Supplementary material for “Temperate agroforestry systems take up more CO_2 compared to conventional crop- or grassland, without consuming more water", submitted to *Biogeosciences*

José Ángel Callejas-Rodelas[1,*], Justus G. V. van Ramshorst[1,2,+], Alexander Knohl[1,3], Sarah Choe[4], Maren Langhof[5], and Christian Markwitz[1]

[1]Bioclimatology, Faculty of Forest Sciences and Forest Ecology, University of Göttingen, Büsgenweg 2, 37077 Göttingen, Germany

[2]Quanterra Systems Ltd., Centenary House, Peninsula Park, Exeter, EX2 7XE, UK

[3]Centre for Biodiversity and Land Use, University of Goettingen, Büsgenweg 2, 37077 Göttingen, Germany

[4]Soil Science of Tropical and Subtropical Ecosystems, Faculty of Forest Sciences and Forest Ecology, University of Göttingen, Büsgenweg 2, 37077 Göttingen, Germany

[5]Institute for Crop and Soil Science, Julius Kühn-Institut (JKI), Federal Research Center for Cultivated Plants, Brunswick, Germany

[*]Currently at Department of Applied Physics, University of Granada (UGR), Avda. de la Fuente Nueva, 18002 Granada, Spain; and Inter-University Institute for Earth System Research (IISTA-CEAMA), Avda. del Mediterráneo, 18006 Granada, Spain.

[+]Currently at Sustainable Water Systems Group, Van Hall Larenstein University of Applied Sciences, Agora 1, 8934 CJ Leeuwarden, the Netherlands

Correspondence: José Ángel Callejas-Rodelas (jangel.calrod@ugr.es)

Instrumentation

Table S1. Meteorological and turbulence variables measured at the stations and instrumental information. Variable names follow the FLUXNET standards (https://fluxnet.org/data/aboutdata/data-variables/, last access 23/01/2025). The information in this table corresponds to the latest measuring setups. Not all variables were available during the whole project duration. Specifically, the lower-cost eddy covariance setups to measure CO_2 molar density and *RH*, were installed in 2019. From 2016 to 2017, a different setup was used, as described in more detail in Section 2.2.1 from the paper (originally in Markwitz and Siebicke, 2019).

Measured variable	**Units**	**Instrument**	**Model**	**Company**
Precipitation, *P*	mm	Precipitation transmitter	5.4032.35.007	Adolf Thies GmbH & Co. KG, Göttingen, Germany
Atmospheric pressure, *PA*	kPa	Baro transmitter	3.1157.10.000	Adolf Thies GmbH & Co. KG, Göttingen, Germany
Net radiation, *NETRAD*	W m^{-2}	Net radiometer	NR Lite2	Kipp & Zonen, B.V., Delft, The Netherlands
Shortwave incoming radiation, *SW_IN*	W m^{-2}	Pyranometer	CMP3	Kipp & Zonen, B.V., Delft, The Netherlands
Shortwave outgoing radiation, *SW_OUT*	W m^{-2}	Pyranometer	CMP3	Kipp & Zonen, B.V., Delft, The Netherlands
Longwave outgoing radiation, *LW_OUT*	W m^{-2}	infra-red temperature sensor	IN510	Omega Engineering GmbH, Deckenpfronn, Germany
Photosynthetic Active Radiation, *PPFD_IN*	μmol_{CO_2} $m^{-2}s^{-1}$	PAR sensor	PQS 1	Kipp & Zonen, B.V., Delft, The Netherlands
Relative humidity, *RH*	%	Hygro-thermo transmitter-compact	1.1105.54.160	Adolf Thies GmbH & Co. KG, Göttingen, Germany
Air temperature, *TA*	°C	Hygro-thermo transmitter-compact	1.1105.54.160	Adolf Thies GmbH & Co. KG, Göttingen, Germany
CO_2 molar density, LC setup, r_c	μmol_{CO_2} $m_{air}{}^{-3}$	Infrared gas analyzer	GMP343	Vaisala Oyj, Helsinki, Finland
RH (from LC setup)	%	Relative humidity cell	HIH-4000	Honey-well, Charlotte, North Carolina, USA
Soil heat flux, *G*	W m^{-2}	Soil heat flux plates	HFP01	Hukseflux, Delft, The Netherlands
Wind velocity components, *u*, *v*, *w*	cm s^{-1}	Ultrasonic anemometer	uSONIC-3 Omni	METEK GmbH, Elmshorn, Germany

Fitting parameters between sum of predictor transformations and predicted fluxes using Alternating Conditional Expectation (ACE)

Table S2 shows the slope and the correlation coefficients of the fit between the sum of the optimal functions and the transformed dependent variable. Generally, the correlation was high and slopes were slightly above 1. This indicated a good explanation of the dependent variables (NEE, ET) by the selected predictors.

Table S2. Slopes and Pearson´s r correlation coefficients of the fit between the transformed variable (NEE or ET) and the sum of the ϕ functions representing the predictors' optimal transformations.

Station	NEE_U50	ET_U50
Dornburg winter AF	1.032 / 0.888	1.014 / 0.942
Dornburg winter OC	1.035 / 0.891	1.014 / 0.947
Dornburg summer AF	1.023 / 0.912	1.010 / 0.961
Dornburg summer OC	1.054 / 0.861	1.018 / 0.931
Forst winter AF	1.018 / 0.925	1.009 / 0.963
Forst winter OC	1.038 / 0.882	1.014 / 0.946
Forst summer AF	1.021 / 0.918	1.015 / 0.945
Forst summer OC	1.029 / 0.896	1.017 / 0.925
Wendhausen winter AF	1.034 / 0.884	1.011 / 0.950
Wendhausen winter OC	1.046 / 0.872	1.013 / 0.942
Wendhausen summer AF	1.018 / 0.931	1.015 / 0.937
Wendhausen summer OC	1.029 / 0.909	1.017 / 0.936
Mariensee AF	1.019 / 0.932	1.008 / 0.964
Mariensee OG	1.016 / 0.934	1.011 / 0.956

Generalized Additive Mixed Models fitting

Two joint models (main-effect joint model and joint-interaction model) were prepared to assess differences between AF and OC or OG, which fitted nonlinear smoothed splines for SPEI and SM, considered responses to Rg, TA and VPD as linear to reduce the risk of complicating the model shape estimate, and considered the year changes as random effects by calculating an intercept which captures the baseline. The main-effect joint model (Eq. 2 in the manuscript) was chosen because it provided more stable and coherent patterns in comparison to the joint-interaction model.

In the main-effect joint model (Eq. 2 in the manuscript) AF and OC/OG shared the same nonlinear-response to SPEI and SM, differing only by a constant shift across the predictor range. In the second model (interaction joint model; Eq. 1), the treatment AF or OC/OG was allowed to interact with the spline terms, therefore they differed in intercept and the response shape along the SPEI and SM ranges. It was represented by the following equation

$$y_i = \beta_0 + f_1(\mathrm{SPEI}_i, \mathrm{treatment}_i) + f_2(\mathrm{SM}_i, \mathrm{treatment}_i) + \beta_2\,\mathrm{Tair}_i + \beta_3\,\mathrm{VPD}_i + \beta_4\,\mathrm{Rg}_i + b_{\mathrm{year}(i)} + \varepsilon_i \tag{1}$$

where y_i is the response for observation i (ith-week of the time series), β_0 is the intercept or baseline level of the response, and $f_1(\mathrm{SPEI}_i)$ and $f_2(\mathrm{SM}_i)$ the nonlinear effects of SPEI and soil moisture (the spline), $f_1(\mathrm{SPEI}_i, \mathrm{treatment}_i)$ and $f_2(\mathrm{SM}_i, \mathrm{treatment}_i)$ allow the response curves to differ between treatments. $\beta_2\,\mathrm{Tair}_i$, $\beta_3\,\mathrm{VPD}_i$, and $\beta_4\,\mathrm{Rg}_i$ are linear effects of TA, VPD and Rg, included as covariates; $b_{\mathrm{year}(i)}$ is the random intercept for year, capturing interannual variability not explained by the fixed effects. Finally, ε_i is the residual error term, representing the remaining unexplained variation.

The joint model was successfully validated as appropriate by fitting a paired model (Eq. 2), which used the difference AF-OC/OG in NEE, ET or WUE as the dependent variable. This second formulation modeled the AF-OC/OG contrast while controlling for shared environmental conditions in each paired week. The comparison between both predictions is shown in Fig. S1 and Table S3.

GAMM quality was evaluated using different indicators, from which the most important ones are the Pearson´s r correlation coefficient, root mean squared error (RMSE), Mean Absolute Error (MAE), convergence (a logic variable), and observed-vs-fitted R^2 (variance explained). Together, these metrics assessed whether the model was stable and converged to produce observed responses. The values shown in Table S3 indicated a good response and convergence of the models, both for the main-effect model without joint interactions and for the comparison of the main-effect model to the paired model as a validation of the former. Only for Mariensee and for NEE vs. both SPEI and SM, the main-effect model did not converge; it was not possible to fit an appropriate response by changing the model parameters. Further investigation would be required to determine the best choice to get a robust model in that case.

Table S3. Diagnostics for the main-effect GAMM fit and for the curve agreement between main effect GAMM and paired-difference GAMM. MEM stands for main-effect model; MPC stands for main-to-paired comparison.

Site	Response	Predictor	r (MPC)	R^2 (MPC)	MAE (MPC)	RMSE (MPC)	N (MEM)	RMSE (MEM)	R^2 (MEM)	Conv (MEM)
Dornburg	ET_U50	SM	0.980	0.960	2.332	2.381	295	4.741	0.722	True
Dornburg	ET_U50	SPEI 90d	0.952	0.906	0.627	0.785	295	4.741	0.722	True
Dornburg	NEE_U50	SM	0.507	0.257	0.469	0.496	295	1.414	0.598	True
Dornburg	NEE_U50	SPEI 90d	0.901	0.811	0.116	0.165	295	1.414	0.598	True
Dornburg	WUE_U50	SM	0.112	0.012	0.548	0.583	295	1.281	0.412	True
Dornburg	WUE_U50	SPEI 90d	0.773	0.598	0.370	0.401	295	1.281	0.412	True
Forst	ET_U50	SM	0.742	0.550	1.359	1.421	401	3.840	0.780	True
Forst	ET_U50	SPEI 90d	0.517	0.267	4.271	4.365	401	3.840	0.780	True
Forst	NEE_U50	SM	0.836	0.700	0.259	0.285	401	1.054	0.479	True
Forst	NEE_U50	SPEI 90d	0.204	0.042	1.115	1.201	401	1.054	0.479	True
Forst	WUE_U50	SM	0.631	0.398	0.272	0.300	401	1.105	0.452	True
Forst	WUE_U50	SPEI 90d	0.437	0.191	0.579	0.667	401	1.105	0.452	True
Mariensee	ET_U50	SM	0.773	0.597	2.336	2.490	198	3.618	0.741	True
Mariensee	ET_U50	SPEI 90d	0.762	0.581	2.114	2.582	198	3.618	0.741	True
Mariensee	NEE_U50	SM	0.929	0.864	0.239	0.260	198	0.815	0.568	False
Mariensee	NEE_U50	SPEI 90d	0.951	0.904	0.187	0.201	198	0.815	0.568	False
Mariensee	WUE_U50	SM	0.907	0.823	0.262	0.319	198	0.989	0.595	True
Mariensee	WUE_U50	SPEI 90d	0.942	0.887	0.304	0.341	198	0.989	0.595	True
Wendhausen	ET_U50	SM	0.781	0.610	2.318	2.555	321	3.870	0.754	True
Wendhausen	ET_U50	SPEI 90d	0.702	0.493	1.209	1.403	321	3.870	0.754	True
Wendhausen	NEE_U50	SM	0.742	0.551	0.310	0.348	321	1.205	0.556	True
Wendhausen	NEE_U50	SPEI 90d	0.667	0.445	0.690	0.759	321	1.205	0.556	True
Wendhausen	WUE_U50	SM	0.761	0.580	0.325	0.406	321	1.008	0.380	True
Wendhausen	WUE_U50	SPEI 90d	0.953	0.909	0.267	0.289	321	1.008	0.380	True

The equation of the paired GAMM reads as follows:

$$\Delta y_i = \alpha_0 + h_1\left(\overline{\mathrm{SPEI}}_i\right) + h_2\left(\overline{\mathrm{SM}}_i\right) + \alpha_1\,\overline{\mathrm{Tair}}_i + \alpha_2\,\overline{\mathrm{VPD}}_i + \alpha_3\,\overline{\mathrm{Rg}}_i + u_{\mathrm{year}(i)} + \eta_i \tag{2}$$

with $\Delta y_i = y_{i,\mathrm{AF}} - y_{i,\mathrm{OC/OG}}$ as the NEE, ET or WUE difference between AF and OC/OG in the ith-week; $h_1(\overline{\mathrm{SPEI}}_i)$ and $h_2(\overline{\mathrm{SM}}_i)$ the smooth nonlinear functions of the paired covariate means, with $\overline{x}_i = (x_{i,\mathrm{AF}} + x_{i,\mathrm{OC/OG}})/2$; α_1, α_2, and α_3 the linear effects of paired-mean meteorological covariates $\overline{\mathrm{Tair}}_i$, $\overline{\mathrm{VPD}}_i$, and $\overline{\mathrm{Rg}}_i$; $u_{\mathrm{year}(i)}$ as the random term accounting for interannual variability; and η_i the residual error.

GAMMs were fitted using the libraries *statsmodels* (Seabold and Perktold, 2010) and *patsy* (Smith et al., 2024) in Python, v. 3.11.7.

The comparison between the response by the main-effect and the paired GAMMs indicated a good agreement between both models (Fig. S1). For most of the sites and variables, the response curve was very similar between models, showing just a shift; in some cases, the response curves were diverging, such as for ET vs. SPEI in Mariensee, or for WUE vs. SM in Forst. As there was generally an agreement between curves, the validation of the main-effect model was considered successful.

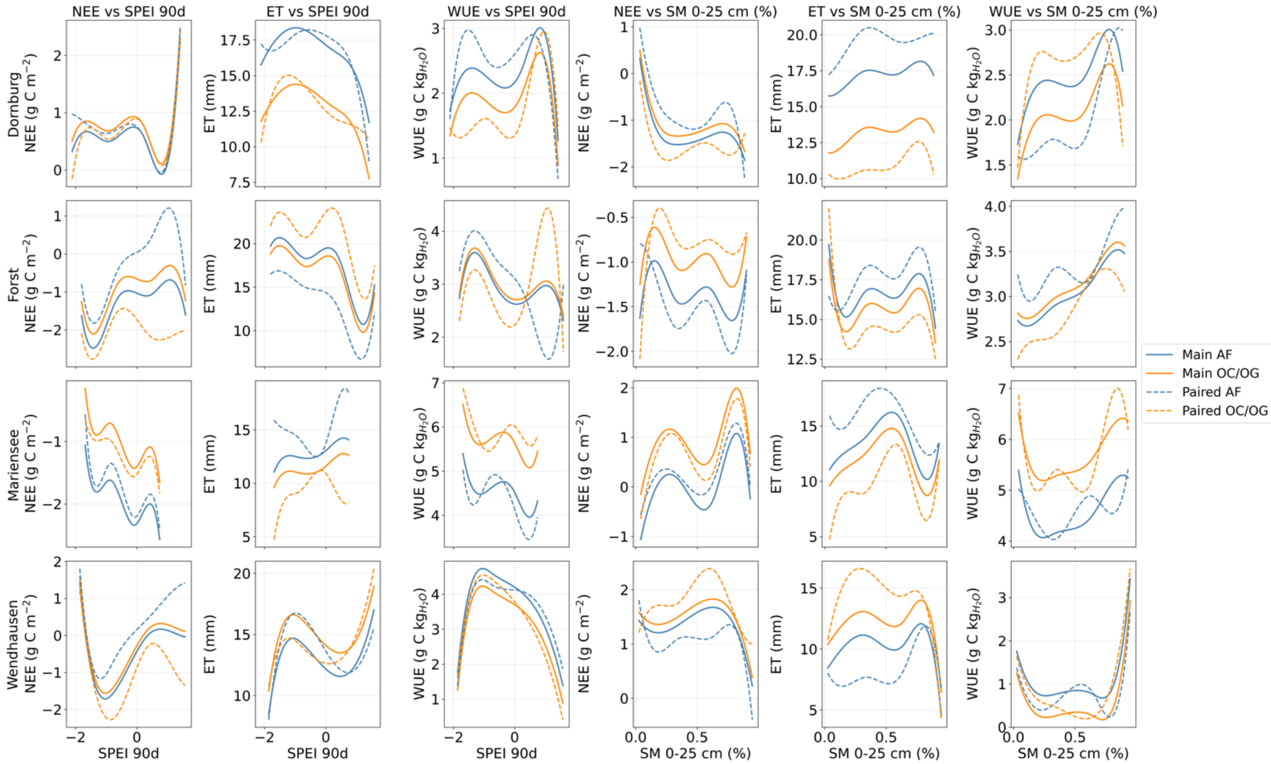

Figure S1. Comparison between response curves (fitted mean) of the main-effect GAMM for AF (blue, solid lines) and OC/OG (orange, solid lines) vs. paired model for AF (blue, dashed lines) and OC/OG (orange, dashed lines), for the behavior of NEE, ET and WUE in relation to SPEI and SM. The paired model curves were reconstructed around the mid-point of the main-effect model.

Meteorological conditions

Meteorological variables exhibited a marked seasonality, typical in these climate zones (Fig. S2): large R_G values, high VPD and warm TA in summer, and low values of the three variables in winter. Precipitation was more abundant in spring with some strong events in summer. SPEI and SM oscillated also with summer and winter transitions, with more dry conditions generally in summer; but there seemed to be a trend: 2019 and 2020 were dry, then conditions became more humid in 2021, then 2022 summer was the driest period, and finally both SPEI and SM increase to indicate more humid conditions in 2023 and 2024, closer to normality for these regions.

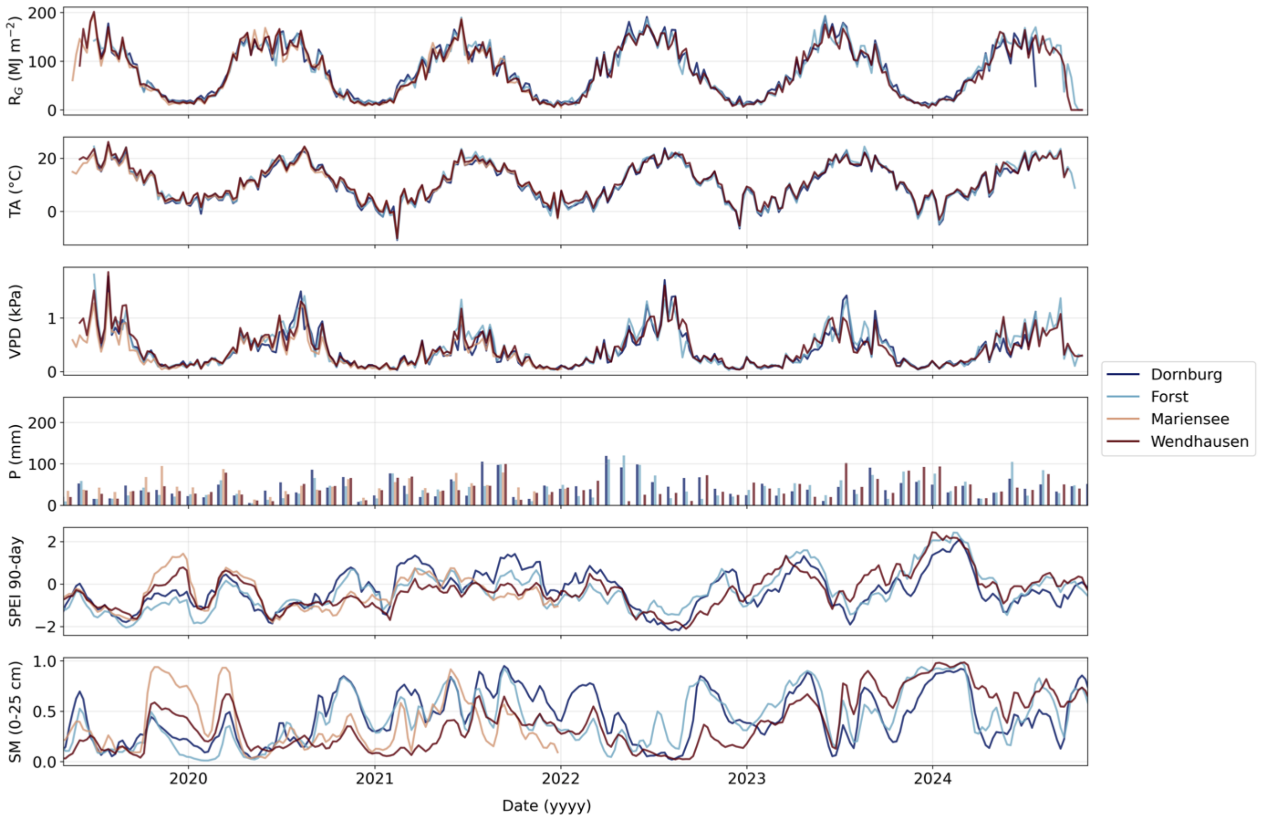

Figure S2. Weekly aggregates of global radiation (R_G; sum), air temperature (TA; mean), vapor pressure deficit (VPD; mean), precipitation (P; sum), standardized precipitation and evaporation index (SPEI; mean) at the 90-days scale, and soil moisture (SM; mean) as relative content, for Dornburg (dark blue line), Forst (light blue line), Mariensee (orange line) and Wendhausen (burgundi line).

Yearly dynamics of GPP and Reco

Cumulative sums of GPP were generally larger for AF than OC/OG, except in 2021 in Mariensee and 2019 in Wendhausen (Fig. S3). Nevertheless, in some years, like 2021 in Forst, 2024 in Dornburg, or 2024 in Wendhausen the differences in GPP between AF and OC were small. This would require further investigation to determine if it is an ecosystem response or a model effect. Nighttime model GPP was larger than daytime model GPP, but the difference AF-OC/OG had a similar ratio for both models, which can be considered an indicator of robustness despite the shift between models.

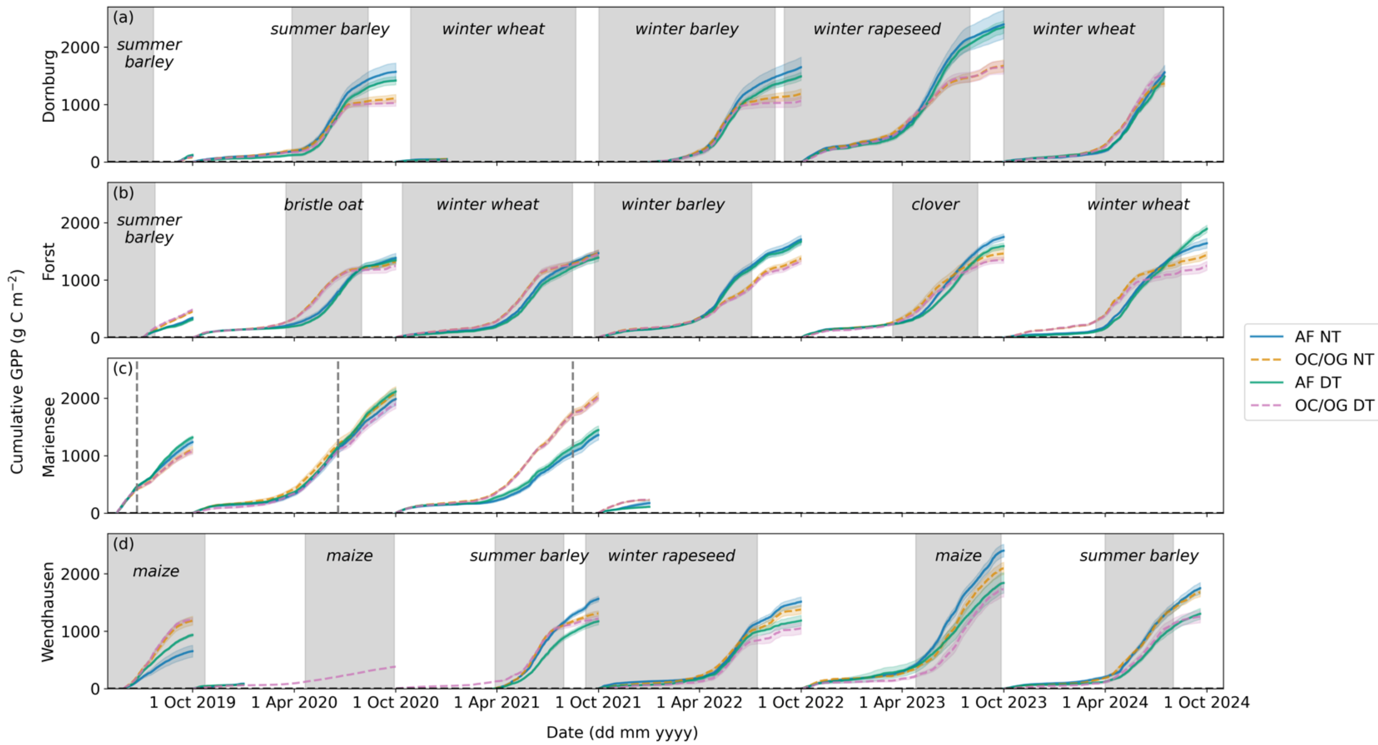

Figure S3. Cumulative sums of gross primary production (GPP) for all the years in the measured period, for Dornburg (a), Forst (b), Mariensee (c) and Wendhausen (d), for AF (blue) and OC or OG (orange). Years were considered from 1 October until 30 September of the following year. 2019 was only partially covered. In Dornburg, Forst and Wendhausen, grey shading corresponds to crops growing season from sowing to harvest, whereas in Mariensee the vertical dashed lines indicate the major grass cuts.

RECO sums were larger at AF than OC in Dornburg, Forst and Wendhausen, with few exceptions (2023 in Forst and 2024 in Wendhausen), but smaller at AF than OG in Mariensee (Fig. S4). The nighttime partitioning model showed the largest RECO in Dornburg, but in Forst it was very similar between both models, and in Wendhausen the daytime partitioning model generally produced a larger RECO. Contrary to GPP, the difference AF-OC/OG was not similar between nighttime and daytime models.

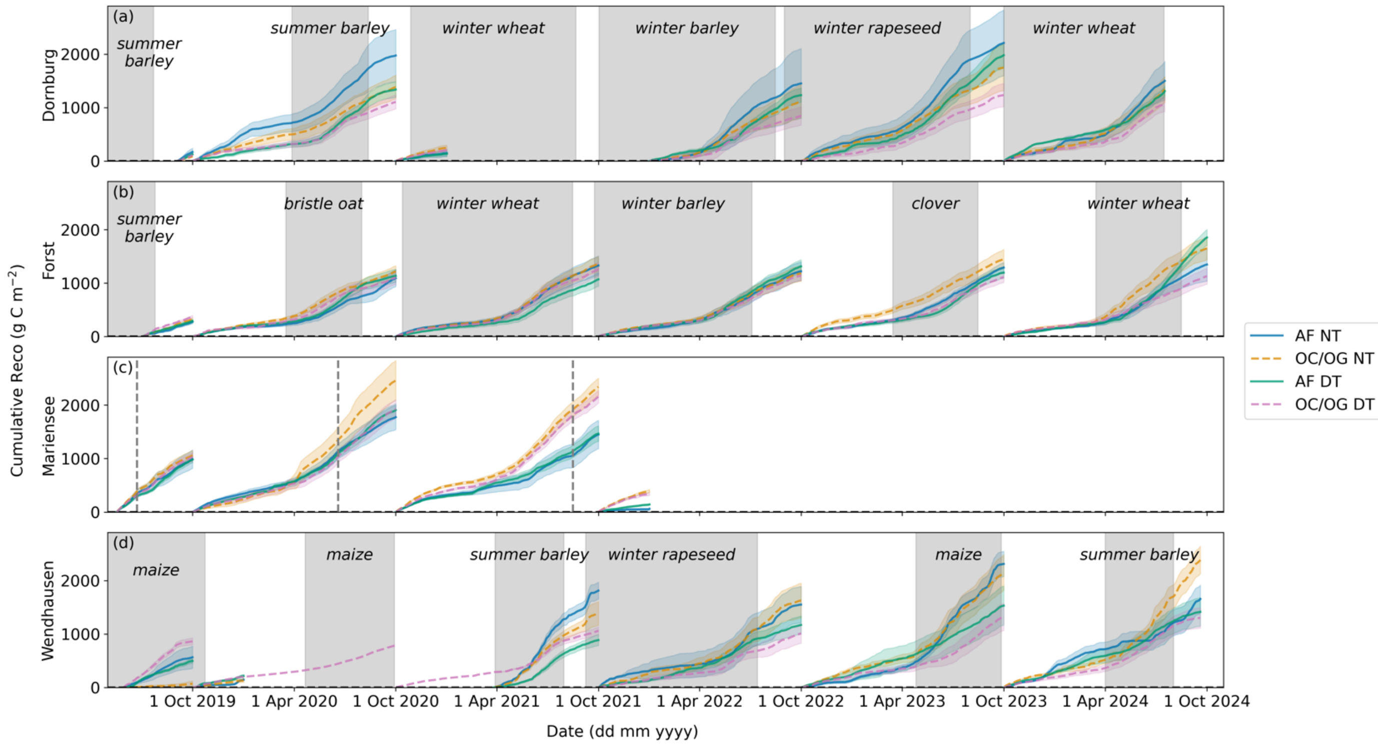

Figure S4. Cumulative sums of ecosystem respiration (RECO) for all the years in the measured period, for Dornburg (a), Forst (b), Mariensee (c) and Wendhausen (d), for AF (blue) and OC or OG (orange). Years were considered from 1 October until 30 September of the following year. 2019 was only partially covered. In Dornburg, Forst and Wendhausen, grey shading corresponds to crops growing season from sowing to harvest, whereas in Mariensee the vertical dashed lines indicate the major grass cuts.

References

Markwitz, C. and Siebicke, L.: Low-Cost Eddy Covariance: A Case Study of Evapotranspiration over Agroforestry in Germany, Atmospheric Measurement Techniques, 12, 4677–4696, https://doi.org/10.5194/amt-12-4677-2019, 2019.

Seabold, S. and Perktold, J.: Statsmodels: Econometric and Statistical Modeling with Python, in: 9th Python in Science Conference, 2010.

Smith, N. J., Wardrop, M., Hudon, C., broessli, Quackenbush, P., Seabold, S., Portnoy, A., Beasley, B., Cameron Davidson-Pilon, Kibirige, H., Leinweber, K., Sheppard, K., Humber, M., Colange, M., Michael Hudson-Doyle, Korenčiak, M., and Gates, T.: Pydata/Patsy: V0.5.6, Zenodo, https://doi.org/10.5281/ZENODO.592075, 2024.