

Supplementary Information

Vegetation biogeography is a main source of uncertainty in modelling the land carbon cycle

Ruiying Zhao^{1*}, Xiangzhong Luo^{1, 2*}, Anthony P. Walker³, Forrest M. Hoffman⁴, Lian Pin Koh²

¹ Department of Geography, National University of Singapore, Singapore

² Center for Nature-based Climate Solutions, National University of Singapore, Singapore

³ Environmental Sciences Division and Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN, USA

⁴ Computational Science and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

* Correspondence to: ruiying@nus.edu.sg; xzluo.remi@nus.edu.sg

Supplementary Table 1. The dynamic global vegetation models (DGVMs) from TRENDY v9, and the availability of their simulations of the area and gross primary productivity (GPP) for different plant functional types (PFTs) ^a.

DGVMs	PFTs									PFT distribution	Note
	EBF	ENF	DBF	DNF	SHR	C3G	C4G	C3C	C4C		
CABLE	YES	YES	YES	YES	YES	YES	YES	NO	NO	Prescribed	
CLASS	YES	YES	YES	YES	NO	YES	YES	YES	YES	Prescribed	
CLM5	YES	YES	YES	YES	YES	YES	YES	YES	YES	Prescribed	C4C includes Corn, Millet, Sorghum and Sugarcane
JSBACH	YES	YES	YES	YES	YES	YES	YES	YES	YES	Prescribed	C3C includes C ₃ crop and C ₃ pasture; C4C includes C ₄ crop and pasture
JULES	YES	YES	YES	YES	YES	YES	YES	YES	YES	Simulated	C3C and C4C area the same as JSBACH
LPJ_GUESS	YES	YES	YES	YES	NO	YES	YES	YES	YES	Simulated	DBF includes boreal needleleaf summergreen, temperate broadleaf summergreen, temperate shade-intolerant broadleaf summergreen
LPX_Bern	YES	YES	YES	YES	NO	YES	YES	YES	YES	Simulated	DBF includes tropical broad raingreen, temperate and boreal broad summergreen
ORCHIDEE	YES	YES	YES	YES	NO	YES	YES	YES	YES	Prescribed	C3C includes cropland C ₃ herbaceous and pasture C ₃ herbaceous; C4C includes cropland C ₄ herbaceous and pasture C ₄ herbaceous
ORCHIDEE-CNP	YES	YES	YES	YES	NO	YES	YES	YES	YES	Prescribed	DBF includes tropical broadleaf raingreen, temperate and boreal broadleaf summergreen
SDGVM	YES	YES	YES	YES	NO	YES	YES	YES	YES	Simulated	Same as ORCHIDEE, but not used, as GPP values are quite low
YIBs	YES	YES	YES	NO	YES	YES	YES	YES	YES	Prescribed	

^a The PFTs include evergreen needleleaf forests (ENF), evergreen broadleaf forests (EBF), deciduous needleleaf forests (DNF), deciduous broadleaf forests (DBF),

shrubs (SHR), C₃ grasses (C3G), C₄ grasses (C4G), C₃ crops (C3C), and C₄ crops (C4C).

Supplementary Table 2. Environmental constraints ^a used to identify types of subtypes for forests, evergreen needleleaf forests (ENF) and C₃ grasses (C3G)¹.

PFTs	Sub PFTs	Environmental constraints
EBF		$T_c > 15.5; \alpha > 0.80$
DBF		$T_c > -15 \text{ \& } T_c < 15.5; GDD_5 > 1200; \alpha > 0.65$
ENF	Temperate ENF (ENFt)	$T_c > -19 \text{ \& } T_c < 5; GDD_5 > 900; \alpha > 0.65$ ^b
	Boreal ENF (ENFb)	$T_c > -35 \text{ \& } T_c < -2; GDD_5 > 350; \alpha > 0.75$
DNF		$T_c < 5; GDD_5 > 350; \alpha > 0.65$
C3G	Warm C3G (C3Gw)	$T_w > 22; \alpha > 0.18$
	Cool C3G (C3Gk)	$GDD_5 > 500; \alpha > 0.33$
	Cold C3G (C3Gc)	$GDD_0 > 100; \alpha > 0.33$

^a The environmental constraints include the mean temperature of the coldest month (T_c) and warmest month (T_w), accumulated growing degree-days above 5°C (GDD_5) and 0°C (GDD_0), and Priestley-Taylor coefficient of annual moisture availability (α).

^b α is calculated as the ratio of actual evapotranspiration (AET) to potential evapotranspiration (PET).

Supplementary Table 3. The PFT area estimates across five remote sensing-based products (unit: 10^6 km^2). Only MODIS and ESA CCI PFT maps have reported part of their own uncertainty. σ refers to inter-product uncertainty (i.e., one standard deviation of the area from five products).

PFTs	MODIS	unc	ESA CCI	unc	CGLS_ LC100	unc	GLC_ FCS30D	unc	GLASS- GLC	unc	σ
EBF	21.56	1.23	14.23	2.85	18.30	NA	15.30	NA	13.97	NA	3.23
ENFt	5.08	1.21	4.17	0.83	7.32	NA	5.24	NA	2.61	NA	1.72
ENFb	4.69	1.11	3.41	0.68	6.20	NA	5.13	NA	8.17	NA	1.79
DBF	16.13	2.79	7.21	1.44	11.68	NA	16.57	NA	9.81	NA	4.04
DNF	4.77	0.83	2.33	0.47	6.04	NA	6.32	NA	4.56	NA	1.58
SHR	14.01	1.73	4.68	0.42	11.33	NA	14.17	NA	18.48	NA	5.08
C3Gw	9.29	0.58	12.73	0.64	8.52	NA	9.96	NA	7.81	NA	1.90
C3Gk	5.84	0.36	12.31	0.62	8.16	NA	6.46	NA	10.38	NA	2.71
C3Gc	3.95	0.25	4.57	0.23	3.87	NA	5.08	NA	3.32	NA	0.68
C3C	7.66	0.71	16.34	0.82	10.90	NA	13.61	NA	7.86	NA	3.74
C4G	17.26	1.09	17.26	1.09	16.62	NA	17.26	NA	17.45	NA	0.32
C4C	3.23	0.30	3.23	0.30	3.47	NA	3.23	NA	3.16	NA	0.12

Note: the remote sensing-based products include MODIS v6.1², ESA CCI v2.0.8³, CGLS_LC100 v3⁴, GLC_FCS30D⁵, and GLASS-GLC⁶. Uncertainties (unc) for MODIS and ESA CCI PFT maps are derived from literature (see Methods). Uncertainties for other products are not available (NA).

Supplementary Table 4. The coefficient of determination (R^2) of emulators for PFT GPP simulations.

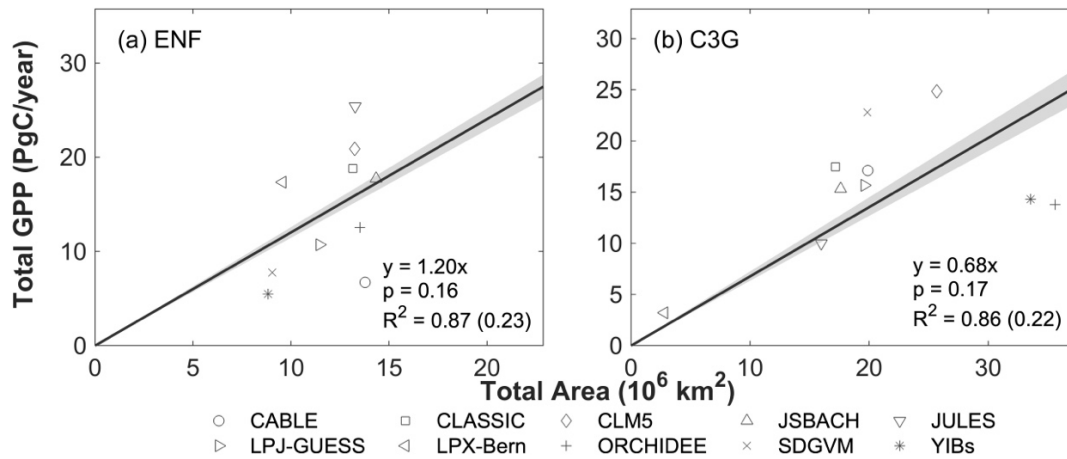
		CABLE	CLASSIC	CLM5	JSBACH	JULES	LPX- Bern	ORCHIDEE	SDGVM	YIBs
EBF	cal	0.98	0.99	0.87	0.99	0.98	0.38	0.93	0.99	0.95
	val	0.98	0.99	0.86	0.99	0.98	0.54	0.93	0.98	0.95
ENFt	cal	0.97	0.98	0.99	0.89	0.94	0.27	0.85	0.97	0.74
	val	0.97	0.98	0.99	0.88	0.94	0.76	0.85	0.97	0.72
ENFb	cal	0.96	0.98	0.98	0.80	0.93	0.95	0.84	0.89	0.83
	val	0.96	0.97	0.97	0.78	0.91	0.95	0.84	0.88	0.82
DBF	cal	0.97	0.99	0.97	0.99	0.98	0.46	0.91	0.95	0.77
	val	0.97	0.99	0.97	0.99	0.98	0.84	0.91	0.95	0.76
DNF	cal	0.98	0.97	0.92	0.87	0.94	0.91	0.61	0.93	0.77
	val	0.98	0.97	0.92	0.85	0.94	0.86	0.61	0.93	0.76
SHR	cal	0.98	0.98	0.95	0.99	0.97	0.44	0.84	0.96	0.66
	val	0.98	0.98	0.95	0.99	0.97	0.38	0.84	0.96	0.66
C3Gw	cal	0.96	1.00	0.96	0.98	0.99	0.77	0.78	0.90	0.70
	val	0.96	0.99	0.96	0.97	0.99	0.60	0.78	0.90	0.69
C3Gk	cal	0.96	0.98	0.91	0.94	0.98	0.44	0.84	0.96	0.59
	val	0.96	0.98	0.91	0.94	0.98	0.38	0.84	0.96	0.59
C3Gc	cal	0.97	0.99	0.91	0.96	0.96	0.86	0.65	0.92	0.80
	val	0.94	0.98	0.91	0.95	0.95	0.86	0.64	0.91	0.77
C3C	cal	0.96	0.96	0.98	0.97	0.99	0.88	0.84	0.95	0.74
	val	0.96	0.95	0.97	0.97	0.99	0.84	0.83	0.95	0.73
C4G	cal	0.97	0.99	0.91	0.96	1.00	0.93	0.91	0.84	0.77
	val	0.97	0.99	0.90	0.96	1.00	0.92	0.91	0.83	0.77
C4C	cal	0.97	0.95	0.90	0.95	0.99	0.84	0.86	0.96	0.80
	val	0.97	0.94	0.90	0.94	0.99	0.94	0.86	0.96	0.78

Note: cal and val indicate the R^2 for training and validation, respectively.

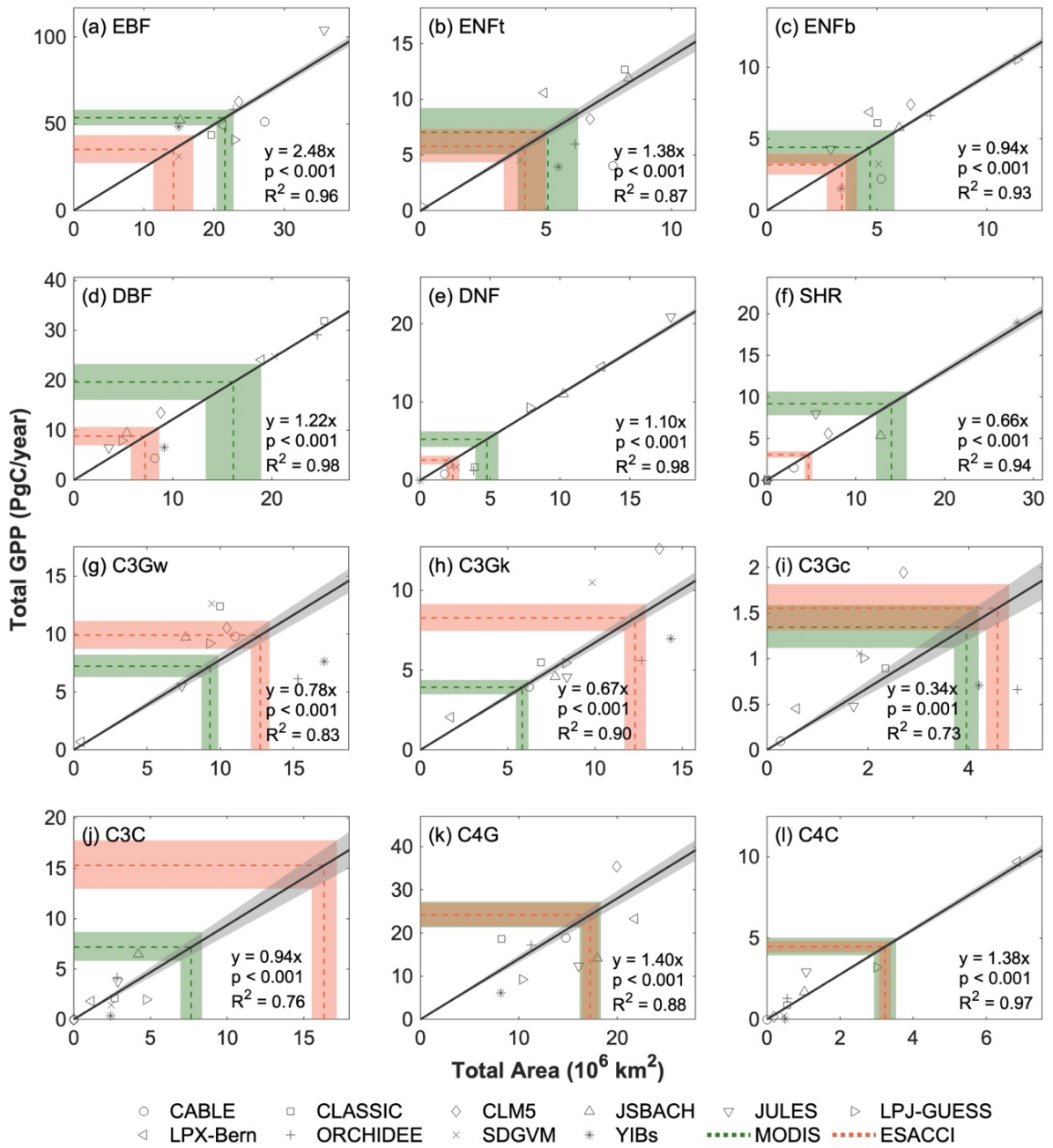
Supplementary Table 5. Contributions of PFT Area, LAI and leaf photosynthetic rate (Aleaf) to inter-model GPP spread quantified by multivariate regression analysis and variance decomposition analysis (Shapley regression).

PFT	<i>r</i>			Shapley values		
	Area	LAI	Aleaf	Area	LAI	Aleaf
EBF	0.839	0.372	0.361	0.672	0.077	0.056
ENFt	0.807	0.617	0.522	0.543	0.266	0.179
ENFb	0.868	0.416	0.246	0.678	0.096	0.032
DBF	0.890	0.230	0.189	0.872	0.047	0.054
DNF	0.856	0.015	-0.066	0.842	0.115	0.030
SHR	0.983	0.282	-0.013	0.936	0.043	-0.001
C3Gw	0.591	0.381	-0.223	0.272	0.162	0.066
C3Gk	0.633	0.494	0.021	0.483	0.338	0.035
C3Gc	0.641	0.348	-0.159	0.351	0.157	0.056
C3C	0.690	0.541	0.496	0.405	0.086	0.104
C4G	0.237	0.460	0.265	0.332	0.465	0.044
C4C	0.312	0.148	0.087	0.521	0.439	0.010
AVG	0.696	0.359	0.144	0.576	0.191	0.055

Note: *r* is the semi-partial correlation coefficients from the multivariate regression and Shapley values indicate contributions of each predictor to total explained variance. Large absolute *r* and Shapley values indicate a larger contribution from the factor. AVG is the average contribution from each factor to the spread of inter-model GPP across all PFTs.

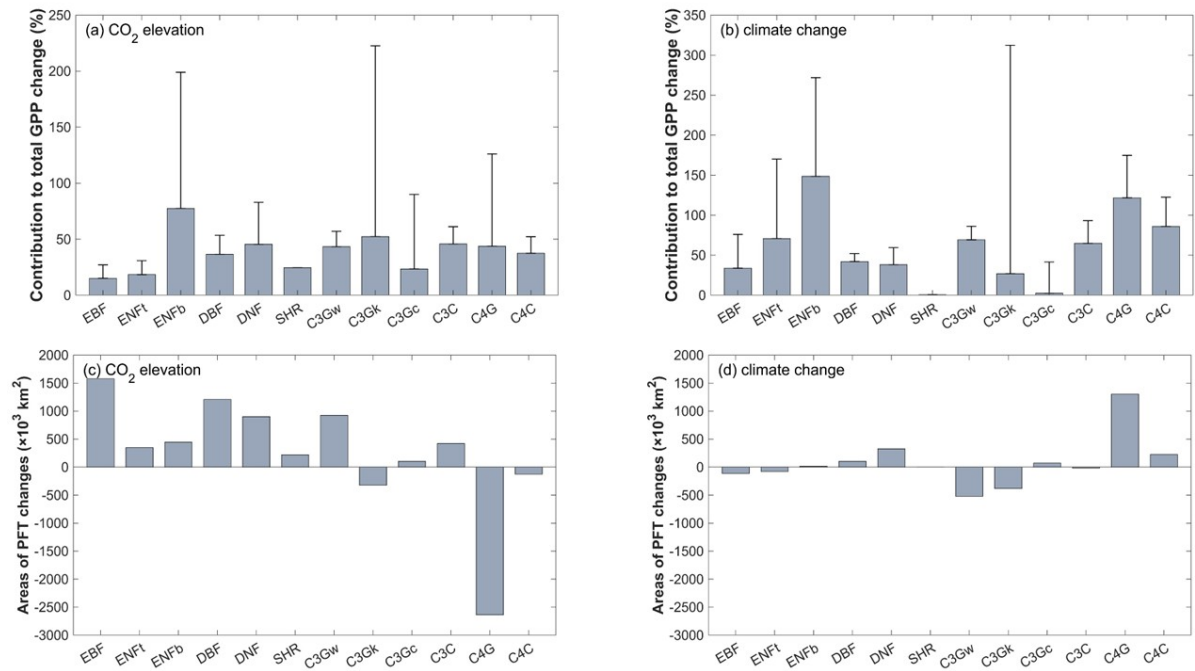


Supplementary Figure 1. The relationships between GPP and total area for ENF and C3G. The panels represent the relationships between GPP and area for (a) evergreen needleleaf forests (ENF), and (b) C₃ grasses (C3G). The uncertainty was quantified as the upper and lower boundary of relationship through a bootstrapping method for the regression (see Method). The regression lines were forced through (0,0) as there would be no GPP if PFT area was zero. The strength of the relationships was assessed by the coefficient of determination (R^2). Meanwhile, we also provided R^2 for the regression lines without forcing through (0,0) in parentheses.

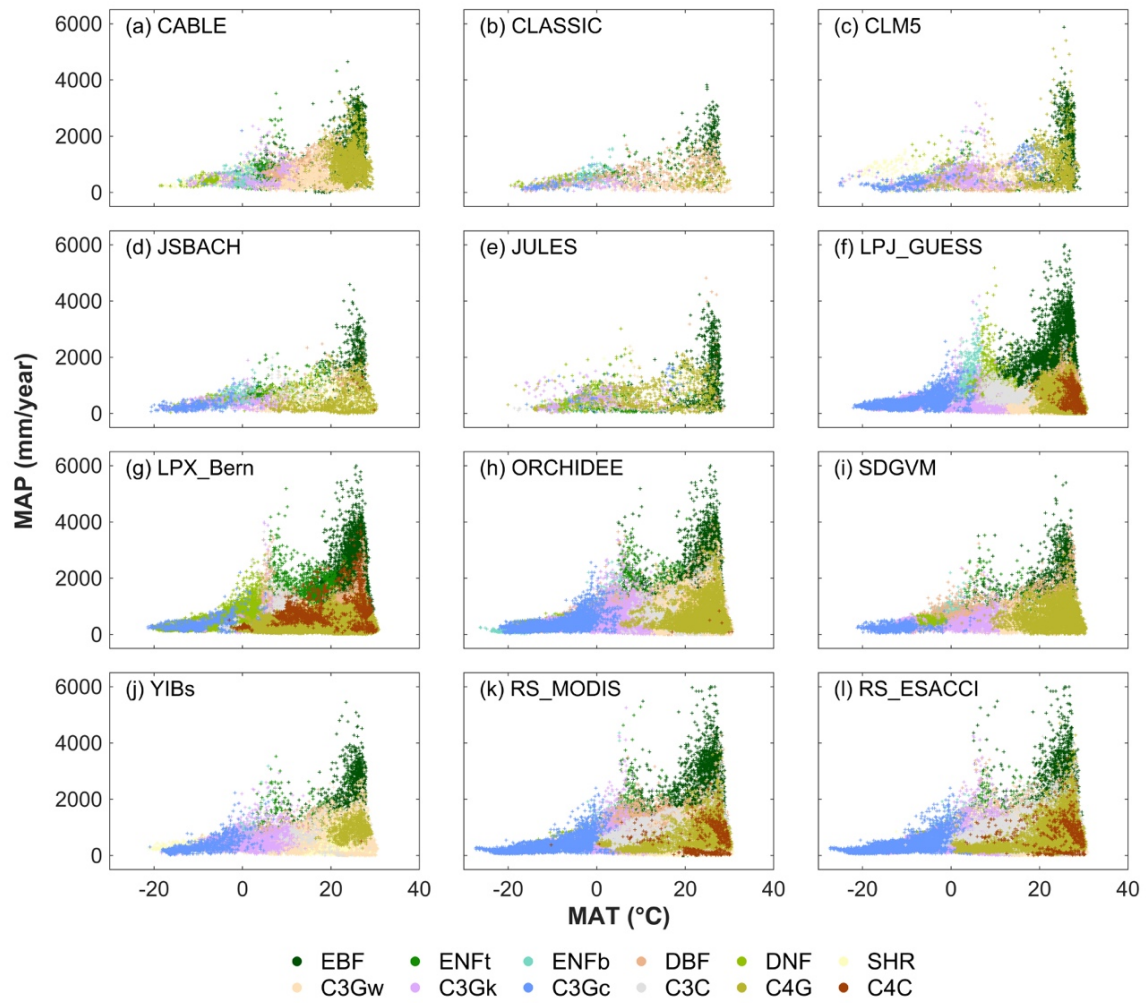


Supplementary Figure 2. The relationships between total GPP and total area for each PFT. Here we presented the results similar to Supplementary Figure 1, but showing the PFT areas observed from MODIS and ESA CCI only. The vertical shadings indicate the one standard deviation of each PFT map derived from existing literature (see Methods). The horizontal green and red lines and shading intervals represent the constrained total GPP for each PFT using MODIS or ESA CCI PFT area, and the uncertainty in GPP propagated from the PFT area. The panels represent the results for **(a)** evergreen broadleaf forests (EBF), **(b)** temperate evergreen needleleaf forests (ENFt), **(c)** boreal evergreen needleleaf forests (ENFb), **(d)** deciduous

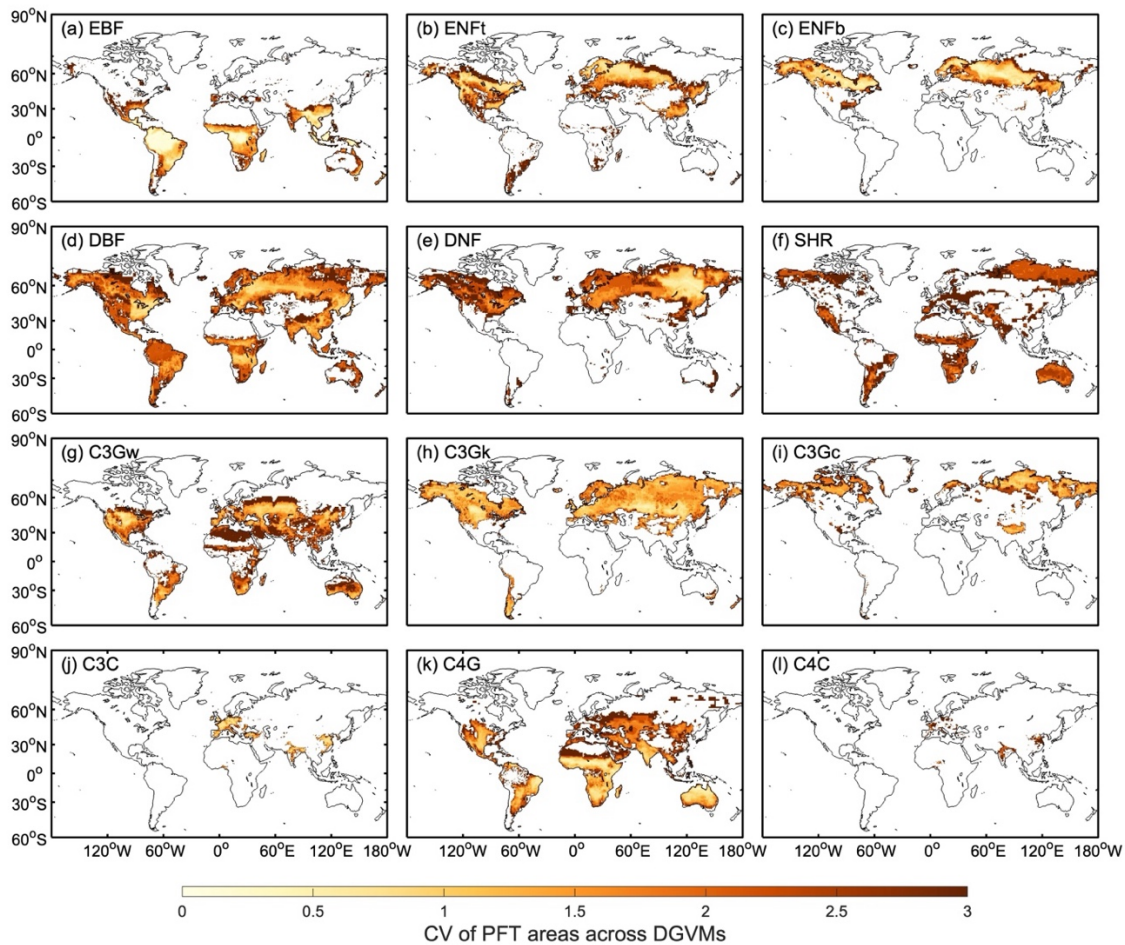
broadleaf forests (DBF), **(e)** deciduous needleleaf forests (DNF), **(f)** shrubs (SHR), **(g)** warm C3 grasses (C3Gw), **(h)** cool C3 grasses (C3Gk), **(i)** cold C3 grasses (C3Gc), **(j)** C3 crops (C3C), **(k)** C4 grasses (C4G), and **(l)** C4 crops (C4C).



Supplementary Figure 3. The contributions of PFT changes to GPP changes from 2001 to 2019. (a) The contribution of elevated CO₂-driven PFT changes to CO₂-driven GPP changes for each PFT; **(b)** The contribution of climate change-driven PFT changes to climate change-driven GPP changes for each PFT. The bar charts include mean and one standard deviation lines, which represent the uncertainties across DGVMs; **(c)** and **(d)** presented the elevated CO₂-driven and climate change-driven area changes for each PFT.

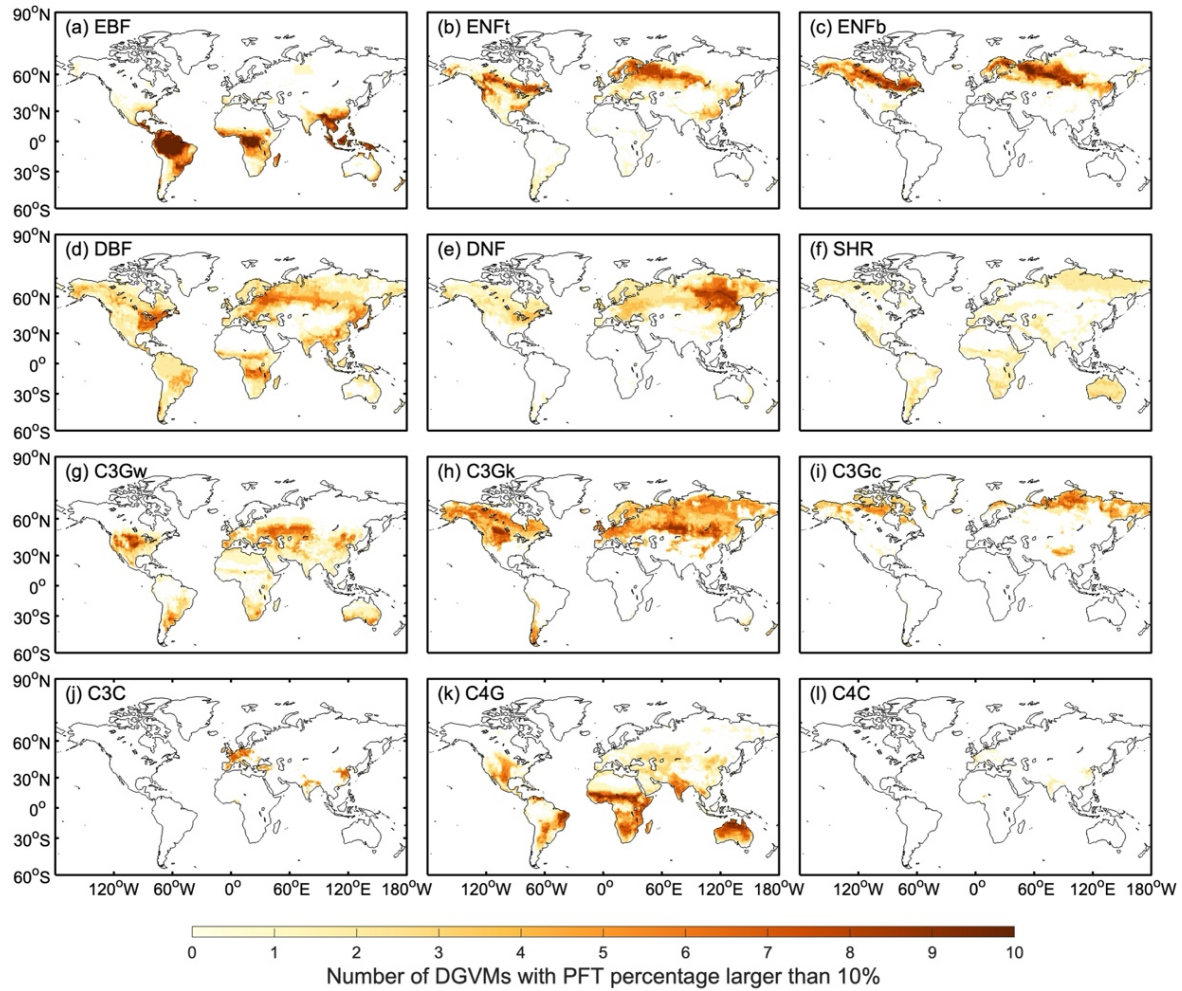


Supplementary Figure 4. The PFT distributions along the gradients of mean annual temperature (MAT) and mean annual precipitation (MAP) from DGVMs and remote sensing-based (RS) products. The (a) to (j) represent the PFT distributions for each DGVM, with the number of points varying due to differences in the spatial resolution of the DGVMs. (k) and (l) represent the RS PFT maps from MODIS (RS_MODIS) and ESA CCI (RS_ESACCI). The abbreviations for PFTs correspond to those in Supplementary Figure 1.

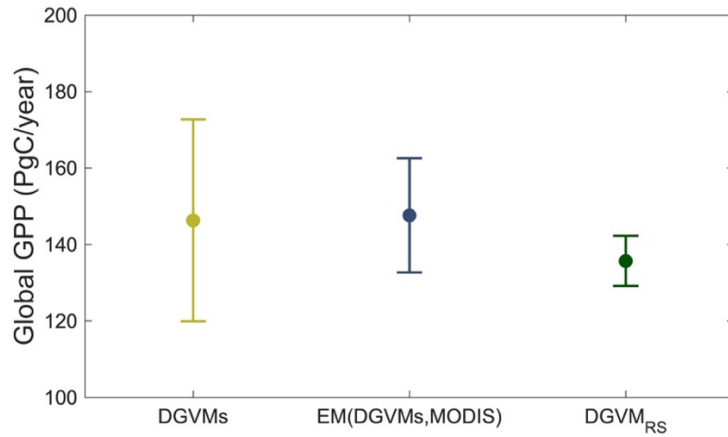


Supplementary Figure 5. The difference in the distribution of PFTs in DGVMs.

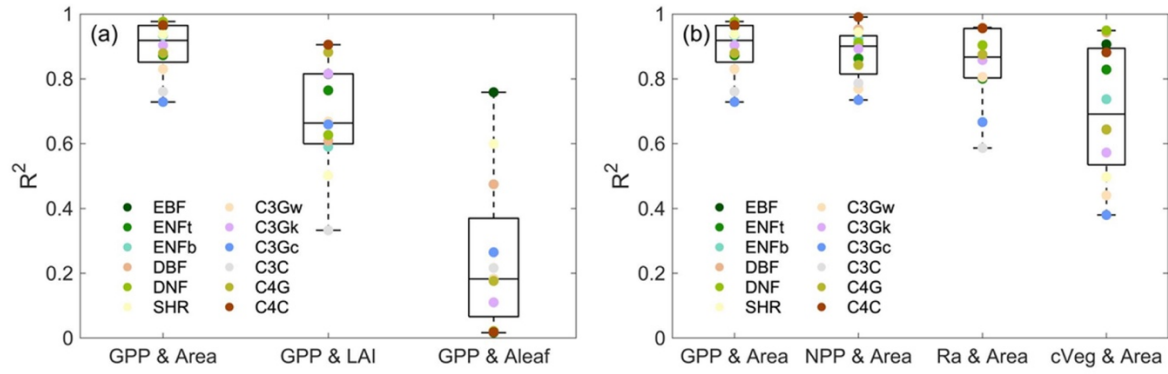
The coefficient of variation (CV) for each PFT distribution across DGVMs is calculated as the ratio of the standard deviation to the mean PFT area across DGVMs, with higher CV values indicating greater uncertainties in PFT simulations from the ensemble of DGVMs. The panels represent the results for **(a)** evergreen broadleaf forests (EBF), **(b)** temperate evergreen needleleaf forests (ENFt), **(c)** boreal evergreen needleleaf forests (ENFb), **(d)** deciduous broadleaf forests (DBF), **(e)** deciduous needleleaf forests (DNF), **(f)** shrubs (SHR), **(g)** warm C3 grasses (C3Gw), **(h)** cool C3 grasses (C3Gk), **(i)** cold C3 grasses (C3Gc), **(j)** C3 crops (C3C), **(k)** C4 grasses (C4G), and **(l)** C4 crops (C4C).



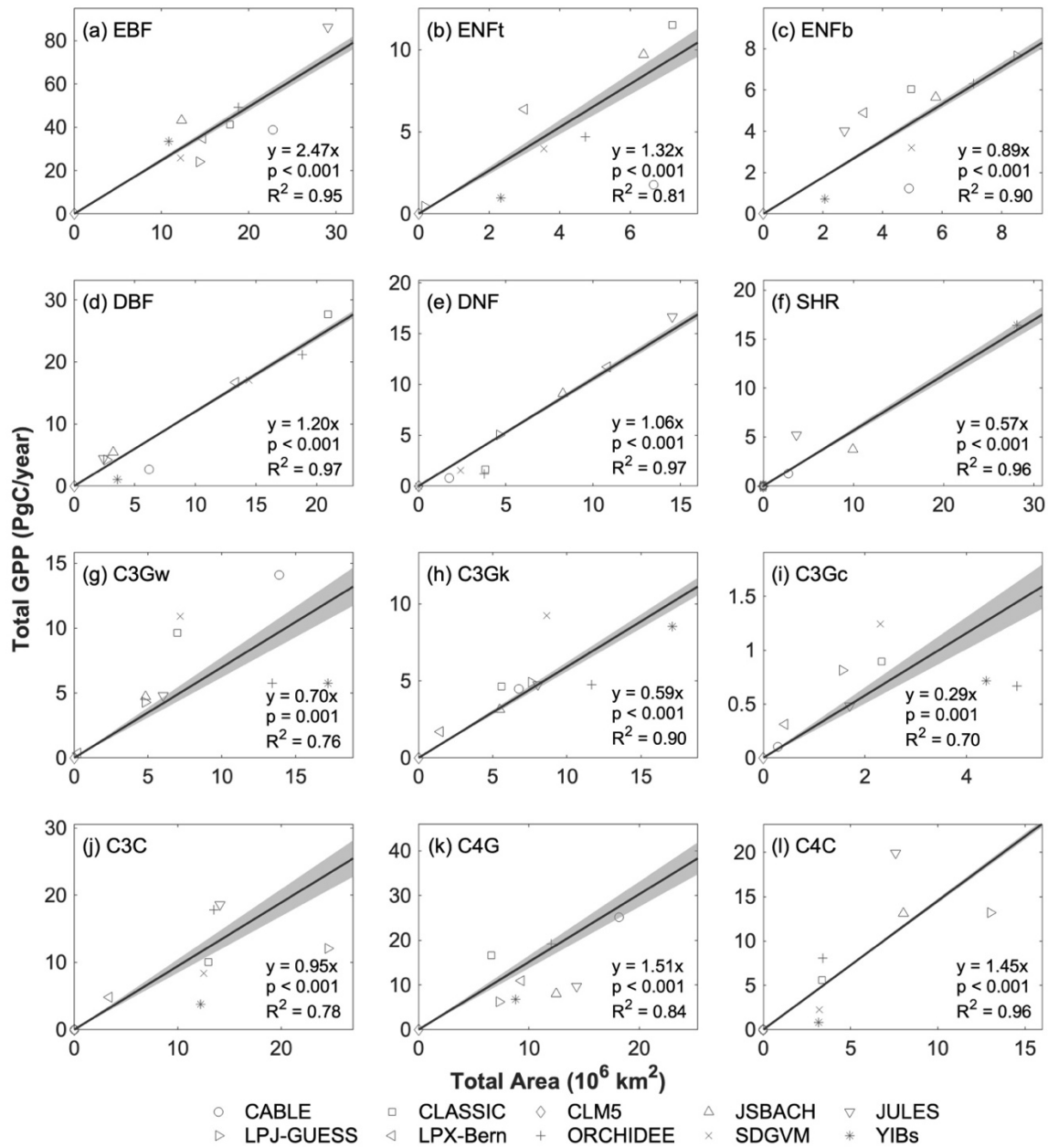
Supplementary Figure 6. The difference in the distribution of PFTs in DGVMs, quantified by another method. Here we presented the number of DGVMs in each grid cell for each PFT. The number of DGVMs where PFT percentage larger than 10% in the same grid, with higher values indicating a greater consistency. The panels represent the results for **(a)** evergreen broadleaf forests (EBF), **(b)** temperate evergreen needleleaf forests (ENFt), **(c)** boreal evergreen needleleaf forests (ENFb), **(d)** deciduous broadleaf forests (DBF), **(e)** deciduous needleleaf forests (DNF), **(f)** shrubs (SHR), **(g)** warm C3 grasses (C3Gw), **(h)** cool C3 grasses (C3Gk), **(i)** cold C3 grasses (C3Gc), **(j)** C3 crops (C3C), **(k)** C4 grasses (C4G), and **(l)** C4 crops (C4C).



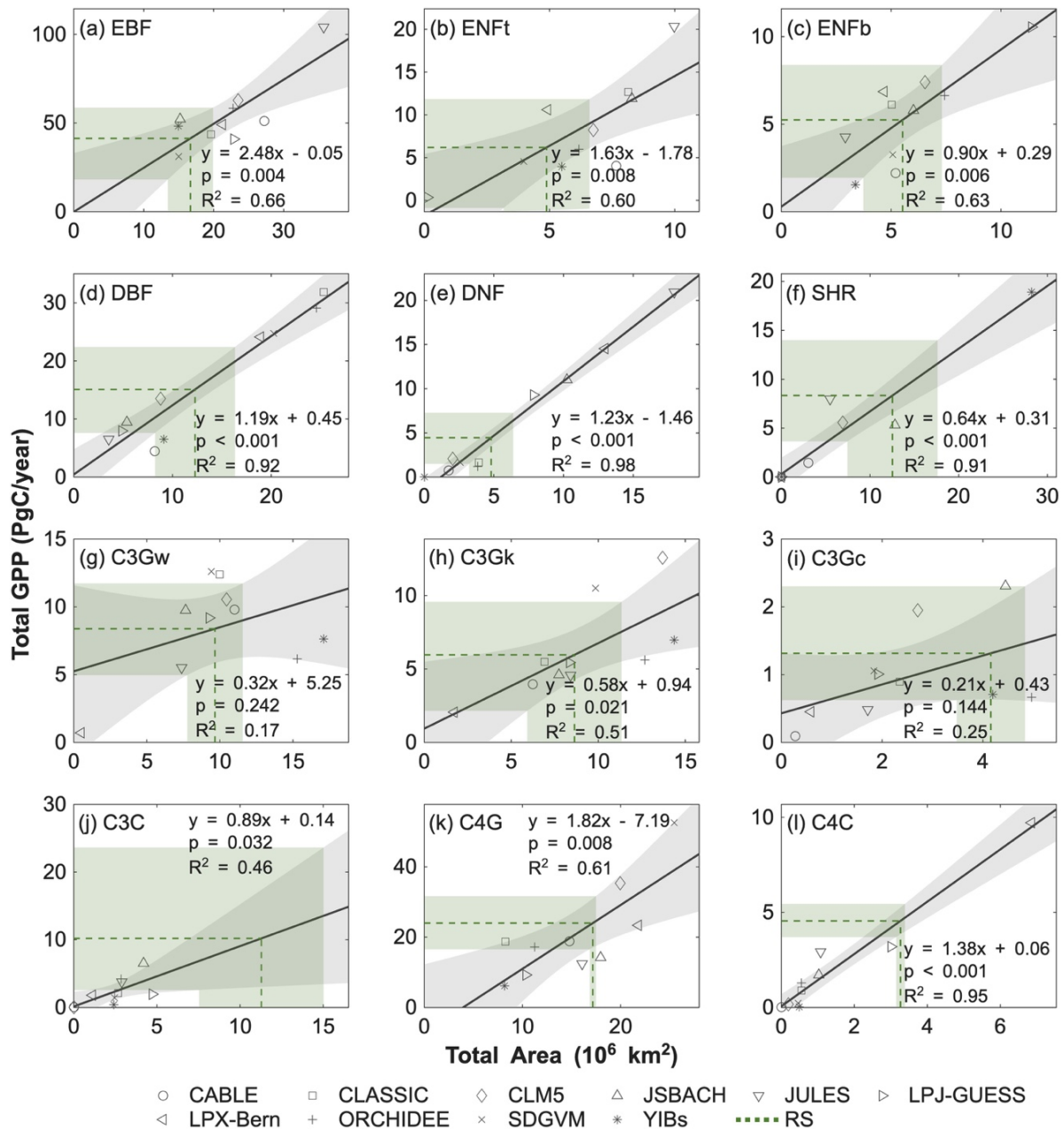
Supplementary Figure 7. The spread of global GPP estimates from original DGVMs and DGVMs forced by MODIS PFT map. The three group of models are 1) DGVMs – the original group of DGVMs, with biased PFT area and location; 2) EM(DGVMs, MODIS) – the emulators of DGVMs forced with MODIS PFT maps, meaning the PFT area and locations are both “corrected”; 3) DGVM_{RS} – the constrained DGVMs simulations, meaning only the PFT area was “corrected” but location remained biased.



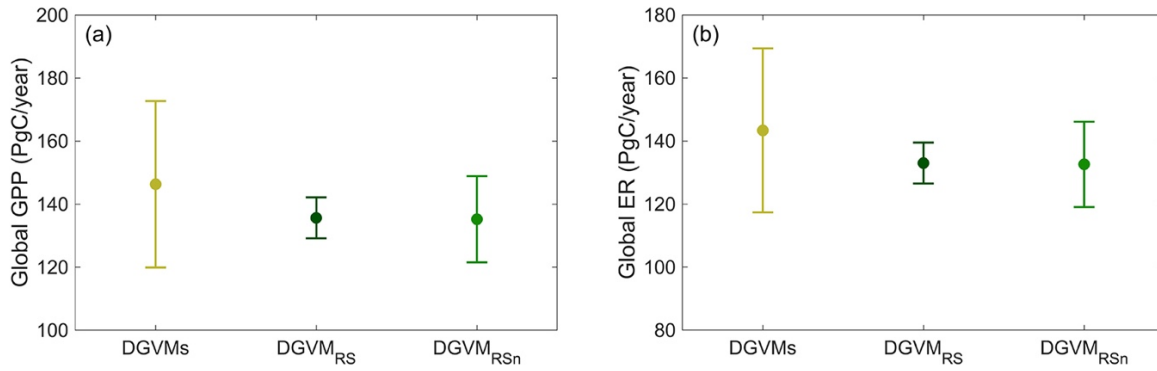
Supplementary Figure 8. Comparison of the relationships between carbon cycle components and PFT characteristics. (a) The coefficient of determination (R^2) for the relationships between PFT total GPP and PFT area (GPP & Area), or leaf area index (GPP & LAI) or leaf photosynthesis rate (GPP & Aleaf). (b) The coefficient of determination (R^2) for the relationships between different carbon fluxes or stocks and PFT area. The relationships include those between PFT-specific area and gross primary productivity (GPP & Area), net primary productivity (NPP & Area), autotrophic respiration (Ra & Area) and vegetation carbon pool (cVeg & Area). The R^2 is the coefficient of determination for the relationships, with higher values indicate a stronger correlation. The abbreviations for PFTs correspond to those in Supplementary Figure 1.



Supplementary Figure 9. The relationships between total GPP and total area for each PFT. Here we presented the results using S3 scenario simulations from TRENDY v9. The panels represent the results for (a) evergreen broadleaf forests (EBF), (b) temperate evergreen needleleaf forests (ENFt), (c) boreal evergreen needleleaf forests (ENFb), (d) deciduous broadleaf forests (DBF), (e) deciduous needleleaf forests (DNF), (f) shrubs (SHR), (g) warm C₃ grasses (C3Gw), (h) cool C₃ grasses (C3Gk), (i) cold C₃ grasses (C3Gc), (j) C₃ crops (C3C), (k) C₄ grasses (C4G), and (l) C₄ crops (C4C).



Supplementary Figure 10. The relationships between total gross primary productivity (GPP) and total area for each PFT. Here we present the result similar to that in Supplementary Figure 1, except that the linear regressions were not forced through zero. The panels represent the results for **(a)** evergreen broadleaf forests (EBF), **(b)** temperate evergreen needleleaf forests (ENFt), **(c)** boreal evergreen needleleaf forests (ENFb), **(d)** deciduous broadleaf forests (DBF), **(e)** deciduous needleleaf forests (DNF), **(f)** shrubs (SHR), **(g)** warm C_3 grasses (C3Gw), **(h)** cool C_3 grasses (C3Gk), **(i)** cold C_3 grasses (C3Gc), **(j)** C_3 crops (C3C), **(k)** C_4 grasses (C4G), and **(l)** C_4 crops (C4C).



Supplementary Figure 11. The constrained global gross primary productivity (GPP) and ecosystem respiration (ER). Here we present the result similar to that in Supplementary Figure 2. Shown are the **(a)** GPP and **(b)** ER spread across models (DGVMs), constrained GPP and ER using regressions forced through zero (DGVM_{RS}), and constrained GPP and ER using regressions not forced through zero (DGVM_{RSn}).

References

1. Prentice, I. C. et al. Special paper: a global biome model based on plant physiology and dominance, soil properties and climate. *J. Biogeogr.* 19, 117–134 (1992).
2. Sulla-Menashe, D. & Friedl, M. A. User guide to collection 6 MODIS land cover (MCD12Q1 and MCD12C1) product. *USGS: Reston, VA, USA*, 1–18 (2018).
3. Harper, K. L. et al. A 29-year time series of annual 300 m resolution plant-functional-type maps for climate models. *Earth Syst. Sci. Data* 15, 1465–1499 (2023).
4. Buchhorn, M. et al. Copernicus global land service: land cover 100m: version 3 globe 2015-2019: algorithm theoretical basis document. *Zenodo* <https://zenodo.org/records/4723924> (2020).
5. Zhang, X. et al. GLC_FCS30D: the first global 30 m land-cover dynamics monitoring product with a fine classification system for the period from 1985 to 2022 generated using dense-time-series Landsat imagery and the continuous change-detection method. *Earth Syst. Sci. Data* 16, 1353–1381 (2024).
6. Liu, H. et al. Annual dynamics of global land cover and its long-term changes from 1982 to 2015. *Earth Syst. Sci. Data* 12, 1217–1243 (2020).