

Document S1 – Supplemental PDF

- Supplemental Tables S1-S6
- Supplemental Figures S1-S9
- Supplemental Notes S1-S4

Supplementary Tables

Table S1 – Reference measurements (RMs). Summary of RMs used from the GROUNDED-EO¹ dataset. Ecoregions refer to the classification of Dinerstein et al.², with the following biome numbers³: 1, Tropical & Subtropical Moist Broadleaf Forests; 2, Tropical & Subtropical Dry Broadleaf Forests; 3, Tropical & Subtropical Coniferous Forests; 4, Temperate Broadleaf & Mixed Forests; 5, Temperate Conifer Forests; 6, Boreal Forests/Taiga; 8, Temperate Grasslands, Savannas & Shrublands; 11, Tundra; 12, Mediterranean Forests, Woodlands & Scrub; 13, Deserts & Xeric Shrublands. The number of plots refers to the number of distinct sampling plots within each field site. Land cover values are adopted from Brown et al.¹, which follow the US National Land Cover Database (NLCD)⁴.

Name	Network	ECO_ID	Biome	Land Cover	#Plots	#Measurements
Bilos	ICOS	648	4	evergreenForest	23	71
Brasschaat	ICOS	664	4	evergreenForest	25	139
Castelporziano 2	ICOS	795	12	mixedForest	25	63
Davos	ICOS	689	5	evergreenForest	24	88
Font-Blanche	ICOS	799	12	evergreenForest	20	31
Fontainebleau-Barbeau	ICOS	664	4	deciduousForest	24	153
Hesse	ICOS	686	4	deciduousForest	4	15
Hohes Holz	ICOS	654	4	deciduousForest	5	149
Hurdal	ICOS	717	6	evergreenForest	11	15
Hyltemossa	ICOS	647	4	evergreenForest	20	124
Hyytiala	ICOS	717	6	evergreenForest	24	106
Norunda	ICOS	679	4	evergreenForest	24	84
Puechabon	ICOS	799	12	evergreenForest	22	50
Renon	ICOS	689	5	evergreenForest	23	31
San Rossore 2	ICOS	795	12	evergreenForest	21	67
Sodankyla	ICOS	717	6	evergreenForest	24	51
Soroe	ICOS	647	4	deciduousForest	24	67
Svartberget	ICOS	717	6	evergreenForest	22	45
Tharandt	ICOS	686	4	evergreenForest	24	31
Vielsalm	ICOS	686	4	mixedForest	23	80
Abbey Road	NEON	352	5	evergreenForest	18	196
Bartlett Experimental Forest	NEON	338	4	deciduousForest	21	182
Blandy Experimental Farm	NEON	331	4	cultivatedCrops	22	240
Caribou-Poker Creeks Research Watershed	NEON	416	11	evergreenForest	23	111
Central Plains Experimental Range	NEON	402	8	grasslandHerbaceous	23	301
Dakota Field Site	NEON	396	8	grasslandHerbaceous	23	234
Dead Lake	NEON	399	8	deciduousForest	23	194
Delta Junction	NEON	375	6	evergreenForest	3	91
Disney Wilderness Preserve	NEON	399	8	pastureHay	16	234
Great Smoky Mountains National Park	NEON	331	4	deciduousForest	23	246
Guanica Forest	NEON	543	2	evergreenForest	20	250
Harvard Forest	NEON	338	4	mixedForest	8	177
Healy	NEON	405	11	shrubScrub	23	96
Jones Ecological Research Center	NEON	399	8	evergreenForest	21	232
Jornada	NEON	428	13	shrubScrub	21	256
Konza Prairie Agroecosystem	NEON	392	8	cultivatedCrops	24	198
Konza Prairie Biological Station	NEON	392	8	grasslandHerbaceous	24	273
Lajas Experimental Station	NEON	543	2	pastureHay	3	270
Lenoir Landing Site	NEON	399	8	deciduousForest	23	198
Lower Teakettle	NEON	366	5	evergreenForest	20	92
Lyndon B. Johnson National Grassland	NEON	390	4	deciduousForest	25	261
Marvin Klemme Range Research Station	NEON	389	8	shrubScrub	20	233
Moab	NEON	429	13	shrubScrub	23	181
Mountain Lake Biological Station	NEON	331	4	deciduousForest	23	195
Niwot Ridge Mountain Research Station	NEON	353	5	grasslandHerbaceous	22	154

Name	Network	ECO_ID	Biome	Land Cover	#Plots	#Measurements
North Sterling	NEON	402	8	cultivatedCrops	19	207
Northern Great Plains Research Laboratory	NEON	396	8	grasslandHerbaceous	23	225
Oak Ridge	NEON	331	4	deciduousForest	28	253
Onaqui Ault	NEON	430	13	shrubScrub	23	233
Ordway Swisher Biological Station	NEON	399	8	evergreenForest	30	259
Pu'u Maka'ala Natural Area Reserve	NEON	623	1	evergreenForest	23	260
Rocky Mountain National Park	NEON	353	5	evergreenForest	25	185
San Joaquin Experimental Range	NEON	423	12	evergreenForest	22	293
Santa Rita	NEON	435	13	shrubScrub	20	291
Smithsonian Conservation Biology Institute	NEON	331	4	deciduousForest	27	272
Smithsonian Environmental Research Center	NEON	399	8	deciduousForest	23	213
Soaproot Saddle	NEON	366	5	evergreenForest	23	151
Steigerwaldt Land Services	NEON	344	4	deciduousForest	13	160
Talladega National Forest	NEON	399	8	evergreenForest	17	186
Toolik Field Station	NEON	411	11	grasslandHerbaceous	8	86
Treehaven	NEON	344	4	deciduousForest	13	166
Underc	NEON	344	4	wetlands	23	189
University of Kansas Field Station	NEON	388	8	deciduousForest	24	274
Utqiaġvik	NEON	407	11	wetlands	20	52
Wind River Experimental Forest	NEON	352	5	evergreenForest	27	166
Woodworth	NEON	396	8	grasslandHerbaceous	25	249
Yellowstone National Park	NEON	367	5	evergreenForest	17	83
Boyagin	TERN	205	12	evergreenForest	1	12
Gingin	TERN	206	12	evergreenForest	2	10
Robson Creek Rainforest	TERN	150	1	evergreenForest	1	3
Warra Tall Eucalypt	TERN	179	4	evergreenForest	1	5

Table S2 – PROSAIL input parameter ranges. Description of predefined parameter ranges and distributions used to generate PROSAIL lookup tables.

$N(\mu, \sigma)|_{min}^{max}$: normal distribution with mean μ and standard deviation σ , truncated at min and max . $U(min, max)$: uniform distribution. $C(x)$: constant value.

Parameters labeled *OPTUNA* indicate distributions optimized during model training using Bayesian optimization (see Table S3). Predefined ranges follow the SNAP toolbox⁵, or implementations used by Estevez et al.⁶, Kovacs et al.⁷ and Wan et al.⁸. Foliar co-distribution uses the SNAP parameterization, except for plant foliar traits derived from co-distributions based on in-situ NEON measurements⁹ (see methods).

Parameter	Abbr.	Unit	SNAP	Foliar Codis.	Estevez et al.	Kovacs et al.	Wan et al.
Leaf variables: PROSPECT-D							
Leaf chlorophyll content	CHL	$\mu\text{g cm}^{-2}$	$N(35, 30) _5^{75}$	<i>Gaussian Copula</i>	$N(35, 30) _5^{75}$	$U(0, 100)$	$N(50, 20) _{15}^{100}$
Leaf carotenoid content	CAR	$\mu\text{g cm}^{-2}$	$CHL / 7$	<i>Gaussian Copula</i>	$CHL / 7$	$N(10, 10) _0^{20}$	$N(7.5, 3.5) _0^{15}$
Leaf mass per area	LMA	g cm^{-2}	$N(0.005, 0.001) _{0.001}^{0.03}$	<i>Gaussian Copula</i>	$N(0.005, 0.001) _{0.001}^{0.03}$	$N(0.005, 0.003) _{0.002}^{0.02}$	$N(0.015, 0.005) _{0.001}^{0.03}$
Equivalent water thickness	EWT	cm	NA	<i>Gaussian Copula</i>	$N(0.02, 0.01) _{0.002}^{0.05}$	$N(0.012, 0.006) _{0.005}^{0.035}$	NA
Relative water content	Cw.rel	–	$N(0.75, 0.08) _{0.6}^{0.85}$	NA	NA	NA	$U(0.55, 0.95)$
Leaf anthocyanin content	ANT	$\mu\text{g cm}^{-2}$	$C(0)$	$C(0)$	$C(0)$	$C(0)$	$C(0)$
Leaf structure parameter	N	–	$N(1.5, 0.03) _{1.2}^{1.8}$	$N(1.5, 0.03) _{1.2}^{1.8}$	$U(1.3, 2.5)$	$N(1.5, 0.05) _{1.0}^{2.7}$	$N(1.75, 0.5) _{1.0}^{2.5}$
Fraction of senescent leaves	BROWN	–	$N(0.0, 0.3) _0^2$	$N(0.0, 0.3) _0^2$	$C(0)$	$C(0)$	$U(0.0, 0.2)$
Canopy variables: 4SAIL							
Leaf area index	LAI	$\text{m}^2 \text{m}^{-2}$	<i>OPTUNA</i>	<i>OPTUNA</i>	<i>OPTUNA</i>	<i>OPTUNA</i>	<i>OPTUNA</i>
Hot spot parameter	q	m m^{-1}	$N(0.2, 0.5) _{0.1}^{0.5}$	$N(0.2, 0.5) _{0.1}^{0.5}$	$C(0.01)$	$C(0.01)$	$C(0.01)$
Average leaf angle	LIDFa	°	$N(60, 30) _{30}^{80}$	$N(60, 30) _{30}^{80}$	$U(40, 70)$	$U(40, 70)$	$N(55, 10) _{30}^{80}$
Leaf distribution function	Type-LIDF	–	2	2	2	2	2
Observation / Illumination conditions: 4SAIL							
Linear comb. of dark and bright soil spectra	psoil	–	$U(0.0, 1.0)$	$N(0.25, 0.6) _{0.0}^{1.0}$	$U(0.0, 1.0)$	$U(0.0, 1.0)$	$U(0.0, 1.0)$
Sun zenith angle	tts	°	<i>GEE S2 metadata</i>	<i>GEE S2 metadata</i>	<i>GEE S2 metadata</i>	<i>GEE S2 metadata</i>	$U(15, 80)$
View zenith angle	tto	°	<i>GEE S2 metadata</i>	<i>GEE S2 metadata</i>	$C(0)$	$U(-10, 10)$	$U(0, 12)$
Relative azimuth angle	psi	°	<i>GEE S2 metadata</i>	<i>GEE S2 metadata</i>	$C(0)$	$C(180)$	$U(0, 180)$

Table S3 – Model-training parameter search spaces. Hybrid RTM inversion optimization parameters sampled by OPTUNA¹⁰ and MLP hyperparameter options evaluated during model training. OPTUNA controlled the simulated training-data design, the composition and modifications to the training LUT, and feature-engineering choices. For each OPTUNA trial, we selected MLP hyperparameters using a nested successive-halving grid search. Table S2 lists details on predefined PROSAIL input ranges and distributions (parameter setup).

Step	Parameter	Choices / Range	Description
<i>OPTUNA parameter sampling</i>			
PROSAIL parameters	num_spectra	[2500, 50000], step size = 2500	Number of samples in the simulated training database
	parameter_setup	['estevez_2022', 'foliar_codistribution', 'kovacs_2023', 'snap_atbd', 'wan_2024_lai']	Selection of predefined PROSAIL input ranges and distributions
	lai_mean	[0, 5], step size = 0.2	Mean of the LAI distribution
	lai_std	[0, 5], step size = 0.2	Standard deviation of the LAI distribution
	rsoil_emit_insitu	['emit', 'insitu']	Source for custom soil background reflectance modification
	rsoil_fraction	[0.1, 1.0], step size = 0.1	Proportion of samples simulated using custom soil reflectance
LUT modifications	additive_noise	[0.0%, 3.0%], step size = 0.5%	Percentage of additive noise added to the simulated reflectances
	multiplicative_noise	[0.0%, 10.0%], step size = 1.0%	Percentage of multiplicative noise added to the simulated reflectances
	p_baresoil_insitu	[0.00, 0.10], step size = 0.01	Proportion of bare soil spectra added from the in-situ OSSL database
	p_baresoil_s2	[0.00, 0.10], step size = 0.01	Proportion of bare soil spectra added from S2
	p_baresoil_emit	[0.00, 0.10], step size = 0.01	Proportion of bare soil spectra added from EMIT Hyperspectral
	p_urban_s2	[0.00, 0.05], step size = 0.005	Proportion of urban spectra added from S2
	p_snow_ice_s2	[0.00, 0.05], step size = 0.005	Proportion of snow and ice spectra added from S2
Feature engineering	nirv_norm	[True, False]	Whether to apply NIRv normalization to reflectances
	transform_target	FAPAR, FCOVER: ['log1p', 'logit', 'None'] LAI: ['log1p', 'standard', 'None']	Transformation method applied to target variables
<i>Nested MLP hyperparameter grid search</i>			
MLP hyperparameters	hidden_layer_sizes	[[5], [10], [5, 5], [10, 5]]	Configurations for hidden layer sizes in the fully connected MLP
	activation	['relu', 'tanh']	Activation function for nodes in hidden layers
	alpha	[0.001, 0.01, 0.1, 1.0]	L2 regularization strength for weight parameters
	learning_rate	['constant', 'adaptive']	Learning rate strategy for model training

Table S4 – S2BIOPHYS model performance. Comparison of S2BIOPHYS and SL2P retrieval performance for LAI, FAPAR and FCOVER across land cover types. For S2BIOPHYS, we report performance for ensemble-averaged retrievals (“mean”) and stacked out-of-sample retrievals (“SOS”), with the lowest RMSE highlighted in bold. The methods section defines the evaluation metrics.

Model	NLCD Land Cover	N	LAIe							FAPAR							FCOVER						
			MAE	R ²	R ² _{global}	RMSE	nRMSE	ME	UAR	MAE	R ²	R ² _{global}	RMSE	nRMSE	ME	UAR	MAE	R ²	R ² _{global}	RMSE	nRMSE	ME	UAR
S2BIOPHYS	All (mean)	11038	0.43	0.77	NaN	0.60	0.45	0.07	0.90	0.10	0.78	NaN	0.15	0.33	0.02	0.70	0.10	0.79	NaN	0.15	0.36	0.01	0.68
	All (SOS)	11038	0.45	0.75	NaN	0.63	0.47	0.07	0.88	0.11	0.75	NaN	0.16	0.36	0.01	0.67	0.11	0.77	NaN	0.15	0.37	0.01	0.67
	cultivatedCrops	498	0.27	0.43	0.78	0.50	0.99	0.14	0.94	0.08	0.78	0.87	0.13	0.53	-0.01	0.76	0.08	0.74	0.85	0.13	0.66	0.01	0.74
	deciduousForest	3010	0.52	0.72	0.82	0.69	0.30	-0.20	0.87	0.11	0.66	0.80	0.15	0.23	-0.03	0.74	0.12	0.68	0.81	0.15	0.24	-0.05	0.70
	evergreenForest	3286	0.55	0.47	0.47	0.69	0.50	0.22	0.85	0.14	0.48	0.50	0.19	0.37	0.05	0.53	0.14	0.51	0.52	0.18	0.40	0.04	0.53
	grasslandHerbaceous	1685	0.30	0.40	0.82	0.45	0.97	0.22	0.95	0.08	0.76	0.86	0.12	0.50	0.02	0.73	0.08	0.78	0.88	0.11	0.56	0.04	0.74
	mixedForest	501	0.43	0.68	0.82	0.56	0.25	-0.23	0.92	0.07	0.77	0.89	0.10	0.15	0.00	0.86	0.08	0.78	0.87	0.11	0.18	-0.01	0.79
	pastureHay	519	0.39	0.24	0.66	0.50	0.72	0.30	0.96	0.10	0.68	0.71	0.14	0.36	-0.00	0.64	0.12	0.66	0.68	0.15	0.44	0.00	0.58
	shrubScrub	1293	0.14	0.79	0.94	0.31	1.26	0.08	0.96	0.04	0.79	0.95	0.09	0.92	0.03	0.90	0.04	0.84	0.96	0.08	0.89	0.02	0.91
	wetlands	246	0.59	0.58	0.63	0.74	0.42	0.21	0.84	0.13	0.66	0.72	0.16	0.28	0.05	0.59	0.13	0.70	0.74	0.16	0.30	0.03	0.58
SL2P	All	11038	0.53	0.68	NaN	0.70	0.53	0.30	0.85	0.14	0.69	NaN	0.18	0.39	-0.04	0.56	0.12	0.74	NaN	0.16	0.40	0.00	0.57
	cultivatedCrops	498	0.49	-0.46	0.44	0.80	1.60	0.44	0.86	0.09	0.76	0.86	0.13	0.56	0.04	0.70	0.09	0.64	0.80	0.15	0.77	0.07	0.69
	deciduousForest	3010	0.50	0.74	0.84	0.66	0.28	0.07	0.89	0.15	0.48	0.69	0.19	0.28	-0.11	0.59	0.12	0.67	0.80	0.16	0.25	-0.05	0.65
	evergreenForest	3286	0.65	0.32	0.32	0.78	0.57	0.40	0.78	0.18	0.30	0.33	0.21	0.43	-0.08	0.39	0.16	0.40	0.41	0.20	0.44	-0.02	0.38
	grasslandHerbaceous	1685	0.53	-0.67	0.49	0.75	1.62	0.50	0.85	0.10	0.68	0.82	0.14	0.57	0.05	0.60	0.11	0.59	0.78	0.15	0.76	0.09	0.54
	mixedForest	501	0.42	0.72	0.85	0.52	0.23	0.01	0.95	0.14	0.31	0.66	0.17	0.26	-0.12	0.56	0.10	0.70	0.82	0.12	0.21	-0.04	0.74
	pastureHay	519	0.72	-0.99	0.11	0.82	1.17	0.70	0.80	0.11	0.67	0.70	0.14	0.37	0.04	0.54	0.14	0.58	0.61	0.17	0.49	0.08	0.42
	shrubScrub	1293	0.23	0.61	0.89	0.43	1.74	0.21	0.93	0.07	0.74	0.94	0.10	1.02	0.06	0.81	0.06	0.74	0.93	0.10	1.15	0.05	0.83
	wetlands	246	0.67	0.49	0.55	0.81	0.46	0.43	0.76	0.14	0.62	0.68	0.17	0.30	-0.04	0.49	0.13	0.67	0.72	0.17	0.32	0.02	0.57
GROUNDED-EO GPR	All	11038	Model not comparable (LAI instead of LAIe)							0.10	0.79	NaN	0.15	0.32	0.02	0.72	Model not available						
	cultivatedCrops	498								0.09	0.73	0.84	0.14	0.60	0.01	0.72							
	deciduousForest	3010								0.11	0.62	0.78	0.16	0.24	-0.01	0.74							
	evergreenForest	3286								0.13	0.56	0.57	0.17	0.34	0.04	0.59							
	grasslandHerbaceous	1685								0.07	0.77	0.87	0.12	0.49	0.01	0.77							
	mixedForest	501								0.07	0.75	0.88	0.10	0.16	0.02	0.89							
	pastureHay	519								0.10	0.70	0.72	0.13	0.35	0.00	0.66							
	shrubScrub	1293								0.04	0.82	0.96	0.08	0.85	0.02	0.90							
	wetlands	246								0.12	0.65	0.71	0.17	0.29	0.06	0.61							

Table S5 – Biome–land-cover specific recoding. Combination of biomes³ (in bold) and simplified NLCD land-cover labels⁴ (in italics) used to define the eight validation groups (see Note S3). Counts indicate the number of reference measurements, ecoregions, and sites represented in each group.

Group	Validation group	Biome: <i>land-cover</i> combination	Reference N	Ecoregions	Sites
1	Cold-climate evergreen forests	Boreal Forests/Taiga, Tundra: <i>Evergreen Forest</i>	377	4	7
2	Open tundra vegetation	Tundra: <i>Grassland, Shrubland, Wetlands</i>	238	4	4
3	Arid shrublands	Deserts & Xeric Shrubland: <i>Shrubland</i>	949	4	4
4	Temperate non-forest vegetation	Temperate Broadleaf & Mixed Forests, Temperate Grasslands, Temperate Savannas & Shrublands: <i>Cropland, Grassland, Shrubland, Wetlands</i>	2762	16	28
5	Temperate deciduous & mixed forests	Temperate Broadleaf & Mixed Forests, Temperate Conifer Forests, Temperate Grasslands, Savannas & Shrublands: <i>Deciduous/Mixed Forest</i>	3319	15	26
6	Temperate evergreen forests	Temperate Broadleaf & Mixed Forests, Temperate Conifer Forests, Temperate Grasslands, Savannas & Shrublands: <i>Evergreen Forest</i>	2034	17	25
7	Tropical evergreen forests	Tropical & Subtropical Dry Broadleaf Forests, Tropical & Subtropical Moist Broadleaf Forests: <i>Evergreen Forest</i>	513	4	3
8	Mediterranean forests & woodlands	Mediterranean Forests, Woodlands & Scrub: <i>Deciduous/Mixed Forest, Evergreen Forest</i>	504	5	7

Table S6 – Performance of biome–land-cover-specific and generic models. Metrics are calculated for ecoregion-blocked out-of-sample retrieval performance for biome–land-cover-specific (Table S5) and generic models. RMSE is specific minus generic, with negative values indicating lower error for the specific model. Group-level R^2 uses the mean reference value within each group in the denominator, whereas R^2_{global} uses the mean across all groups. For pooled results, R^2 and R^2_{global} are identical.

Variable	Validation group	N	RMSE <i>specific</i>	RMSE <i>generic</i>	Δ RMSE	R^2 <i>specific</i>	R^2 <i>generic</i>	R^2_{global} <i>specific</i>	R^2_{global} <i>generic</i>
FAPAR	Cold-climate evergreen forests	377	0.467	0.224	+0.244	-1.930	0.329	-1.530	0.421
	Open tundra vegetation	238	0.223	0.309	-0.086	-0.405	-1.705	0.510	0.057
	Arid shrublands	949	0.046	0.066	-0.019	0.153	-0.705	0.988	0.977
	Temperate non-forest vegetation	2762	0.122	0.133	-0.012	0.808	0.769	0.860	0.832
	Temperate broadleaf & mixed forests	3319	0.145	0.144	+0.000	0.687	0.688	0.822	0.823
	Temperate evergreen forests	2034	0.212	0.205	+0.007	0.308	0.354	0.375	0.417
	Tropical evergreen forests	513	0.229	0.185	+0.044	-0.310	0.145	-0.304	0.148
	Mediterranean forests & woodlands	504	0.183	0.203	-0.020	0.421	0.287	0.456	0.329
	Pooled across validation groups	10696	0.179	0.164	+0.015	0.693	0.741	–	–
FCOVER	Cold-climate evergreen forests	377	0.389	0.226	+0.162	-1.199	0.255	-0.975	0.331
	Open tundra vegetation	238	0.123	0.237	-0.114	0.636	-0.347	0.828	0.366
	Arid shrublands	949	0.038	0.031	+0.007	0.385	0.596	0.991	0.994
	Temperate non-forest vegetation	2762	0.114	0.113	+0.001	0.820	0.823	0.872	0.873
	Temperate broadleaf & mixed forests	3319	0.139	0.144	-0.005	0.737	0.716	0.844	0.832
	Temperate evergreen forests	2034	0.187	0.185	+0.002	0.457	0.466	0.495	0.504
	Tropical evergreen forests	513	0.207	0.194	+0.013	-0.099	0.035	-0.098	0.036
	Mediterranean forests & woodlands	504	0.179	0.182	-0.003	0.406	0.384	0.435	0.414
	Pooled across validation groups	10696	0.159	0.151	+0.008	0.748	0.773	–	–
LAle	Cold-climate evergreen forests	377	1.672	0.887	+0.785	-1.444	0.312	-1.390	0.327
	Open tundra vegetation	238	0.307	0.890	-0.582	0.250	-5.286	0.926	0.376
	Arid shrublands	949	0.073	0.098	-0.025	0.161	-0.502	0.997	0.994
	Temperate non-forest vegetation	2762	0.479	0.504	-0.025	0.676	0.641	0.810	0.789
	Temperate broadleaf & mixed forests	3319	0.657	0.693	-0.036	0.727	0.696	0.835	0.816
	Temperate evergreen forests	2034	0.698	0.744	-0.046	0.469	0.396	0.477	0.405
	Tropical evergreen forests	513	1.041	0.722	+0.318	-0.620	0.220	-0.620	0.220
	Mediterranean forests & woodlands	504	0.685	0.627	+0.058	0.247	0.369	0.424	0.517
	Pooled across validation groups	10696	0.679	0.640	+0.039	0.710	0.743	–	–

Supplementary Figures

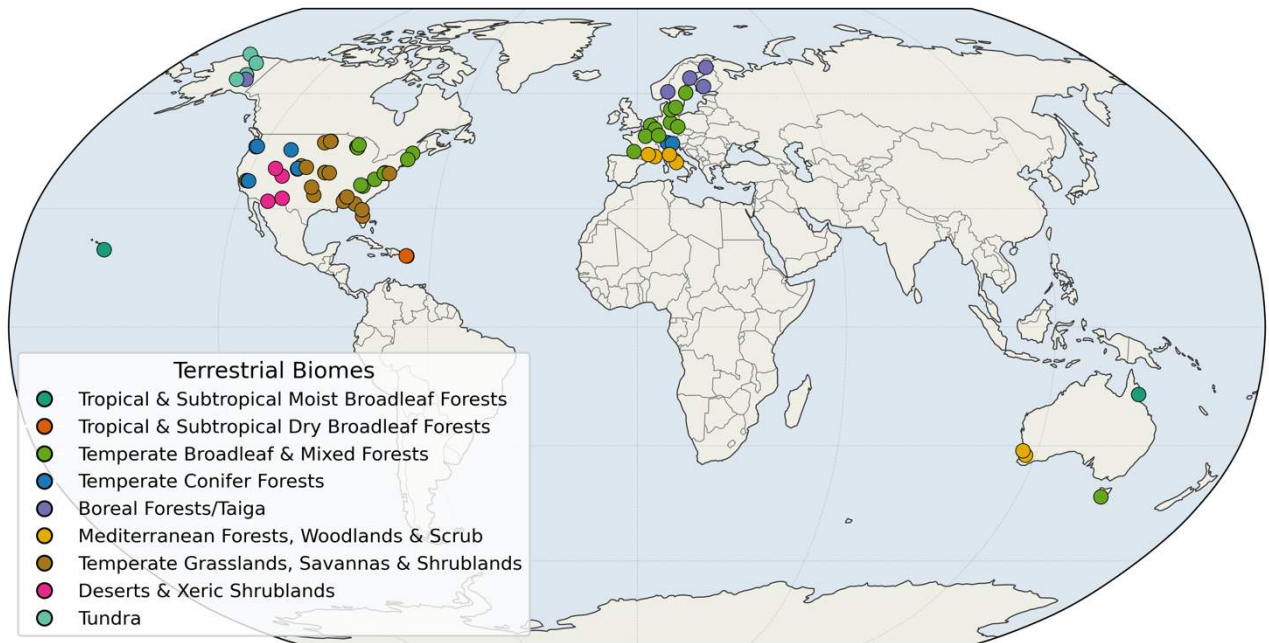


Figure S1 – Distribution of in-situ validation sites. Validation sites (Table S1) are situated in the US (NEON), Europe (ICOS), and Australia (TERN). The centroid of each field site is plotted and colored by the corresponding terrestrial biome³.

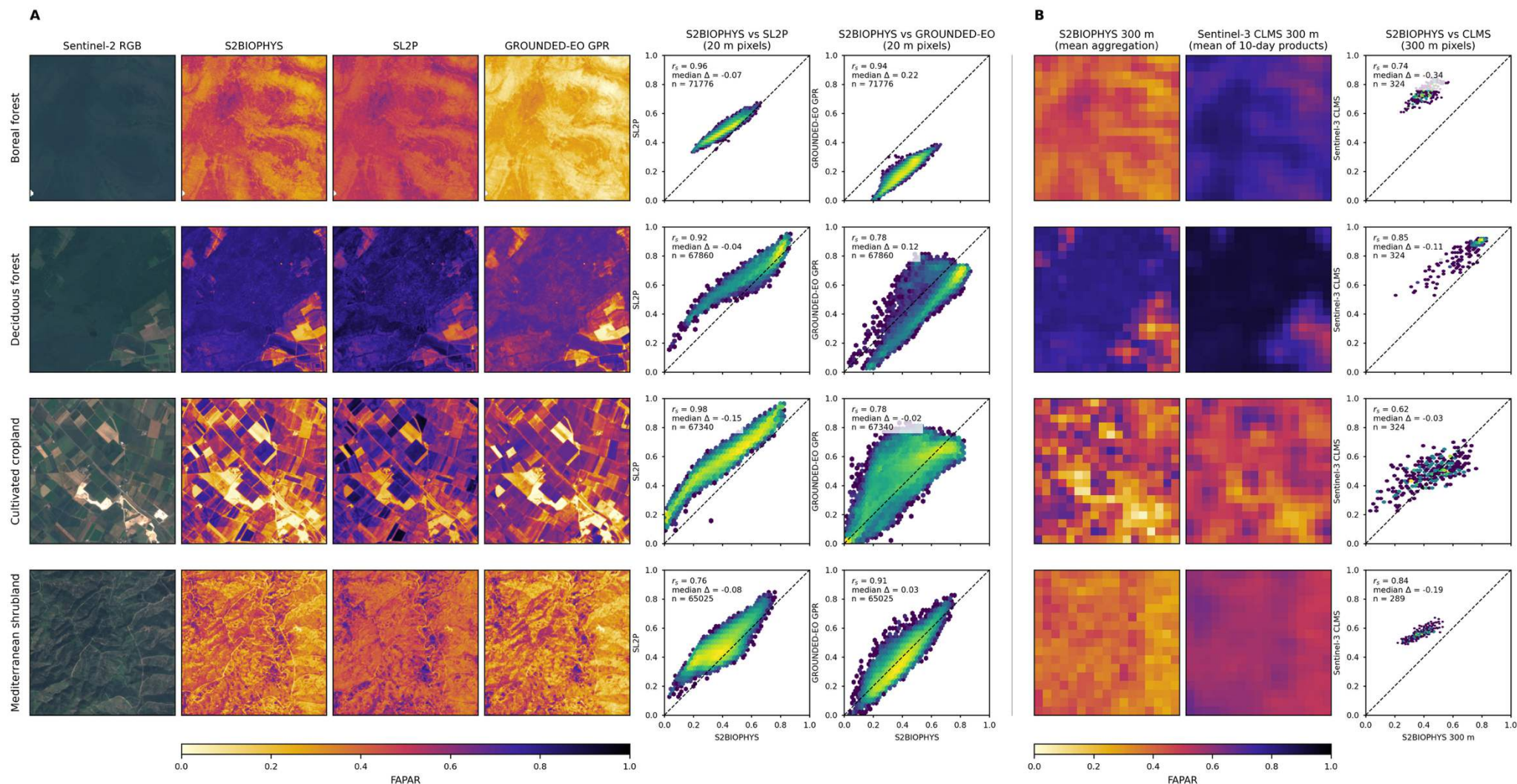


Figure S2 – Map-to-map comparison of FAPAR retrievals. Representative 5×5 km landscapes span boreal forest (Russia; 60.00°N, 125.36°E), deciduous forest (Hainich, Germany; 51.08°N, 10.45°E), cultivated cropland (Beauce, France; 48.30°N, 1.70°E), and Mediterranean shrubland (Sierra Morena, Spain; 38.30°N, 3.80°W). (A) Sentinel-2 RGB imagery and 20 m retrievals during a 1-month period from S2BIOPHYS, SL2P, and GROUNDED-EO GPR, together with pixel-level relationships between S2BIOPHYS and each benchmark method. (B) S2BIOPHYS retrievals are spatially mean-aggregated to 300 m and compared with Sentinel-3 CLMS 300 m, calculated as the temporal mean of all 10-day products within the same period. Pixel-level relationships are shown as hexbin density plots annotated with Spearman's rank correlation coefficient (r_s) and the median signed difference (*median* Δ).

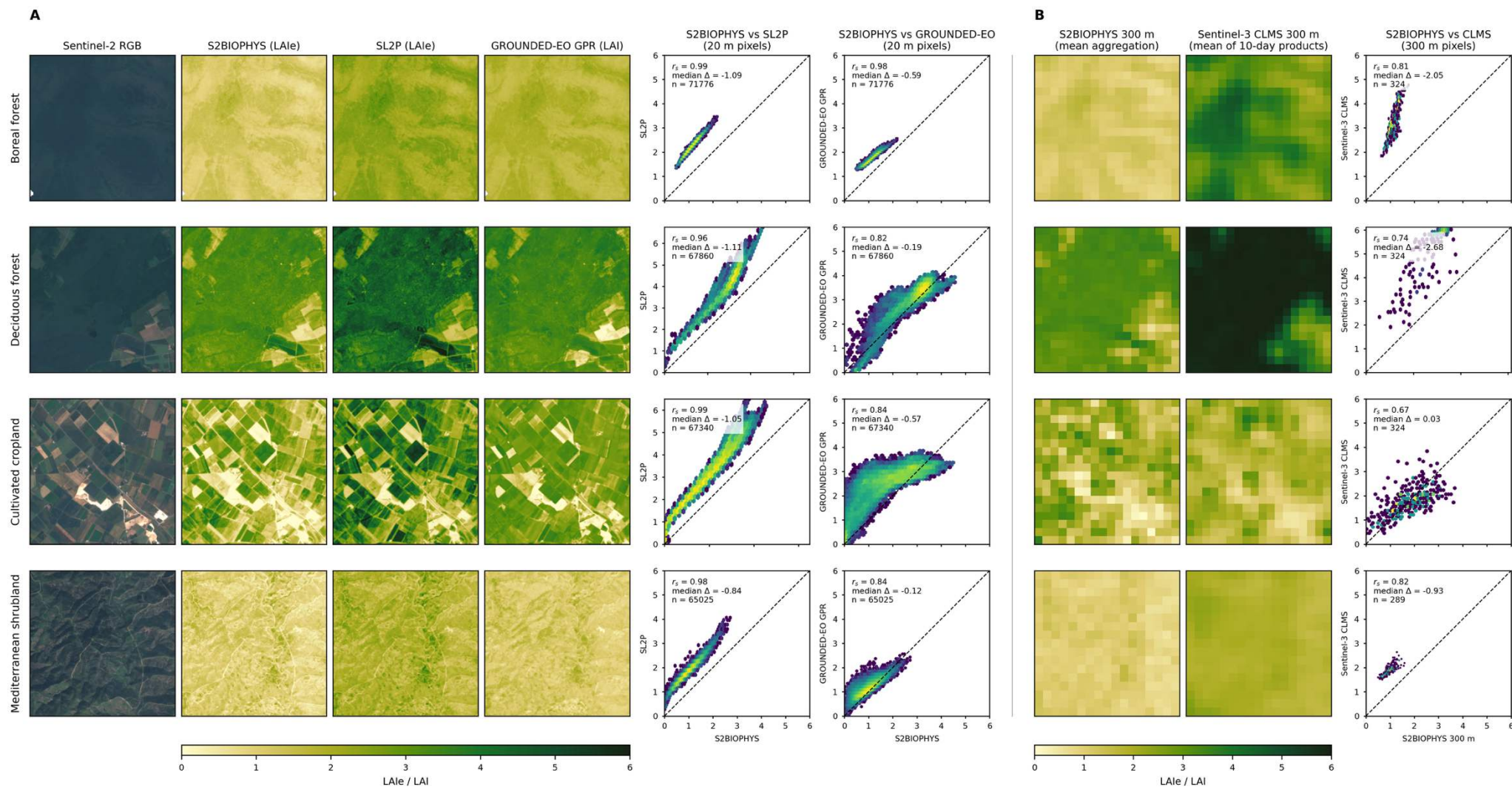


Figure S3 – Map-to-map comparison of LAle retrievals. LAle retrievals from S2BIOPHYS and SL2P and LAI retrievals from GROUNDED-EO GPR are shown for the same landscapes as in Figure S2. We show differences between S2BIOPHYS and the benchmark methods alongside pixel-level relationships. Because GROUNDED-EO GPR and Sentinel-3 CLMS estimate LAI rather than LAle, this comparison primarily assesses correspondence in spatial patterns rather than magnitude. Hexbin plots show pixel-level relationships, Spearman's rank correlation (r_s), and median signed difference (median Δ).

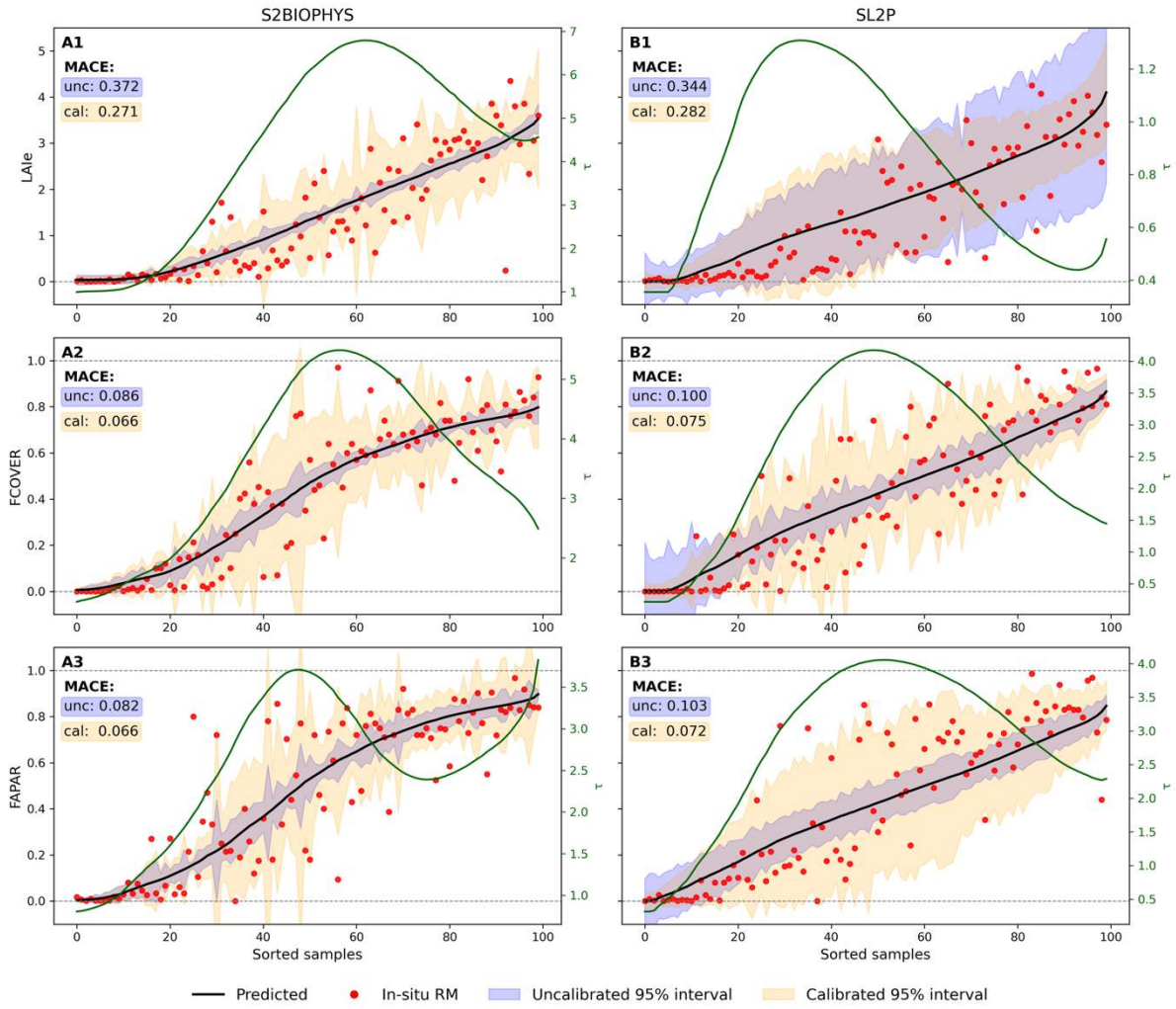


Figure S4 – Uncertainty recalibration. Panels **A1–A3** show results for S2BIOPHYS and panels **B1–B3** for SL2P for LAle (1), FCOVER (2), and FAPAR (3). Predicted values are shown by the black line and in-situ reference measurements (RMs) by red points. Blue and orange shading denotes the uncalibrated and recalibrated 95% prediction intervals, respectively. The green line denotes the learned smooth scaling function, $\tau(\hat{y})$, on the secondary y-axis. Each panel reports MACE before (unc) and after recalibration (cal) across all RMs. We only show 100 samples sorted by predicted values for visualization purposes.

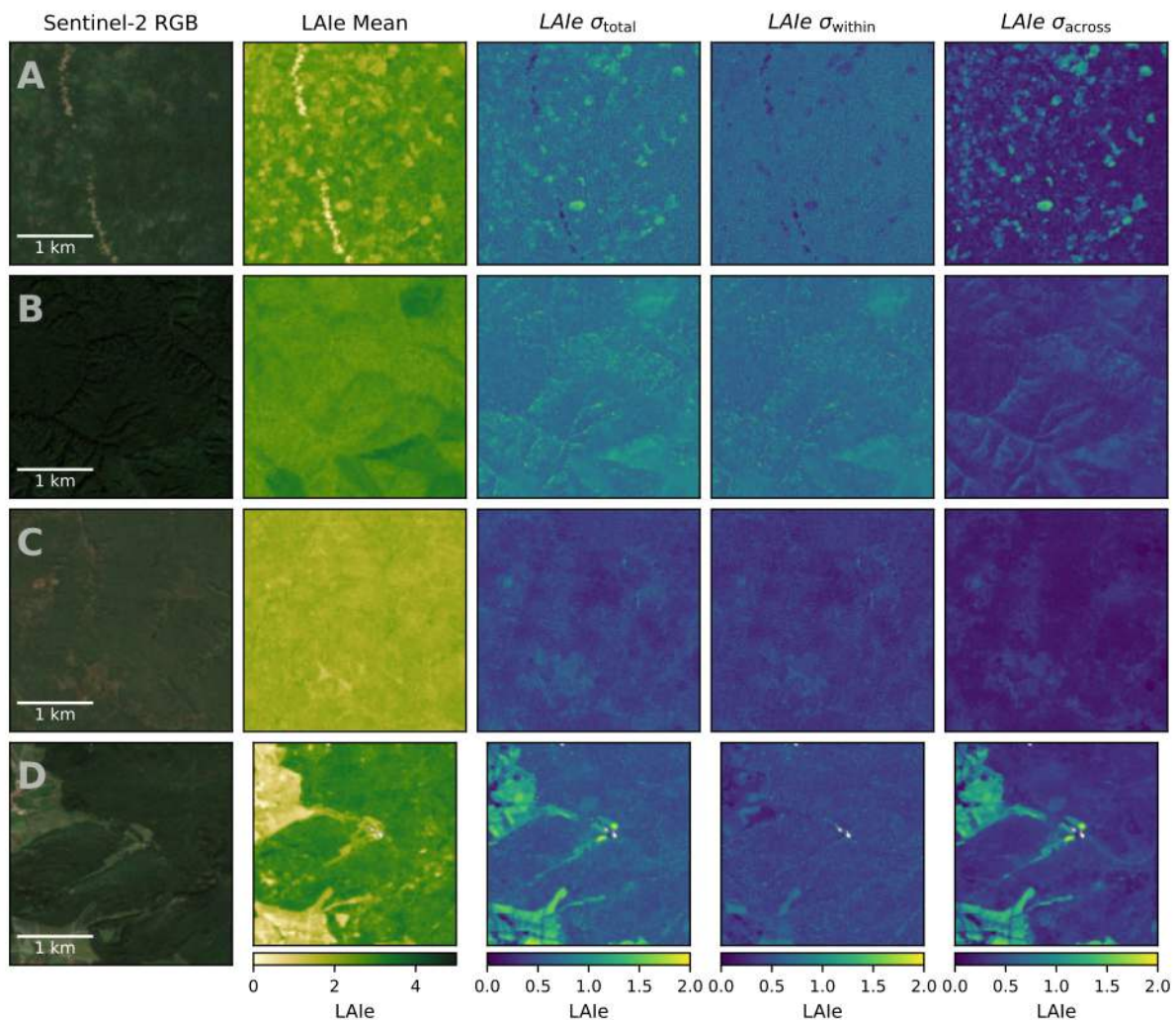


Figure S5 – Decomposition of retrieval uncertainty. (A) shifting agriculture mosaic in Central Africa, (B) temperate rainforest in the Olympic Peninsula (United States), (C) tropical dry Miombo woodland in southern Africa, and (D) temperate deciduous forest-agriculture mosaic in Central Europe. Columns show S2 RGB imagery, mean LAle retrieval, total retrieval uncertainty σ_{total} , within-ensemble uncertainty σ_{within} , and across-acquisition variability σ_{across} for 2022.

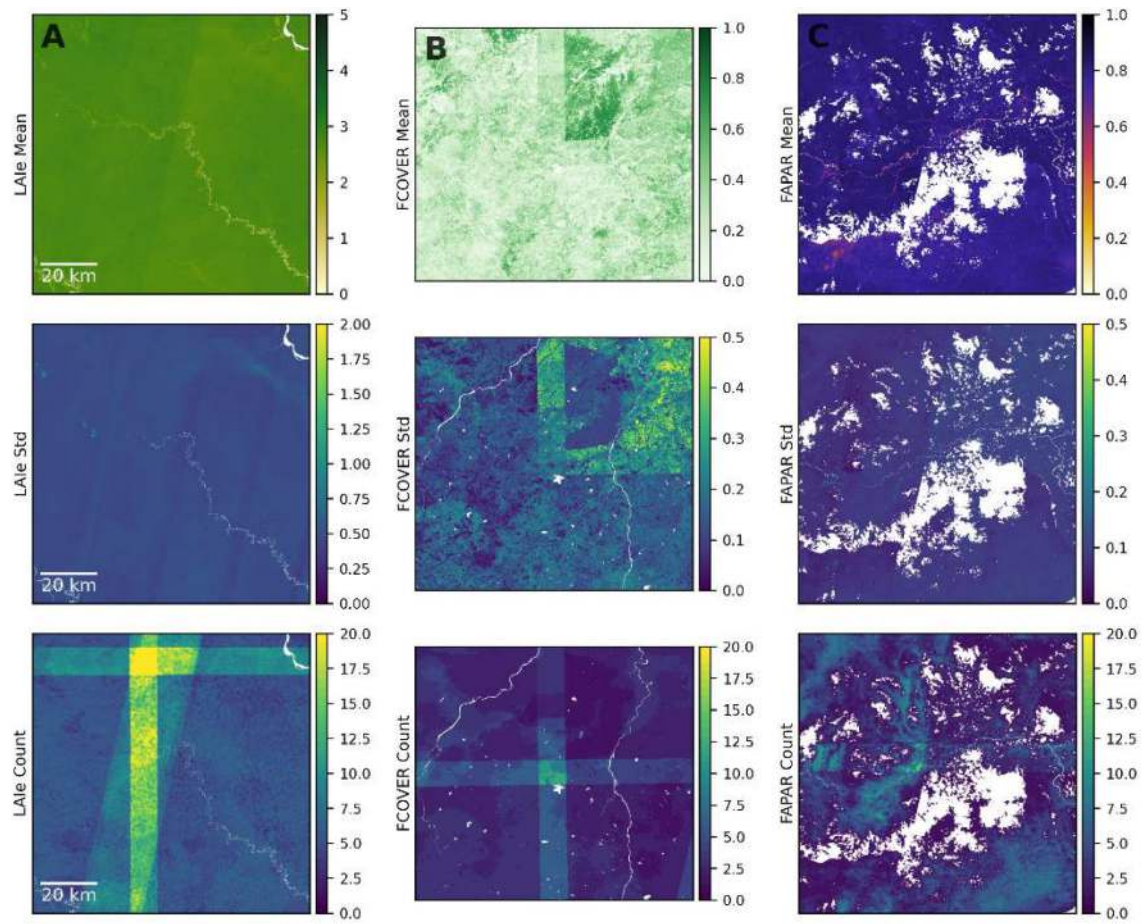


Figure S6 – Examples of artifacts. Mean, uncertainty (Std) and number of observations (Count) for three example regions: **(A)** orbital artifacts in LAIe maps over tropical rainforest in Colombia (1.31° S, 72.17 ° W), **(B)** tiling artifacts in FCOVER over agricultural regions in India (25.26° N, 79.07° E), and **(C)** gaps in FAPAR retrievals due to persistent cloud cover in Borneo, Indonesia (0.81° N, 114.62° E).

Latitudinal Zone	(70, 80]	0.000	0.000	0.004	0.019	0.056	0.064	0.430	0.339	0.078	0.011	0.000	0.000
	(60, 70]	0.001	0.001	0.001	0.003	0.011	0.103	0.383	0.401	0.078	0.018	0.000	0.000
	(50, 60]	0.003	0.001	0.005	0.008	0.055	0.202	0.254	0.335	0.110	0.010	0.006	0.010
	(40, 50]	0.031	0.041	0.015	0.049	0.144	0.114	0.220	0.189	0.136	0.018	0.020	0.022
	(30, 40]	0.122	0.069	0.066	0.087	0.137	0.039	0.062	0.203	0.120	0.035	0.016	0.043
	(20, 30]	0.184	0.120	0.130	0.025	0.028	0.013	0.009	0.069	0.095	0.164	0.100	0.063
	(10, 20]	0.086	0.057	0.038	0.011	0.018	0.008	0.018	0.066	0.258	0.183	0.151	0.106
	(0, 10]	0.169	0.043	0.037	0.050	0.115	0.079	0.064	0.082	0.098	0.147	0.060	0.058
	(-10, 0]	0.096	0.072	0.088	0.118	0.153	0.189	0.102	0.080	0.017	0.019	0.015	0.049
	(-20, -10]	0.040	0.077	0.237	0.312	0.183	0.070	0.034	0.018	0.007	0.005	0.002	0.015
	(-30, -20]	0.077	0.141	0.250	0.119	0.112	0.137	0.071	0.029	0.007	0.024	0.003	0.030
	(-40, -30]	0.054	0.080	0.061	0.071	0.106	0.212	0.214	0.077	0.028	0.056	0.028	0.014
	(-50, -40]	0.021	0.091	0.070	0.096	0.096	0.139	0.209	0.107	0.053	0.016	0.064	0.037
	(-60, -50]	0.097	0.177	0.194	0.145	0.016	0.081	0.000	0.097	0.016	0.048	0.113	0.016
		1	2	3	4	5	6	7	8	9	10	11	12
		Month											

Figure S7 – Normalized frequency counts of S2 images per month across latitudinal zones in 2020. We selected acquisitions to maximize image-level NDVI and minimize cloud cover (see methods), resulting in 151,668 S2 images from 18,343 unique MGRS tiles. Frequency counts were normalized within 10-degree latitudinal zones.

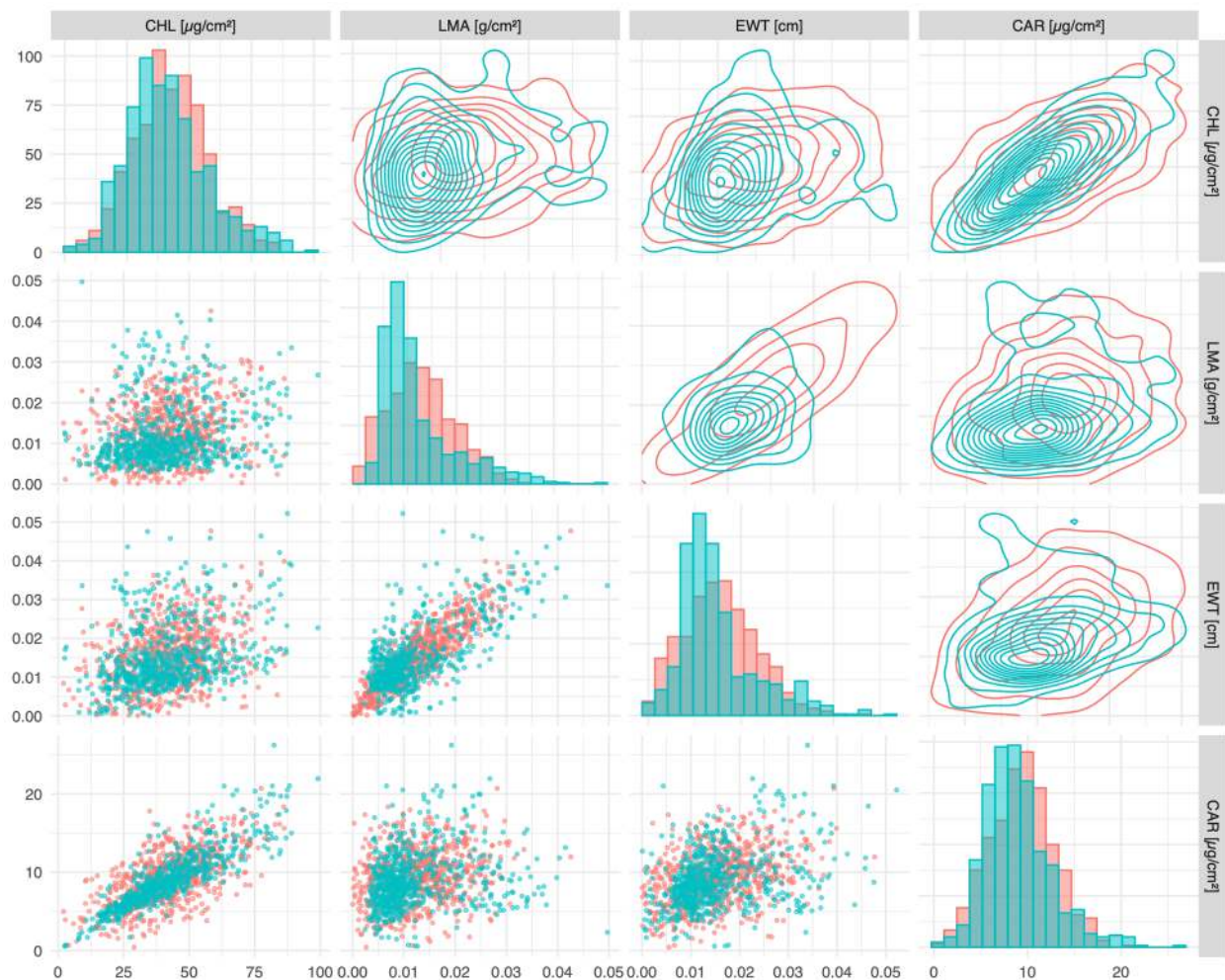


Figure S8 – Pairwise comparison of observed and simulated leaf-trait distributions. NEON in-situ measurements (blue) and samples generated using a Gaussian Copula (red) are compared for leaf chlorophyll content (CHL), leaf carotenoid content (CAR), equivalent water thickness (EWT), and leaf mass per area (LMA). Diagonal panels show histograms, upper panels show bivariate density contours, and lower panels show scatter plots. The simulated samples preserve the covariance structure observed among leaf traits and remain within realistic biological ranges.

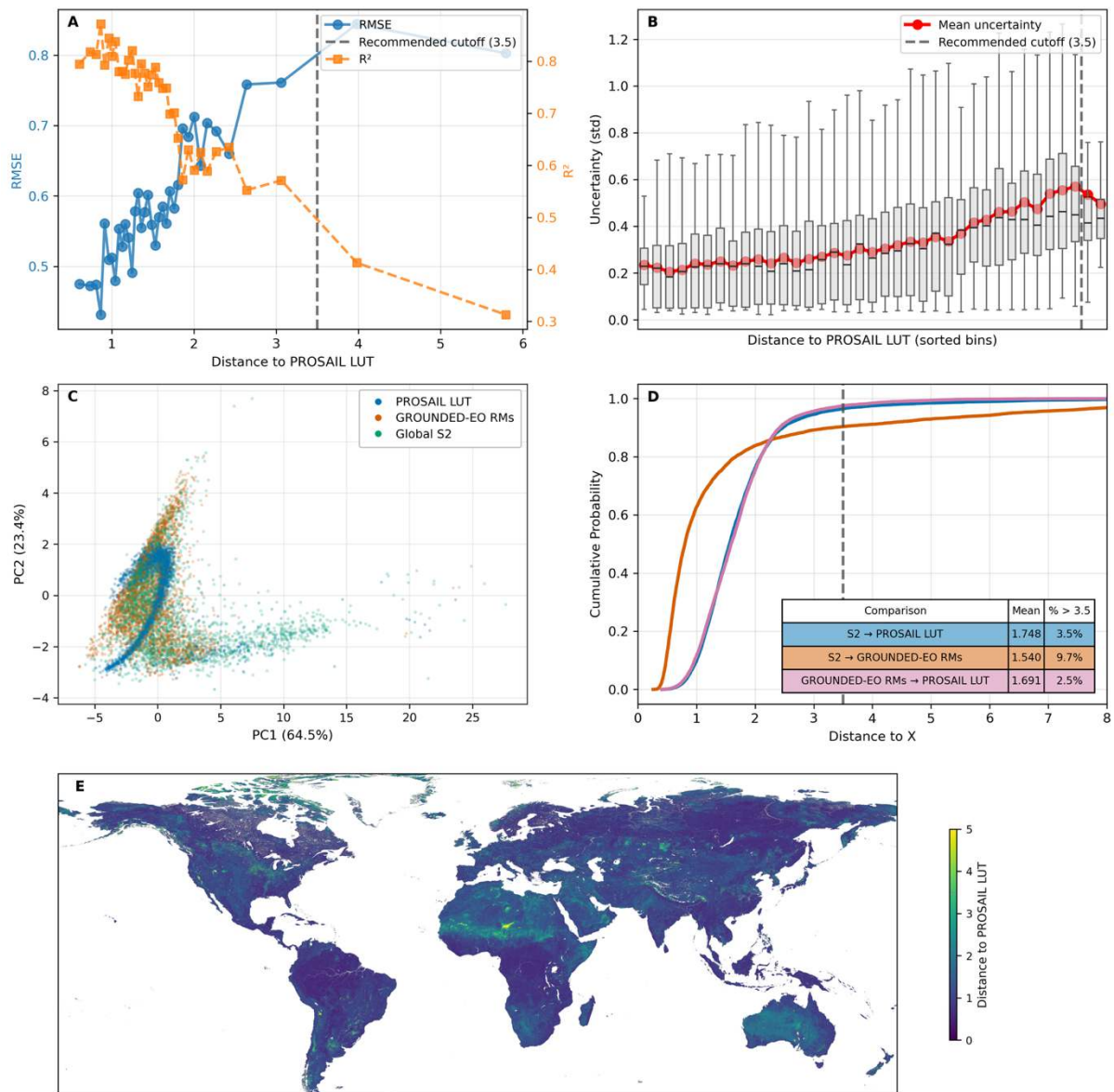


Figure S9 – Spectral domain coverage and extrapolation assessment. **(A)** LAle retrieval performance (RMSE; R^2) evaluated against GROUNDED-EO RMs as a function of mean Manhattan distance to the five nearest pooled LAle LUT spectra in standardized ten-band S2 reflectance space. Metrics are calculated in 300-RM bins. The dashed vertical line indicates the qualitatively determined maximum distance threshold of 3.5. **(B)** Calibrated retrieval uncertainty across the same distance bins, shown as means (red) and boxplots (grey). **(C)** First two principal components showing pooled LAle spectra, GROUNDED-EO RMs, and global S2 reflectance samples. **(D)** Cumulative distributions of average spectral distances for three relationships: S2 → PROSAIL LUT, S2 → GROUNDED-EO RMs, and GROUNDED-EO RMs → PROSAIL LUT, using equal-sized groups. The dashed line indicates the recommended distance threshold of 3.5. The inset table shows mean distances and the percentages of samples exceeding the 3.5 threshold for each comparison. **(E)** Global map of mean

spectral distance to LUT, computed at 100 m resolution with LUTs pooled across LAI_e, FAPAR and FCOVER. See Note S4 for further details.

Supplementary Notes

Note S1 – Landscape heterogeneity differences of LAI_e in forests and tree plantations (see Figure 4A)

Methods. We assessed spatial heterogeneity in annual 20 m LAI_e retrievals within geographically matched plantation and forest windows across Tasmania, Australia. We defined forest type using the JRC Global Map of Forest Types 2020¹¹: plantations were class 20, while the comparison group pooled naturally regenerating forest (class 1) and primary forest (class 10). Candidate windows were restricted to locations where every pixel in a 500 × 500 m window (25 × 25 pixels) belonged to the same JRC class and was classified as tree cover in ESA WorldCover 2021 v2¹². We excluded windows that intersected roads or buildings mapped by OpenStreetMap¹³. We aligned all layers to the 20 m LAI_e grid in EPSG:32755. We matched plantation window centers to forest window centers within 5 km. In addition, no center from a different selected pair could occur within 1 km of either center in an accepted pair. This produced 464 matched pairs, comprising 459 naturally regenerating and five primary-forest controls.

For each window, we calculated mean LAI_e, the sample standard deviation of LAI_e, and adjacent-pixel contrast, defined as the mean absolute difference between horizontally and vertically adjacent valid pixels. We calculated pairwise effects as plantation minus forest. We estimated uncertainty in mean effects using 20,000 percentile-bootstrap resamples of complete matched pairs. Figure 4A shows a representative pair, its LAI_e distributions, and the effects across all matched pairs.

Results. Mean LAI_e was higher in plantations than in matched forests (paired difference: 0.479; 95% bootstrap interval: 0.447–0.512). In contrast, plantations had less variation in LAI_e within windows, in both standard deviation (paired difference: –0.046; –0.058 to –0.033) and differences between adjacent pixels (–0.043; –0.048 to –0.039). Plantations therefore had higher overall LAI_e but less spatial variation than nearby non-plantation forests.

Note S2 – Sensitivity of harvest-related FCOVER change to spatial resolution (see Figure 4C)

Methods. We assessed how spatial resolution affects the FCOVER decline detected within forest-harvest polygons in Sweden. We obtained harvest boundaries and dates from the Swedish Forest Agency distributed through the ArcGIS REST API¹⁴ and selected harvests dated 2020–2024. We used annual precomputed S2BIOPHYS FCOVER products for 2019–2025 at 20 m resolution. We aggregated these maps to 100, 300, and 500 m using average resampling, excluding invalid pixels. For a harvest in year y , we compared FCOVER in years $y - 1$ (pre) and $y + 1$ (post). Polygon means were calculated using *exactextract*¹⁵, weighting valid pixels by their fractional overlap with each polygon. We defined harvest-related change as post-harvest minus pre-harvest FCOVER, so negative values indicate a decline. We divided the retained 231,997 polygons into size bins: smaller than 1 ha, between 1–<2 ha, between 2–5 ha, and larger than 5 ha. For each size bin and resolution, we summarized FCOVER change using the median and interquartile range.

Results. Median FCOVER post-pre changes decreased with coarser spatial resolution (Figure 4C), with the strongest attenuation in smaller harvest polygons. This pattern is consistent with increasing spatial mixing between harvested areas and surrounding land cover: small polygons occupy a smaller fraction of coarse pixels, reducing the change measured within their boundaries. The comparison therefore demonstrates the value of 20 m observations in retaining localized harvest-related changes that spatial aggregation attenuates.

Note S3 – Biome- and vegetation-type-specific model evaluation

Methods. To assess whether vegetation-specific calibration could improve retrieval performance relative to a single global model (henceforth *generic* model), we repeated model calibration for groups defined by combinations of biome and vegetation type. Rather than defining groups by biome or land cover alone, we grouped biome × land-cover combinations representing similar vegetation structure and environmental conditions, while ensuring sufficient data for model calibration and evaluation. Biome assignments followed the RESOLVE ecoregion biome labels. Land-cover classes were derived from the NLCD labels associated with each GROUNDED-EO reference measurement (RM). We simplified the original NLCD classification into six vegetation classes: cropland, deciduous/mixed forest, evergreen forest, grassland, shrubland, and wetland. We then cross-tabulated these vegetation classes with biome assignments and combined reference measurements into eight groups (Table S5). We excluded combinations that could not be assigned to a sufficiently represented group, removing 342 RMs. The resulting grouped dataset contained 10,696 RMs across 44 ecoregions.

We repeated the model calibration described in the main Methods independently for each biome–vegetation group and for a generic model using all RMs assigned to the retained groups. We used the same procedure and parameter search space for all models and traits. All models used the S2BIOPHYS hybrid PROSAIL inversion with a multilayer perceptron. Predictors comprised ten Sentinel-2 reflectance bands and three observation-geometry angles. To reduce computational costs, we used a reduced optimization budget of 1,500 trials per model, comprising 600 random startup trials followed by 900 trials using the tree-structured Parzen estimator (TPE). We calibrated the generic model separately for this experiment using the same reduced optimization budget.

Instead of the five-fold ecoregion-blocked cross-validation used for the main S2BIOPHYS model, we used three folds to accommodate the smaller number of ecoregions within individual biome–vegetation groups. We retained all RMs from the same ecoregion within the same fold.

For the primary comparison, we used stacked out-of-sample (OOS) retrievals, with each RM predicted only when its ecoregion was held out. We evaluated vegetation-specific and generic models on identical RMs within each group using RMSE and R^2 . The group-specific R^2 used the mean of each validation group. The global reference R_g^2 used the mean across all groups. We defined $\Delta\text{RMSE} = \text{RMSE}_{\text{specific}} - \text{RMSE}_{\text{generic}}$, with negative values indicating better vegetation-specific performance.

Results. Overall, the generic model performed better when pooling RMs across all validation groups, with lower pooled RMSE for all three traits (Table S6). Vegetation-specific calibration provided improvements primarily in well-represented groups. For LAI, specific models reduced RMSE in the three largest groups, whereas performance was generally less consistent in groups with fewer RMs (Table S6). This indicates that vegetation-specific calibration can improve retrievals when sufficient reference data are available, but pooling data across vegetation types currently provides more robust overall performance.

Note S4 – Spectral domain coverage and extrapolation assessment (see Figure S9)

Methods. We assessed spectral coverage by comparing modified PROSAIL LUT spectra, S2 reflectances associated with GROUNDED-EO RMs, and global S2 reflectance samples collected from approximately uniformly distributed locations across the terrestrial land surface¹⁶. Each ensemble member has a separate LUT reflecting its optimized simulation settings. We pooled these model-specific LUTs to characterize the ensemble's spectral domain. Distances therefore describe proximity to a pooled reference domain rather than to an individual model's LUT.

For the performance analysis (Figure S9A–B), we standardized the ten reflectance bands to zero mean and unit variance using the combined LUT, RM, and global S2 datasets. We defined distance as the mean Manhattan distance to the five nearest LUT spectra, with each pairwise distance calculated as the sum of absolute band differences. We sorted RMs by distance and calculated ensemble RMSE, R^2 , and mean calibrated uncertainty in 300-RM bins, allowing a smaller final bin. Based on visual inspection, we selected 3.5 as an indicative distance threshold, beyond which retrieval performance declined markedly.

We visualized the domains using the first two principal components (Figure S9C). The cumulative distributions (Figure S9D) used five-neighbor Manhattan distances in the full ten-component PCA space, with LUT and RM reference sets randomly subsampled to equal sizes to account for varying sample sizes.

For the global assessment (Figure S9E), we calculated distances at 100 m resolution using retained S2 observations from 2019–2025. Each of the five realizations pooled a random 25% sample from each ensemble-member LUT across LAI, FAPAR, and FCOVER, then standardized bands to this pooled sample. We randomly ordered all retained growing-season scenes per S2 tile before mosaicking the cloud-masked observations. We then calculated mean Manhattan distances to the five nearest pooled LUT spectra and averaged the five resulting distances.

References

1. Brown, L.A., Fernandes, R., Verrelst, J., Morris, H., Djamai, N., Reyes-Muñoz, P., D.Kovács, D., and Meier, C. (2025). GROUNDED EO: Data-driven Sentinel-2 LAI and FAPAR retrieval using Gaussian processes trained with extensive fiducial reference measurements. *Remote Sensing of Environment* 326, 114797. <https://doi.org/10.1016/j.rse.2025.114797>.
2. Dinerstein, E., Olson, D., Joshi, A., Vynne, C., Burgess, N.D., Wikramanayake, E., Hahn, N., Palminteri, S., Hedao, P., Noss, R., et al. (2017). An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. *BioScience* 67, 534–545. <https://doi.org/10.1093/biosci/bix014>.
3. Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., D'amico, J.A., Itoua, I., Strand, H.E., Morrison, J.C., et al. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience* 51, 933–938. [https://doi.org/10.1641/0006-3568\(2001\)051%5B0933:TEOTWA%5D2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051%5B0933:TEOTWA%5D2.0.CO;2).
4. United States Geological Survey (2026). Annual National Land Cover Database (NLCD) Collection 1 Products. (U.S. Geological Survey). <https://doi.org/10.5066/P94UXNTS>.
5. Weiss, M. (2020). S2ToolBox Level 2 products: LAI, FAPAR, FCOVER (Version 2.1). http://step.esa.int/docs/extra/ATBD_S2ToolBox_V2.1.pdf.
6. Estévez, J., Salinero-Delgado, M., Berger, K., Pipia, L., Rivera-Caicedo, J.P., Woher, M., Reyes-Muñoz, P., Tagliabue, G., Boschetti, M., and Verrelst, J. (2022). Gaussian processes retrieval of crop traits in Google Earth Engine based on Sentinel-2 top-of-atmosphere data. *Remote Sensing of Environment* 273, 112958. <https://doi.org/10.1016/j.rse.2022.112958>.
7. Kovács, D.D., Reyes-Muñoz, P., Salinero-Delgado, M., Mészáros, V.I., Berger, K., and Verrelst, J. (2023). Cloud-Free Global Maps of Essential Vegetation Traits Processed from the TOA Sentinel-3 Catalogue in Google Earth Engine. *Remote Sensing* 15, 3404. <https://doi.org/10.3390/rs15133404>.
8. Wan, L., Ryu, Y., Dechant, B., Hwang, Y., Feng, H., Kang, Y., Jeong, S., Lee, J., Choi, C., and Bae, J. (2024). Correcting confounding canopy structure, biochemistry and soil background effects improves leaf area index estimates across diverse ecosystems from Sentinel-2 imagery. *Remote Sensing of Environment* 309, 114224. <https://doi.org/10.1016/j.rse.2024.114224>.
9. National Ecological Observatory Network (NEON) (2023). Plant foliar traits (DP1.10026.001): RELEASE-2023. 81.4 MB. <https://doi.org/10.48443/tmrs-fb32>.
10. Akiba, T., Sano, S., Yanase, T., Ohta, T., and Masanori Koyama (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining KDD '19*. (Association for Computing Machinery), pp. 2623–2631. <https://doi.org/10.1145/3292500.3330701>.
11. Vancutsem, C., Achard, F., Pekel, J.-F., Vieilledent, G., Carboni, S., Simonetti, D., Gallego, J., Aragão, L.E.O.C., and Nasi, R. (2021). Long-term (1990–2019) monitoring of forest cover changes in the humid tropics. *Science Advances* 7, eabe1603. <https://doi.org/10.1126/sciadv.abe1603>.
12. Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., et al. (2022). ESA WorldCover 10 m 2021 v200. Version v200 (Zenodo). <https://doi.org/10.5281/zenodo.7254221>.
13. OpenStreetMap contributors (2026). OpenStreetMap data for Tasmania, accessed 2026-08-26. <https://download.geofabrik.de/australia-oceania/australia/tasmania-latest-free.gpkg.zip>.
14. Skogsstyrelsen (Swedish Forest Agency) Utförda avverknningar [Completed fellings], accessed 2026-08-26.

<https://geodpags.skogsstyrelsen.se/arcgis/rest/services/Geodataportal/GeodataportalVisaSkogsbruk/MapServer/6>.

15. Baston, D. (2023). exactextract. Version 0.3.0 (Zenodo).
<https://doi.org/10.5281/zenodo.11396420>.
16. González, Á. (2010). Measurement of Areas on a Sphere Using Fibonacci and Latitude–Longitude Lattices. *Math Geosci* 42, 49–64. <https://doi.org/10.1007/s11004-009-9257-x>.