Response to Reviews

EGUSPHERE-2025-3898: Assessing and enhancing Noah-MP land surface modeling over tropical environments

Responses to Reviewer #1:

Comment 1: This manuscript presents a valuable and timely study on the calibration and evaluation of the Noah-MP land surface model in tropical environments, which are critically important to the global climate system. The authors use data from two tropical forest sites (Panama and Malaysia) and one tropical urban site (Singapore) to perform site-specific calibrations, assess model performance across multiple timescales, and identify key parameter sensitivities. The paper is well-structured and well-written, the analysis is logical, and the conclusions are largely supported by the results.

Response 1: Thank you for the encouraging and helpful comments. We have carefully addressed all the comments, including:

- Establishing a new machine learning-based calibration framework to calibrate Noah-MP parameters at both tropical forest sites.
- Separating the data into calibration and validation periods for both tropical forest sites.
- Applying the Shapley value decomposition method to quantify the contribution of each calibrated parameter to the overall improvement in model performance.
- Expanding the discussion on the transferability of the calibration framework and calibrated parameters across tropical forests.

In the item-to-item response below, we have numbered the comments and responses sequentially. Our responses are in blue text and our edits are in red text in the memo below. We also underline line numbers, as well as figures, tables, and equations that have been edited to respond to the suggestions.

Specific comments:

Comment 2: 1. The study employs a univariate sensitivity analysis followed by the selection of parameter values that yield the "lowest bias". Univariate sensitivity analysis does not account for interactions between parameters, which are often significant in complex models like Noah-MP. In addition, was the optimization based on a single variable (e.g., SH), or was a subjective "best-fit" across variables chosen? A more robust discussion is needed to justify why more sophisticated approach of sensitivity analysis and calibration were not employed.

Response 2: Thank you for the helpful comments. We fully agree that univariate sensitivity analysis does not explicitly account for parameter interactions. To address this, we have implemented a machine learning-based emulator calibration in the revised manuscript. Details of this emulator-based calibration are provided in the revised manuscript in section 2.5 (lines 185-298).

We have also compared the performance of Noah-MP using default parameters, the sensitivity analysis parameters from the prior version, and the emulator-based calibration parameters in this revised version. Tables R1 and R2 summarize the evaluation metrics for the Panama BCI and Malaysia PSO sites, respectively, including Root Mean Square Error (RMSE), mean bias (BIAS), and R^2. Overall, the emulator-based calibration shows satisfying performance at both sites.

Specifically, for the Panama BCI site (Table R1 below), all performance metrics of the emulator-based calibration match those of the previous sensitivity-based calibration for latent and sensible heat, and show substantial improvements in RMSE, BIAS, and R^2 for soil moisture compared to both sensitivity-based calibration and default parameterization. These validation results further demonstrate the effectiveness and physical feasibility of the emulator-based calibration.

For the Malaysia PSO site (Table R2 below), although the default configuration shows acceptable RMSE performance, the emulator-based calibration substantially reduces BIAS, indicating improved accuracy in capturing the overall energy balance. Emulator-based calibration also outperforms sensitivity-based calibration in RMSE and R^2 during both the calibration and validation periods.

Table R1. Comparison of Noah-MP simulation performance at the Panama BCI site using default parameters, sensitivity-based calibration (old version), and emulator-based calibration (revised version), for both the calibration and validation periods.

Variable	Data period	Parameter set	RMSE	BIAS	R2
LH (W/m^2)	Calibration	Default	44.00	7.31	0.79
		Sensitivity (old version)	57.70	18.70	0.64
		Emulator (new version)	59.80	20.40	0.61
	Validation	Default	46.70	3.98	0.81
		Sensitivity (old version)	57.50	16.70	0.71
		Emulator (new version)	59.70	19.10	0.69
	Full	Default	45.40	5.64	0.80
		Sensitivity (old version)	57.60	17.70	0.68
		Emulator (new version)	59.70	19.70	0.65
SH (W/m^2)	Calibration	Default	51.20	30.60	0.69
		Sensitivity (old version)	36.90	20.10	0.84
		Emulator (new version)	33.50	17.40	0.87
	Validation	Default	56.90	34.90	0.57
		Sensitivity (old version)	43.10	22.50	0.76
		Emulator (new version)	37.70	19.00	0.81
	Full	Default	54.10	32.70	0.63
		Sensitivity (old version)	40.10	21.30	0.80
		Emulator (new version)	35.70	18.30	0.84
SM (m^3/m^3)	Calibration	Default	0.051	-0.020	0.61
		Sensitivity (old version)	0.043	0.012	0.72
		Emulator (new version)	0.042	0.006	0.73
	Validation	Default	0.066	-0.035	0.40
		Sensitivity (old version)	0.055	0.006	0.58
		Emulator (new version)	0.055	-0.008	0.59
	Full	Default	0.059	-0.028	0.51
		Sensitivity (old version)	0.049	0.009	0.66

		Emulator (new version)	0.049	-0.001	0.67

Table R2. Comparison of Noah-MP simulation performance at the Malaysia PSO site using default parameters, sensitivity-based calibration (old version), and emulator-based calibration (revised version), for both the calibration and validation periods.

Variable	Data period	Parameter set	RMSE	BIAS	R2
LH (W/m^2)	Calibration	Default	52.1	-14.2	0.87
		Sensitivity (old version)	45.7	-3.2	0.90
		Emulator (new version)	48.7	-8.4	0.89
	Validation	Default	53.3	-10.0	0.87
		Sensitivity (old version)	49.6	-0.9	0.89
		Emulator (new version)	50.7	-2.3	0.88
	Full	Default	52.8	-11.7	0.87
		Sensitivity (old version)	48.1	-1.8	0.89
		Emulator (new version)	49.9	-4.7	0.89
SH (W/m^2)	Calibration	Default	46.2	7.6	0.82
		Sensitivity (old version)	51.0	-1.7	0.78
		Emulator (new version)	48.2	3.3	0.80
	Validation	Default	46.0	5.8	0.81
		Sensitivity (old version)	49.9	-1.7	0.78
		Emulator (new version)	49.6	-0.3	0.78
	Full	Default	46.1	6.5	0.82
		Sensitivity (old version)	50.3	-1.7	0.78
		Emulator (new version)	49.0	1.2	0.79
SM (m^3/m^3)	Calibration	Default	0.022	-0.009	0.62
		Sensitivity (old version)	0.022	-0.010	0.63
		Emulator (new version)	0.019	0.005	0.73
	Validation	Default	0.021	-0.014	0.51
		Sensitivity (old version)	0.025	-0.019	0.26
		Emulator (new version)	0.015	-0.001	0.74
	Full	Default	0.021	-0.012	0.58
		Sensitivity (old version)	0.024	-0.015	0.46
		Emulator (new version)	0.017	0.002	0.75

Regarding parameter optimization, since we have changed the calibration method, the previous selection based on "lowest bias" is no longer applicable. In the revised version using emulator-based calibration, the loss function is defined as the sum of the normalized RMSEs for all three variables (sensible heat, latent heat, and soil moisture, Equations 5-7). We selected the parameter combination that minimized this summed normalized RMSE, representing the best compromise across the variables rather than optimizing toward a single target variable. We have clarified this in the revised manuscript: "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)." (lines 274-283).

Comment 3: 2. The analysis for the Singapore urban site feels less comprehensive than for the forest sites. The primary calibration involves adjusting a single parameter, the urban fraction. While the conclusion that the SLUCM's simplified hydrology is a major structural flaw is well-argued and convincing, the paper would be stronger if it clarified whether a sensitivity analysis was also performed for other relevant parameters. If not, the authors should state why. This would provide a more balanced comparison with the methodology used for the forest sites.
Response 3: Thank you for this helpful comment. We agree that the analysis for the urban site is not as comprehensive or methodologically balanced as that for the tropical forest sites. Given this imbalance and considering that the main scientific contributions of this study lie in the tropical forest calibration, we have decided to remove the urban site analysis from the revised manuscript and focus exclusively on tropical forests. Accordingly, we have revised the manuscript title to "Assessing and enhancing Noah-MP land surface modeling over tropical forests using machine learning techniques" (lines 1-2).

In addition, we have substantially strengthened the tropical forest calibration component by (1) introducing an efficient emulator-based calibration framework, (2) separating the datasets into calibration and validation periods to validate the parameters more rigorously, and (3) applying the Shapley value decomposition method to quantify the contribution of each calibrated parameter to the overall improvement in model performance and to rank their relative importance. These revisions improve the methodological rigor and general applicability of the proposed calibration approach for Noah-MP.

Comment 4: 3. Compared to the Panama BCI tropical forest site, Noah-MP fails to capture the seasonal and diurnal cycles of soil moisture at the PSO tropical forest site. The authors should provide a thorough discussion on potential reasons for this model deficiency at the PSO site.
Response 4: Thank you for the helpful comment. We would like to clarify that the apparently poorer performance of Noah-MP in capturing the seasonal and diurnal soil moisture cycles at the Malaysia PSO site arises primarily from differences in the y-axis scaling used in the original Figures 3e-f and Figures 5e-f (now Figures 5e-f and Figures 7e-f). Specifically, for the seasonal cycle, the y-axis interval was 0.05 m^3/m^3 for the Panama BCI site and 0.02 m^3/m^3 for the Malaysia PSO site. For the diurnal cycle, the corresponding intervals were 0.01 m^3/m^3 and 0.005 m^3/m^3, respectively. These differences in axis scaling visually exaggerate the variability at the PSO site and may give the impression of poorer model performance relative to the BCI site. To avoid this potential misinterpretation, we have revised the figures to use consistent y-axis intervals across sites (now Figures 5e-f and Figures 7e-f in the revised manuscript).

Nevertheless, overall, the calibration performs better at the Malaysia PSO site than at the Panama BCI site, and we discuss the potential reasons in the revised manuscript: "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." (lines 450-453) and "Soil moisture shows weaker seasonal variability than at the Panama BCI site and is reasonably well captured by both the default and calibrated configurations." (lines 461-462).

Comment 5: 4. This study clearly demonstrates that the optimal parameter sets for the two tropical forest sites are substantially different. The authors rightly conclude that "calibrated parameter sets are likely not directly transferable across sites" (Line 412). What does this finding mean for applying Noah-MP over larger tropical domains where site-specific data for calibration is unavailable? The paper would be strengthened by adding a more detailed discussion on the challenges of parameter transferability and what this means for the validity of the Plant Functional Type (PFT) concept in tropical regions.
Response 5: Thank you for the helpful comment. We agree that parameter transferability is critical for regional applications and for the broader usefulness of Noah-MP. While it is challenging to define a single, universal "tropical" parameter set due to differences in climate regimes between the Panama BCI site (tropical monsoon climate, Am) and the Malaysia PSO site (tropical rainforest climate, Af), we have implemented three strategies in the revised manuscript to address this limitation and enhance the applicability of Noah-MP across tropical forests:

First, we extended the manuscript by introducing a machine learning-based, computational efficient calibration framework that enables efficient and systematic parameter estimation at new sites. We demonstrate the generality of this framework by applying the emulator-based calibration method to both the Panama BCI and Malaysia PSO sites, and by conducting independent calibration and validation experiments at each site (Figures 4 and 6). This shows that the calibration method itself is transferable, even if the optimized parameter values are not. Details of the machine learning-based emulator calibration are provided in Section 2.5 of the revised manuscript (lines 184-298). We also added a section discussing the 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." (Section 4.1, lines 486-495).

Second, while parameter values are not transferable across sites with differing tropical climates, they may be transferable among sites within similar climatic regimes (e.g., Am and Af). This is supported by our calibration and validation experiments at both sites, where the calibrated parameter sets perform consistently well during independent validation periods (Figures 4 and 6).

Third, we added a new paragraph that explicitly discusses: (1) the limitations of defining a single tropical parameter set within a single plant functional type (PFT), (2) the implications for applying Noah-MP in data-sparse tropical regions, and (3) possible pathways forward, including climate-based or regionalized calibration strategies. Specifically, we have added discussions: "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." (Section 4.2, lines 496-516).

Comment 6: 5. The lines in Fig. 2 and Fig. 4 are too cluttered, making it difficult for intercomparison. Please change it to another format.

Response 6: Thank you for the suggestion. We have revised both Figures 2 and 4 in the updated manuscript (now Figures 4 and 6) to plot 10-day running mean.

Comment 7: 6. Please restructure the table into a three-part layout.

Response 7: We apologize for not fully understanding the request to restructure the table into a three-part layout. However, we have carefully reviewed all tables and ensured that they are now clear, well-structured, and consistently formatted throughout the revised manuscript.

Responses to Reviewer #2:

Comment 1: The manuscript addresses the poor performance of Land Surface Models (LSMs) like Noah-MP in tropical regions due to parameterization and structural limitations. Therefore, the site-specific calibration at three diverse tropical sites (two forest, one urban) is a necessary first step towards improving model performance.

However, in my opinion, the manuscript provides useful site-specific data, but it must be refocused to maximize its scientific utility by synthesizing the calibration results into a transferable tropical parameter set and deepening the methodological analysis of structural and parametric uncertainty.

The work currently functions as a detailed case study collection rather than a definitive enhancement of Noah-MP for the tropical biome. The scientific significance is limited by a lack of synthesis and depth. The authors identify core limitations, the need for regional parameter sets and addressing structural model deficiencies, but fail to incorporate these essential steps into the current work, leaving the most valuable findings unpursued. The methodology for sensitivity analysis and the approach to urban calibration are also overly simplistic for a multi-physics model like Noah-MP.

Response 1: Thanks for the helpful comments. We have carefully addressed all the comments, including:

- Establishing a new machine learning-based calibration framework to calibrate Noah-MP parameters at both tropical forest sites.
- Separating the data into calibration and validation periods for both tropical forest sites.
- Applying the Shapley value decomposition method to quantify the contribution of each calibrated parameter to the overall improvement in model performance.
- Expanding the discussion on the transferability of the calibration framework and calibrated parameters across tropical forests.
- Outlining potential approaches for implementing preferential flow processes into Noah-MP.
- Evaluating the impact of multi-physics options in Noah-MP on the results.

In the item-to-item response below, we have numbered the comments and responses sequentially. Our responses are in blue text and our edits are in red text in the memo below. We also underline line numbers, as well as figures that have been edited to respond to the suggestions.

Major Concerns:

Comment 2: 1. The manuscript correctly identifies that the transferability of calibrated parameters is essential for international readers and regional applications. The mere collection of site-specific parameter values, which are demonstrably different between the Panama and Malaysia forest sites, holds limited significance. The authors must move beyond a "foundational step." Based on this comparison and the sensitivity analysis, it's better for the authors to propose a generalized tropical set of parameters and test this new set against one of the study sites.

Response 2: Thank you for the helpful and constructive comment. We agree that parameter transferability is critical for regional applications and for the broader usefulness of Noah-MP. While it is challenging to define a single, universal "tropical" parameter set due to differences in climate regimes between the Panama BCI site (tropical monsoon climate, Am) and the Malaysia PSO site (tropical rainforest climate, Af), we have implemented three strategies in the revised manuscript to address this limitation and enhance the applicability of Noah-MP across tropical forests:

First, we extended the manuscript by introducing a machine learning-based, computational efficient calibration framework that enables efficient and systematic parameter estimation at new sites. We demonstrate the generality of this framework by applying the emulator-based calibration method to both the Panama BCI and Malaysia PSO sites, and by conducting independent calibration and validation experiments at each site (Figures 4 and 6). This shows that the calibration method itself is transferable, even if the optimized parameter values are not. Details of the machine learning-based emulator calibration are provided in Section 2.5 of the revised manuscript (lines 184-298). We also added a section discussing the 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." (Section 4.1, lines 486-495).

Second, while parameter values are not transferable across sites with differing tropical climates, they may be transferable among sites within similar climatic regimes (e.g., Am and Af). This is supported by our calibration and validation experiments at both sites, where the calibrated parameter sets perform consistently well during independent validation periods (Figures 4 and 6).

Third, we added a new paragraph that explicitly discusses: (1) the limitations of defining a single tropical parameter set within a single plant functional type (PFT), (2) the implications for applying Noah-MP in data-sparse tropical regions, and (3) possible pathways forward, including climate-based or regionalized calibration strategies. Specifically, we have added discussions: "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." (Section 4.2, lines 496-516).

Comment 3: 2. The paper correctly notes critical model limitations, such as the missing simulation of preferential flow in the tropical soil column, which limits soil moisture accuracy. While identifying this is important, simply leaving it in the discussion stage severely limits the study's impact.

Response 3: We appreciate the reviewer's insightful comment and fully agree that the lack of preferential flow representation is a significant limitation, particularly for tropical soils. Implementing and validating a preferential flow scheme within Noah-MP would require substantial model development and additional parameterization and calibration, which are beyond the scope of this study. The primary objective of this work is to develop and demonstrate a robust calibration framework applicable across multiple sites, rather than to achieve the most physically complete soil hydrological representation. Accordingly, we focused on maintaining internal model consistency across all experiments.

Nevertheless, we agree that merely noting this limitation in the discussion is insufficient. To address this concern and strengthen the manuscript, we have expanded the discussion to explicitly identify preferential flow representation as a priority direction for future Noah-MP development and have added a dedicated subsection outlining potential approaches for implementing preferential flow processes into the model in the revised manuscript: "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 incorporate preferential flow pathways into Noah-MP." (Section 4.3.2, lines 534-540).

Comment 4: 3. The methodological approach for both sensitivity analysis and structural investigation is insufficient for a complex model like Noah-MP. The study relies on a form of One-at-a-Time (OAT) or linear univariate sampling. The authors need to acknowledge this constraint and, ideally, perform a more robust Global Sensitivity Analysis (GSA) to quantify the non-linear interaction effects.
Response 4: Thanks for the helpful comment. We fully agree that One-at-a-Time (OAT) or linear univariate sampling is insufficient for parameter calibration in a complex model like Noah-MP. To address this limitation, we have implemented a machine learning-based emulator calibration method for both tropical forest sites in the revised manuscript. Details of this emulator-based calibration are provided in Section 2.5 (lines 184-298).

We have also compared the performance of Noah-MP using default parameters, the sensitivity analysis parameters from the prior version, and the emulator-based calibration parameters in this revised version. Tables R1 and R2 summarize the evaluation metrics for the Panama BCI and Malaysia PSO sites, respectively, including Root Mean Square Error (RMSE), mean bias (BIAS), and R^2. Overall, the emulator-based calibration shows satisfying performance at both sites.

Specifically, for the Panama BCI site (Table R1 below), all performance metrics of the emulator-based calibration match those of the previous sensitivity-based calibration for latent and sensible heat, and show substantial improvements in RMSE, BIAS, and R^2 for soil moisture compared to both sensitivity-based calibration and default parameterization. These validation results further demonstrate the effectiveness and physical feasibility of the emulator-based calibration.

For the Malaysia PSO site (Table R2 below), although the default configuration shows acceptable RMSE performance, the emulator-based calibration substantially reduces BIAS, indicating improved accuracy in capturing the overall energy balance. Emulator-based calibration also outperforms sensitivity-based calibration in RMSE and R^2 during both the calibration and validation periods.

Table R1. Comparison of Noah-MP simulation performance at the Panama BCI site using default parameters, sensitivity-based calibration (old version), and emulator-based calibration (revised version), for both the calibration and validation periods.

Variable	Data period	Parameter set	RMSE	BIAS	R2
LH (W/m^2)	Calibration	Default	44.00	7.31	0.79
		Sensitivity (old version)	57.70	18.70	0.64
		Emulator (new version)	59.80	20.40	0.61
	Validation	Default	46.70	3.98	0.81
		Sensitivity (old version)	57.50	16.70	0.71
		Emulator (new version)	59.70	19.10	0.69
	Full	Default	45.40	5.64	0.80
		Sensitivity (old version)	57.60	17.70	0.68
		Emulator (new version)	59.70	19.70	0.65
SH (W/m^2)	Calibration	Default	51.20	30.60	0.69
		Sensitivity (old version)	36.90	20.10	0.84
		Emulator (new version)	33.50	17.40	0.87
	Validation	Default	56.90	34.90	0.57
		Sensitivity (old version)	43.10	22.50	0.76
		Emulator (new version)	37.70	19.00	0.81
	Full	Default	54.10	32.70	0.63
		Sensitivity (old version)	40.10	21.30	0.80
		Emulator (new version)	35.70	18.30	0.84
SM (m^3/m^3)	Calibration	Default	0.051	-0.020	0.61
		Sensitivity (old version)	0.043	0.012	0.72
		Emulator (new version)	0.042	0.006	0.73
	Validation	Default	0.066	-0.035	0.40
		Sensitivity (old version)	0.055	0.006	0.58
		Emulator (new version)	0.055	-0.008	0.59
	Full	Default	0.059	-0.028	0.51
		Sensitivity (old version)	0.049	0.009	0.66
		Emulator (new version)	0.049	-0.001	0.67

Table R2. Comparison of Noah-MP simulation performance at the Malaysia PSO site using default parameters, sensitivity-based calibration (old version), and emulator-based calibration (revised version), for both the calibration and validation periods.

Variable	Data period	Parameter set	RMSE	BIAS	R2
LH (W/m^2)	Calibration	Default	52.1	-14.2	0.87
		Sensitivity (old version)	45.7	-3.2	0.90
		Emulator (new version)	48.7	-8.4	0.89

	Validation	Default	53.3	-10.0	0.87
		Sensitivity (old version)	49.6	-0.9	0.89
		Emulator (new version)	50.7	-2.3	0.88
	Full	Default	52.8	-11.7	0.87
		Sensitivity (old version)	48.1	-1.8	0.89
		Emulator (new version)	49.9	-4.7	0.89
SH (W/m^2)	Calibration	Default	46.2	7.6	0.82
		Sensitivity (old version)	51.0	-1.7	0.78
		Emulator (new version)	48.2	3.3	0.80
	Validation	Default	46.0	5.8	0.81
		Sensitivity (old version)	49.9	-1.7	0.78
		Emulator (new version)	49.6	-0.3	0.78
	Full	Default	46.1	6.5	0.82
		Sensitivity (old version)	50.3	-1.7	0.78
		Emulator (new version)	49.0	1.2	0.79
SM (m^3/m^3)	Calibration	Default	0.022	-0.009	0.62
		Sensitivity (old version)	0.022	-0.010	0.63
		Emulator (new version)	0.019	0.005	0.73
	Validation	Default	0.021	-0.014	0.51
		Sensitivity (old version)	0.025	-0.019	0.26
		Emulator (new version)	0.015	-0.001	0.74
	Full	Default	0.021	-0.012	0.58
		Sensitivity (old version)	0.024	-0.015	0.46
		Emulator (new version)	0.017	0.002	0.75

Comment 5: 4. The study only performs calibration for a single, fixed configuration of Noah-MP. This ignores the model's core strength: its multi-parameterization design. Better to comment on which combinations of parameterization schemes were tested or why they were excluded, as the scheme choice may have a greater impact on tropical hydrology than parameter tuning alone.

Response 5: We agree that different combinations of Noah-MP's parameterization schemes could potentially influence model results. To address this, we conducted preliminary tests by manually perturbing several key options that likely impact LH and SH (e.g., stomatal resistance, soil moisture factor, and surface resistance options) at both study sites. These tests showed no significant difference in model performance prior to parameter optimization (please refer to Figures R1-R2 below). Therefore, we proceeded with the default Noah-MP physics option set, which has been shown to perform reasonably well in most applications (He et al., 2023), and focused on parameter optimization.

We added the following clarification in the revised manuscript: "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.” (lines 158-163).

Additionally, we acknowledge the potential uncertainty associated with different physics option combinations and have added a discussion on this in the revised manuscript: “This study applies only the default Noah-MP physics option set. Different combinations of physics options may yield different results, which warrants future investigation.” (lines 542-543).

References:
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.
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.

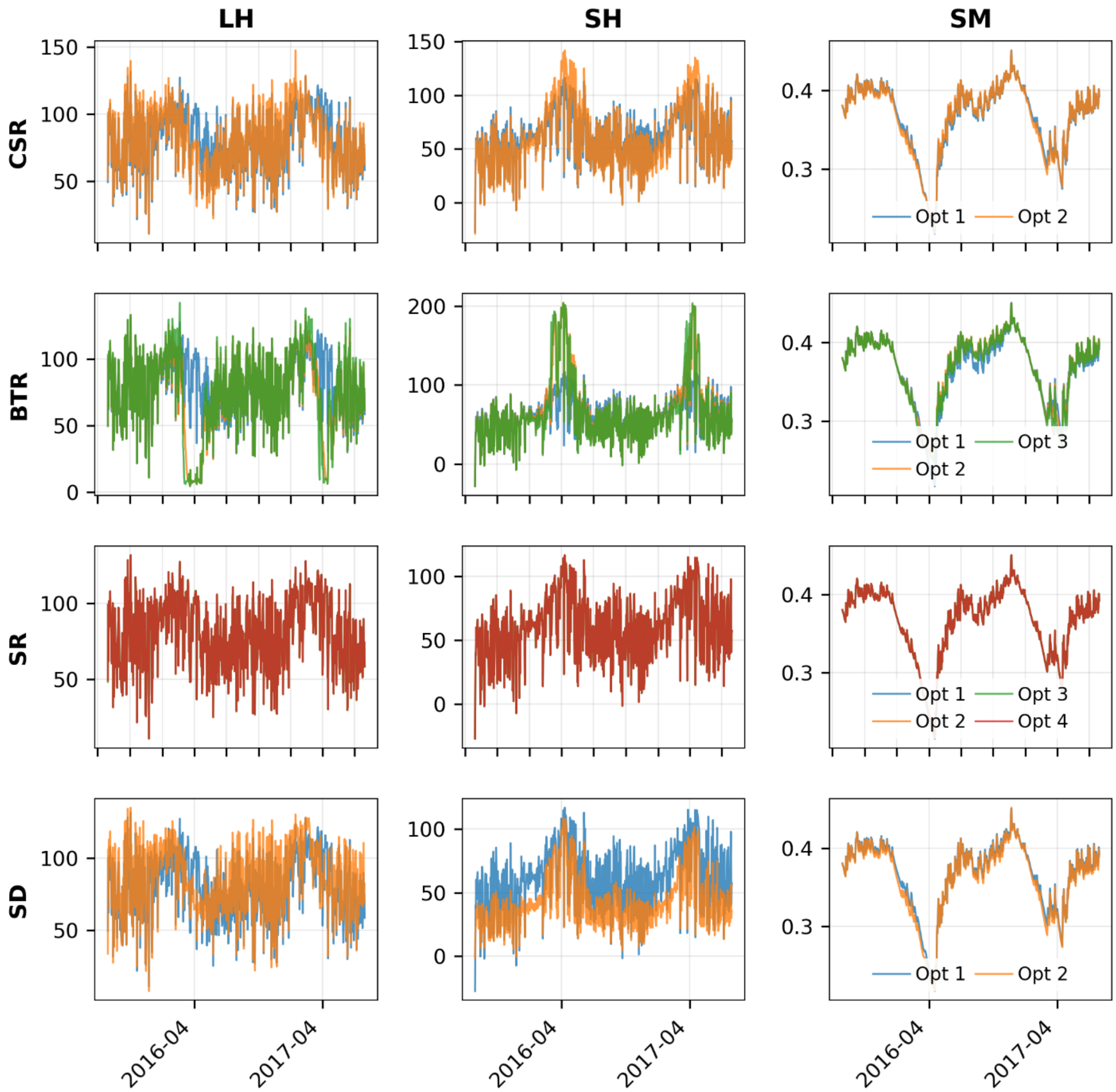

Figure R1: Sensitivity analysis of four Noah-MP physics options at the Panama BCI site. Rows represent four physics options: Canopy Stomatal Resistance (CSR; first row), Beta Factor (BTR; second row), Surface Resistance (SR; third row), and Surface Drag (SD; fourth row). Columns represent the target variables: Latent Heat (LH; left column), Sensible Heat (SH; middle column), and Soil Moisture (SM; bottom column). Colored lines indicate the specific sub-options tested for each physical option.

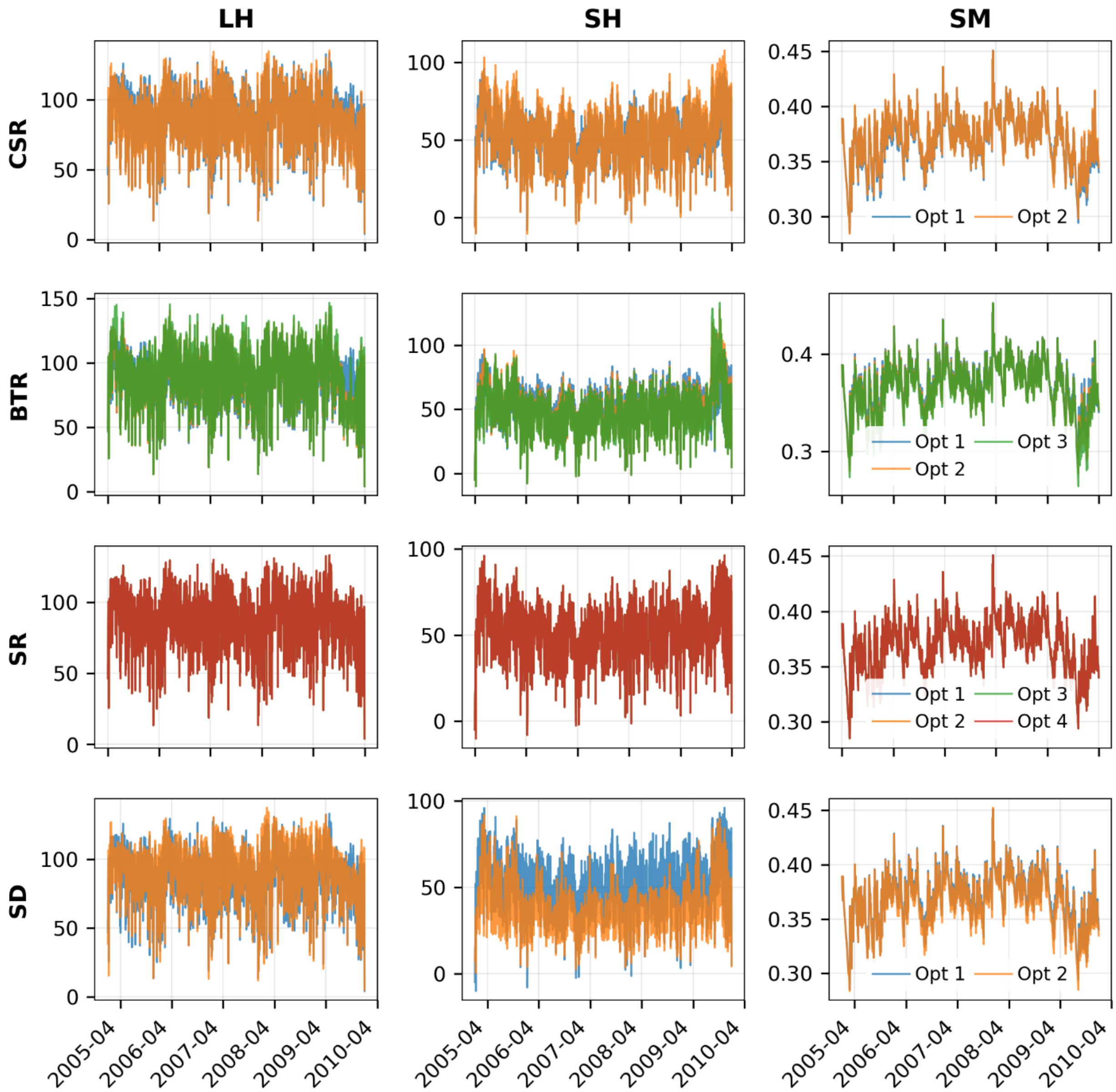

Figure S2: Sensitivity analysis of four Noah-MP physics options at the Malaysia PSO site. Rows represent four physics options: Canopy Stomatal Resistance (CSR; first row), Beta Factor (BTR; second row), Surface Resistance (SR; third row), and Surface Drag (SD; fourth row). Columns represent the target variables: Latent Heat (LH; left column), Sensible Heat (SH; middle column), and Soil Moisture (SM; bottom column). Colored lines indicate the specific sub-options tested for each physical option.

Comment 6: 5. The calibration approach for the Singapore Urban site is scientifically flawed and creates an uneven comparison with the forest sites. The primary adjustment was the Urban Fraction which is an input/site characterization adjustment, not a true calibration of the underlying physical parameters of the SLUCM. The manuscript neglects to optimize crucial urban parameters that directly affect tropical urban physics such as Roof/Wall Albedo, Roughness Length, etc. The authors must clarify why only Urban Fraction was adjusted.
Response 6: Thanks for the helpful comments. We agree that the analysis for the urban site is not as comprehensive or methodologically balanced as that for the tropical forest sites. Given this imbalance and considering that the main scientific contributions of this study lie in the tropical forest calibration, we have decided to remove the urban site analysis from the revised manuscript and focus exclusively on tropical forests. Accordingly, we have revised the manuscript title to "Assessing and enhancing Noah-MP land surface modeling over tropical forests using machine learning techniques" (lines 1-2).

In addition, we have substantially strengthened the tropical forest calibration component by (1) introducing an efficient emulator-based calibration framework, (2) separating the datasets into calibration and validation periods to validate the parameters more rigorously, and (3) applying the Shapley value decomposition method to quantify the contribution of each calibrated parameter to the overall improvement in model performance and to rank their relative importance. These revisions improve the methodological rigor and general applicability of the proposed calibration approach for Noah-MP.

Some specific comments:
Comment 7: 1. Include a summary table comparing model performance metrics (e.g., RMSE, correlation) across all three sites for both default and calibrated simulations. This is essential for a quick, readable assessment of the study's impact.
Response 7: Thanks for the suggestions. We have added a new table summarizing the performance metrics in the revised manuscript (Tables 4 and 6).

Comment 8: 2. "a mean canopy height of 30" lacks units of measurement.
Response 8: We have added a unit of meter in the revised manuscript (line 130).

Comment 9: 3. The description of the calibration method is unclear. Please specify the objective function used for calibration.
Response 9: We have added the objective function in the revised manuscript (line 274-283, Equations 5-7).

Comment 10: 4. "such as such as peat-swamp forests..." Please remove the duplicate "such as."
Response 10: Removed as suggested.