Supplement of Biogeosciences, 21, 4285–4300, 2024
https://doi.org/10.5194/bg-21-4285-2024-supplement

Supplement of

A 2001–2022 global gross primary productivity dataset using an ensemble model based on the random forest method

Xin Chen et al.

Correspondence to: Tiexi Chen (txchen@nuist.edu.cn)

Table S1. Overview of flux sites from ChinaFlux used in this study

ID	Latitude	Longitude	Vegetation type	Study period	References
1	23.1737	112.5344	EBF	2003-2010	Li et al., 2021
2	30.85	91.0833	GRA	2004-2010	Chai et al., 2021
3	31.8068	119.2173	CRO	2015-2018	Zhou et al., 2023
4	32.8	102.55	GRA	2015-2018	Chen et al., 2023
5	33.4997	111.9353	DBF	2017-2018	Niu et al., 2023
6	35.2531	100.6992	GRA	2012-2016	He et al., 2023
7	37.6094	101.3119	GRA	2004-2009	Zhang et al., 2021
8	41.1481	121.2017	CRO	2005-2014	Zhang et al., 2023
9	41.644	110.3315	GRA	2015-2018	Song et al., 2023
10	42.4025	128.0958	DBF	2003-2010	Wu et al., 2021
11	43.3255	116.4032	GRA	2003-2010	Hao et al., 2020
12	45.4167	127.6678	DBF	2016-2018	Wang et al., 2021

Table S2. Parameters for Revise-EC-LUE

	DBF	ENF	EBF	MF	GRA	CRO-C3	CRO-C4	SAV	SHR	WET
ε_{sun}	1.52	2.50	1.96	1.82	1.79	1.77	2.90	1.74	1.27	1.30
ε_{sha}	3.64	4.52	3.33	3.31	3.72	3.82	3.47	4.00	3.61	3.27
θ	29.74	32.71	32.13	32.72	31.15	30.43	19.69	28.87	31.37	32.08
VPD0	16.29	8.86	10.36	12.14	10.52	14.20	15.75	14.21	8.29	15.78

Table S3. Parameters for EC-LUE

	DBF	ENF	EBF	MF	GRA	CRO-C3	CRO-C4	SAV	SHR	WET
ε	2.40	2.84	2.71	2.31	2.43	2.30	3.08	2.30	1.80	1.82

θ	33.08	34.23	29.52	31.49	28.96	32.84	30.29	31.80	30.04	30.19
VPD0	15.89	8.28	8.99	12.56	10.55	12.76	16.21	12.42	9.40	16.79

Table S4. Parameters for kNDVI-GPP

	Formula
DBF	GPP=kNDVI×325.68-0.21
ENF	GPP=kNDVI×625.01+0.24
EBF	GPP=kNDVI×190.38+4.17
MF	GPP=kNDVI×341.14+0.52
GRA	GPP=kNDVI×302.9+0.14
CRO-C3	GPP=kNDVI×334.42-0.71
CRO-C4	GPP=kNDVI×584.7-0.93
SAV	GPP=kNDVI×451.23+0.03
SHR	GPP=kNDVI×422.21+0.14
WET	GPP=kNDVI×281.09+0.38

Table S5. Parameters for NIRv-GPP

	Formula
DBF	GPP=39.31× NIRv-1.83
ENF	GPP=47.08×NIRv-0.77
EBF	GPP=25.56× NIRv+2.17
MF	GPP=37.65× NIRv-1.19
GRA	GPP=33.53×NIRv-0.54
CRO-C3	GPP=36.94× NIRV -1.55
CRO-C4	GPP=59.61× NIRv -1.29
SAV	GPP=44.92× NIRv -1.13
SHR	GPP=24.28× NIRv -0.06

WET	GPP=28.01× NIRv -0.06

Table S6. Parameters for VPM

	DBF	ENF	EBF	MF	GRA	CRO-C3	CRO-C4	SAV	SHR	WET
ε	1.92	3.28	2.82	2.38	2.06	2.03	2.73	2.15	1.92	1.78
Tmax	40.10	40.43	42.46	40.48	41.63	42.18	41.15	41.88	41.34	40.67
Tmin	-1.76	-3.85	-2.16	-3.45	-3.28	-2.92	-0.85	-3.32	-1.64	-3.18
Topt	17.19	15.99	15.97	16.26	15.32	16.40	23.74	21.73	16.51	19.54

Table S7. Parameters for MODIS

	DBF	ENF	EBF	MF	GRA	CRO-C3	CRO-C4	SAV	SHR	WET
ε	1.54	1.35	1.46	1.44	1.39	1.50	1.98	1.52	0.80	1.15
Tmax	10.45	7.41	8.10	9.33	10.54	10.84	10.92	13.56	7.37	11.79
Tmin	-10.39	-9.84	-10.26	-12.03	-11.22	-12.32	-13.21	-6.82	-8.34	-9.44
VPDmax	17.81	45.54	29.98	24.67	52.08	42.70	43.14	30.89	46.10	52.95
VPDmin	7.93	4.36	3.91	5.57	4.74	5.06	7.89	5.73	3.82	6.83

Table S8. Effect of the number of GPP estimate models in the ERF model on model performance

Number of GPP models	2	3	4	5
R^2	0.793±0.024	0.824±0.011	0.836±0.004	0.845±0.001
RMSE	1.798±0.104	1.658±0.052	1.600±0.022	1.556±0.009
Sim/Obs	1±0.001	0.999±0.000	1±0.000	1±0.000

Text S1. Detailed description of six GPP estimate models

1.1 EC-LUE

EC-LUE is a type of light use efficiency model developed by Yuan et al. (2007). The original model was driven by the normalized vegetation index (NDVI), photosynthetically active radiation (PAR), air temperature and the Bowen ratio of sensible latent heat flux. The Bowen ratio of sensible latent heat flux was replaced by VPD to characterize the constraints of atmospheric drought later, and the effect of CO_2 on GPP was integrated at the same time (Yuan et al., 2019). The basic form is as follows

$$GPP = PAR \times FPAR \times \varepsilon \times Cs \times min\,(Ts, Ws) \qquad (1)$$

Where FPAR is the ratio of canopy absorption PAR, ε is the maximum light use efficiency. Cs, Ts and Ws represent the influences of CO_2, temperature and VPD, respectively:

$$\mathrm{Cs} = \frac{Ci-\theta}{Ci+2\theta} \qquad (2)$$

$$\mathrm{Ts} = \frac{(Ta-0)\times(Ta-40)}{(Ta-0)\times(Ta-40)-(Ta-20.33)\times(Ta-20.33)} \qquad (3)$$

$$\mathrm{Ws} = \frac{VPD0}{VPD+VPD0} \qquad (4)$$

In these formulas, Ci and θ respectively represent the leaf internal CO_2 concentration and the CO_2 compensation point during no-dark respiration. Ta is the air temperature, and VPD0 is the empirical coefficient. In this model, θ, ε and VPD0 need to be calibrated using GPP observations of flux towers.

1.2 Revised-EC-LUE

Different from the commonly used light use efficiency model, Revise-EC-LUE divides the canopy into sunlit and shaded leaves, and its effectiveness has been proved in several flux sites. The basic form is as follows:

$$GPP = (\varepsilon_{sun} \times APAR_{sun} + \varepsilon_{sha} \times APAR_{sha}) \times Cs \times min\,(Ts, Ws) \qquad (5)$$

Where ε_{sun} and ε_{sha} represent the maximum light use efficiency of sunlit and shaded leaves respectively, $APAR_{sun}$ and $APAR_{sha}$ represent the PAR absorbed by sunlit and shaded leaves respectively. All the processes can refer to the article of Zheng et al. (2020), Cs, Ts and Ws represent the influences of CO_2, temperature and VPD respectively, and their basic forms are the same as formula (2) - (4). Therefore, the parameters to be calibrated in this model include ε_{sun}, ε_{sha}, θ, and VPD0.

1.3 NIRv-GPP

Near-infrared Vegetation Index (NIRv) proposed by Badgley et al. (2017), is a new vegetation index that approximates the proportion of near-infrared light reflected by vegetation, it has been shown to be directly related to solar-induced fluorescence (SIF) and can be used to estimate GPP. The basic form is as follows

$$GPP = a \times NIRv + b \qquad (6)$$

Where a and b represent the slope and intercept, respectively. NIRv can be calculated using satellite-based red band and near infrared band.

$$\mathrm{NIRv} = \mathrm{NDVI} \times \mathrm{NIR} = \frac{\mathrm{NIR}-\mathrm{R}}{\mathrm{NIR}+\mathrm{R}} \times \mathrm{NIR} \qquad (7)$$

Where NIR and R represent red band and near infrared band respectively.

1.4 kNDVI-GPP

Similar to NIRv, kNDVI is a newly proposed vegetation index (Camps-Valls et al., 2021). In comparison with GPP and SIF, kNDVI always shows stronger correlation than NIRv and NDVI, and has a unique effect in dealing with the saturation of the vegetation index. The form is as follows

$$GPP = a \times kNDVI + b \qquad (8)$$

Where a and b represent the slope and intercept, respectively. kNDVI can be calculated using satellite-based red band and near infrared band.

$$\mathrm{kNDVI} = \tanh\left(\frac{\mathrm{NIR}-\mathrm{R}}{\mathrm{NIR}+\mathrm{R}}\right) \qquad (9)$$

Where NIR and R represent red band and near infrared band respectively.

1.5 MODIS

MODIS-GPP is one of the light use efficiency models (Running et al., 2004), driven by FPAR, temperature and VPD, and its basic form is as follows

$$\mathrm{GPP} = \mathrm{FPAR} \times \varepsilon \times \mathrm{T_s} \times \mathrm{W_s} \qquad (10)$$

Where ε is the maximum light use efficiency. Ts and Ts represent the influences of minimum temperature and VPD, respectively:

$$\mathrm{Ts} = \frac{TMIN - TMIN_{min}}{TMIN_{max} - TMIN_{min}} \qquad (11)$$

$$\mathrm{Ws} = \frac{VPD_{max} - VPD}{VPD_{max} - VPD_{min}} \qquad (12)$$

TMINmax and TMINmin are the daily minimum temperature for $\varepsilon = \varepsilon_{max}$ and $\varepsilon = 0$, respectively, and VPDmax and VPDmin are the VPD for $\varepsilon = 0$ and $\varepsilon = \varepsilon_{max}$, respectively. In this model, ε, TMINmin, TMINmax, VPDmin and VPDmax need to be calibrated using GPP observations from flux towers.

1.6 VPM

The satellite-based VPM (Xiao et al, 2004) uses the product of light use efficiency (LUE, ε), temperature constraint (Ts), water constraint (Ws) and absorbed photosynthetically active radiation by chlorophyll (APARchl) to estimate GPP as follows:

$$\mathrm{GPP} = \mathrm{APARchl} \times \varepsilon \times \mathrm{Ts} \times \mathrm{Ws} \qquad (10)$$

$$\mathrm{APARchl} = (\mathrm{EVI} - 0.1) \times 1.25 \qquad (11)$$

$$\mathrm{Ts} = \frac{(Ta - Tmax) \times (Ta - Tmin)}{(Ta - Tmax) \times (Ta - Tmin) - (Ta - Topt)^2} \qquad (12)$$

$$\mathrm{Ws} = \frac{1 + LSWI}{1 + LSWImax} \qquad (13)$$

Where the T, Tmax, Tmin and Topt refer to the daytime mean temperature, maximum, minimum, and optimum temperature for photosynthesis, respectively.

EVI is the enhanced vegetation index. In this model, ε, Tmax, Tmin and Topt need to be calibrated using GPP observations from flux towers.

The LSWI (Land Surface Water Index) is a good indicator of water stress from the vegetation canopy and soil background. This index is calculated as follows:

$$\mathrm{LSWI} = \frac{NIR-SWIR}{NIR+SWIR} \qquad (14)$$

Where NIR and SWIR represent near infrared band and shortwave infrared band respectively.

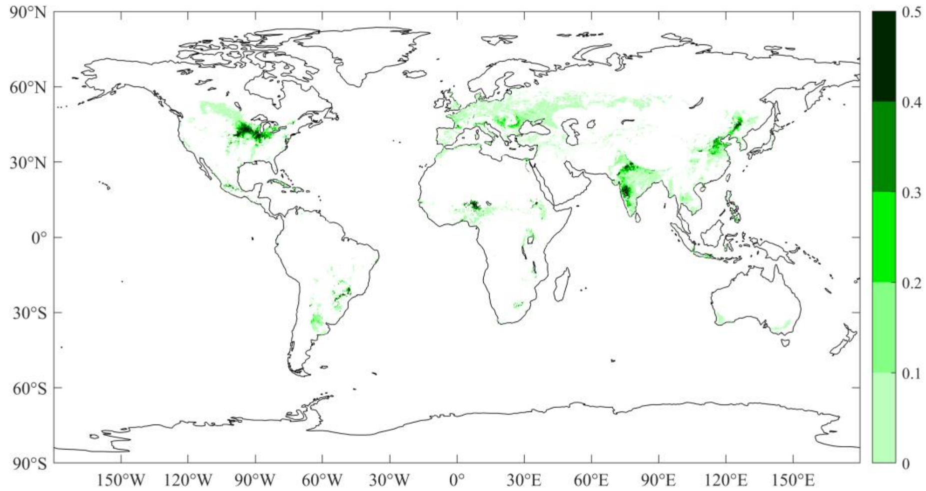

Figure S1. Distribution map of C4 crops.

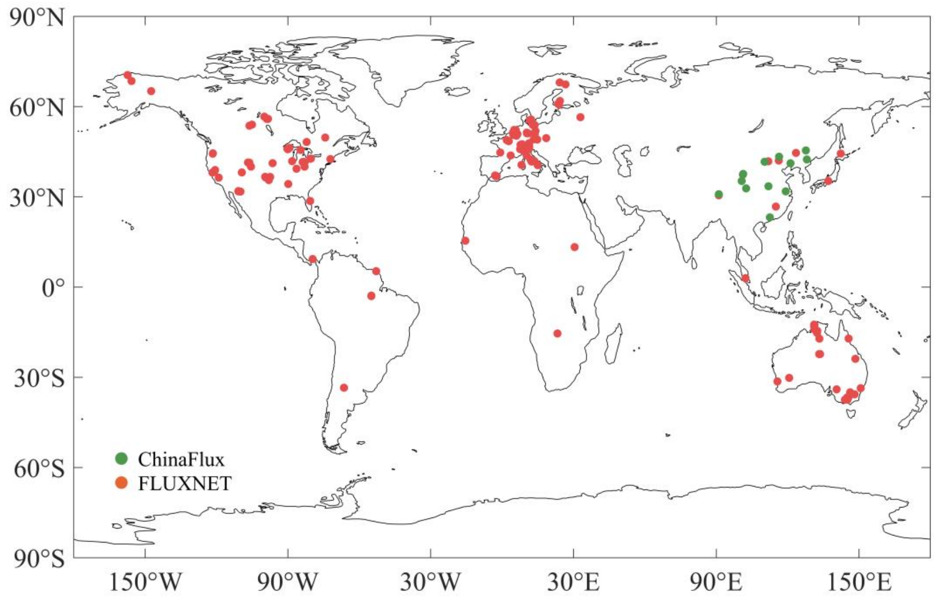

Figure S2. Flux site locations of ChinaFlux and FLUXNET used in this study.

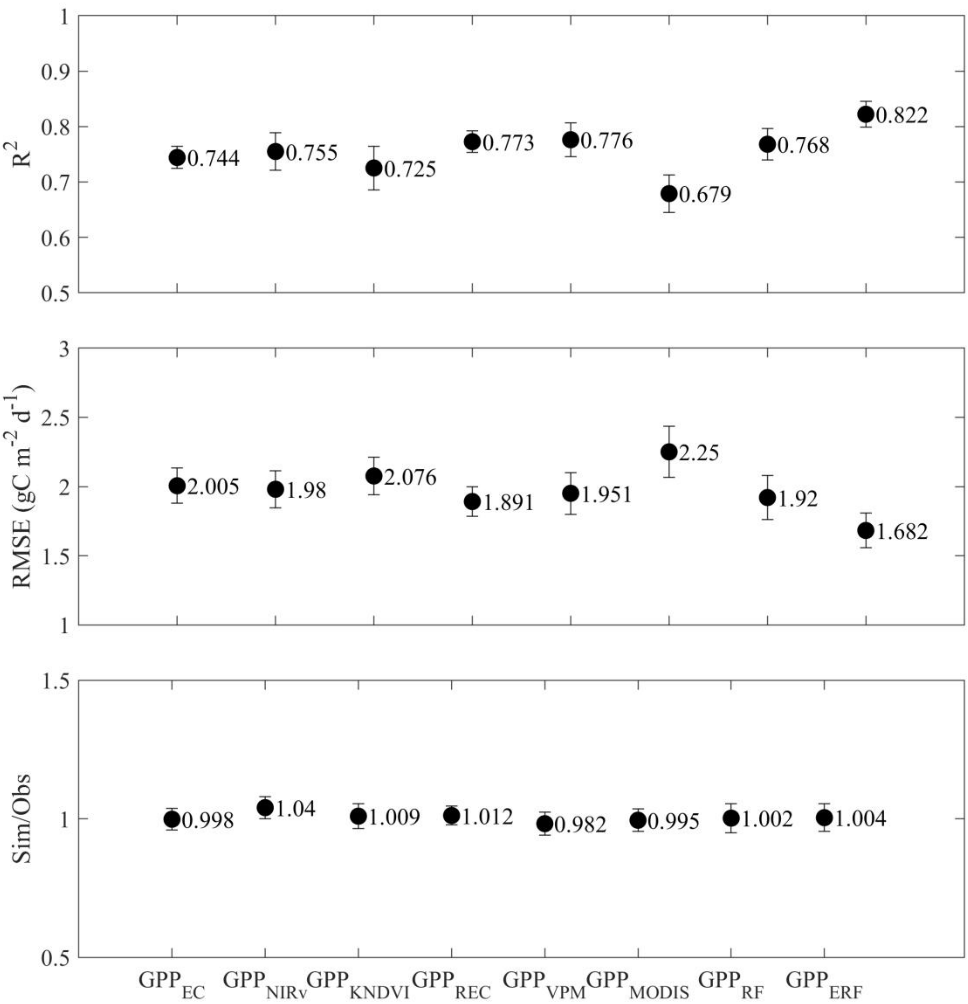

Figure S3. Validation results for each model on all data at 30% of the sites. The black dots represent the mean of the 200 validation results, and the upper and lower boundaries represent the standard deviation.

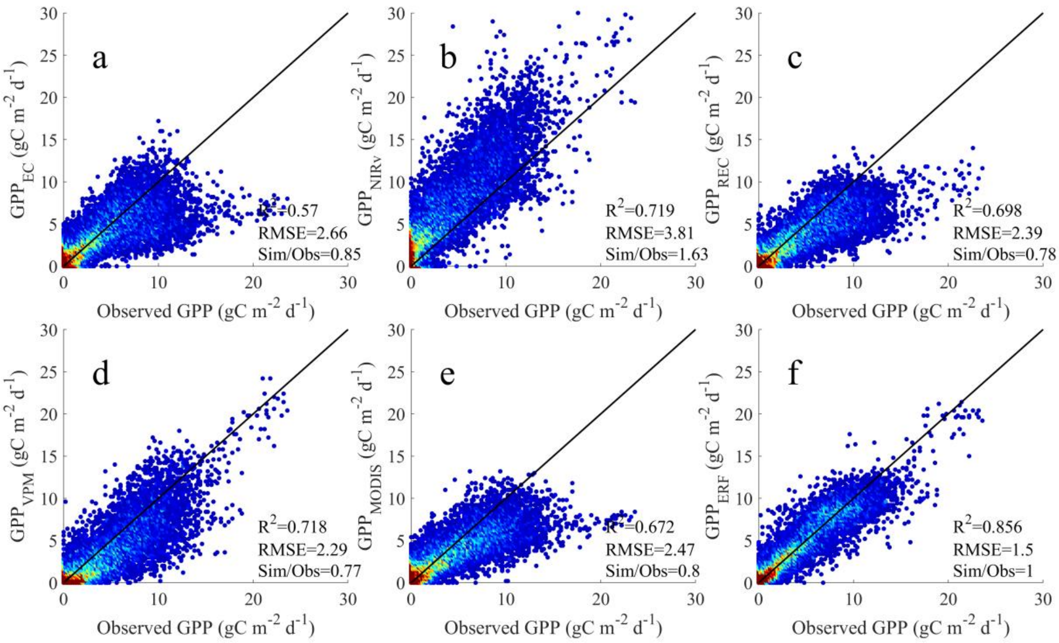

Figure S4. Comparison between the GPP simulations of the six models with original parameters and the GPP observations. a-f represents GPP_{EC}, GPP_{NIRv}, GPP_{REC}, GPP_{VPM}, GPP_{MODIS}, GPP_{ERF}, respectively. The author of kNDVI did not provide model parameters, so this model was abandoned.

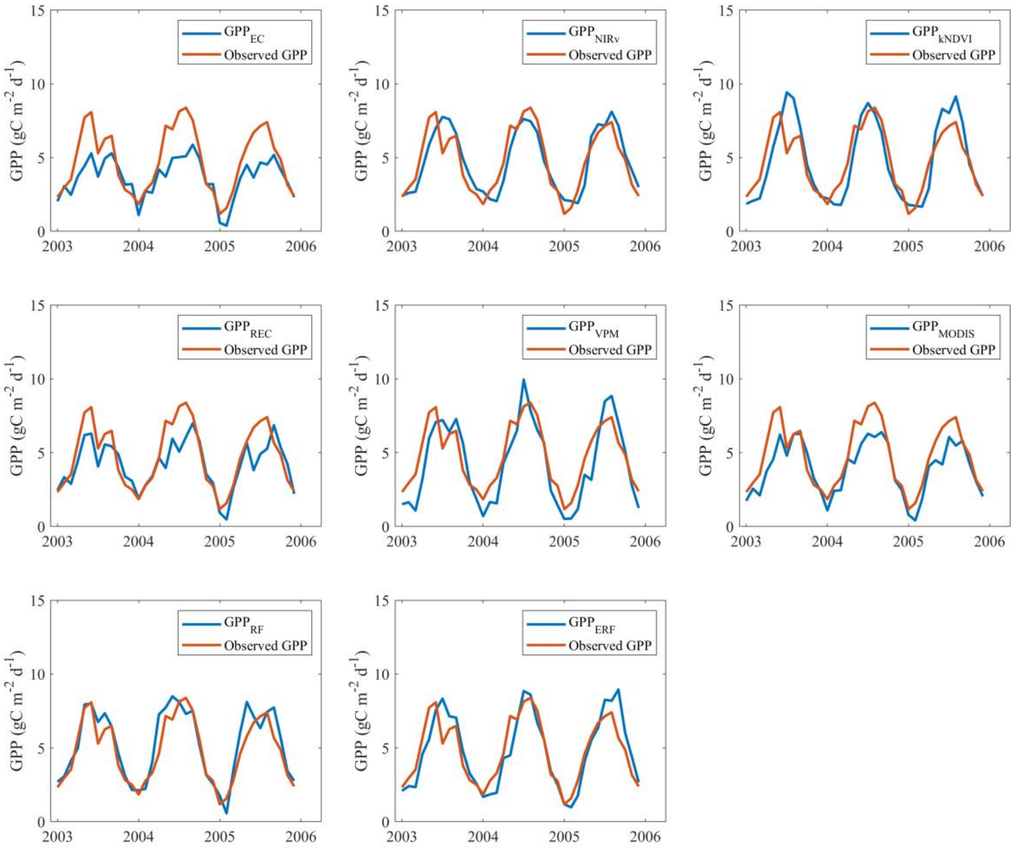

Figure S5. Performance of each GPP estimate model on CN-Qia.

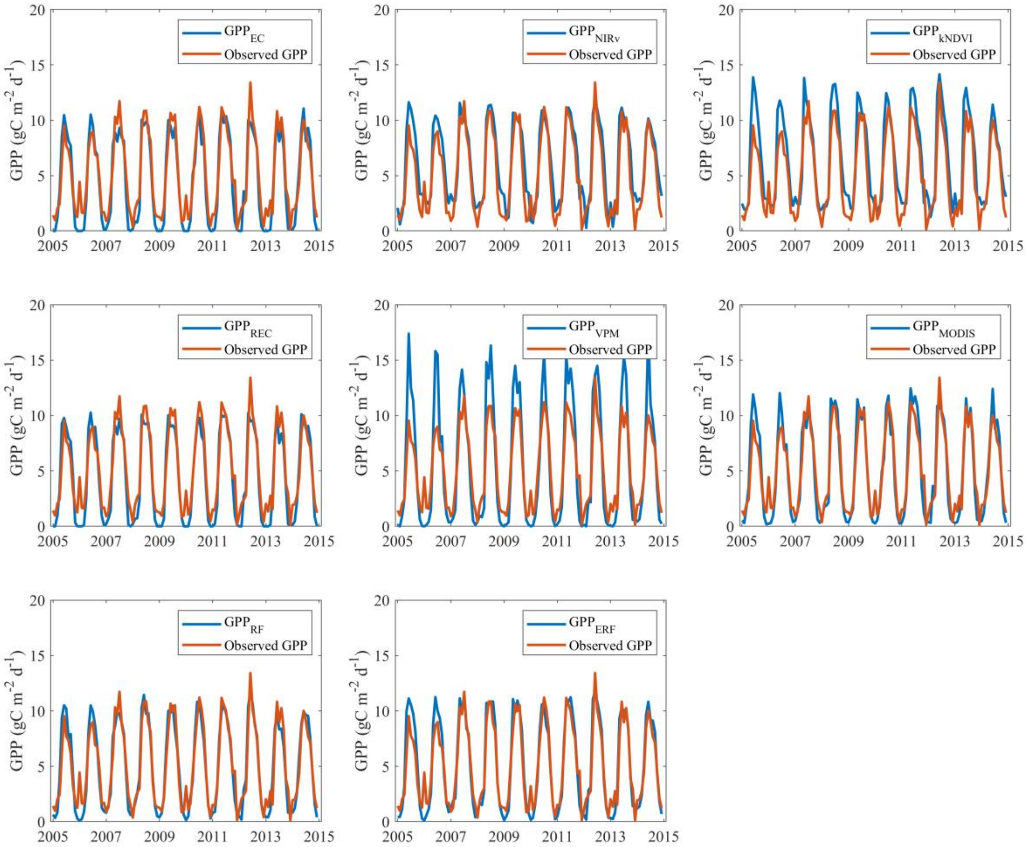

Figure S6. The performance of each GPP estimate model on CH_Lae.

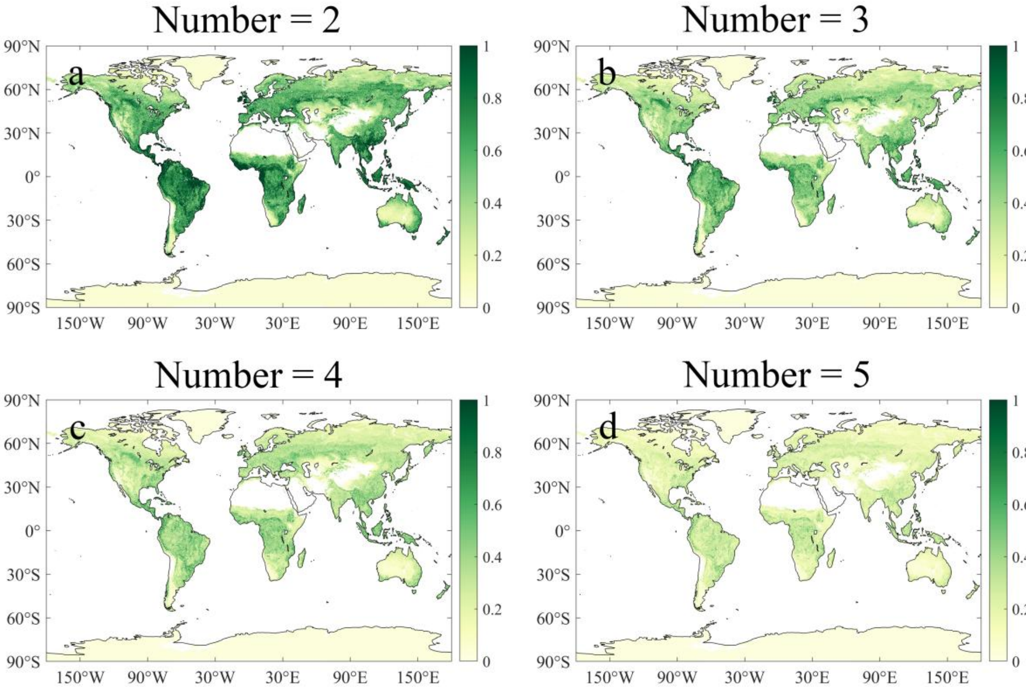

Figure S7. Uncertainty of ERF_GPP caused by the number of features (simulated GPP).

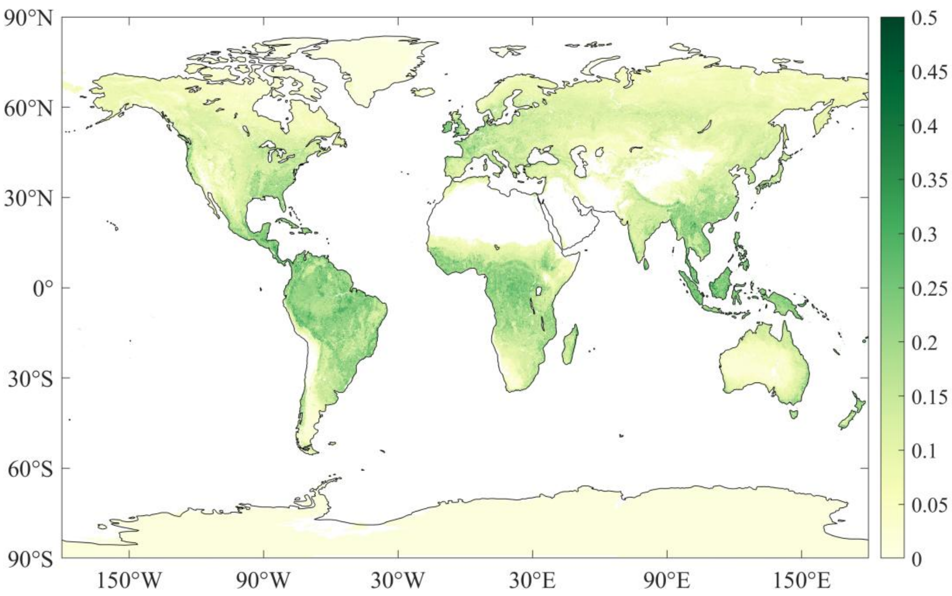

Figure S8. Uncertainty of ERF_GPP caused by the number of GPP observations.

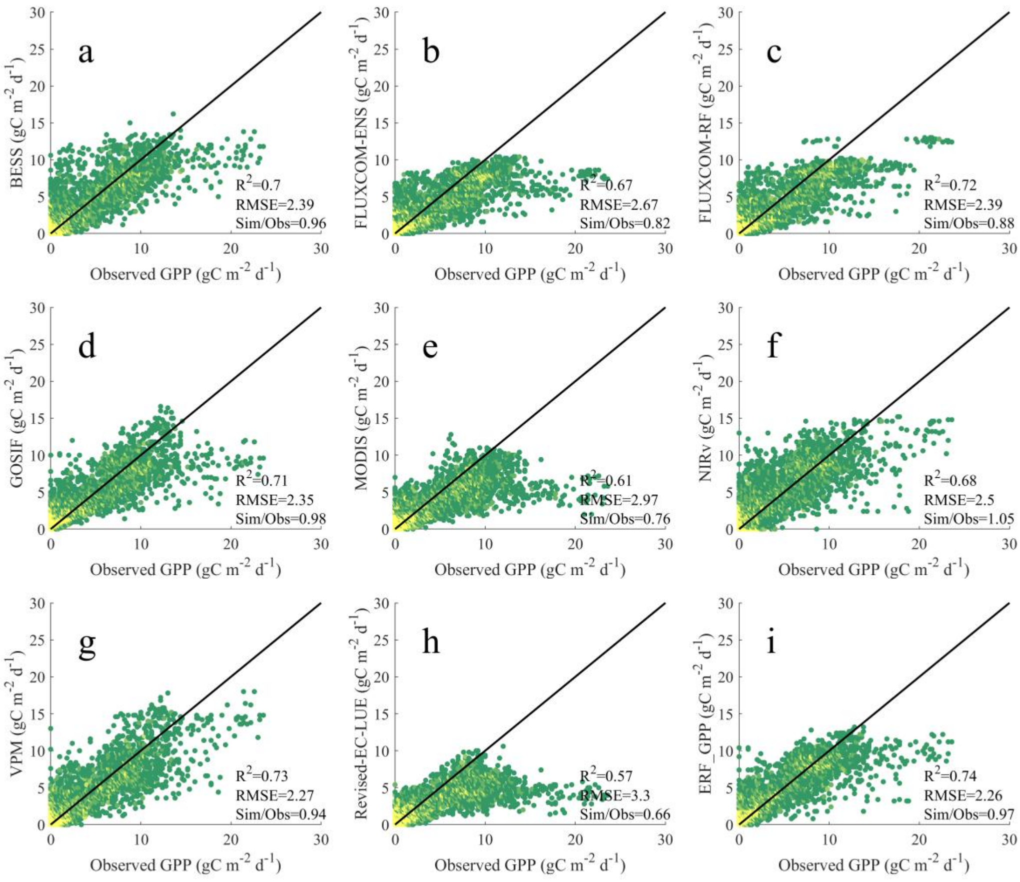

Figure S9. Comparison between the GPP datasets and the GPP observations from FLUXNET. a-i represents BESS, FLUXCOM-ENS, FLUXCOM-RF, GOSIF, MODIS, NIRv, VPM, Revise-EC-LUE, ERF_GPP, respectively.

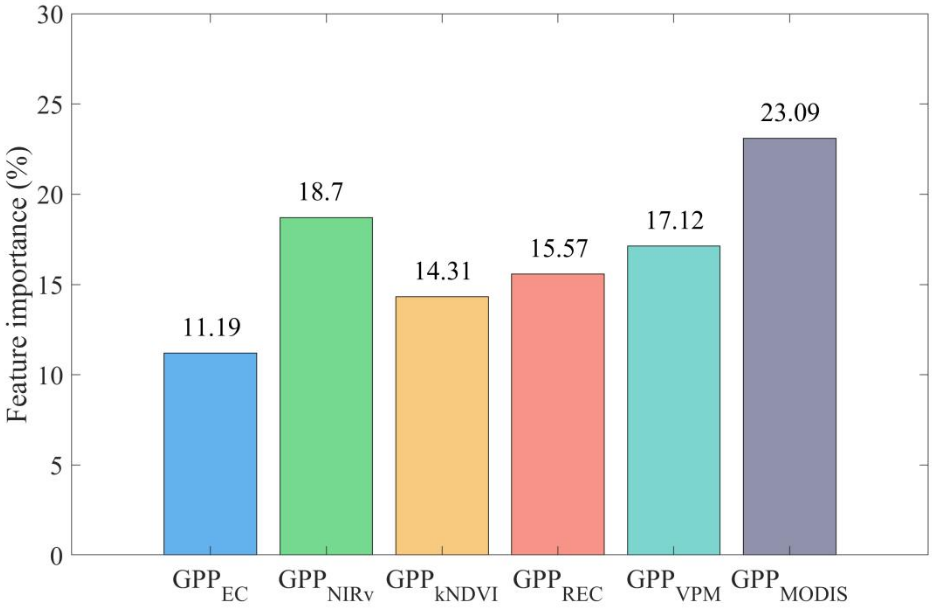

Figure S10. Average of 200 feature importance in the ERF model.

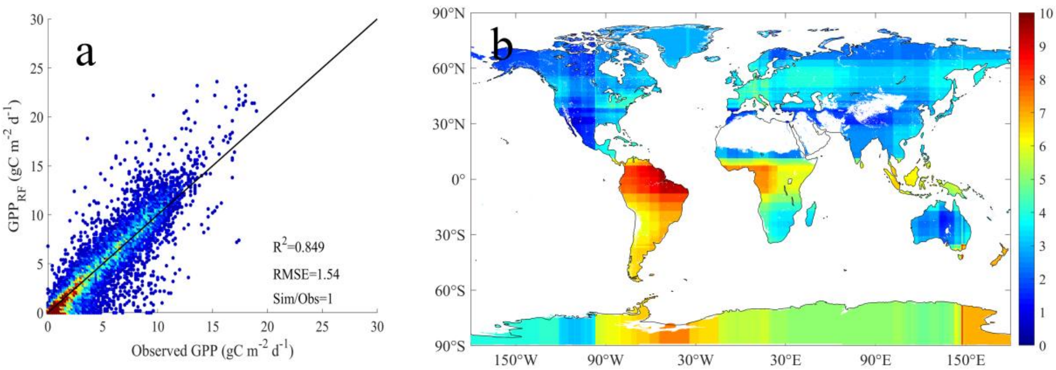

Figure S11. Simulation performance of a random forest model with only longitude, latitude, year, and month. a represents the result of the 5-fold-cross-validation, b is the multi-year mean estimated by the model for 2001-2022.

Reference

Badgley, G., Field, C. B., and Berry, J. A.: Canopy near-infrared reflectance and terrestrial photosynthesis, Science advances, 3, e1602244, https://doi.org/10.1126/sciadv.1602244, 2017.

Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, Á., Walther, S., Duveiller, G., Cescatti, A., Mahecha, M. D., Muñoz-Marí, J., García-Haro, F. J., Guanter, L., Jung, M., Gamon, J. A., Reichstein, M., and Running, S. W.: A unified vegetation index for quantifying the terrestrial biosphere, Science Advances, 7, eabc7447, https://doi.org/10.1126/sciadv.abc7447, 2021.

Chai X., He Y., Shi P., Zhang X., Niu B., Zhang L., and Chen Z. An observation dataset of carbon and water fluxes over alpine meadow in Damxung (2004 – 2010). China Scientific Data, 6(1), https://doi.org/10.11922/csdata.2020.0026.zh, 2021.

Chen, W., Wang, S., and Niu, S. A dataset of carbon, water and heat fluxes of Zoige alpine meadow from 2015 to 2020. China Scientific Data, https://doi.org/10.11922/11-6035.csd.2023.0009.zh, 2023.

Hao Y., Zhang L., Sun X., Yu G., Chen Z., and Wang Y. A dataset of carbon and water fluxes over Xilinhot temperate steppe in Inner Mongolia (2003 – 2010). China Scientific Data, https://doi.org/10.11922/10.11922/csdata.2020.0040.zh, 2020.

He F., Li Q., Chen C., and Zhao L. A dataset of carbon, water and heat fluxes over an Elymus nutans artificial grassland in the Sanjiangyuan Area (2012–2016). China Scientific Data, https://doi.org/10.57760/sciencedb.o00119.00025, 2023.

Li, Y., Yan, J., Meng, Z., Huang, J., Zhang, L., Chen, Z., Liu S., Chu G., Zhang Qi., and Zhang, D. An observation dataset of carbon and water fluxes in a mixed coniferous broad-leaved forest at Dinghushan, Southern China (2003–2010). China Scientific Data, 6(1), https://doi.org/10.11922/csdata.2020.0046.zh, 2021.

Niu, X., Sun, P., Tao, S., Chen, Z., Niu, B., and Liu, S. A dataset of carbon and water fluxes in a natural oak forest of Baotianman in Henan Province (2017-2018). China Scientific Data, https://doi.org/10.11922/11-6035.csd.2023.0124.zh, 2023.

Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., and Hashimoto, H.: A continuous satellite-derived measure of global terrestrial primary production, Bioscience, 54, 547-560, https://doi.org/10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2, 2004.

Song J., Zhou L., Zhou G., Yan Y., and Zhang S. A dataset of carbon and water fluxes of the temperate desert steppe in Damao Banner, Inner Mongolia (2015–2018). China Scientific Data, https://doi.org/10.11922/11-6035.csd.2023.0021.zh, 2023.

Wang X., Hu K., Liu F., Zhu Y., Zhang Q., and Wang C. A dataset of observed carbon fluxes on deciduous broad-leaved forest at the Maoershan Station from 2016 to 2018. China Scientific Data, 8(2), https://doi.org/10.11922/11-6035.csd.2023.0024.zh, 2023.

Wu J., Guan D., Wang A., Yuan F., Diao H., Yu G., Chen Z., and Zhang L. A dataset of carbon and water flux observation over broad-leaved red pine forest in Changbai Mountain (2003–2010). China Scientific Data, 6(1), https://doi.org/10.11922/csdata.2020.0041.zh, 2021.

Xiao, X., Zhang, Q., Braswell, B., Urbanski, S., Boles, S., Wofsy, S., Moore III, B., and Ojima, D.: Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data, Remote sensing of environment, 91, 256-270, https://doi.org/10.1016/j.rse.2004.03.010, 2004.

Yuan, W., Liu, S., Zhou, G., Zhou, G., Tieszen, L. L., Baldocchi, D., Bernhofer, C., Gholz, H., Goldstein, A. H., Goulden, M. L., Hollinger, D. Y., Hu, Y., Law, B. E., Stoy, P. C., Vesala, T., and Wofsy, S. C.: Deriving a light use efficiency model from eddy covariance flux data for predicting daily gross primary production across biomes, Agricultural and Forest Meteorology, 143, 189-207, https://doi.org/10.1016/j.agrformet.2006.12.001, 2007.

Yuan, W., Zheng, Y., Piao, S., Ciais, P., Lombardozzi, D., Wang, Y., Ryu, Y., Chen, G., Dong, W., Hu, Z., Jain, A. K., Jiang, C., Kato, E., Li, S., Lienert, S., Liu, S., Nabel, J. E. M. S., Qin, Z., Quine, T., Sitch, S., Smith, W. K., Wang, F., Wu, C., Xiao, Z., and Yang, S.: Increased atmospheric vapor pressure deficit reduces global vegetation growth, Science Advances, 5, eaax1396, https://doi.org/10.1016/10.1126/sciadv.aax1396, 2019.

Zhang S., Zhou L., Zhou G., Jia Q., Li R., Wang Y. A dataset of carbon and water flux observations in the agricultural ecosystem of spring maize in Jinzhou (2005–2014). China Scientific Data, https://doi.org/10.11922/11-6035.csd.2023.0007.zh, 2023.

Zhang, F., Li, H., Zhao, L., Zhang, L., Chen, Z., Zhu, J., Xu, S., Yang, Y., Zhao, X., Yu, G. and Li, Y. An observation dataset of carbon, water and heat fluxes of alpine wetland in Haibei (2004–2009). China Scientific Data, 6(1), https://doi.org/10.11922/csdata.2020.0033.zh, 2021.

Zheng, Y., Shen, R., Wang, Y., Li, X., Liu, S., Liang, S., Chen, J. M., Ju, W., Zhang, L., and Yuan, W.: Improved estimate of global gross primary production for reproducing its long-term variation, 1982–2017, Earth System Science Data, 12, 2725-2746, https://doi.org/10.5194/essd-12-2725-2020, 2020.

Zhou Y., Zhang Y., Zhu T., and Ju W. A dataset of carbon and water fluxes in the cropland ecosystem at Jurong Station (2015-2020). China Scientific Data, 8(3), https://doi.org/10.11922/11-6035.csd.2023.0072.zh, 2023.