Supplementary Information For "2001-2022 global gross primary productivity dataset using an ensemble model based on random forest"

Xin Chen[1], Tiexi Chen[1,2,3*], Xiaodong Li[4], Yuanfang Chai[5], Shengjie Zhou[1], Renjie Guo[6], Jie Dai[1]

[1]School of Geographical Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, Jiangsu, China.

[2]Qinghai Provincial Key Laboratory of Plateau Climate Change and Corresponding Ecological and Environmental Effects, Qinghai University of Science and Technology, Xining 810016, China

[3]School of Geographical Sciences, Qinghai Normal University, Xining 810008, Qinghai, China.

[4]Qinghai Institute of Meteorological Science, Xining 810008, Qinghai, China.

[5]Department of Earth Sciences, Vrije Universiteit Amsterdam, Boelelaan 1085, 1081 HV, Amsterdam, the Netherlands

[6]Faculty of Geographical Science, Beijing Normal University, Beijing, China.

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	1
2	30.85	91.0833	GRA	2004-2010	2
3	31.8068	119.2173	CRO	2015-2018	3
4	32.8	102.55	GRA	2015-2018	4
5	33.4997	111.9353	DBF	2017-2018	5
6	35.2531	100.6992	GRA	2012-2016	6
7	37.6094	101.3119	GRA	2015-2018	7
8	41.1481	121.2017	CRO	2005-2014	8
9	41.644	110.3315	GRA	2015-2018	9
10	42.4025	128.0958	DBF	2003-2010	10
11	43.3255	116.4032	GRA	2003-2010	11
12	45.4167	127.6678	DBF	2016-2018	12

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 GPP-kNDVI

	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 GPP-NIRv

	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

Text S1. Detailed description of four remote sensing models

1.1 EC-LUE

EC-LUE is a type of light use efficiency model developed by Yuan et al. The original model was driven by 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. The basic form is as follows

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

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

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

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

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

In the formula, Ci and θ respectively represent the CO_2 concentration in the air between leaf cells 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 by GPP of flux tower.

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:

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

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., Cs, Ts and Ws represent the influence 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 GPP- NIRV

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

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

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

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

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

1.4 GPP-KNDVI

Similar to NIRV, KNDVI is also the newly proposed vegetation index. 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 vegetation index. The form is as follows

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

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

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

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

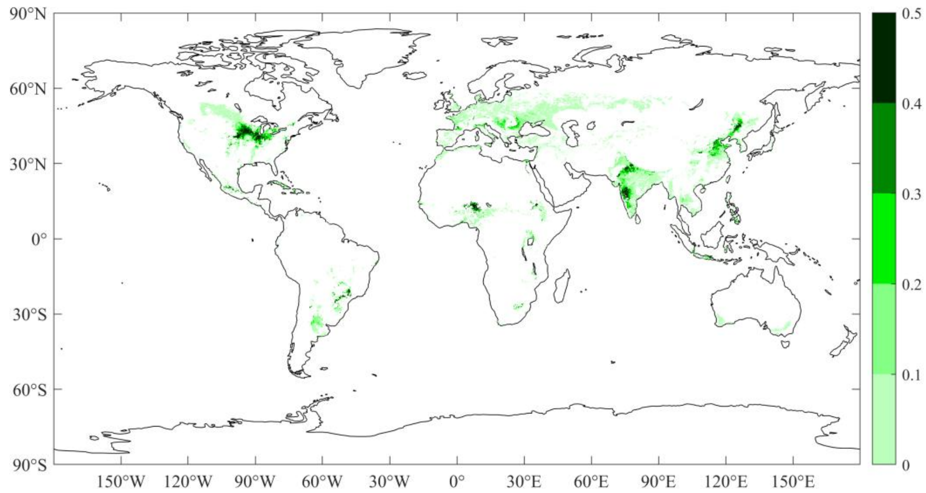

Figure S1. Distribution map of C4 crops.

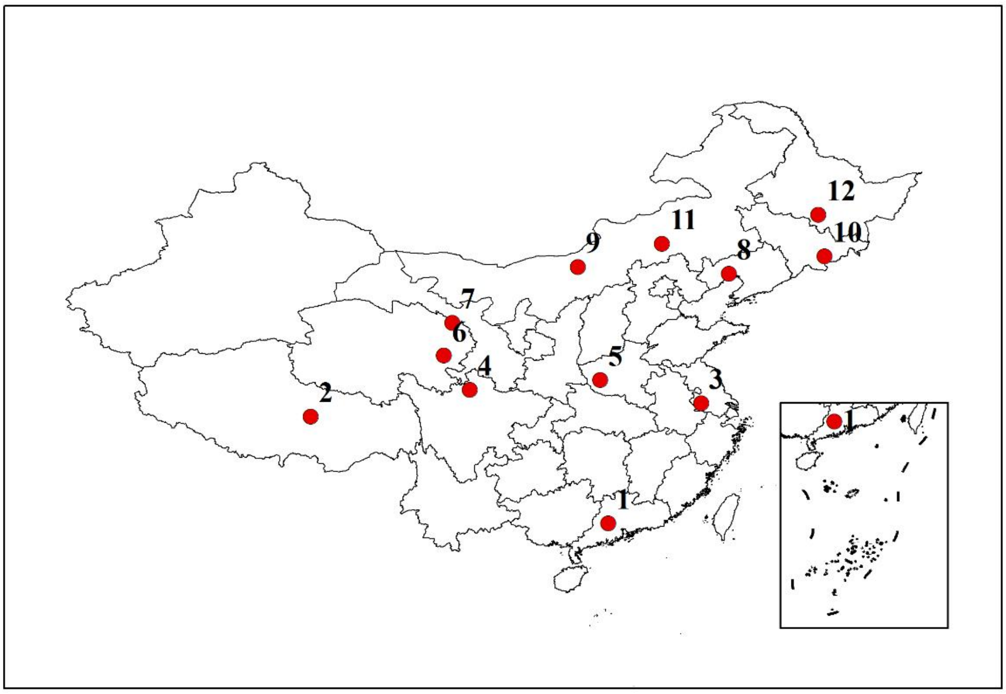

Figure S2. The flux site location of ChinaFlux used in this study.

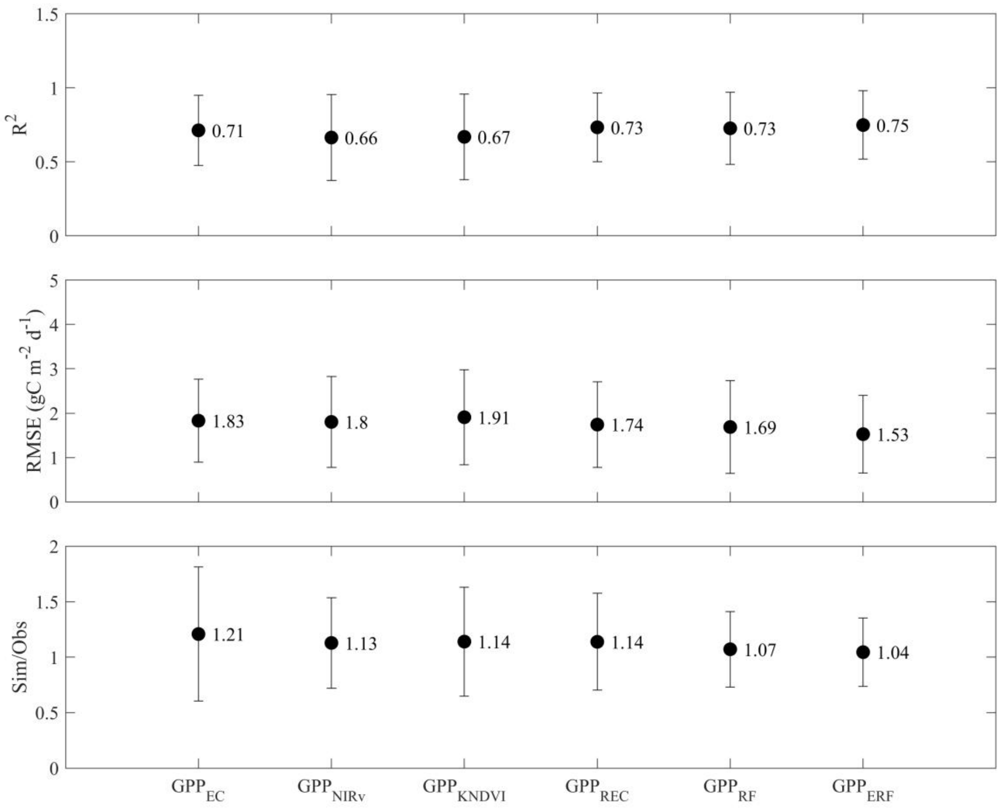

Figure S3. Average model accuracy of six models at all sites

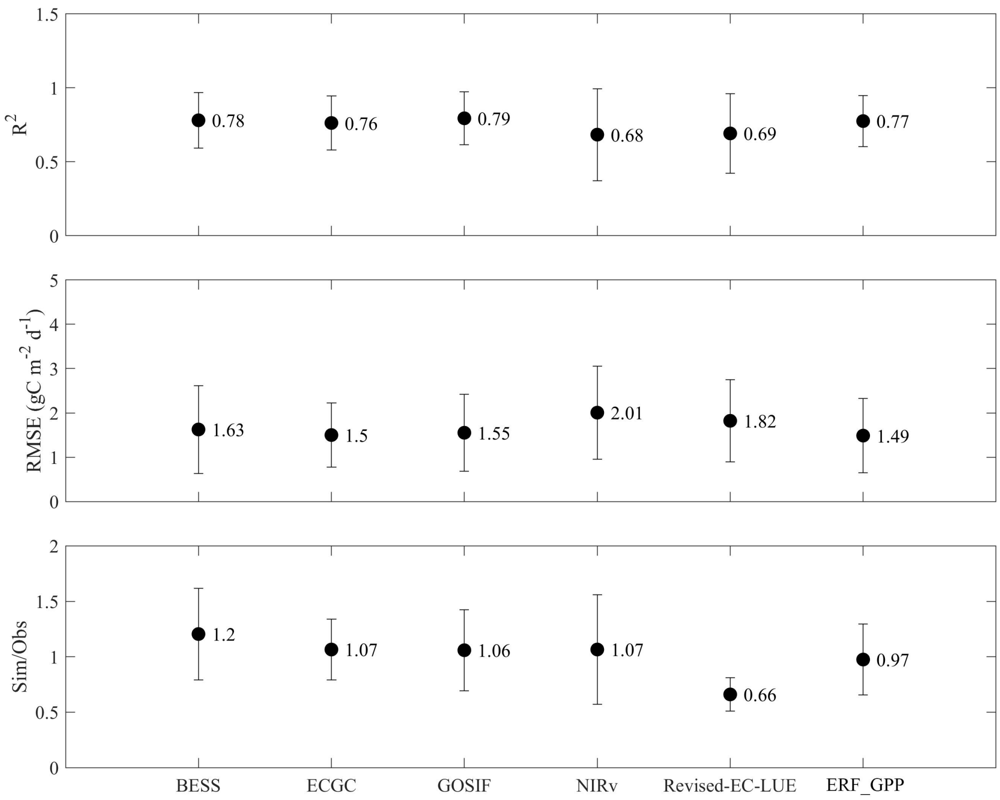

Figure S4. Average model accuracy of GPP datasets at all sites from ChinaFlux

1. An observation dataset of carbon and water fluxes in a mixed coniferous broad-leaved forest at Dinghushan, Southern China (2003 – 2010)
2. Shi, P., and He, Y. (2021). An observation dataset of carbon and water fluxes over alpine meadow in Damxung (2004 – 2010). Science Data Bank.
3. Yanlian, Z., Yongguang, Z., Tingting, Z.H.U., and Weimin, J.U. (2023). A dataset of carbon and water fluxes in the cropland ecosystem at Jurong Station (2015-2020). Science Data Bank.
4. Chen, W., Wang, S., and Niu, S. (2023). A dataset of carbon, water and heat fluxes of Zoige alpine meadow from 2015 to 2020. Science Data Bank.
5. Niu, X., Sun, P., Tao, S., Chen, Z., Niu, B., and Liu, S. (2023). A dataset of carbon and water fluxes in a natural oak forest of Baotianman in Henan Province (2017-2018). Science Data Bank.
6. Fuquan, H.E., Qi, L.I., Dongdong, C., and Liang, Z. (2023). A dataset of carbon, water and heat fluxes over an Elymus nutans artificial grassland in the Sanjiangyuan Area (2012–2016). Science Data Bank.
7. Zhang, F., Li, H., Zhao, L., Zhang, L., Chen, Z., Zhu, J., Xu, S., Yang, Y., Zhao, X., and Yu, G. (2020). An observation dataset of carbon, water and heat fluxes of alpine wetland in Haibei (2004–2009). Science Data Bank.
8. Sen, Z., Li, Z., Guangsheng, Z., Qingyu, J.I.A., Rongping, L.I., and Yu, W. (2023). A dataset

of carbon and water flux observations in the agricultural ecosystem of spring maize in Jinzhou (2005–2014). Science Data Bank.

9. Jiaxin, S., Li, Z., Guangsheng, Z., Yujie, Y.A.N., and Sen, Z. (2023). A dataset of carbon and water fluxes of the temperate desert steppe in Damao Banner, Inner Mongolia (2015–2018). Science Data Bank.

10. A dataset of carbon and water flux observation over broad-leaved red pine forest in Changbai Mountain (2003–2010). Science Data Bank.

11. A dataset of carbon and water fluxes over Xilinhot temperate steppe in Inner Mongolia (2003 – 2010). Science Data Bank.

12. Xingchang, W., Keming, H.U., Fan, L.I.U., Yuan, Z.H.U., Quanzhi, Z., and Chuankuan, W. (2023). A dataset of observed carbon fluxes on deciduous broad-leaved forest at the Maoershan Station from 2016 to 2018. Science Data Bank.