Title
JAX-CanVeg: A Differentiable Land Surface Model

Journal
Water Resources Research, 61(3)

ISSN
0043-1397

Authors
Jiang, Peishi
Kidger, Patrick
Bandai, Toshiyuki
et al.

Publication Date
2025-03-01

DOI
10.1029/2024wr038116

AGU ADVANCING EARTH AND SPACE SCIENCES

Water Resources Research

RESEARCH ARTICLE
10.1029/2024WR038116

Special Collection:
Advancing Interpretable AI/ML Methods for Deeper Insights and Mechanistic Understanding in Earth Sciences: Beyond Predictive Capabilities

JAX-CanVeg: A Differentiable Land Surface Model

Peishi Jiang[1], Patrick Kidger[2], Toshiyuki Bandai[3], Dennis Baldocchi[4], Heping Liu[5], Yi Xiao[1], Qianyu Zhang[5], Carlos Tianxin Wang[4], Carl Steefel[3], and Xingyuan Chen[1]

[1]Atmospheric, Climate, and Earth Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA, [2]Cradle Bio, Zurich, Switzerland, [3]Lawrence Berkeley National Laboratory, Earth and Environmental Sciences Area, Berkeley, CA, USA, [4]Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA, USA, [5]Department of Civil and Environmental Engineering, Washington State University, Pullman, WA, USA

Abstract Land surface models consider the exchange of water, energy, and carbon along the soil-canopy-atmosphere continuum, which is challenging to model due to their complex interdependency and associated challenges in representing and parameterizing them. Differentiable modeling provides a new opportunity to capture these complex interactions by seamlessly hybridizing process-based models with deep neural networks (DNNs), benefiting both worlds, that is, the physical interpretation of process-based models and the learning power of DNNs. Here, we developed a differentiable land model, JAX-CanVeg. The new model builds on the legacy CanVeg by incorporating advanced functionalities through JAX in the graphic processing unit support, automatic differentiation, and integration with DNNs. We demonstrated JAX-CanVeg's hybrid modeling capability by applying the model at four flux tower sites with varying aridity. To this end, we developed a hybrid version of the Ball-Berry equation that emulates the water stress impact on stomatal closure to explore the capability of the hybrid model in (a) improving the simulations of latent heat fluxes (*LE*) and net ecosystem exchange (*NEE*), (b) improving the optimization trade-off when learning observations of both *LE* and *NEE*, and (c) benefiting a multi-layer canopy model setup. Our results show that the proposed hybrid model improved the simulations of *LE* and *NEE* at all sites, with an improved optimization trade-off over the process-based model. Additionally, the multi-layer canopy set benefited hybrid modeling at some sites. Anchored in differentiable modeling, our study provides a new avenue for modeling land-atmosphere interactions by leveraging the benefits of both data-driven learning and process-based modeling.

Plain Language Summary Land-atmosphere interactions involve flux exchanges of carbon, water, and energy. They are important terrestrial ecosystem components that are being gradually modified by the warming climate. Despite the progress in the land surface model development, accurately modeling these interactions still remains a challenge owing to multiple complicated processes going from the canopy top to the soil system. In this paper, we developed a new land surface model that uses a novel modeling approach called differentiable programming to seamlessly integrate process-based models and deep neural networks. The new model, JAX-CanVeg, is consistent with the known ecohydrological processes while flexible to be coupled with neural networks to improve the simulations of water and carbon fluxes. We demonstrated the hybrid modeling capability of JAX-CanVeg by coupling the equation to calculate stomatal conductance with a neural network that quantifies the water stress impact through soil moisture observations. As a proof of concept, applying the hybrid JAX-CanVeg in four different ecosystems improves simulations of latent heat fluxes and net ecosystem exchanges. The improvement shows promise in using the new model to enhance the simulations of terrestrial water and carbon cycling and better facilitate answering research questions related to climate change.

Key Points:

- We developed a differentiable land surface model using a high-performance machine learning package JAX
- We proposed a hybrid Ball-Berry model that accounts for the influence of water stress on stomatal conductance through a deep neural network
- We showed that the hybrid JAX-CanVeg improved the simulations of water and carbon fluxes at four ecosystems with varying aridity

Correspondence to:
P. Jiang,
peishi.jiang@pnnl.gov

Citation:
Jiang, P., Kidger, P., Bandai, T., Baldocchi, D., Liu, H., Xiao, Y., et al. (2025). JAX-CanVeg: A differentiable land surface model. *Water Resources Research*, *61*, e2024WR038116. https://doi.org/10.1029/2024WR038116

Received 5 JUN 2024
Accepted 7 FEB 2025

Author Contributions:
Data curation: Dennis Baldocchi, Heping Liu, Qianyu Zhang, Carlos Tianxin Wang
Formal analysis: Toshiyuki Bandai, Dennis Baldocchi, Yi Xiao, Carl Steefel, Xingyuan Chen
Funding acquisition: Carl Steefel, Xingyuan Chen
Methodology: Toshiyuki Bandai, Yi Xiao, Xingyuan Chen

1. Introduction

Terrestrial ecosystems are dominant regulators of the Earth's climate (Bonan, 2019). Land surface models (LSMs) are simulators that predict carbon, water, and energy cycling from the canopy top to belowground in response to vegetation conditions and meteorology (Bonan, 2019; Fisher & Koven, 2020; Monson & Baldocchi, 2014). Over the past several decades, LSMs have evolved from representing canopies with a big-leaf energy balance approach to multiple-layer canopies that couples water, carbon, and energy fluxes (Drewry et al., 2010; Lawrence et al., 2019; Sellers et al., 1996; Wiltshire et al., 2020; Yokohata et al., 2020). These models have been used to study carbon cycling (Arora et al., 2020; Cox et al., 2000; Fatichi et al., 2019), vegetation's acclimation response to changing climate (Kumarathunge et al., 2019; Mengoli et al., 2022), and surface energy balance (Y. Chen et al., 2016; Laguë et al., 2019).

Project administration: Carl Steefel, Xingyuan Chen
Software: Patrick Kidger, Toshiyuki Bandai, Dennis Baldocchi
Visualization: Dennis Baldocchi, Yi Xiao
Writing – review & editing: Patrick Kidger, Toshiyuki Bandai, Dennis Baldocchi, Heping Liu, Yi Xiao, Qianyu Zhang, Carl Steefel, Xingyuan Chen

LSMs are subject to modeling errors due to the uncertainties in both model parameters and model mechanisms (Medlyn et al., 2015; Tang & Zhuang, 2008). Parameter estimation is usually the first step to constrain the models prior to model simulations (Kemp et al., 2014; Koven et al., 2020; Mäkelä et al., 2019; Post et al., 2017). Model structural biases can lead to systematic prediction errors, even after calibration, stemming primarily from assumptions or simplifications to represent the complex land-atmosphere interactions that are difficult to simulate explicitly (De Kauwe et al., 2014; Walker et al., 2018; Zaehle et al., 2014). For example, the widely adopted Ball-Berry equation (Ball et al., 1987; Collatz et al., 1991) of stomatal conductance calculation assumes that the leaf intercellular air space is saturated with water vapor, with the saturated vapor pressure directly inferred from leaf temperature. The assumption greatly deteriorates the performance of LSMs in arid or semi-arid ecosystems as the Ball-Berry equation that takes the saturated vapor pressure of leaf as input does not account for the regulation of plant water deficit on stomatal conductance (Leuning, 1995). Many methods have been proposed to address the issue (Bonan et al., 2014; Egea et al., 2011; Katul et al., 2009; Medlyn et al., 2011; Verhoef & Egea, 2014; Y.-P. Wang & Leuning, 1998), through either modifying the Ball-Berry equation or explicitly modeling plant hydraulics. Yet, learning the stomatal responses to water stress remains challenging due to the varying ecosystem conditions and the difficulty of measuring canopy water states for constraining the model (Grossiord et al., 2020).

Furthermore, it is challenging to characterize leaf-level properties using ecosystem-level observations. Leaf physiological properties can be best characterized by photosynthetic gas exchange measurements (Busch et al., 2024), yet obtaining these leaf-level observations is labor-intensive. Meanwhile, numerous flux tower sites provide continuous eddy covariance measurements of ecohydrological fluxes at the ecosystem level for multiple years (Baldocchi et al., 2001) and have been widely used to estimate the parameters of LSMs (Chaney et al., 2016; Raoult et al., 2023; Y. P. Wang et al., 2007). Nevertheless, inversion against these ecosystem-level observations oftentimes assumes static parameters that do not capture the potential temporal dynamics in response to the ambient conditions, such as the variability of the parameters of the Ball-Berry model due to plant water stress (Miner & Bauerle, 2017). It, therefore, remains a challenge to use ecosystem-level eddy-covariance measurements to derive the temporal-dependent leaf physiological properties.

Recently, the emergence of differentiable modeling (Innes et al., 2019; Shen et al., 2023) has shown promise in improving the representation of physical models by seamlessly hybridizing process-based models with deep neural network (DNN)–a popular and powerful data-driven tool (Goodfellow et al., 2016). Differentiable modeling has the benefits of both worlds, that is, the physical interpretation of process-based models and the learning power of DNNs. The essential technique that makes differentiable modeling successful is automatic differentiation, which calculates a derivative by using a chain rule to interleave the computation into elementary operations of functions. This makes automatic differentiation efficient and accurate to working precision and distinguishes it from numerical differentiation and symbolic differentiation. Rooted in automatic differentiation, differentiable modeling thus allows optimizing a numerical model (either a process-based model, a DNN, or a hybrid model) via gradient-based methods through backpropagation (Baydin et al., 2018). It is, therefore, a natural choice to hybridize the process-based model with a learnable DNN that either efficiently emulates a computationally expensive subprocess or dynamically estimates model parameters subjective to model forcings or states. Furthermore, the support of graphic processing units (GPU) in many popular deep learning frameworks makes differentiable modeling even more appealing. So far, differentiable models have recently proven useful in speeding up simulation and improving modeling simulations across many scientific domains, such as computational fluid dynamics (Bezgin et al., 2023; Kochkov et al., 2021), molecular modeling (Greener & Jones, 2021), biological modeling (AlQuraishi & Sorger, 2021), etc.

In earth sciences, differentiable modeling has recently gained significant attention (Gelbrecht et al., 2023; Shen et al., 2023) and has been applied in various earth system processes. In ocean science, Häfner et al. (2018) developed a differentiable ocean model, "Veros", using JAX (Bradbury et al., 2018) to enable fast simulation of global ocean wave velocities. In biogeoscience, Aboelyazeed et al. (2023) leveraged differentiable modeling to hybridize a photosynthesis model with neural networks to learn the model parameters. In surface hydrology, various usages of differentiable modeling have been explored by a research group at the Pennsylvania State University. For the first time, Tsai et al. (2021) proposed a differentiable parameter learning framework that accurately and efficiently maps inputs to spatially varying parameters of a process-based model, using a differentiable version of the model as a physical constraint in the training of the map. Later, this idea was extended to other diverse hydrological modeling applications, including graph network-based river routing (Bindas et al., 2024) and bucket hydrological modeling (Song et al., 2023). These successes underscore the potential of

leveraging differentiable modeling, through a hybrid fashion, to further improve the water and carbon simulation of LSM which are subjective to unknown parameters and physical processes.

Therefore, this study aims to explore and demonstrate the capability of differentiable hybrid modeling in improving land surface water, carbon, and energy fluxes. To this end, we developed a differentiable LSM, JAX-CanVeg, that builds upon a legacy LSM, CanVeg (Baldocchi, 1992, 1994, 1999; Baldocchi & Harley, 1995). As our main technical contribution in this effort, the model was written in JAX and thus gained advanced functionalities of GPU support, automatic differentiation, and, notably, integration with deep neural networks. Through JAX-CanVeg, we developed a hybrid version of the Ball-Berry equation (hereafter referred to as the hybrid model in this paper) that leveraged DNN to emulate the influence of water stress on stomatal closure and asked the following questions:

- To what extent does the proposed hybrid model improve the carbon and water fluxes at ecosystems with varying aridity conditions?
- To what extent does the proposed hybrid model improve the trade-off in a multiobjective optimization task when learning both observed water and carbon fluxes?
- To what extent does a multi-layer canopy setup benefit from the proposed hybrid modeling?

To address these questions, we compared the performance of hybrid JAX-CanVeg with the process-based JAX-CanVeg and a pure DNN model when trained against observations of latent heat fluxes and net ecosystem exchanges. The study was performed at four flux tower study sites with varying aridity in the western United States.

2. Methods

We first provided an overview of CanVeg and detailed the specifics of the development of JAX-CanVeg. Then, we introduced the hybrid version of the Ball-Berry equation that leverages DNN to quantify the influence of plant water deficit on stomatal closure through soil moisture observations. Last, we described the modeling design at four flux tower sites with diverse environmental conditions to assess the performance of hybrid modeling using JAX-CanVeg.

2.1. Developing a Differentiable Land Surface Model (LSM): JAX-CanVeg

2.1.1. CanVeg: A Legacy LSM

CanVeg is a one-dimensional, multi-layer biosphere-atmosphere gas exchange model to calculate water, carbon dioxide (CO_2), and heat fluxes from vegetated canopies written in Matlab (Baldocchi, 2023). Over the past 2 decades, CanVeg has been extensively developed and successfully applied to many ecosystem studies (Baldocchi et al., 2002; Lai et al., 2000; Law et al., 2001; Oikawa et al., 2017; Simon et al., 2005; Yan et al., 2023).

By discretizing both canopy and soil systems into multiple layers, CanVeg couples micrometeorological processes with ecophysiological processes, as shown in Figure 1. The micrometeorological modules solve radiative transfer, boundary layer resistance, leaf energy balance, soil energy balance, and scalar concentration profiles across the canopy. Integrated with these environmental variables, the ecophysiological modules solve stomatal conductance, photosynthetic rate, and leaf dark respiration rate. A detailed description of CanVeg is provided in the Appendix.

At each time step, CanVeg represents the processes through a set of fixed point equations as below:

$$\mathbf{x} = \mathbf{F}(\mathbf{x}; \mathbf{w}), \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^{n_x}$ is a vector of model states with size of n_x; $\mathbf{w} \in \mathbb{R}^{n_w}$ is a vector of model parameters with size of n_w; and $\mathbf{F}: \mathbb{R}^{n_x} \times \mathbb{R}^{n_w} \rightarrow \mathbb{R}^{n_x}$ represents the CanVeg model. Here, the vector $\mathbf{x}$ includes multiple ambient, canopy, and soil variables that are solved by CanVeg, including but not limited to, above- and below-ground temperature profiles, incoming radiations, heat fluxes, photosynthetic rate, stomatal conductance, etc. Note that the model states of CanVeg are calculated independently in time because the model primarily uses analytical solutions to obtain model states (e.g., photosynthesis, leaf temperature, leaf latent heat flux, etc.; see the detailed introduction of CanVeg in the Appendix).

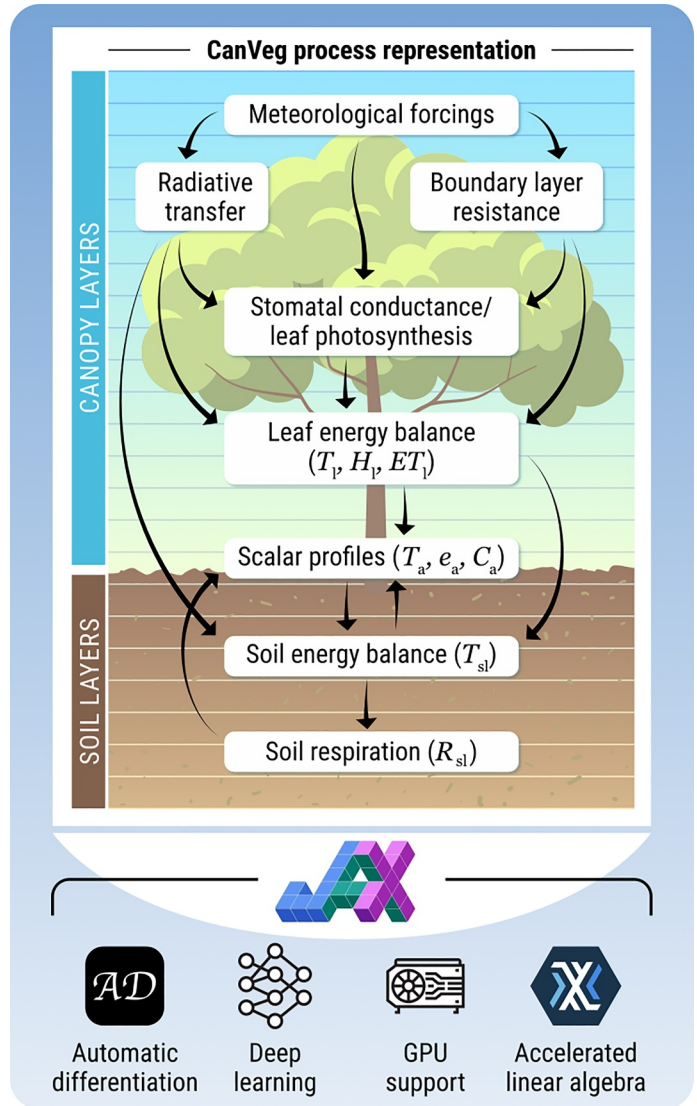

Figure 1. JAX-CanVeg in a nutshell. We developed a differentiable land surface model (LSM) on JAX, named JAX-CanVeg. The new model was built on the legacy CanVeg, which uses meteorological forcings to simulate the water, carbon, and energy fluxes at discretized layers across the canopy and soil layers (see the Appendix for the detailed process representation of CanVeg). Through JAX, the new model incorporates advanced functionalities in the graphic processing unit (GPU) support, high-performance computing through accelerated linear algebra (XLA), automatic differentiation, and, notably, integration with deep neural networks.

The corresponding fixed-point iteration method is used to solve the model equations iteratively until convergence as:

$$\mathbf{x}_{m+1} = \mathbf{F}(\mathbf{x}_m; \mathbf{w}), \tag{2}$$

where m is the iteration index.

2.1.2. JAX-CanVeg: A GPU-Enabled and Differentiable LSM

We have developed JAX-CanVeg by recasting the functionality of CanVeg into JAX (Bradbury et al., 2018), a Python-based scientific machine learning library that has been extensively used in differentiable modeling across a variety of domains (Bezgin et al., 2023; Häfner et al., 2018; Kidger, 2022). Using the capabilities in JAX, the new model now supports automatic differentiation, execution on the GPU, and high-performance computing

Table 1
Comparison Between JAX-CanVeg and the Matlab Version of CanVeg (Baldocchi, 2023)

	Functionality of CanVeg	Automatic differentiation	GPU support	Neural network support/coupling	High performance computing support
JAX-CanVeg	✓	✓	✓	✓	✓(through XLA)
CanVeg	✓	×	×	×	×

Note. (XLA: Accelerated linear algebra).

through accelerated linear algebra (XLA), shown in Figure 1. Most importantly, JAX-CanVeg is a natural platform that affords seamless connection with deep neural networks. An overall comparison between JAX-CanVeg and the Matlab version of CanVeg is given in Table 1. Below, we disclosed a couple of implementation details specific to JAX and the usage of differentiation.

Implementation specifics using JAX. Compared with "traditional" Python programming, JAX imposes additional function design and usage constraints. First, we mostly used pure functions in JAX-CanVeg to enable JAX transformation and the just-in-time compilation supported by XLA. A function is pure when it always returns the same output given the same input and does not allow the usage of a global variable inside the function so that the compiled function is only affected by the input values. Second, we used JAX-specific APIs (abbreviated for Application Programming Interface) for control flows, including `jax.lax.scan` for for-loop, `jax.lax.switch` for switch function, and `jax.lax.cond` for conditional function. Using these APIs allowed a function to be traceable through XLA's just-in-time compilation.

Implicit differentiation. Given the solution $\mathbf{x}$ in Equation 1, the derivative $\frac{\partial \mathbf{x}}{\partial \mathbf{w}}$ is needed in gradient-based optimization. If we apply automatic differentiation naively to calculate $\frac{\partial \mathbf{x}}{\partial \mathbf{w}}$, the iterations can introduce a large computational differentiation graph through the chain rule, thus increasing the memory requirement and computational time. To alleviate this computational burden, we employed the implicit function theorem (Griewank & Walther, 2008) to calculate $\frac{\partial \mathbf{x}}{\partial \mathbf{w}}$.

First, take derivatives with respect to $\mathbf{w}$ at both sides of Equation 1, yielding:

$$\frac{\partial \mathbf{x}}{\partial \mathbf{w}} = \frac{d\mathbf{F}(\mathbf{x}, \mathbf{w})}{d\mathbf{w}}. \tag{3}$$

Given that $\mathbf{x}$ is dependent on $\mathbf{w}$, using the chain rule for a multivariable function then leads to:

$$\frac{\partial \mathbf{x}}{\partial \mathbf{w}} = \frac{\partial \mathbf{F}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{w}} + \frac{\partial \mathbf{F}}{\partial \mathbf{w}}. \tag{4}$$

Next, rearranging the above equation leads to the solution of $\frac{\partial \mathbf{x}}{\partial \mathbf{w}}$:

$$\frac{\partial \mathbf{x}}{\partial \mathbf{w}} = \left[\mathbf{I} - \frac{\partial \mathbf{F}}{\partial \mathbf{x}}\right]^{-1} \frac{\partial \mathbf{F}}{\partial \mathbf{w}}, \tag{5}$$

where $\mathbf{I} \in \mathbb{R}^{n_x \times n_x}$ is an identity matrix.

To facilitate the implementation of automatic differentiation in JAX, we leveraged the Jacobian-Vector Product (JVP), which avoids the explicit construction of a Jacobian matrix and reduces the computational complexity from $O(n_x \times n_w)$ to $O(n_x)$. In a nutshell, the JVP rule simplifies the differentiation process by evaluating the directional derivative that computes the derivative of a function with respect to all inputs by propagating derivatives forward. Here, we calculate $\frac{\partial \mathbf{x}}{\partial \mathbf{w}} \in \mathbb{R}^{n_x \times n_w}$ by multiplying both sides of the equation with a vector $\mathbf{v} \in \mathbb{R}^{n_w \times 1}$. By rearranging Equation 5, we find the following equation:

$$\left[\mathbf{I} - \frac{\partial \mathbf{F}}{\partial \mathbf{x}}\right] \frac{\partial \mathbf{x}}{\partial \mathbf{w}} \mathbf{v} = \frac{\partial \mathbf{F}}{\partial \mathbf{w}} \mathbf{v}. \tag{6}$$

The resulting JVP, $\frac{\partial \mathbf{x}}{\partial \mathbf{w}}\mathbf{v} \in \mathbb{R}^{n_x \times 1}$, was derived through a customized JAX function using `jax.custom_jvp`. This customized function calculated $\frac{\partial \mathbf{x}}{\partial \mathbf{w}}\mathbf{v}$ by solving the linear system Equation 6 as $\mathbf{A}\frac{\partial \mathbf{x}}{\partial \mathbf{w}}\mathbf{v} = \mathbf{b}$ with $\mathbf{A} = \left[\mathbf{I} - \frac{\partial \mathbf{F}}{\partial \mathbf{x}}\right]$ and $\mathbf{b} = \frac{\partial \mathbf{F}}{\partial \mathbf{w}}\mathbf{v}$. Given the defined JVP operation, JAX automatically converts the JVP rule to the corresponding Vector-Jacobian Product rule to enable reverse-mode automatic differentiation. JAX automatically did this neat conversion described in Radul et al. (2023). We verified our implementation by conducting a benchmark test by comparing the implicit function theorem-based automatic differentiation against finite difference-based differentiation using the directional derivative, whose results are shown in Figures A1 and A2 of the Appendix.

Leveraging JAX ecosystem in JAX-CanVeg development. During model development, we leveraged various third-party libraries from the JAX ecosystem. We used `Equinox`–a package enabling object-oriented class-based syntax in JAX (Kidger & Garcia, 2021)–to develop a class-based JAX-CanVeg model. We also used `Optax`–a gradient processing and optimization library for JAX (DeepMind et al., 2020)–to train JAX-CanVeg models through gradient-based optimization. Last, we used `Lineax`–a JAX library for linear solves (Rader et al., 2023)–to solve the linear system of implicit function theorem in Equation 6.

Training and running JAX-CanVeg. Executing the model can be a complicated and daunting process involving many configurations. To alleviate this effort, we suggest training and running the model by reading a JSON-based configuration file provided by users. The file mainly contains four types of configurations:

- *Model configurations*: The basic configurations of the JAX-CanVeg model, such as the time zone and location of the study site, the stomata and leaf angle type of the canopy, the canopy height, the flux tower measurement height, the soil depth, etc;
- *Data configurations*: The file paths of the observed atmospheric forcings and fluxes for both model training and test purposes;
- *Learning configurations*: The model training setup, such as the batch size, the number of training epochs, the loss function, the tunable parameters, the optimizer, etc;
- *Saving configurations*: The file paths where the trained model and the loss values are saved.

A detailed explanation and example codes are available in the `README` file of the open-sourced model repository described in Open Research Section A7.

2.1.3. Hybrid Ball-Berry Equation

Coupled leaf photosynthesis and stomatal conductance module. CanVeg adopts an analytical solution for coupled leaf photosynthesis and stomatal conductance modules (Baldocchi, 1994) at each canopy layer. The photosynthesis process is simulated by the Farquhar model (Farquhar et al., 1980):

$$A = V_c - 0.5V_o - R_d = \min[W_c, W_j]\left(1 - \frac{\Gamma^*}{C_i}\right) - R_d, \tag{7}$$

where A is the photosynthesis rate (μmol m^{-2} s^{-1}); V_c is the carboxylation rate (μmol m^{-2} s^{-1}); V_o is the photorespiration rate (μmol m^{-2} s^{-1}); R_d is the dark respiration rate (μmol m^{-2} s^{-1}); Γ^* is the CO_2 compensation point in the absence of dark respiration (ppm); C_i is the intercellular CO_2 concentration (ppm); and W_c and W_j are the carboxylation rates restricted by Rubisco under saturation of ribulose biphosphate (RuBP) and electron transport for RuBP regeneration, respectively (see Equations A22 and A23 in the Appendix).

Simple conductance relations are used to relate C_i with the surface (C_s) and ambient (C_a) CO_2 concentrations as:

$$C_s = C_a - \frac{A}{g_{b,m}/1.6} \tag{8}$$

$$C_i = C_s - \frac{A}{g_{s,m}/1.6}, \tag{9}$$

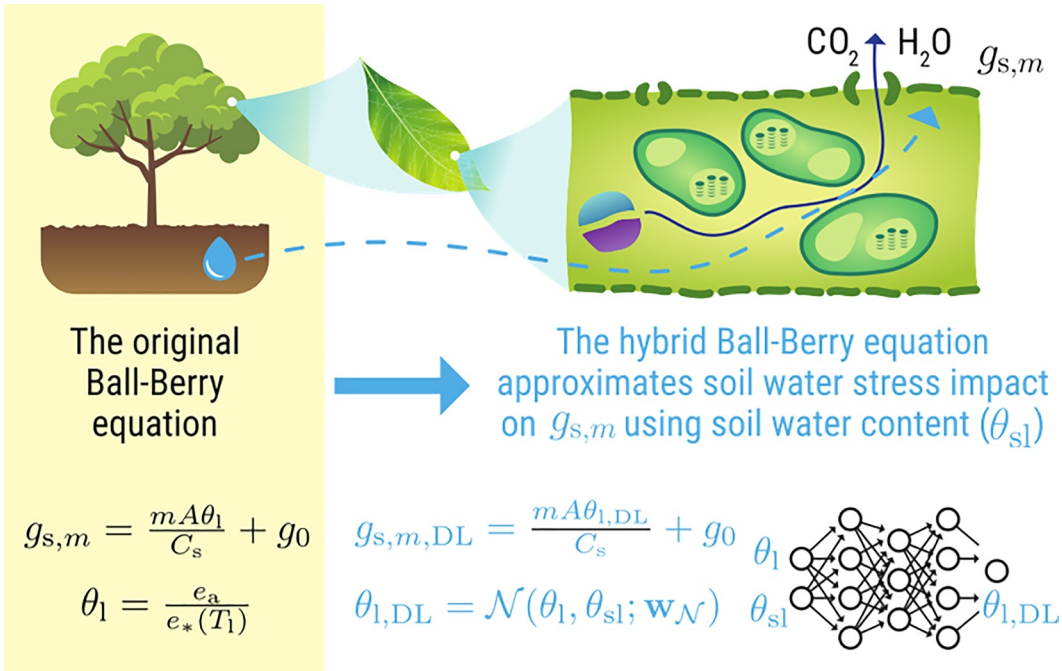

Figure 2. The hybrid Ball-Berry equation in calculating the stomatal conductance $g_{s,m}$. We hybridized the original Ball-Berry equation with a deep neural network (DNN) accounting for the water stress impact on leaf relative humidity $\theta_{l,DL}$ based on observed soil water content θ_{sl}. (C_s: the surface CO_2 concentrations (mol mol^{-1}); m: the dimensionless slope of the Ball-Berry equation; A: the photosynthetic rate (μmol m^{-2} s^{-1}); e_a: the ambient vapor pressure (kPa); $e_*(T_l)$: the saturated vapor pressure (kPa) at the leaf temperature T_l (K); $\mathcal{N}$: a deep neural network; θ_{sl}: the volumetric soil water content (m^3 m^{-3}); and $\mathbf{w_N}$: the learnable parameters of $\mathcal{N}$).

where $g_{b,m}$ and $g_{s,m}$ are the conductances of the leaf laminar boundary layer and the stomata, respectively, in units of mol m^{-2} s^{-1}. C_a is the external input to the photosynthesis model. $g_{b,m}$ is calculated using the flat plate theory (Schuepp, 1993), briefly described in Equation A20 of the Appendix. $g_{s,m}$, that controls the water vapor and CO_2 transfer between leaf and ambient environments, is calculated using the Ball-Berry equation (Ball et al., 1987; Collatz et al., 1991) as (shown in Figure 2):

$$g_{s,m} = \frac{mA\theta_l}{C_s} + g_0, \tag{10}$$

where $g_{s,m}$ is the stomatal conductance (mol m^{-2} s^{-1}); C_s is the surface CO_2 concentration (ppm); A is the photosynthesis rate (μmol m^{-2} s^{-1}); θ_l is the leaf relative humidity; and m and g_0 are the dimensionless slope and zero intercept of the equation. θ_l is calculated as the fractional humidity at the leaf surface:

$$\theta_l = \frac{e_a}{e_*(T_l)}, \tag{11}$$

where e_a is the ambient vapor pressure (kPa); and $e_*(T_l)$ is the saturated vapor pressure (kPa) at the leaf temperature T_l (K). Equations 7–11 leads to a cubic equation of A. We used the resulting analytical solution of A derived in Baldocchi (1994).

Limitation of the vanilla Ball-Berry equation. Equation 11 assumes that the leaf intercellular air space is saturated with water vapor. The assumption limits the response of stomatal conductance to plant and soil water deficit stress (Leuning, 1995). To address this issue, multiple solutions building upon the vanilla Ball-Berry equation have been proposed to account for the water deficit impact (Bonan et al., 2014; Egea et al., 2011; Katul et al., 2009; Medlyn et al., 2011; Verhoef & Egea, 2014; Y.-P. Wang & Leuning, 1998). Ideally, the model needs an explicit representation of the plant hydraulics to simulate its control on stomatal closure, which is currently unavailable in CanVeg.

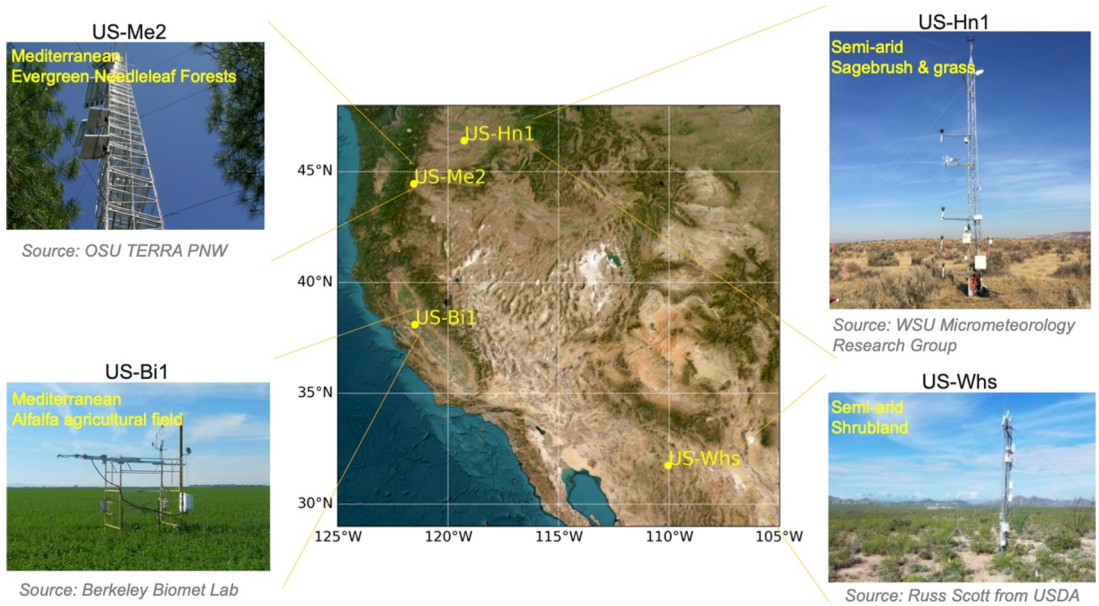

Figure 3. The studied flux tower sites. US-Me2 (upper left): an evergreen needleaf forest ecosystem in Oregon, USA. US-Bi (lower left): an agriculture ecosystem in California, USA. US-Hn1 (upper right): a semi-arid ecosystem in Washington, USA. US-Whs (lower right): a semi-arid shrubland ecosystem in Arizona, USA. (Flux tower photo credits: US-Me2 from the website of Terrestrial Ecosystem Research and Regional Analysis at Oregon State University; US-Bi1 from the Biometeorology Lab at University of California, Berkeley; US-Whs from Russell L Scott at the United States Department of Agriculture; and US-Hn1 from the Micrometeorology Research Group at Washington State University).

Hybrid Ball-Berry equation. In light of parameterizing Equation 10 with soil hydraulic information (Egea et al., 2011; Verhoef & Egea, 2014; Y.-P. Wang & Leuning, 1998), we developed a hybrid Ball-Berry equation that uses observed soil moisture as an approximation of plant water stress to calculate the corresponding stomatal conductance, $g_{s,m,\mathrm{DL}}$, as (Figure 2):

$$g_{s,m,\mathrm{DL}} = \frac{mA\theta_{\mathrm{l,DL}}}{C_s} + g_0, \tag{12}$$

where $\theta_{\mathrm{l,DL}}$ is the leaf relative humidity parameterized using a DNN as:

$$\theta_{\mathrm{l,DL}} = \mathcal{N}(\theta_l, \theta_{\mathrm{sl}}; \mathbf{w_N}), \tag{13}$$

where $\mathcal{N} : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a deep neural network parameterized by $\mathbf{w_N}$ and θ_{sl} is the volumetric soil water content ($m^3\ m^{-3}$). For demonstration purposes, we employed a fully connected neural network for $\mathcal{N}$ with two hidden layers and six neurons in each hidden layer. We used the hyperbolic tangent and sigmoid functions as the nonlinear activations for the hidden and output layers, respectively. Despite the existing methods that incorporate soil hydraulic information by imposing a restricted relational form (Verhoef & Egea, 2014), our hybrid $g_{s,m,\mathrm{DL}}$ that leverages data-driven learning through neural networks is more flexible in capturing the response of stomata closure to water stress.

2.2. Study Sites and Observation Data

Field study sites. We selected four AmerifFlux tower sites in the western United States across various environmental conditions, that is, US-Bi1, US-Me2, US-Whs, and US-Hn1, respectively, shown in Figure 3. US-Bi1 is located at an agricultural land on Bouldin island of the San Joaquin Sacramento Delta region (Hemes et al., 2019; Rey-Sanchez et al., 2022). A perennial alfalfa crop is the major plant at this site. Driven by a Mediterranean climate, this site has an average annual temperature of 16.0°C and an average annual precipitation of 338 mm. US-Me2 is an evergreen needleleaf forest ecosystem situated in Central Oregon (Vickers et al., 2012). This pine tree ecosystem is sustained by a wetter and colder Mediterranean climate with an average annual

Table 2
Characteristics of the Two Flux Tower Sites (MAT, Mean Annual Temperature; MAP, Mean Annual Precipitation; BSk, Cold Semi-Arid Climate; Csa, Hot-Summer Mediterranean Climate; Csb, Warm-Summer Mediterranean Climate)

Site	Location	Climate system	MAT (°C)	MAP (mm)	Flux tower height (m)	Major vegetation	Vegetation height (m)
US-Bi1	CA	Csa	16.0	338	5	Alfalfa	0.8
US-Me2	OR	Csb	6.3	523	34	Ponderosa pine tree	18
US-Whs	AZ	BSk	17.6	320	6.5	Shrub	1
US-Hn1	WA	BSk	12.8	197	5	Sagebrush	1.2

precipitation of 523 mm and an average annual temperature of 6.28°C. Unlike US-Bi1 and US-Me2, the other sites reside in semi-arid dryland ecosystems. US-Whs is a desert shrubland within the Walnut Gulch Experimental Watershed of Arizona (Scott et al., 2015). The site experiences a hotter and drier climate with shrubs as the main canopy coverage. The annual means of temperature and precipitation are 17.6°C and 320 mm, respectively. US-Hn1 is an upland sagebrush-steppe ecosystem located in the Hanford 300 area of Central Washington (Missik et al., 2019). The site is classified as a semiarid climate ecosystem with a mean annual temperature of 12.8°C and receives an average annual precipitation of 197 mm. Table 2 summarizes the key characteristics of the four study sites.

All the sites provide observations of aboveground carbon, water, and energy fluxes and belowground soil temperature and moisture at a resolution of 30-min. We used the following observations to drive JAX-CanVeg: air temperature (T_a), ambient vapor pressure (e_a), air pressure (P_a), ambient CO_2 concentration (C_a), wind speed (WS), friction velocity (u_*), solar radiation (Q), soil temperature at a depth of 15 cm (T_{sl}), and volumetric soil water content at a depth of 15 cm (θ_{sl}). The model simulations were evaluated against observed fluxes, including both latent heat flux (LE) and net ecosystem exchange (NEE). The model also takes the leaf area index (LAI) as the input. The LAI information was obtained from a remote sensing product of the Moderate Resolution Imaging Spectroradiometer (MODIS), MCD15A3H (R. Myneni et al., 2021). We linearly interpolated the 4-day remotely sense LAI to 30-min, consistent with the temporal resolution of flux tower data.

We adopted the observation periods of 01-07-2018 to 30-12-2021 (US-Bi1), 01-07-2016 to 30-06-2020 (US-Me2), 01-02-2012 to 30-06-2015 (US-Whs), and 01-06-2016 to 30-12-2017 (US-Hn1) for this work. At each site, we separated the time into training and test periods shown in Table 3. Figures A3 and A4 plot the time series observations at the four sites.

2.3. Experimental Design

To assess the performance of the hybrid modeling for stomatal conductance calculation in Equation 13, we compared the hybrid JAX-CanVeg using Equation 12 with the process-based JAX-CanVeg using Equation 10 and a pure DNN model at the selected four sites. The observations from the flux towers were used to set up the model. We trained and evaluated the performance of the three models against the observations of both latent heat flux (LE) and net ecosystem exchange (NEE) across a variety of case scenarios by varying canopy layers and the optimization weight assigned to LE and NEE. We designed different model setup and training scenarios to assess the capabilities of the proposed hybrid model in (a) improving the carbon and water fluxes, (b) addressing a muti-optimization task given both carbon and water flux observations, and (c) benefiting a multi-layer canopy model configuration, as described below.

Table 3
Training and Test Periods of Observations at the Four Study Sites

Site	Training time period	Test time period
US-Bi1	07-01-2018 to 30-06-2020	07-01-2020 to 30-12-2021
US-Me2	07-01-2016 to 30-06-2019	07-01-2019 to 30-06-2020
US-Whs	01-02-2012 to 30-06-2014	07-01-2014 to 30-06-2015
US-Hn1	01-06-2016 to 30-06-2017	07-01-2017 to 30-12-2017

2.3.1. Basic Setup of JAX-CanVeg

At each site, we discretized the vertical columns into 50 atmospheric layers from the flux tower to the top of the canopy and 10 soil layers below the ground. We consider both cases of a single big leaf layer (`1L`) and 50 leaf layers (`ML`) extending from the canopy top to the ground surface. Based on field measurements, the canopy heights were approximately 0.8, 18, 1, and 1.2 m for US-Bi1, US-Me2, US-Whs, and US-Hn1, respectively. We used the soil depth of 15 cm where the soil temperature observations were taken. Physiological processes simulated by the Farquhar and the Ball-Berry model

Table 4
Description of the Three Types of Models Used to Predict the Latent Heat Flux (LE) and the Net Ecosystem Exchange (NEE) (DNN, Deep Neural Network)

Model	Description
Process-based model	JAX-CanVeg using the vanilla Ball-Berry equation in Equation 10
Hybrid model	JAX-CanVeg using the hybrid Ball-Berry equation in Equation 12
Pure DNN	A fully connected neural network taking the same inputs of JAX-CanVeg to predict *LE* and *NEE*

(Appendix A3) are parameterized by using the model setup in Baldocchi and Harley (1995); Harley and Baldocchi (1995).

2.3.2. Parameter Learning Design

Table 4 lists the three types of models performed at both sites: the process-based JAX-CanVeg model using the vanilla Ball-Berry equation (Equation 10), the hybrid JAX-CanVeg using its hybrid version (Equation 12), and the pure DNN that takes the same inputs of JAX-CanVeg to predict *LE*. We optimized each model against the observed *LE* and *NEE*. For the process-based model, we estimated the 16 parameters affecting stomatal conductance calculation and sensitive to *LE* and *NEE*, including the slope (m) and intercept (g_0) of the Ball-Berry equation, the leaf's length scale controlling the boundary layer conductance (ls; see Equation A20), the leaf emissivity coefficient (ϵ; see Equations A13 and A35), the leaf quantum yield affecting the potential rate of electron transport (α; see Equation A23), the leaf clumping factor (Ω; see Equation A4), the maximum carboxylation rate by Rubisco ($V_{c,max,25}$) and the maximum rate of electron transport ($J_{max,25}$) at 25°C (used to calculate $V_{c,max}$ in Equation A22 and J_{max} in Equation A24 through the Arrhenius equation), the soil surface emissivity (ϵ_{surf}; see Equation A57), the leaf reflectance and transmittance coefficients of photosynthetic active radiation (PAR) and near-infrared radiation (NIR) (r_{PAR}, r_{NIR}, τ_{PAR}, and τ_{NIR}; see Equations A8 and A9), and the three coefficients of the Q10 power equation for calculating soil respiration (a_r, b_r, and c_r; see Equation A77). The default values of the parameters are listed in Table A1 in the Appendix. For the hybrid model, we estimated the same 16 parameters as the process-based models and additional parameters of the neural network $\mathbf{w_N}$ in Equation 13. The pure DNN model adopted a fully connected neural network that had two hidden layers with six hidden neurons in each layer. Hyperbolic tangent and sigmoid functions were used as the nonlinear activations for the hidden and output layers, respectively. The pure DNN model design is consistent with the neural network $\mathcal{N}$ used in the hybrid JAX-CanVeg (Equation 12) so that the total numbers of tunable parameters of DNN are similar to that of the hybrid model.

We trained each model against both the observed *LE* and *NEE*. We converted this multiobjective optimization problem into minimizing the following weighted loss function:

$$\text{Loss} := \omega \cdot \|\widehat{LE}_o - \widehat{LE}_m\|_2^2 + (1-\omega) \cdot \|\widehat{NEE}_o - \widehat{NEE}_m\|_2^2, \tag{14}$$

where ω is the weighting factor ranging from 0 to 1; $\|\cdot\|_2$ is the $l2$ norm; $\hat{\cdot}$ is the standardization operation that first subtracts the variable by its mean and then divides it by the standard deviation; LE_o and NEE_o are the observed *LE* and *NEE*, respectively; and LE_m and NEE_m is the corresponding model simulations summed across both the canopy and soil systems (see Equations A73 and A75 in the Appendix). Clearly, when $\omega = 1$, the model is trained against the LE_o only. When $\omega = 0$, only NEE_o is used to constrain the model. Each model was optimized over 300 epochs using the Adam algorithm with the default parameter setting in `Optax` (DeepMind et al., 2020). We employed a piecewise constant decay learning scheduler with an initial learning rate of 0.01 and a scaling factor of 0.1 operated at the 50, 100, and 200 epochs. We standardized the inputs of all the deep learning models, including the standalone neural network and the one used by the hybrid model.

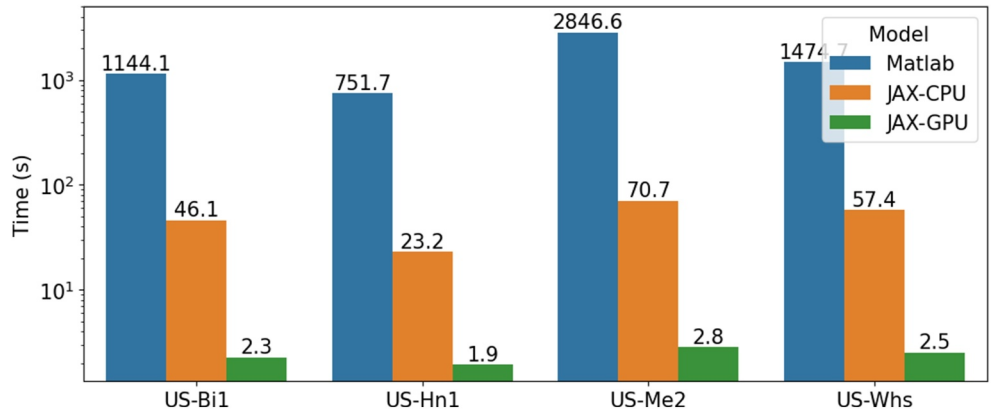

Figure 4. Model simulation time at the four study sites. This barplot shows the simulation time of running process-based CanVeg using the multi-layer canopy setup, that is, the original CanVeg written in Matlab and executed on a central processing unit (CPU), JAX-CanVeg executed on a CPU, and JAX-CanVeg executed on a graphic processing unit (GPU).

2.3.3. Modeling Scenarios

At each site, JAX-CanVeg modeling is subjective to the following three factors:

- whether it is a hybrid model (`Hybrid`) or a process-based model (`PB`);
- whether the model adopts a single layer (`1L`) or 50 layers (`ML`) in the canopy;
- and the value of w used in the loss function Equation 14 (here, we vary ω from 0 to 1 with a discretization interval of 0.1).

This ends up with 44 modeling scenarios of JAX-CanVeg. We denote each scenario with the following format (`Hybrid` or `PB`)-(`ML` or `1L`)-[ω]. For example, a hybrid JAX-CanVeg using multiple canopy layers with $\omega = 0.5$ is denoted as `Hybrid-ML-0.5`. We also train the pure deep learning model with different ω values. These add up to 55 modeling scenarios at each site. Thus, we trained an overall of 220 models for all four sites. The training was performed at four A100 GPUs parallelly, with each accounting for one model.

2.3.4. Evaluation Metrics

All models were trained in the training time periods and evaluated in the test time periods listed in Table 3. We used three metrics to assess the performances of the trained models in simulating *LE* and *NEE*: the Mean Squared Error (*MSE*), the Nash–Sutcliffe model Efficiency coefficient (*NSE*), and the Correlation Coefficient (*CC*), which are given as:

$$MSE_x = \frac{\sum_{i=1}^{N_t}(x_{o,i} - x_{m,i})^2}{N_t}, \tag{15}$$

$$NSE_x = 1 - \frac{\sum_{i=1}^{N_t}(x_{o,i} - x_{m,i})^2}{\sum_{i=1}^{N_t}(x_{o,i} - E[x_o])^2}, \tag{16}$$

$$CC_x = \frac{\sum_{i=1}^{N_t}(x_{o,i} - E[x_o])(x_{m,i} - E[x_m])}{\sqrt{\sum_{i=1}^{N_t}(x_{o,i} - E[x_o])^2}\sqrt{\sum_{i=1}^{N_t}(x_{m,i} - E[x_m])^2}}, \tag{17}$$

where x is the model output to be evaluated, including both *LE* and *NEE*, N_t is the total number of time steps in the test period of a site, and $E[\cdot]$ is the expectation operator. Each metric considers various facets of simulation performance. *MSE* measures the average squared difference between observations and simulations. *NSE* focuses on the modeling capability in capturing the variability of the observations. *CC* assesses the linear relation between the observations and simulations. *NSE* and *CC* are unitless metrics, ranging from `-inf` to 1 and from −1 to 1, respectively.

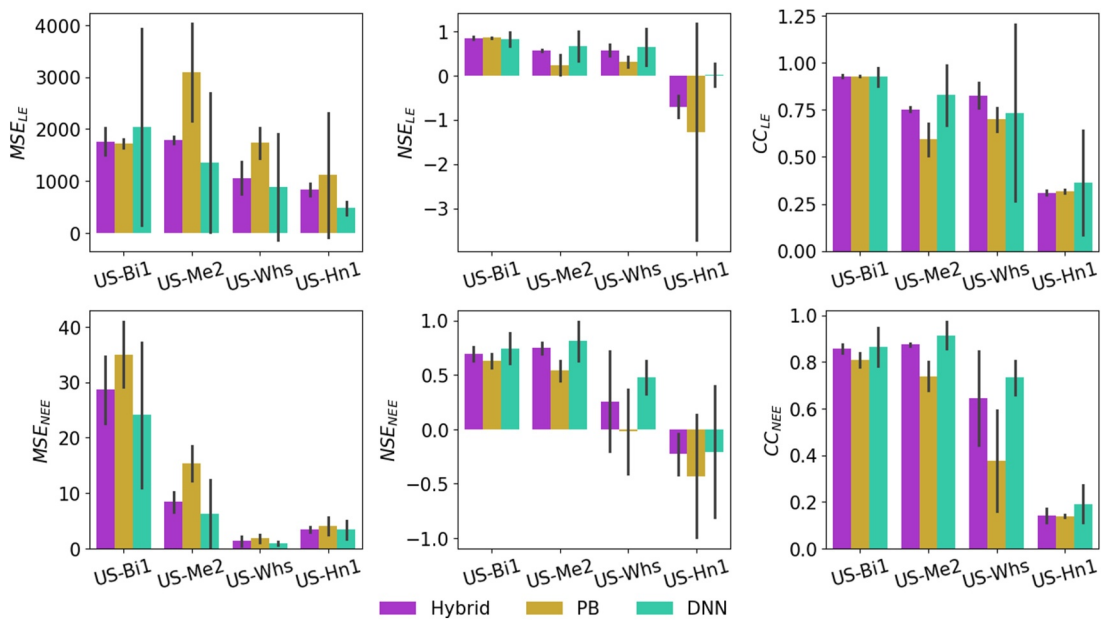

Figure 5. Water and carbon flux simulation performances of the hybrid and process-based JAX-CanVeg and the pure DNN. The barplots show the performance metrics, that is, *MSE*, *NSE*, and *CC* of the process-based JAX-CanVeg (yellow), its hybrid version (violet), and the pure DNN model (cyan) in simulating the latent heat flux (*LE*) and the net ecosystem exchange (*NEE*) for all the four study sites on the test data sets shown in Table 3. (The bar height is the averaged performance across cases using the `ML` setups and all ω values in Equation 14, with the black errorbar representing its standard deviation.

3. Results

3.1. Computational Speedup of JAX-CanVeg Over CanVeg

We compared the computational time of JAX-CanVeg with the original Matlab implementation of CanVeg, using the process-based version with a multi-layer setup (`PB-ML`). The simulations were performed at one A100 GPU and one AMD EPYC 7763 Central Processing Unit (CPU) with JAX version 0.4.19 and Matlab version R2023b. Figure 4 plots the simulation time of running the CanVeg on the CPU, JAX-CanVeg on the CPU, and JAX-CanVeg on the GPU at the four sites. Thanks to the high-performance computing support through XLA, JAX-CanVeg was nearly 20 times faster than CanVeg on the CPU, using around 0.5 to 1-min to complete the simulations compared to approximately 12 to 47-min simulation of CanVeg for both sites. Moreover, running JAX-CanVeg on the GPU achieved even greater computational speedup, taking around 2 s at all sites. Using JAX, the switch between the CPU and the GPU does not modify the existing model and only requires one line of code at the top to configure the computing platform: `import jax; jax.config.update('jax_platform_name', 'gpu')` (or `'cpu'` for the CPU configuration). Meanwhile, running the Matlab function on the GPU is intrusive, necessitating code modifications to specify the data array stored on the computing platform; hence, executing the Matlab version of CanVeg on the GPU is not explored for comparison.

3.2. Capability of Hybrid Modeling in Improving LE/NEE Simulation

Figure 5 shows the barplots of the performance metrics (Equations 15–17) of the process-based JAX-CanVeg (yellow), its hybrid version (violet), and the pure DNN model (cyan) on the test data sets for all the four study sites. The bar height is the averaged metrics across cases using the `ML` canopy setup and all ω values, with the black errorbar representing the standard deviation of the metrics.

For *LE*, the `Hybrid` models outperformed the `PB` models at most sites with varying degrees. Particularly for US-Me2 and US-Whs, the improvement was the greatest, with a significant reduction of MSE_{LE} from around 3,000 to less than 2,000 and from around 2,000 to about 1,000, respectively. The corresponding NSE_{LE} at the two sites also shows significant improvement, increasing up to higher than 0.5, suggesting the enhanced capability of the hybrid modeling in capturing the high values of *LE*. The correlations also saw descent increases. On the other hand, the improvements at the other two sites were limited. For US-Hn1, we observed the decrease of MSE_{LE} and the

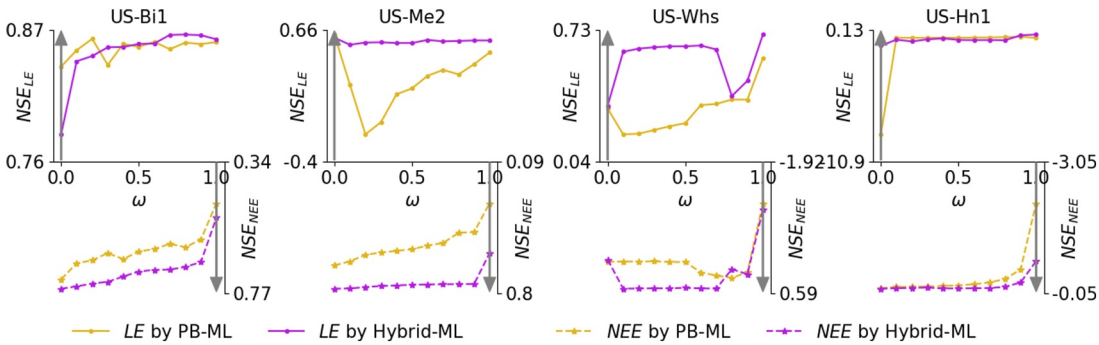

Figure 6. Learning trade-off between *LE* and *NEE*. These plots show the change of *NSE* in Equation 16 over ω used in a multiobjective optimization setting in Equation 14 at the four sites. In each subplot, NSE_{LE} and NSE_{NEE} are plotted in the upper and bottom parts of the frame with yellow and violet lines representing the results of `PB-ML` and `Hybrid-ML` models, respectively. The gray vertical arrow indicates the trend of an improved performance of *NSE*.

increase of NSE_{LE} while CC_{LE} remained almost unchanged. For US-Bi1, the `Hybrid` models is almost no different from the `PB` models, which already demonstrated robust modeling performance with NSE_{LE} and CC_{LE} the highest and closest to 1 among the four sites. The well-calibrated process-based model partially explains the little improvements using the `Hybrid` models.

For *NEE*, we observed the improvements from the `Hybrid` models across all sites. The upgrades were the greatest at US-Me2 and US-Whs, consistent with the results of *LE*. Notably, the `Hybrid` models of US-Me2 not only outperformed the corresponding `PB` models but also induced the highest NSE_{NEE} and CC_{NEE} across all sites. Meanwhile, the improvement of *NEE* at US-Hn1 was still marginal, partially due to the smallest training and test periods of all sites (Table 3) that likely limited the training performance of the model. Compared with the result of *LE*, the major difference was the larger improvement gained at US-Bi1, with enhanced increase and decrease of MSE_{NEE} and NSE_{NEE}.

Comparing JAX-CanVeg modeling (either `Hybrid` and `PB` models) with pure DNN, we found that the pure DNN generally beat the JAX-CanVeg with the mean of its three metrics mostly better than that of JAX-CanVeg at all sites. While it suggests that pure DNN might serve as a benchmark to hybrid modeling, pure DNN had the most tremendous uncertainty, as shown by the largest error bars for metrics of both *LE* and *NEE* at all four sites. These uncertainties resulted from using different weighting factors ω in Equation 14. Compared to the pure DNN, the reduced uncertainties of the JAX-CanVeg trained on various ω illustrates the importance of the physical representation in constraining the modeling.

3.3. Capability of Hybrid Modeling in Improving the Trade-Off in Multiobjective Optimization

A natural follow-up question is whether hybrid modeling can improve the trade-off of training JAX-CanVeg against observations of both *LE* and *NEE* whose interplay is determined by ω in Equation 14. To this end, we plot the change of both NSE_{LE} and NSE_{NEE} over ω in Figure 6. For illustration, we focus on the *ML* setup and *NSE* (see the corresponding plots of *MSE* and *CC* in Figures A5 and A6 of the Appendix). Each subplot of Figure 6 shows the result at each site, with NSE_{LE} and NSE_{NEE} displayed in the up and bottom parts of the plot panel. Results of `Hybrid-ML` and `PB-ML` are plotted in violet and yellow, respectively.

In general, we observed the increase and decrease of NSE_{LE} and NSE_{NEE}, respectively, over ω. There were a few exceptions that did not strictly follow the trends, including NSE_{LE} of `PB-ML`-0.3 at US-Me2 and NSE_{LE} of `Hybrid-ML`-0.8 at US-Whs both of which showed sharp decrease of NSE_{LE}. They were likely due to the complexity of the model with a `ML` setup that increased the loss during the training (not shown here). Despite these cases, the overall increase and decrease of NSE_{LE} and NSE_{NEE} were consistent with the weighting factor ω change.

Comparing `PB-ML` with `Hybrid-ML` showed that the hybrid modeling improved the trade-off of learning two observables by pushing the "frontiers" of the convergence curves. For US-Bi1, though the `PB` and `Hybrid` models yielded similar NSE_{LE} change over ω, NSE_{NEE} were consistently improved by `Hybrid` across all ω. Moreover, for US-Me2 and US-Whs, both NSE_{LE} and NSE_{NEE} got increased for most ω, partially evidenced by

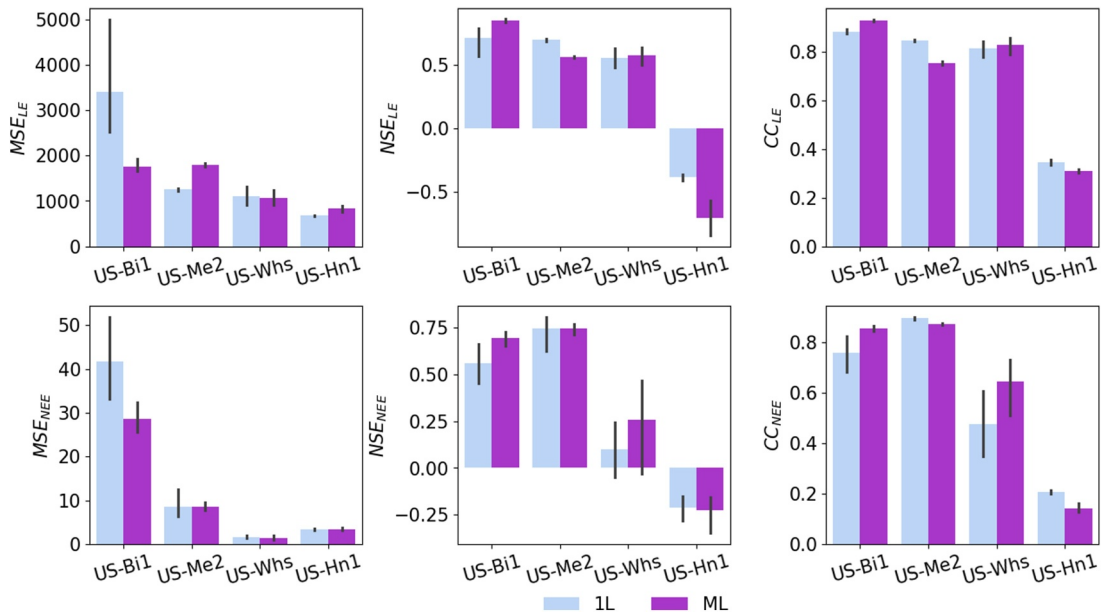

Figure 7. Multi-layer (`ML`) canopy modeling versus single-layer (`1L`) canopy modeling. The barplots show the performance metrics, that is, *MSE*, *NSE*, and *CC* of the `Hybrid-1L` (light blue) and `Hybrid-ML` (violet) in simulating the latent heat flux (*LE*) and the net ecosystem exchange (*NEE*) for all the four study sites on the test data sets shown in Table 3. (The bar height is the averaged performance across cases using different ω values in Equation 14, with the black errorbar representing its standard deviation.)

their best performance gain in Figure 5. US-Hn1 exhibited the least increase of NSE_{LE} and NSE_{NEE} using `Hybrid-ML`, corresponding to its worst modeling performance among all sites (Figure 5).

3.4. Comparing Modeling Performances Using Multi-Layer and Single-Layer Setups

To address our last question on the performance of the proposed hybrid modeling in `ML` modeling, we compared the training results of `Hybrid-ML` with that of `Hybrid-1L`. Figure 7 shows the barplots of the three performance metrics of `Hybrid-1L` (light blue) and `Hybrid-ML` (violet) at the four sites, evaluated against the test data set.

Overall, we found that `ML` improved the modeling of *LE* and *NEE* at some sites but not all. At US-Bi1, `ML` significantly improved the simulation of both observables, with all three metrics better than that of `1L`. At US-Whs, the improvements using `ML` were mostly limited, except for a drastic increase of NSE_{NEE} over that of `1L`. Nevertheless, for US-Me2 and US-Hn1, using `ML` made the performances even worse than `1L`. In sum, `ML` did not necessarily improve the performances of the hybrid modeling.

4. Discussion

4.1. Improved Water and Carbon Responses Captured by the Hybrid Ball-Berry Equation

To illustrate what was learned by the hybrid modeling, we plot in Figure 8 the averaged diurnal signals of *LE*, *NEE*, $g_{s,can}$, A_{can}, and $T_{l,can}$ computed by `PB-ML-0.5` and `Hybrid-ML-0.5` (refer Section 2.3.3 for the definition), averaged across the test time period. $g_{s,can}$, A_{can}, and $T_{l,can}$ are the overall stomatal conductance, photosynthesis, and leaf temperature across the canopy for both shaded and sunlit leaves calculated by Equations A49, A50, and A53 in the Appendix, respectively. We chose $\omega = 0.5$ for illustration because the performances of most `Hybrid` models converged (Figure 6). The estimated parameters of `PB-ML-0.5` and `Hybrid-ML-0.5` were given in Table A1 of the Appendix.

Figure 8 shows that *NEE* simulations got the largest improvement using `Hybrid-ML-0.5` at US-Whs, followed by US-Me2, US-Bi1, and US-Hn1. The *NEE* improvements mostly occurred during the day, where the solar radiations were the largest, and greatly benefited from the corresponding increased A_{can} solved by an analytical solution derived from a coupled the Farquhar model and the Ball-Berry module (Baldocchi, 1994). This was evidenced by the largest increase of A_{can} at US-Whs, followed by US-Me2, US-Bi1, and US-Hn1, corresponding

10.1029/2024WR038116

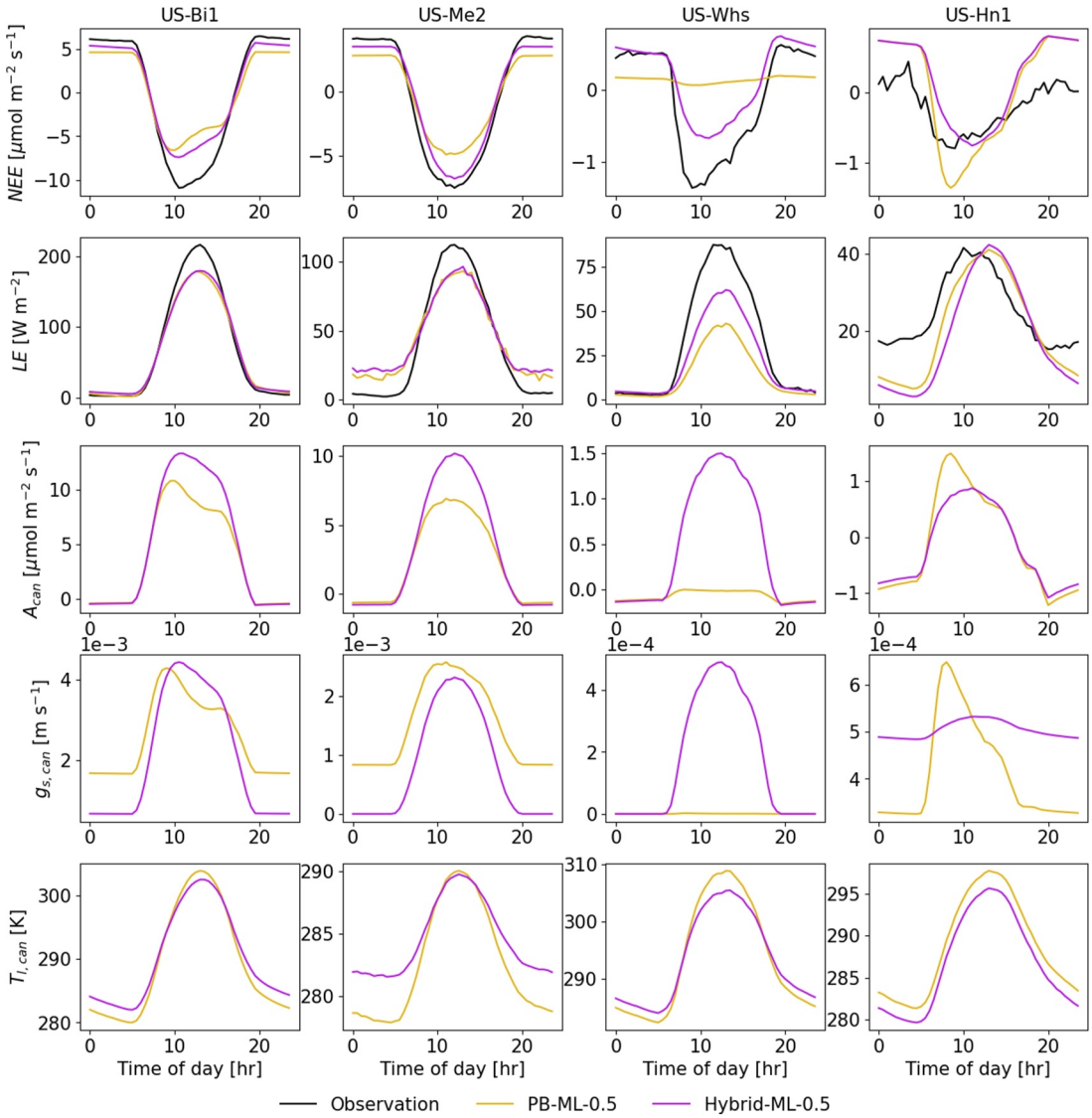

Figure 8. Averaged diurnal signal of JAX-CanVeg simulations. The simulations of `Hybrid-ML`-0.5 (violet) and `PB-ML`-0.5 (yellow) were averaged to 24-hr on the test period at the four sites. Five model outputs are plotted: the net ecosystem exchange (*NEE*), the latent heat fluxes (*LE*), the overall canopy photosynthesis (A_{can}), the overall canopy stomatal conductance ($g_{s,can}$), and the overall canopy leaf temperature ($T_{l,can}$).

to the improvement degree of *NEE* simulation. At the same time, only US-Whs exhibited decent *LE* improvement by using `Hybrid-ML`-0.5 such that the `Hybrid` model better captured the peak of *LE* during the day. This was supported by the larger $g_{s,can}$ over that of `PB-ML`-0.5, suggesting that the `Hybrid` model increased the stomata openness during the day. The other three sites showed different responses of stomata closure when using the hybrid equation. The $g_{s,can}$ of `Hybrid-ML`-0.5 at US-Bi1 and US-Me2 had larger temporal variations from day to night than `PB-ML`-0.5 while the corresponding leaf temperature $T_{l,can}$ showed reduced variations. The enlarged and lowered variations of $g_{s,can}$ and $T_{l,can}$ balanced out each other through Fick's law (Equation A33 of the Appendix), leading to similar simulations of *LE* Overall, such interplay among $g_{s,can}$, A_{can}, and $T_{l,can}$, affected by the introduction of the hybrid Ball-Berry equation, improved the overall simulations of both *NEE* and *LE* (mostly *NEE*) though additional observations of canopy states (e.g., $g_{s,can}$ and $T_{l,can}$) would be helpful to verify the change and further constrain the model.

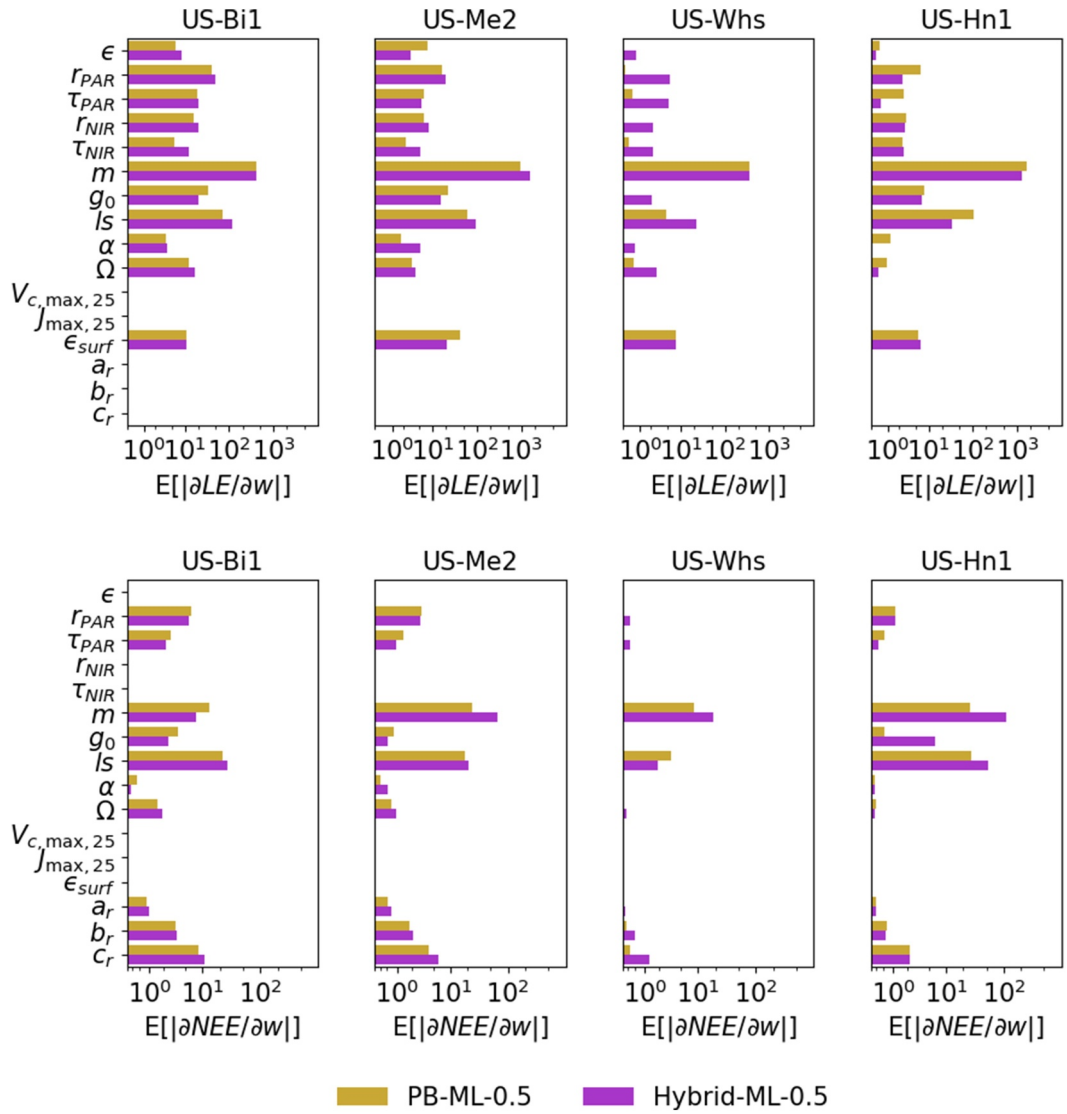

Figure 9. Parameter sensitivity of JAX-CanVeg. The temporally averaged absolute values of the parameter sensitivity with regard to *LE* (top) and *NEE* (bottom), denoted as $E\left[\left|\frac{\partial LE}{\partial w}\right|\right]$ and $E\left[\left|\frac{\partial NEE}{\partial w}\right|\right]$, respectively (where $E[\cdot]$ is the expectation operation and $w \in \mathbf{w}$). The partial derivative was calculated on the test periods of all four sites, using the process-based (yellow) and hybrid (violet) models with a multi-layer (ML) canopy setup and $\omega = 0.5$ of the loss in Equation 14.

4.2. Impact of the Hybrid Model on Parameter Sensitivity

We compared the parameter sensitivity between PB-ML-0.5 and Hybrid-ML-0.5. The barplots in Figure 9 are the averaged absolute values of the partial derivative of *LE* and *NEE*, across the test periods, over each parameter listed in Table A1.

For *LE*, the most sensitive parameter was the slope of the Ball-Berry equation m, followed by other conductance parameters (i.e., g_0 and ls) and the leaf and ground surface reflectance/emissivity/transmittance that control the canopy radiation profile (i.e., ϵ, ϵ_{surf}, r_{PAR}, τ_{PAR}, r_{NIR}, and τ_{NIR}). US-Whs saw drastic increases of sensitivities of many parameters across both conductance parameters and canopy reflectance, emissivity, and transmittance (e.g., g_0, ls, ϵ, r_{PAR}, τ_{PAR}, r_{NIR}, and τ_{NIR}). These changes reflected the improved *LE* simulation of Hybrid-ML-0.5 at this site (Figures 5 and 6). At the other three sites, where the hybrid modeling slightly improved *LE*, PB-ML-0.5 and Hybrid-ML-0.5 provided similar sensitivity result with a modest difference, for example, the increase of the averaged sensitivity for the leaf reflectance, emissivity, and transmittance at US-Bi1.

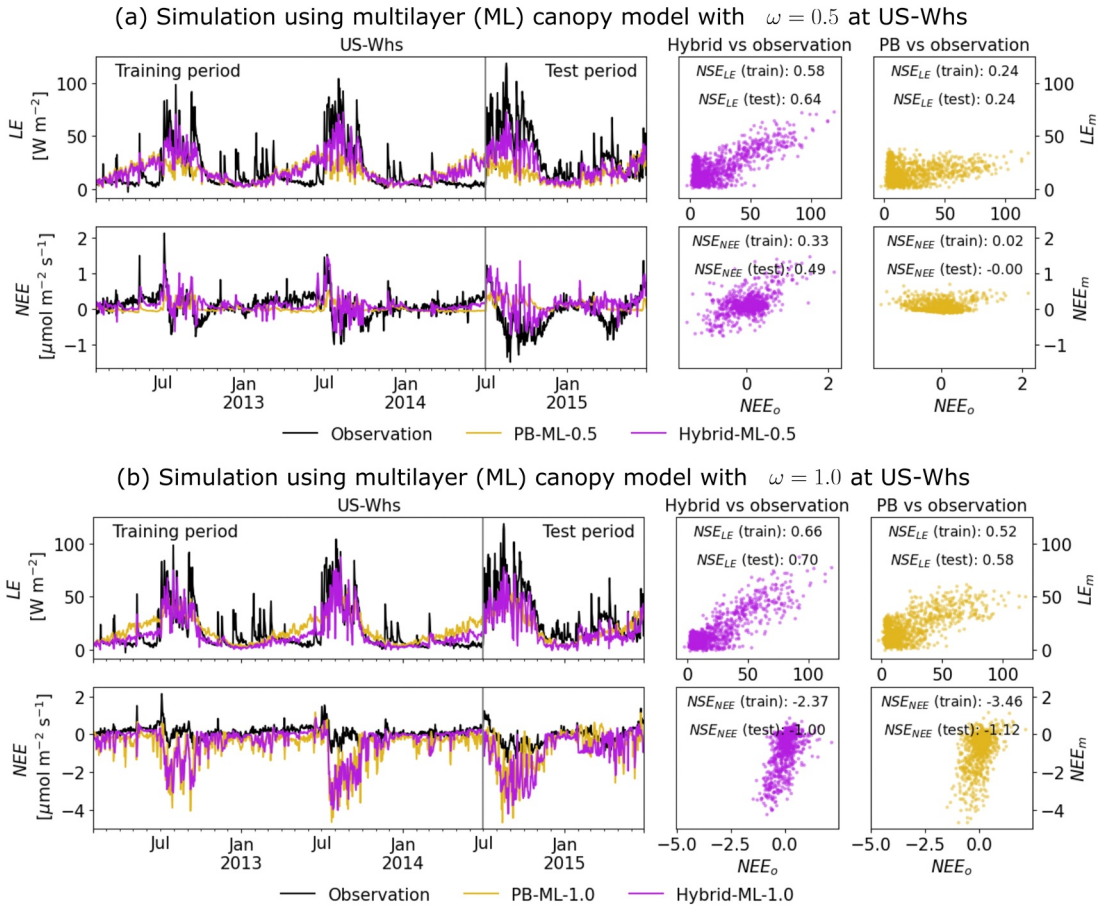

Figure 10. JAX-CanVeg simulation using the multi-layer canopy setup at US-Whs. (a, b) plot the daily averaged simulations of *LE* and *NEE* with both the hybrid (violet) and process-based (yellow) models, using $\omega = 0.5$ and $\omega = 1.0$ in the loss function (Equation 14), respectively.

For *NEE*, the most sensitive parameters switched between *m* and *ls* among the four sites, suggesting the significance of both stomatal conductance and leaf boundary layer conductance in determining *NEE*. We also observed the importance of soil respiration (R_{sl}), as a major component of *NEE* in Equation A75, by the large sensitivities of a_r, b_r, and c_r which are the parameters of the Q10-power model calculating R_{sl} in Equation A77. Among the sites, US-Whs exhibited reduced *ls* and big increases of *m*, b_r, and c_r. The change illustrates that the improved *NEE* at US-Whs was jointly contributed by the regulations of A_{can} and R_{sl}. At US-Me2, the major difference was the increase of *m*, suggesting that A_{can} was the major contributor to the *NEE* improvement shown in Figure 8. We observed slight differences in parameter sensitivity at US-Bi1, corresponding to its limited *NEE* improvement. For US-Hn1, despite the large sensitivity increases of *m*, g_0, and *ls*, the resulting changes in the simulated *NEE* did not greatly bring it closer to the observed *NEE*.

4.3. The Trade-Off in Model Optimization Between Water and Carbon Fluxes

Although the proposed `Hybrid` model facilitated reducing the trade-off when trained on both *LE* and *NEE*, the trade-off was not alleviated (Figure 6). This was evidenced by the best performance of either *LE* or *NEE* with $\omega = 1.0$ and $\omega = 0.0$, respectively. The trade-off could be attributed to unknown physical processes poorly represented in the model. To further investigate the potential cause of trade-off, we looked into the modeling performance of US-Whs and US-Hn1, both of which are dryland ecosystems, but one with the greatest improvement using the hybrid setup (US-Whs) and the other with the least improvement (US-Hn1).

Figure 10 plots the daily averaged simulations of *LE* and *NEE* at US-Whs of both `Hybrid`- and `PB` models using the `ML` canopy setup, with the results of $\omega = 0.5$ and $\omega = 1.0$ shown in Figures 10a and 10b, respectively. On one hand, `Hybrid-ML`-0.5 yielded the best balance between *LE* and *NEE* with the corresponding *NSE* greatly larger

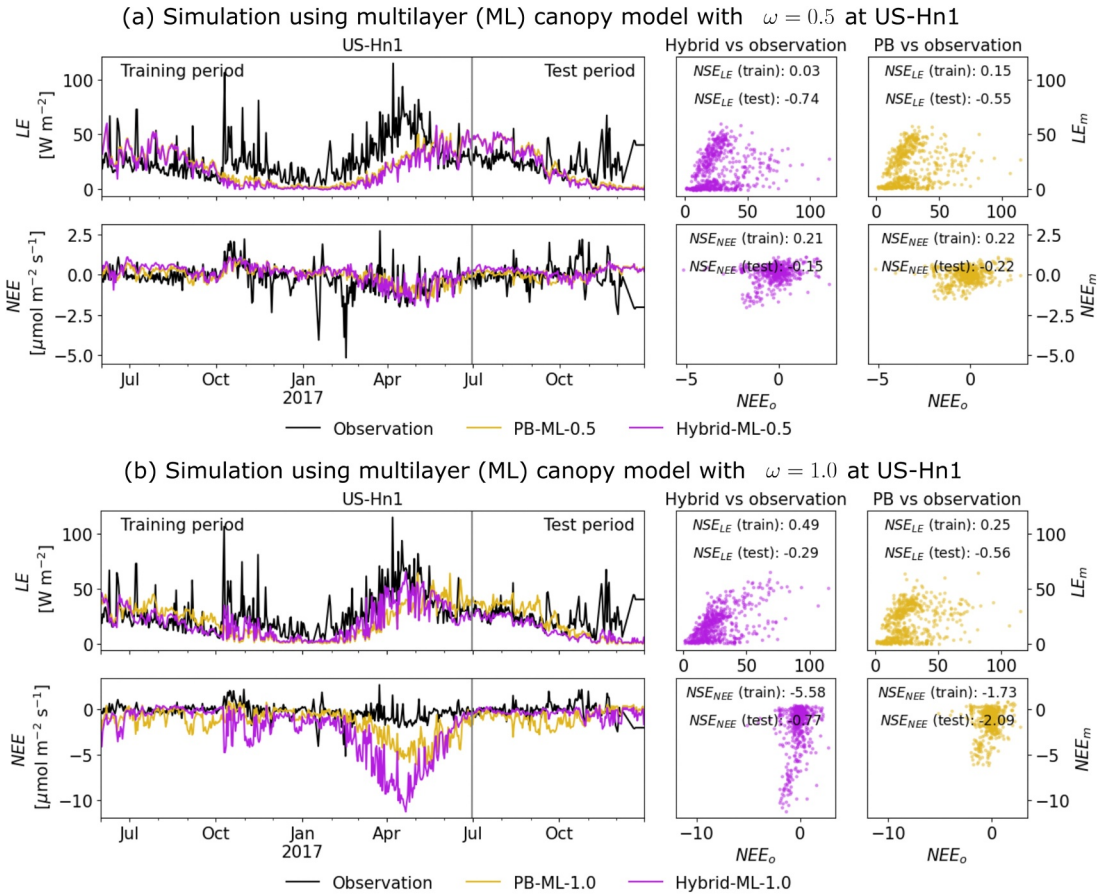

Figure 11. JAX-CanVeg simulation using the multi-layer canopy setup at US-Hn1. (a, b) plot the daily averaged simulations of *LE* and *NEE* with both the hybrid (violet) and process-based (yellow) models, using $\omega = 0.5$ and $\omega = 1.0$ in the loss function (Equation 14), respectively.

than that of PB-ML-0.5 (Figure 10a). However, the model overestimated the low *LE* during spring each year (i.e., March to July). On the other hand, when only trained on *LE*, Hybrid-ML-1.0 generated better *LE* than Hybrid-ML-0.5 with NSE_{LE} increasing from 0.64 to 0.70 on the test period (Figure 10b). Particularly, the low *LE* in springtime was better captured by Hybrid-ML-1.0. However, the improved *LE* came with the cost of the poor performance of *NEE* simulation with $NSE_{NEE} = -1.0$. The large negative values of *NEE* in the summer corresponded to the large *LE*, suggesting large photosynthesis A_{can} was produced during summer. Clearly, when only *LE* observations were available, *NEE* was hard to simulate well due to the unknown contribution of soil respiration (R_{sl}; see Equation A75), which was calculated by a simple Q10-power equation in Equation A77. The simplified representation of R_{sl} in turn affected the interplay between *LE* and *NEE*, leading to the trade-off of the simulations of the two, especially in dryland ecosystems where canopy coverage is limited compared to wetter or forest regions (e.g., US-Bi1 and US-Me2).

The role of soil carbon cycling representation in governing the interplay between *LE* and *NEE* is probably more important at US-Hn1. Figure 11 shows the corresponding plot at US-Hn1 with the top and bottom representing the simulations using $\omega = 0.5$ and $\omega = 1.0$, respectively. When the observations *LE* and *NEE* played equal roles in the training (i.e., $\omega = 0.5$ in Figure 11a), the corresponding simulations barely captured any observed variations using either the Hybrid or the PB model, with all *NSE* lower than 0.25. Nevertheless, when the model was only constrained by the observed *LE* (Figure 11b), the simulation of *LE* was greatly improved using the Hybrid model. Hybrid-ML-1.0 was able to capture the large *LE* trend from March of 2017 to September of 2018. Yet, similar to US-Whs, it came with the cost of large negative *NEE* due to the increased stomata and A_{can}, which were not well balanced by the simulated R_{sl}. The reduced improvement using the hybrid modeling suggests a greater

role of the soil carbon cycling that is not well captured by JAX-CanVeg in *NEE* simulation at US-Hn1 than US-Whs.

However, the performances of hybrid modeling were less affected by the soil carbon cycling dynamics at the other two wetter sites, US-Bi1 and US-Me2. The corresponding simulations plots in Figures A7 and A8 of the Appendix show that the *LE* simulation performances of the two hybrid models with $\omega = 0.5$ and $\omega = 1.0$ were similar at both sites while `Hybrid-ML-0.5` yielded better *NEE* simulations than `Hybrid-ML-1.0`. The result indirectly suggests the oversimplified representation of soil carbon dynamics has limited impacts on *NEE*, which were mostly attributed to the photosynthesis A_{can} at the two sites.

4.4. Need of the Multi-Layer Canopy Setup in Differentiable Hybrid Modeling

Although the `ML` canopy setup did not increase the performance of our `Hybrid` models at all sites (Figure 7), using `ML` canopy setup could potentially improve `Hybrid` models for the following reasons. First, we observed decent improvements using `ML` canopy at US-Bi1 and US-Whs, particularly for the simulation of *NEE*. Second, the slightly reduced performances at US-Me2 and US-Hn1 could be attributed to the same default model parameters at both sites. These parameters, adopted from Harley and Baldocchi (1995) and Baldocchi and Harley (1995), were likely less ideal initials that made the training hard for the `ML` model whose numerical representation is more complicated than the `1L` model. Third, the proposed hybrid Ball-Berry equation in Equation 12 did not explicitly capture the vertical gradient of leaf water potential, which potentially limited the benefits of a `ML` model during model training. Last, running JAX-CanVeg on GPU was fast, with only around 2 s for all sites on multiple years (Figure 4). The GPU speedup significantly alleviates the computational burden of using a `ML` setup. As Bonan et al. (2021) demonstrated the potential of using `ML` model setup to move beyond the widely adopted current big-leaf models, future research is needed to further explore the possibility of using hybrid modeling to enhance water and carbon fluxes in a `ML` canopy setup through better parameter initializations and carefully designed hybrid components capturing the vertical variation across the canopy.

5. Conclusion

We developed a differentiable land surface model, JAX-CanVeg, by recasting a legacy model, CanVeg, into a scientific machine learning programming platform, JAX. We demonstrated the hybrid modeling capability of JAX-CanVeg by proposing a hybrid Ball-Berry equation that integrates the original equation with a DNN encoding the impact of water stress on stomatal closure through observed soil moisture. Through applications at four flux tower sites, we found that the hybrid model consistently improved the simulations of *LE* and *NEE* over the process-based model with varying degrees. When trained against both observed *LE* and *NEE* using different weighting factors ω in the loss function (Equation 14), the hybrid models generally reduced the optimization trade-off with the simulation performance better than the process-based models. Further, we found that not all sites benefited from hybrid modeling in a multi-layer canopy setup–only two sites showed improvement over a single-layer configuration–likely due to the training challenges. A combination of suboptimal initial parameters and the complexity of the multi-layer setup may have made the training process difficult. Potential future research may focus on (a) improving the soil carbon representation, potentially through a hybrid way, to reduce the optimization trade-off further and (b) explore an improved way to train a hybrid multi-layer canopy model. We anticipate that the new differentiable modeling framework provides a new avenue for modeling land-atmospheric interactions by leveraging the benefits of both data-driven learning and process-based modeling (M. Chen et al., 2023).

Appendix A: An Overview of CanVeg

CanVeg is a one-dimensional, multi-layer biosphere-atmosphere gas exchange model to calculate water, carbon dioxide (CO_2), and heat fluxes from vegetated canopies. The model couples micrometeorological processes with ecophysiological processes. The micrometeorological modules solve the radiative transfer across the canopy, leaf and soil energy balances, and scalar concentration profiles. Integrated with these ambient states, the ecophysiological modules solve leaf photosynthesis, stomatal conductance, transpiration, and respiration processes. Below we delve into the equations for radiative transfer, photosynthesis and stomatal conductance, leaf energy balance, and soil energy balance. More details of the model are reported in (Baldocchi, 1992, 1994, 1999; Baldocchi & Harley, 1995).

The overall objective of the model is to calculate the passive scalars (i.e., water, CO_2, and heat) across both the canopy and its ambient surroundings. The diffusion process across leaf surfaces between the internal and ambient scalar concentrations or heat content can be described using Fick's law:

$$S(Con_a(z)) = -a(z)\frac{Con_a(z) - Con_l(z)}{1/g_{b,m}(z) + 1/g_{s,m}(z)}, \tag{A1}$$

where z is the vertical depth; a is the leaf area density ($m^2\ m^{-3}$); Con_a and Con_l are the ambient and leaf internal scalar concentrations or heat content, for example, water vapor pressure e, CO_2 concentration, and temperature T; S is the flux serving a source and sink term; $g_{b,m}$ is the boundary layer conductance to molecular diffusion (mol $m^{-2}\ s^{-1}$); and $g_{s,m}$ is the stomatal conductance in mole (mol $m^{-2}\ s^{-1}$).

A1. Canopy Radiative Transfer

The model divided the above-ground canopy into M layers to better capture the energy, carbon, and water fluxes across the canopy. Delineating the radiation profiles of the canopy is critical to calculate photosynthesis, leaf energy balance, and soil energy balance. Generally, the model takes in the observed global shortwave radiation (Q_{TOT}; W m^{-2}) and separates it into photosynthetic active radiation (PAR, denoted as Q_{PAR}) and near-infrared radiation (NIR, denoted as Q_{NIR}), both of which are furthered decomposed into the direct beam and diffusive components:

$$Q_{TOT} = Q_{PAR} + Q_{NIR} \tag{A2}$$

$$= Q_{PAR,bm} + Q_{PAR,df} + Q_{NIR,bm} + Q_{NIR,df}, \tag{A3}$$

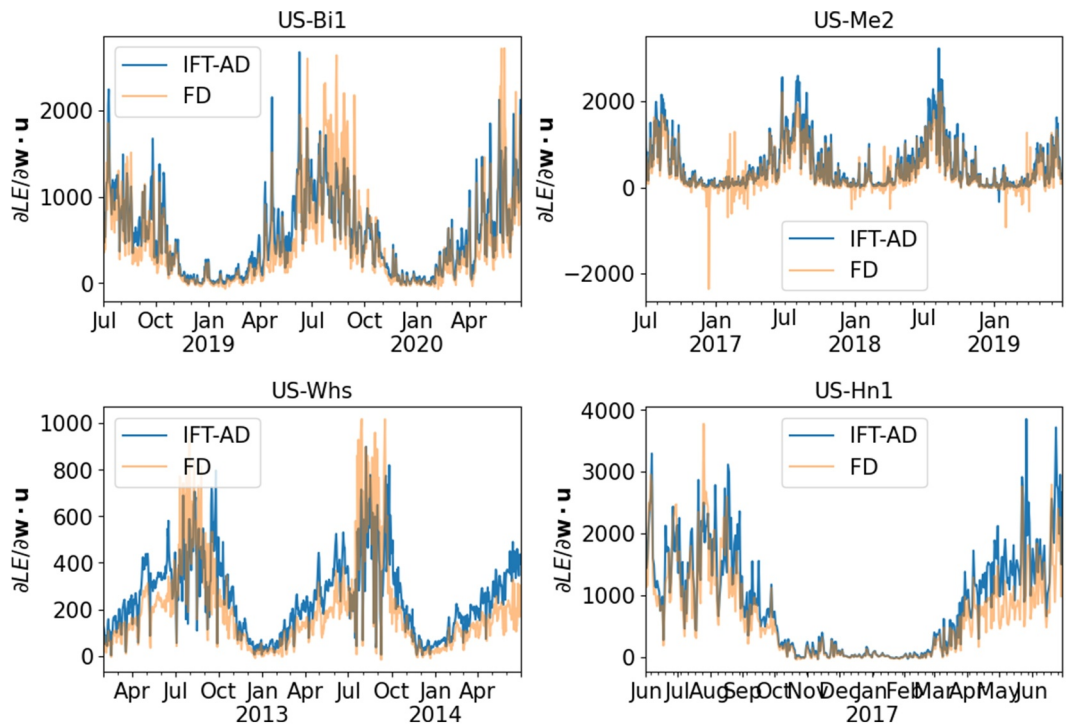

Figure A1. Comparison between implicit function theorem-based automatic differentiation (IFT-AD) and finite difference-based differentiation (FD) using the directional derivative of LE using the process-based JAX-CanVeg on the default parameters $\mathbf{w}$ listed in Table A1. (The directional derivative of IFT-AD is calculated as $\frac{\partial LE}{\partial \mathbf{w}} \cdot \mathbf{u}$ where $\mathbf{u} = [1, \ldots, 1]_{n_w}$ is a unit vector with size n_w; and the directional derivative of FD is calculated as $\frac{F_{LE}(\mathbf{x}, \mathbf{w}+h\mathbf{u}) - F_{LE}(\mathbf{x}, \mathbf{w})}{h}$, where $h = 0.01$ and $F_{LE} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_w} \to \mathbb{R}$ is the function of JAX-CanVeg to calculate LE.

Table A1
The Default and Estimated Parameters of JAX-CanVeg for a Multi-Layer (`ML`) *Setup With the Weighting Factor* $\omega = 0.5$

	Default	US-Bi1 PB	US-Bi1 Hybrid	US-Me2 PB	US-Me2 Hybrid	US-Whs PB	US-Whs Hybrid	US-Hn1 PB	US-Hn1 Hybrid
ϵ	9.8e−01	6.1e−01	4.1e−01	8.2e−02	3.4e−02	8.9e−01	6.7e−01	8.6e−01	9.9e−01
r_{PAR}	5.0e−02	3.3e−03	5.0e−05	1.7e−01	1.6e−01	2.5e−01	1.8e−01	2.4e−01	2.4e−01
τ_{PAR}	5.0e−02	3.9e−01	2.5e−01	2.0e−01	5.0e−05	2.8e−01	1.9e−01	1.4e−03	4.4e−03
r_{NIR}	6.0e−01	3.3e−03	3.2e−01	2.1e−01	6.0e−01	5.2e−01	5.8e−01	6.6e−01	7.5e−01
τ_{NIR}	2.0e−01	5.0e−05	5.0e−05	1.5e−01	2.2e−01	1.5e−01	1.6e−01	2.4e−01	3.5e−01
m	5.0e−02	6.2e−02	2.4e−02	1.5e−02	5.0e−05	5.0e−05	5.0e−05	6.9e−03	1.0e−02
g_0	4.0e−02	8.4e−02	1.0e−01	1.0e−01	1.0e−01	1.0e−01	1.0e−01	1.0e−01	9.3e−02
ls	2.2e−01	1.8e−01	1.7e−01	1.1e−01	1.2e−01	7.7e−03	1.1e−01	3.3e−02	1.7e−02
α	8.2e+00	8.5e+00	8.4e+00	8.2e+00	8.1e+00	8.2e+00	8.0e+00	8.0e+00	7.9e+00
Ω	9.5e−01	1.0e+00	1.0e+00	7.9e−01	8.7e−01	7.4e−01	7.6e−01	7.1e−01	7.5e−01
$V_{c,\max,25}$	1.7e+02	1.7e+02	1.7e+02	1.7e+02	1.7e+02	1.7e+02	1.7e+02	1.7e+02	1.7e+02
$J_{\max,25}$	2.6e+02	2.6e+02	2.6e+02	2.6e+02	2.6e+02	2.6e+02	2.6e+02	2.6e+02	2.6e+02
ϵ_{surf}	9.8e−01	5.6e−01	5.4e−01	2.3e−01	7.4e−01	1.0e+00	8.6e−01	1.0e+00	1.0e+00
a_r	5.0e+00	5.7e+00	5.7e+00	5.2e+00	5.4e+00	4.5e+00	4.9e+00	4.9e+00	5.0e+00
b_r	1.7e+00	1.0e+00	1.5e+00	8.5e−01	1.3e+00	1.2e+00	1.6e+00	1.6e+00	1.7e+00
c_r	8.0e−01	1.1e−01	1.8e−01	4.1e−01	3.3e−01	1.1e+00	8.2e−01	8.5e−01	8.4e−01

Note. (m: The slope of the Ball-Berry equation; g_0: The intercept of the Ball-Berry equation; ls: The leaf's length scale controlling the boundary layer conductance used in Equation A20, ϵ: The leaf emissivity coefficient used in Equations A13 and A35; α: The leaf quantum yield affecting the potential rate of electron transport used in Equation A23); Ω: The leaf clumping factor used in Equation A4; $V_{c,\max,25}$ and $J_{\max,25}$: The maximum carboxylation rate by Rubisco and the maximum rate of electron transport at 25 °C used to calculate $V_{c,\max}$ in Equation A22 and $J_{\max}$ in Equation A24 through the Arrhenius equation, respectively; ϵ_{surf}: The soil surface emissivity used in Equation A57; r_{PAR}, r_{NIR}, τ_{PAR}, and τ_{NIR}: The leaf reflectance and transmittance coefficients of photosynthetic active radiation (PAR) and near-infrared radiation (NIR) used in Equations A8 and A9; and a_r, b_r, and c_r: The three coefficients of the Q10 power equation for calculating soil respiration used in Equation A77.

where bm and df refer to the beam and diffusive components. Each component in Equation A3 is calculated using the method described in Weiss and Norman (1985). The canopy radiative transfer uses different components of the incoming radiation to further calculate the radiation profiles of both sunlit and shaded leaves in each canopy layer.

Probability of light penetration. Given the incoming short radiation on the top of the canopy, the probability of the light radiation penetration is calculated using a Markov model to account for the leaf clumping effect (Myneni et al., 1989):

$$P_i = \exp\left(-\frac{\mathcal{L}_i \mathcal{G} \Omega}{\sin\beta}\right), \tag{A4}$$

where $i \in [1, \ldots, M]$ is the index of the canopy layer with $i = 1$ standing for the lowest layer; $\mathcal{L}_i$ is the cumulative leaf area index from the top of the canopy to the ith layer; Ω is the leaf clumping factor and is within (0, 1); β is the solar elevation angle; and $\mathcal{G}$ is the foliage orientation function depending on the given leaf angle distribution. Here, we uniformly divide a total leaf area index, $\mathcal{L}$, into each canopy layer such that:

$$\Delta\mathcal{L}_i = \frac{\mathcal{L}}{M} \tag{A5}$$

$$\mathcal{L}_i = \mathcal{L} - \sum_{j=1}^{i} \Delta\mathcal{L}_j, \tag{A6}$$

where $\Delta\mathcal{L}_i$ is the leaf area index at the $i$the layer.

Shortwave direct beam radiation. The direct beam radiation of a given waveband that reaches the *i*th canopy layer can be calculated as the product of the incoming beam radiation of the waveband and the probability of beam penetration at that layer:

$$Q_{i,*,\mathrm{bm}} = P_i \cdot Q_{*,\mathrm{bm}}, \tag{A7}$$

where $*$ is a general notation of the two wavebands (i.e., PAR and NIR).

Shortwave diffuse radiation. The scattering of light within the canopy is calculated using the Norman model (Norman & Campbell, 1989) as:

$$Q^{\downarrow}_{i,*,\mathrm{df}} = Q^{\downarrow}_{i+1,*,\mathrm{df}}[l_{\mathrm{df},i+1} + (1 - l_{\mathrm{df},i+1})\tau_*] + Q^{\uparrow}_{i,*,\mathrm{df}}(1 - l_{\mathrm{df},i+1})\, r_* + Q_{i+1,*,\mathrm{bm}}(1 - l_{\mathrm{bm},i+1})\tau_* \tag{A8}$$

$$Q^{\uparrow}_{i+1,*,\mathrm{df}} = Q^{\uparrow}_{i,*,\mathrm{df}}[l_{\mathrm{df},i+1} + (1 - l_{\mathrm{df},i+1})\tau_*] + Q^{\downarrow}_{i+1,*,\mathrm{df}}(1 - l_{\mathrm{df},i+1})\, r_* + Q_{i+1,*,\mathrm{bm}}(1 - l_{\mathrm{bm},i+1})\, r_*, \tag{A9}$$

where the superscripts $\downarrow$ and $\uparrow$ represent the downward and upward direction of the diffuse radiation; τ_* and r_* are the leaf transmittance and reflectance coefficients, respectively, of either PAR or NIR; $l_{\mathrm{bm},i}$ is the direct beam transmittance through $\Delta\mathcal{L}_i$ obtained by $l_{\mathrm{bm},i} = \exp(-\Delta\mathcal{L}_i \mathcal{G}\Omega/\sin\beta)$; and $l_{\mathrm{df},i}$ is the diffuse transmittance through layer i with leaf area index $\Delta\mathcal{L}_i$ obtained by summing $\exp(-\Delta\mathcal{L}_i \mathcal{G}\Omega/\sin\beta)$ over all sky angles β. Here, the downward shortwave diffuse radiation at the canopy top $Q^{\downarrow}_{M,*,\mathrm{df}}$ is given by $Q_{*,\mathrm{df}}$.

Longwave radiation. The longwave radiation profile across the canopy is also calculated by using the Norman algorithm based on the incoming longwave radiation from the sky, L_{sky}, which is calculated using the following function Crawford and Duchon (1999):

$$L_{\mathrm{sky}} = L_{\mathrm{sky,clear}}\left(1 - \frac{Q_{\mathrm{TOT}}}{Q_{\mathrm{TOT,clear}}}\right) + \frac{Q_{\mathrm{TOT}}}{Q_{\mathrm{TOT,clear}}}\sigma T_{\mathrm{a}}^4, \tag{A10}$$

where $Q_{\mathrm{TOT,clear}}$ is the theoretical clear sky downward solar radiation (W m^{-2}) calculated by Allen (1998); T_{a} is the air temperature (K); $L_{\mathrm{sky,clear}}$ is the clear sky downwelling longwave radiation determined by the method introduced in Brunt (1932):

$$L_{\mathrm{sky,clear}} = \left(0.605 + 0.048 e_{\mathrm{a}}^{1/2}\right)\sigma T_{\mathrm{a}}^4. \tag{A11}$$

Given L_{sky}, we used the Norman algorithm to further delineate the canopy radiation profile as:

$$L^{\downarrow}_i = L^{\downarrow}_{i+1}[l_{\mathrm{df},i+1} + (1 - l_{\mathrm{df},i+1})\, s] + L^{\uparrow}_i(1 - l_{\mathrm{df},i+1})\, s + \epsilon\sigma T^4_{l,i+1}(1 - l_{\mathrm{df},i+1}) \tag{A12}$$

$$L^{\uparrow}_{i+1} = L^{\uparrow}_i[l_{\mathrm{df},i+1} + (1 - l_{\mathrm{df},i+1})\, s] + L^{\downarrow}_{i+1}(1 - l_{\mathrm{df},i+1})\, s + \epsilon\sigma T^4_{l,i+1}(1 - l_{\mathrm{df},i+1}), \tag{A13}$$

where ϵ is the leaf emissivity coefficient and s is the scattering coefficient of L, given as $(1 - \epsilon)/2$. Note that the above equation replaces the direct beam scattering with emitted radiation based on leaf temperature. L_{sky} is set as the downward radiation at the canopy top, $L^{\downarrow}_M$.

Radiations absorbed by sunlit and shaded leaves. For both PAR and NIR, the maximum potential radiations that reaches sunlit and shaded leaves are then calculated. For shaded leaves, the radiation is the sum of the upward and downward diffusive radiations absorbed by the leaf and can be calculated as:

$$Q_{i,*,\mathrm{shade}} = \left(Q^{\downarrow}_{i,*,\mathrm{df}} + Q^{\uparrow}_{i,*,\mathrm{df}}\right)(1 - \tau_* - r_*). \tag{A14}$$

Sunlit leaves additionally absorb the unscattered direct beam radiation, which is the direct beam radiation that reaches the layer adjusted by the leaf/sun angles and absorbed by the leaf. This is given as:

$$Q_{i,*,\text{sun}} = Q_{i,*,\text{bm}}\frac{\mathcal{G}}{\sin\beta}(1-\tau_* - r_*) + Q_{i,*,\text{shade}}. \tag{A15}$$

The canopy longwave radiation contributes equally to the radiation budget of the two types of leaves as:

$$L_{i,\text{shade}} = L_i^{\downarrow} + L_i^{\uparrow} \tag{A16}$$

$$L_{i,\text{sun}} = L_i^{\downarrow} + L_i^{\uparrow}. \tag{A17}$$

Total radiation at each canopy layer. The total radiations of sunlit and shaded leaves at each layer is the sum of the longwave and shortwave radiation components such that:

$$Q_{\text{in},i,\text{shade}} = Q_{i,\text{PAR,shade}} + Q_{i,\text{NIR,shade}} + \epsilon L_{i,\text{shade}} \tag{A18}$$

$$Q_{\text{in},i,\text{sun}} = Q_{i,\text{PAR,sun}} + Q_{i,\text{NIR,sun}} + \epsilon L_{i,\text{sun}}. \tag{A19}$$

A2. Leaf Boundary Layer Conductance

The model used the flat plate theory (Schuepp, 1993) to compute the leaf boundary layer conductance $g_{\text{b},m}$ in Equation A1 as

$$g_{\text{b},m} = \frac{d\text{Sh}}{ls}, \tag{A20}$$

where ls is a leaf's length scale; Sh is the Sherwood number; and d is the molecular diffusivity using the calculation in Massman (1998).

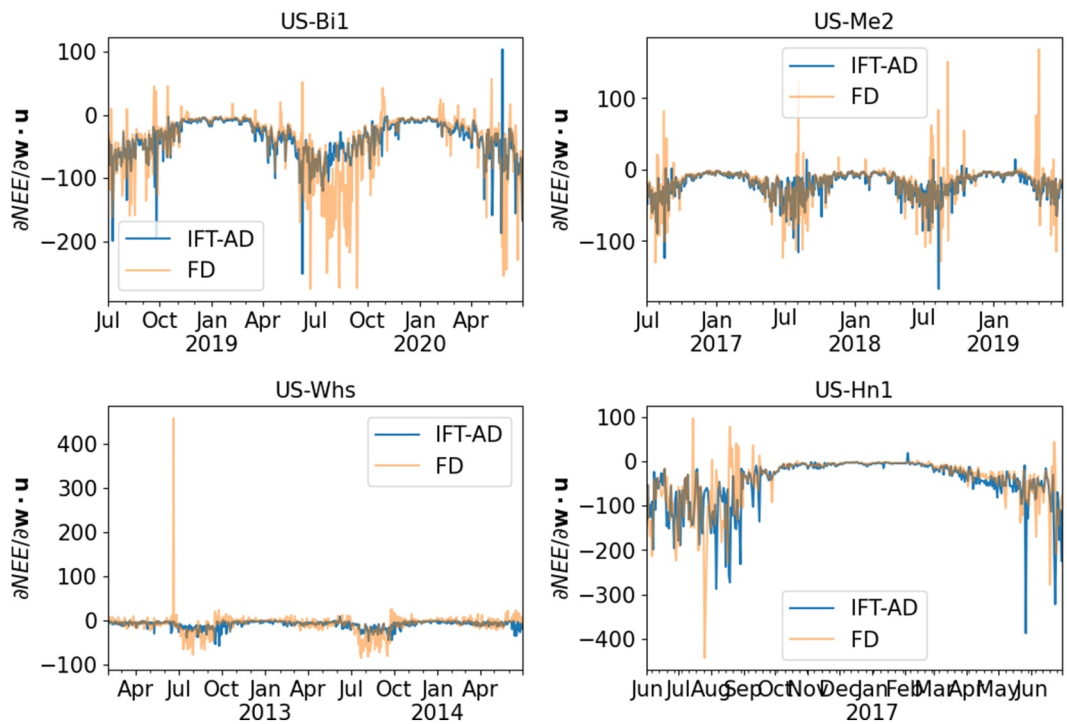

Figure A2. Comparison between implicit function theorem-based automatic differentiation (IFT-AD) and finite difference-based differentiation (FD) using the directional derivative of *NEE* using the process-based JAX-CanVeg on the default parameters **w** listed in Table A1. (The directional derivative of IFT-AD is calculated as $\frac{\partial NEE}{\partial \mathbf{w}} \cdot \mathbf{u}$ where $\mathbf{u} = [1, \ldots, 1]_{n_w}$ is a unit vector with size n_w; and the directional derivative of FD is calculated as $\frac{F_{NEE}(\mathbf{x}, \mathbf{w}+h\mathbf{u}) - F_{NEE}(\mathbf{x}, \mathbf{w})}{h}$, where $h = 0.01$ and $F_{NEE} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_w} \to \mathbb{R}$ is the function of JAX-CanVeg to calculate *NEE*.

A3. Photosynthesis and Stomatal Conductance

The photosynthesis process is simulated by the Farquhar model (Farquhar et al., 1980) given by:

$$A = V_c - 0.5V_o - R_d = \min[W_c, W_j]\left(1 - \frac{\Gamma^*}{C_i}\right) - R_d, \tag{A21}$$

where A is the photosynthesis rate (μmol m^{-2} s^{-1}); V_c is the carboxylation rate (μmol m^{-2} s^{-1}); V_o is the photorespiration rate (μmol m^{-2} s^{-1}); R_d is the dark respiration rate (μmol m^{-2} s^{-1}); Γ^* is the CO_2 compensation point in the absence of dark respiration (ppm); C_i is the intercellular CO_2 concentration (ppm); and W_c and W_j are the carboxylation rates restricted by the Rubisco under saturation of ribulose biphosphate (RuBP) and electron transport for RuBP regeneration, respectively. The formulations of W_c and W_j follow the Michaelis-Menton type model as:

$$W_c = \frac{V_{cmax} C_i}{C_i + K_c\left(1 + \frac{[O_2]}{K_o}\right)} \tag{A22}$$

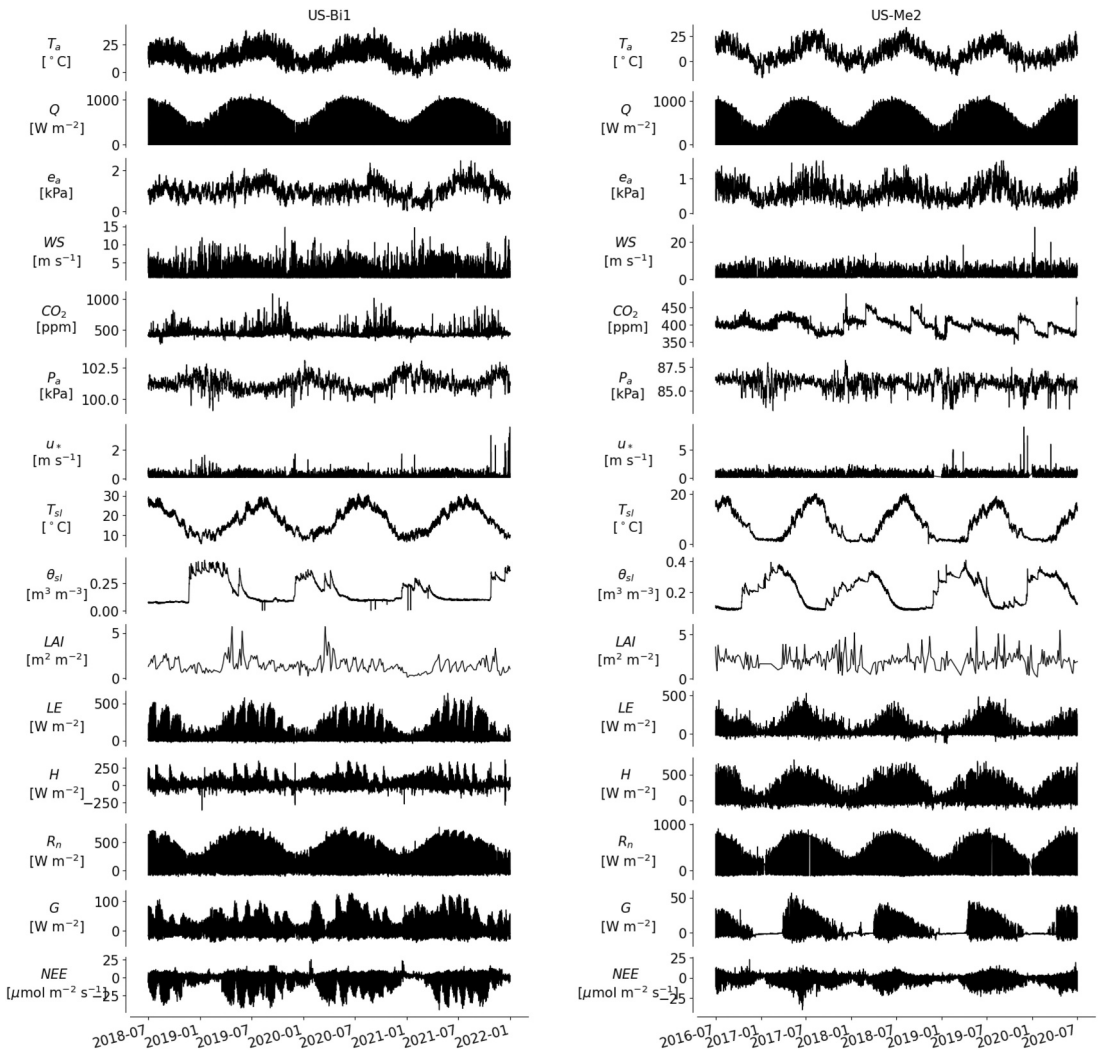

Figure A3. Time series of flux tower observations and MODIS leaf area index at both flux tower sites US-Bi1 (left) and US-Me2 (right), including air temperature (T_a, °C), ambient vapor pressure (e_a, kPa), air pressure (P_a, hPa), ambient CO_2 concentration (C_a, ppm), wind speed (WS, m s^{-1}), friction velocity (u_*, m s^{-1}), solar radiation (Q, W m^{-2}), soil temperature (T_s, °C), volumetric soil water content (θ_s, m^3 m^{-3}), latent heat flux (LE, W m^{-2}), net ecosystem exchange (NEE, μmol m^{-2} s^{-1}), net radiation (R_n, W m^{-2}), sensible heat flux (H, W m^{-2}), and ground heat flux (G, W m^{-2}).

$$W_j = \frac{JC_i}{4C_i + 8\Gamma^*}, \tag{A23}$$

where K_o and K_c are the Michaelis-Menton coefficients for O_2 and CO_2, respectively; $[O_2]$ is the oxygen concentration; V_{cmax} is the maximum carboxylation rate by Rubisco (μmol m^{-2} s^{-1}); and J is the potential rate of electron transport and is given as:

$$J = \frac{\alpha I}{\sqrt{1 + \frac{\alpha^2 I^2}{J_{max}^2}}}, \tag{A24}$$

where $I = 4.6 \cdot Q_{PAR,*}$ is the incident photosynthetically active photon flux (mmols m^{-2} s^{-1}) and converted from $Q_{PAR,*}$ in Equation A15, α is the quantum yield (mol e$^-$ mol^{-1} quanta) and J_{max} is the maximum rate of electron transport (μmol m^{-2} s^{-1}), respectively. Values of J_{max}, V_{cmax}, K_c, K_o, and R_d are temperature dependent. K_c, K_o, and R_d are calculated using the Arrhenius equation that delineates an exponential increase with the temperature (see Equation 10 in Bernacchi et al. (2013)). J_{max} and V_{cmax} are calculated using the modified Arrhenius equation to account for the value decrease at extreme temperatures (see Equation 11 in Bernacchi et al. (2013)).

Simple conductance relations are used to relate C_i with the surface (C_s) and ambient (C_a) CO_2 concentrations as:

$$C_s = C_a - \frac{A}{g_{b,m}/1.6} \tag{A25}$$

$$C_i = C_s - \frac{A}{g_{s,m}/1.6}, \tag{A26}$$

where $g_{b,m}$ and $g_{s,m}$ are the conductances of the leaf laminar boundary layer and the stomata, respectively, in units of mol m^{-2} s^{-1}. C_a and $g_{b,m}$ are external inputs to the photosynthesis model calculated in Equation A20. $g_{s,m}$ is calculated using the Ball-Berry model as:

$$g_{s,m} = \frac{mA\theta_l}{C_s} + g_0, \tag{A27}$$

where θ_l is the leaf relative humidity; and m and g_0 are the model slope and intercept, respectively. m and g_0 are tunable parameters. Here, we obtain θ_l by using both an equation with saturation assumption and a hybrid model that links soil moisture content to account for the plant water deficit (see Equation 13 in the main manuscript).

Equations A21–A27 lead to a cubic equation of A. We used the resulting analytical solution of A derived in Baldocchi (1994).

Fluxes intergrated across canopy. The total fluxes from the canopy is the weighted sum of the fluxes from each canopy layer timed by the leaf area index in the layer:

$$A_{can} = \sum_{i=1}^{M} (A_{sun,i} P_{sun,i} + A_{shade,i} P_{shade,i}) \Delta \mathcal{L}_i \tag{A28}$$

$$R_{d,can} = \sum_{i=1}^{M} (R_{d,sun,i} P_{sun,i} + R_{d,shade,i} P_{shade,i}) \Delta \mathcal{L}_i, \tag{A29}$$

where $P_{sun,i}$ is the probability of sunlit or beam portion of the leave and calculated as the derivative of P_i with respect to leaf area index $\mathcal{L}_i$, adjusted for the sunlit leaf angles:

$$P_{sun,i} = -\frac{\sin\beta}{\mathcal{G}} \frac{dP_0}{d\mathcal{L}_i} = \Omega \exp\left(-\frac{\mathcal{L}_i \mathcal{G} \Omega}{\sin\beta}\right). \tag{A30}$$

Correspondingly, the remaining part is the shaded portion:

$$P_{\text{shade},i} = 1 - P_{\text{sun},i}. \tag{A31}$$

A4. Leaf Energy Balance

The energy balance of a two-sided leaf is conserved among the net radiation $R_{\text{n,l}}$ (W m^{-2}), the latent heat flux LE_{l} (W m^{-2}), the sensible heat flux H (W m^{-2}), the incoming radiation $Q_{\text{in,l}}$ (W m^{-2}), and the outgoing longwave radiation L_{l} (W m^{-2}). For simplicity, we omit the layer index i and shaded or sunlit options and express the leaf energy balance as below:

$$R_{\text{n,l}} = Q_{\text{in,l}} - L_{\text{l}} = LE_{\text{l}} + H_{\text{l}} \tag{A32}$$

$$LE_{\text{l}} = \frac{0.622\lambda\rho_{\text{a}}g_{\text{w}}}{P_{\text{a}}}(e_{\text{s}}(T_{\text{l}}) - e_{\text{a}}) \tag{A33}$$

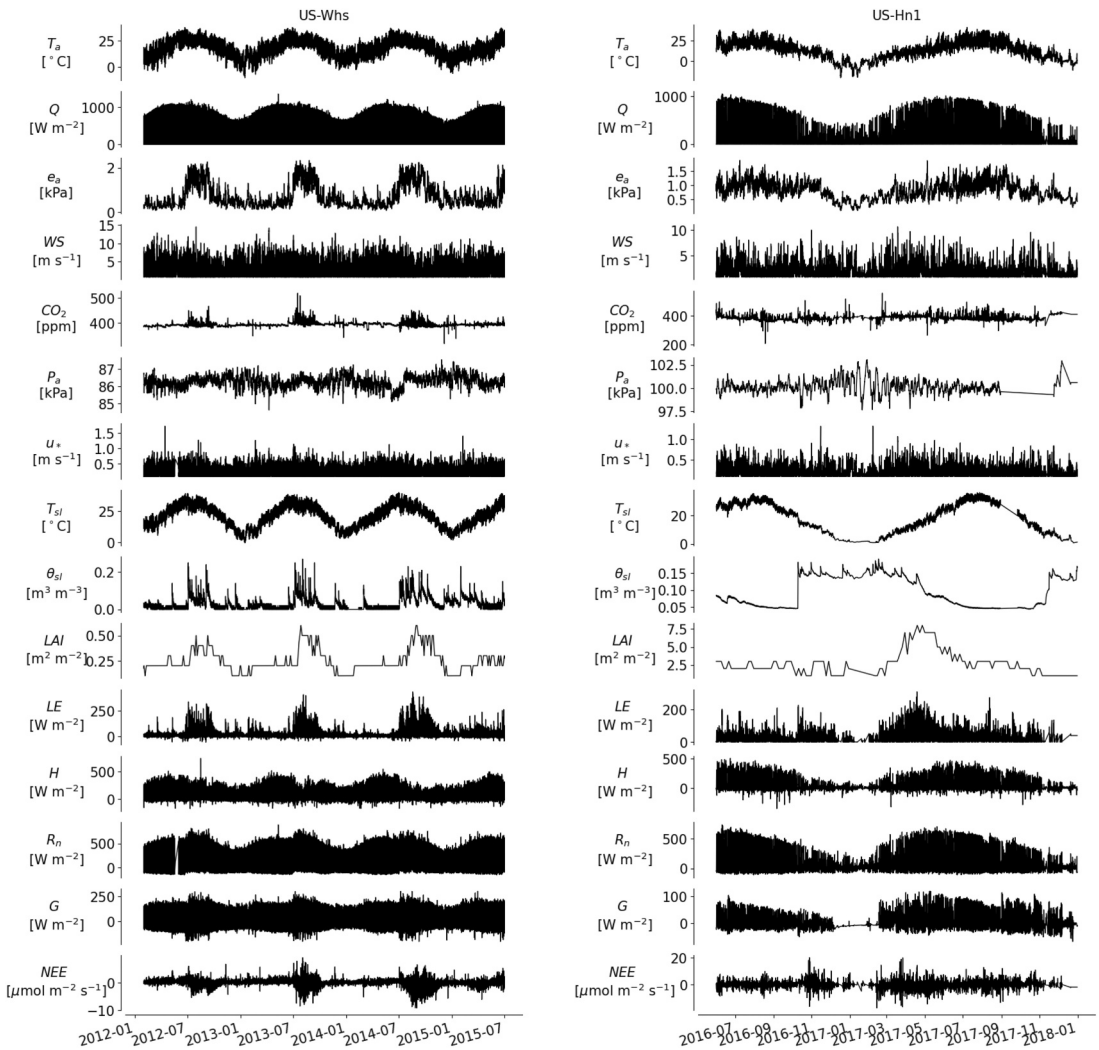

Figure A4. Time series of flux tower observations and MODIS leaf area index at both flux tower sites US-Whs (left) and US-Hn1 (right), including air temperature (T_a, °C), ambient vapor pressure (e_a, kPa), air pressure (P_a, hPa), ambient CO_2 concentration (C_a, ppm), wind speed (WS, m s^{-1}), friction velocity (u_*, m s^{-1}), solar radiation (Q, W m^{-2}), soil temperature (T_s, °C), volumetric soil water content (θ_s, m^3 m^{-3}), latent heat flux (LE, W m^{-2}), net ecosystem exchange (NEE, μmol m^{-2} s^{-1}), net radiation (R_n, W m^{-2}), sensible heat flux (H, W m^{-2}), and ground heat flux (G, W m^{-2}).

$$H_l = 2\rho_a C_p g_h (T_l - T_a) \tag{A34}$$

$$L_l = 2\epsilon\sigma T_l^4, \tag{A35}$$

where ρ_a is air density; λ is the latent heat of vapourization; g_w is the conductance for water vapor transfer (m s^{-1}); g_h is the conductance for sensible heat transfer (m s^{-1}); P_a is air pressure (kPa); T_l is the leaf temperature (K); T_a is the air temperature (K); e_a is the ambient vapor pressure (kPa); e_s is the saturated vapor pressure (kPa); and C_p is the specific heat of air. The incoming radiation Q_{in} is calculated in the radiative transfer subroutine (Equations A18 and A19). Here, we omit the subscripts for the layer index and the leaf type for convenience.

Solution for leaf temperature T_l. By using the Taylor Series expansion to approximate $\epsilon\sigma T_l^4$ and $e_s(T_l)$ through Clausis Claperyon, up to the second order, we can derive the quadratic equation to solve the difference between T_l and T_a as below (Paw U & Gao, 1988):

$$a_1 \Delta T^2 + b_1 \Delta T + c_1 = 0. \tag{A36}$$

The coefficients are given as:

$$a_1 = 12\epsilon\sigma T_a^2 + \frac{d^2 e_s(T_a)}{dT_a^2}\frac{\gamma_{LE_l}}{2} \tag{A37}$$

$$b_1 = 8\epsilon\sigma T_a^2 + \gamma_{H_l} + \gamma_{LE_l}\frac{de_s(T_a)}{dT_a} \tag{A38}$$

$$c_1 = -Q_{in,l} + L_a + \gamma_{LE_l}(e_s(T_a) - e_a), \tag{A39}$$

where $\gamma_{H_l} = 2\rho_a C_p g_h$, $\gamma_{LE_l} = \frac{0.622\lambda\rho_a g_w}{P_a}$, and $L_a = 2\epsilon\sigma T_a^4$. Once ΔT is solved, we can obtain $T_l = T_a + \Delta T$ and thus obtain H, L, R_n based on Equations A32, A34, and A35, respectively.

Solution for leaf latent heat flux LE_l. A similar quadratic equation can be derived for LE as below for checking the energy balance in Equation A32:

$$a_2 LE_l^2 + b_2 LE_l + c_2 = 0. \tag{A40}$$

The coefficients a_2, b_2, and c_2 depend on the leaf type. For hypostomatous leaves, the conductance of water vapor transfer g_w is defined as $g_w = \frac{g_s g_b}{g_s + g_b}$ with g_s and g_b as stomatal conductance (m s^{-1}) and aerodynamic conductance for water vapor transfer (m s^{-1}), respectively. Here, the conversions from $g_{s,m}$ and $g_{b,m}$ are performed based on:

$$g_s = g_{s,m} T_l \frac{101.3 \times 0.022624}{273.15 P_a} \tag{A41}$$

$$g_b = g_{b,m} T_l \frac{101.3 \times 0.022624}{273.15 P_a}. \tag{A42}$$

Correspondingly, the coefficients in Equation A40 are given as

$$a_2 = \gamma_{LE_l}\frac{d^2 e_s(T_a)}{dT_a^2}\frac{1}{4\gamma} \tag{A43}$$

$$b_2 = -\gamma - \gamma_{LE_l}\frac{de_s(T_a)}{dT_a} - 2a_2(Q_{in,l} - L_a) \tag{A44}$$

$$c_2 = a_2\left(Q_{in,l}^2 - 2Q_{in,l}L_a + L_a^2\right) + \gamma_{LE_l}\left((e_s(T_a) - e_a)\gamma + \frac{de_s(T_a)}{dT_a}(Q_{in,l} - L_a)\right), \tag{A45}$$

where $\gamma = \gamma_{H_l} + 8\epsilon\sigma T_a^3$.

For amphistomatous leaves, g_w is defined as $g_w = \frac{g_{top}+g_{bottom}}{g_{top}g_{bottom}}$, where $g_{top} = g_{bottom} = \frac{1}{g_b + g_s/2}$. The coefficients are derived as:

$$a_2 = \gamma_{LE_l}\frac{d^2 e_s(T_a)}{dT_a^2}\frac{1}{2\gamma} \tag{A46}$$

$$b_2 = -\gamma - \gamma_{LE_l}\frac{de_s(T_a)}{dT_a} - 2a_2\frac{Q_{in,l}}{\gamma} + 2a_2\frac{L_a}{\gamma} \tag{A47}$$

$$c_2 = a_2\left(Q_{in,l}^2 - 2Q_{in,l}L_a + L_a^2\right) + \gamma_{LE_l}\left((e_s(T_a) - e_a)\gamma + \frac{de_s(T_a)}{dT_a}(Q_{in,l} - L_a)\right). \tag{A48}$$

States and fluxes integrated across canopy. The total states and fluxes from the canopy is the weighted sum of the fluxes from sunlit and shaded leaves at each canopy layer timed by the leaf area index in the layer:

$$g_{s,can} = \sum_{i=1}^{M}\left(g_{s,sun,i}P_{sun,i} + g_{s,shade,i}P_{shade,i}\right)\Delta\mathcal{L}_i \tag{A49}$$

$$T_{l,can} = \sum_{i=1}^{M}\left(T_{l,sun,i}P_{sun,i} + T_{l,shade,i}P_{shade,i}\right)\Delta\mathcal{L}_i \tag{A50}$$

$$R_{n,can} = \sum_{i=1}^{M}\left(R_{n,l,sun,i}P_{sun,i} + R_{n,l,shade,i}P_{shade,i}\right)\Delta\mathcal{L}_i \tag{A51}$$

$$H_{can} = \sum_{i=1}^{M}\left(H_{l,sun,i}P_{sun,i} + H_{l,shade,i}P_{shade,i}\right)\Delta\mathcal{L}_i \tag{A52}$$

$$LE_{can} = \sum_{i=1}^{M}\left(LE_{l,sun,i}P_{sun,i} + LE_{l,shade,i}P_{shade,i}\right)\Delta\mathcal{L}_i. \tag{A53}$$

A5. Soil Energy Balance

The soil column is divided into $N + 1$ layers with $T_{sl,i}$ representing the soil temperature at the ith layer ($i \in [0, \ldots, N + 1]$). We denote $T_{surf} = T_{sl,0}$ as the ground surface temperature. The soil energy balance is solved by coupling the surface energy balance with soil heat transfer.

Surface energy balance. At the soil surface (one-sided), the energy balance is conserved as:

$$R_{n,surf} = Q_{in,surf} - L_{surf} = LE_{surf} + H_{surf} + G \tag{A54}$$

$$LE_{surf} = \frac{0.622\lambda\rho_a g_{w,surf}}{P_a}(e_s(T_{surf}) - e_a) \tag{A55}$$

$$H_{surf} = 2\rho_a C_p g_{h,surf}(T_{surf} - T_a) \tag{A56}$$

$$L_{surf} = \epsilon_{surf}\sigma T_{surf}^4, \tag{A57}$$

$$Q_{in,surf} = Q_{1,PAR,bm} + Q_{1,PAR,df}^{\downarrow} - Q_{1,PAR,df}^{\uparrow} + Q_{1,NIR,bm} + Q_{1,NIR,df}^{\downarrow} - Q_{1,NIR,df}^{\uparrow} + \epsilon L_1^{\downarrow} \tag{A58}$$

where the subscript surf refers to the states or energy flux components at the ground surface; T_{surf} is the surface temperature (K); ϵ_{surf} is the soil surface emissivity; G is the ground heat flux (W m^{-2}); and $Q_{in,s}$ is the overall incoming radiation reaching the ground (W m^{-2}). The sensible heat conductance of soil, $g_{h,surf}$, is calculated using

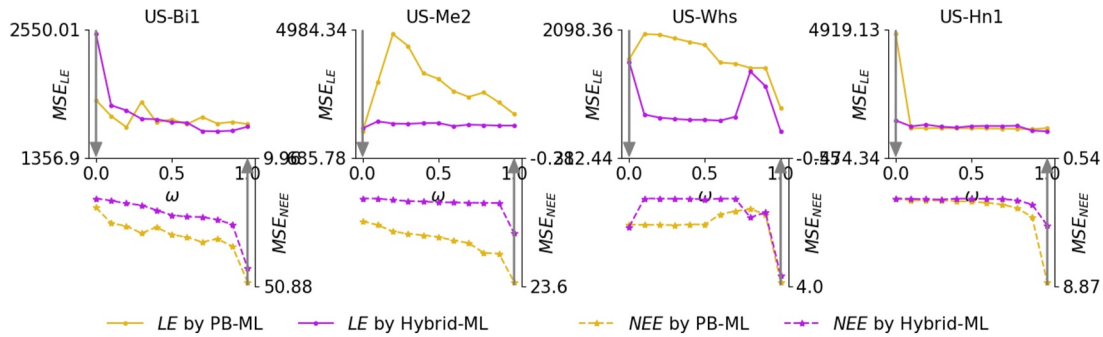

Figure A5. Learning trade-off between *LE* and *NEE*. These plots show the change of *MSE* in Equation 15 over ω used in a multi-objective optimization setting in Equation 14 at the four sites. In each subplot, NSE_{LE} and NSE_{NEE} are plotted in the upper and bottom parts of the frame with yellow and violet lines representing the results of `PB-ML` and `Hybrid-ML` models, respectively. The gray vertical arrow indicates the trend of an improved performance of *MSE*.

the Daamen and Simmonds methods (Daamen & Simmonds, 1996). The latent heat conductance, $g_{\mathrm{w,surf}}$, is calculated as $\frac{1}{1/g_{\mathrm{h,surf}}+1/g_{\mathrm{H2O,surf}}}$ with the surface water conductance $g_{\mathrm{H2O,surf}}$ computed using the Kondo model (Kondo et al., 1990).

Having the soil temperature profile solved (see below), we can calculate the ground heat flux G as:

$$G^t = k_{\mathrm{sl},1}\left(T^t_{\mathrm{sl},1} - T^t_{\mathrm{sl},2}\right) + c_{p,\mathrm{sl}}\left(T^t_{\mathrm{sl},1} - T^{t-1}_{\mathrm{sl},1}\right), \tag{A59}$$

where the superscript t is the time index; $k_{\mathrm{sl},1}$ is the top soil thermal conductivity (W m^{-2} K^{-1}); $c_{p,\mathrm{sl}}$ is the top soil heat capacity rate (W m^{-2} K^{-1}).

Solution for surface latent heat flux LE_{surf}. Following the linearization idea of Equation A33, LE_{surf} can be obtained by solving a quadratic equation as:

$$a_3 LE^2_{\mathrm{surf}} + b_3 LE_{\mathrm{surf}} + c_3 = 0, \tag{A60}$$

where a_3, b_3, and c_3 are the coefficients that are derived as (based on the linearization of both Clausis Claperyon and Stefan Bolzmann equations up to the second order):

$$a_3 = \gamma_{LE_{\mathrm{surf}}} \frac{d^2 e_{\mathrm{s}}(T_{\mathrm{a}})}{dT^2_{\mathrm{a}}} \frac{1}{2\gamma_{\mathrm{surf}}} \tag{A61}$$

$$b_3 = -\gamma_{\mathrm{surf}} - \gamma_{LE_{\mathrm{surf}}} \frac{de_{\mathrm{s}}(T_{\mathrm{a}})}{dT_{\mathrm{a}}} - 2a_3\left(Q_{\mathrm{in,surf}} - L_{\mathrm{a}} - G\right) \tag{A62}$$

$$\begin{aligned} c_3 = {} & a_3\left(Q^2_{\mathrm{in,surf}} + L^2_{\mathrm{a}} + G^2 - 2Q_{\mathrm{in,surf}}L_{\mathrm{a}} - 2Q_{\mathrm{in,surf}}G + 2GL_{\mathrm{a}}\right) \\ & + \gamma_{LE_{\mathrm{surf}}}\left(e_{\mathrm{s}}(T_{\mathrm{a}}) - e_{\mathrm{a}}\right)\gamma_{\mathrm{surf}} + \frac{de_{\mathrm{s}}(T_{\mathrm{a}})}{dT_{\mathrm{a}}}\left(Q_{\mathrm{in,surf}} - L_{\mathrm{a}} - G\right)\Big), \end{aligned} \tag{A63}$$

where $\gamma_{\mathrm{surf}} = \rho_{\mathrm{a}} C_p g_{\mathrm{h,surf}} + 4\epsilon_{\mathrm{s}}\sigma T^3_{\mathrm{a}}$; $\gamma_{LE_{\mathrm{surf}}} = \frac{0.622\lambda\rho_{\mathrm{a}} g_{\mathrm{w,surf}}}{P_{\mathrm{a}}}$.

Solution for surface temperature T_{surf}. Similar to the derivation of T_1, we can derive the quadratic equation to solve the difference between T_{surf} and T_{a} such that:

$$a_4 \Delta T^2_{\mathrm{surf}} + b_4 \Delta T_{\mathrm{surf}} + c_4 = 0. \tag{A64}$$

The coefficients are given as:

$$a_4 = 6\epsilon_{\text{surf}}\sigma T_{\text{a}}^2 + \frac{d^2 e_{\text{s}}(T_{\text{a}})}{dT_{\text{a}}^2}\frac{\gamma_{LE,surf}}{2} \tag{A65}$$

$$b_4 = 4\epsilon\sigma T_{\text{a}}^2 + \gamma_{H,\text{surf}} + \gamma_{LE_{\text{surf}}}\frac{de_{\text{s}}(T_{\text{a}})}{dT_{\text{a}}} \tag{A66}$$

$$c_4 = -Q_{\text{in,surf}} + L_{\text{a}} + G + \gamma_{LE_{\text{surf}}}(e_{\text{s}}(T_{\text{a}}) - e_{\text{a}}), \tag{A67}$$

where $\gamma_{H,\text{surf}} = \rho_{\text{a}} C_p g_{\text{h,surf}}$. Subsequently, the derived ΔT_{surf} allows the calculation of T_{surf}, H_{surf}, L_{surf}, and $R_{\text{n,surf}}$.

Soil heat transfer. A transient, one-dimensional heat conduction in the soil requires:

$$c_{\text{v}}\frac{\partial T_{\text{sl}}}{\partial t} = -\frac{\partial F}{\partial z} = \frac{\partial}{\partial z}\left(K\frac{\partial T_{\text{sl}}}{\partial z}\right), \tag{A68}$$

where c_{v} is volumetric heat capacity (J m^{-3} K^{-1}); T_{sl} is the soil temperature (K); K is soil thermal conductivity (W m^{-1} K^{-1}); and F is the heat flux by conduction (W m^{-2}) following the Fourier's law.

The soil column is discretized into N evenly divided layers. Let $\bar{K}_j = \eta K_j^{t+1} + (1-\eta)K_j^t$ with $\eta \in [0,1]$ and $j \in [0,\ldots,N]$. The discretized K_j^t represents K between layers $j+1$ and j at the time step t. We can further discretize the partial differential equation (Equation A68) into:

$$c_{\text{v}}\frac{T_{\text{sl},j}^{t+1} - T_{\text{sl},j}^t}{\Delta t}\frac{z_{j+1} - z_{j-1}}{2} = \bar{K}_j\frac{\left[\eta T_{\text{sl},j+1}^{t+1} + (1-\eta)T_{\text{sl},j+1}^t\right] - \left[\eta T_{\text{sl},j}^{t+1} + (1-\eta)T_{\text{sl},j}^t\right]}{z_{j+1} - z_j} - \bar{K}_{j-1}\frac{\left[\eta T_{\text{sl}}^{t+1} + (1-\eta)T_{\text{sl},j}^t\right] - \left[\eta T_{\text{sl},j-1}^{t+1} + (1-\eta)T_{\text{sl},j-1}^t\right]}{z_j - z_{j-1}}. \tag{A69}$$

Equation A69 is a general solution for temporal difference. When $\eta = 0$, Equation A69 yields a forward difference solution while it adopts a backward difference solution when $\eta = 1$.

Solving Equation A69 requires the rearrangement of the equation into the following linear equations

$$a_{\text{sl},j}T_{\text{sl},j-1}^{t+1} + b_{\text{sl},j}T_{\text{sl},j}^{t+1} + c_{\text{sl},j}T_{\text{sl},j+1}^{t+1} = d_{\text{sl},j}, \tag{A70}$$

producing a matrix of equations for solving $T_{\text{sl},i}^{t+1}$ with $a_{\text{sl},j}$, $b_{\text{sl},j}$, $c_{\text{sl},j}$, and $d_{\text{sl},j}$ the coefficients. We used the Thomas Algorithm (Bittelli et al., 2015) to solve the set of simultaneous equations with the conditions $a_{\text{sl},1} = 0$ and $c_{\text{sl},N} = 0$.

The top and bottom boundary conditions adopt the simulated and observed soil temperature. The top boundary condition $T_{\text{sl},0}^t$ uses the simulated surface temperature at the previous time step T_{surf}^{t-1}. The bottom temperature $T_{\text{sl},N+1}^t$ uses the observed temperature from sensors.

A6. Turbulence and Diffusion

Using Fick's law in Equation A1 requires getting the ambient concentration or heat content Con_{a}. The model employs a Lagrangian random walk method to resolve Con_{a} at any height z as below (Baldocchi, 1992):

$$Con_{\text{a}}(z) - Con_{\text{a}}(r) = \sum_{i=1}^{M} S_i(Con_{\text{l}}(z_i))\, D_{z,z_i}\Delta_{z_i}, \tag{A71}$$

where Con_{l} is the source of the particles from the canopy; r is a reference level (e.g., flux tower); Δ_{z_i} is the incremental height of the $i$the layer; and D_{z,z_i} is a dispersion matrix in units of s m^{-1} calculated using the algorithm of Thomson (1987). Equation A71 describes that the concentration between the reference height r and any

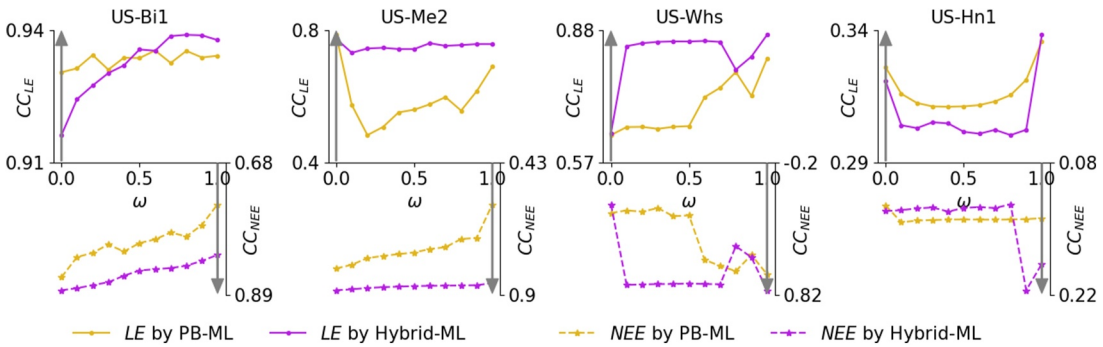

Figure A6. Learning trade-off between *LE* and *NEE*. These plots show the change of *CC* in Equation 17 over ω used in a multi-objective optimization setting in Equation 14 at the four sites. In each subplot, NSE_{LE} and NSE_{NEE} are plotted in the upper and bottom parts of the frame with yellow and violet lines representing the results of `PB-ML` and `Hybrid-ML` models, respectively. The gray vertical arrow indicates the trend of an improved performance of *CC*.

arbitrary height z at the ambient condition is the sum of the contributions of materials from different layers of the canopy (Baldocchi et al., 1999).

A7. Calculation of Overall Fluxes

The energy fluxes reaching the top of canopy is the sum of the fluxes from both canopy and soil as:

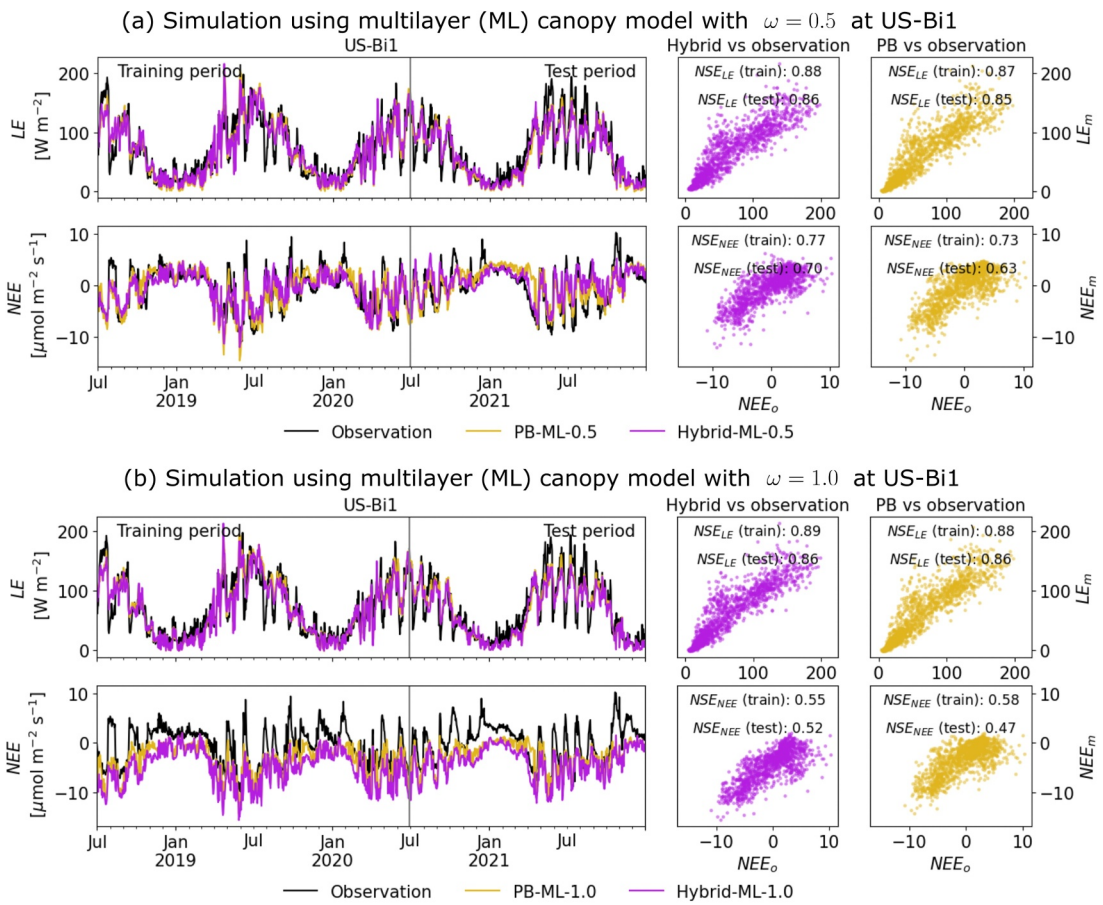

Figure A7. JAX-CanVeg simulation using the multi-layer canopy setup at US-Bi1. (a, b) plot the daily averaged simulations of *LE* and *NEE* with both the hybrid (violet) and process-based (yellow) models, using $\omega = 0.5$ and $\omega = 1.0$ in the loss function (Equation 14), respectively.

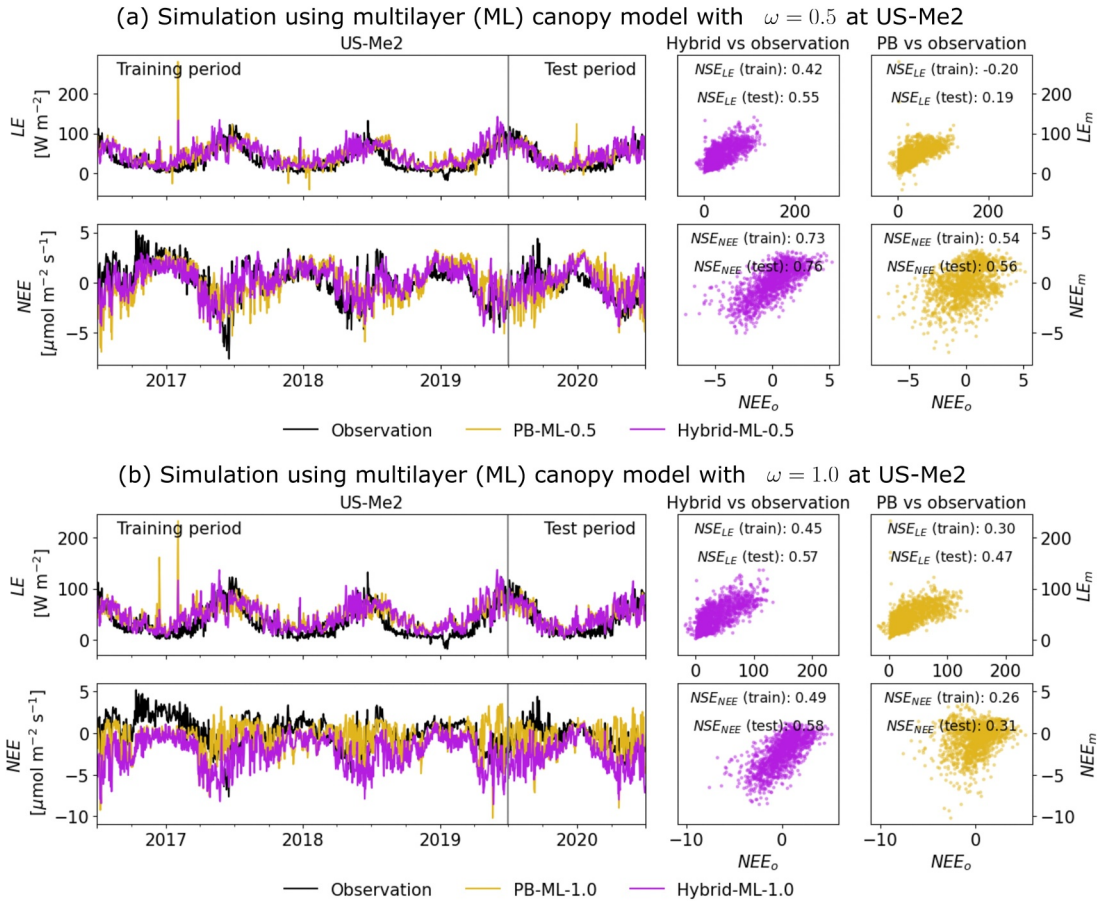

Figure A8. JAX-CanVeg simulation using the multi-layer canopy setup at US-Me2. (a, b) plot the daily averaged simulations of *LE* and *NEE* with both the hybrid (violet) and process-based (yellow) models, using $\omega = 0.5$ and $\omega = 1.0$ in the loss function (Equation 14), respectively.

$$R_{\mathrm{n}} = R_{\mathrm{n,can}} + R_{\mathrm{n,surf}} \tag{A72}$$

$$LE = LE_{\mathrm{can}} + LE_{\mathrm{surf}} \tag{A73}$$

$$H = H_{\mathrm{can}} + H_{\mathrm{surf}}. \tag{A74}$$

The corresponding carbon flux is calculated as the net ecosystem exchange *NEE* such that:

$$NEE = R_{\mathrm{sl}} - GPP, \tag{A75}$$

$$GPP = A_{\mathrm{can}} - R_{\mathrm{d,can}}, \tag{A76}$$

where *GPP* is the gross primary production (μmol m^{-2} s^{-1}) and R_{sl} is the soil respiration (μmol m^{-2} s^{-1}). R_{sl} is calculated by using a Q10-power model (B. Wang et al., 2014) through soil temperature and soil moisture as:

$$R_{\mathrm{sl}} = a_{\mathrm{r}} \cdot b_{\mathrm{r}}^{\frac{(T_{\mathrm{sl}}-273.15)-10}{10}} \cdot {\theta_{\mathrm{sl}}}^{c_{\mathrm{r}}}, \tag{A77}$$

where θ_{sl} is the volumetric soil water content (m^3 m^{-3}) and a_{r}, b_{r}, and c_{r} are tunable parameters.

Data Availability Statement

The source code of JAX-CanVeg is available and maintained at https://github.com/pnnl/JAX-CanVeg. A code copy is also downloadable from Zenodo at Jiang et al. (2024).

Acknowledgments

The authors thank Nate Mcdowell from Pacific Northwest National Laboratory and three reviewers for their insightful suggestions to improve this work. This work was funded by both the ExaSheds project and the Laboratory Directed Research and Development Program at Pacific Northwest National Laboratory. The Laboratory Directed Research and Development Program at Pacific Northwest National Laboratory is a multiprogram national laboratory operated by Battelle for the U.S. Department of Energy. The ExaSheds project was supported by the United States Department of Energy, Office of Science, Office of Biological and Environmental Research, Earth and Environmental Systems Sciences Division, Data Management Program, under Award Number DE-AC02-05CH11231. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science of the United States Department of Energy under contract DE-AC02-05CH11231. Pacific Northwest National Laboratory is operated for the DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the United States Department of Energy or the United States Government. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paidup, irrevocable, world-wide license to publish, or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan, last access: 4 April 2024).

References

Aboelyazeed, D., Xu, C., Hoffman, F. M., Liu, J., Jones, A. W., Rackauckas, C., et al. (2023). A differentiable, physics-informed ecosystem modeling and learning framework for large-scale inverse problems: Demonstration with photosynthesis simulations. *Biogeosciences*, *20*(13), 2671–2692. https://doi.org/10.5194/bg-20-2671-2023

Allen, R. G. (1998). Crop evapotranspiration. *FAO irrigation and drainage paper*, *56*, 60–64.

AlQuraishi, M., & Sorger, P. K. (2021). Differentiable biology: Using deep learning for biophysics-based and data-driven modeling of molecular mechanisms. *Nature Methods*, *18*(10), 1169–1180. https://doi.org/10.1038/s41592-021-01283-4

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., et al. (2020). Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models. *Biogeosciences*, *17*(16), 4173–4222. https://doi.org/10.5194/bg-17-4173-2020

Baldocchi, D. D. (1992). A Lagrangian random-walk model for simulating water vapor, CO_2 and sensible heat flux densities and scalar profiles over and within a soybean canopy. *Boundary-Layer Meteorology*, *61*(1–2), 113–144. https://doi.org/10.1007/BF02033998

Baldocchi, D. D. (1994). An analytical solution for coupled leaf photosynthesis and stomatal conductance models. *Tree Physiology*, *14*(7–8-9), 1069–1079. https://doi.org/10.1093/treephys/14.7-8-9.1069

Baldocchi, D. D. (2023). Canveg – Matlab version. Retrieved from https://github.com/baldocchi/CanVeg

Baldocchi, D. D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., et al. (2001). Fluxnet: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society*, *82*(11), 2415–2434. https://doi.org/10.1175/1520-0477(2001)082⟨2415:FANTTS⟩2.3.CO;2

Baldocchi, D. D., Fuentes, J. D., Bowling, D. R., Turnipseed, A. A., & Monson, R. K. (1999). Scaling isoprene fluxes from leaves to canopies: Test cases over a boreal aspen and a mixed species temperate forest. *Journal of Applied Meteorology*, *38*(7), 885–898. https://doi.org/10.1175/1520-0450(1999)038⟨0885:SIFFLT⟩2.0.CO;2

Baldocchi, D. D., & Harley, P. C. (1995). Scaling carbon dioxide and water vapour exchange from leaf to canopy in a deciduous forest. ii. model testing and application. *Plant, Cell and Environment*, *18*(10), 1157–1173. https://doi.org/10.1111/j.1365-3040.1995.tb00626.x

Baldocchi, D. D., Wilson, K. B., & Gu, L. (2002). How the environment, canopy structure and canopy physiological functioning influence carbon, water and energy fluxes of a temperate broad-leaved deciduous forest—An assessment with the biophysical model CANOAK. *Tree Physiology*, *22*(15–16), 1065–1077. https://doi.org/10.1093/treephys/22.15-16.1065

Ball, J. T., Woodrow, I. E., & Berry, J. A. (1987). A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions. In J. Biggins (Ed.), *Progress in photosynthesis research: Volume 4 proceedings of the viith international congress on photosynthesis providence, Rhode Island, USA, august 10–15, 1986* (pp. 221–224). Springer Netherlands. https://doi.org/10.1007/978-94-017-0519-6_48

Baydin, A. G., Pearlmutter, B. A., Radul, A. A., & Siskind, J. M. (2018). Automatic differentiation in machine learning: A survey. *Journal of Machine Learning Research*, *18*(153), 1–43. Retrieved from http://jmlr.org/papers/v18/17-468.html

Bernacchi, C. J., Bagley, J. E., Serbin, S. P., Ruiz-vera, U. M., Rosenthal, D. M., & Vanloocke, A. (2013). Modelling c3 photosynthesis from the chloroplast to the ecosystem. *Plant, Cell and Environment*, *36*(9), 1641–1657. https://doi.org/10.1111/pce.12118

Bezgin, D. A., Buhendwa, A. B., & Adams, N. A. (2023). Jax-fluids: A fully-differentiable high-order computational fluid dynamics solver for compressible two-phase flows. *Computer Physics Communications*, *282*, 108527. https://doi.org/10.1016/j.cpc.2022.108527

Bindas, T., Tsai, W.-P., Liu, J., Rahmani, F., Feng, D., Bian, Y., et al. (2024). Improving river routing using a differentiable muskingum-cunge model and physics-informed machine learning. *Water Resources Research*, *60*(1), e2023WR035337. https://doi.org/10.1029/2023WR035337

Bittelli, M., Campbell, G. S., & Tomei, F. (2015). *Soil physics with python: Transport in the soil-plant-atmosphere system*. OUP.

Bonan, G. B. (2019). *Climate change and terrestrial ecosystem modeling*. Cambridge University Press.

Bonan, G. B., Patton, E. G., Finnigan, J. J., Baldocchi, D. D., & Harman, I. N. (2021). Moving beyond the incorrect but useful paradigm: Reevaluating big-leaf and multilayer plant canopies to model biosphere-atmosphere fluxes – A review. *Agricultural and Forest Meteorology*, *306*, 108435. https://doi.org/10.1016/j.agrformet.2021.108435

Bonan, G. B., Williams, M., Fisher, R. A., & Oleson, K. W. (2014). Modeling stomatal conductance in the earth system: Linking leaf water-use efficiency and water transport along the soil–plant–atmosphere continuum. *Geoscientific Model Development*, *7*(5), 2193–2222. https://doi.org/10.5194/gmd-7-2193-2014

Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., et al. (2018). Jax: Composable transformations of Python+NumPy programs. Retrieved from http://github.com/google/jax

Brunt, D. (1932). Notes on radiation in the atmosphere. I. *Quarterly Journal of the Royal Meteorological Society*, *58*(247), 389–420. https://doi.org/10.1002/qj.49705824704

Busch, F. A., Ainsworth, E. A., Amtmann, A., Cavanagh, A. P., Driever, S. M., Ferguson, J. N., et al. (2024). A guide to photosynthetic gas exchange measurements: Fundamental principles, best practice and potential pitfalls. *Plant, Cell and Environment*, *47*(9), 3344–3364. https://doi.org/10.1111/pce.14815

Chaney, N. W., Herman, J. D., Ek, M. B., & Wood, E. F. (2016). Deriving global parameter estimates for the noah land surface model using fluxnet and machine learning. *Journal of Geophysical Research: Atmospheres*, *121*(22), 13218–13235. https://doi.org/10.1002/2016JD024821

Chen, M., Qian, Z., Boers, N., Jakeman, A. J., Kettner, A. J., Brandt, M., et al. (2023). Iterative integration of deep learning in hybrid earth surface system modelling. *Nature Reviews Earth & Environment*, *4*(8), 568–581. https://doi.org/10.1038/s43017-023-00452-7

Chen, Y., Ryder, J., Bastrikov, V., McGrath, M. J., Naudts, K., Otto, J., et al. (2016). Evaluating the performance of land surface model orchidee-can v1.0 on water and energy flux estimation with a single- and multi-layer energy budget scheme. *Geoscientific Model Development*, *9*(9), 2951–2972. https://doi.org/10.5194/gmd-9-2951-2016

Collatz, G., Ball, J., Grivet, C., & Berry, J. A. (1991). Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: A model that includes a laminar boundary layer. *Agricultural and Forest Meteorology*, *54*(2), 107–136. https://doi.org/10.1016/0168-1923(91)90002-8

Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A., & Totterdell, I. J. (2000). Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, *408*(6809), 184–187. https://doi.org/10.1038/35041539

Crawford, T. M., & Duchon, C. E. (1999). An improved parameterization for estimating effective atmospheric emissivity for use in calculating daytime downwelling longwave radiation. *Journal of Applied Meteorology*, *38*(4), 474–480. https://doi.org/10.1175/1520-0450(1999)038⟨0474:AIPFEE⟩2.0.CO;2

Daamen, C. C., & Simmonds, L. P. (1996). Measurement of evaporation from bare soil and its estimation using surface resistance. *Water Resources Research*, *32*(5), 1393–1402. https://doi.org/10.1029/96WR00268

DeepMind, Babuschkin, I., Baumli, K., Bell, A., Bhupatiraju, S., Bruce, J., et al. (2020). The DeepMind JAX ecosystem. Retrieved from http://github.com/google-deepmind

De Kauwe, M. G., Medlyn, B. E., Zaehle, S., Walker, A. P., Dietze, M. C., Wang, Y.-P., et al. (2014). Where does the carbon go? A model–data intercomparison of vegetation carbon allocation and turnover processes at two temperate forest free-air CO_2 enrichment sites. *New Phytologist*, *203*(3), 883–899. https://doi.org/10.1111/nph.12847

Drewry, D. T., Kumar, P., Long, S., Bernacchi, C., Liang, X.-Z., & Sivapalan, M. (2010). Ecohydrological responses of dense canopies to environmental variability: 1. Interplay between vertical structure and photosynthetic pathway. *Journal of Geophysical Research*, *115*(G4). https://doi.org/10.1029/2010JG001340

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

Farquhar, G., von Caemmerer, S., & Berry, J. (1980). A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 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

Fisher, R. A., & Koven, C. D. (2020). Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. *Journal of Advances in Modeling Earth Systems*, *12*(4), e2018MS001453. https://doi.org/10.1029/2018MS001453

Gelbrecht, M., White, A., Bathiany, S., & Boers, N. (2023). Differentiable programming for earth system modeling. *Geoscientific Model Development*, *16*(11), 3123–3135. https://doi.org/10.5194/gmd-16-3123-2023

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. Retrieved from http://www.deeplearningbook.org

Greener, J. G., & Jones, D. T. (2021). Differentiable molecular simulation can learn all the parameters in a coarse-grained force field for proteins. *PLoS One*, *16*(9), e0256990. https://doi.org/10.1371/journal.pone.0256990

Griewank, A., & Walther, A. (2008). *Evaluating derivatives* (2nd ed.). Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898717761

Grossiord, C., Buckley, T. N., Cernusak, L. A., Novick, K. A., Poulter, B., Siegwolf, R. T. W., et al. (2020). Plant responses to rising vapor pressure deficit. *New Phytologist*, *226*(6), 1550–1566. https://doi.org/10.1111/nph.16485

Häfner, D., Jacobsen, R. L., Eden, C., Kristensen, M. R. B., Jochum, M., Nuterman, R., & Vinter, B. (2018). Veros v0.1 – A fast and versatile ocean simulator in pure Python. *Geoscientific Model Development*, *11*(8), 3299–3312. https://doi.org/10.5194/gmd-11-3299-2018

Harley, P. C., & Baldocchi, D. D. (1995). Scaling carbon dioxide and water vapour exchange from leaf to canopy in a deciduous forest. I. Leaf model parametrization. *Plant, Cell and Environment*, *18*(10), 1146–1156. https://doi.org/10.1111/j.1365-3040.1995.tb00625.x

Hemes, K. S., Chamberlain, S. D., Eichelmann, E., Anthony, T., Valach, A., Kasak, K., et al. (2019). Assessing the carbon and climate benefit of restoring degraded agricultural peat soils to managed wetlands. *Agricultural and Forest Meteorology*, *268*, 202–214. https://doi.org/10.1016/j.agrformet.2019.01.017

Innes, M., Edelman, A., Fischer, K., Rackauckas, C., Saba, E., Shah, V. B., & Tebbutt, W. (2019). A differentiable programming system to bridge machine learning and scientific computing. *CoRR*. Retrieved from http://arxiv.org/abs/1907.07587

Jiang, P., Kidger, P., Bandai, T., Baldocchi, D., Liu, H., Xiao, Y., et al. (2024). JAX-CanVeg: A differentiable land surface model. *Zenodo*. https://doi.org/10.5281/zenodo.13936757

Katul, G., Manzoni, S., Palmroth, S., & Oren, R. (2009). A stomatal optimization theory to describe the effects of atmospheric CO_2 on leaf photosynthesis and transpiration. *Annals of Botany*, *105*(3), 431–442. https://doi.org/10.1093/aob/mcp292

Kemp, S., Scholze, M., Ziehn, T., & Kaminski, T. (2014). Limiting the parameter space in the Carbon Cycle Data Assimilation System (CCDAS). *Geoscientific Model Development*, *7*(4), 1609–1619. https://doi.org/10.5194/gmd-7-1609-2014

Kidger, P. (2022). On neural differential equations. Retrieved from https://arxiv.org/abs/2202.02435

Kidger, P., & Garcia, C. (2021). Equinox: Neural networks in JAX via callable PyTrees and filtered transformations. In *Differentiable Programming workshop at Neural Information Processing Systems 2021*.

Kochkov, D., Smith, J. A., Alieva, A., Wang, Q., Brenner, M. P., & Hoyer, S. (2021). Machine learning–accelerated computational fluid dynamics. *Proceedings of the National Academy of Sciences*, *118*(21), e2101784118. https://doi.org/10.1073/pnas.2101784118

Kondo, J., Saigusa, N., & Sato, T. (1990). A parameterization of evaporation from bare soil surfaces. *Journal of Applied Meteorology and Climatology*, *29*(5), 385–389. https://doi.org/10.1175/1520-0450(1990)029⟨0385:APOEFB⟩2.0.CO;2

Koven, C. D., Knox, R. G., Fisher, R. A., Chambers, J. Q., Christoffersen, B. O., Davies, S. J., et al. (2020). Benchmarking and parameter sensitivity of physiological and vegetation dynamics using the Functionally Assembled Terrestrial Ecosystem Simulator (FATES) at barro Colorado Island, Panama. *Biogeosciences*, *17*(11), 3017–3044. https://doi.org/10.5194/bg-17-3017-2020

Kumarathunge, D. P., Medlyn, B. E., Drake, J. E., Tjoelker, M. G., Aspinwall, M. J., Battaglia, M., et al. (2019). Acclimation and adaptation components of the temperature dependence of plant photosynthesis at the global scale. *New Phytologist*, *222*(2), 768–784. https://doi.org/10.1111/nph.15668

Laguë, M. M., Bonan, G. B., & Swann, A. L. S. (2019). Separating the impact of individual land surface properties on the terrestrial surface energy budget in both the coupled and uncoupled land–atmosphere system. *Journal of Climate*, *32*(18), 5725–5744. https://doi.org/10.1175/JCLI-D-18-0812.1

Lai, C.-T., Katul, G., Oren, R., Ellsworth, D., & Schäfer, K. (2000). Modeling CO_2 and water vapor turbulent flux distributions within a forest canopy. *Journal of Geophysical Research*, *105*(D21), 26333–26351. https://doi.org/10.1029/2000JD900468

Law, B. E., Cescatti, A., & Baldocchi, D. D. (2001). Leaf area distribution and radiative transfer in open-canopy forests: Implications for mass and energy exchange. *Tree Physiology*, *21*(12–13), 777–787. https://doi.org/10.1093/treephys/21.12-13.777

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G. B., 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

Leuning, R. (1995). A critical appraisal of a combined stomatal-photosynthesis model for C_3 plants. *Plant, Cell and Environment*, *18*(4), 339–355. https://doi.org/10.1111/j.1365-3040.1995.tb00370.x

Mäkelä, J., Knauer, J., Aurela, M., Black, A., Heimann, M., Kobayashi, H., et al. (2019). Parameter calibration and stomatal conductance formulation comparison for boreal forests with adaptive population importance sampler in the land surface model JSBACH. *Geoscientific Model Development*, *12*(9), 4075–4098. https://doi.org/10.5194/gmd-12-4075-2019

Massman, W. (1998). A review of the molecular diffusivities of H_2O, CO_2, CH_4, CO, O_3, SO_2, NH_3, N2O, No, and NO_2 in air, O_2 and N_2 near Stp. *Atmospheric Environment*, *32*(6), 1111–1127. https://doi.org/10.1016/S1352-2310(97)00391-9

Medlyn, B. E., Duursma, R. A., Eamus, D., Ellsworth, D. S., Prentice, I. C., Barton, C. V., et al. (2011). Reconciling the optimal and empirical approaches to modelling stomatal conductance. *Global Change Biology*, *17*(6), 2134–2144. https://doi.org/10.1111/j.1365-2486.2010.02375.x

Medlyn, B. E., Zaehle, S., De Kauwe, M. G., Walker, A. P., Dietze, M. C., Hanson, P. J., et al. (2015). Using ecosystem experiments to improve vegetation models. *Nature Climate Change*, *5*(6), 528–534. https://doi.org/10.1038/nclimate2621

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

Miner, G. L., & Bauerle, W. L. (2017). Seasonal variability of the parameters of the ball–berry model of stomatal conductance in maize (Zea Mays l.) and sunflower (Helianthus annuus l.) under well-watered and water-stressed conditions. *Plant, Cell and Environment*, *40*(9), 1874–1886. https://doi.org/10.1111/pce.12990

Missik, J. E. C., Liu, H., Gao, Z., Huang, M., Chen, X., Arntzen, E., et al. (2019). Groundwater-river water exchange enhances growing season evapotranspiration and carbon uptake in a semiarid riparian ecosystem. *Journal of Geophysical Research: Biogeosciences*, *124*(1), 99–114. https://doi.org/10.1029/2018JG004666

Monson, R., & Baldocchi, D. D. (2014). *Terrestrial biosphere-atmosphere fluxes*. Cambridge University Press.

Myneni, R. B., Ross, J., & Asrar, G. (1989). A review on the theory of photon transport in leaf canopies. *Agricultural and Forest Meteorology*, *45*(1), 1–153. https://doi.org/10.1016/0168-1923(89)90002-6

Myneni, R., P, T., & Knyazikhin, Y. (2021). MCD15A3H MODIS/Terra+Aqua leaf area index/FPAR 4-day l4 global 500m SIN Grid v061 [Dataset]. *NASA Eosdis land processes DAAC*. https://doi.org/10.5067/MODIS/MCD15A3H.061

Norman, J. M., & Campbell, G. S. (1989). Canopy structure. In R. W. Pearcy, J. R. Ehleringer, H. A. Mooney, & P. W. Rundel (Eds.), *Plant physiological ecology: Field methods and instrumentation* (pp. 301–325). Springer Netherlands. https://doi.org/10.1007/978-94-009-2221-1_14

Oikawa, P., Sturtevant, C., Knox, S., Verfaillie, J., Huang, Y., & Baldocchi, D. (2017). Revisiting the partitioning of net ecosystem exchange of CO_2 into photosynthesis and respiration with simultaneous flux measurements of $^{13}CO_2$ and CO_2, soil respiration and a biophysical model, Canveg. *Agricultural and Forest Meteorology*, *234–235*, 149–163. https://doi.org/10.1016/j.agrformet.2016.12.016

Paw U, K. T., & Gao, W. (1988). Applications of solutions to non-linear energy budget equations. *Agricultural and Forest Meteorology*, *43*(2), 121–145. https://doi.org/10.1016/0168-1923(88)90087-1

Post, H., Vrugt, J. A., Fox, A., Vereecken, H., & Hendricks Franssen, H.-J. (2017). Estimation of community land model parameters for an improved assessment of net carbon fluxes at European sites. *Journal of Geophysical Research: Biogeosciences*, *122*(3), 661–689. https://doi.org/10.1002/2015JG003297

Rader, J., Lyons, T., & Kidger, P. (2023). Lineax: Unified linear solves and linear least-squares in JAX and Equinox. In *AI for science workshop at neural information processing systems 2023*. arXiv:2311.17283.

Radul, A., Paszke, A., Frostig, R., Johnson, M. J., & Maclaurin, D. (2023). You only linearize once: Tangents transpose to gradients. *Proceedings of the ACM on Programming Languages*, *7*(POPL), 1246–1274. https://doi.org/10.1145/3571236

Raoult, N., Edouard-Rambaut, L.-A., Vuichard, N., Bastrikov, V., Lansø, A. S., Guenet, B., & Peylin, P. (2023). Using free air CO_2 enrichment data to constrain land surface model projections of the terrestrial carbon cycle. *EGUsphere*, *2023*, 1–26. https://doi.org/10.5194/egusphere-2023-360

Rey-Sanchez, C., Wang, C. T., Szutu, D., Shortt, R., Chamberlain, S. D., Verfaillie, J., & Baldocchi, D. D. (2022). AmeriFlux FLUXNET-1F US-Bi1 Bouldin island alfalfa, Ver. 3–5, AmeriFlux AMP [Dataset]. https://doi.org/10.17190/AMF/1871134

Schuepp, P. H. (1993). Tansley review No. 59. leaf boundary layers. *New Phytologist*, *125*(3), 477–507. https://doi.org/10.1111/j.1469-8137.1993.tb03898.x

Scott, R. L., Biederman, J. A., Hamerlynck, E. P., & Barron-Gafford, G. A. (2015). The carbon balance pivot point of southwestern U.S. Semiarid ecosystems: Insights from the 21st century drought. *Journal of Geophysical Research: Biogeosciences*, *120*(12), 2612–2624. https://doi.org/10.1002/2015JG003181

Sellers, P., Randall, D., Collatz, G., Berry, J., Field, C., Dazlich, D., et al. (1996). A revised land surface parameterization (SIB2) for atmospheric GCMS. Part I: Model formulation. *Journal of Climate*, *9*(4), 676–705. https://doi.org/10.1175/1520-0442(1996)009⟨0676:ARLSPF⟩2.0.CO;2

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., et al. (2023). Differentiable modelling to unify machine learning and physical models for geosciences. *Nature Reviews Earth & Environment*, *4*(8), 552–567. https://doi.org/10.1038/s43017-023-00450-9

Simon, E., Meixner, F. X., Ganzeveld, L., & Kesselmeier, J. (2005). Coupled carbon-water exchange of the amazon rain forest, i. model description, parameterization and sensitivity analysis. *Biogeosciences*, *2*(3), 231–253. https://doi.org/10.5194/bg-2-231-2005

Song, Y., Knoben, W. J. M., Clark, M. P., Feng, D., Lawson, K. E., & Shen, C. (2023). When ancient numerical demons meet physics-informed machine learning: Adjoint-based gradients for implicit differentiable modeling. *Hydrology and Earth System Sciences Discussions*, *2023*, 1–35. https://doi.org/10.5194/hess-2023-258

Tang, J., & Zhuang, Q. (2008). Equifinality in parameterization of process-based biogeochemistry models: A significant uncertainty source to the estimation of regional carbon dynamics. *Journal of Geophysical Research*, *113*(G4). https://doi.org/10.1029/2008JG000757

Thomson, D. J. (1987). Criteria for the selection of stochastic models of particle trajectories in turbulent flows. *Journal of Fluid Mechanics*, *180*(-1), 529–556. https://doi.org/10.1017/S0022112087001940

Tsai, W.-P., Feng, D., Pan, M., Beck, H., Lawson, K., Yang, Y., et al. (2021). From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling. *Nature Communications*, *12*(1), 1–13. https://doi.org/10.1038/s41467-021-26107-z

Verhoef, A., & Egea, G. (2014). Modeling plant transpiration under limited soil water: Comparison of different plant and soil hydraulic parameterizations and preliminary implications for their use in land surface models. *Agricultural and Forest Meteorology*, *191*, 22–32. https://doi.org/10.1016/j.agrformet.2014.02.009

Vickers, D., Thomas, C. K., Pettijohn, C., Martin, J. G., & Law, B. E. (2012). Five years of carbon fluxes and inherent water-use efficiency at two semi-arid pine forests with different disturbance histories. *Tellus B: Chemical and Physical Meteorology*, *64*(1), 17159. https://doi.org/10.3402/tellusb.v64i0.17159

Walker, A. P., Ye, M., Lu, D., De Kauwe, M. G., Gu, L., Medlyn, B. E., et al. (2018). The Multi-Assumption Architecture and Testbed (MAAT v1.0): R code for generating ensembles with dynamic model structure and analysis of epistemic uncertainty from multiple sources. *Geoscientific Model Development*, *11*(8), 3159–3185. https://doi.org/10.5194/gmd-11-3159-2018

Wang, B., Zha, T. S., Jia, X., Wu, B., Zhang, Y. Q., & Qin, S. G. (2014). Soil moisture modifies the response of soil respiration to temperature in a desert shrub ecosystem. *Biogeosciences*, *11*(2), 259–268. https://doi.org/10.5194/bg-11-259-2014

Wang, Y. P., Baldocchi, D. D., Leuning, R., Falge, E., & Vesala, T. (2007). Estimating parameters in a land-surface model by applying nonlinear inversion to eddy covariance flux measurements from eight Fluxnet sites. *Global Change Biology*, *13*(3), 652–670. https://doi.org/10.1111/j.1365-2486.2006.01225.x

Wang, Y.-P., & Leuning, R. (1998). A two-leaf model for canopy conductance, photosynthesis and partitioning of available energy I: Model description and comparison with a multi-layered model. *Agricultural and Forest Meteorology*, *91*(1), 89–111. https://doi.org/10.1016/S0168-1923(98)00061-6

Weiss, A., & Norman, J. (1985). Partitioning solar radiation into direct and diffuse, visible and near-infrared components. *Agricultural and Forest Meteorology*, *34*(2), 205–213. https://doi.org/10.1016/0168-1923(85)90020-6

Wiltshire, A. J., Duran Rojas, M. C., Edwards, J. M., Gedney, N., Harper, A. B., Hartley, A. J., et al. (2020). Jules-gl7: The global land configuration of the joint UK land environment simulator version 7.0 and 7.2. *Geoscientific Model Development*, *13*(2), 483–505. https://doi.org/10.5194/gmd-13-483-2020

Yan, Y., Klosterhalfen, A., Moyano, F., Cuntz, M., Manning, A. C., & Knohl, A. (2023). A modeling approach to investigate drivers, variability and uncertainties in O_2 fluxes and O_2: CO_2 exchange ratios in a temperate forest. *Biogeosciences*, *20*(19), 4087–4107. https://doi.org/10.5194/bg-20-4087-2023

Yokohata, T., Kinoshita, T., Sakurai, G., Pokhrel, Y., Ito, A., Okada, M., et al. (2020). Miroc-integ-land version 1: A global biogeochemical land surface model with human water management, crop growth, and land-use change. *Geoscientific Model Development*, *13*(10), 4713–4747. https://doi.org/10.5194/gmd-13-4713-2020

Zaehle, S., Medlyn, B. E., De Kauwe, M. G., Walker, A. P., Dietze, M. C., Hickler, T., et al. (2014). Evaluation of 11 terrestrial carbon–nitrogen cycle models against observations from two temperate free-air CO_2 enrichment studies. *New Phytologist*, *202*(3), 803–822. https://doi.org/10.1111/nph.12697