Supplementary Material for Enhancing Parameter Calibration in Land Surface Models Using a Multi-Task Surrogate Model within a Differentiable Parameter Learning Framework

Wenpeng Xie[1], Hongmei Li[2], Kei Yoshimura[1,3]

[1]Institute of Industrial Sciences, The University of Tokyo, Kashiwa, Japan.

[2]School of Civil and Environmental Engineering, Cornell University, USA

[3]Earth Observation Research Center, Japan Aerospace Exploration Agency, Tsukuba, Japan.

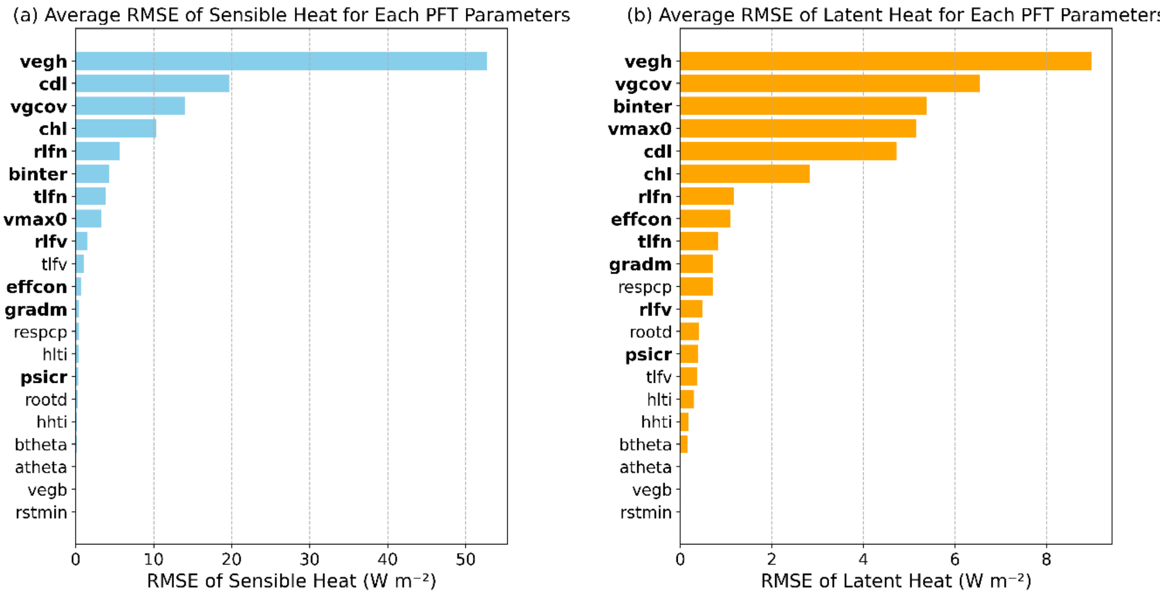

Figure S1. Comparison of average RMSEs between default and perturbed flux simulations for sensible and latent heat fluxes across PFT parameters(PLUMBER2, ES-LgS Site). The parameters in bold indicate calibration targets.

Table S1. MdPL parameter calibration results for Evergreen Forest. Unconstrained estimates falling outside the prescribed parameter ranges were clipped to the corresponding bounds; the resulting boundary values are shown in bold. Note that all reported parameter values are constrained within their predefined physical ranges and correspond to the values used in the ILS simulations.

Evergreen Forest	default	CN-Din	PT-Esp	PT-Mi1
vegh	35	20.22	6.60	**0.5**
rlfv	0.1	**0.05**	0.05	0.16
rlfn	0.45	0.42	0.34	0.63
tlfn	0.25	0.33	0.14	0.25
vgcov	1	0.96	0.11	0.84
cdl	0.11	0.37	0.29	0.41
chl	0.0274	**0.01**	**0.01**	0.0131
vmax0	6.00E-05	1.23E-05	4.36E-05	2.15E-06
gradm	9	10.93	**1**	3.98
binter	0.01	**0.001**	0.02	0.01
effcon	0.08	**0.01**	0.03	0.06
psicr	-200	-272.41	-313.03	-115.21

Table S2. MdPL parameter calibration results for Cultivation. Unconstrained estimates falling outside the prescribed parameter ranges were clipped to the corresponding bounds; the resulting boundary values are shown in bold.

Cultivation	default	DK-Fou	DE-Seh	IE-Ca1	IT-CA2	IT-Bci
vegh	1	**0.5**	2.06	2.19	**0.5**	1.13
rlfv	0.11	0.11	0.14	0.07	0.18	**0.05**
rlfn	0.58	0.23	0.45	0.22	0.33	0.30
tlfn	0.25	0.19	0.30	**0.1**	0.23	0.13
vgcov	1	0.49	0.78	0.86	0.86	0.82
cdl	0.098	0.212	0.310	0.442	0.252	0.110
chl	0.0246	0.0324	0.0152	**0.01**	0.0268	0.0432
vmax0	6.00E-05	2.73E-05	6.57E-05	1.00E-06	5.41E-05	5.04E-05
gradm	9	4.69	13.30	9.54	10.61	11.01
binter	0.01	**0.001**	0.07	**0.001**	0.05	0.05
effcon	0.08	0.06	0.01	0.04	0.06	0.03
psicr	-200	-370.56	-237.23	-367.18	-171.21	-215.11

Table S3. MdPL parameter calibration results for Grassland. Unconstrained estimates falling outside the prescribed parameter ranges were clipped to the corresponding bounds; the resulting boundary values are shown in bold.

Grassland	default	DK-Lva	PL-wet	CN-Dan	CN-Du2	IE-Dri
vegh	1	6.59	3.01	0.63	**0.5**	0.61
rlfv	0.11	0.17	**0.05**	**0.05**	**0.05**	0.12
rlfn	0.58	0.41	0.3	0.43	0.33	0.43
tlfn	0.25	0.27	0.1	0.15	**0.1**	0.20
vgcov	1	0.59	0.53	0.57	0.57	0.71
cdl	9.82E-2	0.270	0.081	0.242	0.146	0.207
chl	2.46E-2	0.0600	**0.01**	0.0106	0.0327	**0.01**
vmax0	6.00E-05	**1.00E-06**	2.81E-05	4.85E-05	4.51E-05	2.22E-05
gradm	4	1.05	5.51	9.84	8.62	5.05
binter	0.04	0.02	0.04	0.04	0.04	0.06
effcon	0.05	0.12	0.08	0.07	0.04	0.07
psicr	-200	-258.58	-484.86	-243.03	-327.12	-269.51

Table S4. MdPL parameter calibration results for Woodland. Unconstrained estimates falling outside the prescribed parameter ranges were clipped to the corresponding bounds; the resulting boundary values are shown in bold.

Woodland	default	AR-SLu	JP-SMF	CN-Cha	DE-Meh	US-Bar
vegh	20	9.84	27.23	1.28	4.44	27.86
rlfv	0.07	0.09	**0.05**	**0.05**	0.12	**0.05**
rlfn	0.4	0.34	0.40	0.42	0.42	0.42
tlfn	0.15	0.12	**0.1**	0.19	**0.1**	0.12
vgcov	1	0.60	0.71	0.99	1.00	0.90
cdl	0.111	0.34	0.387	0.108	0.275	0.337
chl	0.0277	0.01	**0.01**	0.0216	0.0291	0.0320
vmax0	6.00E-05	1.00E-5	6.25E-05	5.83E-05	3.17E-05	7.73E-06
gradm	7.5	7.6	3.99	10.35	3.63	**1**
binter	0.01	0.002	0.05	0.03	0.02	0.01
effcon	0.08	0.18	0.03	0.04	0.07	**0.01**
psicr	-200	-62.92	-187.25	-244.42	-496.00	-389.32

Supplementary Section S1. Comparison between multi-task MdPL and single-task dPL

To compare the performance of the multi-task MdPL surrogate (ILS_MdPL) with the single-task dPL surrogate (ILS_dPL), we selected four evergreen forest sites and performed parameter calibration using both approaches. The evaluation metrics obtained (**Table S5**) indicated that both models outperformed the default parameter set in terms of RMSE and R. For sensible heat, the average RMSE obtained for ILS_MdPL was slightly higher than that obtained for the ILS_dPL. However, ILS_MdPL showed a lower RMSE for latent heat than ILS_dPL. For R, ILS_MdPL outperformed ILS_dPL in terms of both sensible and latent heat fluxes. These observations indicated that even after sharing parameters across multiple tasks, ILS_MdPL still maintained strong collaborative effects between key variables, excelling in latent heat flux simulations. The parameter calibration results for each site are presented in **Table S6**.

Table S5. Evaluation metrics (RMSE and R) for parameter calibration performance based on the ILS_MdPL and ILS_dPL frameworks at four evergreen forest sites.

Site Name	**ILS_ORI**				**ILS_MdPL**				**ILS_dPL**			
	RMSE		R		RMSE		R		RMSE		R	
	Sensible heat ($W m^{-2}$)	Latent heat ($W m^{-2}$)	Sensible heat	Latent heat	Sensible heat ($W m^{-2}$)	Latent heat ($W m^{-2}$)	Sensible heat	Latent heat	Sensible heat ($W m^{-2}$)	Latent heat ($W m^{-2}$)	Sensible heat	Latent heat
CN-Din	47.49	58.06	0.84	0.78	37.26	48.3	0.84	0.81	37.21	54.73	0.82	0.74
ID-Pag	52.38	60.26	0.85	0.87	24.37	39.78	0.92	0.96	23.74	40.46	0.91	0.95
PT-Esp	70.09	43.62	0.78	0.77	66.65	39.11	0.82	0.80	63.15	40.24	0.83	0.78
PT-Mi1	55.43	27.12	0.92	0.63	43.01	24.03	0.92	0.71	41.61	22.57	0.92	0.62
Mean	**56.35**	**47.26**	**0.85**	**0.76**	**42.82**	**37.8**	**0.88**	**0.82**	**41.43**	**39.5**	**0.87**	**0.77**

Note: The comparison of model performance with respect to the simulation of sensible heat and latent heat fluxes was relative to the default parameter set (ILS_ORI).

Table S6. dPL parameter calibration results for Evergreen Forest. Numbers shown in bold font exceed the specified lower or upper limits and have been set to the corresponding bound.

Evergreen Forest From dPL	default	CN-Din	ID-Pag	PT-Esp	PT-Mi1
vegh	35	24.26	3.21	20.87	3.88
rlfv	0.1	0.08	0.15	**0.05**	0.14
rlfn	0.45	0.41	0.41	0.34	0.41
tlfn	0.25	0.26	0.19	0.21	0.27
vgcov	1	0.61	0.82	0.22	0.58
cdl	0.11	0.39	0.21	0.35	0.44
chl	0.0274	**0.01**	**0.01**	0.0155	0.0123
vmax0	6.00E-05	5.55E-05	1.55E-05	6.51E-05	**1E-06**
gradm	9	3.82	**1**	6.09	**1**
binter	0.01	0.01	0.05	0.03	0.01
effcon	0.08	**0.01**	**0.01**	0.07	0.08
psicr	-200	-294.71	**-500**	-276.42	-410.23

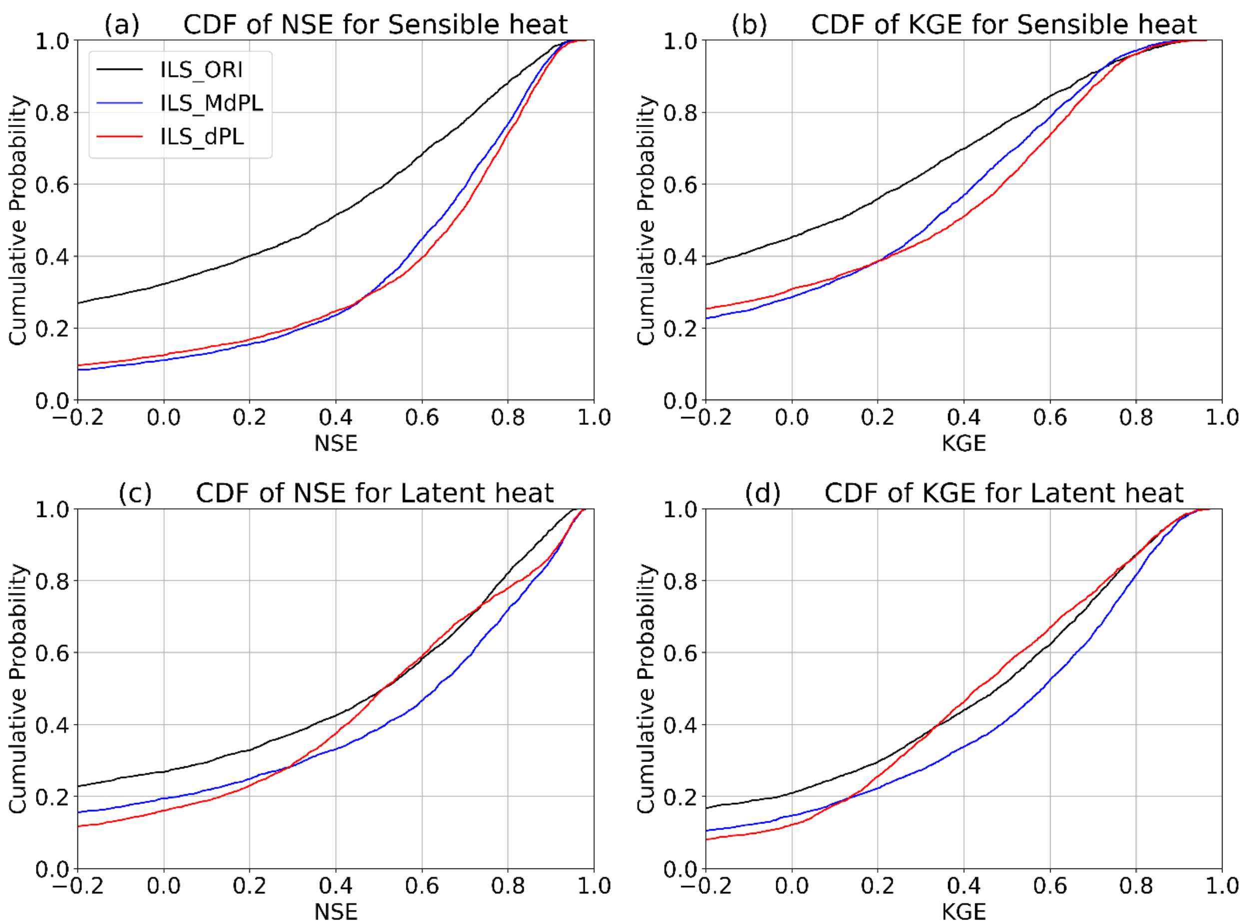

Figure S2. Cumulative distribution functions (CDFs) of daily NSE and KGE for sensible and latent heat outputs across across the four evergreen forest sites. (a) NSE- and (b) KGE-based sensible heat flux performance comparison between multi-task (ILS_MdPL) and single-task (ILS_dPL) surrogate models. (c) NSE- and (d) KGE-based latent heat flux performance comparison between multi-task (ILS_MdPL) and single-task (ILS_dPL) surrogate models. In all panels, curves shifted to the right indicate better performance.

Supplementary Section S2. Transferability analysis using LOOCV

To assess the transferability and generalizability of MdPL-derived parameter calibrations, we separately performed leave-one-out cross-validation (ILS_MdPL_LOOCV) for three PFT classes (Lumumba et al., 2024): evergreen forest (n = 4; PT-Mi1, CN-Din, ID-Pag, PT-Esp), Woodland (n = 5; AR-SLu, DE-Meh, US-Bar, JP-SMF, CN-Cha), and Cultivation land (n = 3; DK-Fou, IE-Ca1, DE-Seh), and reported all monthly KGE results as mean ± STD over the corresponding n experiments. In each LOOCV experiment, one site was held out for evaluation, while the remaining sites within the same PFT class were used for parameter learning. Thereafter, each LOOCV-derived parameter set was used to drive ILS simulations, and the resulting mean ± STD values of monthly KGE for sensible heat and latent heat fluxes were compared against the KGE values obtained for both the original ILS parameterization (ILS_ORI) and site-specific MdPL parameterization (ILS_MdPL), as summarized in **Table S7**.

Table S7. Mean ± STD of KGE values for sensible heat and latent heat fluxes simulated using the: (1) original ILS parameter set (ILS_ORI), (2) leave-one-out–calibrated MdPL (ILS_MdPL_LOOCV), and (3) site-specific MdPL (ILS_MdPL), averaged over LOOCV experiments (n = 4 for Evergreen forest; n = 5 for Woodland; n = 3 for Cultivation sites).

	Sensible heat			Latent heat		
	ILS_ORI	ILS_MdPL_LOOCV	ILS_MdPL	ILS_ORI	ILS_MdPL_LOOCV	ILS_MdPL
Evergreen Forest (n = 4)	-0.05±0.98	0.32±0.40	**0.42**±0.47	0.30±0.40	0.63±0.17	**0.73**±0.05
Woodland (n = 5)	0.29±0.24	0.32±0.18	**0.45**±0.38	**0.65**±0.21	0.54±0.20	0.63±0.18
Cultivation (n = 3)	-0.29±0.66	-0.40±0.39	**0.49**±0.13	-0.38±1.03	-0.65±1.28	**0.42**±0.42

Table S7 shows that transferability was strongly dependent on PFT and target variable. For evergreen forest sites, ILS_MdPL_LOOCV improved mean monthly KGE relative to ILS_ORI for both sensible heat (0.32 vs. -0.05) and latent heat (0.63 vs. 0.30), although it remained below site-specific ILS_MdPL (0.42 and 0.73, respectively). For woodland sites, LOOCV produced a small improvement for sensible heat (0.32 vs. 0.29) but lower performance for latent heat (0.54 vs. 0.65). For cultivation sites, LOOCV yielded lower mean KGE than ILS_ORI for both sensible (-0.40 vs. -0.29) and latent (-0.65 vs. -0.38) heat fluxes, whereas site-specific ILS_MdPL achieved positive mean KGE values (0.49 and 0.42, respectively). These results indicate that transferability was PFT- and variable-dependent and was weakest for cultivation sites (Martin & Isaac, 2018).

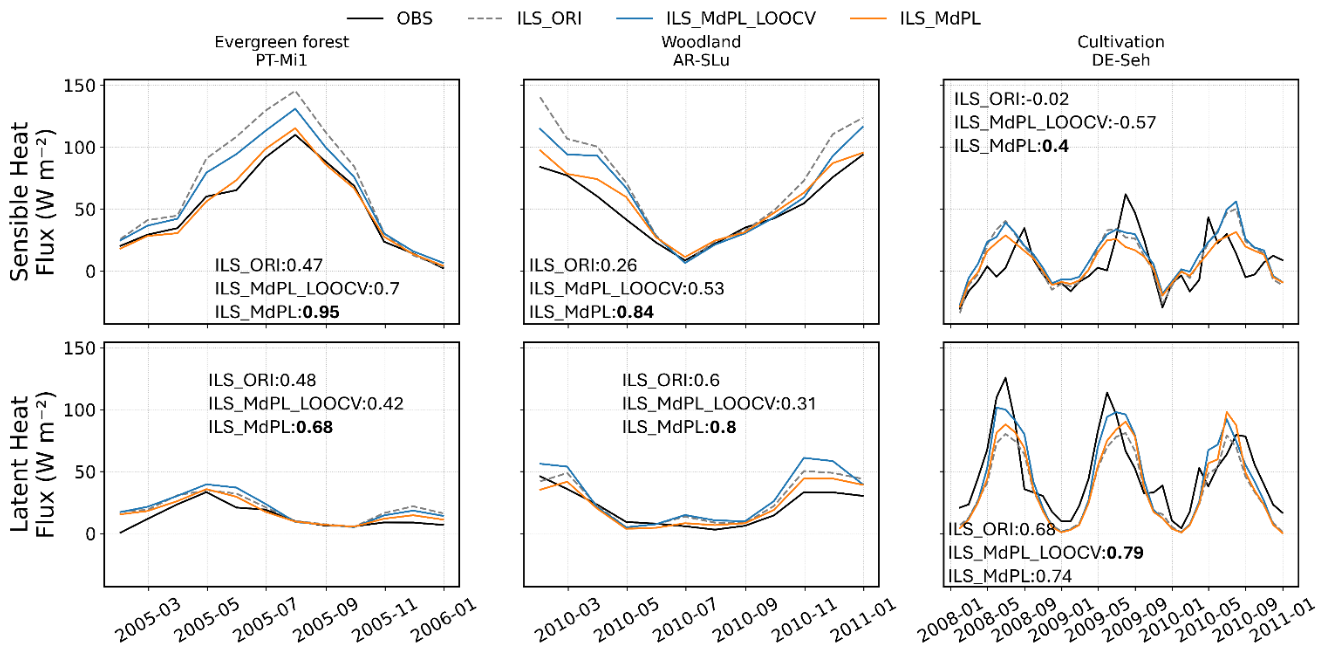

Figure S3. Monthly time series of sensible heat (top) and latent heat (bottom) simulated by ILS_ORI, ILS_MdPL_LOOCV, and ILS_MdPL across three representative PFT sites. The corresponding Kling–Gupta efficiency (KGE) values are shown within each panel for reference.

Fig. S3 provides representative monthly time series for one site from each PFT class and illustrates the site-level variability underlying the mean results in Table S7. The results do not show a universal performance ranking among ILS_ORI, ILS_MdPL_LOOCV, and ILS_MdPL. For the evergreen forest and woodland examples, LOOCV improved some sensible heat simulations but did not consistently outperform ILS_ORI for latent heat. For the cultivation example, both ILS_ORI and ILS_MdPL_LOOCV showed poor performance, whereas site-specific ILS_MdPL achieved a clear improvement. Together with Table S7, these examples show that parameter transfer was more successful for some forest-site and flux combinations but was not uniformly beneficial, while transferability was particularly limited for cultivation sites. This pattern may reflect the greater heterogeneity of managed systems, where management practices, phenological variability, and site-specific conditions can reduce cross-site parameter transferability (Hoppe et al., 2024).

Despite these promising results, this study had some limitations. First, the current calibration primarily targeted energy fluxes (sensible and latent heat), whereas other critical LSM outputs, such as surface temperature and soil moisture content, were not evaluated. Additionally, the ILS model employed fixed PFT parameters, which may not adequately reflect temporal variability in vegetation, particularly for cultivation sites influenced by human management activities. The increased complexity of the multitask surrogate model, even though beneficial in terms of performance, may also pose challenges in terms of computational cost and overfitting in data-scarce regions. Finally, the PFT- and variable-dependent LOOCV results highlighted a key limitation of MdPL: the benefits of cross-site parameter transfer depend on the similarity of site characteristics represented in the training data. Transferability was particularly limited for cultivation sites in this experiment.