Table S1 Site information

No.	Site	Latitude	Longitude	PFT	Year_Start	Year_End	Reference
1	US-Bar[A]	44.06	-71.29	DBF	2004	2005	Ouimette et al. (2018)
2	US-Cwt[A]	35.06	-83.43	DBF	2011	2015	Oishi et al. (2020)
3	US-Dk2[A]	35.97	-79.10	DBF	2003	2005	Oishi et al. (2018)
4	US-MOz[A]	38.74	-92.20	DBF	2004	2017	Gu et al. (2016)
5	US-LPH[A]	42.54	-72.19	DBF	2002	2004	Hadley (2016)
6	US-Wi1[A]	46.73	-91.23	DBF	2003		Chen et al. (2020)
7	US-Wi8[A]	46.72	-91.25	DBF	2002		Chen et al. (2020)
8	US-Ha1	42.54	-72.17	DBF	2000	2016	Urbanski et al. (2007)
9	US-MMS	39.32	-86.41	DBF	2000	2014	Schmid et al. (2000)
10	US-Oho	41.55	-83.84	DBF	2004	2013	Noormets et al. (2008)
11	US-UMB	45.56	-84.71	DBF	2000	2014	Gough et al. (2013)
12	US-UMd	45.56	-84.70	DBF	2007	2014	Gough et al. (2013)
13	US-WCr	45.81	-90.08	DBF	2000	2014	Cook et al. (2004)
14	US-Wi3	46.63	-91.10	DBF	2002	2004	Chen et al. (2020)

DBF, deciduous broadleaf forests; The superscript 'A' means sites from AmeriFlux network.

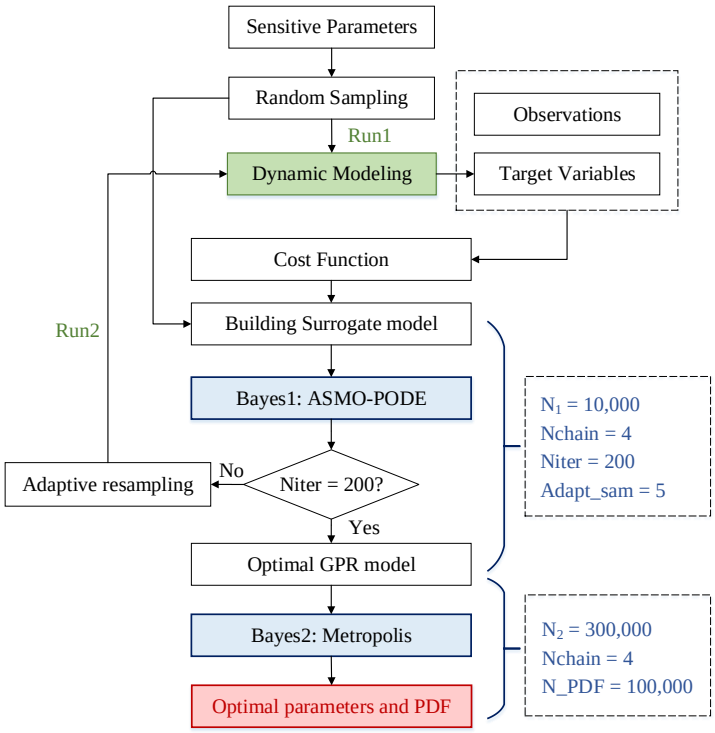

Figure S1. Framework of modified adaptive surrogate modeling (MASM). N_1 and N_2 mean length of a Markov chain in the two steps respectively. Nchain represents the number of Markov chains, Niter is the number of iterations, and Adapt_sam is the number of parameter samples in each adaptive resampling iterations. N_PDF means using the last 100,000 combinations to calculate the posterior parameter distributions of all sensitive parameters.

Table. S2 Environmental variables for building XGBoost model. Variables with 'C' means constant values; 'Y' represnets using annual mean values from all years; 'GS' represents using growing season mean values from all years. 'fraction' refers to the percentage of deciduous forests in each grid based on land cover data from NLCF (30m).

Types	Variable name	Source	C	Y	GS	Types	Variable name	Source	C	Y	GS
Climate	temperature (T)	Daymet		√	√	Soil	sand content	GSDE	√		
	precipitation (prec)			√			clay content		√		
	vapor pressure deficit (vpd)			√	√	Grid	latitude (lat)	-	√		
	daylength (dayl)			√			longitude (lon)	-	√		
	shortwave radiation (srad)			√	√		elevation (dem)		√		
	pressure (press)	NLDAS		√			fraction (frac)	NLCF	√		
	wind speed (vs)	GRIDMET		√		Vegetation variables	GPP	GLASS		√	
	specific humidity (sph)			√			LAI			√	
	cloud cover (cld)	NARR		√			photosynthetically active radiation (PAR)			√	√

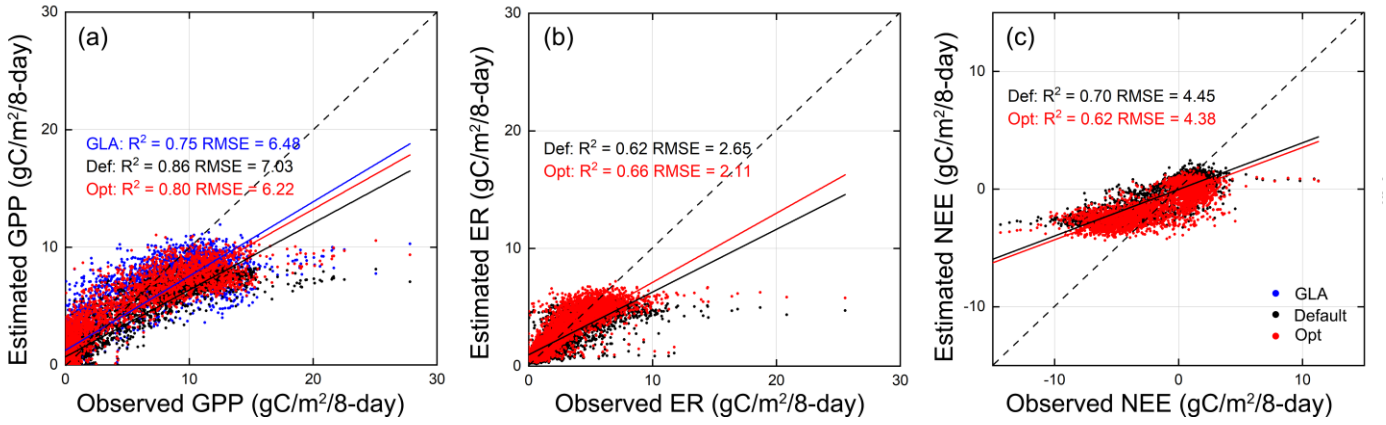

Figure S2. Scatter plots of observed 8-day carbon fluxes versus estimated 8-day carbon fluxes before and after optimization. 'GLA' represents GLASS GPP, 'Default' represents carbon fluxes estimated from previous model with default parameters, 'Opt' represents values from the optimized model with calibrated parameters.

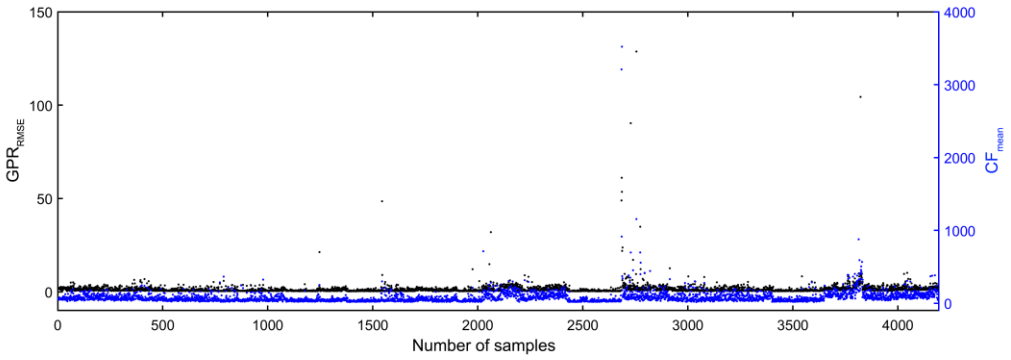

Figure S3. Errors of all the samples when running MASM method.

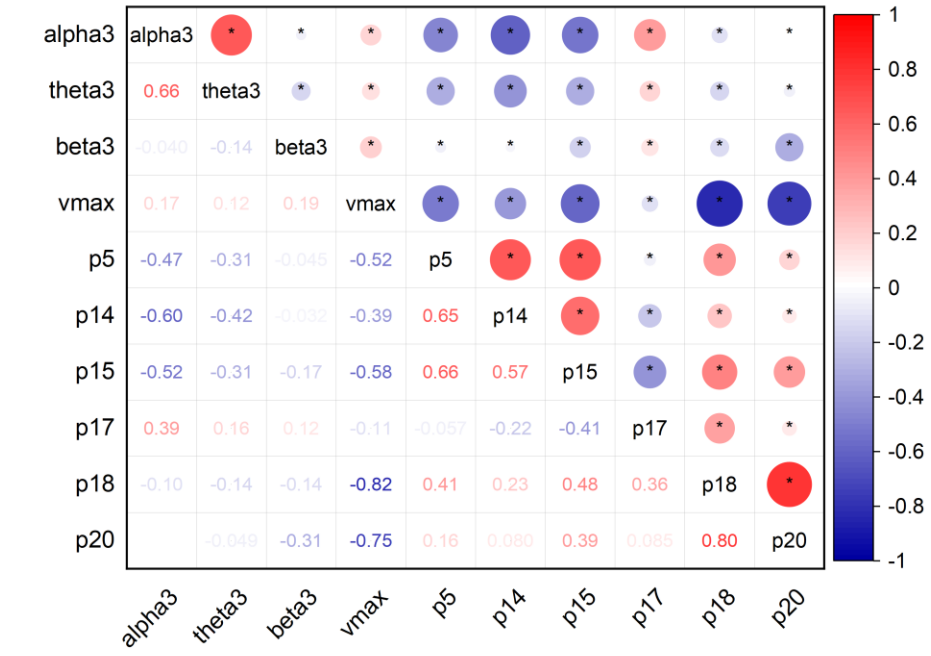

Figure S4. The 2D-correlation map between optimal parameters for the deciduous forests in the EUS. The upper triangle uses the circle size and color to indicate the degree of correlation between two parameters. Higher correlation is expressed as a darker and larger circle, with blue indicating negative correlation and red indicating positive correlation. The lower triangle shows the values of the correlation coefficient, and colors have the same meaning as the upper triangle. Asterrisk indicates a significant correlation (p<=0.05).

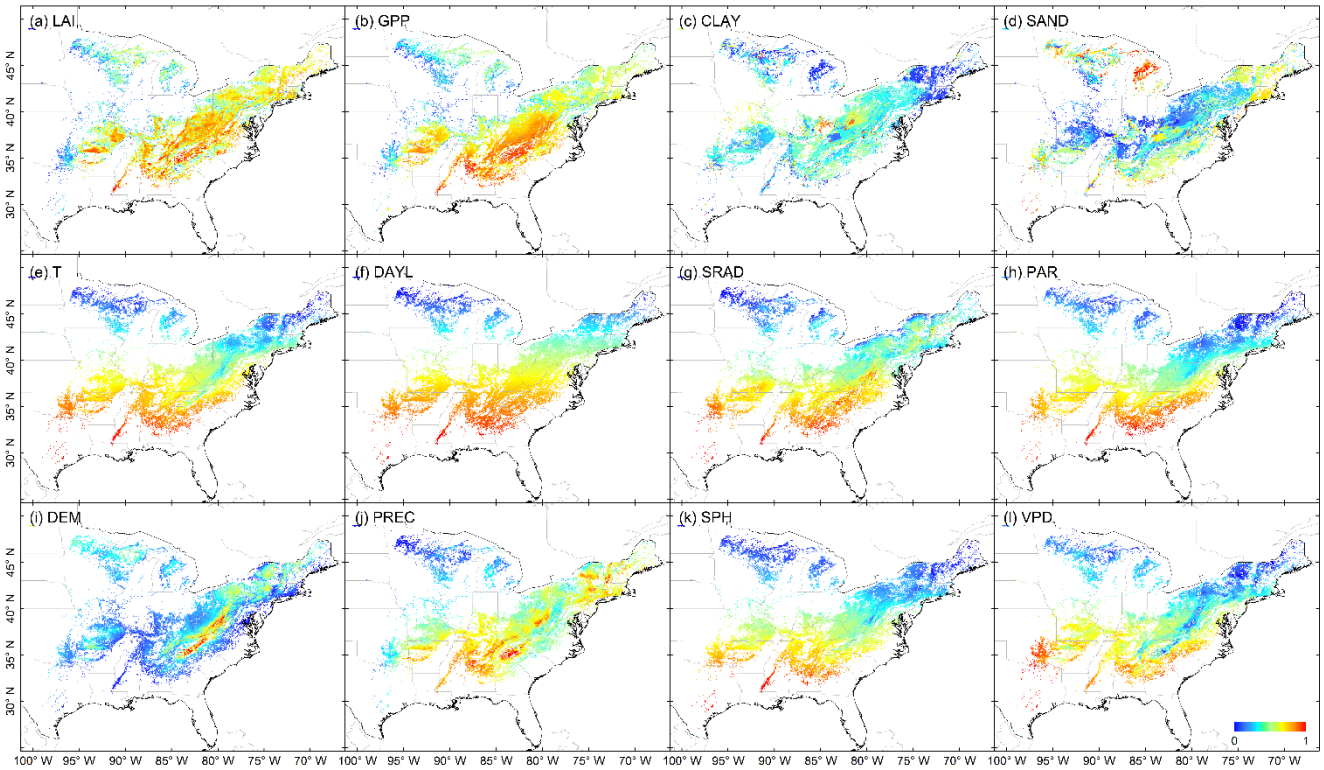

Figure S5. Spatial distribution of several key environmental variables used in the XGBoost approach. All variables have been normalized to 0-1. The explanation of variable abbreviations can be found in Table S2.