperature in the top layer, following Hansson et al. (2004). The hydraulic properties for the van Genuchten functions (see Appendix Equation A17 and A18) were $\nu_p = 0.535$, $\theta_{\text{res}} = 0.05$, $n = 1.48$, $\alpha = 1.11\ \text{m}^{-1}$, and $K_{\text{sat}} = 3.2 \times 10^{-6}\ \text{m s}^{-1}$. Thermal properties were approximated using the figures in Hansson et al. (2004). We set $\nu_{\text{quartz, ss}} = 0.6$, $\nu_{\text{minerals, ss}} = 0$, $\nu_{\text{om, ss}} = 0.4$, and $\nu_{\text{gravel, ss}} = 0$. The resolution is $\Delta z = 1\text{cm}$.

In Figure 3, we show the comparison of the simulation with data. The agreement is comparable to that found by Dall'Amico (2011), Hansson et al. (2004), Painter (2011), which we would expect, as ClimaLand's soil model is similar to those models in many ways, but we do not have access to goodness of fit metrics from the other models. The mean absolute error for ClimaLand was computed by interpolating the simulation output to the depths at which the measurements were taken; the average value over the entire domain was 0.019 m^3m^{-3} after 12 hr, 0.013 m^3m^{-3} after 24 hr, and 0.018 m^3m^{-3} after 50 hr. Importantly, the phenomenon of cryosuction is observed: liquid water is drawn toward and above the freezing front from deeper in the soil, where it then freezes. This results in a higher water content above the front than below it, as shown in Figure 3.

We also compared the model to an analytic result for soil freezing, obtained under the assumption of $K_{\text{sat}} = 0$. These results are presented in Appendix A7.

4.1.2. Bare Soil Evaporation

Lehmann et al. (2008) present experiments of a sand column undergoing evaporation. The sand starts fully saturated, so that the Phase I evaporation is at the maximum rate, but it undergoes a transition to Phase II evaporation after the surface becomes sufficiently dry. The data from the experiment was provided to us by the authors and is shown in Figure 4, along with the predicted evaporation from two ClimaLand simulations. The simulations are identical except for the vertical resolution. Both evaporation as a function of time and as a function of water mass loss are shown.

To carry out the simulations, the standalone soil model was used (i.e., without a canopy or snow model). Physical parameters for the soil were taken from Lehmann et al. (2008) with some modifications as follows, or else representative values for sand were used. We used $\nu_p = 0.43$, $\theta_{\text{res}} = 0.043$, $n = 8.91$, $\alpha = 3\ \text{m}^{-1}$, $K_{\text{sat}} = 2.6 \times 10^{-6}\ \text{m s}^{-1}$, and $\nu_{ss,\text{quartz}} = 1.0$. The values of α and n are similar to those inferred from the

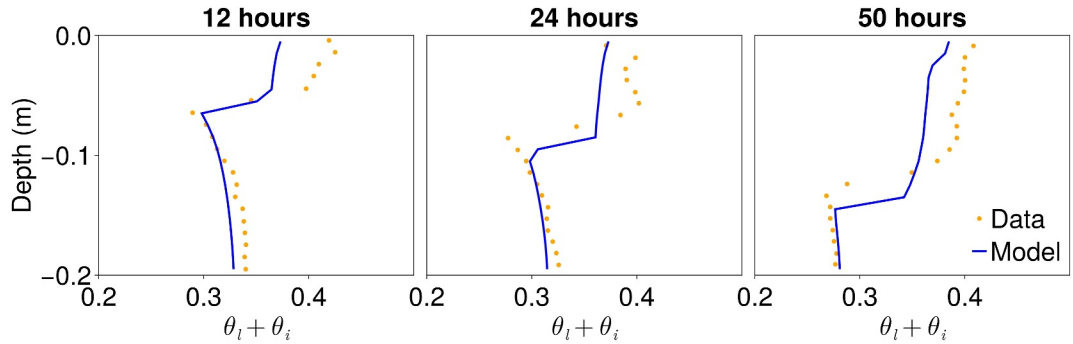

Figure 3. Comparison of Climate Modeling Alliance simulation (blue) with laboratory data (orange) of a column undergoing freezing from above. The left, middle, and right panels show the total volumetric water content as a function of depth at 12, 24, and 50 hr after the experiment started, with a mean absolute error in each case of 0.019 m^3m^{-3}, 0.01 $3m^3m^{-3}$, and 0.018 m^3m^{-3}, respectively. See text for a discussion.

measured air entry potential and capillary width in Lehmann et al. (2008), 7.4 m^{-1} and 6.34, respectively; however, these parameters have a large effect on the timing of the transition so they have been adjusted for better agreement with the data. To compute the turbulent fluxes, we set $z_{0,m} = z_{0,b} = 0.01$ m (see Appendix E2 for definitions). The relative humidity and temperature of the air were taken as reported, air pressure was set to atmospheric pressure, and we adjusted wind speed in order to match the potential evaporation rate at the start of the experiment. The domain was 0.35 m deep, as reported, and we used 28 or 7 equally sized layers. The timestep was 900 s for both simulations. The choice to run the simulation with two resolutions was in order to test the resolution-dependence of the evaporation scheme. As seen in Figure 4, both simulations accurately reproduce a smoothed version of the measured evaporation rate, both as a function of time and a function of mass loss. The mean absolute error for the high resolution simulation is 0.41 mm day^{-1}, while for the low resolution simulation it is 0.48 mm day^{-1}; the Kling-Gupta efficiency for the high-resolution simulation is 0.79, and for the low resolution simulation it is 0.74.

The evaporation scheme that we implement (Lehmann et al., 2018) has a smooth dependence on soil moisture. This is different from a model such as Swenson et al. (2012), which has a "switch" in terms of a critical soil moisture, above which the evaporation is at the maximal rate (soil resistance is zero), and below which the evaporation rate steadily drops to a constant (soil resistance ramps up to constant). We found that the latter was better able to match the experimental data under study here. However, it performed worse compared to our adopted schemes in global runs. It is likely that calibration of the "switch" type model would enable it to perform better; the Lehmann et al. (2018) model we adopted has no free parameters. This requires further study.

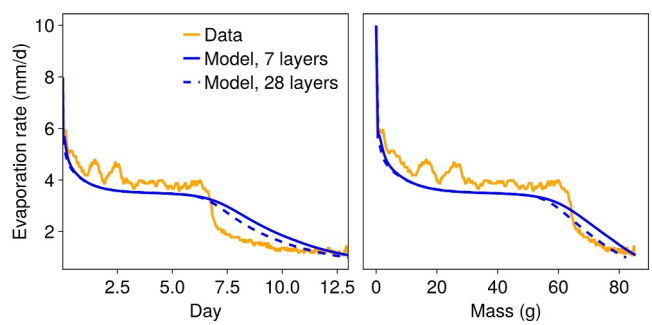

Figure 4. Comparison of two ClimaLand standalone soil simulations (blue) with laboratory data (orange) of a column undergoing evaporation. The dashed blue line and solid blue line indicate different model resolutions, as noted in the legend. Left panel: evaporation rate as a function of time. Right panel: evaporation rate as a function of mass loss. The mean absolute error for the high resolution simulation is 0.41 mm day^{-1}, while for the low resolution simulation it is 0.48 mm day^{-1}; the Kling-Gupta efficiency for the high-resolution simulation is 0.79, and for the low resolution simulation it is 0.74.

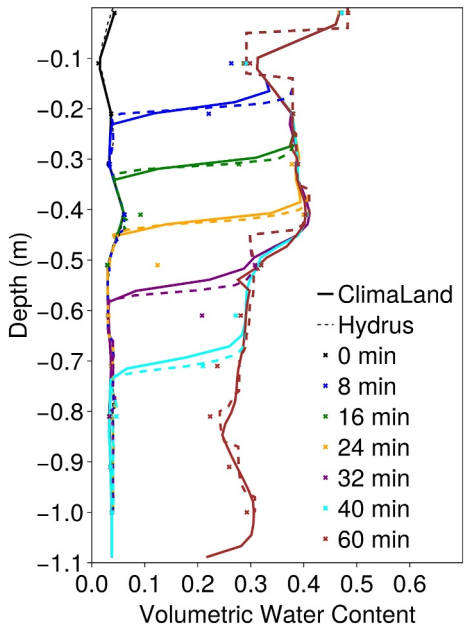

Figure 5. Comparison of a ClimaLand soil model simulation, a Hydrus-1D simulation, and measured data from an infiltration experiment at a field site in Alberta, Canada, in a multi-layered coarse textured soil column. The different colors represent different times since the start of the experiment. The points mark data, solid lines mark Climate Modeling Alliance simulation results, and dashed lines represent the Hydrus-1D results. This demonstrates that the ClimaLand model can adequately represent the flow of water in multi-layered soils.

4.2. Site-Level Data

4.2.1. Layered Soil: Fort McMurray Field Site

As reported by Zettl (2011), infiltration experiments were carried out at several field sites near Fort McMurray in Alberta, Canada. During each experiment, a constant ponding height between 5 and 10 cm was maintained, and the resulting soil moisture as a function of depth into the soil was measured at 10 cm intervals, every 4 min, down to 1.1 m. Once the experiment was complete, each site was excavated, and soil properties (ν_p, θ_{res}, n, α, K_{sat}) were determined at 2–10 cm intervals (Zettl, 2011). The values of K_{sat} were refined in a followup study by Huang (2011), which used simulations of the sites using the code Hydrus-1D (Simunek et al., 2005). The final set of soil hydraulic parameters, including the optimized K_{sat}, is reported in Table 1 of Huang (2011).

We obtained the data and Hydrus-1D simulation results for site SV62 from the authors (Barbour, L., and Huang, M.) and published them (Deck et al., 2025). Using the soil properties for each layer (16 in total) and the boundary conditions reported in Huang (2011) (constant head at the top of the profile, free drainage at the bottom), we carried out an infiltration simulation with a 1.1 m deep heterogeneous soil column. The initial conditions were taken from a linear interpolation of the data at the start of the experiment; all measured soil parameters are also linearly interpolated to the model grid.

Results are shown in Figure 5. The simulation yields similar results as the Hydrus-1D simulation; both differ from the data in the sharpness of the wetting front. This was noted in Huang (2011) and hypothesized to be due to preferential flow, which is not modeled well by a single-porosity model. Although the soils at this site were all coarse-textured, these results indicate that the model can be used to model infiltration realistically in finely layered soils, given representative van Genuchten parameters and K_{sat}.

4.2.2. Col de Porte

Col de Porte is an alpine snow monitoring site located in the French Chartreuse mountains at an elevation of 1325 m with decades of in situ data (Lejeune et al., 2019); the data from the site have been used as part of snow model intercomparison projects (Krinner et al., 2018; Ménard et al., 2019). We modeled 3 years of snow evolution (2010–2012, inclusive) using the ClimaLand standalone snow model. This version of the model ignores interactions with soil and approximates the ground heat flux (Appendix Equation B3) as $G = 0$. The snow albedo is set fixed to the median value observed at the site over the 3 years, and initial conditions for S, z_{snow}, and T are taken from the measurements, assuming $S_l = 0$.

Figure 6 compares the model results for snow water, snow depth, and surface temperature with observations at the site. As mentioned in Section 2.2.2 and Appendix B1, the evolution of the three prognostic variables does not guarantee that the temperature is below (or equal to) the freezing point. While the phase change flux acts to melt snow when the temperature is predicted in an Euler step to be larger than the freezing temperature, our actual time-stepper is not forward Euler, and excursions above the freezing point are observed. This will require future study. Despite this, for this particular location during these years, the uncalibrated ClimaLand snow model realistically simulates the snow state as a function of time. The mean absolute error is 0.07 m (SWE), 0.11 m (depth), and 4.2° K (temperature). Simulated snow water equivalent has a high bias, presumably because the model approximates the snow surface temperature with the bulk temperature. This would lead to delayed melting (observable in spring) as well as less melting overall, which in turn leads to less loss of water from the snowpack.

4.2.3. US-Var Fluxtower Site

The Ameriflux US-Var site is located at the Vaira ranch in the lower elevation foothills of the Sierra Nevada mountain range (Ma et al., 2007). The vegetation consists mostly of C3 grasses up to 0.5 m high; the soil is 0.5 m

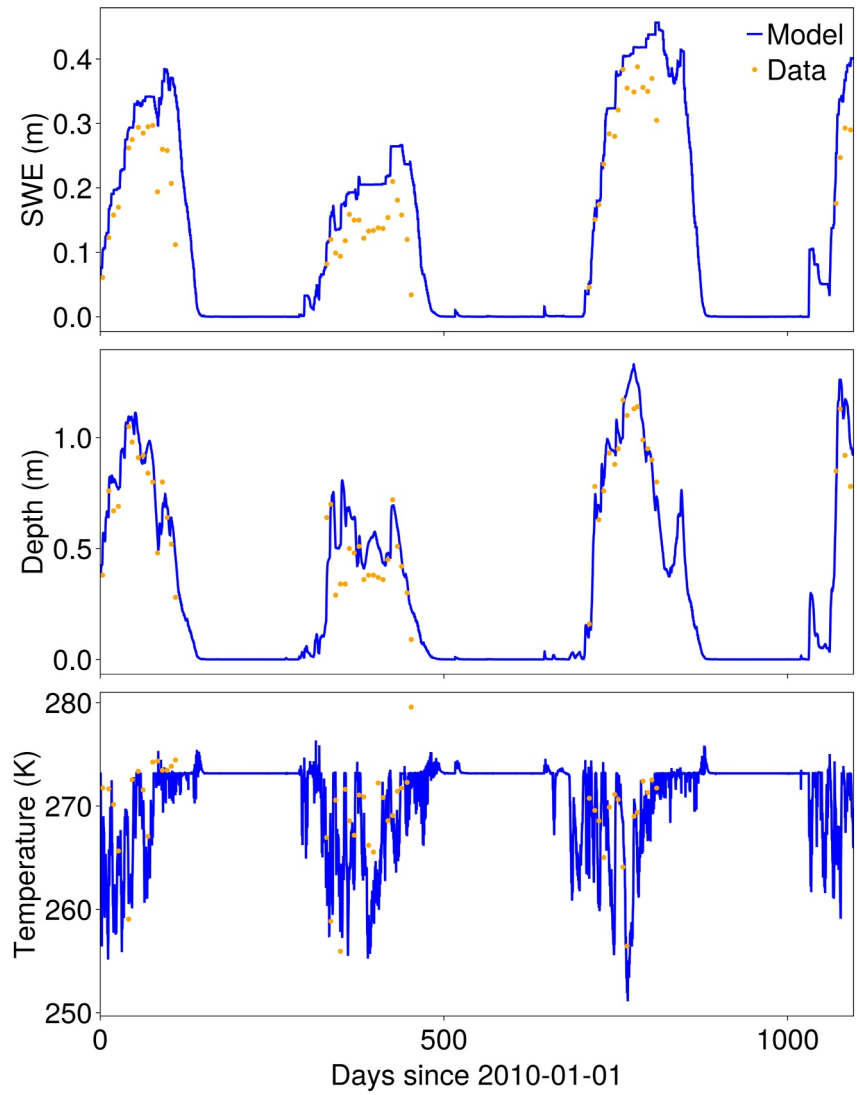

Figure 6. Comparison of a ClimaLand snow model simulation (blue) forced with meteorological data from Col de Porte, and compared to data (orange) measured at the site. Top: snow water content as a function of time, middle: snow depth as a function of time, bottom: snow temperature as a function of time. The mean absolute error is 0.07 m (SWE), 0.11 m (depth), and 4.2°K (temperature). See text for discussion.

deep rocky silt loam, with composition 30% sand, 57% silt, and 13% clay (Xu & Baldocchi, 2004). We simulated the soil and vegetation at this location using ClimaLand during the year 2003; meteorological forcings were provided in the Ameriflux data. Owing to the local climate at the site, the total reported precipitation was assumed to be rain.

For the soil, we used parameters of $\nu_p = 0.5$, $\theta_{\text{res}} = 0.0$, $n = 1.6$, $\alpha = 2\ \text{m}^{-1}$, $S_s = 10^{-3}\ \text{m}^{-1}$, and $K_{\text{sat}} = 1.25 \times 10^{-6}\ \text{m s}^{-1}$. The soil albedo in the NIR and PAR bands was taken to be independent of moisture, with a value of 0.35 for both; the emissivity was 0.98, and the roughness lengths were $z_{0m} = z_{0b} = 0.01$ m.

We used $g_1 = 166\ \text{Pa}^{1/2}$, $\text{V}_{\text{cmax}} = 2.5 \times 10^{-5}\ \text{molCO}_2\ \text{m}^{-2}\text{s}^{-1}$, and $\rho ch_{\text{canopy}} = 745\ \text{J m}^{-2}\ \text{K}^{-1}$. Radiative parameters for the vegetation were a clumping index $\Omega = 0.75$, a leaf angle distribution parameter of 0.5, an albedo in the PAR band of 0.11, an albedo in the NIR band of 0.45, a transmissivity in the PAR band of 0.05, a transmissivity in the NIR band of 0.34, and an emissivity of 0.97. Finally, for the plant hydraulics portion of the code, we used moisture stress parameters of $p_c = -3 \times 10^5\ \text{Pa}, s_c = 4 \times 10^{-6}\ \text{Pa}^{-1}$, hydraulic conductivity parameters of $K_{\text{sat, canopy}} = 2 \times 10^{-8}\ \text{m s}^{-1}$, $\psi_{63} = -275$ m, and a Weibull exponent of 4. For the retention

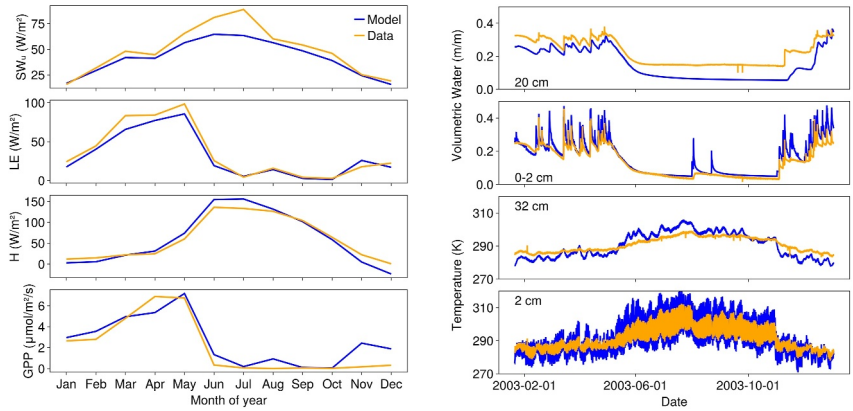

Figure 7. ClimaLand simulations (blue) at the US-Var Fluxtower site compared with observations made in 2003 (orange). Left: monthly averages of fluxes, right: soil water content and temperature at different depths. Note that outgoing longwave radiation is not measured at the site, so only the upwelling shortwave radiation is shown. For the monthly fluxes, we found a mean absolute error of 7 W m^{-2} (SW_u), 11 W m^{-2} (H), 6 W m^{-2} (LE), and 0.8 μmol m^{-2} s^{-1} (gross primary productivity.

curve, we used $a_1 = 4.9 \times 10^{-4}$ m^{-1} and $a_2 = 9.8 \times 10^{-5}$ m^{-1}. The plant porosity was $\nu_p = 8.9 \times 10^{-3}$. For the surface fluxes, we set $z_{0m} = 0.13 h_{\text{canopy}}$, with $h_{\text{canopy}} = 0.5$ m, and $z_{0b} = 0.1 z_{0m}$. A timestep of 10 min was used, with a vertical resolution at the surface of 3.5 cm.

Figure 7 compares the simulation results with the Ameriflux Fluxnet data provided by Ma et al. (2023). For the soil moisture, we found that no single set of van Genuchten parameters, porosity, and residual fraction was able to produce good fits at both the surface and at 20 cm depths; soil parameters that change with depth may be more realistic. Soil temperature also proved challenging to match; the simulations generally show larger variation compared to the data. As discussed in Appendix E3, we model the turbulent fluxes from vegetated surfaces as a weighted average of fluxes from the canopy, bare soil, and bare snow. For heavily vegetated surfaces, this approximation will be poorer, but, as this site is a grassland, we expect it to be adequate. For the monthly fluxes, we found a mean absolute error of 7 W m^{-2} (SW_u), 11 W m^{-2} (H), 6 W m^{-2} (LE), and 0.8 μmol m^{-2} s^{-1} (gross primary productivity).

4.3. Global Simulation

The ClimaLand model was used to simulate soil, snow, and canopy dynamics globally while forced by prescribed atmospheric conditions. Atmospheric conditions at 1° resolution included wind speed at 10 m, dewpoint temperature at 2 m, air temperature at 2 m, air pressure at 2 m, downwelling short and long wave radiative fluxes, and precipitation (split into rain and snow) obtained from ERA5 reanalysis data for the years 2000–2020 (Hersbach et al., 2023). Although the forcing data was available every hour, the actual atmospheric conditions applied to ClimaLand at each timestep are computed using a linear interpolation between the conditions at intervals of 3 hours. The leaf area index was prescribed using a linear interpolation of MODIS data from the same time period (Myneni et al., 2015; Y. Wang et al., 2022), with a native 8-day interval. Linear interpolation in space was used to create the forcing on the model grid.

The land model itself requires parameter maps of spatially varying (but assumed temporally constant) soil and vegetation properties. These maps are interpolated to 1° resolution. The soil van Genuchten parameters, porosity, residual water content, and saturated conductivity were obtained from Gupta et al. (2020, 2022). Soil composition data was obtained from SoilGrids (Hengl et al., 2017). The parameters for the TOPMODEL runoff scheme were derived from a high-resolution map of topographic index (Marthews et al., 2015). The specific storativity was set to a global constant of 0.001 m^{-1}. A map of soil color was obtained from P. J. Lawrence and Chase (2007); the methodology for combining soil colors and soil parameters was obtained from Braghiere, Wang, et al. (2023). All vegetation parameters maps were obtained from the Community Land Model surface parameters files (D. M. Lawrence et al., 2019) by determining the dominant plant functional type (PFT) at each ClimaLand simulation coordinate and using the vegetation parameter of that PFT; these parameters are described further in Appendix G.

Canopy height is treated as a constant, stem area index was set to zero, and root area index is treated as a constant, $\mathcal{A}_{\text{roots}} = 1$.

The ClimaLand simulation domain used has a resolution of 1° in each horizontal direction, with 15 layers in the vertical. The top layer width is 5 cm, the domain depth is 50 m, and the layer widths increase with depth such that the bottom layer is 10 m in width. The simulation was advanced in time using an ARS111 time-stepper with a timestep of 7.5 min; three N iterations were carried out during the implicit solve for soil water, soil energy, and canopy temperature.

Calibration of land models has been a persistent challenge due to the number of free parameters within land models and due to the cost of forward model runs. Ensemble Kalman inversion and variants thereof have been developed and tested for solving inverse problems with computationally expensive forward models and noisy data, without requiring derivatives (Dunbar et al., 2022; Huang, Schneider, & Stuart, 2022; Iglesias et al., 2013; Kovachki & Stuart, 2019; Schneider et al., 2022). With these approaches, multiple members of an ensemble simulate observations using the forward model, with different parameter values, and compare those with real observations in a loss function. At the next iteration, a new set of parameter values is used, based on both the prior distributions for these parameters as well as cross covariances among ensemble members, which approximate derivatives of the loss function. A pipeline for calibrating the global ClimaLand model using ensemble Kalman inversion methods is included in version 1 of ClimaLand. This pipeline was used to calibrate a subset of the free parameters of the ClimaLand model. For this example, we used the P-model parameterization for photosynthesis and stomatal conductance, the piecewise soil moisture stress function, and a snow albedo parameterization, which depends on zenith angle and snow bulk density.

4.3.1. Parameter Calibration

We calibrated the P-model parameters $c^\star$, a unitless cost factor for electron transport, and β, the unitless ratio of carboxylation to transpiration costs at standard temperature. Priors for these parameters were taken from reported values of $c^\star = 0.41 \pm 0.11$ (H. Wang et al., 2017) and $\beta = 141 \pm 2.7$ (Stocker et al., 2020); these values have also been used in global simulations with the P-model using the land model Noah-MP Ren et al. (2025). We calibrated the parameters of the snow albedo model depending on zenith angle and bulk snow density (Appendix D1). Additionally, we calibrated the exponent c appearing in the soil moisture stress parameterization β_m, using a prior centered on a value of 1, corresponding to a linear increase from completely stressed to unstressed as soil moisture θ_l increases from the residual (θ_r) to saturated (ν_p) water fractions, as

$$\beta_m = \min\left[\int \left[\frac{\theta_l - \theta_r}{\nu_p - \theta_r}\right]^c P(z)dz, 1\right] \tag{40}$$

where $P(z)$ is the root distribution function (Section 2.3.1). The calibration used a loss function proportional to the mean squared error of the latent heat flux summed with a mean squared error of the upwelling shortwave flux. To be more precise, the loss was computed as

$$\mathcal{L}(\theta) \propto \sum_t \int d\lambda \int \cos\phi d\phi \left[\left(\text{LE}(\lambda,\phi,t;\theta) - \widehat{\text{LE}}(\lambda,\phi,t)\right)^2 + \left(\text{SW}_u(\lambda,\phi,t;\theta) - \widehat{\text{SW}}_u(\lambda,\phi,t)\right)^2\right], \tag{41}$$

where $\text{LE}(\lambda,\phi,t;\theta)_i$ is the model output of latent heat flux at longitude λ and latitude ϕ averaged over a 3 month period starting at any time t corresponding to a day of year $\in$ (1 December, 1 March, 1 June, 1 September), θ indicates the model parameters, and $\widehat{\text{LE}}(\lambda,\phi,t)$ indicates the same quantity but from ERA5 reanalysis. In this example, all observables at all locations in space are treated equally in the loss function, that is, the covariance matrix is proportional to the identity matrix. Eight iterations of Unscented Kalman Inversion (Huang, Schneider, & Stuart, 2022) were carried out, each consisting of a 27 month simulation, with the first 3 months discarded as spinup. For example, iteration 0 used the data from 1 September 2000 to 31 November 2002, with the first 3 months discarded. The final completed iteration used the same days of the year, for the years of 2014 and 2016. We assessed that calibration had converged by monitoring the value of the loss function. These simulations were carried out on the Derecho supercomputer (Computational and Information Systems Laboratory, 2023)

Table 3
Calibrated Values for the P-Model (Photosynthesis and Stomatal Conductance Parameterization), Snow Albedo Parameterization, and Soil Moisture Stress Parameterization

Parameters	Calibrated value	Parameterization
$c^\star$	0.30	P-model
β	192	P-model
c	0.47	Piecewise soil moisture stress
α_0	0.59	Snow albedo
$\Delta\alpha$	0.40	Snow albedo
k	1.96	Snow albedo
β_0	0.97	Snow albedo

Note. All parameters are unitless. Details of how the values were calibrated are given in the main text.

using 16 A100 GPUs (for 16 ensemble members), and took 12 hr to complete. The resulting parameter values are given in Table 3.

4.3.2. Results

Figure 8 shows the simulated net radiative fluxes at the land surface, the sensible heat flux at the land surface, and the latent heat flux at the land surface. In each case, we show the global map of the fluxes averaged over the period 1 March 2017–1 March 2020, the global map of the corresponding ERA5 variable averaged over the same period, and the mean difference between the two (ClimaLand—ERA5). The root mean squared error and bias computed using the time averages of the fluxes are $(10.3, -0.8)$ W m^{-2} (SW$_n$), $(9.0, 5.4)$ W m^{-2} (LW$_n$), $(11.1, -1.9)$ W m^{-2} (LE), and $(14.7, 7.8)$ W m^{-2} (H). The Bowen ratio in many parts of the world is incorrect (opposite sign bias in H and LE), and overall the land surface is too bright, and not emitting enough longwave radiation. The errors seen are due to a combination of structural error in the model and remaining uncalibrated parameters.

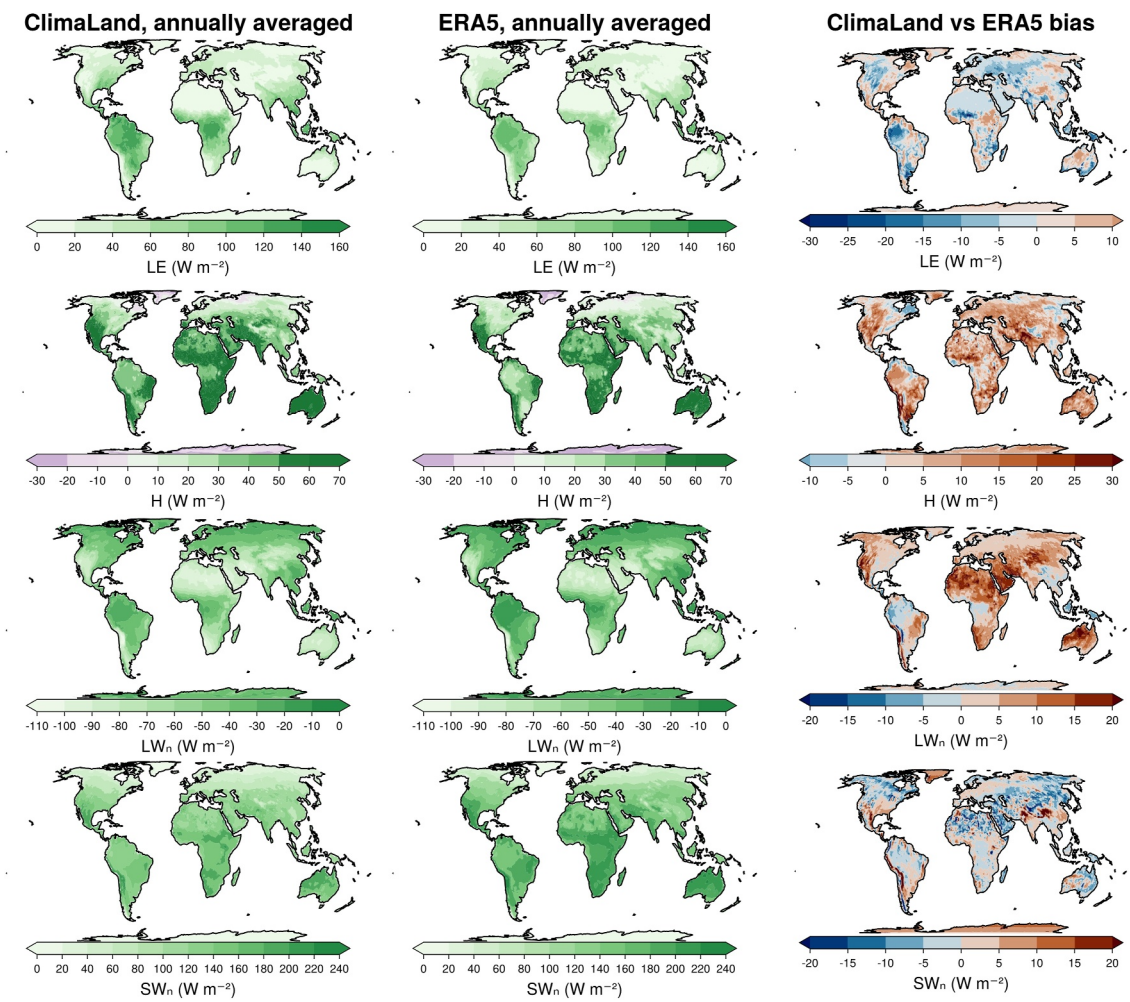

Figure 8. Global output from ClimaLand compared with ERA5. Left: ClimaLand annual average of latent heat, sensible heat, net longwave radiation, and net shortwave radiation (by row). Middle: ERA5 annual average of the same fluxes (by row). Right: Bias (ClimaLand—ERA5) in global averages for the same fluxes (by row). See text for discussion.

5. Discussion

ClimaLand is designed for flexibility. The model can be run with each component in standalone mode, for example, the canopy model with prescribed soil state, or with components integrated into a single land model, taking into account interactions between the components. This allows the use of the ClimaLand code for simulating textbook examples, highly controlled lab experiments, site-level studies (e.g., flux tower locations), and global land dynamics.

5.1. Enabling Data-Driven Parameterizations

The choice of the Julia programming language was originally motivated by Julia's unique (at the time) feature of combining high-level code, including easy integration with ML libraries, speed, and portability to different computer architectures. Enabled by the latter, ClimaLand also supports a wide range of computing resources from single CPUs to multi-GPU setups, possibly distributed across multiple nodes. The superior energy efficiency of GPUs compared to CPUs makes them a more sustainable choice for computationally intensive tasks.

As demonstrated by ClimaLand's snow depth model, it is straightforward to integrate parameterizations based on neural networks into ClimaLand. These may be trained offline, if data exist for the specific process of interest, as for the snow depth model (Charbonneau et al., 2025). However, it is also possible to calibrate these parameterizations online, that is, by carrying out forward runs and simulating an observed flux. Other areas where such approaches can be taken to improve LSMs include land-use changes (e.g., J. Wang et al., 2022), river-modeling (e.g., Lima et al., 2024), stomatal conductance parameterizations (e.g., ElGhawi et al., 2023; Saunders et al., 2021), and subgrid contributions to runoff modeling. Calibration of seven parameters of ClimaLand using Ensemble Kalman Inversion was described in Section 4.3.1; these calibration runs took collectively 12 hr on 16 A100 GPUs, running in parallel. The same approach can be used for calibration of ML model parameters within ClimaLand.

5.2. Hydrological Modeling

M. P. Clark et al. (2015) reviewed the hydrological processes on land that affect large-scale energy, water, and carbon fluxes and which are not typically represented well in LSMs; the authors suggested specific model improvements that LSMs could implement to address these shortcomings. Here we discuss areas where ClimaLand has incorporated these suggestions and where there is clear room for future work.

Saturated zone modeling: Many LSMs use a form of Richards' equation that does not permit modeling of saturated soils. ClimaLand solves the mixed-form of Richards' equation using the augmented liquid fraction (Endrizzi et al., 2014; Woodward & Dawson, 2000), allowing it to simulate saturated soils. ELM is another land model which solves for variably-saturated soil (Bisht, Riley, Hammond, & Lorenzetti, 2018).

Groundwater flow modeling and lateral flow: LSMs generally solve Richards' equation in the vertical direction only; alternatively, ELM is an example of an LSM which solves for three dimensional flow (Bisht, Riley, Wainwright, et al., 2018; Qiu et al., 2024). Although simulations with ClimaLand currently set lateral flow to zero, so that columns are independent of each other, the underlying numerical discretization of the equations is 3D. By building ClimaLand with this spatial discretization to begin with, the infrastructure for interconnectivity is already established, allowing for future implementation of lateral energy and water flow in the soil.

Improving evapotranspiration: ClimaLand explicitly resolves hydraulic gradients within the soil-plant-atmosphere system based on D. M. Lawrence et al. (2019) and treats canopy water content as a prognostic variable.

Planned improvements include improving modeling of soil moisture by accounting for macropores and preferred paths in soil with a dual-porosity model (Gerke & Van Genuchten, 1993). Additionally, the runoff scheme will be an active area of development, as it lacks realistic stream-aquifer interactions, subgrid lateral flow, and surface moisture effects on infiltration capacity (M. P. Clark et al., 2015; Entekhabi & Eagleson, 1989). This is an area where data-driven (machine learned) models may play a central role.

5.3. Vegetation Modeling

Traditional LSMs rely on empirical parameterizations, often using fixed plant functional types (PFTs), to represent vegetation traits (G. B. Bonan et al., 2002). While this approach simplifies the parameter space, it limits the ability to capture the dynamic adaptation of vegetation to climate variability. Moreover, variability within a single PFT often exceeds variability between different PFTs, leading to inaccuracies in model predictions (Cranko Page et al., 2024; Y. Wang et al., 2025).

Recent advances, such as optimality-based modeling frameworks (Franklin et al., 2020; Harrison et al., 2021), offer an alternative by predicting plant traits as emergent properties of environmental constraints, improving model generalization. These models enable vegetation to adjust key physiological traits, such as V_{cmax}, in response to temperature, light, and water availability (H. Wang et al., 2017; Stocker et al., 2020). These approaches have shown success in simulating seasonal variations in leaf area index (Cai et al., 2025; Norton et al., 2023) and stomatal conductance, as demonstrated in an early version of the ClimaLand model (Y. Wang et al., 2021). Another benefit of such models is that they eliminate the need for high-resolution parameter maps, which may not always be available. ClimaLand currently implements an option for an optimality-based model for V_{cmax} and g_1, based on Stocker et al. (2020). Future developments in ClimaLand will reduce reliance on static PFT classifications using optimality models where applicable, while implementing learned functions of environmental conditions in other cases.

Recent research also highlights the importance of spectral surface reflectance for radiative transfer (Braghiere, Wang, et al., 2023). Most LSMs approximate reflectance using broadband values, but Braghiere, Wang, et al. (2023) showed that this introduces biases, affecting energy fluxes and photosynthesis. ClimaLand will incorporate hyperspectral surface albedo to improve energy balance simulations and better integrate remote sensing data (Braghiere et al., 2021).

5.4. Carbon Pools

Terrestrial carbon is primarily distributed between two major pools: live biomass (comprising vegetation, where carbon constitutes approximately half of the dry mass) and dead carbon (including soil organic matter, dead plant material, and dissolved organic and inorganic carbon) (IPCC, 2023). The net carbon flux on land—the balance between carbon uptake and release—is determined by changes in these pools over time. Vegetation carbon in LSMs is further subdivided into functionally distinct compartments, including leaf, stem, root, and non-structural carbohydrates (Fatichi et al., 2019). Net Primary Productivity (NPP) is allocated among these compartments according to species-specific strategies, with each component exhibiting its own turnover rate. For soil carbon, key inputs include leaf litter and root turnover, while the dominant output is microbial respiration (Tao et al., 2023).

These processes are important to model because the global state of carbon pools is sensitive to climate change and land use change. The current scientific consensus is that the land has been acting as a sink of carbon (Friedlingstein et al., 2025), but it remains unclear if this will persist in the future. Currently, ClimaLand prescribes these carbon pools, but future iterations will incorporate prognostic carbon pools, enabling dynamic responses to environmental changes. This transition will improve model accuracy by allowing carbon stocks to evolve in response to climate drivers and land-use changes.

6. Conclusions

We have presented the first version of ClimaLand, a new LSM. ClimaLand simulates the dynamics of water, energy, and carbon fluxes on land and their interaction with the atmosphere. It is designed to overcome the limitations of many existing LSMs by leveraging GPUs for computational efficiency and by providing a modular architecture that enables integration with ML libraries. The development of ClimaLand was motivated by the need for a computationally efficient LSM that can be used to simulate the land surface at high resolution and to incorporate ML parameterizations. Its modularity also lends itself well to scientific experimentation and exploration and makes the model easily extensible, for example, through the incorporation of optimality models.

The soil model simulates flow and phase changes of water in both saturated and unsaturated zones. The snow model is a bulk model that simulates the accumulation and melt of snow, including an ML model for snow thickness. The canopy model is a modified big-leaf model that simulates the exchange of water, energy, and

carbon between the canopy and the atmosphere. ClimaLand has been validated against benchmarks with known behavior, experimental data, and field data. The model results are in good agreement with the data, demonstrating that ClimaLand can realistically simulate many aspects of Earth's land surface.

Future versions of ClimaLand will include additional processes, such as lateral water and energy transport in the soil and optimality-based vegetation models. The model will also be extended with further ML components and calibrated against observations to improve its accuracy.

Appendix A: A Soil Model

A1. Soil Composition

We take soil to consist of a dry soil matrix (composed of sand, clay, coarse fragments/gravel, etc.) with pore spaces that may contain liquid water, ice, and gases (moist air). Because the primary constituents of soil are essentially incompressible, the composition of soil is commonly expressed in terms of volume fractions (e.g., volume of liquid water per volume of soil), which we denote as $(\theta_l, \theta_i, \theta_g, \theta_{ds})$, for liquid water, ice, gases, and dry soil. These volumetric fractions satisfy

$$\nu_p = \theta_g + \theta_l + \theta_i = 1 - \theta_{ds}, \tag{A1}$$

where ν_p is the porosity of the soil.

The dry soil portion of the bulk soil in turn consists of constituents that can have different thermodynamic and hydraulic properties. Let ν_j indicate the dry volume fraction of constituent χ_j in the total volume of soil, with χ_j labeling the constituents, for example,

$$\chi_j \in \{\text{sand}, \text{clay}, \text{silt}, \text{gravel}, \text{organic matter}, \text{bedrock}\}.$$

We also let $\nu_{ss,j}$ denote the volume fraction of each constituent of dry soil relative to the soil solids only. Then we have

$$\frac{\sum_j \nu_j}{1 - \nu_p} = \sum_j \nu_{ss,j} = 1.$$

The dry soil composition depends on a space coordinate, and it may also vary on long (decadal and longer) timescales. In the ClimaLand model, the volume fractions $\nu_{ss,j}$ of constituents of dry soil are prescribed parameter fields. We denote the set $\{\nu_{ss,j}\}$ as $\boldsymbol{\nu}$.

A2. Thermal Conductivity

The thermal conductivity $\kappa(\boldsymbol{\theta}; \boldsymbol{\nu})$ of soil varies depending on mineral composition, organic matter content, porosity, and the water content of soils. We model the thermal conductivity as a weighted mean of the conductivities of dry soil, κ_{dry}, and water saturated soil, κ_{sat},

$$\kappa = K_e \kappa_{\text{sat}} + (1 - K_e)\kappa_{\text{dry}}. \tag{A2}$$

The weighting factor is the dimensionless Kersten number ($0 \le K_e \le 1$), an empirical function that monotonically increases with soil moisture content (Dai et al., 2019; Farouki, 1981).

Balland and Arp (2005) found the following formulation for the Kersten number through fits to empirical data:

$$K_e = K_{e,\text{unfrozen}} = S_r^{(1+\nu_{ss,\text{om}} - \alpha\nu_{ss,\text{quartz}} - \nu_{ss,\text{gravel}})/2} \times \left(\left[1 + e^{-\beta S_r}\right]^{-3} - \left(\frac{1 - S_r}{2}\right)^3 \right)^{1-\nu_{ss,\text{om}}} \quad \text{if } \theta_i = 0. \tag{A3}$$

$$K_e = K_{e,\text{frozen}} = S_r^{1+\nu_{\text{om},ss}} \qquad \text{if } \theta_i > 0. \tag{A4}$$

Here,

$$S_r = \frac{\theta_l + \theta_i}{\nu_p} \tag{A5}$$

is the relative saturation. The scale parameters $\alpha \approx 0.24 \pm 0.04$ and $\beta \approx 18.1 \pm 1.1$ are adjustable parameters determined on the basis of soil measurements. Dai et al. (2019) found this formulation to perform well in simulating soil temperatures across a range of locations.

This formulation is not continuous at $\theta_i = 0$, and implementing a model that is (e.g., with simpler parametrized models (Painter, 2011) or by combining $K_{e,\text{frozen}}$ and $K_{e,\text{unfrozen}}$ with continuous weights) will be investigated in the future.

The dry thermal conductivity $\kappa_{\text{dry}}(\boldsymbol{\nu})$ is usually estimated on the basis of soil composition data and conductivity models (Dai et al., 2019). In the Balland and Arp (2005) model, it is given as

$$\kappa_{\text{dry}} = \frac{(a\kappa_{\text{solid}} - \kappa_{\text{air}})\rho_b + \kappa_{\text{air}}\rho_{ds}}{\rho_{ds} - (1-a)\rho_b}, \tag{A6}$$

where $a = 0.053$ is a unitless empirical constant,

$$\rho_b = (1 - \nu_p)\rho_{ds} \tag{A7}$$

is the bulk density of the dry soil, and ρ_{ds} is the particle density of the soil.

The thermal conductivity of the soil solids is given by

$$\kappa_{\text{solid}}(\boldsymbol{\nu}) = \kappa_{\text{om}}^{\nu_{ss,\text{om}}} \kappa_{\text{sand}}^{\nu_{ss,\text{quartz}}} \kappa_{\text{gravel}}^{\nu_{ss,\text{gravel}}} \kappa_{\text{minerals}}^{1-\nu_{ss,\text{om}}-\nu_{ss,\text{quartz}}-\nu_{ss,\text{gravel}}}. \tag{A8}$$

The saturated thermal conductivity κ_{sat} is given as a geometric mean of the individual components. Following Balland and Arp (2005),

$$\kappa_{\text{sat,frozen}}(\boldsymbol{\nu}, \theta) = \kappa_{\text{solid}}^{1-\nu_p} \kappa_{\text{ice}}^{\nu_p - \theta_l} \kappa_{\text{liq}}^{\theta_l} \tag{A9}$$

$$\kappa_{\text{sat,unfrozen}}(\boldsymbol{\nu}, \theta) = \kappa_{\text{solid}}^{1-\nu_p} \kappa_{\text{liq}}^{\nu_p}, \tag{A10}$$

which we then combine using a geometric mean as

$$\kappa_{\text{sat}}(\boldsymbol{\nu}, \theta) = \kappa_{\text{sat, unfrozen}}^{\theta_l/\nu_p} \kappa_{\text{sat, unfrozen}}^{\theta_l/\nu_p}. \tag{A11}$$

A3. Pressure Head

The prognostic variables of the soil hydrology model are the augmented liquid fraction ϑ_l and the ice fraction θ_i. The pressure head ψ and the conductivity K need to be expressed in terms of these variables to close the model equations.

In unsaturated soil, the pressure head is equal to the matric potential (suction head) ψ_m, while in saturated soils, the pressure head can be solved for by inverting the definition of the augmented liquid fraction (3), assuming the soil is saturated when $\theta_l = \nu_{\text{eff}}$, where ν_{eff} is the effective porosity defined as

$$\nu_{\text{eff}} = \nu_p - \theta_i. \tag{A12}$$

We define the volumetric liquid fraction as

$$\theta_l = \begin{cases} \vartheta_l & \text{for } \vartheta_l < \nu_{\text{eff}}, \\ \nu_{\text{eff}} & \text{for } \vartheta_l \geq \nu_{\text{eff}}, \end{cases} \tag{A13}$$

and the pressure head becomes

$$\psi(\boldsymbol{\theta};\boldsymbol{\nu}) = \begin{cases} \psi_m(\boldsymbol{\theta};\boldsymbol{\nu}) & \text{for } \vartheta_l < \nu_{\text{eff}}, \\ (\vartheta_l - \nu_{\text{eff}})/S_s & \text{for } \vartheta_l \geq \nu_{\text{eff}}. \end{cases} \tag{A14}$$

This formulation ensures that $\theta_l + \theta_i \leq \nu_p$ and that the pressure change occurs when the soil is full of water (even if some of that water is frozen).

The matric potential $\psi_m(\boldsymbol{\theta};\boldsymbol{\nu})$ as a function of water content in the unsaturated zone remains to be specified. The functional forms of ψ_m and of the hydraulic conductivity $K(\boldsymbol{\theta};\boldsymbol{\nu})$ are related, so a hydraulic conductivity K can be calculated once the matric potential ψ_m is known (Mualem, 1976). These functions typically are arguments of the saturation, defined in terms of the porosity ν_p, volumetric water content θ_l, and $\theta_{\text{res}} = \theta_{\text{res}}(\boldsymbol{\nu})$, the residual water fraction in inaccessible pore spaces:

$$S_l = \left(\frac{\theta_l - \theta_{\text{res}}}{\nu_p - \theta_{\text{res}}}\right). \tag{A15}$$

When ice is present, this definition can be extended to the effective saturation

$$S_{l,\text{eff}} = \left(\frac{\theta_l - \theta_{\text{res}}}{\nu_{\text{eff}} - \theta_{\text{res}}}\right). \tag{A16}$$

The van Genuchten (1980) formulation for the matric potential is

$$\psi_m(x) = -\alpha^{-1} x^{-1/(nm)} \left(1 - x^{1/m}\right)^{1/n}, \tag{A17}$$

where, following Bear (2018) and Painter and Karra (2014), the argument x is set to $S_{l,\text{eff}}$. This is partly practical: the matric potential function evaluated at an argument of unity should be zero, so if we use ν_p in the denominator of the effective saturation, the pressure head would not be continuous at $\nu = \nu_{\text{eff}}$ in partially frozen soil. Additionally, Painter and Karra (2014) argue that capillary pressure $P_{\text{gas}} - P_l$ should be related to the actual available pore space, rather than the pore space that would exist if no ice were present. Therefore, this definition is an attractive choice when ice is present, and if θ_i is zero, it reduces to the standard definition.

In addition to the parameters entering the effective saturation, this depends on two fitting parameters which reflect the soil composition $\boldsymbol{\nu}$: (a) an exponent n, with $m = 1 - 1/n$; and (b) an inverse reference potential $\alpha > 0\ (\text{m}^{-1})$.

A4. Hydraulic Conductivity

A hydraulic conductivity that is consistent with the matric potential Equation A17 is

$$K(T,\theta_i,x) = \Theta(T)\Gamma(\theta_i)\, K_{\text{sat}} \times \begin{cases} x^{1/2}\left[1 - \left(1 - x^{1/m}\right)^m\right]^2 & \text{for } x < 1 \\ 1 & \text{for } x \geq 1, \end{cases} \tag{A18}$$

where

- $K_{\text{sat}} = K_{\text{sat}}(\boldsymbol{\nu})$ is the hydraulic conductivity at saturation with liquid water when no ice is present;
- $\Theta(T)$ is a function that models the temperature dependence of the conductivity; and

- $\Gamma(\theta_i)$ is an impedance factor that may be included to model reduced hydraulic conductivities in frozen soils (Lundin, 1990).

Following Bear (2018), we use as the argument $x = S_l$, that is, the definition of saturation with the porosity ν_p rather than the effective saturation $S_{l,\mathrm{eff}}$, because K_{sat} is given for ice-free soil.

Physically, the hydraulic conductivity is related to the kinematic viscosity of liquid water. Because the viscosity varies with temperature—it decreases by about 65% from 0°C to 20°C—the hydraulic conductivity generally is an increasing function of temperature. To model this temperature dependence, we represent the hydraulic conductivity not only as a function of effective saturation S_l but also as an empirical function of temperature,

$$\Theta(T) = \exp[\gamma(T - T_{\mathrm{ref}})]. \tag{A19}$$

Here,

$$\gamma = \frac{T_1}{(T_2 - T_{\mathrm{ref}})^2} \approx 2.64 \times 10^{-2}\ \mathrm{K}^{-1} \tag{A20}$$

is an empirical factor, with $T_1 = 507.88$ K and $T_2 = 149.3$ K, and T_{ref} is the reference temperature at which $\Theta = 1$. The reference temperature T_{ref} may be taken to be the annual-mean temperature at the site in question when tabulated values for the saturated hydraulic conductivity are used. The value $\gamma \approx 2.64 \times 10^{-2}\ \mathrm{K}^{-1}$ is obtained with $T_{\mathrm{ref}} = 288$ K and implies a 30% increase in hydraulic conductivity for a 10 K temperature increase.

A more general expression for the temperature dependence of the hydraulic conductivity is

$$\Theta(T) \propto \exp\left(-\frac{T_1}{T - T_2}\right)$$

with empirical constants T_1 and T_2. This derives from a standard expression for the temperature dependence of the viscosity of water, neglecting the small $\left(O(10^{-3})\right)$ changes in the density of liquid water over typical soil temperatures. The expression (A19) with the coefficient (A20) comes from a linearization of the exponent around a reference temperature T_{ref}, which should be within the range of typical soil temperatures, so that variations around it are small.

The hydraulic conductivity K decreases as ice fraction increases. Even with $\Gamma = 1$, freezing decreases the volumetric liquid fraction relative to porosity and hence decreases the conductivity (Watanabe & Flury, 2008). This is consistent with treating freezing and drying as hydraulically equivalent. To further reduce the hydraulic conductivity, an empirical impedance factor (Hansson et al., 2004; Lundin, 1990; Swenson et al., 2012)

$$\Gamma(\theta_i) = 10^{-\Omega\theta_i/(\theta_l + \theta_i)} \tag{A21}$$

is sometimes employed, though there is debate about if such a factor is warranted (Kurylyk & Watanabe, 2013). We follow Hansson et al. (2004) and use the impedance parameter $\Omega = 7$.

A5. Runoff Parameterizations

Surface runoff of water occurs when the rate of water influx exceeds the capacity of the soil to take up the water (Margulis, 2017). It is often discussed in terms of saturation excess (Dunne et al., 1975; Hewlett & Hibbert, 1967), when any influx of water on saturated soil is taken to be runoff (zero infiltration) and of infiltration excess (Freeze, 1974; Horton, 1933), when the rate of influx exceeds the rate of soil uptake but some water still infiltrates the soil. The former fits into the latter definition assuming that the infiltration capacity of saturated soil is zero. The challenge for coarse-resolution land models is that the surface soil moisture, which governs the infiltration capacity, has subgrid variations due to local topography, vegetation patterns, soil type, etc (Entekhabi & Eagleson, 1989).

Unresolved lateral subgrid-scale water fluxes, in either the groundwater table or the unsaturated zone, also produce loss of water from the soil (Margulis, 2017); for example, this occurs where the water table intersects the surface, such as at a stream. This is referred to as subsurface runoff and also must be parametrized.

We have currently implemented a runoff parameterization for both surface and subsurface runoff that models subgrid variability in topography (Niu et al., 2005), based on the work of Beven and Kirkby (1979) and Sivapalan et al. (1987). According to Niu et al. (2005), the surface infiltration I and surface runoff R_s are given by

$$I = \left(1 - f_{\text{sat}}\right) \max(F_{\text{w,sfc}}, I_c), \tag{A22}$$

$$R_s = F_{\text{w,sfc}} - I, \tag{A23}$$

where I_c is the infiltration capacity $\left(\text{m s}^{-1}\right)$, which we parameterize as

$$I_c = -K_{\text{sat,sfc}}, \tag{A24}$$

and where

$$f_{\text{sat}} = f_{\text{max}} \exp\left(-f_{\text{over}} d_\nabla / 2\right) \tag{A25}$$

is the saturated fraction of the grid cell, where no infiltration is permitted, $f_{\text{over}} = 3.28\ \text{m}^{-1}$ is a parameter that can be calibrated, and $d_\nabla = z_\nabla - z_{\text{sfc}}$ is the depth to the water table, with $z = z_\nabla$ the location of the water table closest to the surface. The quantity f_{max} is a spatially varying parameter computed from a map of the topographic index at high resolution. It is the maximum saturated fraction of a grid cell, computed from the cumulative distribution of the high resolution topographic index per low resolution grid cell. Improved parameterizations for the infiltration capacity include its dependence on surface moisture (Entekhabi & Eagleson, 1989) but are not yet implemented.

Note that f_{sat} reaches a constant maximum value equal to f_{max} when the grid mean water table reaches the surface. This corresponds to a saturated soil surface at the grid level. Because generally $f_{\text{max}} < 1$, the infiltration I does not necessarily go to 0 as the soil saturates, that is, as $\theta_{l,\text{sfc}} + \theta_{i,\text{sfc}} \to \nu_{p,\text{sfc}}$. This reflects subgrid-scale variations: regions can be drier than the grid mean and hence take up water even if the grid mean is saturated. The mixed-form of Richards' equation is able to account for such effects since the augmented liquid fraction can be larger than porosity.

Following Niu et al. (2005), the subsurface runoff R_{ss} and sink terms $S_{ss,w}$ and $S_{ss,e}$ are parametrized as

$$R_{ss} = R_{sb} \exp\left(-f_{\text{over}} d_\nabla\right), \tag{A26}$$

$$S_{ss,w} = -\frac{R_{ss}}{\max(h_\nabla, \epsilon)} \mathcal{H}(\theta_i + \theta_l - \nu_p), \tag{A27}$$

$$S_{ss,e} = \rho_l e_l S_{ss,w}, \tag{A28}$$

where $R_{sb} = 1.48 \times 10^{-7}\ \text{m s}^{-1}$ is an empirical and adjustable parameter and h_∇ is the height of the water table (m), defined as

$$h_\nabla = \int_{z_{\text{min}}}^{z_{\text{sfc}}} \mathcal{H}(\theta_i(z) + \theta_l(z) - \nu_p(z))\, dz, \tag{A29}$$

and

$$d_\nabla = z_{\text{sfc}} - z_{\text{min}} - h_\nabla, \tag{A30}$$

with z_{min} and z_{sfc} as the z coordinates of the bottom and top (surface) of the soil, respectively.

A6. Heterotrophic Respiration

Carbon dioxide can be produced by respiration and diffuses in the porous air space of the soil. There are two respiration sources for CO_2 in the current version of ClimaLand's soil model: autotrophic root respiration (not described) and heterotrophic microbial decomposition of soil organic carbon; the latter component is computed from the Dual Arrhenius and Michaelis-Menten (DAMM) kinetics model (Davidson et al., 2012), described in what follows.

In the DAMM model, the heterotrophic respiration source term S_C $\left(\text{kgC m}^{-3}\ \text{s}^{-1}\right)$ in Equation 15 depends on two environmental drivers: soil temperature T and soil moisture (θ_l, θ_i). Additionally, it depends on the availability of soluble carbon, $\rho_{C,s}$ $\left(\text{kgC m}^{-3}\right)$ and oxygen, $\rho_{O,s}$ $\left(\text{kgO m}^{-3}\right)$.

The respiration rate S_C is expressed as:

$$S_C = V_{\text{pot}} \cdot \frac{\rho_{C,s}}{kM_{C,s} + \rho_{C,s}} \cdot \frac{\rho_{O,s}}{kM_{O,s} + \rho_{O,s}} \tag{A31}$$

where V_{pot} $\left(\text{kgC m}^{-3}\ \text{s}^{-1}\right)$ is the potential rate of respiration, and the second and third terms represent the availability of soluble carbon and oxygen, respectively. The constants $kM_{C,s}$ $\left(\text{kgC m}^{-3}\right)$ and $kM_{O,s}$ $\left(\text{kgO m}^{-3}\right)$ are the Michaelis constants for substrate and oxygen, respectively. The potential respiration rate V_{pot} is expressed as

$$V_{\text{pot}} = V_{\max} \exp\left(-\frac{E_a}{RT}\right), \tag{A32}$$

where $V_{\max}$ $\left(\text{kgC m}^{-3}\ \text{s}^{-1}\right)$ is the maximum, E_a $\left(\text{J mol}^{-1}\right)$ is the activation energy of the reaction, and R is the gas constant.

The mass per volume of soluble carbon substrates is affected by soil water content, and specifically by diffusion of substrates through soil water films. Using these underlying principles, $\rho_{C,s}$ is parameterized as

$$\rho_{C,s} = f_{C,s}\rho_{C,\text{som}}\tilde{D}_{C,s,\text{liq}}\theta_l^3 \tag{A33}$$

where $\rho_{C,\text{som}}$ $\left(\text{kgC m}^{-3}\right)$ is the total mass per volume of soil organic carbon (prescribed, currently temporally and spatially constant); $f_{C,s}$ is the fraction of soil organic carbon that is soluble; and $\tilde{D}_{C,s,\text{liq}}$ is a dimensionless diffusivity of the soluble carbon in liquid.

The mass per volume of soluble O_2 is calculated as

$$\rho_{O,s} = \tilde{D}_{O,g}\rho_O\theta_g^{4/3}, \tag{A34}$$

where $\tilde{D}_{O,g}$ is a dimensionless diffusivity of O_2 in air, ρ_O is the mass per volume of O_2 in air in soil pores (currently prescribed), and $\theta_g = \nu_p - \theta_l - \theta_i$ is the volumetric fraction of gases in the pore space.

The parameterization for the diffusivity D_C (m s^{-2}) appearing in Equation 15 is taken from Ryan et al. (2018):

$$D_C = D_{C,0} \cdot \left[2\theta_{g,100}^3 + 0.04\theta_{g,100}\right] \cdot \left(\frac{\theta_g}{\theta_{g,100}}\right)^{2+\frac{3}{b}}, \tag{A35}$$

where

$$D_{C,0} = D_{C,\text{ref}} \cdot \left(\frac{T}{T_{\text{ref}}}\right)^{1.75} \frac{P_{\text{ref}}}{P_a}, \tag{A36}$$

Table A1
Variables and Parameters of the Soil CO_2 Model

Category	Description	Units	Definition
Prognostic			
ρ_C	Mass per volume of carbon in the pore space of soil	kgC m^{-3}	
Diagnostic			
S_C	Heterotrophic respiration	kgC m^{-3} s^{-1}	Equation A31
$F_{C,sfc}$	CO_2 efflux at the soil surface	kgC m^{-2} s^{-1}	Equation A37
$\rho_{C,s}$	Mass per volume of soluble carbon	kgC m^{-3}	Equation A33
$\rho_{O,s}$	Mass per volume of soluble oxygen	kgO m^{-3}	Equation A34
V_{pot}	Potential respiration rate	kgC m^{-3} s^{-1}	Equation A32
D_C	Diffusivity of CO_2 in the pore space	m s^{-2}	Equation A35
Parameters			
V_{max}	Maximum respiration rate	kgC m^{-3} s^{-1}	
E_a	Activation energy	J mol^{-1}	
$kM_{C,s}$	Michaelis parameter for soil	kgC m^{-3}	
$kM_{O,s}$	Michaelis parameter for O_2	kgO m^{-3}	
$f_{C,s}$	Fraction soil organic carbon which is soluble	–	
$\theta_{g,100}$	Volumetric air fraction when $\psi = -1$m	–	
$\tilde{D}_{C,s,liq}$	Dimensionless diffusivity of soluble carbon in liquid	–	
$\tilde{D}_{O,g}$	Dimensionless diffusivity of oxygen in air	–	
Prescribed input			
ρ_O	Mass per volume of O_2 in soil air space	kgO m^{-3}	
$\rho_{C,som}$	Soil organic carbon	kgC m^{-3}	
$\rho_{C,a}$	CO_2 mass per volume in atmosphere above soil	kgC m^{-3}	

with $T_{ref} = 273.15$ K, $P_{ref} = 1013.25$ hPa, P_a the (prescribed) atmospheric pressure at a height close to but above the land surface. and $D_{C,ref} = 1.39 \times 10^{-5}$ m s^{-2}. The parameter b derives from Brooks and Corey (1964) and is estimated from the van Genuchten parameters (n, m) as $b \approx (1 + m)/(nm^2)$; $\theta_{g,100}$ is the volumetric gas fraction when the matric potential is $\psi = -1$ m.

Finally, one of the predictions of the soil CO_2 model is the flux of CO_2 at the surface of the soil. Using a finite difference approximation, this is given by

$$F_{C,sfc} = -D_C(z_{sfc} - \Delta z/2)\frac{\rho_{C,a} - \rho_C(z_{sfc} - \Delta z/2)}{\Delta z/2}, \tag{A37}$$

where $\rho_{C,a}$ is the mass per volume of carbon due to CO_2 in the atmosphere, Δz is the width of the top layer of the soil model, and $z_{sfc} - \Delta z/2$ is the coordinate of the center of the first layer of the soil model. Variables and parameters for the DAMM model are listed in Table A1.

A7. Freezing Front: Comparison to an Analytic Solution

Under certain assumptions, an analytic solution for a phase change front propagating into a medium exists (due to Neumann; see, e.g., Carslaw and Jaeger (1959)). A semi-infinite domain is split into two regions, one frozen and one thawed, which are separated by a freezing (or thawing) interface that moves in time. Thermal properties are discontinuous across the interface, but are assumed to be constant within each region. Any differences in density between the frozen and melted states are ignored. Importantly, there is no flow within the medium. Assuming a

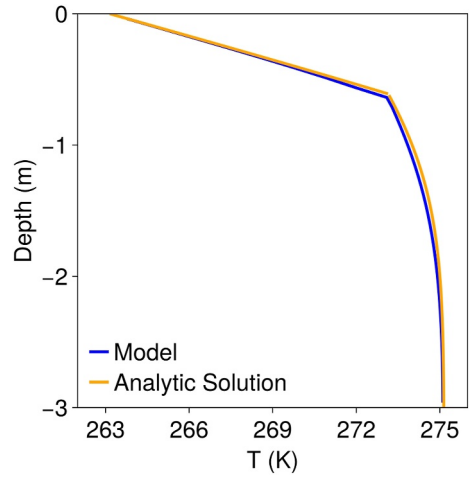

Figure A1. Temperature profile of a freezing front simulation under the assumption of zero water flow. Despite not meeting all the assumptions of the analytic solution, the soil model produces a realistic temperature profile (blue) compared with an analytic solution (orange).

spatially constant initial temperature and a temporally constant surface temperature, the temperature profile as a function of time and depth can be derived. The solution is given in Appendix C of Dall'Amico (2011).

We simulated this problem using the following soil parameters: $\nu_{ss,\text{quartz}} = 0.2$, $\nu_{ss,\text{minerals}} = 0.6$, $\nu_{ss,\text{om}} = 0.2$, $K_{\text{sat}} = 0$, $\nu_p = 0.535$, and $\theta_{\text{res}} = 0$. The initial temperature was 275.15 K, and the surface boundary condition is 263.15 K. The initial water profile was $\theta_l = 0.33$. The domain is 3 m deep in order to mimic the semi-infinite domain of the problem on timescales of the simulation (20 days).

The results of the model are compared with the analytic solution in Figure A1. There is general agreement between the analytic and numerical solutions. In computing the analytic solution, we assumed that the water is entirely frozen or in liquid form on either side of the front; the actual liquid content in phase equilibrium depends on T, and hence will change within the region above and below the front, violating the assumption of constant thermal properties.

Appendix B: Snow Model Parameterizations

B1. Phase Changes in Snow

Following Semtner (1976), we first compute the net energy flux for the snow; if the predicted energy after one (Euler) timestep corresponds to a temperature larger than the freezing temperature, we compute the difference in energy between the predicted energy and the energy at the freezing point. This difference divided by the timestep and the energy per unit mass required to convert ice to liquid water is the phase change mass flux. Similarly, if $S_l > 0$ and the predicted energy corresponds to a temperature colder than the phase change temperature, a phase change mass flux is computed to allow for refreezing.

B2. Liquid Water Loss

Liquid water may be retained in the pores of the snowpack or lost as runoff at the bottom of the snowpack, where it can infiltrate the soil. Following Bartelt and Lehning (2002), we parameterize the maximum volumetric fraction of liquid water that a snowpack can hold as $\theta_{l,c}$, corresponding to a mass fraction of $q_{l,c} = \theta_{l,c}\rho_l/\rho_{\text{snow}}$. We assume that liquid water in excess of this will leave the snowpack on a timescale of

$$\tau = \max\left(\Delta t, z_{\text{snow}} K_{\text{sat,snow}}^{-1}\right), \tag{B1}$$

where Δt is the timestep of the model, z_{snow} is the snowpack depth and $K_{\text{sat,snow}}$ is the saturated hydraulic conductivity for the flow of liquid water through a porous snowpack. Thus, we model the liquid mass fraction as a relaxation of q_l to $q_{l,c}$,

$$\rho_l R_{\text{snow}} = \rho_l \frac{(q_l - q_{l,c})\, S}{\tau} \mathcal{H}(q_l - q_{l,c}). \tag{B2}$$

B3. Snow/Soil Energy Flux

We assume that the energy flux at the snow/soil interface G is proportional to the difference in temperature between the soil and snow near the interface, and we assume that the flux near the interface is continuous: G is equal to the diffusive flux just interior to and within the snow and just interior to and within the soil. Then we can show that

$$G = -\kappa_{\text{eff}} \frac{T - T_{\text{soil}}}{\Delta z_{\text{eff}}} \tag{B3}$$

$$\kappa_{\text{eff}} = \frac{\kappa_{\text{soil}} \kappa_{\text{snow}}}{(\kappa_{\text{snow}} \Delta z_{\text{soil}}/2 + \kappa_{\text{soil}} z_{\text{snow}}/2)} \frac{\Delta z_{\text{soil}} + z_{\text{snow}}}{2} \tag{B4}$$

$$\Delta z_{\text{eff}} = \frac{\Delta z_{\text{soil}} + z_{\text{snow}}}{2}, \tag{B5}$$

where Δz_{soil} is taken to be the distance between the top layer center and the surface, and T_{soil} is the temperature of the soil in the top layer.

The thermal conductivity is modeled as in Jordan (1991) as

$$\kappa_{\text{snow}} = \kappa_{\text{air}} + \left[0.07 \left(\frac{\rho_{\text{snow}}}{\rho_{\text{ice}}} \right) + 0.93 \left(\frac{\rho_{\text{snow}}}{\rho_{\text{ice}}} \right)^2 \right] (\kappa_{\text{ice}} - \kappa_{\text{air}}), \tag{B6}$$

where the (unitless) coefficients are determined by an empirical fit, and we have $\kappa_{\text{snow}} = \kappa_{\text{ice}}$ when $\rho_{\text{snow}} = \rho_{\text{ice}}$.

B4. Snow Cover Fraction

The snow cover fraction is parametrized, following Wu and Wu (2004), as

$$\sigma = \min\left(1, \frac{a z'}{z' + 1}\right), \tag{B7}$$

where $z' = z_{\text{snow}}/c$, and a (unitless) and c (m) are free parameters that depend on the horizontal resolution of the simulation. For 1.5° resolution, $c \approx 0.1$ m and $a \approx 2$. Improving this parameterization is a goal for the next version of ClimaLand.

Appendix C: P-Model Implementation

To evaluate the P-model and obtain gross primary production, net primary production, and the stomatal conductance, the following input is required: absorbed canopy photosynthetic active radiation APAR (moles of photons m^{-2} s^{-1}), vapor pressure deficit VPD (Pa), canopy temperature T (K) and air pressure P (Pa). For a derivation of the following, please see H. Wang et al. (2017).

C1. Core Photosynthesis Equations

As in the Farquhar-von Caemmerer-Berry model (Farquhar et al., 1980), we have

$$A_c = V_{\text{cmax}} m_c \tag{C1}$$

$$A_j = \frac{J}{4} m_j \tag{C2}$$

$$\text{GPP} = \min(A_c, A_j) \tag{C3}$$

$$A_n = \text{GPP} - R_d, \tag{C4}$$

where A_c is the Rubisco-limited assimilation (mol CO_2 m^{-2} s^{-1}), V_{cmax} is the maximum rate of carboxylation (mol CO_2 m^{-2} s^{-1}), A_j is the light-limited assimilation (mol CO_2 m^{-2} s^{-1}), J is the electron transport rate (mol electrons m^{-2} s^{-1}; the factor of 4 converts from moles of electrons to moles of CO_2), GPP (mol CO_2 m^{-2} s^{-1}) is the gross primary production, and A_n is the net primary production (mol CO_2 m^{-2} s^{-1}), which has been corrected for dark respiration R_d (mol CO_2 m^{-2} s^{-1}). We also have

$$\begin{aligned} m_c &= \begin{cases} \dfrac{c_i - \Gamma^\star}{c_i + K} & C3 \\ 1 & C4 \end{cases} \\ m_j &= \begin{cases} \dfrac{c_i - \Gamma^\star}{c_i + 2\Gamma^\star} & C3 \\ 1 & C4, \end{cases} \end{aligned} \tag{C5}$$

where $\Gamma^\star$ is the CO_2 compensation point (Pa), K is the Michelis-Menton ratio (defined below; Pa), and c_i is the intercellular CO_2 concentration (Pa); C3 and C4 indicate the photosynthesis mechanism.

In order to compute $\Gamma^\star$ and K, we use an Arrhenius function defined as

$$f_{\text{Arr}}(T; \Delta H) = \exp(\Delta H(T - T_{25})/(T_{25} R T)), \tag{C6}$$

where $T_{25} = 298.15\text{K}$, R is the universal gas constant (J $\text{mol}^{-1}\text{K}^{-1}$), and ΔH is an activation energy (J mol^{-1}) which depends on the rate being modified. We then compute

$$\begin{aligned} \Gamma^\star &= \Gamma^\star_{25} \frac{P}{101325\text{Pa}} f_{\text{Arr}}\left(T; \Delta H \Gamma^\star\right) \\ K_c &= K_{c,25} f_{\text{Arr}}(T; \Delta H K_c) \\ K_o &= K_{o,25} f_{\text{Arr}}(T; \Delta H K_o) \\ K &= K_c \left[1 + \frac{o_i P}{K_o}\right], \end{aligned} \tag{C7}$$

where o_i is the intercellular oxygen concentration (mol mol^{-1}), treated as constant, and $\Delta H K_c, \Delta H K_o, \Delta H \Gamma^\star$ are all constants (J mol^{-1}).

The maximum rate of carboxylation V_{cmax}, the maximum rate of electron transport rate J_{max}, and the dark respiration R_d are also computed from values at T_{25} as

$$V_{\text{cmax}} = V_{\text{cmax},25} f'_{\text{Arr}}(T; \Delta H V_{\text{cmax}}, \Delta S V_{\text{cmax}}, \Delta H_{d,V}) \tag{C8}$$

$$J_{\text{max}} = J_{\text{max},25} f'_{\text{Arr}}(T; \Delta H J_{\text{max}}, \Delta S J_{\text{max}}, \Delta H_{d,J}) \tag{C9}$$

$$R_d = b V_{\text{cmax},25} f_{\text{Rd}}(T; \Delta H R_d, \Delta S R_d, \Delta H_{d,R}), \tag{C10}$$

Table C1
Parameters of the P-Model

Parameter	Value	Equation	Source
$\Gamma^\star_{25}$	4.332 Pa	Equation C7	Stocker et al. (2020)
$K_{c,25}$	39.97 Pa	Equation C7	Stocker et al. (2020)
$K_{o,25}$	27,480 Pa	Equation C7	Stocker et al. (2020)
$\Delta H\Gamma^\star$	37,830 J mol^{-1}	Equation C7	Stocker et al. (2020)
ΔHK_c	79,430 J mol^{-1}	Equation C7	Stocker et al. (2020)
ΔHK_o	36,380J mol^{-1}	Equation C7	Stocker et al. (2020)
ΔHV_{cmax}	71,513 J mol^{-1}	Equation C8	Kattge and Knorr (2007)
ΔSV_{cmax}	649 J mol^{-1}K^{-1}	Equation C8	Kattge and Knorr (2007)
$\Delta H_{d,V}$	200,000 J mol^{-1}	Equation C8	Kattge and Knorr (2007)
ΔHJ_{max}	49,884 J mol^{-1}	Equation C8	Kattge and Knorr (2007)
ΔSJ_{max}	646 J mol^{-1}K^{-1}	Equation C8	Kattge and Knorr (2007)
$\Delta H_{d,J}$	200,000 J mol^{-1}	Equation C8	Kattge and Knorr (2007)
o_i	0.2095	Equation C7	G. Bonan (2019)
b	0.015	Equation C10	G. Bonan (2019)
ϕ_{a0} C3/C4	0.030624	Equation C15	Stocker et al. (2020)
ϕ_{a1} C3/C4	0.001914 K^{-1}	Equation C15	Stocker et al. (2020)
ϕ_{a2} C3/C4	−0.00002958 K^{-2}	Equation C15	Stocker et al. (2020)
β	198	Equation C17	Calibrated
$c^\star$	0.30	Equation C20	Calibrated

where b is a constant and $f'_{\text{Arr}}(T;\Delta S,\Delta H_d)$ reflects activation and deactivation at higher temperature (Kattge & Knorr, 2007; Leuning, 2002), with rate specific parameters:

$$f'_{\text{Arr}}(T;\Delta H,\Delta S,\Delta H_d) = f_{\text{Arr}}(T;\Delta H)\frac{1+\exp[(T_{25}\Delta S-\Delta H_d)/(RT_{25})]}{1+\exp[(T\Delta S-\Delta H_d)/(RT)]}, \tag{C11}$$

and

$$f_{\text{Rd}}(T;) = \exp\left[b(T-T_{25})+c\left(T^2-T_{25}^2\right)\right], \tag{C12}$$

with $b = 0.1012\text{K}^{-1}, c = -0.0005\text{K}^{-2}$ given in Heskel et al. (2016). Currently, we do not distinguish between C3 and C4 plants in the dark respiration term.

The intercellular CO_2 concentration is computed in the P-model as

$$c_i = \frac{\xi c_a + \Gamma^\star\sqrt{\text{VPD}}}{\xi+\sqrt{\text{VPD}}}, \tag{C13}$$

where c_a is the atmospheric CO_2 pressure (Pa) and ξ (Pa) will be defined in the section below. Finally, the electron transport rate J is

$$J = \frac{4\phi_0\text{APAR}}{\sqrt{1+\left(4\phi_0\text{APAR}/J_{\text{max}}\right)^2}}, \tag{C14}$$

where ϕ_0 is the quantum yield defined as

$$\phi_o = \phi_{a0} + \phi_{a1}(T - T_{25}) + \phi_{a2}(T - T_{25})^2, \tag{C15}$$

where $\phi_{a0}, \phi_{a1}, \phi_{a2}$ are constants defined the same for C3 and C4 photosynthesis. Constant parameters for the P-model are defined in Table C1.

The canopy level stomatal conductance in units of mol H_2O $m^2 s^{-1}$ is given as

$$g_s = \frac{1.6A_n}{c_a - c_i}. \tag{C16}$$

C2. Obtaining $V_{cmax,25}$, $J_{max,25}$ and ξ

The optimal values of ξ, $V_{cmax,25}$, and $J_{max,25}$ are computed at a given instant as

$$\xi_{opt} = \left[\frac{\beta(K + \Gamma^\star)}{1.6\eta^\star}\right]^{1/2}, \tag{C17}$$

where β is a free parameter and $\eta^\star$ is the ratio of the viscosity of water at temperature T to its value at T_{25},

$$\begin{aligned} V_{cmax,opt} &= \frac{\beta_m \phi_0 m' \text{APAR}}{m_c}, \\ V_{cmax,25,opt} &= \frac{V_{cmax,opt}}{f'_{Arr}(T; \Delta HV_{cmax}, \Delta SV_{cmax}, \Delta H_{d,V})}, \end{aligned} \tag{C18}$$

and

$$\begin{aligned} J_{max,opt} &= \frac{4\phi_0 \text{APAR}}{\sqrt{\left(m_j/(\beta_m m')\right)^2 - 1}}, \\ J_{max,25,opt} &= \frac{J_{max}}{f'_{Arr}(T; \Delta HJ_{max}, \Delta SJ_{max}, \Delta H_{d,J})}, \end{aligned} \tag{C19}$$

where β_m is a dimensionless moisture stress parameter and

$$m' = m_j\sqrt{1 - (c^\star/m_j)^{2/3}}, \tag{C20}$$

with m_j and m_c given in Equation (C5), and $c^\star$ a free parameter. Note that the placement of β_m in the $J_{max,opt}$ term is required by enforcing $A_c = A_j$ at the optimal point.

Because plants cannot immediately adjust to the optimal point, the values for $\xi, V_{cmax,25}, J_{max,25}$ used in computing photosynthesis at any given time are lagging the optimal values defined above. Currently, we follow Mengoli et al. (2022) and use

$$\xi^{i+1} = \alpha\xi^{i+1}_{opt} + (1 - \alpha)\xi^i \tag{C21}$$

$$V^{i+1}_{cmax,25} = \alpha V^{i+1}_{cmax,25,opt} + (1 - \alpha)V^i_{cmax,25} \tag{C22}$$

$$J^{i+1}_{max,25} = \alpha J^{i+1}_{max,25,opt} + (1 - \alpha)J^i_{max,25}, \tag{C23}$$

where the update occurs at local noon, that is, once per day so that the index i indicates a day counter, and $\alpha = 0.067$, which corresponds to a memory timescale of 15 days.

Appendix D: Radiative Transfer

D1. Soil and Snow Radiation

The net radiative energy flux for bare soil or snow is

$$\begin{aligned} R_n &= \mathrm{SW}_d - \mathrm{SW}_u + \mathrm{LW}_d - \mathrm{LW}_u \\ &= \mathrm{SW}_n + \mathrm{LW}_n, \end{aligned} \tag{D1}$$

where SW_n and LW_n are the net shortwave (solar radiation) and longwave (thermal radiation) radiative energy fluxes. In the energy flux boundary condition for soil Equation 19, the net radiative flux is weighted by the bare soil area fraction $(1 - \sigma)$, where σ is the snow cover fraction. For snow, the net radiation R_n is weighted by the snow cover fraction (σ), in the flux in Equation 26.

The net shortwave radiation SW_n at the bare ground surface is computed as

$$\mathrm{SW}_n = \int_0^\infty d\lambda f_{\mathrm{trans}}(\lambda, \dots) u_{\mathrm{SW},d}(\lambda, \dots)(1 - \alpha_{\mathrm{ground}}(\lambda, \dots)), \tag{D2}$$

where λ indicates a wavelength of light, f_{trans} is the fraction transmitted by the canopy, $u_{\mathrm{SW},d}$ is the downwelling radiative energy flux density (per wavelength increment) at the land surface above the canopy, and α_{ground} is the soil or snow albedo. The quantities $f_{\mathrm{trans}}, u_{\mathrm{SW},d}$, and α_{ground} are all functions of wavelength. If no canopy is present, $f_{\mathrm{trans}} = 1$ (see D2).

In the above, we assume zero transmissivity of radiation into the soil or snow at depth. As the transmitted fraction in snow is less than 5% for snowpacks of 10 cm depth (Perovich, 2007), this is a negligible effect.

The soil albedo is computed as a linear combination of a wet and dry soil albedo (D. M. Lawrence et al., 2019; Braghiere, Wang, et al., 2023). We obtained maps of the soil color from NCAR (P. J. Lawrence & Chase, 2007), and values of the soil albedo in the different wavelength bands for wet soil and dry soil, by soil color, following Braghiere et al. (2021). Then, we approximate the soil albedo as

$$\alpha_{\mathrm{soil}}(\lambda, S_l) = S_l \alpha_{\mathrm{soil,wet}}(\lambda) + (1 - S_l)\alpha_{\mathrm{soil,dry}}(\lambda), \tag{D3}$$

where S_l is the effective saturation, defined in Equation A15. The snow albedo changes with wavelength, snow morphology, impurities, and snow age (Flanner et al., 2021). In this first version of ClimaLand, it is approximated as a global constant with a calibratable value or as a simple function of zenith angle and snow bulk density,

$$\alpha_{\mathrm{snow}} = f(\rho_{\mathrm{snow}})[\alpha_0 + \Delta\alpha \exp(-k\mu)], \tag{D4}$$

where α_0, $\Delta\alpha$, and k are constants, $\mu = \cos\theta_s$ is the cosine of the zenith angle θ_s, and

$$f(\rho_{\mathrm{snow}}) = 1 - \beta(\rho_{\mathrm{snow}}/\rho_{\mathrm{liq}} - \rho_{\mathrm{snow,min}}/\rho_{\mathrm{liq}}), \tag{D5}$$

where $0 < \beta < 1$ and $\rho_{\mathrm{snow,min}}$ are constants. The net longwave radiation at the surface of the ground is

$$\mathrm{LW}_n = \epsilon_{\mathrm{ground}}(\mathrm{LW}_{d,\mathrm{canopy}} - \sigma_{\mathrm{SB}} T^4_{\mathrm{ground,\ sfc}}), \tag{D6}$$

where again ground refers to soil or snow, $\epsilon_{\mathrm{soil}} = 0.96$ is the emissivity of the soil (assumed constant), $\epsilon_{\mathrm{snow}} = 0.97$ is the emissivity of the snow (assumed constant)σ_{SB} is the Stefan-Boltzmann constant, $T_{\mathrm{ground,sfc}}$ is the surface temperature of the soil or snow, and $\mathrm{LW}_{d,\mathrm{canopy}}$ is the downwelling longwave radiation underneath the canopy (equal to the downwelling longwave radiation from the atmosphere if no vegetation is present). The term $\epsilon_{\mathrm{ground}}$ is applied to the downwelling longwave radiation because absorptance is numerically equal to emissivity through the Kirchhoff law. The downwelling longwave radiation is computed in D2.

D2. Canopy Radiation

As with the soil and snow components, we model the radiative transfer in the canopy in shortwave and longwave bands. The fraction of shortwave radiation that is transmitted by the canopy f_{trans} is a key output of this module, as this dictates the shortwave radiation available for the ground components, soil and snow. The reflected shortwave radiation, a fraction f_{refl} of the incoming radiation, dictates the upwelling shortwave radiation from the land surface, and the absorbed shortwave radiation, a fraction f_{abs} affects what is available for photosynthesis and heating of the canopy.

D2.1. Shortwave Radiation

The net shortwave radiation for the canopy is

$$\text{SW}_{n,\text{canopy}} = \int_0^\infty d\lambda f_{\text{abs}}(\lambda, \dots) u_{\text{SW},d}(\lambda, \dots), \tag{D7}$$

where λ indicates a wavelength of light, f_{abs} is the fraction absorbed by the canopy and $u_{\text{SW},d}$ is the downwelling radiative energy flux density (per wavelength increment) at the land surface above the canopy.

The net upwelling shortwave radiation from the land surface is

$$\text{SW}_{\text{u}} = \int_0^\infty d\lambda f_{\text{refl}}(\lambda, \dots) u_{\text{SW},d}(\lambda, \dots), \tag{D8}$$

where f_{refl} is the fraction reflected by the land surface.

Currently, the radiation field is discretized into two bands: Photosynthetically Active Radiation (PAR, 400–700 nm) and Near Infrared (NIR, 700–2,500 nm).

D2.2. Extinction Coefficient

The extinction coefficient is defined following Campbell (1998) in terms of the zenith angle θ_s as

$$\mathcal{K} = \frac{G(\theta_s)}{(\cos\theta_s + \epsilon)}, \tag{D9}$$

where $G(\theta_s)$ is the projection of leaf area in the direction of the solar beam (Ross, 1981), and the small value ϵ prevents division by zero. In ClimaLand, one can either supply a constant value of $G(\theta_s)$, or use the semi-empirical Ross-Goudriaan function (Goudriaan, 1977)

$$\begin{aligned} G(\theta_s) &= \phi + 0.88(1 - 2\phi)\cos\theta_s \\ \phi &= 0.5 - 0.63\chi_l - 0.33\chi_l^2, \end{aligned} \tag{D10}$$

where χ_l is the Ross index, which varies with vegetation type.

D2.3. Beer-Lambert Law

The absorbed, transmitted, and reflected shortwave radiation fractions from the canopy, f_{abs},f_{trans},f_{refl}, in any wavelength band, are calculated following the Beer-Lambert law:

$$\begin{aligned} f_{\text{abs},\,\lambda} &= (1 - \alpha_\lambda)\left(1 - e^{-\mathcal{K}\mathcal{A}_{\text{leaf}}\Omega}\right) \\ f_{\text{trans},\,\lambda} &= e^{-\mathcal{K}\mathcal{A}_{\text{leaf}}\Omega} \\ f_{\text{refl},\lambda} &= 1 - f_{\text{abs},\lambda} - f_{\text{trans},\lambda}(1 - \alpha_{\text{ground},\lambda}), \end{aligned} \tag{D11}$$

where λ indexes wavelength bands, α_λ is the leaf albedo, $\mathcal{K}$ is the vegetation extinction coefficient, $\mathcal{A}_{\text{leaf}}$ is the leaf area index, and Ω is the clumping index following Braghiere et al. (2021). The albedo of the ground is computed as

$$\alpha_{\text{ground},\lambda} = \alpha_{\text{soil},\lambda}(1-\sigma) + \sigma\alpha_{\text{snow},\lambda}, \tag{D12}$$

where σ is the snow cover fraction.

D2.4. Radiative Transfer in a Vertically Resolved Canopy

In the Beer-Lambert scheme, the canopy is treated as a single bulk layer with certain physical properties ($\mathcal{A}_{\text{leaf}}, \alpha_\lambda$). Following Sellers (1985), we also provide an option to resolve vertical structure in the canopy and simulate the effects of multiple scattering by leaves, branches, and stems. As with the Beer-Lambert scheme, this involves the two-stream approximation which splits the radiation into an upwelling and downwelling stream. The radiation is further decomposed into a direct and diffuse component (Räisäenen, 2002). The scheme of Sellers (1985) is adjusted to account for the clumping index following Braghiere et al. (2019, 2021).

As with the Beer-Lambert scheme, the fraction of shortwave radiation absorbed, reflected, and transmitted to the ground is computed per band, but also for diffuse and direct beams. The ground albedo is computed using Equation (D12), but also for diffuse and direct radiation separately. The transmissivity of the leaves τ_l is also required for each waveband.

D2.5. Longwave Radiation

The downward longwave radiation at the top of the canopy is denoted by LW_d. Below the canopy, it is given by

$$\text{LW}_{d,\text{canopy}} = (1-\epsilon_{\text{canopy}})\,\text{LW}_d + \epsilon_{\text{canopy}}\sigma_{\text{SB}}T^4_{\text{canopy}}, \tag{D13}$$

where $\epsilon_{\text{canopy}} = 0.97\left(1 - e^{-\mathcal{A}}\right)$ is the emissivity of the canopy (D. M. Lawrence et al., 2019).

The upwelling longwave radiation above the soil or snow is given by

$$\text{LW}_{u,\text{soil}} = \epsilon_{\text{soil}}\sigma_{SB}T^4_{\text{soil}} + (1-\epsilon_{\text{soil}})\,\text{LW}_{d,\text{canopy}} \tag{D14}$$

where the same equation holds for snow if soil variables are replaced by the equivalent snow variables. The net LW radiation absorbed by the canopy is

$$\text{LW}_{n,\text{canopy}} = \epsilon_{\text{canopy}}\left[\text{LW}_d - 2\sigma_{\text{SB}}T^4_{\text{canopy}} + (1-\sigma)\text{LW}_{u,\text{soil}} + \sigma\text{LW}_{\text{u, snow}}\right] \tag{D15}$$

so that the upwelling long wave radiation from the land surface is

$$\text{LW}_u = (1-\epsilon_{\text{canopy}})[(1-\sigma)\text{LW}_{u,\text{soil}} + \sigma\text{LW}_{\text{u,snow}}] + \epsilon_{\text{canopy}}\sigma_{\text{SB}}T^4_{\text{canopy}} \tag{D16}$$

From this, we can compute an effective surface temperature T_{eff} for the land surface as

$$\sigma_{\text{SB}}T^4_{\text{eff}} \equiv \text{LW}_{\text{u}}. \tag{D17}$$

Appendix E: Turbulent Surface Fluxes

The boundary fluxes between the land surface and the atmosphere include turbulent exchange fluxes of energy and water. When running with a prescribed atmosphere, we assume that the temporally and spatially varying atmospheric temperature $T_a(\Delta z_a)$, air density $\rho_a(\Delta z_a)$, water vapor specific humidity $q_a(\Delta z_a)$, and wind speed $u_a(\Delta z_a)$ tangential to the surface are prescribed at a height Δz_a close to but above the land surface. In reality, and

when the land model is coupled to an atmosphere, the atmospheric conditions are determined through interactions of the atmosphere with the land.

The surface fluxes of water, latent heat, and sensible heat are given by bulk exchange laws, which are first-order finite difference approximations of flux-gradient relationships (Yatunin et al., 2025),

$$\begin{aligned} E &= -\rho_s \frac{q_a - q_s}{r_{\mathrm{ae}}} \\ LE &= L_{v,0} E \\ H &= -\rho_s \frac{\mathrm{DSE}_a - \mathrm{DSE}_s}{r_{\mathrm{ae}}}, \end{aligned} \tag{E1}$$

where q_s, DSE_s and ρ_s are the surface specific humidity, dry static energy, and air density, with

$$\mathrm{DSE} = c_{pm}(T - T_0) + \Phi, \tag{E2}$$

where c_{pm} is the specific heat at constant pressure of moist air and $\Phi = gz$ is the gravitational potential energy of the air.

The aerodynamic resistance $r_{ae}\ \left(\mathrm{sm}^{-1}\right)$ is defined as

$$r_{\mathrm{ae}} = \frac{1}{C_h u_a}, \tag{E3}$$

with C_h the (dimensionless) drag coefficient for heat transport as determined by Monin-Obukhov Similarity Theory. The drag coefficient C_h depends on the functional form of the similarity profile, as well as the roughness lengths and displacement height of the surface, the height at which the wind speed reaches zero. Below, we will see how the surface variables, roughness lengths, and displacement heights change for bare soil, bare snow, or canopy covered land.

While ClimaLand tracks water content and energy, it does not explicitly model q_s, ρ_s, or T_s (needed for the dry static energy). For all components, the surface air density is approximated using adiabatic extrapolation as

$$\rho_s \approx \rho_a \left(\frac{T_s}{T_a}\right)^{c_{vm}(q_a)/R_m(q_a)}, \tag{E4}$$

where c_{vm} is the specific heat at constant volume and R_m is the gas constant of the moist air in the atmosphere. Further parameterizations are required for q_s and T_s.

E1. Bare Snow

The bare snow roughness length for momentum is fixed at 0.01 m and the roughness length for scalars at 0.007 m. The displacement height is set to zero.

We estimate the specific humidity at the surface of the snow q_s by averaging the saturated specific humidity over ice and water using the mass fraction of liquid water q_l,

$$q_s = q_l q_{\mathrm{sat}}(T_s, \rho_s) + (1 - q_l)\, q_{\mathrm{sat}}(T_s, \rho_s), \tag{E5}$$

where the saturation specific humidity is computed over ice or liquid as appropriate.

The surface temperature is currently set to be equal to the bulk temperature,

$$T_s = T. \tag{E6}$$

This is a very coarse approximation, and in the future we will either solve for the surface temperature using a flux balance approach or introduce additional prognostic variables (You et al., 2014).

E2. Bare Soil

The bare soil roughness length for momentum is fixed at 0.01 m and the roughness length for scalars at 0.007 m. The displacement height is set to zero.

For the surface temperature, we use the value at the center of the top layer (a diagnostic variable of the model), which is an acceptable approximation given that the vertical resolution is $\mathcal{O}$(cm) near the surface: $T_s = T_{\text{soil,sfc}}$.

For simplicity, we model the mass fluxes of water vapor due to sublimation E_i and evaporation E_l, but not both at the same time.

E2.1. Sublimation

When the temperature is below the depressed freezing temperature T_f^* (Equation 13c), only sublimation is considered. We model $q_s = q_{\text{sat}}(T_s, \rho_s)$, the saturated specific humidity over ice at temperature T_s and in air of density ρ_s, that is,

$$E_i = -\rho_s \frac{q_a - q_{\text{sat}}(T_s, \rho_s)}{r_{\text{ae}}}$$
$$E_l = 0. \tag{E7}$$

E2.2. Evaporation

As the soil surface dries under evaporation, a hydraulic gradient is established between the surface and deeper, wetter layers. Initially, this gradient drives Darcy flow of water to the surface, maintaining a maximum evaporation rate (Phase I evaporation, where the surface is saturated); as more of the soil dries near the surface, these flow paths cease and the evaporation is instead sustained by diffusion of water vapor from deeper in the soil to the surface (Phase 2 evaporation, e.g., Or et al. (2013)). Lehmann et al. (2008, 2018) derived evaporation from soil to be

$$E_l = E_{l,\max} \frac{\rho_l K_{\text{eff}} \frac{\Delta h}{\Delta z}\Big|_{\text{cap}}}{E_{l,\max} + \rho_l K_{\text{eff}} \frac{\Delta h}{\Delta z}\Big|_{\text{cap}}}, \tag{E8}$$

where $E_{l,\max}$ is the maximum evaporation rate computed as if $q_s = q_{\text{sat}}$ and the capillary flux magnitude which drives this flow is

$$K_{\text{eff}} \frac{\Delta h}{\Delta z}\Big|_{\text{cap}} = 4K_{\text{sfc}} \left[1 + \frac{E_{l,\max}}{\rho_l 4K(S_c)}\right]. \tag{E9}$$

Here $K(S_c)$ is the hydraulic conductivity evaluated at the critical saturation S_c below which the flow paths become disconnected. This was derived (Lehmann et al., 2008) to be

$$S_c = \left(1 + \left(\frac{n-1}{n}\right)^{1-2n}\right)^{-m}. \tag{E10}$$

In summary, then, when $T > T_f^*$ (the depressed freezing point temperature), we have

$$E_l = \begin{cases} E_{l,\max}\dfrac{\rho_l K_{\text{eff}}\left.\dfrac{\Delta h}{\Delta z}\right|_{\text{cap}}}{E_{l,\max} + \rho_l K_{\text{eff}}\left.\dfrac{\Delta h}{\Delta z}\right|_{\text{cap}}} & q_{\text{sat}}\left(T_s,\rho_s\right) > q_a \\ -\rho_s \dfrac{q_a - q_{\text{sat}}\left(T_s,\rho_s\right)}{r_{ae}} & \text{otherwise} \end{cases} \tag{E11}$$

$$E_i = 0. \tag{E12}$$

The latent heat flux is computed using E_l, while the sensible heat flux is approximated using the solution to the Monin-Obukhov solve with $q_s = q_{\text{sat}}\left(T_s,\rho_s\right)$.

E3. Vegetated Surfaces

To simulate turbulent surface exchange with the atmosphere from vegetated surfaces, at least two general approaches are possible. One treats each component of the land surface independently: bare soil and snow interact with the atmosphere as described above, and the vegetation also interacts with the atmosphere in a similar fashion. Three sets of surface fluxes are computed. This approach (referred to as parallel fluxes in G. Bonan (2019)) may be sufficient for sparse vegetation where the vegetation does not alter the interactions between the soil/snow and the atmosphere much from standard surface theory. A more realistic approach for vegetated surfaces models fluxes between vegetation, soil, and snow between a common surface layer, and further accounting for the impact of the canopy on soil and snow fluxes. In the common surface layer model, the fluxes between the surface layer and the atmosphere are then equated with the sum of the fluxes between the land components and the surface layer, implicitly determining the unknown surface layer temperature and humidity (G. Bonan, 2019; D. M. Lawrence et al., 2019). ClimaLand version 1 uses parallel fluxes, but this scheme will be revisited as needed in the future.

The net vapor flux magnitude is

$$E_{\text{total}} = (1-\sigma)E_{\text{soil}} + \sigma E_{\text{snow}} + T, \tag{E13}$$

where T, the canopy transpiration, already accounts for the leaf area index. Evaporation from water on leaves is not modeled. The net heat flux is

$$(H+L)_{\text{total}} = (1-\sigma)(H+L)_{\text{soil}} + \sigma(H+L)_{\text{snow}} + (H+L)_{\text{canopy}}, \tag{E14}$$

where L is defined as in Equation E1 and H_{canopy} already account for the stem and leaf area index.

Focusing on vapor fluxes from the canopy, we will assume that the specific humidity is known at some location indicated with a prime, q'. We model the flux between this location and the canopy air (denoted with s) as a diffusive flux with some known resistance r'_q, that is,

$$E' = -\rho_s \frac{q_s - q'}{r'_q}. \tag{E15}$$

Assuming that the fluxes are continuous, we set $E' = E$ and eliminate the unknown q_s to find

$$E = -\rho_s \frac{q_a - q'}{r'_q + r_{\text{ae}}}. \tag{E16}$$

In Equation (E16), we set $q' = q_{\text{sat}}\left(T_{\text{canopy}},\rho_s\right)$, and

$$r'_q = r_{\text{stomata}} + r_b, \tag{E17}$$

where r_{stomata} is the inverse of the stomatal conductance upscaled to the canopy (weighted by $\mathcal{A}_{\text{leaf}}$), and r_b is the leaf boundary layer resistance (D. M. Lawrence et al., 2019) also upscaled to the canopy,

$$r_b = \frac{1}{\mathcal{A}_{\text{leaf}}} \frac{1}{C_v} \left(\frac{u_\star}{d_{\text{leaf}}} \right)^{-1/2}, \tag{E18}$$

with $C_v \left(\text{m s}^{-1/2} \right)$ the turbulent transfer coefficient, assumed to be constant, and d_{leaf} is a characteristic length-scale (constant, 0.04 m). The two resistances are added in parallel, while the resistance from each unit of area is added in series (hence the multiplicative factor of $\mathcal{A}_{\text{leaf}}$).

The same exact flux balance approach is carried out for sensible heat fluxes, that is,

$$H = -\rho_s \frac{\text{DSE}_a - \text{DSE}'}{r'_T + r_{\text{ae}}}. \tag{E19}$$

We compute DSE′ using $T' = T_{\text{canopy}}$ and upscale to the canopy instead as

$$r'_T = r_b \frac{\mathcal{A}_{\text{leaf}}}{\mathcal{A}}, \tag{E20}$$

since both stem and leaf areas can contribute to the sensible heat flux.

Appendix F: Jacobian Approximation

As described in Section 3, we used a mixed implicit/explicit (IMEX) time-stepping scheme to solve the land system. To solve the implicit tendencies, we use Newton's method with an approximate Jacobian. Here we describe the Jacobian approximation currently used in the model. Note that this section describes the Jacobian used for the full, integrated land model. The approximations used for the standalone `RichardsModel` and `EnergyHydrology` soil models and for the standalone `CanopyModel` are similar but contain only the terms solved for in those models.

We begin with the general equations we solve for explicitly- and implicitly-stepped variables in the model. These arise from discretizing ordinary differential equations with respect to time; for simplicity, here, we consider only implicit and explicit Euler steps, and focus on two scalar variables x and y. However, the Jacobian derived can also be used in multi-stage methods. We will denote the continuous time derivative terms, or tendencies, appearing on the right-hand side of a differential equation as either implicit or explicit, depending on how we choose to timestep them. Terms which require implicit stepping are also referred to as stiff.

Consider the following discrete equation for evolving a variable y, which only has an explicitly-stepped tendency $\mathcal{T}_{\text{exp}}$:

$$y^{n+1} = y^n + \mathcal{T}_{\text{exp}}(x^n, y^n, t^n)\Delta t, \tag{F1}$$

where y^{n+1} indicates the value of y at time $t^{n+1} = t^n + \Delta t$, and y^n indicates the value of y at time t^n. The latter is assumed to be known.

The equivalent equation for a variable x with tendencies that are stepped both implicitly and explicitly is:

$$x^{n+1} - \Delta t \mathcal{T}_{\text{imp}}\left(x^{n+1}, y^n, t^n\right) = x^n + \mathcal{T}_{\text{exp}}(x^n, y^n, t^n)\Delta t, \tag{F2}$$

where we evaluate the implicit tendency $\mathcal{T}_{\text{imp}}$ at the (unknown) value x^{n+1}. We can rewrite Equation (F2) as

$$H\left(x^{n+1}\right) = x^{n+1} - \Delta t \mathcal{T}_{\text{imp}}\left(x^{n+1}, y^n\right) - x^n - \mathcal{T}_{\text{exp}}(x^n, y^n)\Delta t = 0; \tag{F3}$$

solving for x^{n+1} involves finding the root of this nonlinear Equation F3. We do this iteratively using Newton's method:

$$H\left(x^{n+1,k}\right) + \left.\frac{\partial H}{\partial x^{n+1}}\right|_{x^{n+1,k}} \delta x^{n+1,k+1} = 0, \tag{F4}$$

where k refers to the Newton iteration, with $x^{n+1,1} = x^n$ and

$$\frac{\partial H}{\partial x^{n+1}}|_{x^{n+1,k}} = 1 - \Delta t \frac{\partial \mathcal{T}_{\text{imp}}}{\partial x^{n+1}}|_{x^{n+1,k}} \equiv 1 - J\left(x^{n+1,k}\right)\Delta t \tag{F5}$$

$$\delta x^{n+1,k+1} = x^{n+1,k+1} - x^{n+1,k} \tag{F6}$$

Now we can plug these terms into (F4):

$$\left(1 - \Delta t J\left(x^{n+1,k}\right)\right)\delta x^{n+1,k+1} = -H\left(x^{n+1,k}\right). \tag{F7}$$

More generally, x and y become vectors. For the land model, we replace x with $\boldsymbol{x}$, denoting the implicitly-stepped prognostic variables:

$$\boldsymbol{x} = \left(\vartheta_l^1, \ldots, \vartheta_l^N, \rho e_{int}^1, \ldots, \rho e_{int}^N, T_{\text{canopy}}\right).$$

Here N is the number of vertical levels in the domain, so that the dimensionality of x is $2N + 1$. We also have the analogous vector of explicitly-stepped variables $\boldsymbol{y}$, with dimensionality $2N + 5$:

$$\boldsymbol{y} = \left(\theta_i^1, \ldots \theta_i^N, \rho_{\text{C}}^1, \ldots \rho_{\text{C}}^N, \theta_{\text{stem}}, \theta_{\text{leaf}}, S, S_l, U\right).$$

In the vector case, J becomes a square matrix, with elements $J_{i,j}$ defined as

$$J_{ij}\left(x^{n+1,k}\right) = \left.\frac{\partial \mathcal{T}_{\text{imp},i}\left(x^{n+1,k}\right)}{\partial x_j}\right|_{x^{n+1,k}}. \tag{F8}$$

The indices i,j run from $1, \ldots, 4N + 6$, that is, over all variables in the model. Equation (F7) then becomes a system of linear equations,

$$\left(I_{i,j} - \Delta t J_{ij}\left(x^{n+1,k}\right)\right)\delta x_j^{n+1,k+1} = -H_i\left(x^{n+1,k}\right), \tag{F9}$$

where $I_{i,j} = \delta_{i,j}$ is an element of the identity matrix, and $\delta_{i,j}$ is the Kronecker delta.

Once the form of J is known for the problem at hand, each Newton iteration requires solving this system of linear equations for $\delta x^{n+1,k+1}$. The frequency at which the Jacobian is updated during the solve can be changed in the code, but by default it is updated at every Newton iteration. The Jacobian in its exact form is a dense matrix, which makes Equation (F9) expensive to solve for large N. It is numerically easier to solve if we approximate it to have a simpler form, such as tridiagonal. However, this approximation, if poor, will affect convergence.

In the Jacobian approximation implemented for the full land model, we include the terms described below, and exclude all other terms, approximating them as zero. Note that in component models without all of these prognostic variables, terms containing the absent variables are not included in the Jacobian. The non-zero terms are evaluated by discretizing the following

$$\frac{\partial \mathcal{T}_{\text{imp\_}\vartheta_l}}{\partial \vartheta_l} = \frac{\partial}{\partial z} K \frac{\partial^2 \psi}{\partial z \partial \vartheta_l} \tag{F10}$$

$$\frac{\partial \mathcal{T}_{\text{imp\_}\rho e_{int}}}{\partial \rho e_{int}} = \frac{\partial}{\partial z}\kappa\frac{\partial}{\partial z}\frac{1}{\rho_s c_s} \tag{F11}$$

$$\frac{\partial \mathcal{T}_{\text{imp\_}\rho e_{int}}}{\partial \vartheta_l} = \frac{\partial}{\partial z}K\rho_l e_l\frac{\partial^2 \psi}{\partial z \partial \vartheta_l} \tag{F12}$$

$$\frac{\partial \mathcal{T}_{\text{imp\_}T_{\text{canopy}}}}{\partial T_{\text{canopy}}} = \frac{\frac{\partial R_n}{\partial T_{\text{canopy}}} - \frac{\partial H}{\partial T_{\text{canopy}}} - \frac{\partial L}{\partial q_{\text{sat}}}\cdot\frac{\partial q_{\text{sat}}}{\partial T_{\text{canopy}}}}{\rho c h \cdot \mathcal{A}} \tag{F13}$$

Here, $\mathcal{T}_{\text{imp\_}x}$ is the implicit tendency used to advance the prognostic variable x. The first 3 equations in this set are part of the soil model, and the last equation comes from the canopy model. The variables on the right side of the soil equations are defined in Table 1, and the variables in the canopy equation are defined in Section 2.3.2. Exceptions are $\rho_l e_l$, which is the volumetric internal energy of liquid water, and q_{sat}, which is the humidity over water at a particular temperature and pressure (used in place of q_{canopy} here).

In terms of structure, the Jacobian matrix has one block for each term $\frac{\partial \mathcal{T}_{\text{imp\_}x_i}}{\partial x_j}$, where x_i, x_j are any prognostic variables. The blocks for the 3 nonzero equations from the soil model have a tridiagonal matrix structure, since the value at each vertical layer in the soil depends on 3 layers of soil: itself, the layer directly above, and the layer directly below. The block for the non-zero equation from the canopy model has a diagonal structure, since the canopy model does not have depth and is evaluated only at the surface. When these blocks are combined together, the overall Jacobian is nearly a diagonal matrix, with the exception of the non-zero off-diagonal term $\frac{\partial \mathcal{T}_{\text{imp\_}\rho e_{int}}}{\partial \vartheta_l}$. The sparsity of this matrix makes the equation $A\boldsymbol{x} = \boldsymbol{b}$ in Equation (F9) numerically easier to solve than if we were to use the dense Jacobian without any approximations.

We currently ignore contributions to the Jacobian from surface boundary conditions of the soil except in the case of Dirichlet conditions on ϑ_l. These contributions are straightforward to add and should improve the stability of the timestepping, since most of the stiffness in the equations comes from the thin layers near the surface. For more details about the timestepping and numerical methods used by CliMA models, see Yatunin et al. (2025).

Appendix G: Dimensionality and Source of Model Parameters

Model parameters are summarized in Tables G1–G3. The spatial dimensionality is stated from the point of view of global simulations.

Table G1
Parameters of ClimaLand's Soil Model, Version 1

Parameter	Spatial dimension	Source
ν_p, θ_{res}, K_{sat} α, n	3	Fixed (Gupta et al., 2020, 2022)
$\boldsymbol{\nu}$	3	Fixed (Hengl et al., 2017)
S_s	0	To be calibrated
f_{over}	0	To be calibrated
R_{sb}	0	To be calibrated
$\alpha_{\text{soil, wet}}, \alpha_{\text{soil, dry}}$	2	Fixed (D. M. Lawrence et al., 2019)
ϵ_{soil}	0	To be calibrated
$z_{0,m}, z_{0,b}, d$	0	Fixed
$d_{d,s}$	0	To be calibrated
α, β (Kersten model)	0	Fixed (Balland & Arp, 2005)
a (κ_{dry} model)	0	Fixed (Balland & Arp, 2005)
γ, Ω (factors adjusting K)	0	Fixed (Hansson et al., 2004)

Note. Physical constants, such as the density of water, will not be calibrated and are given in Table 1

Table G2
Key Parameters of ClimaLand's Canopy Model, Version 1

Parameter	Spatial dimension	Source
g_0	0	Fixed (G. Bonan, 2019)
g_1	2	To be calibrated (D. M. Lawrence et al., 2019)
$\chi, \alpha_{\mathrm{PAR}}, \alpha_{\mathrm{NIR}}, \tau_{\mathrm{PAR}}, \tau_{\mathrm{NIR}}, \epsilon, \Omega_{cl}$	2	To be calibrated (D. M. Lawrence et al., 2019)
$\mathrm{V}_{\mathrm{cmax},25}$	2	To be calibrated (D. M. Lawrence et al., 2019)
C3 versus C4 flag	2	Fixed (D. M. Lawrence et al., 2019)
s_c, p_c (moisture stress factors)	0	To be calibrated
θ_j	0	To be calibrated
ϕ	0	To be calibrated
$K_{c,25}, K_{o,25}$	0	Fixed (G. Bonan, 2019)
Γ^*_{25}	0	Fixed (G. Bonan, 2019)
$\Delta H_{kc,ko,V_{\mathrm{cmax}},\Gamma^*,J_{\mathrm{max}},R_d}$	0	Fixed (G. Bonan, 2019)
o_i	0	Fixed
f	0	Fixed
$J_{\mathrm{max},25}/\mathrm{V}_{\mathrm{cmax},25}$	0	Fixed
PFT map	2	Fixed (D. M. Lawrence et al., 2019)
ρch	0	To be calibrated
$h_{\mathrm{stem}}, h_{\mathrm{leaf}}, \mathrm{SAI}$	0	Fixed
factor relating RAI to SAI and LAI	0	fixed
Rooting depth	2	Fixed (D. M. Lawrence et al., 2019)
$K_{\mathrm{sat}}, \psi_{63}, c$ (Weibull permeability)	0	To be calibrated
a_1, a_2, ν, S_s (retention curve parameters)	0	To be calibrated
LAI	2	Prescribed (MODIS or ERA5)

Note. All 2D fields come from plant functional type maps.

Table G3
Parameters of ClimaLand's Snow Model, Version 1

Parameter	Spatial dimension	Source
Neural depth parameters	0	Fixed (Charbonneau et al., 2025)
$z_{0,m}, z_{0,b}, d$	0	Fixed
α_{snow}	0	To be calibrated
ϵ_{snow}	0	Fixed
θ_r	0	To be calibrated
K_{sat}	0	To be calibrated

Note. Physical constants, such as the density of water, will not be calibrated and are given in Table 2.

Conflict of Interest

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

Data Availability Statement

All code used in the examples is available (Deck et al., 2025), and the required data will download automatically when scripts are run, with the exception of hourly forcing data from ERA5 due to file size constraints and the

layered soil infiltration measurement data. The ERA5 data was obtained from Hersbach et al. (2023). The measurement data used in the layered soil infiltration example is checked into the repository on a protected branch of Deck et al. (2025): https://github.com/CliMA/ClimaLand.jl/tree/paper/layered_soil_plots.

Acknowledgments
This research was supported by Schmidt Sciences, LLC, and by the Resnick Sustainability Institute at Caltech. Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. We thank Lee Barbour and Mingbin Huang for providing the SV62 site data for the layered soil experiment, Dani Or and Peter Lehmann for providing the bare soil evaporation data, Yongjiu Dai and Nan Wei for helping us to understand the SoilGrids data, and Natan Holtzmann for aiding our initial flux tower simulations. We would also like to thank Elias Massoud, John Worden, Alex Norton, Victoria Meyer and Paul Levine for early discussions regarding ClimaLand. We acknowledge high-performance computing support from Caltech's Resnick High Performance Computing Center and the Derecho system (Computational and Information Systems Laboratory, 2023) provided by the National Center for Atmospheric Research (NCAR), sponsored by the National Science Foundation.

References

Ameriflux. (2015). Remote sensing. Retrieved from https://ameriflux.lbl.gov
Anderson, E. A. (1976). *A point energy and mass balance model of a snow cover*. Stanford University.
Balland, V., & Arp, P. A. (2005). Modeling soil thermal conductivities over a wide range of conditions. *Journal of Environmental Engineering and Science*, *4*(6), 549–558. https://doi.org/10.1139/s05-007
Bartelt, P., & Lehning, M. (2002). A physical SNOWPACK model for the Swiss avalanche warning: Part I: Numerical model. *Cold Regions Science and Technology*, *35*(3), 123–145. https://doi.org/10.1016/s0165-232x(02)00074-5
Bear, J. (2018). *Modeling phenomena of flow and transport in porous media*. Springer.
Beucler, T., Pritchard, M., Rasp, S., Ott, J., Baldi, P., & Gentine, P. (2021). Enforcing analytic constraints in neural networks emulating physical systems. *Physical Review Letters*, *126*(9), 098302. https://doi.org/10.1103/physrevlett.126.098302
Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Journal*, *24*(1), 43–69. https://doi.org/10.1080/02626667909491834
Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, *59*(1), 65–98. https://doi.org/10.1137/141000671
Bisht, G., Riley, W. J., Hammond, G. E., & Lorenzetti, D. M. (2018). Development and evaluation of a variably saturated flow model in the global E3SM Land Model (ELM) version 1.0. *Geoscientific Model Development*, *11*(10), 4085–4102. https://doi.org/10.5194/gmd-11-4085-2018
Bisht, G., Riley, W. J., Wainwright, H. M., Dafflon, B., Yuan, F., & Romanovsky, V. E. (2018). Impacts of microtopographic snow redistribution and lateral subsurface processes on hydrologic and thermal states in an Arctic polygonal ground ecosystem: A case study using ELM-3D v1. 0. *Geoscientific Model Development*, *11*(1), 61–76. https://doi.org/10.5194/gmd-11-61-2018
Bonan, G. (2019). *Climate change and terrestrial ecosystem modeling*. Cambridge Univ. Press.
Bonan, G. B., Levis, S., Kergoat, L., & Oleson, K. W. (2002). Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models. *Global Biogeochemical Cycles*, *16*(2), 5-1. https://doi.org/10.1029/2000GB001360
Bonan, G. B., Lucier, O., Coen, D. R., Foster, A. C., Shuman, J. K., Laguë, M. M., et al. (2024). Reimagining Earth in the Earth system. *Journal of Advances in Modeling Earth Systems*, *16*(8), e2023MS004017. https://doi.org/10.1029/2023MS004017
Braghiere, R. K., Fisher, J. B., Miner, K. R., Miller, C. E., Worden, J. R., Schimel, D. S., & Frankenberg, C. (2023). Tipping point in North American Arctic-boreal carbon sink persists in new generation Earth system models despite reduced uncertainty. *Environmental Research Letters*, *18*(2), 025008. https://doi.org/10.1088/1748-9326/acb226
Braghiere, R. K., Quaife, T., Black, E., He, L., & Chen, J. M. (2019). Underestimation of global photosynthesis in Earth system models due to representation of vegetation structure. *Global Biogeochemical Cycles*, *33*(11), 1358–1369. https://doi.org/10.1029/2018GB006135
Braghiere, R. K., Wang, Y., Doughty, R., Sousa, D., Magney, T., Widlowski, J. L., et al. (2021). Accounting for canopy structure improves hyperspectral radiative transfer and sun-induced chlorophyll fluorescence representations in a new generation Earth system model. *Remote Sensing of Environment*, *261*, 112497. https://doi.org/10.1016/J.RSE.2021.112497
Braghiere, R. K., Wang, Y., Gagné-Landmann, A., Brodrick, P. G., Bloom, A. A., Norton, A. J., et al. (2023). The importance of hyperspectral soil albedo information for improving Earth system model projections. *AGU Advances*, *4*, e2023AV000910. https://doi.org/10.1029/2023av000910
Brooks, R. J., & Corey, A. T. (1964). *Hydraulic properties of porous media (Tech. Rep.)* (Vol. 3). Colorado State University.
Cai, W., Zhu, Z., Harrison, S. P., Ryu, Y., Wang, H., Zhou, B., & Prentice, I. C. (2025). A unifying principle for global greenness patterns and trends. *Communications Earth & Environment*, *6*(1), 19. https://doi.org/10.1038/s43247-025-01992-0
Carslaw, H. S., & Jaeger, J. C. (1959). Conduction of heat in solids.
Charbonneau, A., Deck, K., & Schneider, T. (2025). A physics-constrained neural differential equation framework for data-driven snowpack simulation. *Artificial Intelligence for the Earth Systems*, *4*(3), 240040. https://doi.org/10.1175/aies-d-24-0040.1
Christopoulos, C., Lopez-Gomez, I., Beucler, T., Cohen, Y., Kawczynski, C., Dunbar, O. R. A., & Schneider, T. (2024). Online learning of entrainment closures in a hybrid machine learning parameterization. *Journal of Advances in Modeling Earth Systems*, *16*, e2024MS004485. https://doi.org/10.1029/2024MS004485
Clark, D., Mercado, L., Sitch, S., Jones, C. D., Gedney, N., Best, M., et al. (2011). The Joint UK Land Environment Simulator (JULES), model description. Part 2: Carbon fluxes and vegetation dynamics. *Geoscientific Model Development*, *4*(3), 701–722. https://doi.org/10.5194/gmd-4-701-2011
Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., et al. (2015). Improving the representation of hydrologic processes in Earth system models. *Water Resources Research*, *51*(8), 5929–5956. https://doi.org/10.1002/2015wr017096
Computational and Information Systems Laboratory. (2023). Derecho: HPE Cray EX System (University Community Computing) [Hardware]. https://doi.org/10.5065/qx9a-pg09
Cranko Page, J., Abramowitz, G., De Kauwe, M. G., & Pitman, A. J. (2024). Are plant functional types fit for purpose? *Geophysical Research Letters*, *51*(1), e2023GL104962. https://doi.org/10.1029/2023gl104962
Dagon, K., Sanderson, B. M., Fisher, R. A., & Lawrence, D. M. (2020). A machine learning approach to emulation and biophysical parameter estimation with the Community Land Model, version 5. Advances in Statistical climatology. *Meteorology and Oceanography*, *6*(2), 223–244. https://doi.org/10.5194/ascmo-6-223-2020
Dai, Y., Wei, N., Yuan, H., Zhang, S., Shangguan, W., Liu, S., et al. (2019). Evaluation of soil thermal conductivity schemes for use in land surface modeling. *Journal of Advances in Modeling Earth Systems*, *11*, 3454–3473. https://doi.org/10.1029/2019ms001723
Dall'Amico, E., Endrizzi, S., Gruber, S., & Rigon, R. (2011). A robust and energy-conserving model of freezing variably-saturated soil. *The Cryosphere*, *5*(2), 469–484. https://doi.org/10.5194/tc-5-469-2011
Danabasoglu, G., Lamarque, J., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., et al. (2020). The Community Earth System Model Version 2 (CESM2). *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS001916. https://doi.org/10.1029/2019MS001916
Davidson, E. A., Samanta, S., Caramori, S. S., & Savage, K. (2012). The Dual Arrhenius and Michaelis–Menten kinetics model for decomposition of soil organic matter at hourly to seasonal time scales. *Global Change Biology*, *18*(1), 371–384. https://doi.org/10.1111/J.1365-2486.2011.02546.X

Deck, K., Sloan, J., Bozzola, G., Renchon, A. A., Phan, K., Reddy, T., et al. (2025). ClimaLand.jl: V1.0.0 [Software]. *Zenodo*. https://doi.org/10.5281/ZENODO.17373248

Dingman, S. L. (2015). *Physical hydrology* (3rd ed.). Waveland Press.

Dunbar, O. R. A., Lopez-Gomez, I., Garbuno-Iñigo, A., Huang, D. Z., Bach, E., & Wu, J.-l. (2022). EnsembleKalmanProcesses.jl: Derivative-free ensemble-based model calibration. *Journal of Open Source Software*, *7*(80), 4869. https://doi.org/10.21105/joss.04869

Dunne, T., Moore, T., & Taylor, C. (1975). Recognition and prediction of runoff-producing zones in humid regions. *Hydrological Sciences Bulletin*, *20*(3), 305–327.

Duursma, R. A., & Medlyn, B. E. (2012). MAESPA: A model to study interactions between water limitation, environmental drivers and vegetation function at tree and stand levels, with an example application to CO2× drought interactions. *Geoscientific Model Development*, *5*(4), 919–940. https://doi.org/10.5194/gmd-5-919-2012

Egea, G., Verhoef, A., & Vidale, P. L. (2011). Towards an improved and more flexible representation of water stress in coupled photosynthesis–stomatal conductance models. *Agricultural and Forest Meteorology*, *151*(10), 1370–1384. https://doi.org/10.1016/j.agrformet.2011.05.019

ElGhawi, R., Kraft, B., Reimers, C., Reichstein, M., Körner, M., Gentine, P., & Winkler, A. J. (2023). Hybrid modeling of evapotranspiration: Inferring stomatal and aerodynamic resistances using combined physics-based and machine learning. *Environmental Research Letters*, *18*(3), 034039. https://doi.org/10.1088/1748-9326/acbbe0

Endrizzi, S., Dall'Amico, M., & Rigon, R. (2014). GEOtop 2.0: Simulating the combined energy and water balance at and below the land surface accounting for soil freezing, snow cover and terrain effects. *Geoscientific Model Development*, *7*(6), 2831–2857. https://doi.org/10.5194/gmd-7-2831-2014

Entekhabi, D., & Eagleson, P. S. (1989). Land surface hydrology parameterization for atmospheric general circulation models including subgrid scale spatial variability. *Journal of Climate*, *2*(8), 816–831. https://doi.org/10.1175/1520-0442(1989)002<0816:lshpfa>2.0.co;2

Eyring, V., Collins, W. D., Gentine, P., Barnes, E. A., Barreiro, M., Beucler, T., et al. (2024). Pushing the frontiers in climate modelling and analysis with machine learning. *Nature Climate Change*, *14*(9), 916–928. https://doi.org/10.1038/s41558-024-02095-y

Fang, J., & Gentine, P. (2024). Exploring optimal complexity for water stress representation in terrestrial carbon models: A hybrid-machine learning model approach. *Journal of Advances in Modeling Earth Systems*, *16*(12), e2024MS004308. https://doi.org/10.1029/2024ms004308

Färber, C., Plessow, H., Mischel, S., Kratzert, F., Addor, N., Shalev, G., & Looser, U. (2024). GRDC-Caravan: Extending caravan with data from the Global Runoff Data Centre. *Earth System Science Data Discussions*, *2024*, 1–17.

Farouki, O. T. (1981). *Thermal properties of soils (Tech. Rep.). Hanover, NH: United States army corps of engineers*. Cold Regions Research and Engineering Laboratory.

Farquhar, G. D., von Caemmerer, S. V., & Berry, J. A. (1980). A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. *Planta*, *149*(1), 78–90. https://doi.org/10.1007/bf00386231

Fatichi, S., Pappas, C., Zscheischler, J., & Leuzinger, S. (2019). Modelling carbon sources and sinks in terrestrial vegetation. *New Phytologist*, *221*(2), 652–668. https://doi.org/10.1111/NPH.15451

Fischer, E. M., Lawrence, D. M., & Sanderson, B. M. (2011). Quantifying uncertainties in projections of extremes—A perturbed land surface parameter experiment. *Climate Dynamics*, *37*(7–8), 1381–1398. https://doi.org/10.1007/s00382-010-0915-y

Flanner, M. G., Arnheim, J., Cook, J. M., Dang, C., He, C., Huang, X., et al. (2021). SNICAR-AD v3: A community tool for modeling spectral snow albedo. *Geoscientific Model Development Discussions*, *2021*(12), 1–49. https://doi.org/10.5194/gmd-14-7673-2021

Franklin, O., Harrison, S. P., Dewar, R., Farrior, C. E., Åke, B., Dieckmann, U., et al. (2020). Organizing principles for vegetation dynamics. *Nature Plants*, *6*(5), 444–453. https://doi.org/10.1038/s41477-020-0655-x

Freeze, R. A. (1974). Streamflow generation. *Reviews of Geophysics*, *12*(4), 627–647. https://doi.org/10.1029/rg012i004p00627

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C., Hauck, J., et al. (2023). Global carbon budget 2023. *Earth System Science Data*, *15*(12), 5301–5369. https://doi.org/10.5194/ESSD-15-5301-2023

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Landschützer, P., et al. (2025). Global carbon budget 2024. *Earth System Science Data*, *17*(3), 965–1039. https://doi.org/10.5194/essd-17-965-2025

Gerke, H. H., & Van Genuchten, M. T. (1993). A dual-porosity model for simulating the preferential movement of water and solutes in structured porous media. *Water Resources Research*, *29*(2), 305–319. https://doi.org/10.1029/92wr02339

Goudriaan, J. (1977). *Crop micrometeorology: A simulation study*. Wageningen University and Research.

Gupta, S., Hengl, T., Lehmann, P., Bonetti, S., & Or, D. (2020). SoilKsatDB: Global soil saturated hydraulic conductivity measurements for geoscience applications [Dataset]. *Earth System Science Data Discussions*, *2020*, 1–26. https://doi.org/10.5194/essd-13-1593-2021

Gupta, S., Papritz, A., Lehmann, P., Hengl, T., Bonetti, S., & Or, D. (2022). Global soil hydraulic properties dataset based on legacy site observations and robust parameterization [Dataset]. *Scientific Data*, *9*(1), 444. https://doi.org/10.1038/s41597-022-01481-5

Hansson, K., Šimůnek, J., Mizoguchi, M., Lundin, L.-C., & van Genuchten, M. T. (2004). Water flow and heat transport in frozen soil: Numerical solution and freeze–thaw applications. *Vadose Zone Journal*, *3*, 693–704.

Harrison, S. P., Cramer, W., Franklin, O., Prentice, I. C., Wang, H., Brännström, Å., et al. (2021). Eco-evolutionary optimality as a means to improve vegetation and land-surface models. *New Phytologist*, *231*(6), 2125–2141. https://doi.org/10.1111/nph.17558

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One*, *12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Sabater, J. M., et al. (2023). ERA5 hourly data on pressure levels from 1940 to present [Dataset]. https://doi.org/10.24381/cds.adbb2d47

Heskel, M. A., O'Sullivan, O. S., Reich, P. B., Tjoelker, M. G., Weerasinghe, L. K., Penillard, A., et al. (2016). Convergence in the temperature response of leaf respiration across biomes and plant functional types. *Proceedings of the National Academy of Sciences*, *113*(14), 3832–3837. https://doi.org/10.1073/pnas.1520282113

Hewlett, J. D., & Hibbert, A. R. (1967). Factors affecting the response of small watersheds to precipitation in humid areas. *Forest hydrology*, *1*, 275–290.

Horton, R. E. (1933). The role of infiltration in the hydrologic cycle. *Eos, Transactions American Geophysical Union*, *14*(1), 446–460.

Huang, D. Z., Schneider, T., & Stuart, A. M. (2022). Iterated Kalman methodology for inverse problems. *Journal of Computational Physics*, *463*, 111262. https://doi.org/10.1016/j.jcp.2022.111262

Huang, M. E. A., Lee Barbour, S., Elshorbagy, A., Zettl, J. D., & Cheng Si, B. (2011). Infiltration and drainage processes in multi-layered coarse soils. *Canadian Journal of Soil Science*, *91*(2), 169–183. https://doi.org/10.4141/cjss09118

Iglesias, M. A., Law, K. J., & Stuart, A. M. (2013). Ensemble Kalman methods for inverse problems. *Inverse Problems*, *29*(4), 045001. https://doi.org/10.1088/0266-5611/29/4/045001

IPCC. (2023). Global carbon and other biogeochemical cycles and feedbacks. In *Climate change 2021—The physical science basis* (pp. 673–816). https://doi.org/10.1017/9781009157896.007

Kattge, J., & Knorr, W. (2007). Temperature acclimation in a biochemical model of photosynthesis: A reanalysis of data from 36 species. *Plant, Cell and Environment*, *30*(9), 1176–1190. https://doi.org/10.1111/j.1365-3040.2007.01690.x

Kovachki, N. B., & Stuart, A. M. (2019). Ensemble Kalman inversion: A derivative-free technique for machine learning tasks. *Inverse Problems*, *35*(9), 095005. https://doi.org/10.1088/1361-6420/ab1c3a

Kraft, B., Jung, M., Körner, M., Koirala, S., & Reichstein, M. (2021). Towards hybrid modeling of the global hydrological cycle. *Hydrology and Earth System Sciences Discussions*, *2021*, 1–40.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*, *22*(11), 6005–6022. https://doi.org/10.5194/hess-22-6005-2018

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., & Nearing, G. (2019). Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, *55*(12), 11344–11354. https://doi.org/10.1029/2019wr026065

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences*, *23*(12), 5089–5110. https://doi.org/10.5194/hess-23-5089-2019

Krinner, G., Derksen, C., Essery, R., Flanner, M., Hagemann, S., Clark, M., et al. (2018). ESM-SnowMIP: Assessing snow models and quantifying snow-related climate feedbacks. *Geoscientific Model Development*, *11*(12), 5027–5049. https://doi.org/10.5194/gmd-11-5027-2018

Kurylyk, B. L., & Watanabe, K. (2013). The mathematical representation of freezing and thawing processes in variably-saturated, non-deformable soils. *Advances in Water Resources*, *60*, 160–177. https://doi.org/10.1016/j.advwatres.2013.07.016

Lai, C.-Y., Hassanzadeh, P., Sheshadri, A., Sonnewald, M., Ferrari, R., & Balaji, V. (2025). Machine learning for climate physics and simulations. *Annual Review of Condensed Matter Physics*, *16*(1), 343–365. https://doi.org/10.1146/annurev-conmatphys-043024-114758

Landerer, F. W., & Swenson, S. (2012). Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research*, *48*(4), W04531. https://doi.org/10.1029/2011wr011453

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., et al. (2019). 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. https://doi.org/10.1029/2018ms001583

Lawrence, P. J., & Chase, T. N. (2007). Representing a new MODIS consistent land surface in the Community Land Model (CLM 3.0). *Journal of Geophysical Research*, *112*(G1), G01023. https://doi.org/10.1029/2006JG000168

Lee, J.-E., Berry, J. A., van der Tol, C., Yang, X., Guanter, L., Damm, A., et al. (2015). Simulations of chlorophyll fluorescence incorporated into the Community Land Model version 4. *Global Change Biology*, *21*(9), 3469–3477. https://doi.org/10.1111/gcb.12948

Lehmann, P., Assouline, S., & Or, D. (2008). Characteristic lengths affecting evaporative drying of porous media. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, *77*(5), 056309. https://doi.org/10.1103/physreve.77.056309

Lehmann, P., Merlin, O., Gentine, P., & Or, D. (2018). Soil texture effects on surface resistance to bare-soil evaporation. *Geophysical Research Letters*, *45*(19), 10–398. https://doi.org/10.1029/2018gl078803

Lejeune, Y., Dumont, M., Panel, J.-M., Lafaysse, M., Lapalus, P., Le Gac, E., et al. (2019). 57 years (1960–2017) of snow and meteorological observations from a mid-altitude mountain site (Col de Porte, France, 1325 m of altitude). *Earth System Science Data*, *11*(1), 71–88. https://doi.org/10.5194/essd-11-71-2019

Leuning, R. (2002). Temperature dependence of two parameters in a photosynthesis model. *Plant, Cell and Environment*, *25*(9), 1205–1210. https://doi.org/10.1046/j.1365-3040.2002.00898.x

Lima, M., Deck, K., Dunbar, O. R., & Schneider, T. (2024). Toward routing river water in land surface models with recurrent neural networks. *arXiv preprint arXiv:2404.14212*.

Longo, M., Knox, R. G., Medvigy, D. M., Levine, N. M., Dietze, M. C., Kim, Y., et al. (2019). The biophysics, ecology, and biogeochemistry of functionally diverse, vertically and horizontally heterogeneous ecosystems: The Ecosystem Demography model, version 2.2—Part 1: Model description. *Geoscientific Model Development*, *12*(10), 4309–4346. https://doi.org/10.5194/gmd-12-4309-2019

Lopez-Gomez, I., Christopoulos, C., Ervik, H. L., Dunbar, O. R. A., Cohen, Y., & Schneider, T. (2022). Training physics-based machine-learning parameterizations with gradient-free ensemble Kalman methods. *Journal of Advances in Modeling Earth Systems*, *14*(8), e2022MS003105. https://doi.org/10.1029/2022ms003105

Lundin, L.-C. (1990). Hydraulic properties in an operational model of frozen soil. *Journal of Hydrology*, *118*(1–4), 289–310. https://doi.org/10.1016/0022-1694(90)90264-x

Ma, S., Baldocchi, D. D., Xu, L., & Hehn, T. (2007). Inter-annual variability in carbon dioxide exchange of an oak/grass savanna and open grassland in California. *Agricultural and Forest Meteorology*, *147*(3–4), 157–171. https://doi.org/10.1016/j.agrformet.2007.07.008

Ma, S., Xu, L., Verfaillie, J., & Baldocchi, D. (2023). *AmeriFlux FLUXNET-1F US-Var vaira ranch-ione (technical report)*. Lawrence Berkeley National Laboratory (LBNL).

Margulis, S. (2017). Introduction to hydrology.

Marthews, T., Dadson, S., Lehner, B., Abele, S., & Gedney, N. (2015). High-resolution global topographic index values for use in large-scale hydrological modelling. *Hydrology and Earth System Sciences*, *19*(1), 91–104. https://doi.org/10.5194/hess-19-91-2015

Medlyn, B. E., Barton, C., Broadmeadow, M., Ceulemans, R., De Angelis, P., Forstreuter, M., et al. (2001). Stomatal conductance of forest species after long-term exposure to elevated CO2 concentration: A synthesis. *New Phytologist*, *149*(2), 247–264. https://doi.org/10.1046/j.1469-8137.2001.00028.x

Ménard, C. B., Essery, R., Barr, A., Bartlett, P., Derry, J., Dumont, M., et al. (2019). Meteorological and evaluation datasets for snow modelling at 10 reference sites: Description of in situ and bias-corrected reanalysis data. *Earth System Science Data*, *11*(2), 865–880. https://doi.org/10.5194/essd-11-865-2019

Mengoli, G., Agustí-Panareda, A., Boussetta, S., Harrison, S. P., Trotta, C., & Prentice, I. C. (2022). Ecosystem photosynthesis in land-surface models: A first-principles approach incorporating acclimation. *Journal of Advances in Modeling Earth Systems*, *14*(1), e2021MS002767. https://doi.org/10.1029/2021ms002767

Mizoguchi, M. (1990). *Water, heat and salt transport in freezing soil*. Ph.D. thesis. (In Japanese.).

Mualem, Y. (1976). A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resources Research*, *12*(3), 513–522. https://doi.org/10.1029/wr012i003p00513

Myneni, R., Knyazikhin, Y., & Park, T. (2015). MOD15A2H MODIS/Terra leaf area index/FPAR 8-day l4 global 500m SIN grid V006. https://doi.org/10.5067/MODIS/MCD15A2H.006

Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., et al. (2024). Global prediction of extreme floods in ungauged watersheds. *Nature*, *627*(8004), 559–563. https://doi.org/10.1038/s41586-024-07145-1

Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., & Gulden, L. E. (2005). A simple TOPMODEL-based runoff parameterization (SIMTOP) for use in global climate models. *Journal of Geophysical Research*, *110*(D21). https://doi.org/10.1029/2005jd006111

Norton, A. J., Bloom, A. A., Parazoo, N. C., Levine, P. A., Ma, S., Braghiere, R. K., & Smallman, T. L. (2023). Improved process representation of leaf phenology significantly shifts climate sensitivity of ecosystem carbon balance. *Biogeosciences*, *20*(12), 2455–2484. https://doi.org/10.5194/BG-20-2455-2023

Or, D., Lehmann, P., Shahraeeni, E., & Shokri, N. (2013). Advances in soil evaporation physics—a review. *Vadose Zone Journal*, *12*(4), 1–16. https://doi.org/10.2136/vzj2012.0163

Painter, S. L. (2011). Three-phase numerical model of water migration in partially frozen geological media: Model formulation, validation, and applications. *Computational Geosciences*, *15*(1), 69–85. https://doi.org/10.1007/s10596-010-9197-z

Painter, S. L., Coon, E. T., Atchley, A. L., Berndt, M., Garimella, R., Moulton, J. D., et al. (2016). Integrated surface/subsurface permafrost thermal hydrology: Model formulation and proof-of-concept simulations. *Water Resources Research*, *52*(8), 6062–6077. https://doi.org/10.1002/2015wr018427

Painter, S. L., & Karra, S. (2014). Constitutive model for unfrozen water content in subfreezing unsaturated soils. *Vadose Zone Journal*, *13*(4), 1–8. https://doi.org/10.2136/vzj2013.04.0071

Perezhogin, P., Zhang, C., Adcroft, A., Fernandez-Granda, C., & Zanna, L. (2024). A stable implementation of a data-driven scale-aware mesoscale parameterization. *Journal of Advances in Modeling Earth Systems*, *16*(10), e2023MS004104. https://doi.org/10.1029/2023ms004104

Perovich, D. K. (2007). Light reflection and transmission by a temperate snow cover. *Journal of Glaciology*, *53*(181), 201–210. https://doi.org/10.3189/172756507782202919

Qiu, H., Bisht, G., Li, L., Hao, D., & Xu, D. (2024). Development of inter-grid-cell lateral unsaturated and saturated flow model in the E3SM Land Model (v2. 0). *Geoscientific Model Development*, *17*(1), 143–167. https://doi.org/10.5194/gmd-17-143-2024

Räisäenen, P. (2002). Two-stream approximations revisited: A new improvement and tests with GCM data. *Quarterly Journal of the Royal Meteorological Society*, *128*(585), 2397–2416. https://doi.org/10.1256/qj.01.161

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., & Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Ren, Y., Wang, H., Harrison, S. P., Prentice, I. C., Mengoli, G., Zhao, L., et al. (2025). Incorporating the acclimation of photosynthesis and leaf respiration in the Noah-MP land surface model: Model development and evaluation. *Journal of Advances in Modeling Earth Systems*, *17*(3), e2024MS004599. https://doi.org/10.1029/2024ms004599

Ross, J. (1981). The radiation regime and Architecture of plant stands. https://doi.org/10.1007/978-94-009-8647-3

Ryan, E. M., Ogle, K., Kropp, H., Samuels-Crow, K. E., Carrillo, Y., & Pendall, E. (2018). Modeling soil CO2 production and transport with dynamic source and diffusion terms: Testing the steady-state assumption using DETECT v1. 0. *Geoscientific Model Development*, *11*(5), 1909–1928. https://doi.org/10.5194/gmd-11-1909-2018

Saunders, A., Drew, D. M., & Brink, W. (2021). Machine learning models perform better than traditional empirical models for stomatal conductance when applied to multiple tree species across different forest biomes. *Trees, Forests and People*, *6*, 100139. https://doi.org/10.1016/j.tfp.2021.100139

Schaaf, C., & Wang, Z. (2015). MCD43A1 MODIS/Terra+ Aqua BRDF/Albedo model parameters daily L3 global-500m v006. In *NASA EOSDIS Land Processes Distributed Active Archive Center (DAAC) data set, MCD43A1-006*.

Schädel, C., Rogers, B. M., Lawrence, D. M., Koven, C. D., Brovkin, V., Burke, E. J., et al. (2024). Earth system models must include permafrost carbon processes. *Nature Climate Change*, *14*(2), 114–116. https://doi.org/10.1038/s41558-023-01909-9

Schneider, T., Lan, S., Stuart, A., & Teixeira, J. (2017). Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophysical Research Letters*, *44*(24), 12396–12417. https://doi.org/10.1002/2017GL076101

Schneider, T., Stuart, A. M., & Wu, J.-L. (2022). Ensemble Kalman inversion for sparse learning of dynamical systems from time-averaged data. *Journal of Computational Physics*, *470*, 111559. https://doi.org/10.1016/j.jcp.2022.111559

Schneider, T., Teixeira, J., Bretherton, C. S., Brient, F., Pressel, K. G., Schär, C., & Siebesma, A. P. (2017). Climate goals and computing the future of clouds. *Nature Climate Change*, *7*(1), 3–5. https://doi.org/10.1038/nclimate3190

Sellers, P. (1985). Canopy reflectance, photosynthesis and transpiration. *International Journal of Remote Sensing*, *6*(8), 1335–1372. https://doi.org/10.1080/01431168508948283

Semtner, A. J., Jr. (1976). A model for the thermodynamic growth of sea ice in numerical investigations of climate. *Journal of Physical Oceanography*, *6*(3), 379–389. https://doi.org/10.1175/1520-0485(1976)006<0379:amfttg>2.0.co;2

Simunek, J., Van Genuchten, M. T., & Sejna, M. (2005). The HYDRUS-1d software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media. In *University of California-Riverside research reports* (Vol. 3, pp. 1–240).

Sivapalan, M., Beven, K., & Wood, E. F. (1987). On hydrologic similarity: 2. a scaled model of storm runoff production. *Water Resources Research*, *23*(12), 2266–2278. https://doi.org/10.1029/wr023i012p02266

Stocker, B. D., Wang, H., Smith, N. G., Harrison, S. P., Keenan, T. F., Sandoval, D., et al. (2020). P-model v1.0: An optimality-based light use efficiency model for simulating ecosystem gross primary production. *Geoscientific Model Development*, *13*(3), 1545–1581. https://doi.org/10.5194/gmd-13-1545-2020

Swenson, S. C., Lawrence, D. M., & Lee, H. (2012). Improved simulation of the terrestrial hydrological cycle in permafrost regions by the Community Land Model. *Journal of Advances in Modeling Earth Systems*, *4*(3), M08002. https://doi.org/10.1029/2012ms000165

Tao, F., Huang, Y., Hungate, B. A., Manzoni, S., Frey, S. D., Schmidt, M. W., et al. (2023). Microbial carbon use efficiency promotes global soil carbon storage. *Nature*, *618*(7967), 981–985. https://doi.org/10.1038/s41586-023-06042-3

Tarboton, D. G., & Luce, C. H. (1996). *Utah energy balance snow accumulation and melt model (UEB)*. Citeseer.

Tuzet, A., Perrier, A., & Leuning, R. (2003). A coupled model of stomatal conductance, photosynthesis and transpiration. *Plant, Cell and Environment*, *26*(7), 1097–1116. https://doi.org/10.1046/j.1365-3040.2003.01035.x

van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Amer., J.*, 892–898.

Walko, R. L., Band, L. E., Baron, J., Kittel, T. G., Lammers, R., Lee, T. J., et al. (2000). Coupled atmosphere–biophysics–hydrology models for environmental modeling. *Journal of Applied Meteorology*, *39*(6), 931–944. https://doi.org/10.1175/1520-0450(2000)039<0931:cabhmf>2.0.co;2

Wang, D., Schwartz, P., Yuan, F., Thornton, P., & Zheng, W. (2022). Toward ultrahigh-resolution E3SM land modeling on exascale computers. *Computing in Science & Engineering*, *24*(6), 44–53. https://doi.org/10.1109/MCSE.2022.3218990

Wang, H., Prentice, I. C., Keenan, T. F., Davis, T. W., Wright, I. J., Cornwell, W. K., et al. (2017). Towards a universal model for carbon dioxide uptake by plants. *Nature Plants*, *3*(9), 734–741. https://doi.org/10.1038/s41477-017-0006-8

Wang, J., Bretz, M., Dewan, M. A. A., & Delavar, M. A. (2022). Machine learning in modelling land-use and land cover-change (lulcc): Current status, challenges and prospects. *Science of the Total Environment*, *822*, 153559. https://doi.org/10.1016/j.scitotenv.2022.153559

Wang, Y., Braghiere, R. K., Fischer, W. W., Yao, Y., Shen, Z., Schneider, T., et al. (2025). Impacts of leaf traits on vegetation optical properties in Earth system modeling. *Nature Communications*, *16*(1), 4968. https://doi.org/10.1038/s41467-025-60149-x

Wang, Y., Braghiere, R. K., Longo, M., Norton, A. J., Köhler, P., Doughty, R., et al. (2023). Modeling global vegetation gross primary productivity, transpiration and hyperspectral canopy radiative transfer simultaneously using a next generation land surface model—CliMA Land. *Journal of Advances in Modeling Earth Systems*, *15*(3), e2021MS002964. https://doi.org/10.1029/2021MS002964

Wang, Y., Braghiere, R. K., Yin, Y., Yao, Y., Hao, D., & Frankenberg, C. (2024). Beyond the visible: Accounting for ultraviolet and far-red radiation in vegetation productivity and surface energy budgets. *Global Change Biology*, *30*(5), e17346. https://doi.org/10.1111/gcb.17346

Wang, Y., Köhler, P., Braghiere, R. K., Longo, M., Doughty, R., Bloom, A. A., & Frankenberg, C. (2022). GriddingMachine, a database and software for earth system modeling at global and regional scales. *Scientific Data*, *9*(1), 258. https://doi.org/10.1038/s41597-022-01346-x

Wang, Y., Köhler, P., He, L., Doughty, R., Braghiere, R. K., Wood, J. D., & Frankenberg, C. (2021). Testing stomatal models at the stand level in deciduous angiosperm and evergreen gymnosperm forests using CliMA Land (v0.1). *Geoscientific Model Development*, *14*(11), 6741–6763. https://doi.org/10.5194/GMD-14-6741-2021

Watanabe, K., & Flury, M. (2008). Capillary bundle model of hydraulic conductivity for frozen soil. *Water Resources Research*, *44*(12), W12402. https://doi.org/10.1029/2008wr007012

Watanabe, K., Kito, T., Wake, T., & Sakai, M. (2011). Freezing experiments on unsaturated sand, loam and silt loam. *Annals of Glaciology*, *52*(58), 37–43. https://doi.org/10.3189/172756411797252220

Woodward, C. S., & Dawson, C. N. (2000). Analysis of expanded mixed finite element methods for a nonlinear parabolic equation modeling flow into variably saturated porous media. *SIAM Journal on Numerical Analysis*, *37*(3), 701–724. https://doi.org/10.1137/s0036142996311040

Wu, T., & Wu, G. (2004). An empirical formula to compute snow cover fraction in GCMs. *Advances in Atmospheric Sciences*, *21*(4), 529–535. https://doi.org/10.1007/bf02915720

Xu, L., & Baldocchi, D. D. (2004). Seasonal variation in carbon dioxide exchange over a mediterranean annual grassland in California. *Agricultural and Forest Meteorology*, *123*(1–2), 79–96. https://doi.org/10.1016/j.agrformet.2003.10.004

Yatunin, D., Byrne, S., Kawczynski, C., Kandala, S., Bozzola, G., Sridhar, A., et al. (2025). The CliMA atmosphere dynamical core: Concepts, numerics, and scaling. *Journal of Advances in Modeling Earth Systems*. https://doi.org/10.22541/essoar.173940262.23304403/v1

You, J., Tarboton, D. G., & Luce, C. H. (2014). Modeling the snow surface temperature with a one-layer energy balance snowmelt model. *Hydrology and Earth System Sciences*, *18*(12), 5061–5076. https://doi.org/10.5194/hess-18-5061-2014

Zaehle, S., Sitch, S., Smith, B., & Hatterman, F. (2005). Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics. *Global Biogeochemical Cycles*, *19*(3). https://doi.org/10.1029/2004GB002395

Zarakas, C. M., Kennedy, D., Dagon, K., Lawrence, D. M., Liu, A., Bonan, G., et al. (2024). Land processes can substantially impact the mean climate state. *Geophysical Research Letters*, *51*(21), e2024GL108372. https://doi.org/10.1029/2024GL108372

Zettl, J. E. A., Lee Barbour, S., Huang, M., Si, B., & Leskiw, L. A. (2011). Influence of textural layering on field capacity of coarse soils. *Canadian Journal of Soil Science*, *91*(2), 133–147. https://doi.org/10.4141/cjss09117

Zhao, W. L., Gentine, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., et al. (2019). Physics-constrained machine learning of evapotranspiration. *Geophysical Research Letters*, *46*(24), 14496–14507. https://doi.org/10.1029/2019GL085291