Supplementary

A machine learning approach targeting parameter estimation for plant functional type coexistence modeling using ELM-FATES (v2.0)

Lingcheng Li[1], Yilin Fang[2], Zhonghua Zheng[3], Mingjie Shi[1], Marcos Longo[4], Charles D. Koven[4], Jennifer A. Holm[4], Rosie A. Fisher[5], Nate G. McDowell[1,6], Jeffrey Chambers[4], L. Ruby Leung[1]

1. Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, WA, USA
2. Earth System and Science Division, Pacific Northwest National Laboratory, Richland, WA, USA
3. Department of Earth and Environmental Sciences, The University of Manchester, Manchester, UK
4. Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
5. CICERO Center for International Climate and Environmental Research, Oslo, Norway
6. School of Biological Sciences, Washington State University, PO Box 644236, Pullman, WA, USA

Correspondence to: Lingcheng Li (lingcheng.li@pnnl.gov)

Table S1 Monthly and annual climatology of observations (SD: standard deviation)

Month	ET (mm)	ET's SD	GPP (umol/m2/s)	GPP's SD	SH (W/m2)	SH's SD	BW	BW's SD
1	88.28	6.42	10.96	0.58	25.73	9.15	0.31	0.08
2	69.94	14.56	9.70	1.07	23.28	3.04	0.27	0.10
3	81.87	16.08	10.03	0.33	20.55	3.17	0.27	0.05
4	69.15	19.03	8.96	0.44	25.42	5.56	0.34	0.09
5	79.61	20.47	9.98	1.91	25.29	5.89	0.35	0.14
6	94.73	5.73	8.15	0.64	26.47	4.72	0.28	0.06
7	91.43	12.18	8.21	1.08	27.79	5.12	0.32	0.07
8	107.87	23.51	7.89	0.85	34.04	10.24	0.35	0.11
9	104.79	23.25	8.88	0.92	35.77	3.92	0.44	0.09
10	101.14	8.72	10.44	0.88	33.81	4.69	0.41	0.15
11	86.83	14.55	11.02	0.98	30.83	6.40	0.39	0.06
12	81.27	15.70	10.97	0.84	24.03	2.39	0.34	0.04
Annual mean	88.08	15.02	9.60	0.88	27.75	5.36	0.34	0.09

Table S2 comparisons between Exp-1 and Exp-2

Category	BR_{e2t} ∈[0.1, 0.9]	|AGB_bias| < 15%	|GPP_bias| < 15%	|ET_bias| < 15%	|SH_bias| < 15%	|BW_bias| < 15%	Exp-1		Exp-2	
							count	percent	count	percent
Late							130	8.7%	22	1.5%
Coexistence							309	20.6%	61	4.1%
Early							1059	70.6%	1417	94.5%
All dead							2	0.1%	0	0.0%
Total							1500		1500	
Add observation constraints	+						309	20.6%	61	4.1%
	+	+					98	6.5%	44	2.9%
	+	+	+				85	5.7%	52	3.5%
	+	+	+	+			23	1.5%	11	0.7%
	+	+	+	+	+		23	1.5%	7	0.5%
	+	+	+	+	+	+	21	1.4%	6	0.4%

"+" means adding specific constraint to filter the experiments

Table S3 comparisons between Exp-1 and Exp-3

Category	BR_{e2t} ∈[0.1, 0.9]	|AGB_bias| < 15%	|GPP_bias| < 15%	|ET_bias| < 15%	|SH_bias| < 15%	|BW_bias| < 15%	Exp-1		Exp-3		Ratio
							count	percent	count	percent	
Late							130	8.7%	174	11.6%	1.3
Coexistence							309	20.6%	1097	73.1%	3.6
Early							1059	70.6%	229	15.3%	0.2
All dead							2	0.1%	0	0.0%	
Total							1500		1500		
Add observation constraints	+						309	20.6%	1097	73.1%	3.6
	+	+					98	6.5%	620	41.3%	6.3
	+	+	+				85	5.7%	618	41.2%	7.3
	+	+	+	+			23	1.5%	572	38.1%	24.9
	+	+	+	+	+		23	1.5%	502	33.5%	21.8
	+	+	+	+	+	+	21	1.4%	495	33.0%	23.6

"+" means adding specific constraint to filter the experiments

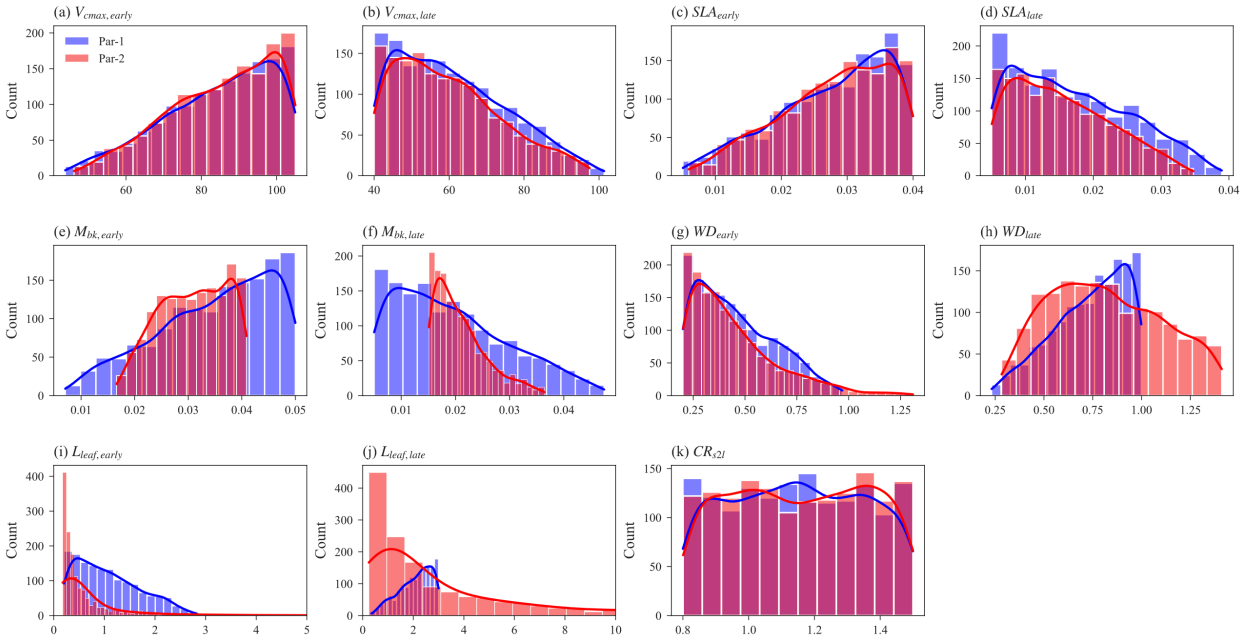

Figure S1. Distribution of ELM-FATES parameters for Exp-1 (Par-1) and Exp-2 (Par-2).

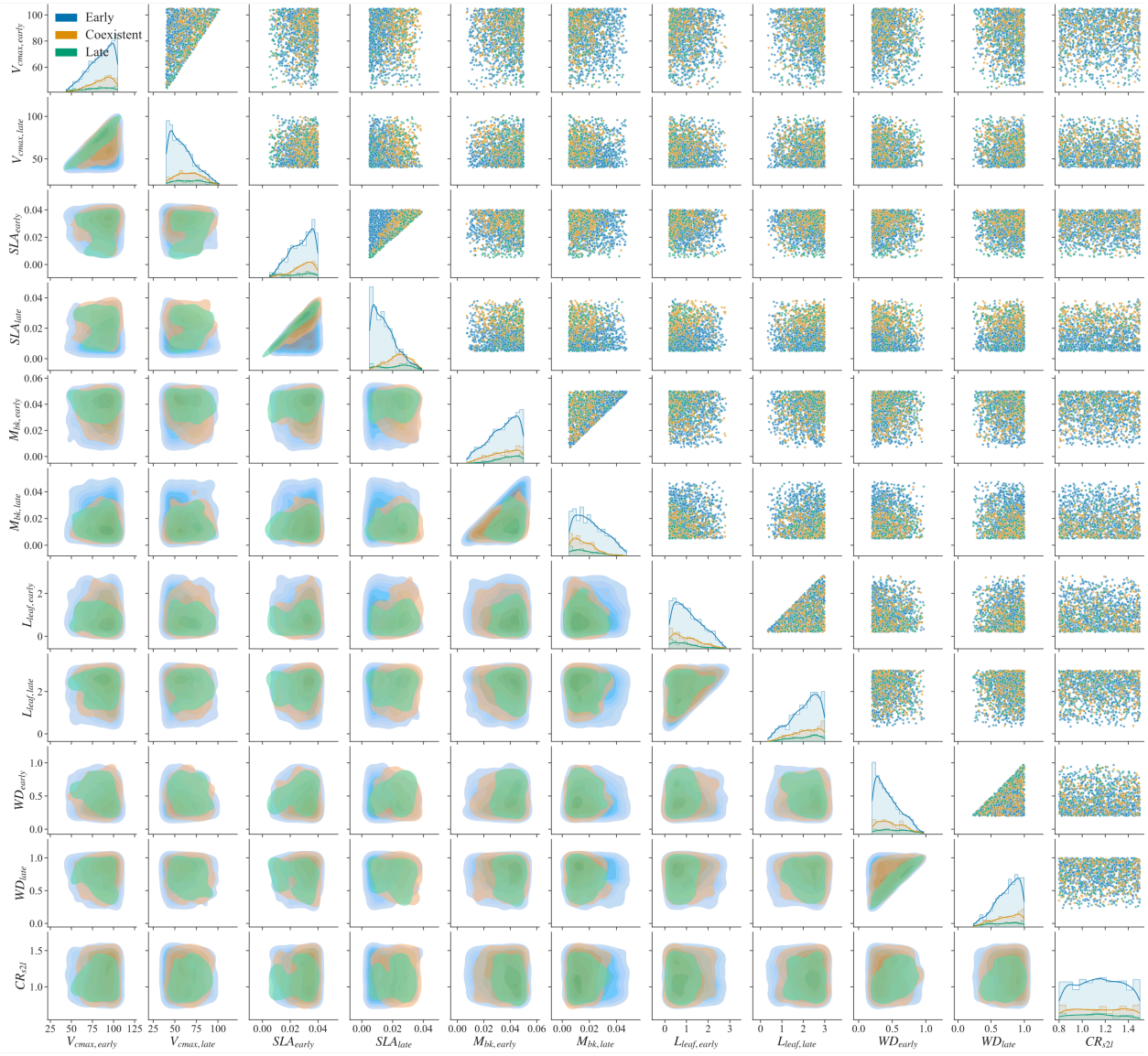

Figure S2. Pairgrid plots of parameters in Par-1 grouped by BR_{e2t}. The upper subplots are scatter plots, the bottom subplots are corresponding kernel density estimate plots, and the diagonal plots are the distribution for each parameter. These parameters are presented in three groups, i.e., blue for the early cases, orange for the coexistent cases, and green for the late cases.

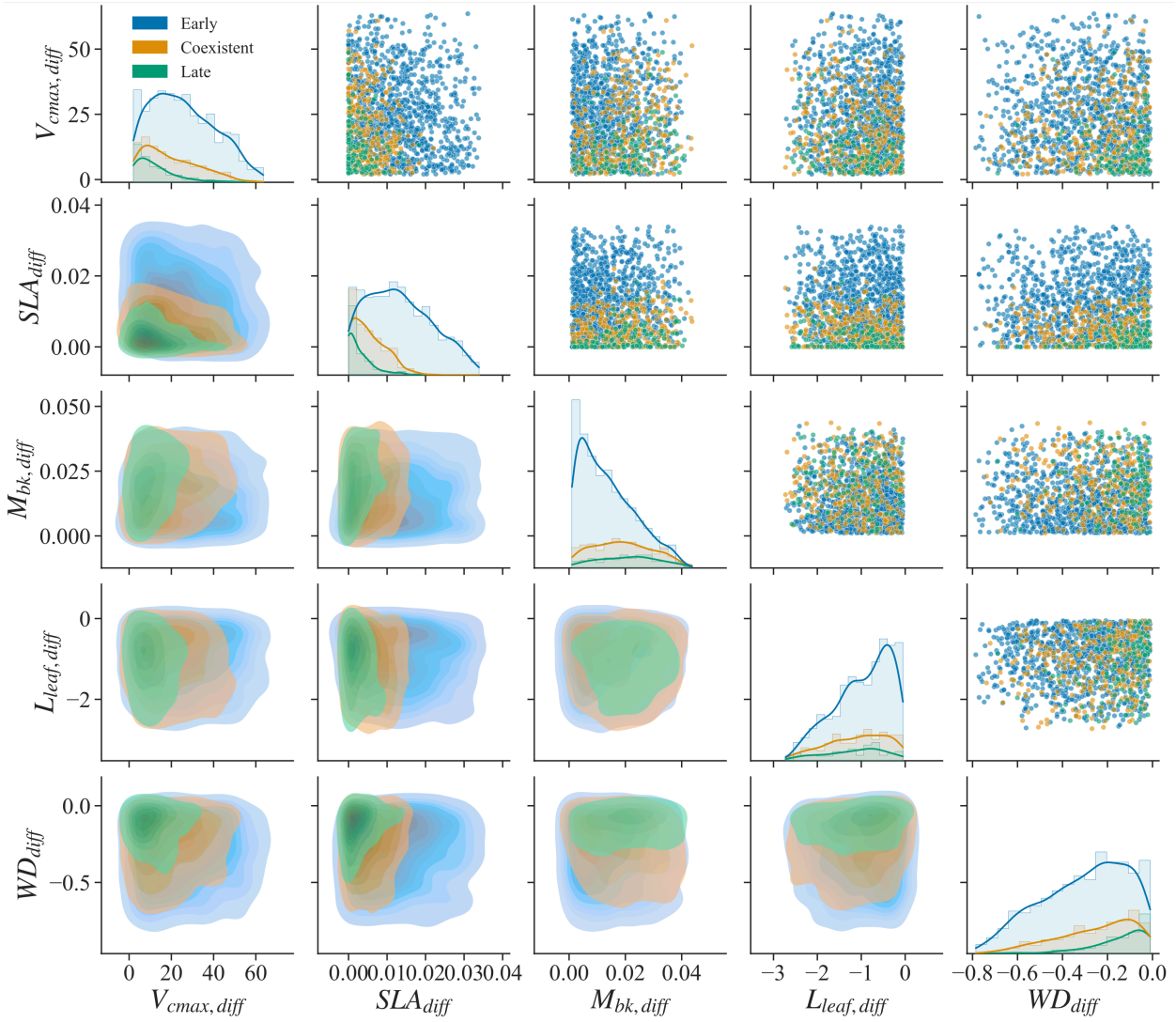

Figure S3. Same as Figure S2, but for the parameter difference between early and late PFT (i.e., diff = early – late).

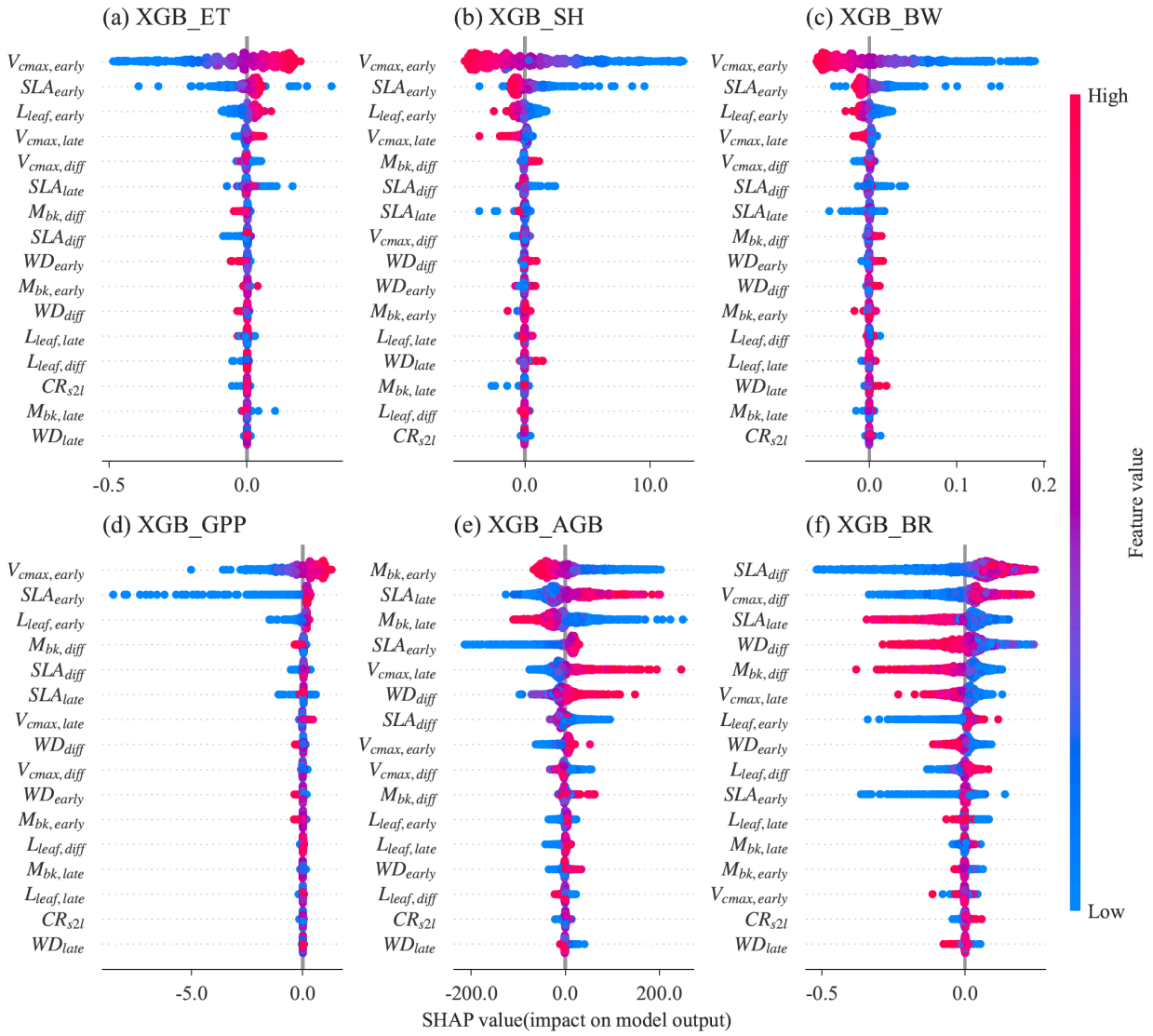

Figure S4. SHAP bee swarm plots for different XGBoost models to show the distribution of the impacts of each feature on the model output. Each dot corresponds to a FATES simulation sample. The x-axis is the SHAP value, which represents the impact of a specific feature (y-axis) on XGBoost model's prediction. The dot color represents the parameter's value from low to high. A positive (negative) SHAP value means that the specific parameter value pushes the model output above (below) the base value (the average model output over the training dataset). For each parameter (y-axis) in each XGBoost model, the Pearson correlation coefficient is calculated between parameter values and corresponding SHAP values, which is displayed in Figure 5 of the main text.

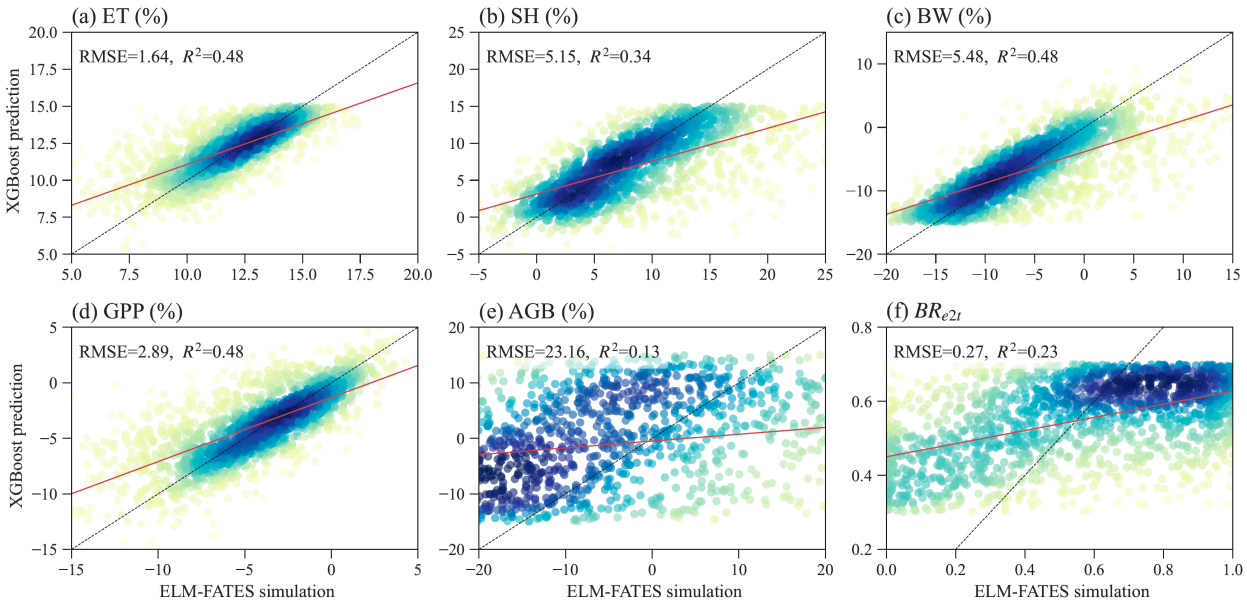

Figure S5. Comparison between XGBoost surrogate models' predictions and ELM-FATES simulations in Exp-3, (a) ET, (b) SH, (c) BW, (d) GPP, (e) AGB, and (f) BR_{e2t}.

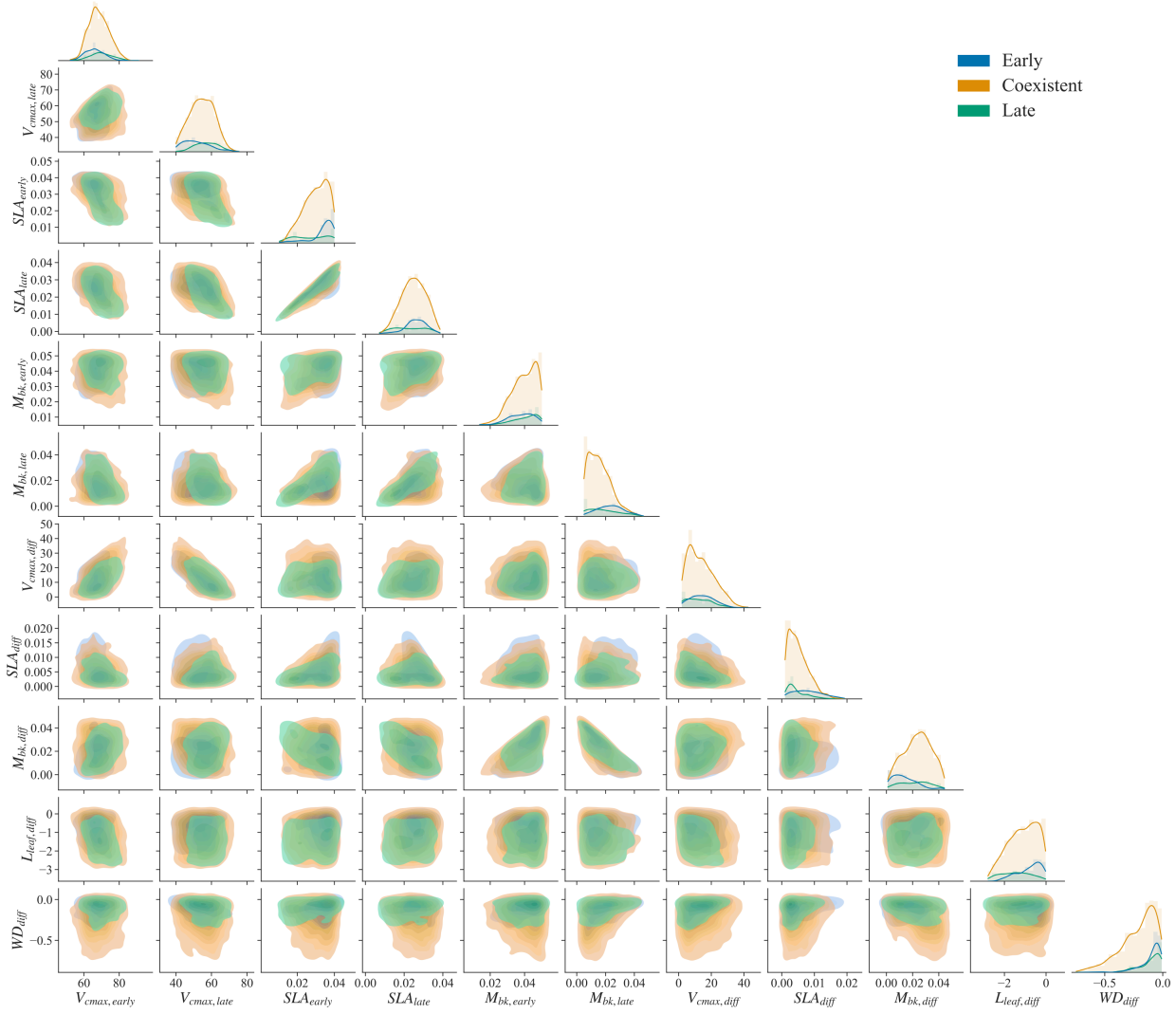

Figure S6. Comparison of parameter or parameter difference in Par-3 (Only eleven features are shown here). The bottom subplots are kernel density estimate plots, and the diagonal plots are the distribution for each parameter. There are three groups, i.e., blue, orange and green for the early, coexistent, and late cases.

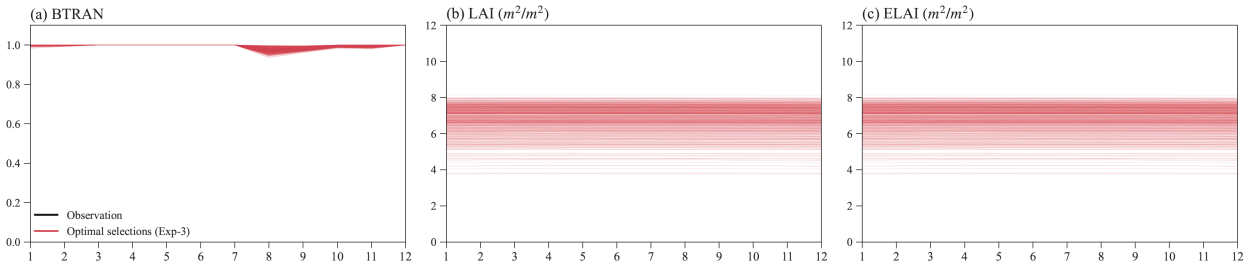

Figure S7. Mean monthly optimal 495 ELM-FATES simulations. (a) BTRAN, plant water stress factor with a valid range between 0 (full water stress) and 1 (no water stress). (b) LAI. Each red curve represents one ELM-FATES simulation.