

Assessing and enhancing Noah-MP land surface modeling over tropical forests using machine learning techniques

Yanyan Cheng^{1,2*}, Yaomin Wang³, Kalli Furtado², Cenlin He⁴, Fei Chen⁵, Alan Ziegler⁶, Song Chen², Matteo Detto^{7,8}, Yuna Mao⁹, Baoxiang Pan¹⁰, Yoshiko Kosugi¹¹, Marryanna Lion¹², Shoji Noguchi¹³, Satoru Takanashi¹⁴, Lulie Melling¹⁵, Baoqing Zhang¹⁶

1 Asian School of the Environment, Nanyang Technological University, Singapore, Singapore

2 Centre for Climate Research Singapore, Singapore, Singapore

3 Department of Civil and Environmental Engineering, National University of Singapore, Singapore, Singapore

4 NSF National Center for Atmospheric Research, Boulder, Colorado, USA

5 Division of Environment and Sustainability, The Hong Kong University of Science and Technology, Hong Kong, Hong Kong, China

6 Andaman Coastal Research Station for Development, Faculty of Fisheries, Kasetsart University, Ranong, Thailand

7 Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey, USA

8 Smithsonian Tropical Research Institute, Panama City, Panama

9 State Key Laboratory of Earth Surface Processes and Disaster Risk Reduction, Faculty of Geographical Science, Beijing Normal University, Beijing, China

10 Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

11 Laboratory of Forest Hydrology, Division of Forest and Biomaterials Science, Graduate School of Agriculture, Kyoto University, Kyoto, Japan

12 Forest Research Institute Malaysia, Kepong, Selangor, Malaysia

13 Tohoku Research Center, Forestry and Forest Products Research Institute, Iwate, Japan

14 Kansai Research Center, Forestry and Forest Products Research Institute, Kyoto, Japan

15 UN Sustainable Development Solutions Network, Asia Headquarters, Sunway University, Bandar Sunway, Selangor, Malaysia

16 Key Laboratory of West China's Environmental Systems (Ministry of Education), College of Earth and Environmental Sciences, Lanzhou University, Lanzhou, Gansu, China

Correspondence to: Yanyan Cheng (yanyan.cheng@ntu.edu.sg)

Abstract. Tropical land-surface processes play a key role in Earth system dynamics, yet evaluation and calibration of land surface models in these regions remain limited. This study addresses this gap through site-specific calibration of the Noah with Multi-Parameterizations (Noah-MP) land surface model at two tropical forest sites: the Panama site with a tropical monsoon climate and the Malaysia site with a tropical rainforest climate. We develop an efficient machine learning-based calibration framework that combines emulator construction with parameter optimization. Site-specific calibration improves the model's ability to simulate key variables, including latent and sensible heat fluxes as well as soil moisture, particularly at daily and seasonal scales. The emulator-based framework enables efficient calibration and shows robust performance during independent validation periods at each site. While the calibrated parameters are not directly transferable between the two sites with different climate regimes, they perform effectively during unseen validation periods at their respective sites, suggesting the potential transferability of parameters across similar climates. Remaining challenges include simulating nighttime sensible heat fluxes, balancing the optimization of latent and sensible heat fluxes, and capturing seasonal soil moisture dynamics. These limitations may stem from structural simplifications in Noah-MP, such as the lack of multi-species vegetation modeling, soil organic layer representation, and detailed tropical subsurface hydrology. Our results demonstrate how the calibration framework can effectively improve Noah-MP performance in tropical forests and provide guidance for future development priorities, which

[can support](#) broader efforts to generalize model calibration strategies and improve Earth system model fidelity in data-scarce, climatically distinct regions.

40 1 Introduction

Land surface processes play a vital role in modulating regional and global climate dynamics through mechanisms such as evapotranspiration, the exchange of momentum, albedo, and greenhouse gas emissions (Bonan, 1995; Cheruy et al., 2014; Cox et al., 1999, 2000; Crossley et al., 2000; Dickinson, 1983; Lawrence et al., 2012; Lin et al., 2017; Mahmood et al., 2014; Pielke et al., 2011; Pitman et al., 2009; Unger, 2014; Van Weverberg et al., 2018; Teo et al., 2022; Yan et al., 2022; Zhang et al., 2026). These processes are particularly critical in tropical regions, which act as key drivers of global climate systems (Gentine et al., 2019). Tropical forests, for instance, are the largest terrestrial carbon sinks, regulating the Earth's climate by cycling vast amounts of carbon, water and energy (Pan et al., 2011; Schlesinger and Jasechko, 2014). Additionally, tropical land surfaces significantly impact local weather patterns, monsoon circulation, and extreme weather events such as tropical cyclones and intense convective storms (Gentine et al., 2019; Li et al., 2016). Understanding these processes is essential for predicting climate change and variability, managing water resources, and conserving biodiversity in these ecologically vital regions.

Recent rapid urbanization, deforestation, forestation, agricultural expansion, and wildfires across the tropical regions have profoundly altered land cover and land use patterns (Lambin et al., 2003; Vetrta and Cochrane, 2020). These changes have significant implications for surface energy, water, and carbon cycles, as they modify surface albedo, heat fluxes, soil moisture, vegetation covers, and local climate dynamics (Hasler et al., 2024; Sun et al., 2021). For example, revegetation can enhance evapotranspiration, leading to a cooling effect (Shao and Zhang, 2019). However, reduced surface albedo associated with revegetation may increase temperature by absorbing more shortwave radiation, potentially outweighing the cooling effect of increased evapotranspiration (Zhang et al., 2022). Such changes are particularly concerning in tropical regions, which are prone to extreme weather events like tropical cyclones, monsoon disturbances, and intense convective storms. Land surface processes can influence the development, intensity, and tracks of these events through land-atmospheric interactions (Ma et al., 2013). Accurate representation of these changes in land surface and land-atmosphere interactive processes is therefore crucial for improving early warning systems, disaster preparedness, and sustainable land management strategies in tropical regions.

Land surface models (LSMs) are essential tools for simulating these complex processes, predicting their changes, and supporting land management interventions and policy maker decisions (Bonan and Doney, 2018; Weiskopf et al, 2022). Over the years, LSMs have evolved from simplified representations of land surface biophysics to sophisticated frameworks that integrate biophysical, biogeochemical, hydrological, and ecological processes (Fisher and Koven, 2020). They are widely used to assess the impacts of land-use changes on water, energy, and carbon cycles, evaluate land-atmosphere interactions, and inform strategies for managing the water-food-energy-carbon nexus (Cheng et al., 2022; Cheng et al., 2024a).

Despite recent advancements, the complex LSMs still require extensive testing and calibration across diverse climate regimes. This issue relates to the fact that default parameterizations, which are often developed using mid-latitude observational data, may not translate well to other biomes with fundamentally different environmental conditions (Lu et al., 2018, 2024). Each region (tropical, temperate, boreal, and arid) differs in key environmental and biological processes, and land cover types such as forests, grasslands, and urban areas exhibit distinct vegetation dynamics and land-atmosphere interactions. Consequently, using default configurations could lead to poor representation of temporal variability in surface fluxes (e.g., latent and sensible heat, photosynthesis) and misrepresentation of biogeochemical and hydrological processes (e.g., soil moisture and streamflow dynamics) (Lu et al., 2018, 2024).

This challenge is pronounced in tropical regions, where high biodiversity, complex canopy structures, and distinct wet-dry seasonality complicate model performance. For example, representing tropical forest diversity in land surface models remains challenging due to limited understanding of the mechanisms underlying forest functional diversity. In particular, few studies have calibrated the ecological coexistence of tropical tree species, which are the processes that allow multiple species with differing traits to persist together over time within the same ecosystem (Cheng et al., 2024b). Moreover, the critical role of soil hydrological processes in shaping tropical ecosystem behaviour has been underexplored in model calibration efforts (Cheng et al., 2022; 2023; 2024b).

To address these limitations, site- or region-specific calibration is essential for improving model accuracy and ensuring reliable simulations at local, regional, and global scales (Cheng et al., 2023; Huang et al., 2016; Lu and Ricciuto, 2019; Müller et al., 2015; R. Sun et al., 2021). Particularly important is the joint tuning of soil and vegetation parameters and validation across multiple timescales, especially in regions with pronounced dry-wet seasonality such as the tropics (Ogden et al., 2013).

Effective calibration of LSMs requires optimization approaches that can efficiently explore high-dimensional parameter spaces and avoid becoming trapped in local optima arising from strong nonlinearity and complex parameter interactions (Gupta et al., 1999). To address these challenges, population-based global optimization algorithms have been widely adopted, particularly in hydrological applications. Prominent examples include the Shuffled Complex Evolution (SCE-UA) (Duan et al., 1992) and Differential Evolution (DE) (Storn and Price, 1997), which is particularly robust for navigating the rugged, non-differentiable response surfaces typical of LSMs.

However, these global search algorithms typically require tens of thousands of model evaluations to converge. The computational burden and wall-clock time increase significantly when these simulations cannot be fully parallelized, because each iteration depends on the results of the previous set of simulations. For process-rich LSMs, such computational demands are often prohibitive. To address this challenge, recent studies have increasingly employed machine learning (ML)-based emulators (or surrogates) to reproduce model behaviours with high fidelity but at a fraction of the computational cost, thereby enabling the extensive sampling required for robust global optimization (Lu et al., 2018; Sun et al., 2021; Lu et al., 2024).

The objective of this study is to assess and enhance the performance of the widely-used Noah with Multi-Parameterization options (Noah-MP) LSM in tropical regions. Noah-MP stands out for its versatility and comprehensive treatment of land surface processes (Chang et al., 2020; He et al., 2023a; Niu et al., 2011; Yang et al., 2011). Its multi-physics framework

facilitates the incorporation of diverse parameterizations, enabling its applications across a wide range of environmental conditions. Its seamless integration with numerical weather prediction (NWP) models, such as the Weather Research and Forecasting (WRF) model, the Multiscale Prediction and Analysis System (MPAS), and NOAA’s Unified Forecast System (UFS), significantly enhances its utility in climate and weather simulations. However, despite its widespread use, Noah-MP applications remain understudied in tropical environments, resulting in limited application and systematic evaluation under the distinct hydroclimatic, ecological, and land-atmosphere coupling regimes characteristic of the tropics. In this study, we evaluate and calibrate Noah-MP against observational data from two distinct tropical forest sites in Panama (tropical monsoon climate) and Malaysia (tropical rainforest climate). We develop an effective machine learning-based calibration framework for Noah-MP that combines emulator construction with parameter optimization. Using this framework, we assess the model’s ability to simulate diurnal, daily, and seasonal dynamics of land surface energy fluxes and soil moisture with calibrated parameters. Recognizing that Noah-MP’s performance is sensitive to the choice of key vegetation and soil parameters, we explicitly calibrate these two types of parameters and quantify their individual contributions to overall model performance improvements. We further discuss the transferability of the emulator-based calibration framework and calibrated parameters across different climate regimes, as well as future development directions for Noah-MP.

2 Methods and data

2.1 Study area and field data for model evaluation and calibration

We use observational data from two tropical sites in Panama and Malaysia (Figure 1, Table 1) to calibrate Noah-MP and improve its performance over tropical forests. Specifically, we evaluate and improve Noah-MP’s performance in simulating latent heat (LH), sensible heat (SH), and soil moisture (SM), which are key variables influencing land-atmosphere interactions.

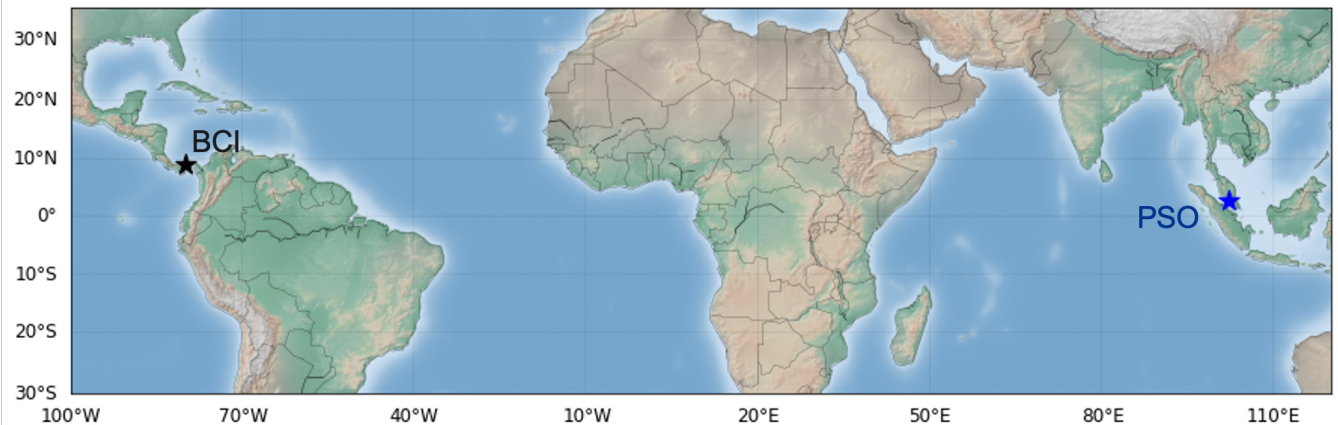


Figure 1. Locations of the two tropical forest study sites in Barro Colorado Island, Panama (BCI, black star) and Pasoh Forest Reserve, Malaysia (PSO, blue star).

Table 1. Detailed information for the two tropical forest study sites.

	Panama BCI	Malaysia PSO
Latitude	9.151	2.9667
Longitude	-79.855	102.3
Vegetation	Evergreen broadleaf forest	Evergreen broadleaf forest
Data availability	20150730 ~ 20170831	20030101 ~ 20091231
Wind speed (m/s)	2.94	1.86
Temperature (°C)	25.5	25.3
Humidity	89.0%	82.9%
Air pressure (mb)	988	989
Shortwave radiation (W/m ²)	195	198
Longwave radiation (W/m ²)	428	419
Precipitation (mm/year)	2025	1790
Latent heat (W/m ²)	71.2	98.6
Sensible heat (W/m ²)	30.9	45.0

2.1.1 Panama tropical forest site

The Barro Colorado Island (BCI) site in Panama is located at 9.151 °N, 79.855 °W. It is a primary lowland (elevation 120 m) semi-deciduous tropical forest site with a mean canopy height of 30 [m](#) and few emergent trees (Cheng et al., 2022; Detto et al., 2015; Detto and Pacala, 2022; Fang et al., 2022). It has distinct dry (mid-December to mid-April) and wet (late-April to early December) seasons (Ogden et al., 2013). The site exhibits a tropical monsoon climate (Am), characterized by a short dry season. The soils are mostly well weathered kaolinitic Oxisols. Meteorological and flux data were collected from a 45 m tower located on the top-plateau at a half-hour temporal resolution (Detto and Pacala, 2022). The mean annual precipitation, wind speed, air temperature, relative humidity, air pressure, shortwave radiation, and longwave radiation during 2015 to 2017 are 2025 mm, 2.94 m/s, 25.5 °C, 89%, 98.8 kPa, 195 W/m², and 428 W/m², respectively (Table 1). The soil moisture is measured in the top 15 cm of the soil (Cheng et al., 2022; Detto and Pacala, 2022).

2.1.2 Malaysia tropical forest site

The Pasoh Forest Reserve (PFR), known as the PSO AsiaFlux site, is a tropical forest located in Malaysia at 2.9667 °N, 102.3 °E. This lowland forest (140 m elevation) boasts a canopy height of 30 ~ 40 m, with emergent trees exceeding 45 m. The site

is notable for being located within a dry zone of Peninsular Malaysia and receives the lowest annual rainfall among adjacent south-eastern tropical rainforests (Lion et al., 2017; Noguchi et al., 2016). The soils are dominated by Ultisols and Oxisols, which are Haplic Acrisol and Xanthic Ferralsol in FAO classification. The site has a tropical rainforest climate (Af), with consistently high rainfall throughout the year.

The observed data were measured at a height of 54 m on the flux tower at a half-hour temporal resolution (Lion et al., 2017). The mean annual precipitation, wind speed, air temperature, relative humidity, air pressure, shortwave radiation, and longwave radiation from 2003 to 2009 are 1790 mm, 1.86 m/s, 25.3 °C, 82.9%, 98.9 kPa, 198 W/m², 419 W/m², respectively (Table 1). The volumetric soil water content is calculated as the average of observations from nine TDR (time domain reflectometry) sensors placed at depths of 10, 20, and 30 cm at three locations around the flux tower (Noguchi et al., 2016).

2.2 Noah-MP

The community Noah-MP LSM (He et al., 2023a, 2023b; Niu et al., 2011) is designed for a wide range of applications, including uncoupled studies of hydrometeorological and ecohydrological processes, as well as coupled numerical weather prediction and decadal climate simulations. It can operate from point to global scales. The land grid in Noah-MP is divided into two sub-grid tiles: vegetated and non-vegetated areas, determined by vegetation cover fraction. The biogeophysical and biogeochemical processes are treated separately for these two tiles. Noah-MP employs a “big-leaf” canopy approach, with canopy properties varying according to the type of vegetation (He et al., 2023a). More detailed descriptions of Noah-MP have been provided elsewhere (Chen et al., 2011; He et al., 2023c).

The core strength of Noah-MP lies in its multi-parameterization design, allowing testing and use of different combinations of parameterization schemes (Li et al., 2020; 2022; Gan et al., 2019). Prior to calibration, we tested several physical options that could potentially impact LH and SH, including stomatal resistance, soil moisture factor, and surface resistance options. No significant difference in model performance was observed at either site (Figures S1-S2); therefore, we adopted the default Noah-MP physics option set (He et al., 2023a) for parameter optimization.

2.3 Model simulations

We configure single-point Noah-MP simulations at the two study sites and calibrate the model at the daily scale. For the Panama tropical forest site, the vegetation type is set as evergreen broadleaf forest and the soil type as clay. The simulated soil moisture for the top 10 cm is used as the closest available depth for validation against the observations at the site in Panama. At the Malaysia site, the vegetation type is also evergreen broadleaf forest, with silty clay soil. The simulated soil moisture averaged over the top 30 cm is used to validate against the observed soil water content at the site.

2.4 Calibrated parameters

The key vegetation and soil parameters for calibration are selected based on the past experience and literature reports (Arsenault et al., 2018) (Table 2), which include the maximum rate of carboxylation at 25 degC (VCMX25), top of canopy

(HVT), bottom of canopy (HVB), momentum roughness length (Z0MVT), empirical canopy wind absorption parameter (CWPVT), saturated value of soil moisture (MAXSMC), saturated hydraulic conductivity (SATDK), wilting point (WLTSMC), and field capacity (REFSMC). For vegetation parameters, VCMX25 regulates the photosynthesis rate and stomatal resistance, and HVT and HVB determine canopy thickness. CWPVT affects turbulent heat transfer below canopies, while Z0MVT controls sensible and latent heat fluxes above the canopy. For soil parameters, MAXSMC, WLTSMC, and REFSMC determine soil water retention and the amount of water available for plant uptake, while SATDK controls the infiltration rate.

180

Table 2. Noah-MP parameters used in the sensitivity analysis.

Parameter	Full name	Relevant process	Bound	Reference
VCMX25	Maximum rate of carboxylation at 25 degC (umol CO ₂ /m ² /s)	Photosynthesis rate	30 ~ 120	Domingues et al., (2005); Chen et al., (2008)
HVT	Top of canopy (m)	Plant transpiration	30 ~ 55	Kosugi et al., (2012)
HVB	Bottom of canopy (m)	Plant transpiration	5 ~ 20	Kosugi et al., (2012)
CWPVT	Empirical canopy wind absorption parameter	In and below canopies sensible heat	0.15 ~ 5.35	Goudriaan (1977)
Z0MVT	Momentum roughness length (m)	Above-canopy sensible heat	0.3 ~ 2.0	He et al., (2023)
WLTSMC	Wilting point (m ³ /m ³)	Soil moisture	0.02 ~ 0.26	Marryanna et al., (2019)
REFSMC	Field capacity (m ³ /m ³)	Soil moisture	0.15 ~ 0.45	Litt et al., 2020 (2020)
MAXSMC	Saturated value of soil moisture (m ³ /m ³)	Soil moisture	0.45 ~ 0.75	Litt et al., 2020 (2020)
SATDK	Saturated hydraulic conductivity (m/s)	Soil moisture	$1 \times 10^{-6} \sim 5 \times 10^{-4}$	Marryanna et al., (2019)

2.5 ML-based calibration framework

2.5.1 Overview

Figure 2 illustrates the algorithmic workflow of the proposed ML-based calibration framework, which integrates machine-learning-based emulation with a population-based global optimization algorithm to efficiently estimate optimal Noah-MP

parameters while preserving the model’s physical behavior. By replacing repeated expensive model evaluations with a fast emulator, the framework substantially reduces computational cost while maintaining accuracy and robustness.

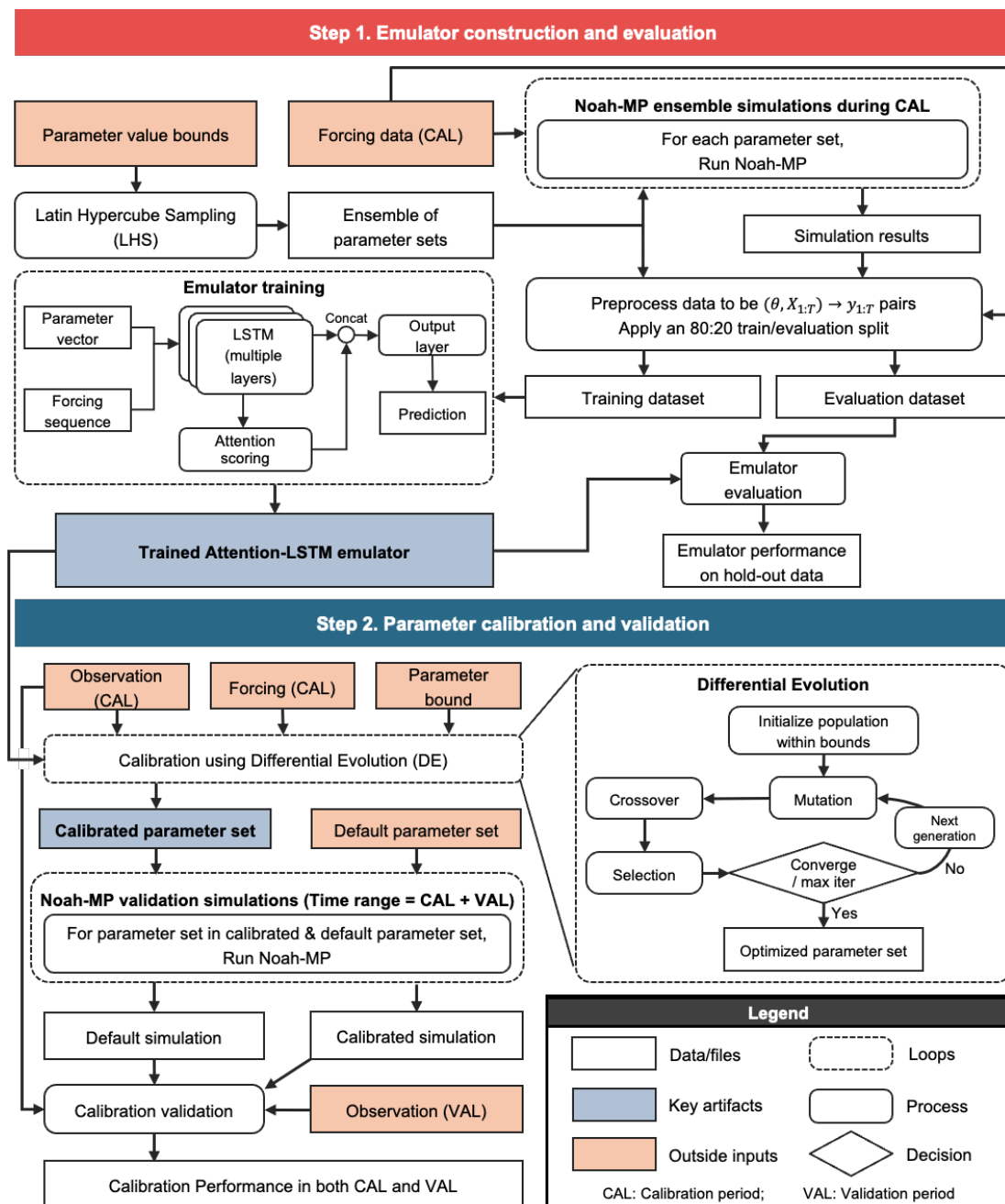


Figure 2. The algorithmic workflow of the two-step calibration framework. The $(\theta, X_{1:T}) \rightarrow Y_{1:T}$ pairs in Step 1 represent the mapping from parameter vector θ and meteorological forcing $X_{1:T}$ to the Noah-MP simulated outputs $Y_{1:T}$.

The framework consists of two main steps. First, an ML-based emulator of Noah-MP is constructed to approximate the mapping from model parameters and meteorological forcing data to simulated land surface variables at each study site. For this purpose, we perform 1,000 ensemble Noah-MP simulations using 1,000 different parameter combinations sampled from Latin Hypercube Sampling and site-specific climate forcing. For each site, 80% of these simulations are used to train the emulator, while the remaining 20% are reserved for independent evaluation of emulator performance.

Second, the trained emulator is used to calibrate model parameters at each study site. Observational data are divided into calibration and validation periods. Parameters are optimized using observations from the calibration period only, and the optimized parameter set is then applied in full simulations with the original Noah-MP model over the entire study period. Model performance is evaluated separately for the calibration and validation periods using standard metrics. Details of the two-step calibration workflow, including emulator construction, evaluation, and calibration procedures, are described below.

2.5.2 Emulator Construction and Evaluation

2.5.2.1 Emulator construction

We employ an Attention-Long Short-Term Memory (LSTM) neural network to construct the emulator, which is designed to capture the processes of the Noah-MP land surface model. LSTM networks are well suited for representing short- to medium-term temporal features (Kratzert et al., 2018), such as the memory of soil moisture states, while the attention mechanism further enhances generalization by dynamically weighting the importance of different historical time steps (e.g., antecedent precipitation events) when predicting current states.

The emulator learns a statistical approximation of Noah-MP behaviour, mapping model parameters and meteorological forcing data to simulated land surface states and fluxes. In this study, the emulator inputs consist of eight dynamic meteorological forcing variables, including 2m temperature (T2M), rain rate (RAINRATE), longwave radiation (LWFORC), shortwave radiation (SWFORC), 2m specific humidity (Q2D), surface pressure (PSFC), 2m U-wind component (U2D), and 2m V-wind component (V2D), consistent with the parent Noah-MP model. In addition, nine model parameters, including both vegetation and soil parameters targeted for calibration (Table 2), are also provided as inputs.

The emulator is trained to predict a comprehensive set of 29 output variables (Figures S3 and S4), including the three primary calibration targets, which are latent heat (LH), sensible heat (SH), and soil moisture (SM), as well as auxiliary diagnostic variables (e.g., ground heat flux, runoff components, and temperature states). Higher loss weights are assigned to the primary targets, while auxiliary variables act as process-aware constraints that respect the underlying conservation laws of energy and mass.

Training data for the emulator are generated by sampling 1,000 parameter combinations within prescribed bounds (Table 2) and running Noah-MP once for each parameter set, using identical site-specific forcing and model configuration. Each simulation produces time series of both primary target variables (LH, SH, and SM) and auxiliary variables (Figures S3 and S4). The resulting dataset comprises 1,000 paired samples of $(\theta, X_{1:T}) \rightarrow Y_{1:T}$, where θ denotes the parameter vector, $X_{1:T}$ the

climate forcing, and $Y_{1:T}$ the Noah-MP simulated outputs. Because observational data are separated into calibration and validation periods for subsequent parameter calibration (Section 2.5.3), emulator training uses the calibration-period forcing and the corresponding Noah-MP simulations. This dataset is subsequently split into training and evaluation subsets with an 80:20 ratio.

To prevent training instability caused by differing physical magnitudes, all input and output variables are standardized using Z-score normalization. The training loss aggregates weighted RMSE across all predicted variables, allowing the network to learn the coupled evolution of energy and water processes while accounting for differences in physical units. The loss function, L_{train} , is defined as:

$$L_{train} = \sum_{v \in V} w_v \times \sqrt{\frac{1}{N} \frac{1}{T} \sum_{n=1}^N \sum_{t=1}^T (\hat{x}_{v,n,t} - x_{v,n,t})^2} \quad \sum_{v \in V} w_v = 1, w_v > 0 \quad (1)$$

where V represents the set of all emulator-predicted variables (29 variables in this study), N is the batch size ($N = 16$ in this study), T is the total number of time steps, $\hat{x}_{v,n,t}$ and $x_{v,n,t}$ are normalized values of each variable from the emulator-predictions and Noah-MP-simulations, respectively, and w_v is variable-specific weighting factor. Both $\hat{x}_{v,n,t}$ and $x_{v,n,t}$ are normalized values to account for disparate physical magnitudes across variables (e.g., W/m² for heat fluxes versus m³/m³ for soil moisture) and to ensure numerical stability.

2.5.2.2 Emulator evaluation

Before using the machine-learning-based emulator for parameter calibration, its predictive performance is assessed using the evaluation subset (20% of ensemble Noah-MP simulations). Unlike the training loss (Equation 1), emulator skill is assessed using denormalized outputs in physical units for each primary target variable (LH, SH, and SM).

The metrics of evaluating emulator performance include the coefficient of determination (R^2 , Equation 2), root mean square error (RMSE, Equation 3), and percent bias (PBIAS, Equation 4). Only emulators demonstrating sufficient accuracy are used in subsequent calibration in Section 2.5.3, ensuring that parameter optimization is based on a reliable approximation of Noah-MP behaviour.

$$R^2 = \frac{1}{M} \cdot \sum_{m=1}^M \left(1 - \frac{\sum_{t=1}^T (y_{m,t} - \hat{y}_{m,t})^2}{\sum_{t=1}^T (y_{m,t} - \bar{y}_m)^2} \right), \quad \bar{y}_m = \frac{1}{T} \sum_{t=1}^T y_{m,t} \quad (2)$$

where y and $\hat{y}$ are original values of each variable from the emulator-predictions and Noah-MP-simulations, respectively, and M is the ensemble size used for emulator evaluation (20% of the ensemble Noah-MP simulations in this study).

$$RMSE = \frac{1}{M} \cdot \sum_{m=1}^M \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{y}_{m,t} - y_{m,t})^2} \quad (3)$$

$$PBIAS = \frac{1}{M} \cdot \sum_{m=1}^M 100 \cdot \frac{\sum_{t=1}^T (\hat{y}_{m,t} - y_{m,t})}{\sum_{t=1}^T (y_{m,t}) + \varepsilon} \quad (4)$$

2.5.3 Parameter calibration and validation

To ensure the robustness of the calibrated parameters and assess their transferability to unseen temporal conditions, we separate the observational data into calibration and validation periods, conduct the calibration using only the forcing data and observations from the calibration period, and reserve the remaining independent and unseen data for subsequent validation. For the Panama BCI site, the calibration period is 2015-07-30 to 2016-07-29 and the validation period is 2016-07-30 to 2017-07-30. For the Malaysia PSO site, the calibration period is 2005-01-01 to 2006-12-31 and the validation period is 2007-01-01 to 2009-12-30.

2.5.3.1 Parameter calibration

Following the emulator construction and evaluation, parameter calibration is performed using the Differential Evolution (DE) algorithm (Storn and Price, 1997). DE is a population-based optimization method well suited to highly non-linear and non-differentiable objectives typical of land-surface model calibration.

Starting from an initial population sampled within parameter bounds, DE iteratively proposes new candidates through mutation and crossover operations and retains those that improve the objective function. In our implementation, DE iteratively generates populations of parameter vectors. For each candidate vector, the emulator predicts the corresponding model outputs which are then evaluated against in situ observations using the objective function defined in Equation 7. This objective function is formulated to minimize discrepancies between emulator-predicted fluxes and the observed fluxes. This cycle repeats until convergence criteria are met or the maximum number of generations is reached.

Multiple independent DE runs with different random seeds are conducted to reduce sensitivity to stochastic effects. This yields an ensemble of calibrated parameter sets, from which we report the best-performing solution. This multi-run design also provides a practical diagnostic of identifiability: parameters that consistently converge to similar values across independent runs are more likely to be well constrained by the data and objective.

Calibration is formulated as a joint optimization problem to explicitly account for interactions and trade-offs among energy and water fluxes. For each primary variable, RMSE is first computed in physical units over all valid (non-missing) time steps (Equation 5) and then normalized by the observed dynamic range to obtain a dimensionless error (Equation 6). The overall calibration loss is defined as a weighted sum of normalized RMSE across the three primary target variables (LH, SH, and SM) (Equation 7).

$$RMSE_p = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{z}_{p,t} - z_{p,t})^2} \quad (5)$$

where p is the primary target variable (LH, SH, or SM), and z and $\hat{z}$ are flux tower observations and Noah-MP-simulations, respectively.

$$NRMSE_p = \frac{RMSE_p}{\max(z_p) - \min(z_p) + \varepsilon} \quad (6)$$

$$L_{calibration} = \sum_{p \in P} w_p \times NRMSE_p, \sum_{p \in P} w_p = 1, w_p > 0 \quad (7)$$

where P represents the set of primary target variables for calibration (LH, SH, and SM in this study), and w_p is the variable-specific weighting factor.

This $L_{calibration}$ metric aggregates the errors of different variables by normalizing them, thereby resolving the issue of disparate physical magnitudes. The weighting factor w_p allows for flexible prioritization of specific physical processes based on site-specific conditions or measurement reliability. For instance, if observed SH of the study site is suspected to be biased, lower weights can be assigned to SH to ensure the model prioritizes the more credible sources of guidance.

For parameters with prior knowledge, fixed values can be set manually before calibration to impose additional physical constraints. For example, at the PSO site, LAI is fixed at the field-observed value of 6.52 m²/m² rather than calibrated.

2.5.3.2 Parameter validation

Because the emulator is an approximation, the calibrated parameters are validated by applying them in simulations with the original Noah-MP model. The calibrated parameter set is therefore used to rerun Noah-MP over the complete study period.

Model performance is evaluated separately for the calibration and validation periods using standard metrics computed in physical units, including RMSE, R², and bias (BIAS; Equation 8). Performance during the calibration assesses whether the emulator-based calibration identifies parameterizations that translate to the physical model, while performance during the independent validation period evaluates generalization and robustness.

$$BIAS = \frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t) \quad (8)$$

2.6 Contribution analysis

To quantify the contribution of each calibrated parameter (or parameter group) to the overall improvement in model performance, we employ Shapley value decomposition, a game-theoretic approach originally developed for fair allocation problems (Shapley, 1953). This method provides a unique and theoretically fair way to distribute the total gain among the parameters.

Formally, let $u(A)$ represent the performance metric (RMSE in this study) achieved when a subset of parameters A is set to their calibrated values, while the remaining parameters are held at their default values. The Shapley value ϕ_i for a parameter i is defined as the weighted average of its contributions across all possible parameter combinations:

$$\phi_i = \sum_{A \subseteq Z \setminus \{i\}} \frac{|A|! (j - |A| - 1)!}{j!} [u(A \cup \{i\}) - u(A)] \quad (9)$$

where Z is the set of all calibrated parameters, and j is the number of parameter groups.

For n parameter groups, complete enumeration of all combinations requires 2^n model evaluations, corresponding to every possible combination of default and calibrated parameter states. In this study, two parameters (HVT and HVB) are correlated, with HVB constrained to be below HVT, reducing the original nine parameters to eight parameter groups. Accordingly, we conducted 256 model simulations. For each simulation, performance metrics (RMSE in this study) are calculated against observational data. Shapley values are then computed to decompose the total performance change into individual parameter contributions.

Finally, to quantify the absolute influences of each parameter, we define its importance as the absolute value of its Shapley value (Equation 10). This approach allows direct comparison of the relative impact of each parameter on the model performance improvement.

$$S_i = |\phi_i| \quad (10)$$

3. Results

We first analyse the emulator validation results due to its fundamental functionality of the calibration. For each study site, we examine the optimal parameter values and their individual contributions to model performance to identify the most influential parameters. We also examine how calibrating these parameters affects day-to-day variability in energy and water fluxes, as well as seasonal and diurnal cycles.

3.1 Emulator performance

Prior to calibration, the predictive accuracy of the emulator is rigorously evaluated using the independent validation subset, consisting of 20% of the ensemble Noah-MP simulations not used during training. The emulator demonstrates high fidelity in

reproducing the dynamics of the original Noah-MP model (Figure 3), achieving high agreement for all target variables (Figure S3-4). Specifically, R^2 reaches 0.95 for all the three primary target variables including LH, SH, and SM (Figure 3), while RMSE and PBIAS remain low, indicating that the emulator effectively captures both the magnitude and temporal variability of the Noah-MP simulations. These results demonstrate that the emulator possesses sufficient precision to capture the complex relationships between meteorological forcing, model parameters, and the resulting Noah-MP outputs, making it a reliable tool for subsequent model calibration.

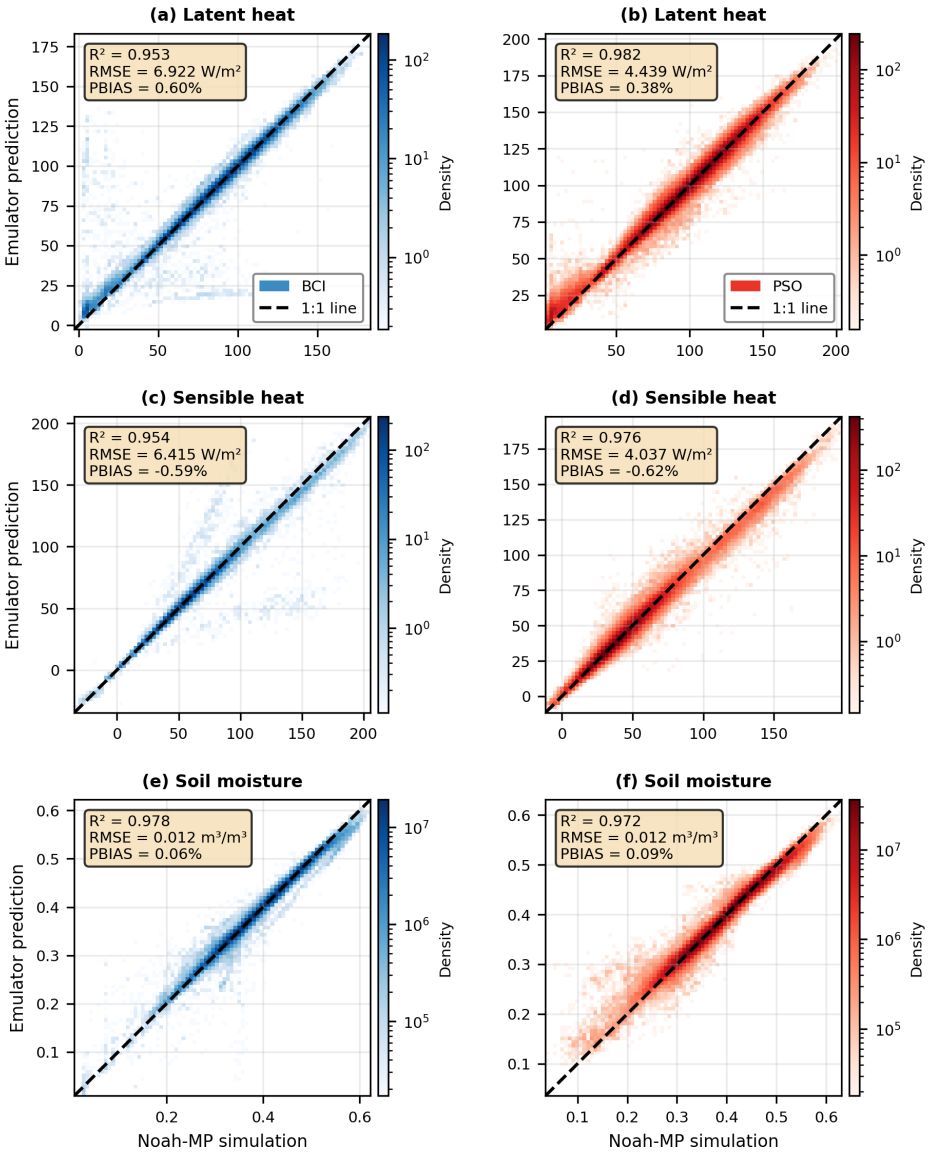


Figure 2. Evaluation of the trained Noah-MP emulator for the three primary target variables (a and b) latent heat, (c and d) sensible heat, and (e and f) soil moisture at the Panama BCI site (left column) and the Malaysia PSO site (right column).

335 **3.2 Model performance in simulating energy and water fluxes**

3.2.1 Panama tropical forest site

3.2.1.1 Calibration and contribution analysis results

Different model parameters exert distinct influences on energy and water cycling processes (Table 3). At the Panama BCI tropical forest site, we find that the most important parameter for LH and SH is the maximum rate of carboxylation at 25 degC (VCMX25) (Table 3). VCMX25 controls the photosynthesis rate and stomatal conductance, and thereby also plays a dominant role in regulating carbon assimilation and evapotranspiration. The calibrated VCMX25 value (114.6 $\mu\text{mol}/\text{CO}_2/\text{m}^2/\text{s}$) is consistent with values reported for Panama BCI tropical forests in previous studies (Cheng et al., 2024b).

While VCMX25 modulates both heat fluxes primarily through its influence on LH, with corresponding effects on SH via surface energy partitioning, MAXSMC and REFSMC also play a key role in regulating LH and SH, as they determine the soil water storage available for plants to uptake. In addition, Z0MVT directly regulates above-canopy SH flux, which induces a latent heat response by cooling the land surface.

Table 3. Default and calibrated parameter values, along with parameter importance rankings, at the Panama BCI tropical forest site.

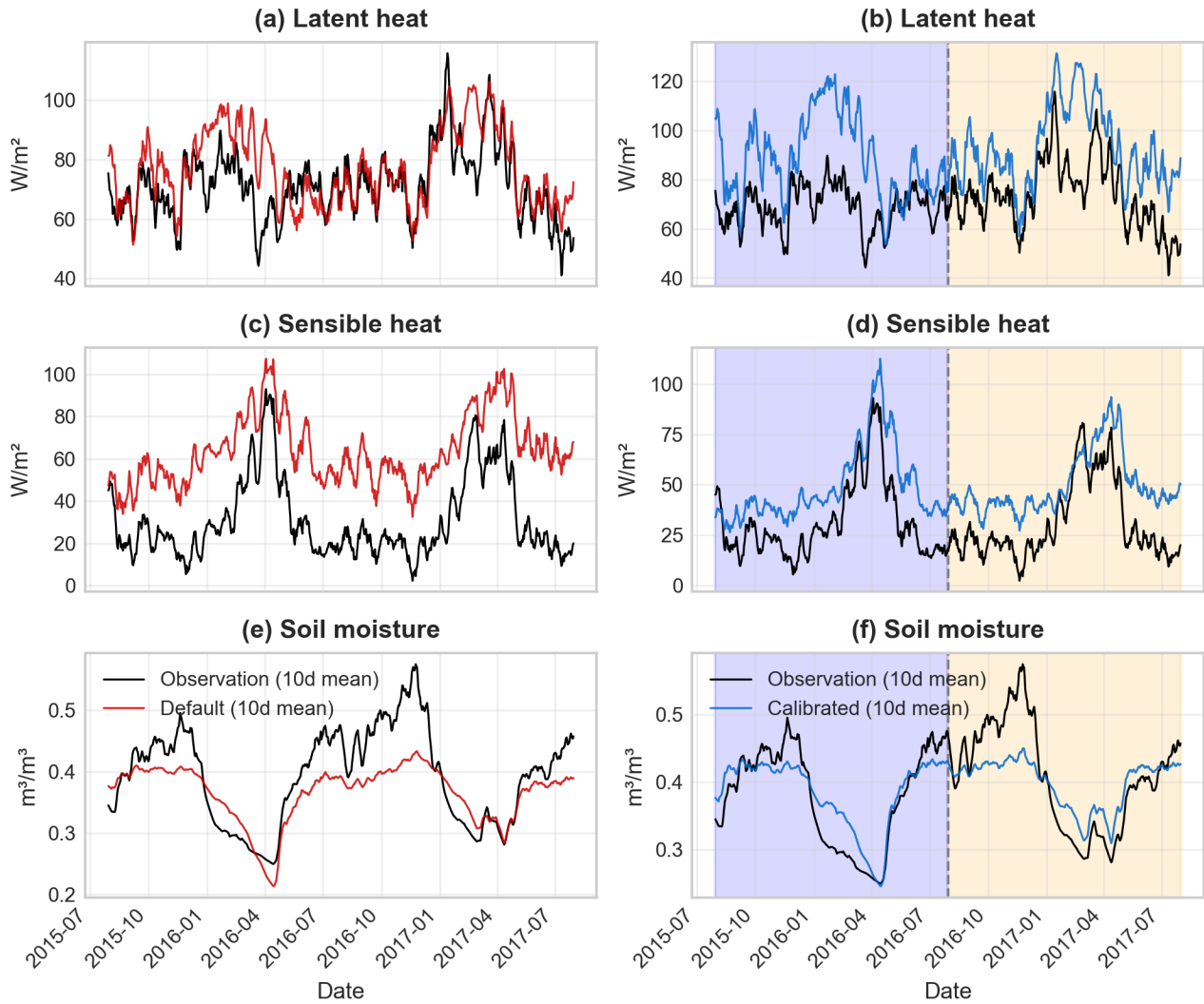
Parameter	Default value	Calibrated value	Rank of importance for LH	Rank of importance for SH
VCMX25 ($\mu\text{mol CO}_2/\text{m}^2/\text{s}$)	60	114.6	1	1
HVT (m)	20	35.5	8	7
HVB (m)	8	13.9	8	7
CWPVT (-)	0.67	4.04	7	8
Z0MVT (m)	1.1	0.35	4	2
WLTSMC (m^3/m^3)	0.138	0.207	3	4
REFSMC (m^3/m^3)	0.412	0.397	6	6
MAXSMC (m^3/m^3)	0.468	0.538	2	3
SATDK (m/s)	9.74×10^{-7}	1.66×10^{-5}	5	5

350 **3.2.1.2 Daily-scale performance**

The calibration substantially improves SH simulation at the daily scale during both the calibration and validation periods, demonstrating the effectiveness of the calibrated parameters (Figure 4, Table 4). Specifically, during the calibration period, the mean bias of SH decreases from 30.6 W/m^2 under the default parameterization to 17.4 W/m^2 under the calibrated parameterization. During the validation period, the mean bias of SH decreases from 34.9 W/m^2 to 19.0 W/m^2 . However, this

360

improvement comes with a trade-off in LH performance: over the full period, the mean bias of LH increases from 5.6 W/m² for the default parameterization to 19.7 W/m² for the calibrated parameterization, and R² for LH decreases from 0.80 to 0.65 (Table 4). Nevertheless, the gains in SH are substantial, with mean bias decreasing by 44% (from 32.7 W/m² to 18.3 W/m²) and R² increasing from 0.63 to 0.84 over the full period. In addition, the R² under the calibrated parameterization remains stable between the calibration (0.87) and validation (0.81) periods.



365

Figure 3. Observations (black line) and daily Noah-MP simulation results at the Panama BCI tropical forest site for (a-b) latent heat, (c-d) sensible heat, and (e-f) soil moisture, using default parameters (left column, red line) and calibrated parameters (right column, blue line). The data are separated into calibration (purple background) and validation (yellow background) periods in panels (b), (d), and (f). Both simulated and observed time series are plotted as 10-day running means for clarity.

Table 4. Performance metrics (RMSE, MAE, BIAS, and R²) of simulation results at the Panama BCI tropical forest site for latent heat (LH), sensible heat (SH), and soil moisture (SM), using default and calibrated parameters during calibration, validation, and full periods with the original Noah-MP model.

<u>Variable</u>	<u>Data period</u>	<u>Parameter set</u>	<u>RMSE</u>	<u>BIAS</u>	<u>R²</u>
<u>LH (W/m²)</u>	<u>Calibration</u>	<u>Default</u>	<u>44.0</u>	<u>7.3</u>	<u>0.79</u>
		<u>Calibrated</u>	<u>59.8</u>	<u>20.4</u>	<u>0.61</u>
	<u>Validation</u>	<u>Default</u>	<u>46.7</u>	<u>4.0</u>	<u>0.81</u>
		<u>Calibrated</u>	<u>59.7</u>	<u>19.1</u>	<u>0.69</u>
	<u>Full</u>	<u>Default</u>	<u>45.4</u>	<u>5.6</u>	<u>0.8</u>
		<u>Calibrated</u>	<u>59.7</u>	<u>19.7</u>	<u>0.65</u>
<u>SH (W/m²)</u>	<u>Calibration</u>	<u>Default</u>	<u>51.2</u>	<u>30.6</u>	<u>0.69</u>
		<u>Calibrated</u>	<u>33.5</u>	<u>17.4</u>	<u>0.87</u>
	<u>Validation</u>	<u>Default</u>	<u>56.9</u>	<u>34.9</u>	<u>0.57</u>
		<u>Calibrated</u>	<u>37.7</u>	<u>19.0</u>	<u>0.81</u>
	<u>Full</u>	<u>Default</u>	<u>54.1</u>	<u>32.7</u>	<u>0.63</u>
		<u>Calibrated</u>	<u>35.7</u>	<u>18.3</u>	<u>0.84</u>
<u>SM (m³/m³)</u>	<u>Calibration</u>	<u>Default</u>	<u>0.051</u>	<u>-0.020</u>	<u>0.61</u>
		<u>Calibrated</u>	<u>0.042</u>	<u>0.006</u>	<u>0.73</u>
	<u>Validation</u>	<u>Default</u>	<u>0.066</u>	<u>-0.035</u>	<u>0.40</u>
		<u>Calibrated</u>	<u>0.055</u>	<u>-0.008</u>	<u>0.59</u>
	<u>Full</u>	<u>Default</u>	<u>0.059</u>	<u>-0.028</u>	<u>0.51</u>
		<u>Calibrated</u>	<u>0.049</u>	<u>-0.001</u>	<u>0.67</u>

The calibrated parameters improve the simulation of daily soil moisture at the Panama tropical forest site during both the calibration and validation periods (Figure 3, Table 4). During the calibration period, the mean bias decreases from -0.020 m³/m³ for default parameterization to 0.006 m³/m³ for calibrated parameterization, while during the validation period it decreases from -0.035 m³/m³ to -0.008 m³/m³. Over the full period, the mean bias decreases from -0.028 m³/m³ to -0.001 m³/m³. RMSE decreases by 16.9%, from 0.059 m³/m³ to 0.049 m³/m³, with reductions of 17.6% during calibration and 16.7% during validation. R² values increase from 0.51 under the default parameterization to 0.67 under the calibrated parameterization, rising from 0.61 to 0.73 during calibration and from 0.40 to 0.59 during validation (Table 4).

3.2.1.3 Seasonal and diurnal-scale performance

Beyond the daily scale, Noah-MP effectively captures the seasonal and diurnal dynamics of surface energy and water fluxes (Figure 4). Specifically, it captures the seasonal decline in LH (Figure 5a) and soil moisture (Figure 5e), as well as the corresponding peak in SH around April (Figure 4a). However, the model does not fully capture the peak in soil moisture from September to November (Figure 4e), consistent with its deficiencies in simulating daily soil moisture peaks during the wet season (Figure 4f). The underestimated wet-season soil moisture may be caused by the deficiencies in subsurface water transport representation, such as the absence of preferential flow simulation in Noah-MP, which is known to govern seasonal sub-surface fluxes of water in tropical rainforests by rapidly transporting water downward (Cheng et al., 2017, 2018, 2019). Preferential flow paths generated by tree root penetration, termite activity, earthworm burrows, or other soil microporosity significantly alter the partitioning of water and energy, affecting not only runoff, but also water storage and plant water uptake in tropical forests. Therefore, their absence in Noah-MP could be an important source of structural error which would further hinder calibrations based on heat fluxes. This critical process has received insufficient attention in the context of land surface modelling and deserves future model development efforts.

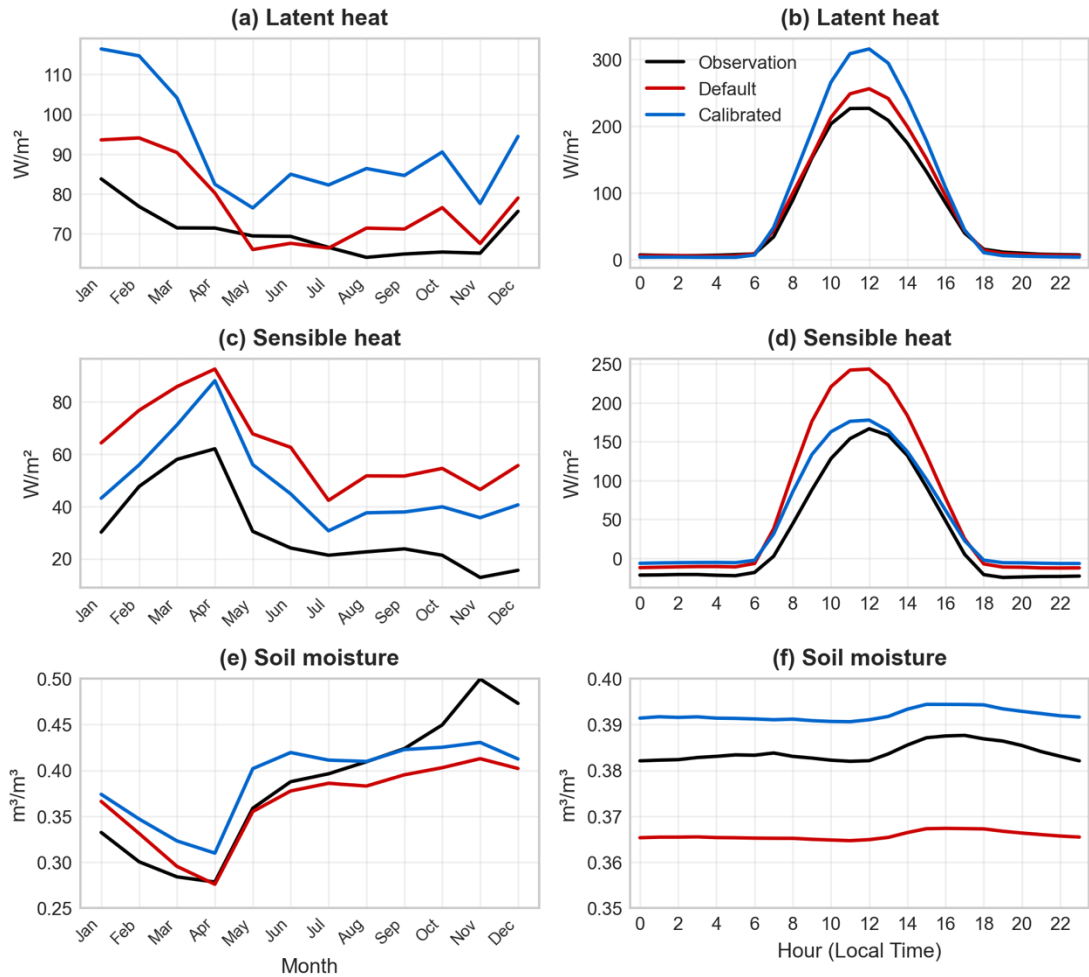


Figure 4. Observation (black line) and Noah-MP simulation results at the Panama BCI tropical forest site for (a-b) latent heat, (c-d) sensible heat, and (e-f) soil moisture, using default parameters (red line) and calibrated parameters (blue line) for seasonal (left column) and diurnal (right column) cycles.

395 The trade-off in optimizing LH and SH at the daily scale (Figure 4) also persists at the seasonal and diurnal scales (Figure 5). LH overestimation is most pronounced between January and April, as the dry season develops but simulated LH remains relatively high. We note from Figures 4e-f that the dry down of soil moisture at these times of year is poorly captured, with both the calibrated and uncalibrated models overestimating soil moisture, which may contribute to LH biases. Additional factors include potential biases in downward solar radiation measurements from the flux tower. If it is overestimated, this could contribute to the overall overestimation of both latent and sensible heat fluxes.

400 Another possible source of the discrepancy between calibrated LH and SH may be Noah-MP’s inability to simulate forest functional diversity, which are prevalent in tropical forests and critical for capturing forest response to environmental variability (Cheng et al., 2024b). Addressing this limitation may require the use of more advanced ecosystem demography models to better capture the interactions of multi-species coexistence in tropical ecosystems.

405 The model persistently overestimates nighttime SH (Figure 4d), corresponding to overestimated nighttime soil temperatures (Figure S5). This discrepancy likely stems from the absence of a soil organic layer parameterization in Noah-MP. Organic layers, with low thermal conductivity and high heat capacity, buffer both daytime heating and nocturnal cooling. Without this buffering effect, the model likely exaggerates diurnal soil temperature amplitudes, inflating both daytime peaks and nighttime minima. This structural limitation, recognized in previous evaluations (Zhang et al., 2021), contributes to systematic SH biases and highlights a priority area for future model development.

3.2.2 Malaysia tropical forest site

3.2.2.1 Contribution analysis results

415 The single most important parameters for calibrating LH and SH at the Malaysia PSO tropical forest site are VCMX25 and WLTSMC, due to their roles in controlling photosynthesis and soil water availability (Table 5). Other soil parameters, including WLTSMC and SATDK, also strongly influence Noah-MP simulations of energy and water fluxes by regulating the amount of soil moisture available for plant uptake. LAI is prescribed based on site measurements rather than calibrated, but it is also important for both LH and SH due to its role in controlling vegetation phenology.

420 **Table 5.** Default and calibrated parameter values, along with parameter importance rankings, at the Malaysia PSO tropical forest site.

<u>Parameter</u>	<u>Default value</u>	<u>Calibrated value</u>	<u>Rank of importance for LH</u>	<u>Rank of importance for SH</u>

<u>VCMX25 (umol CO₂/m²/s)</u>	<u>60</u>	<u>88.9</u>	<u>2</u>	<u>1</u>
<u>HVT (m)</u>	<u>20</u>	<u>35.7</u>	<u>8</u>	<u>8</u>
<u>HVB (m)</u>	<u>8</u>	<u>18.0</u>	<u>8</u>	<u>8</u>
<u>CWPVT (-)</u>	<u>0.67</u>	<u>0.35</u>	<u>7</u>	<u>7</u>
<u>Z0MVT (m)</u>	<u>1.1</u>	<u>1.16</u>	<u>9</u>	<u>9</u>
<u>LAI (m²/m²)</u>	<u>4.5</u>	<u>6.52^a</u>	<u>3</u>	<u>4</u>
<u>WLTSMC (m³/m³)</u>	<u>0.126</u>	<u>0.2580</u>	<u>1</u>	<u>2</u>
<u>REFSMC (m³/m³)</u>	<u>0.404</u>	<u>0.4456</u>	<u>5</u>	<u>3</u>
<u>MAXSMC (m³/m³)</u>	<u>0.468</u>	<u>0.4861</u>	<u>6</u>	<u>6</u>
<u>SATDK (m/s)</u>	<u>1.34 × 10⁻⁶</u>	<u>2.0 × 10⁻⁶</u>	<u>4</u>	<u>5</u>

a. The modified LAI value of 6.52 m²/m² is adopted from field observation at the PSO site (Kosugi et al., 2012) and is prescribed prior to the calibration process rather than being calibrated.

3.2.2.2 Daily-scale performance

425 The calibrated parameters substantially improve the simulation of daily energy and water fluxes at the Malaysia PSO tropical forest site, during both the calibration and validation periods (Figure 5, Table 6). Unlike the Panama BCI site, where
improvements in latent and sensible heat exhibit a trade-off, calibration at the PSO site leads to simultaneous improvements
in both fluxes. During the calibration period, the mean bias of LH decreases from -14.2 W/m² under the default
parameterization to -8.4 W/m² after calibration, while mean bias of SH decreases from -7.6 W/m² to -3.3 W/m². During the
430 validation period, mean LH bias decreases from -10.0 W/m² to -2.3 W/m², and mean SH bias decreases from 5.8 W/m² to -0.3
W/m². Over the full period, the mean bias of LH decreases from -11.7 W/m² to -4.7 W/m², and mean SH bias decreases from
6.5 W/m² to 1.2 W/m².

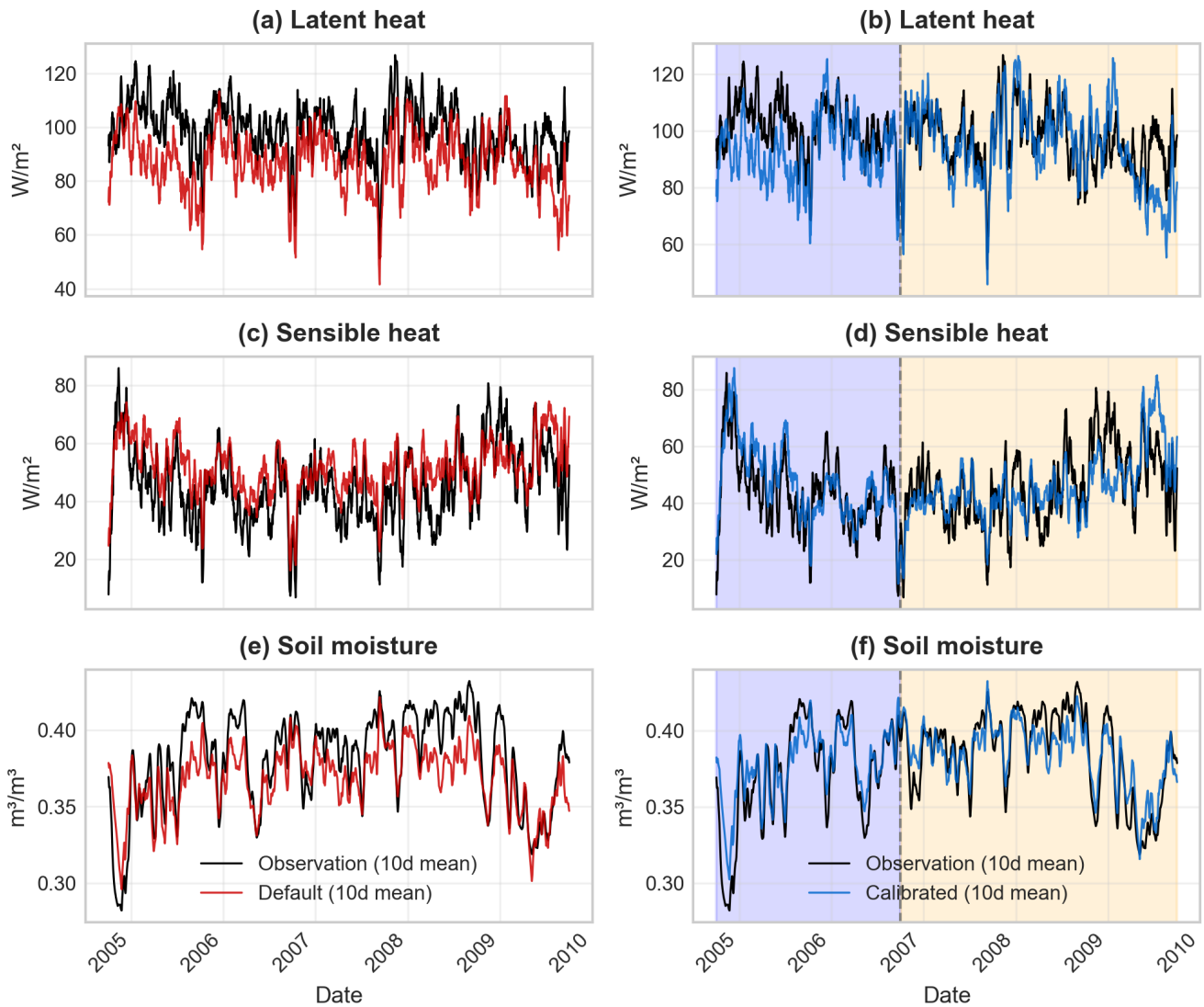


Figure 5. Observations (black line) and daily Noah-MP simulation results at the Malaysia PSO tropical forest site for (a-b) latent heat, (c-d) sensible heat, and (e-f) soil moisture, using default parameters (left column, red line) and calibrated parameters (right column, blue line). The data are separated into calibration (purple background) and validation (yellow background) periods in panels (b), (d), and (f). Both simulated and observed time series are plotted as 10-day running means for clarity.

Table 6. Performance metrics (RMSE, MAE, BIAS, and R^2) of simulation results at the Malaysia PSO tropical forest site for latent heat (LH), sensible heat (SH), and soil moisture (SM), using default parameters and calibrated parameters during calibration, validation, and full periods with the original Noah-MP model.

<u>Variable</u>	<u>Data period</u>	<u>Parameter set</u>	<u>RMSE</u>	<u>BIAS</u>	<u>R^2</u>
<u>LH (W/m²)</u>	<u>Calibration</u>	<u>Default</u>	<u>52.1</u>	<u>-14.2</u>	<u>0.87</u>
		<u>Calibrated</u>	<u>48.7</u>	<u>-8.4</u>	<u>0.89</u>
	<u>Validation</u>	<u>Default</u>	<u>53.3</u>	<u>-10.0</u>	<u>0.87</u>
		<u>Calibrated</u>	<u>50.7</u>	<u>-2.3</u>	<u>0.88</u>
	<u>Full</u>	<u>Default</u>	<u>52.8</u>	<u>-11.7</u>	<u>0.87</u>
		<u>Calibrated</u>	<u>49.9</u>	<u>-4.7</u>	<u>0.89</u>
<u>SH (W/m²)</u>	<u>Calibration</u>	<u>Default</u>	<u>46.2</u>	<u>7.6</u>	<u>0.82</u>
		<u>Calibrated</u>	<u>48.2</u>	<u>3.3</u>	<u>0.80</u>
	<u>Validation</u>	<u>Default</u>	<u>46.0</u>	<u>5.8</u>	<u>0.81</u>
		<u>Calibrated</u>	<u>49.6</u>	<u>-0.3</u>	<u>0.78</u>
	<u>Full</u>	<u>Default</u>	<u>46.1</u>	<u>6.5</u>	<u>0.82</u>
		<u>Calibrated</u>	<u>49.0</u>	<u>1.2</u>	<u>0.79</u>
<u>SM (m³/m³)</u>	<u>Calibration</u>	<u>Default</u>	<u>0.022</u>	<u>-0.009</u>	<u>0.62</u>
		<u>Calibrated</u>	<u>0.019</u>	<u>0.005</u>	<u>0.73</u>
	<u>Validation</u>	<u>Default</u>	<u>0.021</u>	<u>-0.014</u>	<u>0.51</u>
		<u>Calibrated</u>	<u>0.015</u>	<u>-0.001</u>	<u>0.74</u>
	<u>Full</u>	<u>Default</u>	<u>0.021</u>	<u>-0.012</u>	<u>0.58</u>
		<u>Calibrated</u>	<u>0.017</u>	<u>0.002</u>	<u>0.75</u>

For soil moisture, the emulator-based calibration significantly outperforms the default configuration, particularly during the validation period. The default configuration exhibits reduced stability, with R^2 decreasing from 0.62 during the calibration period to 0.51 during validation, while the mean bias increases from -0.009 m³/m³ to -0.014 m³/m³. In contrast, simulations using calibrated parameter values perform better during the calibration period, achieving a lower RMSE (0.019 m³/m³ compared to 0.022 m³/m³ for default parametrization) and a higher R^2 (0.73) than default parameterization (0.62). During the validation period, the calibrated configuration not only maintains its predictive skill but further improves, yielding the lowest RMSE of 0.015 m³/m³ and a stable R^2 of 0.74. Overall, calibration attempts can be considered more successful for PSO in Malaysia than for BCI in Panama, possibly because the weaker seasonality at the Malaysia site means that the models do not need to capture distinct wet-dry transitions.

3.2.2.3 Seasonal and diurnal-scale performance

In addition to daily dynamics, Noah-MP adequately captures both seasonal and diurnal dynamics of energy and water fluxes at the PSO site (Figure 6). At the seasonal scale, both model configurations capture the decline in LH from April to July and from September to December (Figures 6a and 7b). Calibration greatly improves the simulation of LH, by removing a substantial underestimate that is present for all months with the default parameters (Figure 7a). SH is similarly improved, except between February and April, when the decrease in SH due to calibration leads to an underestimation; however, the seasonal dynamics from May to December are well captured and substantially improved relative to the default parameterization. Soil moisture shows weaker seasonal variability than at the Panama BCI site and is reasonably well captured by both the default and calibrated configurations.

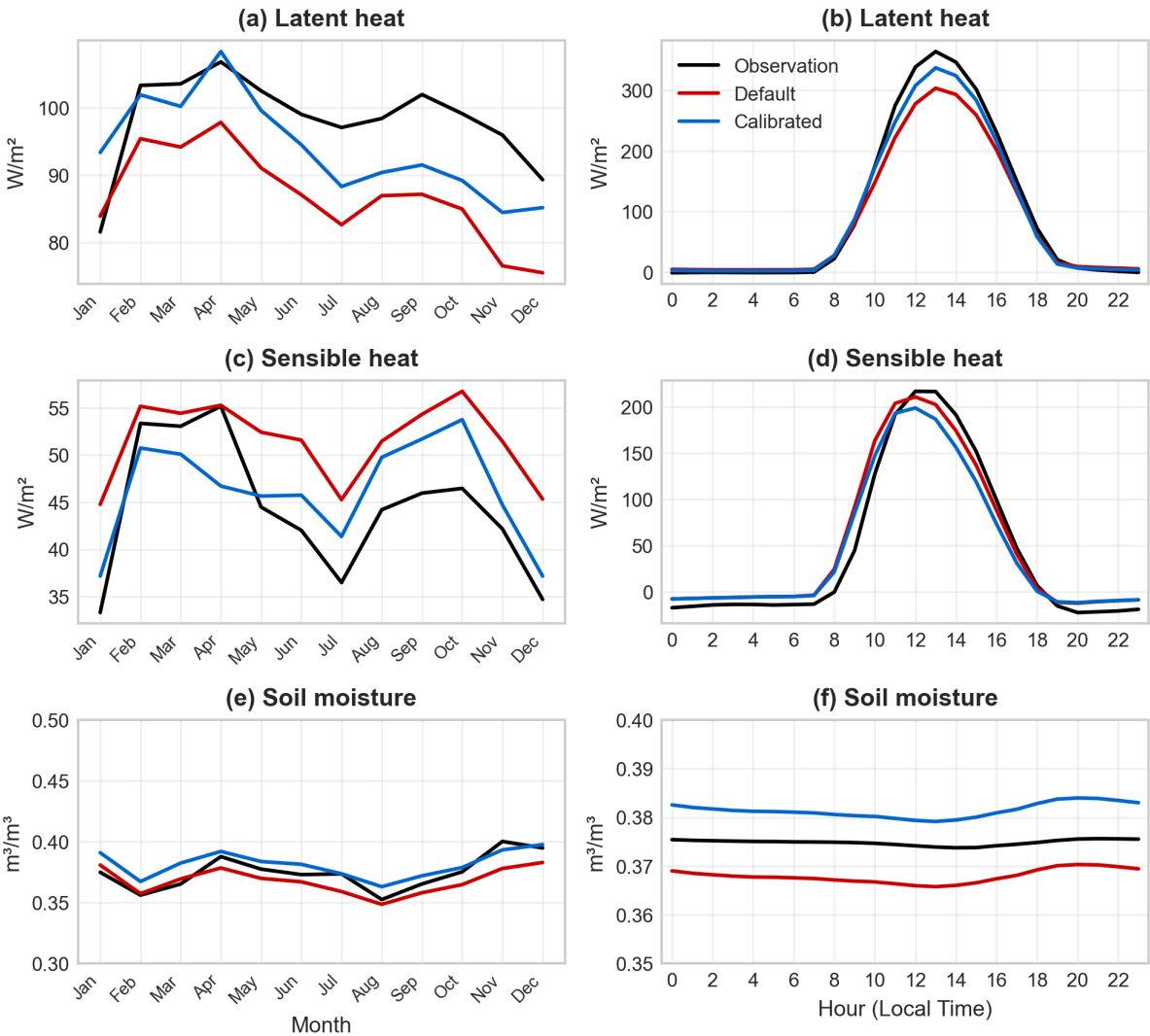


Figure 6. Observation (black line) and Noah-MP simulation results at the PSO tropical forest site for (a-b) latent heat, (c-d) sensible heat, and (e-f) soil moisture, using default parameters (red line) and calibrated parameters (blue line) for seasonal (left column) and diurnal (right column) cycles.

At the diurnal scale, Noah-MP with calibrated parameterization better captures the peaks in LH (Figure 6b) around noon time compared to default parameterization. These improvements indicate that calibration at the daily timescale translates into enhanced model performance at sub-daily scales. Interestingly, for SH, the peak of the diurnal cycle is underestimated after calibration, even though the overall performance on daily and seasonal timescales is improved. This stems from the period from February to April, when the calibrated SH values are further reduced relative to the default, resulting in greater underestimation compared to observations. The variation in model performance in capturing the diurnal cycle across seasons clearly reveals this contrast. Specifically, from February to April, calibration increases the SH mean bias from $+1.1 \text{ W/m}^2$ (default) to -6.6 W/m^2 . While in the remaining months, calibration reduces the mean bias from 8.2 W/m^2 (default) to 0.02 W/m^2 .

The notable success of co-calibration of both heat fluxes at PSO in Malaysia, compared to the conflicting results for LH and SH at BCI in Panama, is perhaps due to the large differences in climate and vegetation conditions between the two locations. Specifically, Panama has a tropical monsoon climate (Am), characterized by distinct dry and wet seasons, while Malaysia experiences a tropical rainforest climate (Af), with consistently high rainfall throughout the year. Furthermore, despite receiving less rainfall, the PSO site exhibits higher LH than the BCI site (Table 1), suggesting that structural vegetation differences, such as LAI (represented at PSO by the sensitivity of the model to LAI) and canopy wind effects, are playing a significant role in energy fluxes. For example, the positive LH biases at PSO with default parameters may be partially a result of underestimated damping of in-canopy winds by the relatively denser canopy, in which case calibrating LAI and CWPVT can be expected to improve both LH and SH. As a result, the parameters are not transferable between the two sites.

4. Additional Discussions

4.1 Transferability of the emulator-based calibration framework

The generalization and transferability of the calibration framework, including the emulator and optimization strategy, are critical for practical and broader applications. The emulator-based calibration framework developed in this study is inherently flexible and can be applied to any site with available observations. While the current implementation requires constructing a new emulator and performing calibration separately for each site, this workflow remains computationally efficient and scalable compared to traditional calibration approaches. A promising direction for future development is to incorporate static geographical attributes (e.g., elevation, latitude, longitude, soil properties, or vegetation type and properties) as explicit inputs to the emulator. Conditioning the emulator on such spatially specific factors would allow it to better represent spatial

heterogeneity and potentially support emulation across multiple locations, reducing the need to retrain site-specific emulators from scratch.

4.2 Parameter transferability across different climate regimes

Parameter transferability is critical for applying Noah-MP across large tropical domains, particularly in regions where site-specific calibration data are unavailable. In this study, we focus on two representative tropical forest sites: the Panama BCI site, which has a tropical monsoon climate (Am), and the Malaysia PSO site, which has a tropical rainforest climate (Af). Results show that parameters calibrated at the Panama BCI site cannot be directly applied to the Malaysia PSO site. This lack of parameter transferability between the two sites suggests that, although the current plant functional type (PFT) framework in Noah-MP captures first-order ecosystem differences, it may be insufficient to represent sub-PFT ecohydrological and physiological heterogeneity in tropical forests, which is critical for accurately simulating soil moisture and surface energy fluxes. Nevertheless, the site-specific calibrated parameter sets perform consistently well during independent validation periods at their respective sites. These findings indicate that, while calibrated parameters are not transferable across sites with different tropical climate regimes, they are potentially transferable among tropical forest sites with similar climate conditions. This finding highlights the potential value of climate regime-based calibration strategies in data-scarce tropical environments. In addition, we identified a consistent set of influential parameters across both sites, particularly those controlling photosynthesis and soil water storage, which could be prioritized for calibration in similar tropical settings.

Beyond site-level calibration, regional calibration offers a practical alternative for parameter estimation across large tropical domains. Instead of relying on parameter sets calibrated at individual locations, which often lack transferability, regional calibration can estimate parameters using multi-site or spatially distributed constraints to estimate parameters more robustly and achieve better performance across the region. The emulator-based framework presented in this study provides a practical foundation for such regional calibration efforts. Where large-scale in situ measurements are scarce, remote sensing products and reanalysis datasets can be leveraged to support regional calibration efforts.

Future work could extend calibration efforts to broader tropical ecosystems, such as peatland forests and mangroves, which have distinct carbon and water cycles. Incorporating these ecosystems into regional calibration frameworks could further improve the accuracy of regional weather forecasts and the reliability of regional climate projections across tropical areas. Such regional calibration efforts would also enhance understanding of land surface processes under the complex tropical climatic and environmental conditions and provide a foundation for applications such as evaluating the climate impacts of land-use change.

4.3 Future model development directions

4.3.1 Inclusion of missing vegetation and soil processes into Noah-MP

525 Noah-MP's inability to simulate the coexistence of multiple tropical forest species may contribute to discrepancies between calibrated LH and SH. Tropical forests are characterized by high biodiversity, including species with varying ages, sizes, and functional traits. These factors are essential for accurately representing energy, water, and carbon cycling as well as ecosystem functions (Cheng et al., 2024b). Incorporating more advanced ecosystem demography models into Noah-MP would allow for better simulation of these ecosystem dynamics. Additionally, the absence of an organic soil layer can contribute to the model's
530 overestimation of nighttime soil temperatures, highlighting another important area for future development.

4.3.2 Inclusion of missing hydrological processes into Noah-MP

The absence of preferential flow pathways, which are formed by tree roots and earthworm burrows, is another likely source of structural error in Noah-MP. These pathways are prevalent in tropical regions and play a crucial role in rainfall partitioning and hydrological responses (Cheng et al., 2017; 2018). Their exclusion may undermine model calibration efforts for surface
535 heat fluxes. Despite their importance, these processes remain underrepresented in land surface models and should be prioritized in future model improvement efforts. Macropores can be characterized by their size and number, and their distributions are commonly described using normal or log-normal functions (Edwards et al., 1988; Munyankusi et al., 1994; Cheng et al., 2017, 2018). The volumetric flow rate within an individual macropore can be calculated using Poiseuille's law for pipe flow (Sutera and Skalak, 1993; Cheng et al., 2017; 2018). This information provides a basis for future model development efforts to
540 incorporate preferential flow pathways into Noah-MP.

4.4 Model benchmarking and comparative evaluation

This study applies only the default Noah-MP physics option set. Different combinations of physics options may yield different results, which warrants future investigation. Comparative studies could evaluate Noah-MP's performance against other land
545 surface models, such as the Joint UK Land Environment Simulator (JULES) and the Community Terrestrial System Model (CTSM). These comparisons would provide valuable insights into how different land surface schemes influence model performance in tropical regions.

5 Conclusions

This study serves as a foundational step toward enhancing Noah-MP's performance in tropical forests by establishing a
550 machine learning-based calibration framework to calibrate the model using data from two tropical forest sites in Malaysia (tropical rainforest climate) and Panama (tropical monsoon climate). This study demonstrates the applicability of the calibration framework, showing that site-specific calibration using this framework significantly improves Noah-MP's ability

to simulate land surface processes in tropical forests relative to its default parameterization, particularly for latent and sensible heat fluxes and soil moisture at daily and seasonal timescales.

Calibrated parameter sets are not directly transferable across the two study sites, even within similar biome, due to differences in canopy structure, land cover composition, and seasonal behaviour of water and energy fluxes. Nevertheless, the site-specific calibrated parameter sets perform consistently well during independent validation periods at their respective sites. These results suggest that, although calibrated parameters are not directly transferable across sites with different tropical climate regimes, they may be transferable among sites with similar climate conditions. Contribution analyses highlight a consistent set of key vegetation and soil parameters in tropical forests, such as VCMX25, WLTSMC, and SATDK, that often dominate model performance and could be prioritized for tuning in similar settings. Moreover, the emulator-based calibration framework could be extended to support regional-scale calibration efforts.

Despite improvements, persistent model biases reveal structural limitations that cannot always be resolved through calibration alone. Noah-MP fails to adequately capture wet-season soil moisture peaks and latent heat variability in Panama tropical forests under a tropical monsoon climate, likely due to the deficiencies in subsurface water transport representation and the lack of multi-species vegetation representation. Persistent overestimation of nighttime sensible heat further underscores structural deficiencies, such as the omission of organic soil layers that buffer diurnal temperature variations. Furthermore, efforts to improve one energy flux often degrade another, reflecting energy partitioning trade-offs and internal coupling limitations, particularly under strong seasonal forcing, as seen in Panama. These findings underscore the importance of regionally adaptive calibration strategies, while also pointing to the need for structural model development in tropical hydrology and vegetation to ensure reliable land-atmosphere simulations in data-scarce, climate-sensitive tropical regions.

Acknowledgments: This research is partially supported by the Ministry of Education, Singapore, under its Academic Research Fund Tier 1, RS43/25, and by the Start-Up grant in NTU. We thank Hailin Yan (now at Bureau of Meteorology, Australia) and Jerry Liu (Centre for Climate Research Singapore) for technical support. The collection of the PSO data used in this study was supported by the JSPS KAKENHI (Grant number 24K01813).

Code and data availability statement: The current version of Noah-MP model is available from <https://github.com/NCAR/hrldas>. The exact version of the model used to produce the results used in this paper is archived on repository under 10.5281/zenodo.16780672 (Cheng 2025a), The flux tower data at the Malaysia PSO site is available at: https://asiaflux.net/index.php?page_id=93. The flux tower data at the Singapore site is available at: <https://doi.org/10.5281/zenodo.7104984> (Lipson et al., 2022). The data produced by the simulations in this study is available from <https://zenodo.org/records/16022303> (Cheng 2025b).

Competing interests: C.H. is member of the editorial board of Geoscientific Model Development.

Author contribution: YC, YW, KF, CH, FC, and AZ contributed to conceptualization, methodology, and investigation. YC and YW led the simulation, investigation, data curation, and visualization. MD, YM, BP, YK, ML, SN, ST, LM, and BZ provided resources. YC led the writing of this manuscript with contributions from all authors. All authors read and approved the final manuscript.

References

Arsenault, K. R., Nearing, G. S., Wang, S., Yatheendradas, S., and Peters-Lidard, C. D. (2018). Parameter sensitivity of the Noah-MP land surface model with dynamic vegetation. *Journal of Hydrometeorology*, 19(5), 815–830. <https://doi.org/10.1175/JHM-D-17-0205.1>

Bonan. (1995). Land-Atmosphere interactions for climate system Models: coupling biophysical, biogeochemical, and ecosystem dynamical processes. *Remote Sensing of Environment*, 51(1), 57–73. [https://doi.org/10.1016/0034-4257\(94\)00065-U](https://doi.org/10.1016/0034-4257(94)00065-U)

Bonan, and Doney, S. (2018). Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models. *Science*, 359(6375). <https://doi.org/10.1126/science.aam8328>

Chang, M., Liao, W., Wang, X., Zhang, Q., Chen, W., Wu, Z., and Hu, Z. (2020). An optimal ensemble of the Noah-MP land surface model for simulating surface heat fluxes over a typical subtropical forest in South China. *Agricultural and Forest Meteorology*, 281(November 2019), 107815. <https://doi.org/10.1016/j.agrformet.2019.107815>

Chen, D. X., Ban, X. Q., Li, Y. De, Xiao, W. F., Luo, T. S., Lin, M. X., and Xu, H. (2008). Responses of gas exchange to neighborhood interference in leaves of teak (*Tectona grandis* L. f.) in a tropical plantation forest. *Shengtai Xuebao/ Acta Ecologica Sinica*, 28(9), 4059–4069. [https://doi.org/10.1016/s1872-2032\(08\)60078-5](https://doi.org/10.1016/s1872-2032(08)60078-5)

Cheng, Ogden, F., and Zhu, J. (2017). Earthworms and tree roots: A model study of the effect of preferential flow paths on runoff generation and groundwater recharge in steep, saprolitic, tropical lowland catchments. *Water Resources Research*, 53(7), 5400–5419. <https://doi.org/10.1002/2016WR020258>

Cheng, Y, Huang, M., Chen, M., Guan, K., Bernacchi, C., Peng, B., and Tan, Z. (2020). Parameterizing Perennial Bioenergy Crops in Version 5 of the Community Land Model Based on Site-Level Observations in the Central Midwestern United States. *Journal of Advances in Modeling Earth Systems*, 12(1), 1–24. <https://doi.org/10.1029/2019MS001719>

Cheng, Y, Leung, L. R., Huang, M., Koven, C., Detto, M., Knox, R., et al. (2022). Modeling the joint effects of vegetation characteristics and soil properties on ecosystem dynamics in a Panama tropical forest. *Journal of Advances in Modeling Earth Systems*. <https://doi.org/10.1029/2021ms002603>

Cheng, Y., Ogden, F. L., Zhu, J., and Bretfeld, M. (2018). Land use dependent preferential flow paths affect hydrological response of steep tropical lowland catchments with saprolitic soils. *Water Resources Research*.

<https://doi.org/10.1029/2017WR021875>

- Cheng, Y., Ogden, F., and Zhu, J. (2019). Characterization of sudden and sustained base flow jump hydrologic behaviour in the humid seasonal tropics of the Panama Canal Watershed. *Hydrological Processes*, (April), 1–14. <https://doi.org/10.1002/hyp.13604>
- Cheng, Y., Huang, M., Lawrence, D. M., Calvin, K., Lombardozzi, D. L., Sinha, E., et al. (2022). Future bioenergy expansion could alter carbon sequestration potential and exacerbate water stress in the United States. *Science Advances*, 8(18), 1–14. <https://doi.org/DOI: 10.1126/sciadv.abm8237>
- Cheng, Y., Xia, W., Detto, M., and Shoemaker, C. A. (2023). A framework to calibrate ecosystem demography models within Earth system models using parallel surrogate global optimization. *Water Resources Research*. <https://doi.org/10.1029/2022WR032945>
- Cheng, Y., Lawrence, D. M., Pan, M., Zhang, B., Graham, N. T., Lawrence, P. J., et al. (2024a). A bioenergy-focused versus a reforestation-focused mitigation pathway yields disparate carbon storage and climate responses. *Proceedings of the National Academy of Sciences*, 121(7), 1–11. <https://doi.org/10.1073/pnas.2306775121>
- Cheng, Y., Wang, W., Detto, M., Fisher, R., and Shoemaker, C. (2024b). Calibrating Tropical Forest Coexistence in Ecosystem Demography Models Using Multi-Objective Optimization Through Population-Based Parallel Surrogate Search. *Journal of Advances in Modeling Earth Systems*, 16(8), 1–18. <https://doi.org/10.1029/2023MS004195>
- Cheng, Y., (2025a). Noah-MP.Cheng. Zenodo. <https://zenodo.org/records/16780672>
- Cheng, Y., (2025b). Simulation results for “Assessing and enhancing Noah-MP land surface modeling over tropical environments”. Zenodo. <https://zenodo.org/records/16022303>
- Cheruy, F., Dufresne, J. L., Hourdin, F., and Ducharne, A. (2014). Role of clouds and land-atmosphere coupling in midlatitude continental summer warm biases and climate change amplification in CMIP5 simulations. *Geophysical Research Letters*, 41(18), 6493–6500. <https://doi.org/10.1002/2014GL061145>
- Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, P. R., and Smith, J. (1999). The impact of new land surface physics on the GCM simulation of climate and climate sensitivity. *Climate Dynamics*, 15(3), 183–203. <https://doi.org/10.1007/s003820050276>
- Cox, P. M., Betts, R. A., Jones, C. D., and Spall, S. A. (2000). Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, 408(November), 184–187.
- Crossley, J. F., Polcher, J., Cox, P. M., Gedney, N., and Planton, S. (2000). Uncertainties linked to land-surface processes in climate change simulations. *Climate Dynamics*, 16(12), 949–961. <https://doi.org/10.1007/s003820000092>
- Detto, M., and Pacala, S. W. (2022). Plant hydraulics, stomatal control, and the response of a tropical forest to water stress over multiple temporal scales. *Global Change Biology*, 28(14), 4359–4376. <https://doi.org/10.1111/gcb.16179>
- Detto, M., Asner, G. P., Muller-Landau, H. C., and Oliver Sonnentag. (2015). Spatial variability in tropical forest leaf area density from multireturn lidar and modeling. *Journal of Geophysical Research: Biogeosciences*, 120, 965–978. <https://doi.org/10.1002/2014JG002774>.Received

- Duan, Q., Sorooshian, S., & Gupta, V. K. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28(4), 1015-1031.
- Dickinson, R. E. (1983). Land surface processes and climate—surface albedos and energy balance. *Advances in Geophysics*, 25(C), 305–353. [https://doi.org/10.1016/S0065-2687\(08\)60176-4](https://doi.org/10.1016/S0065-2687(08)60176-4)
- Domingues, T. F., Berry, J. A., Martinelli, L. A., Ometto, J. P. H. B., and Ehleringer, J. R. (2005). Parameterization of canopy structure and leaf-level gas exchange for an Eastern Amazonian tropical rain forest (Tapajós national forest, Pará, Brazil). *Earth Interactions*, 9(17). <https://doi.org/10.1175/EI149.1>
- Fang, Y., Leung, L. R., Koven, C. D., Bisht, G., Detto, M., Cheng, Y., et al. (2022). Modeling the topographic influence on aboveground biomass using a coupled model of hillslope hydrology and ecosystem dynamics. *Geoscientific Model Development*, 15(20), 7879–7901. <https://doi.org/10.5194/gmd-15-7879-2022>
- Fisher, R., and Koven, C. (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). <https://doi.org/10.1029/2018ms001453>
- Gan, Y., Liang, X. Z., Duan, Q., Chen, F., Li, J., & Zhang, Y. (2019). Assessment and reduction of the physical parameterization uncertainty for Noah-MP land surface model. *Water Resources Research*, 55(7), 5518-5538.
- Gentine, P., Massmann, A., Lintner, B. R., Alemohammad, S. H., Fu, R., Green, J. K., et al. (2019). Land–atmosphere interactions in the tropics – a review, 4171–4197.
- Goudriaan, J. (1977). Crop micrometeorology: a simulation study. *Wageningen University and Research*.
- Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1999). Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *Journal of Hydrologic Engineering*, 4(2), 135-143.
- Hasler, N., Williams, C. A., Denney, V. C., Ellis, P. W., Shrestha, S., Hart, D. E. T., et al. (2024). Accounting for albedo change to identify climate-positive tree cover restoration. *Nature Communications*, 1–11. <https://doi.org/10.1038/s41467-024-46577-1>
- He, C., Valayamkunnath, P., Barlage, M., Chen, F., Gochis, D., Cabell, R., et al. (2023a). Modernizing the open-source community Noah with multi-parameterization options (Noah-MP) land surface model (version 5.0) with enhanced modularity, interoperability, and applicability. *Geoscientific Model Development*, 16(17), 5131–5151. <https://doi.org/10.5194/gmd-16-5131-2023>
- He, C., Chen, F., Barlage, M., Yang, Z. L., Wegiel, J. W., Niu, G. Y., et al. (2023b). Enhancing the Community Noah-MP Land Model Capabilities for Earth Sciences and Applications. *Bulletin of the American Meteorological Society*, 104(11), E2023–E2029. <https://doi.org/10.1175/BAMS-D-23-0249.1>
- He, C., Valayamkunnath, P., Barlage, M., Chen, F., Yang, Z., Niyogi, D., and Michael, E. (2023c). The Community Noah-MP Land Surface Modeling System Technical Description Version 5.0, 284. <https://doi.org/https://doi.org/10.5065/ew8g-yr95>
- Huang, M., Ray, J., Hou, Z., Ren, H., Liu, Y., and Swiler, L. (2016). On the applicability of surrogate-based Markov chain

Monte Carlo-Bayesian inversion to the Community Land Model: Case studies at flux tower sites. *Journal of Geophysical Research: Atmospheres*, 121, 7548–7563. <https://doi.org/https://doi.org/10.1002/2015JD024339>

Kennedy, D., Dagon, K., Lawrence, D. M., Fisher, R. A., Benjamin, M., Collier, N., et al. (2024). One-at-a-time Parameter Perturbation Ensemble of the Community Land Model , version 5.1.

690 Kosugi, Y., Takanashi, S., Yokoyama, N., Philip, E., and Kamakura, M. (2012). Vertical variation in leaf gas exchange parameters for a Southeast Asian tropical rainforest in Peninsular Malaysia. *Journal of Plant Research*, 125(6), 735–748. <https://doi.org/10.1007/s10265-012-0495-5>

Kratzert F., Klotz D., Brenner C., Schulz K., and Herrnegger M., Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*, 22.11: 6005-6022. 2018.

695 Lambin, E. F., Geist, H. J., and Lepers, E. (2003). Dynamics of Land-Use and Land-Cover Change in Tropical Regions. *Annual Review of Environment and Resources*, 28(1), 205–241. <https://doi.org/10.1146/annurev.energy.28.050302.105459>

Lawrence, P., Feddema, J., Bonan, G., Meehl, G., O'Neill, B., Oleson, K., et al. (2012). Simulating the biogeochemical and biogeophysical impacts of transient land cover change and wood harvest in the Community Climate System Model (CCSM4) from 1850 to 2100. *Journal of Climate*, 25(9), 3071–3095. <https://doi.org/10.1175/JCLI-D-11-00256.1>

700 Li, J., Miao, C., Zhang, G., Fang, Y. H., Shangguan, W., & Niu, G. Y. (2022). Global evaluation of the Noah-MP land surface model and suggestions for selecting parameterization schemes. *Journal of Geophysical Research: Atmospheres*, 127(5), e2021JD035753.

Li, J., Chen, F., Lu, X., Gong, W., Zhang, G., & Gan, Y. (2020). Quantifying contributions of uncertainties in physical parameterization schemes and model parameters to overall errors in Noah-MP dynamic vegetation modeling. *Journal of Advances in Modeling Earth Systems*, 12(7), e2019MS001914.

705 Lin, Y., Dong, W., Zhang, M., Xie, Y., Xue, W., Huang, J., and Luo, Y. (2017). Causes of model dry and warm bias over central U.S. and impact on climate projections. *Nature Communications*, 8(1), 1–8. <https://doi.org/10.1038/s41467-017-01040-2>

Lion, M., Kosugi, Y., Takanashi, S., Noguchi, S., Itoh, M., Katsuyama, M., et al. (2017). Evapotranspiration and water source of a tropical rainforest in peninsular Malaysia. *Hydrological Processes*, 31(24), 4338–4353. <https://doi.org/10.1002/hyp.11360>

Litt, G. F., Ogden, F. L., Mojica, A., Hendrickx, J. M. H., Kempema, E. W., Gardner, C. B., et al. (2020). Land cover effects on soil infiltration capacity measured using plot scale rainfall simulation in steep tropical lowlands of Central Panama. *Hydrological Processes*, 34(4), 878–897. <https://doi.org/10.1002/hyp.13605>

715 Lu, D., and Ricciuto, D. (2019). Efficient surrogate modeling methods for large-scale Earth system models based on machine learning techniques. *Geoscientific Model Development*, 1791–1807. <https://doi.org/https://doi.org/10.5194/gmd-12-1791-2019>

Lu, D., Ricciuto, D., Stoyanov, M., and Gu, L. (2018). Calibration of the E3SM Land Model Using Surrogate-Based Global Optimization. *Journal of Advances in Modeling Earth Systems*, 10(6), 1337–1356.

- 720 <https://doi.org/10.1002/2017MS001134>
- Lu, D., Liu, Y., Zhang, Z., Bao, F., and Zhang, G. (2024). A Diffusion-Based Uncertainty Quantification Method to Advance E3SM Land Model Calibration. *Journal of Geophysical Research: Machine Learning and Computation*, 1(3). <https://doi.org/10.1029/2024jh000234>
- Ma, H. Y., Mechoso, C. R., Xue, Y., Xiao, H., Neelin, J. D., and Ji, X. (2013). On the connection between continental-scale
725 land surface processes and the tropical climate in a coupled ocean-atmosphere-land system. *Journal of Climate*, 26(22), 9006–9025. <https://doi.org/10.1175/JCLI-D-12-00819.1>
- Mahmood, R., Pielke, R. A., Hubbard, K. G., Niyogi, D., Dirmeyer, P. A., Mcalpine, C., et al. (2014). Land cover changes and their biogeophysical effects on climate. *International Journal of Climatology*, 34(4), 929–953. <https://doi.org/10.1002/joc.3736>
- 730 Marryanna, L., S. Noguchi, Y. Kosugi, K. Niiyama, M. Itoh, T. Sato, S. Takanashi, S. Siti-Aisah, and K. A.-R. (2019). Spatial Distribution of Soil Moisture and Its Influence on Stand Structure in a Lowland Dipterocarp Forest in Peninsular Malaysia. *Journal of Tropical Forest Science*, 31(2), 135–150.
- Müller, J., Paudel, R., Shoemaker, C. A., Woodbury, J., Wang, Y., and Mahowald, N. (2015). CH4 parameter estimation in CLM4.5bgc using surrogate global optimization. *Geoscientific Model Development*, 8(10), 3285–3310.
735 <https://doi.org/10.5194/gmd-8-3285-2015>
- Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research Atmospheres*, 116(12), 1–19. <https://doi.org/10.1029/2010JD015139>
- Noguchi, S., Kosugi, Y., Takanashi, S., Tani, M., Niiyama, K., Siti Aisah, S., and Lion, M. (2016). Long-term variation in soil
740 moisture in Pasoh forest reserve, a lowland tropical rainforest in Malaysia. *Journal of Tropical Forest Science*, 28(5), 324–333.
- Ogden, F. L., Crouch, T. D., Stallard, R. F., and Hall, J. S. (2013). Effect of land cover and use on dry season river runoff, runoff efficiency, and peak storm runoff in the seasonal tropics of Central Panama. *Water Resources Research*, 49(12), 8443–8462. <https://doi.org/10.1002/2013WR013956>
- 745 Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., and Phillip, O. L. (2011). A Large and Persistent Carbon Sink in the World’s Forests. *Science*, 333(6045), 988–993. Retrieved from 10.1126/science.1201609
- Pielke, R. A., Pitman, A., Niyogi, D., Mahmood, R., McAlpine, C., Hossain, F., et al. (2011). Land use/land cover changes and climate: Modeling analysis and observational evidence. *Wiley Interdisciplinary Reviews: Climate Change*, 2(6), 828–850. <https://doi.org/10.1002/wcc.144>
- 750 Pitman, A. J., De Noblet-Ducoudré, N., Cruz, F. T., Davin, E. L., Bonan, G. B., Brovkin, V., et al. (2009). Uncertainties in climate responses to past land cover change: First results from the LUCID intercomparison study. *Geophysical Research Letters*, 36(14), 1–6. <https://doi.org/10.1029/2009GL039076>
- Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games*, 2(28), 307–317.

- Storn, R., & Price, K. (1997). Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization*, 11, 341–359. <https://doi.org/libproxy1.nus.edu.sg/10.1023/A:1008202821328>
- Schlesinger, W. H., and Jasechko, S. (2014). Transpiration in the global water cycle. *Agricultural and Forest Meteorology*, 189–190, 115–117. <https://doi.org/10.1016/j.agrformet.2014.01.011>
- Shao, R., and Zhang, B. (2019). Estimating the Increase in Regional Evaporative Water Consumption as a Result of Vegetation Restoration Over the Loess Plateau, China, *Journal of Geophysical Research: Atmospheres*, 783–802. <https://doi.org/10.1029/2019JD031295>
- Sun, R., Duan, Q., and Huo, X. (2021). Multi-Objective Adaptive Surrogate Modeling-Based Optimization for Distributed Environmental Models Based on Grid Sampling. *Water Resources Research*, 57(11). <https://doi.org/10.1029/2020WR028740>
- Sun, W., Wang, B., and Wang, Y. (2021). Parameterizing Subgrid Variations of Land Surface Heat Fluxes to the Atmosphere Improves Boreal Summer Land Precipitation Simulation With the NCAR CESM1 . 2 *Geophysical Research Letters*. *Geophysical Research Letters*, 48, 1–10. <https://doi.org/10.1029/2020GL090715>
- Teo, H. C., Raghavan, S. V., He, X., Zeng, Z., Cheng, Y., Luo, X., Lechner, A. M., Ashfold, M. J., Lamba, A., Sreekar, R., Zheng, Q., Chen, A. and Koh, L. P.: Large-scale reforestation can increase water yield and reduce drought risk for water-insecure regions in the Asia-Pacific, *Glob. Chang. Biol.*, (August), 6385–6403, doi:10.1111/gcb.16404, 2022.
- Unger, N. (2014). Human land-use-driven reduction of forest volatiles cools global climate. *Nature Climate Change*, 4(10), 907–910. <https://doi.org/10.1038/nclimate2347>
- Vetrita, Y., and Cochrane, M. A. (2020). Fire frequency and related land-use and land-cover changes in Indonesia’s Peatlands. *Remote Sensing*, 12(1). <https://doi.org/10.3390/RS12010005>
- Van Weverberg, K., Morcrette, C. J., Petch, J., Klein, S. A., Ma, H. Y., Zhang, C., et al. (2018). CAUSES: Attribution of Surface Radiation Biases in NWP and Climate Models near the U.S. Southern Great Plains. *Journal of Geophysical Research: Atmospheres*, 123(7), 3612–3644. <https://doi.org/10.1002/2017JD027188>
- Weiskopf, S. R., Harm, Z. V, Miller, B. W., Myers, B. J. E., Pereira, L., Cecilia, M., Arce-plata, M. I., Blanchard, J. L., Ferrier, S., Fulton, E. A., Harfoot, M., Isbell, F., Johnson, J. A., Mori, A. S., Weng, E., Isabel, M. and Rosa, D.: Increasing the uptake of ecological model results in policy decisions to improve biodiversity outcomes, *Environ. Model. Softw.*, 149(January), doi:10.1016/j.envsoft.2022.105318, 2022.
- Yan, H., Sun, N., Eldardiry, H., Thurber, T. B., Reed, P. M., Malek, K., Gupta, R., Kennedy, D., Swenson, S. C., Hou, Z., Cheng, Y. and Rice, J. S.: Large Ensemble Diagnostic Evaluation of Hydrologic Parameter Uncertainty in the Community Land Model Version Journal of Advances in Modeling Earth Systems, *J. Adv. Model. Earth Syst.*, 15, 1–23, doi:10.1029/2022MS003312, 2022.
- Yang, Z. L., Niu, G. Y., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 2 . Evaluation over global river basins, 116, 1–16.

<https://doi.org/10.1029/2010JD015140>

Zhang, B., Tian, L., Yang, Y., and He, X. (2022). Revegetation Does Not Decrease Water Yield in the Loess Plateau of China.

790 *Geophysical Research Letters*, 49(9). <https://doi.org/10.1029/2022GL098025>

Zhang, G., Chen, Y., and Li, J. (2021). Effects of organic soil in the Noah-MP land-surface model on simulated skin and soil temperature profiles and surface energy exchanges for China. *Atmospheric Research*, 249(September 2020). <https://doi.org/10.1016/j.atmosres.2020.105284>

795 Zhang, J., Zhang, B., Cheng, Y., Wang, X., Yang, F. and Zhao, X. (2026): Revegetation Rebalances Water Resources by Enhancing Rainwater to Increase Vegetation Carrying Capacity in China's Loess Plateau, *Water Resources Research*, 1–18, doi:10.1029/2025WR040307, 2026.