Table S1. Site characteristics of the flux tower observations.

Site	Latitude (°N)	Longitude (°E)	IGBP	Period	Type
BE-Bra	51.31	4.52	ENF	2021–2024	I
BE-Vie	50.30	6.00	MF	2021–2024	I
CH-Dav	46.82	9.86	ENF	2020–2024	I
CZ-Lnz	48.68	16.95	DBF	2022–2024	II
DE-Har	47.93	7.60	MF	2020–2024	II
DE-HoH	52.09	11.22	DBF	2019–2024	II
DE-Msr	47.81	11.46	ENF	2021–2024	II
DE-Tha	50.96	13.57	ENF	2021–2024	I
DK-Sor	55.49	11.64	DBF	2022–2024	I
FI-Hyy	61.85	24.29	ENF	2019–2024	I
FI-Ken	67.99	24.24	ENF	2020–2023	II
FI-Var	67.75	29.61	ENF	2017–2024	II
FR-Bil	44.49	-0.96	ENF	2020–2024	II
FR-Fon	48.48	2.78	DBF	2020–2024	I
FR-Hes	48.67	7.06	DBF	2022–2024	II
FR-Pue	43.74	3.60	EBF	2022–2024	I
IT-BFt	45.20	10.74	DBF	2019–2024	II
IT-Ren	46.59	11.43	ENF	2022–2024	II
IT-SR2	43.73	10.29	ENF	2020–2023	I
IT-TrF	45.82	7.56	DNF	2013–2023	II
SE-Htm	56.10	13.42	ENF	2018–2024	II
SE-Nor	60.09	17.48	ENF	2019–2023	II
SE-Svb	64.26	19.77	ENF	2019–2024	II

Notes: Ecosystem types, classified following the International Geosphere–Biosphere Programme (IGBP) scheme, include evergreen needleleaf forest (ENF), evergreen broadleaf forest (EBF), deciduous needleleaf forest (DNF), deciduous broadleaf forest (DBF) and mixed forest (MF). Type I sites provide measurement height and canopy height data for flux tower locations reported in Wang et al. (2020), which were used in the analysis of the Gs–VPD relationship. Type II sites were not included in the Gs–VPD analysis.

Reference: Wang, L., Tian, F., Han, S., & Wei, Z. (2020). Determinants of the asymmetric parameter in the generalized complementary principle of evaporation. Water Resources Research, 56(9), e2019WR026570.

Table S2. Threshold estimates and segment slopes of SHAP-VPD relationships across variables, radiation regimes, and analysis settings.

Analysis group	τ_1	τ_1_CI_low	τ_1_CI_high	τ_2	τ_2_CI_low	τ_2_CI_high	β_L	β_I	β_H
NEE, all SW conditions	3.90	3.87	3.93	20.84	20.59	21.29	-0.41	0.39	0.12
NEE, SW < 200 W m^{-2}	3.09	2.84	3.13	19.75	18.71	20.74	-0.31	0.20	-0.01
NEE, SW 200-400 W m^{-2}	3.98	3.93	4.42	18.67	18.21	19.17	-0.10	0.42	0.09
NEE, SW > 400 W m^{-2}	4.62	4.56	5.24	19.42	19.14	19.82	-0.03	0.47	0.16
Daytime-mean NEE	3.66	2.98	4.01	9.34	7.83	13.40	-0.03	0.25	0.05
NEE, DBF	3.80	3.25	3.88	15.87	15.45	17.43	-0.55	0.51	0.28
NEE, ENF	4.53	4.22	4.57	19.09	18.65	19.64	-0.25	0.28	0.07
GPP, daytime partitioning	6.80	6.24	7.38	20.38	19.43	21.12	-0.07	-0.43	-0.17
GPP, nighttime partitioning	3.96	3.92	3.99	18.76	18.55	19.24	0.26	-0.43	-0.11

Notes: τ_1 and τ_2 denote the first and second thresholds identified from the piecewise linear fit, respectively. τ_1_CI_low and τ_1_CI_high represent the lower and upper bounds of the 95% confidence interval for τ_1, and τ_2_CI_low and τ_2_CI_high represent the corresponding bounds for τ_2. All threshold estimates and their 95% confidence interval bounds are expressed in hPa. β_L, β_I, and β_H denote the slopes of the low-, intermediate-, and high-VPD segments, respectively, expressed as change in SHAP value per hPa. Except for Daytime-mean NEE, all analyses were based on hourly data. “Daytime partitioning” and “nighttime partitioning” refer to the two flux-partitioning methods used to derive GPP. SW denotes shortwave radiation; DBF, deciduous broadleaf forest; ENF, evergreen needleleaf forest; NEE, net ecosystem exchange; and GPP, gross primary productivity.

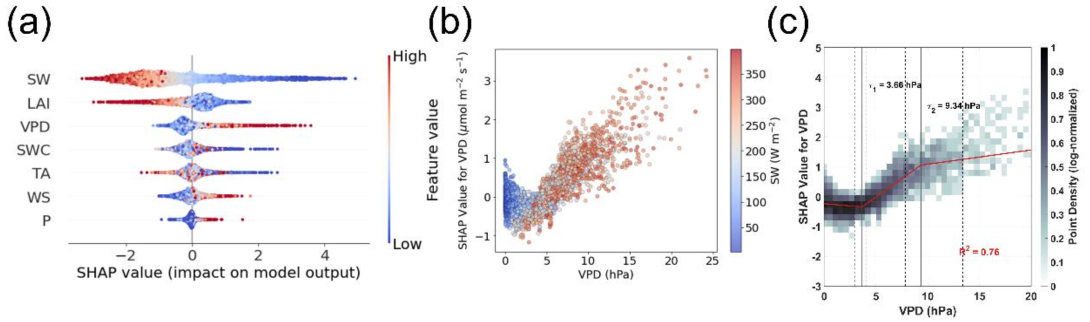

Fig. S1. SHAP values of net ecosystem carbon exchange (NEE) drivers. (a) SHAP summary bee swarm plots, where each dot represents a site-level daily daytime average sample. The x-axis position indicates the contribution of a given variable to the XGBoost model prediction for that sample. When multiple dots overlap at the same x position, they accumulate to reflect density. Dot color indicates the feature value, with blue representing low values and red representing high values. (b) SHAP dependence plot showing the relationship between VPD and its SHAP values, stratified by the gradient of incoming shortwave radiation (SW). The XGBoost models achieved validation R^2 values of 0.62. (c) Heatmap of point density for the relationship between daily daytime VPD and its SHAP values for NEE. In each panel, color indicates the log-normalized point density within each bin, with darker colors representing higher data concentration. The red curve shows a three-segment continuous piecewise linear fit to the binned median SHAP response. Two thresholds, τ_1 and τ_2, are identified from the fitted response and separate the three response regimes. Solid vertical black lines indicate the estimated thresholds, and dashed vertical black lines indicate their 95% bootstrap confidence intervals. Threshold values are labeled in black. The coefficient of determination (R^2), shown in red, quantifies the fit between the modeled response and the original SHAP values.

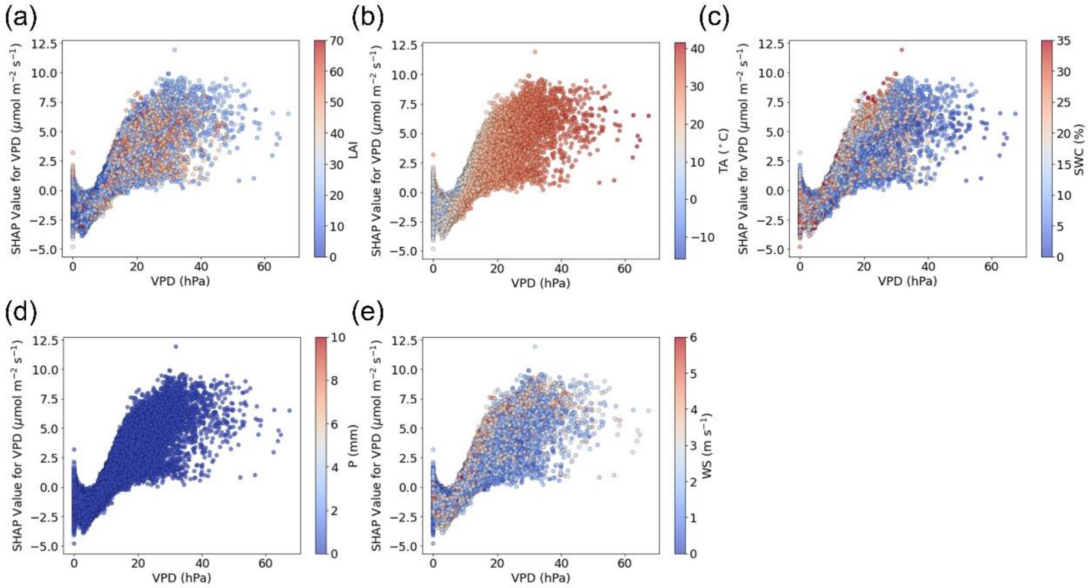

Fig. S2. SHAP dependence plot showing the relationship between VPD and its SHAP values, stratified by the gradient of leaf area index (LAI, a), air temperature (TA, b), soil water content (SWC, c), precipitation (P, d), and wind speed (WS, e). The XGBoost models achieved validation R^2 values of 0.72.

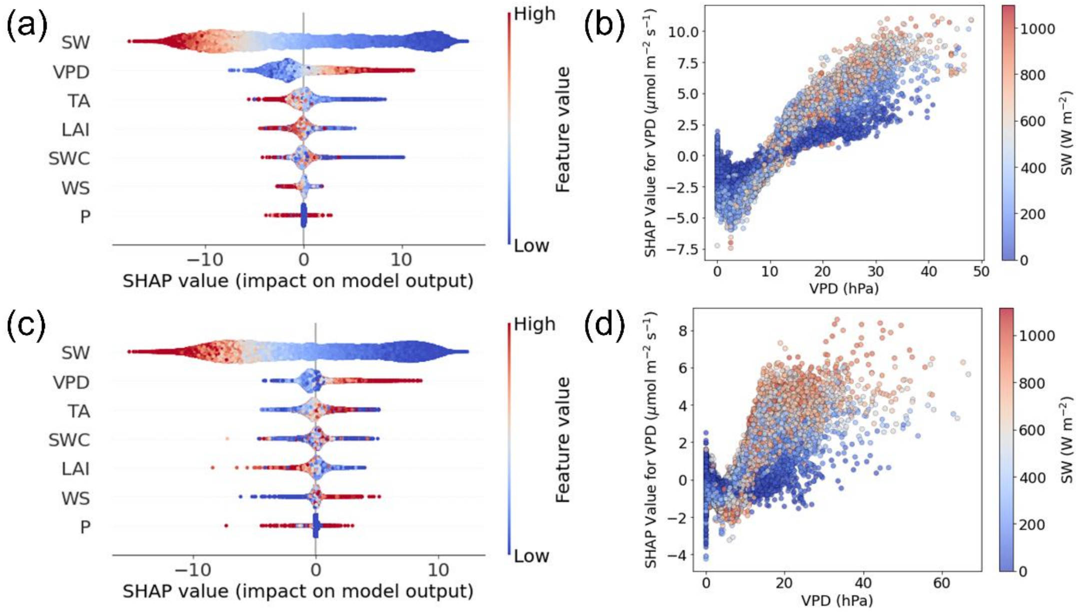

Fig. S3. (a, c) SHAP summary bee swarm plots in DBF (a) and ENF (c) ecosystems, where each dot represents a site-level half-hourly sample. The x-axis position indicates the contribution of a given variable to the XGBoost model prediction for that sample. When multiple dots overlap at the same x position, they accumulate to reflect density. Dot color indicates the feature value, with blue representing low values and red representing high values. (b, d) SHAP dependence plot showing the relationship between VPD and its SHAP in DBF (b) and ENF (d) ecosystems, stratified by the gradient of incoming shortwave radiation (SW).The XGBoost models achieved validation R^2 values of 0.76 for DBF and 0.74 for ENF ecosystems.

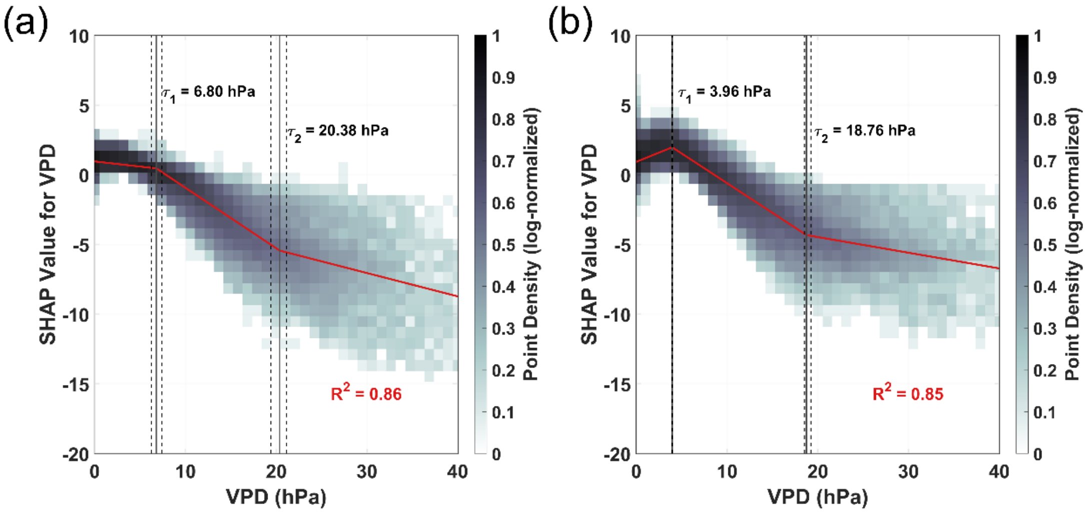

Fig. S4. Heatmap of point density for the relationship between VPD and its SHAP values for GPP derived from the daytime partitioning method (a) and nighttime partitioning method (b). In each panel, color indicates the log-normalized point density within each bin, with darker colors representing higher data concentration. The red curve shows a three-segment continuous piecewise linear fit to the binned median SHAP response. Two thresholds, τ_1 and τ_2, are identified from the fitted response and separate the three response regimes. Solid vertical black lines indicate the estimated thresholds, and dashed vertical black lines indicate their 95% bootstrap confidence intervals. Threshold values are labeled in black. The coefficient of determination (R^2), shown in red, quantifies the fit between the modeled response and the original SHAP values.

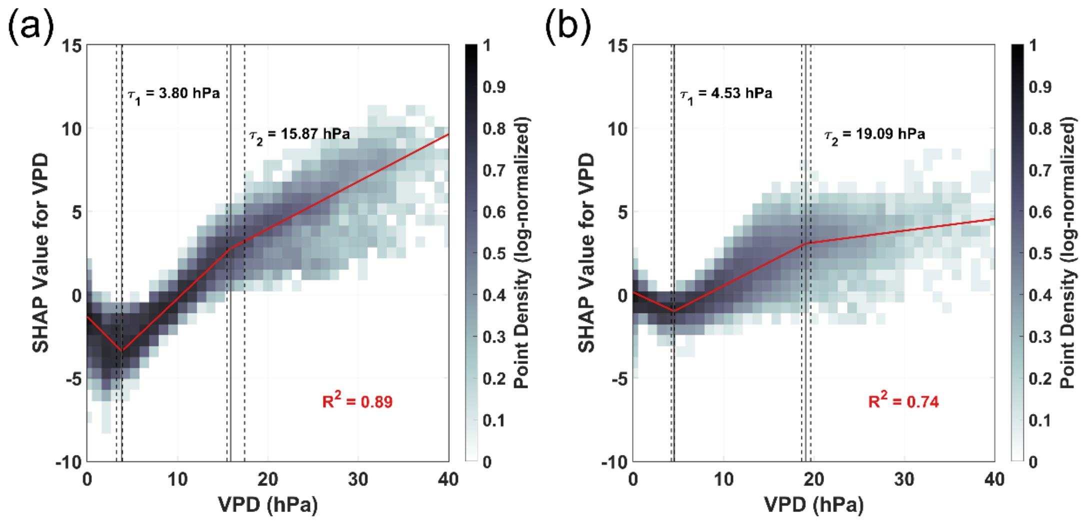

Fig. S5. Heatmap of point density for the relationship between VPD and its SHAP values for NEE in DBF (a) and ENF (c) ecosystems. In each panel, color indicates the log-normalized point density within each bin, with darker colors representing higher data concentration. The red curve shows a three-segment continuous piecewise linear fit to the binned median SHAP response. Two thresholds, τ_1 and τ_2, are identified from the fitted response and separate the three response regimes. Solid vertical black lines indicate the estimated thresholds, and dashed vertical black lines indicate their 95% bootstrap confidence intervals. Threshold values are labeled in black. The coefficient of determination (R^2), shown in red, quantifies the fit between the modeled response and the original SHAP values.

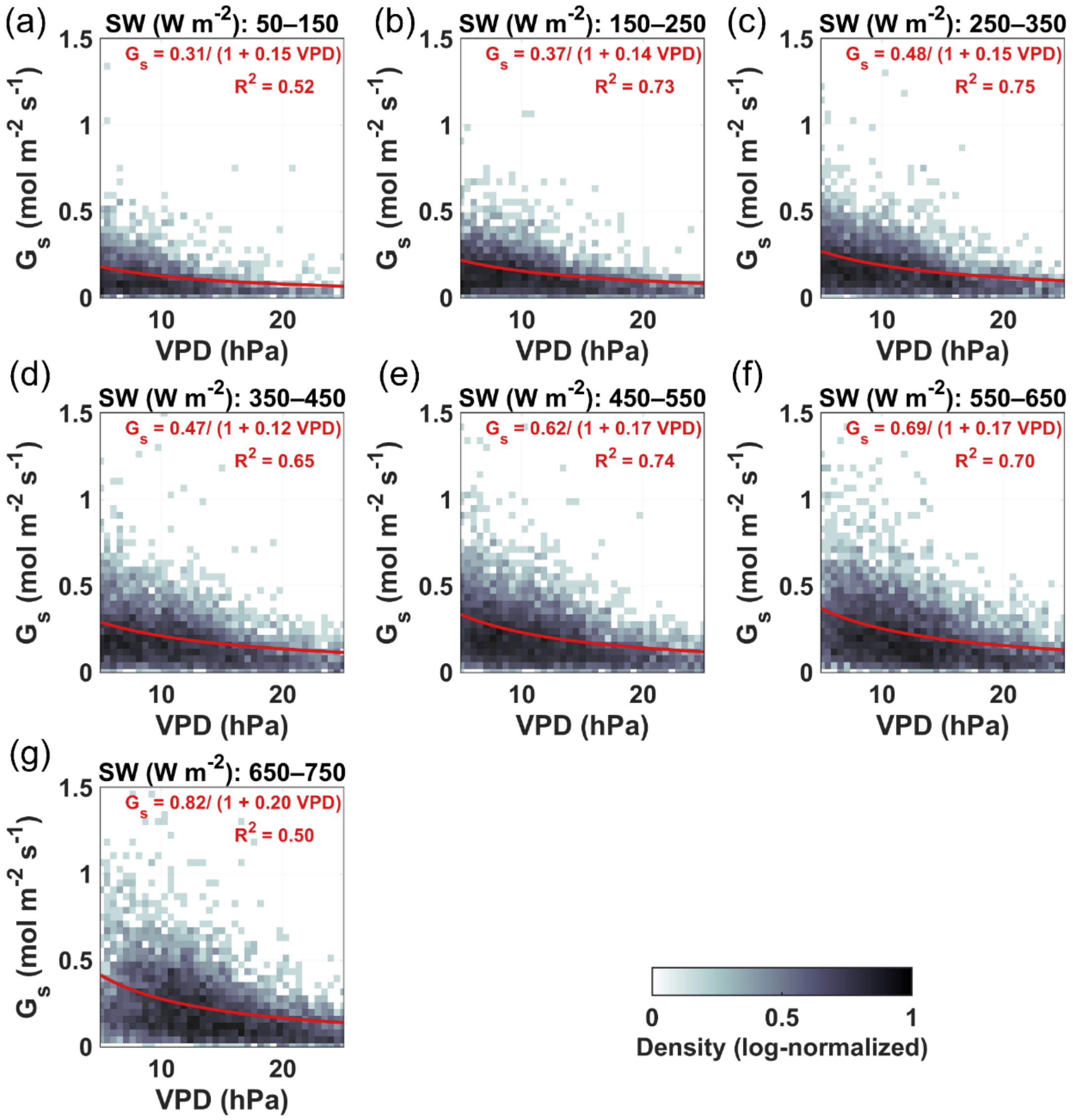

Fig. S6. Relationship between VPD and G_s across different SW bins. Each panel shows the binned distribution of data points, where color indicates the log-normalized density within each bin, with darker colors representing higher concentrations. Red lines show the fitted curves, with the corresponding fitted equations and R^2 values shown in the upper part of each panel. The R^2 values were calculated using binned median responses to reduce the influence of uneven data density.