

The 19th China-US Carbon Consortium Annual Workshop

Estimating Global GPP Using a Machine Learning Approach

Tiexi Chen (Prof. Dr.)

School of Geographical Sciences, Nanjing University of Information Science and Technology

School of Ecology and Environmental Science, Xining University

txchen@nuist.edu.cn

2023.7.27 Nanjing

The focuses:

- Identify and quantify (contributions) the drivers of vegetation changes over regional and global scales
- Quantify the key fluxes of the terrestrial ecosystems at reginal and global scales

Guo, R., **Chen, T.***, et al., (2023). Estimating Global GPP from the Plant Functional Type Perspective Using a Machine Learning Approach. **JGR-Biogeosciences**

Zhou, S., **Chen, T.***, et al., (2022): The Impact of Cropland Abandonment of Post-Soviet Countries on the Terrestrial Carbon Cycle Based on Optimizing the Cropland Distribution Map, **Biology**

Chen, X., **Chen, T.***, Shu, Y.*, Yan, Q., Han, Q., Wei, X., et al. (2021). A framework to assess the potential uncertainties of three FPAR products. **JGR-Biogeosciences**

Chen, T.*, et al., (2014). Global cropland monthly gross primary production in the year 2000. **Biogeosciences**

Chen, T.*, et al., (2011). Evaluation of cropland maximum light use efficiency using eddy flux measurements in North America and Europe. **Geophysical Research Letters**

Chen, T.*, et al., (2022): Land management explains the contrasting greening pattern across China-Russia border based on Paired Land Use Experiment approach, **JGR-Biogeosciences**,

Chen, T.*, et al., (2022): Land Management Contributes significantly to observed Vegetation Browning in Syria during 2001–2018, **Biogeosciences**,

张林林,. . . ,**陈铁喜***(2022).土地管理对植被变绿的潜在贡献——以中国东北农业区为例[J].**生态学报**

Chen, X; **Chen, T***, et al., (2021). The Ongoing Greening in Southwest China despite Severe Droughts and Drying Trends. **Remote Sensing**,

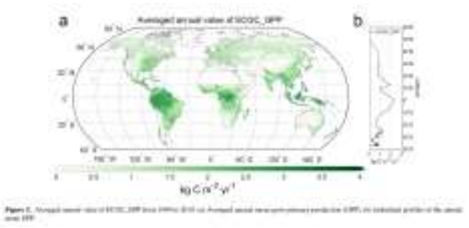
Chen, T.* et al., (2021). The Greening and Wetting of the Sahel Have Leveled off since about 1999 in Relation to SST. **Remote Sensing**

Chen, T., et al., (2016). Asymmetric NDVI trends of the two cropping seasons in the Huai River basin. **Remote Sensing Letters**

Chen, T.*, et al., (2014). Using satellite based soil moisture to quantify the water driven variability in NDVI: A case study over mainland Australia. **Remote Sensing of Environment**

Chen, T., et al., (2013) A global analysis of the impact of drought on net primary productivity. **Hydrology and Earth System Sciences**

Innovation 1: Establish a Paired Land Use Experiment (PLUE) method to analyze the impact of regional-scale land management on vegetation change (Chen, T.* 2022a,b)



Innovation 2: Developed the global multi-vegetation type farmland GPP (Chen, T.* 2014) and the global terrestrial ecosystem GPP dataset based on random forest (Guo, R., Chen, T.*, et al., (2023)).

the Paired Land Use Experiments (PLUE) theory in driver identification of regional vegetation change

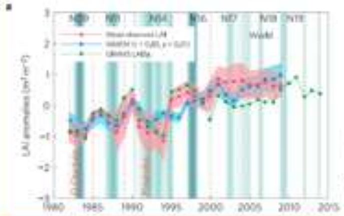
Tiexi CHEN and co-authors[ref1,2]

School of Geographical Sciences, Nanjing University of Information Science and Technology, Nanjing 210044, China.

txchen@nuist.edu.cn; web: https://faculty.nuist.edu.cn/chentiexi/zh_CN/

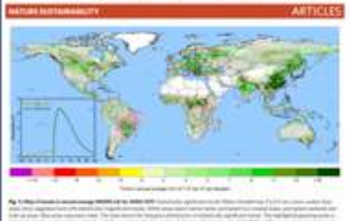
Abstract: How to identify the drivers of regional-scale vegetation change, especially to distinguish between climate change and human activities remains a great challenge. Modeling studies show that the CO2 fertilization effect plays a dominant role, but the significant greening contribution of farmland areas at the global scale seems to indicate that **land management changes (LMC)** activities have a huge impact. **This study proposes the theory of Paired Land Use Experiment (PLUE), which selects areas with large differences in land management and consistent climate change to achieve "control" of climate change and attribute the difference in vegetation change to on the LMC.** The PLUE method can directly identify land management activities other than climate elements from observations at the regional scale, which is helpful for further research on the driving forces of long-term vegetation change trends.

Have we known the drivers of greening/browning well?




nature climate change LETTERS

Greening of the Earth and its drivers




ARTICLES nature sustainability

China and India lead in greening of the world through land-use management

Modeling vs Observational Analysis: Both made significant contributions, but the problem has not been well solved yet, **why?**

Modeling

Advantages: easy to quantify each driving factor by controlling variables and setting different scenarios

Disadvantages: lack or incomplete process cannot be fully expressed, which is especially prominent in land management changes

Observational Analysis

Advantages: Based on observational data, contains all real processes

Disadvantages: Difficult to both identify and quantify drivers

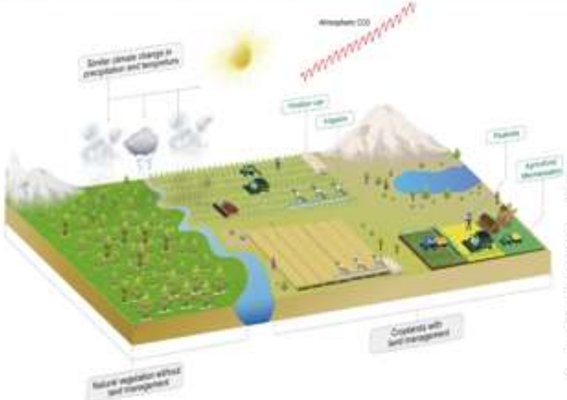
the Paired Land Use Experiments (PLUE) theory

Scientific question

Is it possible to identify and quantify the climatic and anthropogenic factors in the drivers of vegetation change from observations at the regional scale?

Our hypothesis

If the climate change in a certain area is basically the same, and there are significant differences in land use or land management, the influence of human factors can be identified and the contribution can be quantified under the condition of controlling "climate change".



The premise of PLUE is essentially based on the "natural experiment" approach: to infer and quantify the effects of a treatment while other independent variables are controlled based on a natural configuration. In the case of PLUE, we try to assess the effects of land management on the dynamics of LAI by controlling climate factors under a natural cross-border configuration.

Schematic diagram of the Paired Land Use Experiment

The general procedure of PLUE can be described as follows: select a region which is large enough but still with a roughly homogeneous climate environment. Two parts of such a region have different land use practices, and typical examples are natural vegetation and managed lands (such as croplands), or two managed lands with different intensity levels. At least one part has a stable land cover type.

Therefore, these two regions could be treated as a PLUE with identical climate change forcing. The difference of vegetation response to environment could be contributed to land use change and land management change. Two objectives are expected: first, to identify the significance of the impact of human activities with natural control experiment of climate environment change. Second, to quantify the contributions of human activities by abstracting natural variations (a base line of the climatic influences). While, the premise required by the second goal is not easy to achieve in reality and is based on certain assumptions.

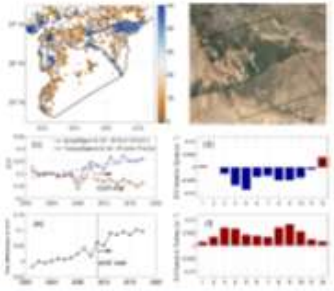
References of PLUE

Chen, T.*, Dolman, H., Sun, Z., Gao, H., Miao, L., Wei, X., Li, C., Han, Q., Shi, T., Wang, G., Zhou, S., Liang, C., and Chen, X. (2022): Land management explains the contrasting greening pattern across China-Russia border based on Paired Land Use Experiment approach, **JGR-Biogeosciences**, 127, e2021JG006659. https://doi.org/10.1029/2021JG006659

Chen, T.*, Guo, R., Yan, Q., Xin Chen, Zhou, S., Chen, X., Liang, C., Wei, X., and Dolman, H.(2022): Land Management Contributes significantly to observed Vegetation Browning in Syria during 2001–2018, ***Biogeosciences***, 19, 1515–1525. https://doi.org/10.5194/bg-19-1515-2022

CASE 1, the browning of Syria?

Spatial patterns of interannual trends during 2001–2018 for (a) EVI in the growing period, (b) EVI of croplands, (c) EVI of natural vegetation

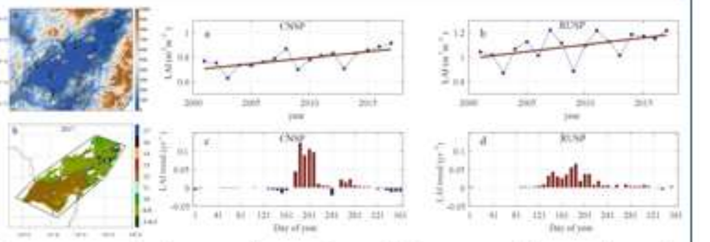


The PLUE analysis over the Khabur River basin across Syria and Turkey

(a) The percentage of the total cropland area used for irrigation for the reference year of 2005. (b) The Khabur River basin and cropland distribution. The vertex coordinates of the parallelogram are [37.5 N, 40.5 E], [36.75 N, 41 E], [37.25 N, 39.5 E] and [36.5 N, 40 E]. (c) Annual EVI series of cropland on the Syrian side and the Turkish side with linear fitting. (d) EVI trends of each month of cropland on the Syrian side. (e) EVI difference between the Turkish and Syrian sides of the basin (Turkey's EVI minus Syria's EVI). (f) EVI trends of each month of cropland on the Turkish side.

The LMC triggered by social unrest and civil war in Syria are responsible for the browning of northern regions. LMC includes insufficient irrigation and lack of seeds, fertilizers, pesticides and field management

CASE 2, Contrasting Greening Pattern Across China-Russia Border



The PLUE analysis over the Sanjiang Plain across China and Russia

the Sanjiang Plain has distinct land management practices across the border-intensified agricultural development on China side (CNSP) versus largely little-disturbed natural vegetation on Russia side (RUSP). Different LMC practices lead to notably different seasonal variability in vegetation changes. LMC in CNSP side contains dry croplands to paddy fields, agriculture mechanization and the usage of fertilizer and pesticide.

Current global warming is driven by human

It is unequivocal that human influence has warmed the atmosphere, ocean and land. Widespread and rapid changes in the atmosphere, ocean, cryosphere and biosphere have occurred.

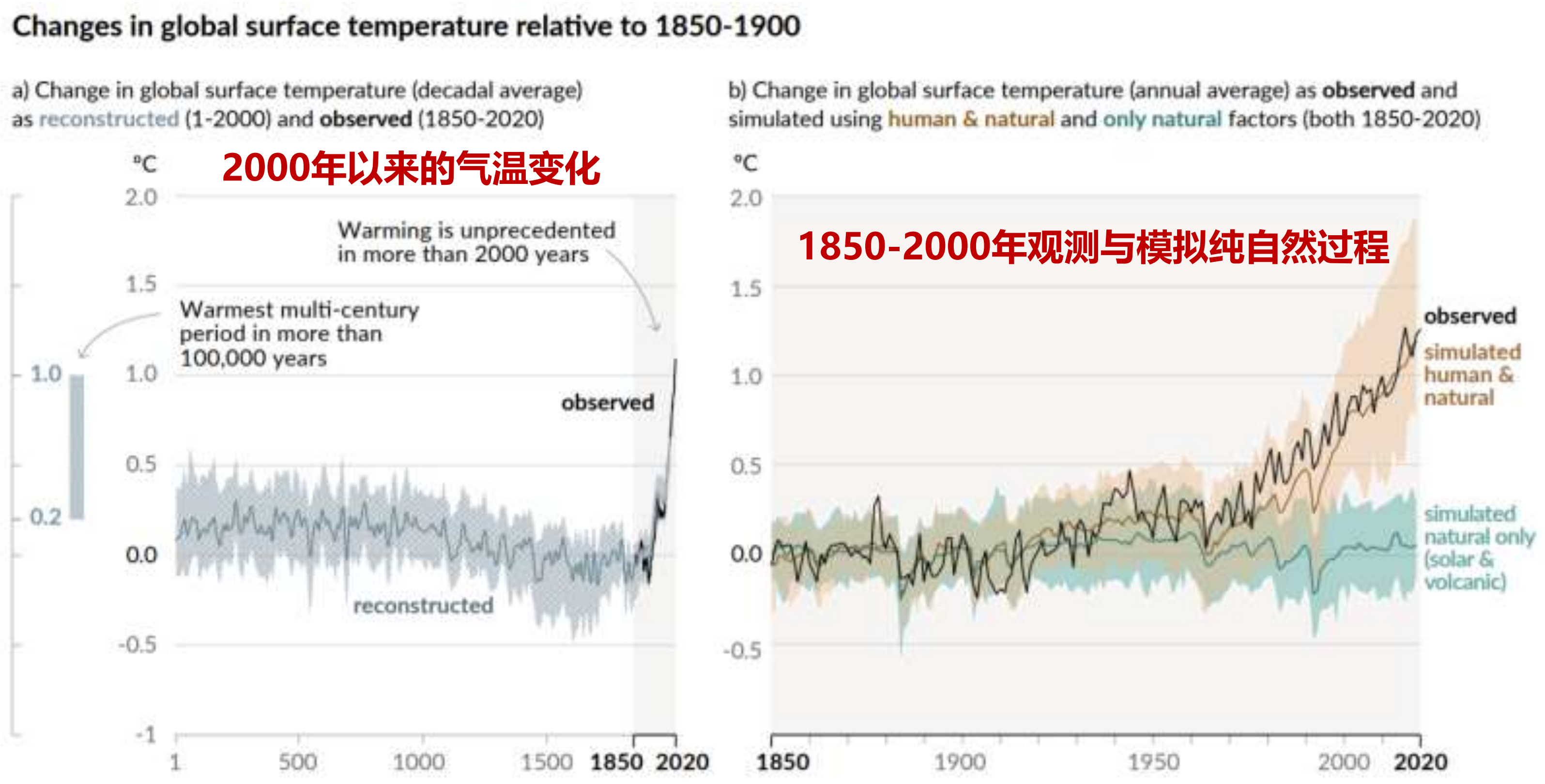


Figure SPM.1: History of global temperature change and causes of recent warming.

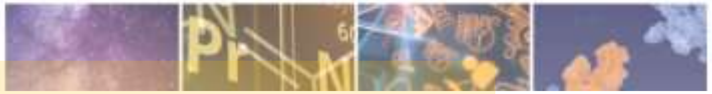
What Don’t We Know? 2005 & 2021

NUIST

BOOKLET · SPONSORED BY SHANGHAI JIAO TONG UNIVERSITY (SJTU)

125 questions: Exploration and discovery

14 MAY 2021



Ecology

Can we stop global climate change?

Climate change is one of the most pressing, complex, and frightening challenges facing us today, and scientists agree that ending it hinges on two major issues, both of which are now being addressed. The first roadblock is associated with the amount of climate data we are able to collect and share. We still lack a global climate observational system. We also need more investment in climate data infrastructure. Additionally, we contend with a lack of coordination and planning, with diverse actors spanning different countries, governments (national and local), sectors, and agencies. Moreover, much of our approach to climate change has been reactive rather than proactive. To truly stop climate change, more robust risk management systems must be established that complement and transform the work of environmental scientists, before further climate crises unfold.

The other major barrier we must overcome to stop climate change is our dependence on fossil fuels for most of our energy needs. If we can harness and utilize more green energy, such as solar, wind, geothermal, and nuclear, we have a fighting chance. Many nations, including China, the United States, and others around the world, are financing advanced research in this realm, although some of the concerns yet to be addressed include adjusting electrical grids to be able to manage the unpredictability of green energy as well as energy storage.

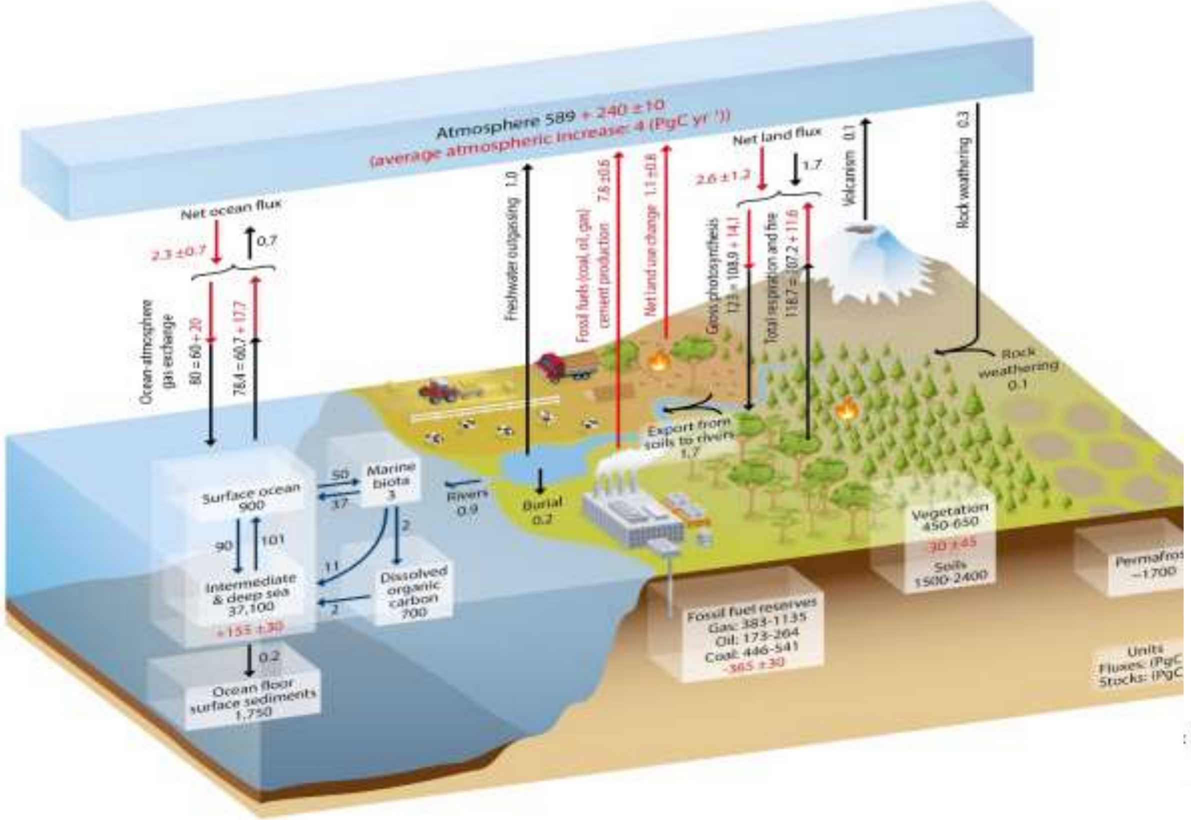
But we are not hobbled. We are in the midst of an extraordinary technological revolution in data science, computing, and energy science. We know how to build tools to collect valuable environmental data. Computer scientists are working with ecologists to apply unique artificial intelligence, deep-learning, and machine-learning techniques to Earth observation systems. And more funding for green energy research is becoming available across sectors. We have the capability: We can reduce our dependence on fossil fuels. We can practice better stewardship of the planet. With the help of the public and professionals alike, we may be able to arrive at a solution that is feasible, useful, and realistic.

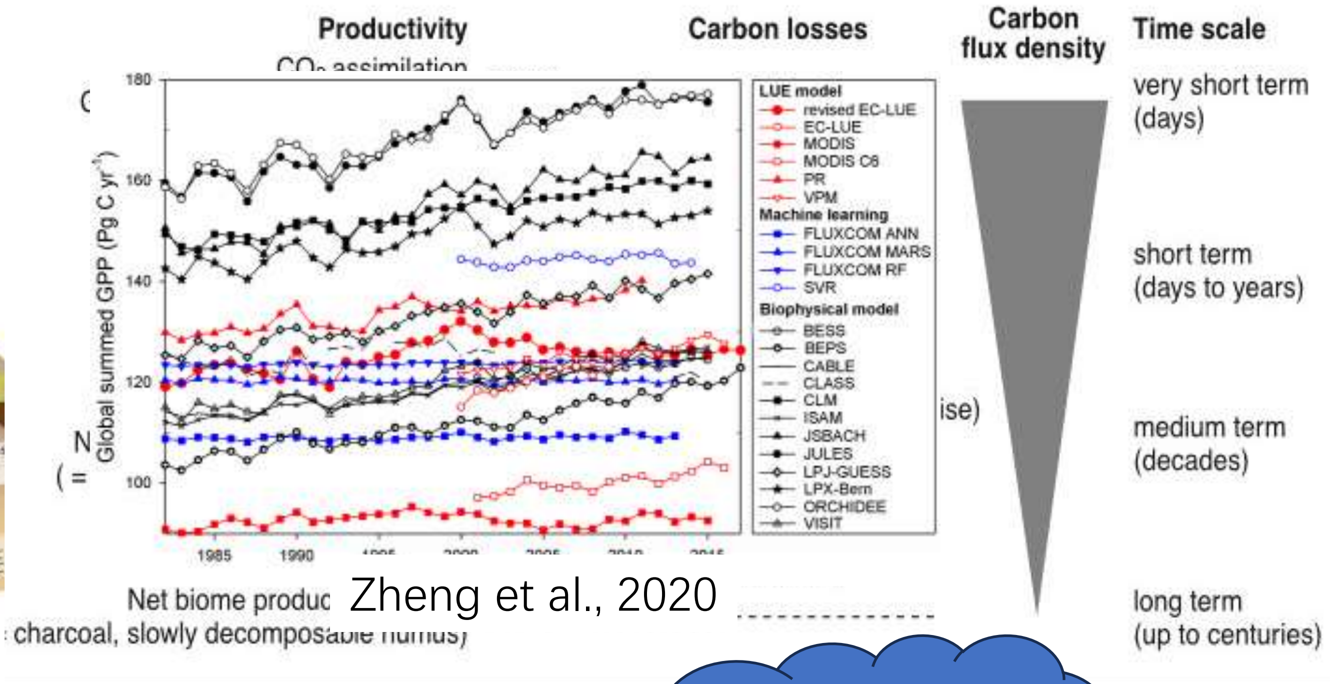
Where do we put all the excess carbon dioxide?

The 2019 global emissions of carbon dioxide (CO_2) were estimated to be approximately 33.1 billion metric tons, according to the U.S. Energy Information Administration, for which the United States is responsible for 15.4%. CO_2 arrives in the atmosphere mostly by two means: natural sources, such as human and animal exhalation and waste, and human actions, largely from energy production, such as burning coal, oil, and natural gas. One of the chief drivers of climate change research focuses on geologic and biologic carbon sequestration methods, where CO_2 is stored in underground geologic formations or in organic materials such as vegetation, soils, and woody products, as well as in aquatic environments. As the U.S. Geological Survey notes, “by encouraging the growth of plants—particularly larger plants like trees—advocates of biologic sequestration hope to help remove CO_2 from the atmosphere.”

What happens if all the ice on the planet melts?

If all the ice on the planet melts, sea level will rise 70 meters (230 feet), and every coastal city on the planet will flood.





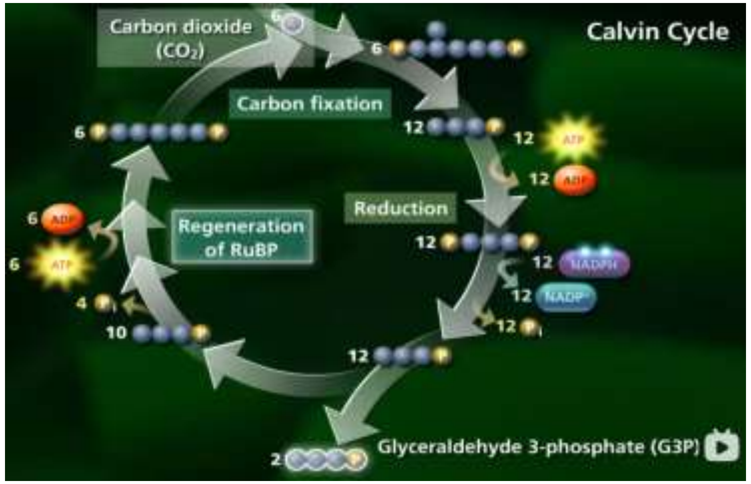
Zheng et al., 2020

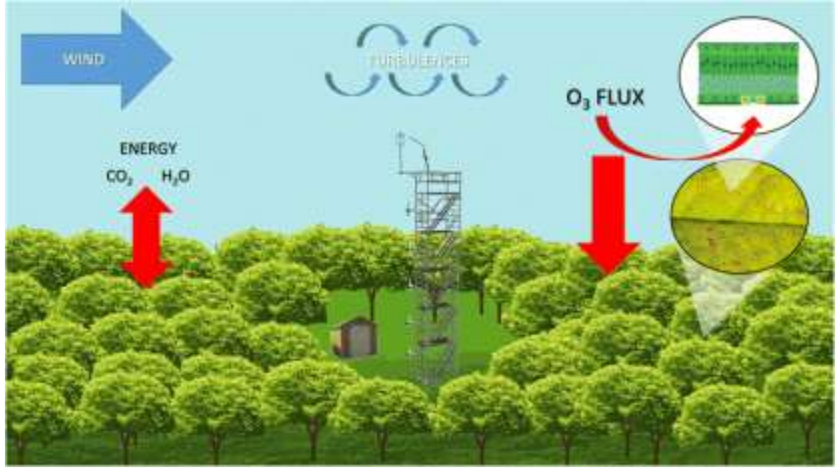
we're still stuck here

Carbon cycle research: the flux of each step in the cycle

The target indices of carbon sources and sinks are NEE and NBP, but GPP and NPP estimates still have huge errors.

Gross Primary production (GPP) :GPP is the synthesis of organic compounds from atmospheric or aqueous carbon dioxide. **Generally, GPP refers to the total amount of photosynthesis in terrestrial ecosystems.**

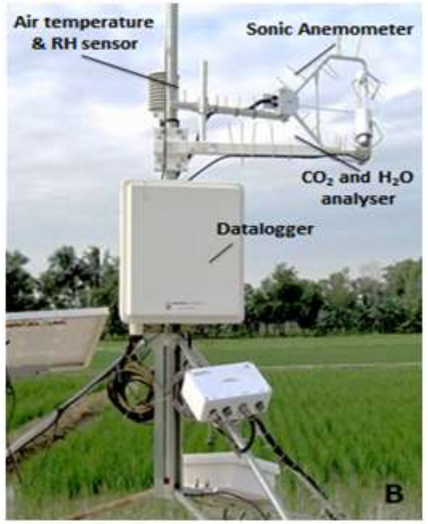




Leaf scale observations

Site scale observation

Regional and global scale estimation



$$GPP = PAR \times fPAR \times \varepsilon^*_{GPP} \times T(\varepsilon) \times W(\varepsilon)$$

$$GPP = (LUE / SIFyield) * SIF$$

Photosynthesis 光合作用 $A = \min(w_e, w_c, w_s)$

$$w_e = \begin{cases} \dfrac{(c_i - \Gamma^*)4.6 * APAR * \alpha}{c_i + 2\Gamma^*} \\ 4.6 * APAR * \alpha \end{cases} \quad w_c = \begin{cases} \dfrac{V_{cmax}(c_i - \Gamma^*)}{c_i + K_c(1 + o_i/K_o)} \\ V_{cmax} \end{cases} \quad w_s = \begin{cases} 0.5 * V_{cmax} & \text{for C3 plants} \\ \varepsilon * \dfrac{c_i}{P_{atm}} * V_{cmax} & \text{for C4 plants} \end{cases}$$

look up table of PFTs

NUIST

Plant functional type (PFT) : is a system used by climatologists to classify plants according to their physical, phylogenetic and phenological characteristics as part of an overall effort to develop a vegetation model for use in land use studies and climate models。 **Model parameters are often assigned separately based on PFTs, the spatial distribution of PFTs is the core input of the model.**

Table 2 Mean and standard deviation (SD) of V_{cmax} at the growing temperature (V_{cmaxTg}) and normalized to 25 °C (V_{cmax25}) for different plant functional types (PTF) calculated from the TROPOMI and ecological optimality theory (EOT) products in comparison with two ground-based databases (Smith et al., 2019 and Kattge et al., 2009).

		TROPOMI		EOT		Smith 2019		Kattge 2009	
PFT	(μmol m^{-2}s^{-1})	Mean	SD	Mean	SD	Mean	SD	Mean	SD
ENF	V_{cmax25}	32.36	12.51	60.66	7.19	53.70	26.95	62.50	24.70
	V_{cmaxTg}	7.31	3.62	13.68	2.97	17.43	11.13		
EBF	V_{cmax25}	46.89	13.02	54.55	6.79	45.83	23.27	43.80	16.83
	V_{cmaxTg}	44.22	15.98	50.88	12.19	37.12	23.59		
DNF	V_{cmax25}	44.38	8.93	60.50	5.05	44.82	23.34	39.10	11.70
	V_{cmaxTg}	10.95	2.58	14.93	2.09	11.59	6.28		
DBF	V_{cmax25}	44.42	16.42	59.60	6.31	51.31	25.06	57.70	21.20
	V_{cmaxTg}	18.12	17.07	22.68	15.68	24.31	20.72		
SHR	V_{cmax25}	53.30	13.60	61.37	7.55	50.63	27.75	57.85	19.55
	V_{cmaxTg}	13.21	11.24	15.76	14.54	31.88	27.80		
GRS	V_{cmax25}	74.74	22.76	69.45	12.37	82.70	47.86	78.20	31.10
	V_{cmaxTg}	49.30	40.10	41.42	27.85	21.65	18.25		
CRP	V_{cmax25}	87.57	17.42	62.12	9.59	90.21	32.13	100.70	36.60
	V_{cmaxTg}	54.83	37.14	39.63	26.72	42.11	22.64		

Chen, J. M., Wang, R., Liu, Y., He, L., Croft, H., Luo, X., ... & Dong, N. (2022). Global datasets of leaf photosynthetic capacity for ecological and earth system research. *Earth System Science Data*, *14*(9), 4077-4093.

Table 2.2. *Biome-Property-Look-Up-Table (BPLUT) for MODIS GPP/NPP algorithm with NCEP-DOE reanalysis II and the Collection5 FPAR/LAI as inputs. The full names for the University of Maryland land cover classification system (UMD_VEG_LC) in MCDLCHKM dataset (fieldname: Land_Cover_Type_1) are, Evergreen Needleleaf Forest (ENF), Evergreen Broadleaf Forest (EBF), Deciduous Needleleaf Forest (DNF), Deciduous Broadleaf Forest (DBF), Mixed forests (MF), Closed Shrublands (CShrub), Open Shrublands (OShrub), Woody Savannas (WSavanna), Savannas (Savanna), Grassland (Grass), and Croplands (Crop).*

UMD_VEG_LC	ENF	EBF	DNF	DBF	MF	CShrub	OShrub	WSavanna	Savanna	Grass	Crop
LUEmax (KgC/m^2/d/MJ)	0.000962	0.001268	0.001086	0.001165	0.001051	0.001281	0.000841	0.001239	0.001206	0.000860	0.001044
Tmin_min (C)	-8.00	-8.00	-8.00	-6.00	-7.00	-8.00	-8.00	-8.00	-8.00	-8.00	-8.00
Tmin_max (C)	8.31	9.09	10.44	9.94	9.50	8.61	8.80	11.39	11.39	12.02	12.02
VPD_min (Pa)	650.0	800.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0	650.0
VPD_max (Pa)	4600.0	3100.0	2300.0	1650.0	2400.0	4700.0	4800.0	3200.0	3100.0	5300.0	4300.0
SLA (LAI/KgC)	14.1	25.9	15.5	21.8	21.5	9.0	11.5	27.4	27.1	37.5	30.4
Q_{10}*	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
froot_leaf_ratio	1.2	1.1	1.7	1.1	1.1	1.0	1.3	1.8	1.8	2.6	2.0
livewood_leaf_ratio	0.182	0.162	0.165	0.203	0.203	0.079	0.040	0.091	0.051	0.000	0.000
leaf_mr_base	0.00604	0.00604	0.00815	0.00778	0.00778	0.00869	0.00519	0.00869	0.00869	0.0098	0.0098
froot_mr_base	0.00519	0.00519	0.00519	0.00519	0.00519	0.00519	0.00519	0.00519	0.00519	0.00819	0.00819
livewood_mr_base	0.00397	0.00397	0.00397	0.00371	0.00371	0.00436	0.00218	0.00312	0.00100	0.00000	0.00000

*: The constant Q_{10} = 2.0 is applied to fine roots and live wood, while for leaves, a temperature acclimation Q_{10} value is used as described in Equation.

Running, S. W., & Zhao, M. (2019). User's guide daily GPP and annual NPP (MOD17A2H/A3H) and year-end gap-filled (MOD17A2HGF/A3HGF) products NASA Earth Observing System MODIS Land Algorithm (For Collection 6). Process. DAAC, 490, 1-37.

Are further distinctions required within one PFT??

NUIST

Question: Does cropland qualify for a single PFT classification?

Fields vary widely in crop characteristics, unlike the natural attributes of other PFTs

I. The plant types are much more homogeneous than natural ecosystems due to management practice of farmers.

II. The plant types change much faster than natural ecosystems due to crop rotation schemes used, which means the land cover type does not uniquely determine plant types as in more natural ecosystems.

III. Sowing, ploughing and harvesting activities change the ecosystems in croplands abruptly and leave land fallow for long periods, sometimes even in the growing season.

IV. Multiple cropping is an important way to increase yield and bring unique phenological characteristics.

V. Agricultural modernization, including pesticide spread, fertilization, seed improvement, and the development of agricultural machinery.

Question: Does cropland qualify for a single PFT classification?

Fields vary widely in crop characteristics, unlike the natural attributes of other PFTs

I. f farmers.

II. n means the

III. and fallow

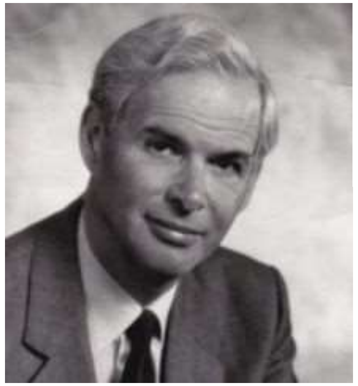
IV. s.

V. lopment of agricultural machinery.

The light use efficiency (LUE) model

NUIST

The LUE model was built by John Monteith (Monteith,1972, Monteith and Moss,1977)



Solar radiation and productivity in tropical ecosystems
JL Monteith - Journal of applied ecology, 1972 - JSTOR
In thermodynamic terms, ecosystems are machines supplied with energy from an ex-ternal source, usually the sun. When the input of energy to an ecosystem is exactly equal to its total ...
☆ 保存 99 引用 被引用次数：3103 相关文章 所有 4 个版本

Climate and the efficiency of crop production in Britain
JL Monteith - ... transactions of the royal society of London ..., 1977 - royalsocietypublishing.org
... Such analysis has already been applied to primary production in the tropics (Monteith 1972). This paper contains a similar analysis for Britain, with special em phasis on climatic ...
☆ 保存 99 引用 被引用次数：4665 相关文章 所有 7 个版本

Evaporation and environment
JL Monteith - Symposia of the society for experimental ..., 1965 - repository.rothamsted.ac.uk
A turgid leaf exposed to bright sunshine can transpire an amount of water several times its own weight during a summer day. Rapid evaporation is sustained by a supply of heat from the ...
☆ 保存 99 引用 被引用次数：8400 相关文章 所有 5 个版本

植物吸收光　　比例系数

$$\text{Growth} = \text{APAR} \times \varepsilon$$

模型演进

Tucker, 1979; Asrar et al., 1984; Sellers, 1985; Potter et al., 1993; Running et al., 2000 …

光合有效辐射　　最大光合利用效率

$$\text{GPP (or NPP)} = \text{PAR} \times \text{fPAR} \times \varepsilon^*_{GPP}\ (\text{or } \varepsilon^*_{NPP}) \times T(\varepsilon) \times \boxed{W(\varepsilon)}$$

光合有效辐射吸收率（植被指数）　　环境约束（温度+水分）

Built the look-up table

NUIST

Types: we divided the cropland into 26 different type based on the survey gridded dataset MIRCA2000(monthly irrigated and rainfed crop areas; Portmann et al., 2010)。

$$RMSE = [\frac{1}{N}\sum_{n=1}^{N}(GPP_{CASA} - GPP_{FLUXNET})^2]^{1/2}$$

FLUXNET only can cover 8 types which account about 55% global cropland area。 The rest 18 types are from literature. How?

Convert the parameter

$$\varepsilon^*_{GPP\_FLUXNET} = 0.6757 \times \varepsilon^*_{GPP\_literature} + 0.1252$$

The citations of Monteith's two papers in google scholar

Climate and the efficiency of crop production in Britain [and discussion]

JL **Monteith**, CJ Moss - Philosophical Transactions of ..., 1977 - rstb.royalsocietypublishing.org

Abstract The efficiency of crop production is defined in thermodynamic terms as the ratio of energy output (carbohydrate) to energy input (solar radiation). Temperature and water supply are the main climatic constraints on efficiency. Over most of Britain, the radiation ...

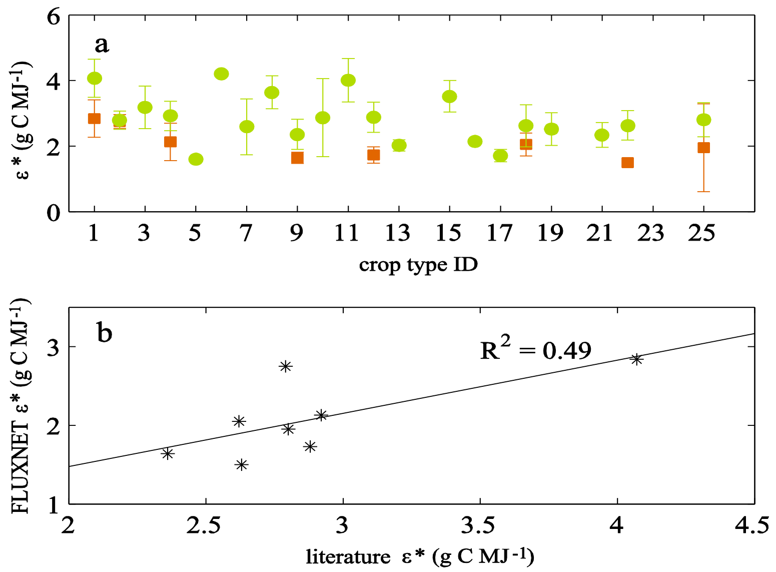
Cited by 2334 Related articles All 7 versions Cite Save

Solar radiation and productivity in tropical ecosystems

JL **Monteith** - Journal of applied ecology, 1972 - JSTOR

In thermodynamic terms, ecosystems are machines supplied with energy from an external source, usually the sun. When the input of energy to an ecosystem is exactly equal to its total output of energy, the state of equilibrium which exists is a special case of the First Law of ...

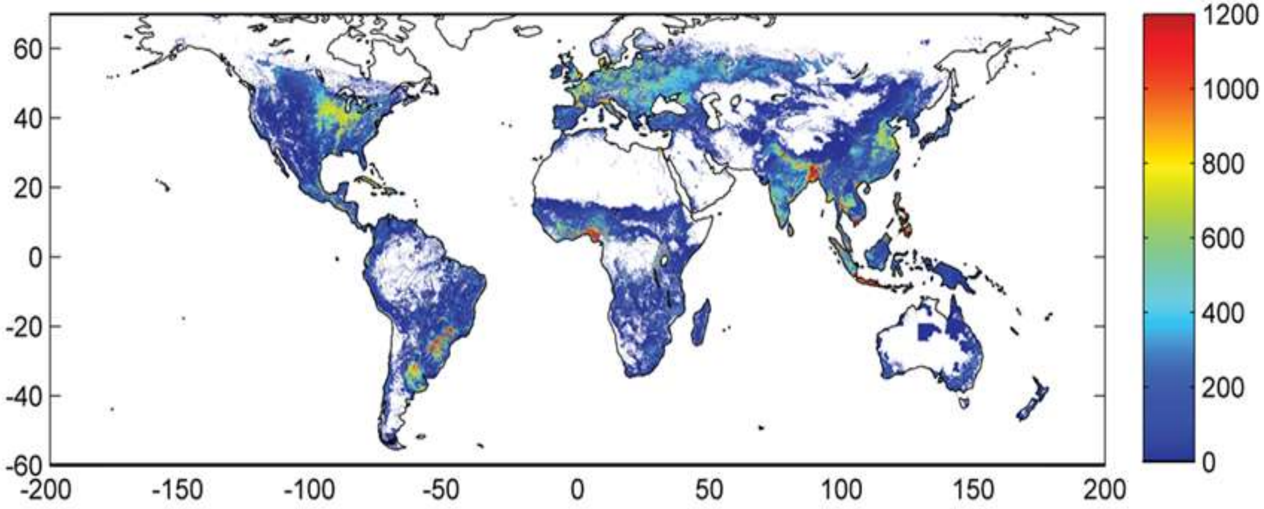
Cited by 1244 Related articles All 5 versions Cite Save



Global Cropland GPP Estimates

NUIST

Global cropland GPP in 2000 is 11.05 Pg C.

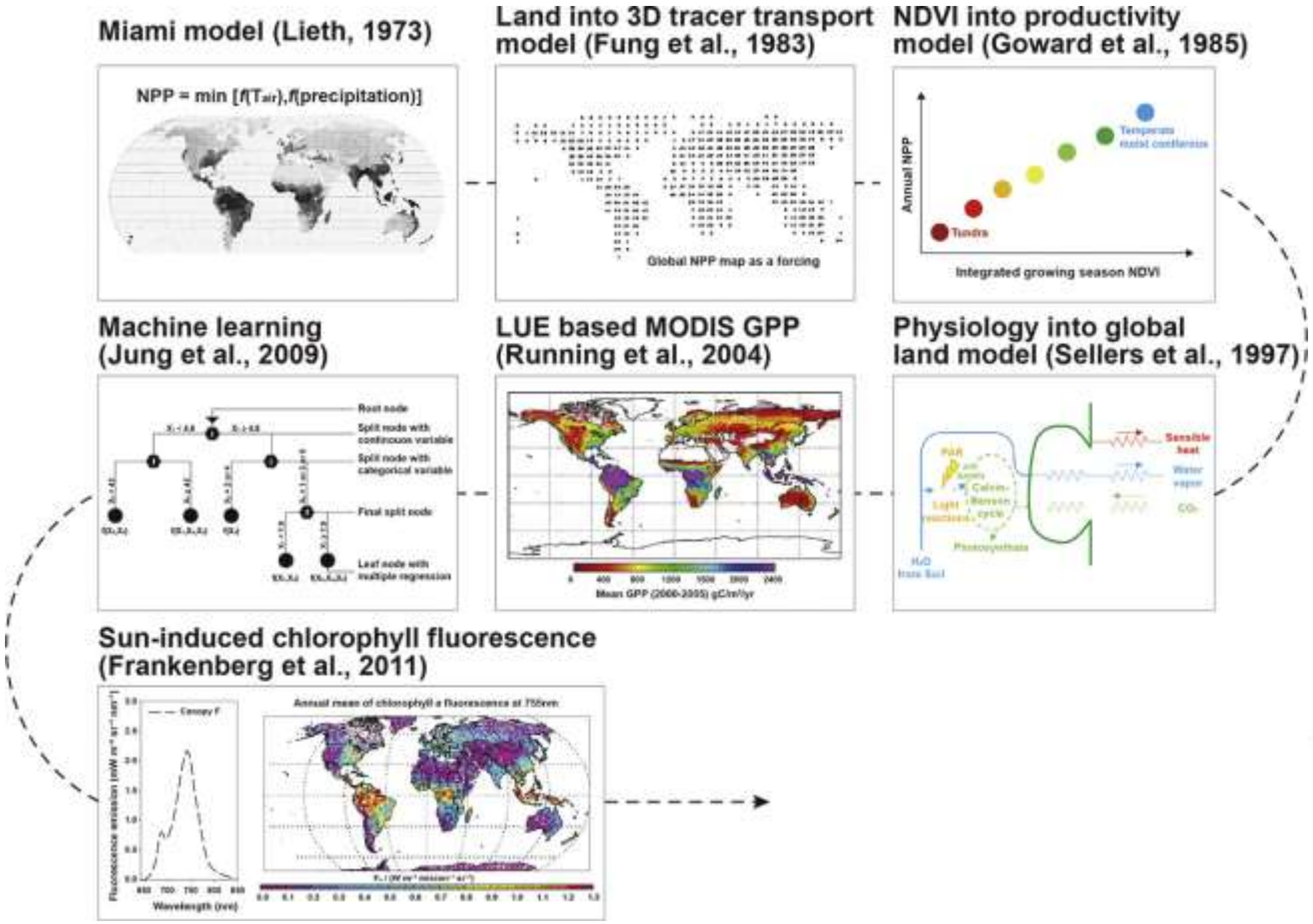


The spatial distribution of global crop land GPP.

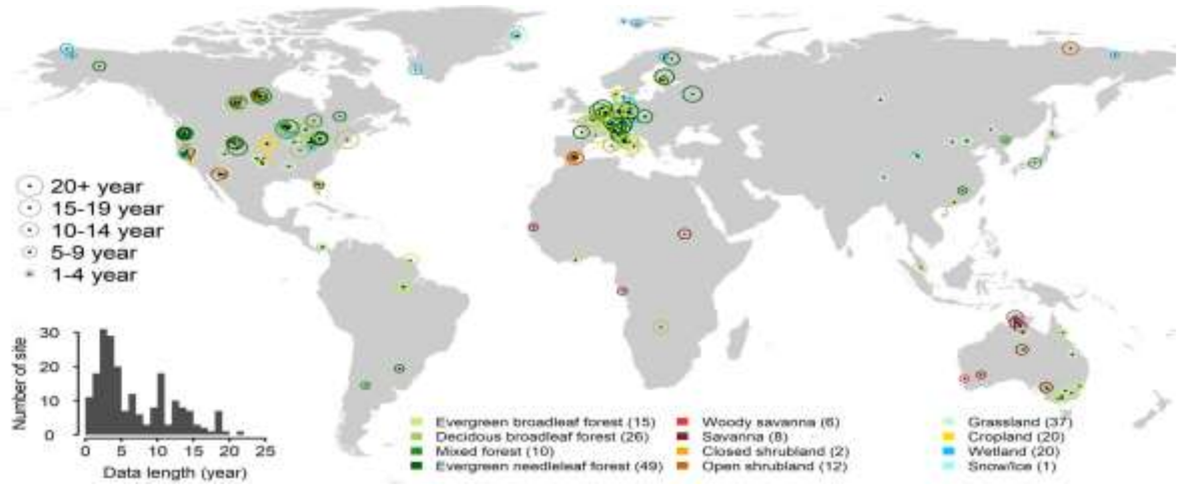
ID	crop types	GPP(Pg C yr^{-1})
1	Maize	1.545
2	Rice	1.514
3	Fodder grasses	1.389
4	Wheat	1.384
5	Others perennial	0.795
6	Cassava	0.612
7	Others annual.	0.508
8	Sugarcane	0.494
9	Soybeans	0.491
10	Pulses	0.353
11	Sorghum	0.272
12	Barley	0.26
13	Oil palm	0.21
14	Coffee	0.158
15	Millet	0.134
16	Cocoa	0.132
17	Cotton	0.123
18	Rape seed	0.115
19	Sunflower	0.112
20	Rye	0.109
21	Groundnuts	0.105
22	Potatoes	0.091
23	Citrus	0.064
24	Grapes	0.041
25	Sugar beet	0.04
26	Date palm	0.001
	global	11.05

Global Scale GPP Estimates

NUIST



The representative work of global GPP estimation, in which the algorithm based on machine learning is the first time to summarize the global flux network data FLUXCOM, and its estimated value is generally regarded as a reference.



全球涡度通量站点分布图

Ryu, Y., Berry, J. A., & Baldocchi, D. D. (2019). What is global photosynthesis? History, uncertainties and opportunities. *Remote sensing of environment*, *223*, 95-114.

Global Scale GPP Estimates

NUIST

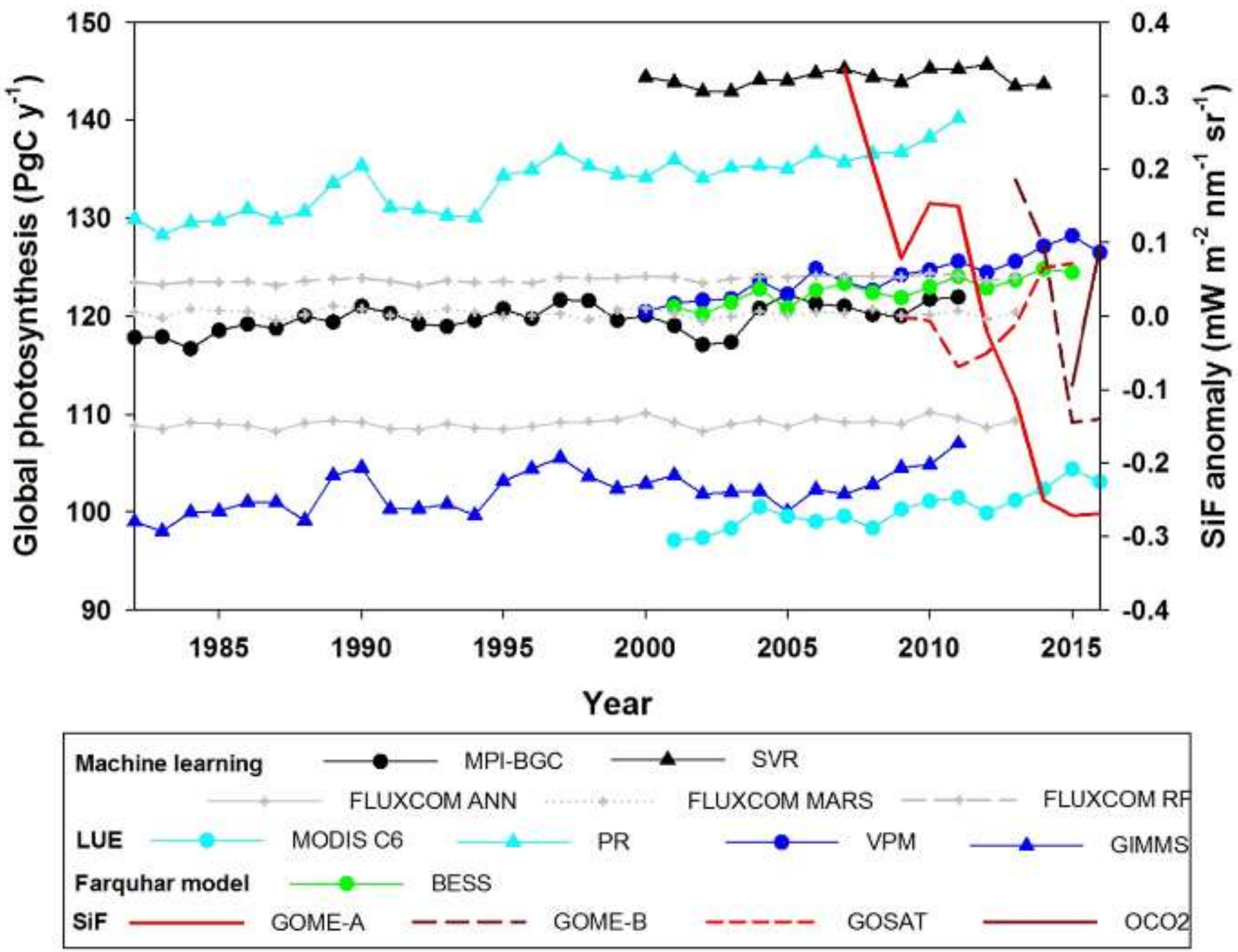


Fig. 3. Time series of annual global photosynthesis and SiF estimates from different remote sensing based products. SiF datasets present annual anomaly values adopted from (Luo et al., 2018). Data sources include: MPI-BGC (Jung et al., 2010), SVR (Kondo et al., 2015), FLUXCOM (Jung et al., 2017), MODIS C6 GPP product (Running et al., 2004), PR (Keenan et al., 2016), VPM (Zhang et al., 2017b), GIMMS (Smith et al., 2016), BESS (Jiang and Ryu, 2016), GOME (Joiner et al., 2013), GOSAT (Frankenberg et al., 2011), and OCO2 (Sun et al., 2017).

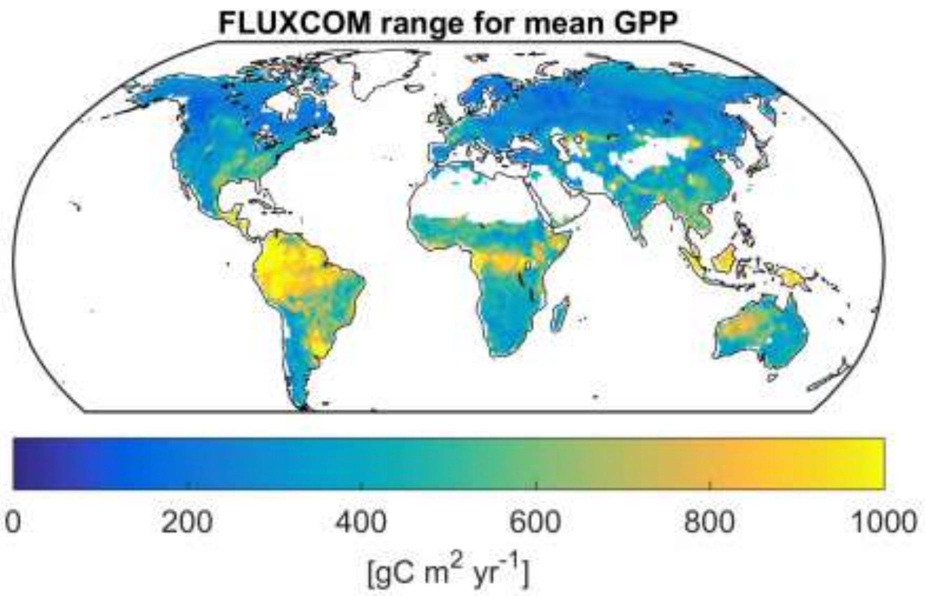
As can be seen from the gray curve in the right figure, the FLUXCOM estimate obviously lacks inter-annual fluctuations!

Ryu, Y., Berry, J. A., & Baldocchi, D. D. (2019). What is global photosynthesis? History, uncertainties and opportunities. *Remote sensing of environment*, *223*, 95-114.

Global Scale GPP Estimates

NUIST

references	Main results
jung2011	Use the MTE method to estimate global GPP
yao2017	Use the RF method with Asian eddy flux sites to estimate China GPP
jung2020	Generation of global-scale GPP by fusion of multiple machine learning methods
huang2020	Introduce PFT for GPP classification training, and verify the important influence of water body index on GPP estimation
schlund2020	Uncertainty analysis of machine learning accuracy, comparison of various GPP products, pointing out the importance of farmland to GPP estimation



FLUXCOM global GPP（jung2020）

Previous studies have clearly pointed out that the sparse sites is the main limitation in machine learning estimation of global-scale GPP. The uncertainty of global GPP estimation reduced by quantitative site data has not been effectively resolved.

Global Scale GPP Estimates

NUIST

FluxCom | FluxCom

Home | Download | Upload | Registration | Administration

Available Files >> **File Details** >> Data Usage Agreement >> Download

File Details

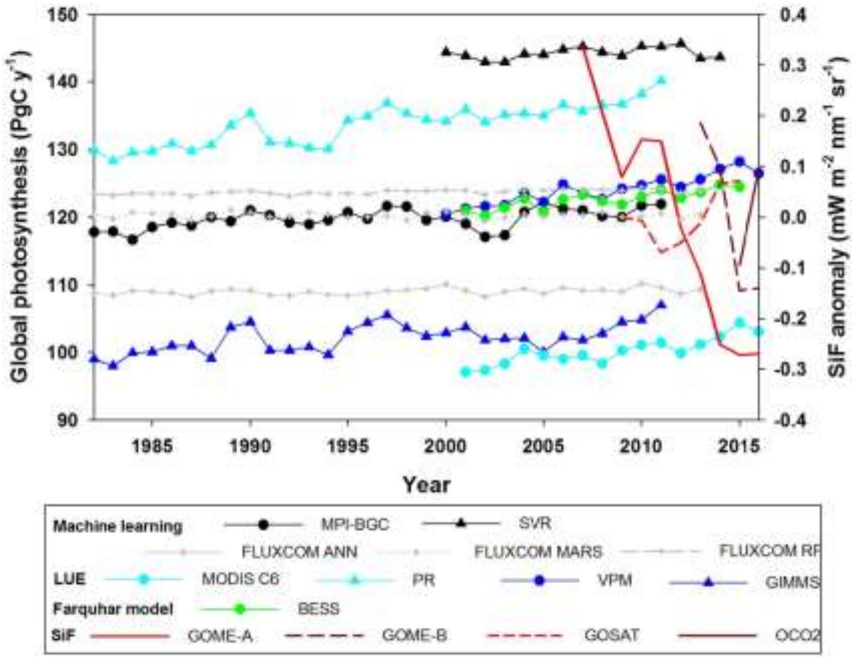
(-1: yes; 0: no)

FileID	259
Title	FLUXCOM (RS+METEO) Global Land Carbon Fluxes using CRUNCEP climate data
File Name	
Version	v1
Description	Please cite both, Jung et al., 2016 and Tramontana et al., 2016, when using any of the data for publications. This data set provides 1) subfolder "raw": global land carbon fluxes (GPP, TER) on daily to annual resolution from 1980-2013 generated by 3 machine learning methods (RF, ANN, MARS) which were forced with CRUNCEPv6 meteorological data and mean seasonal cycles of several MODIS based variables (RS+METEO setup, see Tramontana et al., 2016 for details); 2) subfolder "AnomaliesClim": detrended carbon flux anomalies driven by air temperature (Tair), water availability (WAI2), shortwave radiation (Rg) as described in Jung et al., 2016. Two variants of GPP and TER refer to 2 flux partitioning methods, where _HB refers to Lasslop et al., 2010, and no specifier refers to Reichstein et al. 2005. Usage notes: Long-term trends and interannual variations in this data set originate exclusively from direct effects of changing climate; i.e. the climate forcing data set CRUNCEPv6 used here. Results from using other climate forcing data will also be available. Potential effects due to e.g. CO2 fertilization or vegetation greening are not accounted for. Magnitudes of interannual variations appear to be too small in the datasets, and a normalization of the anomalies is recommended when analysing interannual varibility. Analysing long-term mean TER is not recommended due to a likely bias of mean annual TER; the same hold for mean annual NEE computed from TER-GPP. The choice of flux partitioning variant is usually not critical; we would give a slight preference to Reichstein et al., 2005 variants.
Publ. Date	2020-01-20
Public	-1
NcDump	
Workpackage	
TaskNo	10.17871/F
DOI	10.17871/FLUXCOM_RS_METEO_CRUNCEPv6_1980_2013_v1
GeoLocation	global (90°N-90°S; 180°W-180°E), 0.5° spatial resolution
Owner	Jung
Projects	FluxCom

Download

(Registration required)

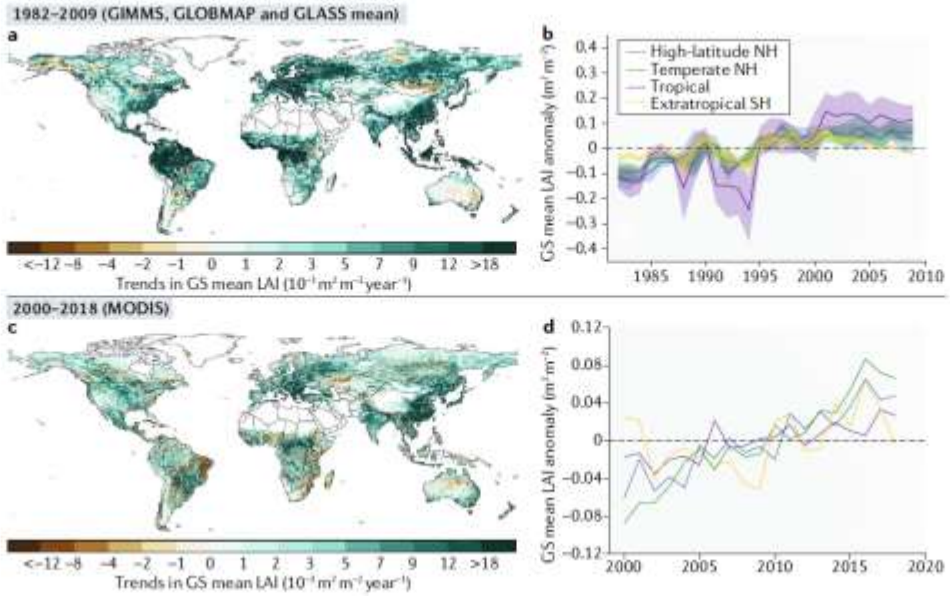
FLUXCOM data information



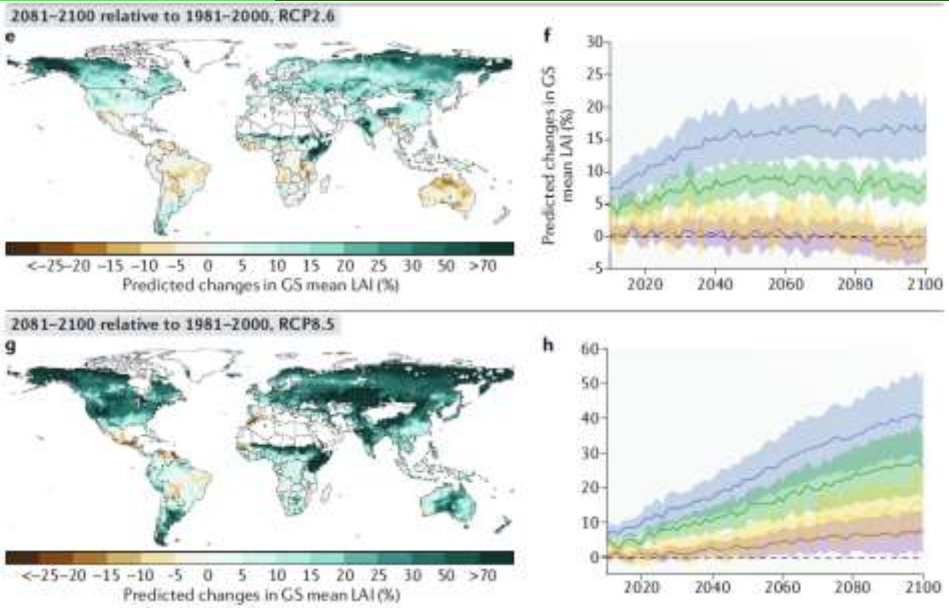
The trend of the global total of FLUXCOM_GPP is not obvious, and there is no interannual fluctuation, which is inconsistent with many evidences from observation and remote sensing. The potential reason is that the data set is not based on different PFTs for model training, and the multi-feature mixture amplifies the environmental differences between PFTs, thereby reducing the impact of CO2 fertilization and vegetation greening.

Significant increase in greenness of global vegetation

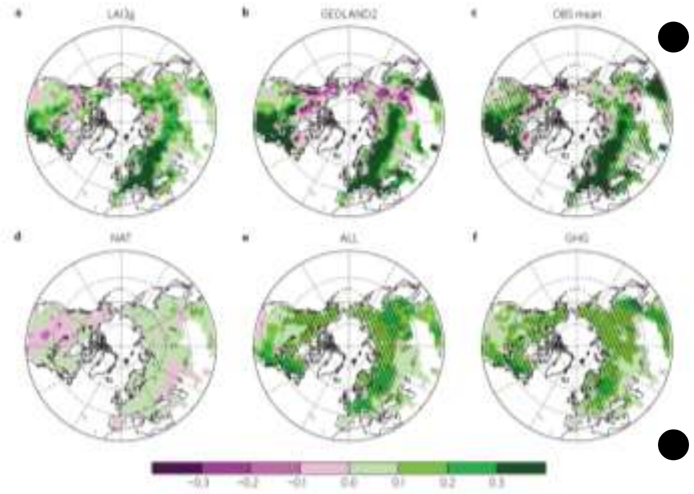
NUIST



基于遥感植被指数的变绿现象 Piao 2019 Nature Reviews



未来气候变化与植被变化预估 Piao 2019 Nature Reviews



北半球变绿趋势1982 - 2011, Mao 2016 NCC

- Vegetation "greening" generally refers to the trend of increasing vegetation greenness on an interannual scale. In applications, remote sensing vegetation indices are often used, including leaf area index LAI, normalized difference vegetation index NDVI, and enhanced vegetation index EVI.
- As an important indicator of vegetation productivity, GPP has become an indisputable fact that the annual total has increased year by year with the global greening phenomenon.

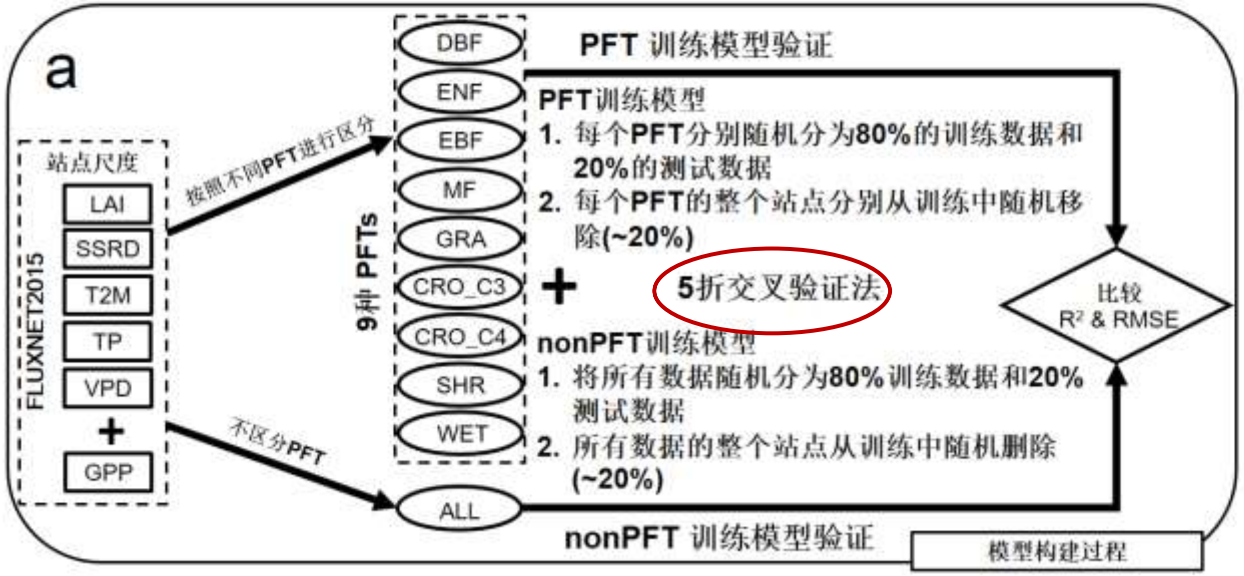
Core scientific question: How to improve GPP estimation of large-scale terrestrial ecosystems based on limited observations?

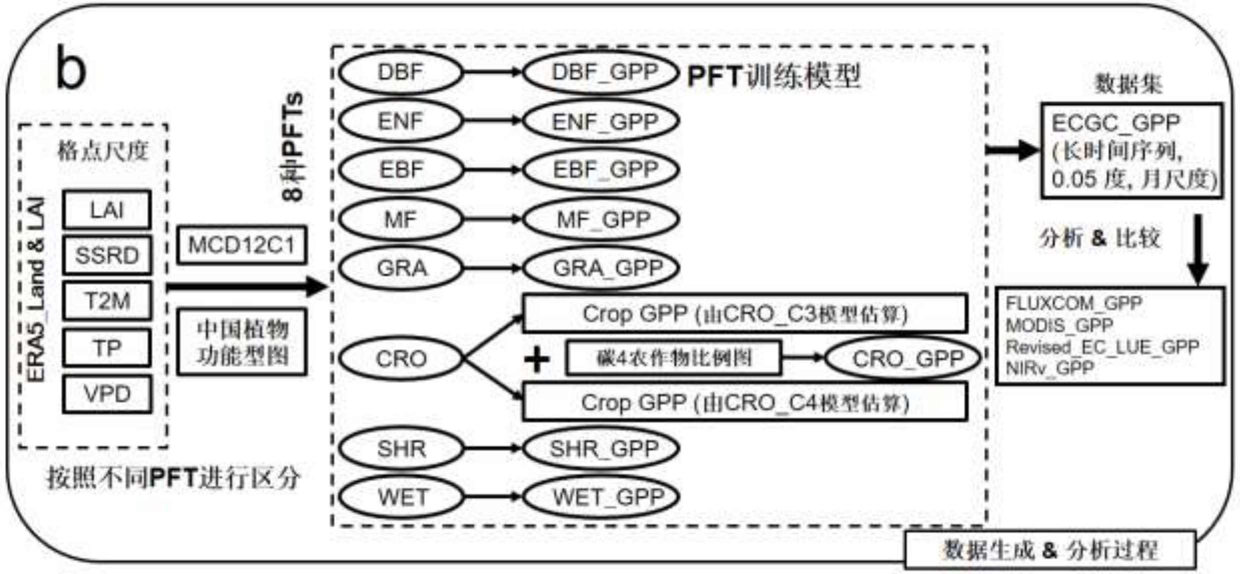
Two obvious concerns:

- The lack of interannual fluctuations in global GPP estimated by traditional machine learning is contrary to the objective fact of global greening. How to improve the interannual variation trend of GPP?

- The functional types of vegetation vary greatly. Whether machine learning training can be trained independently for different types, especially C4/C3 vegetation types.

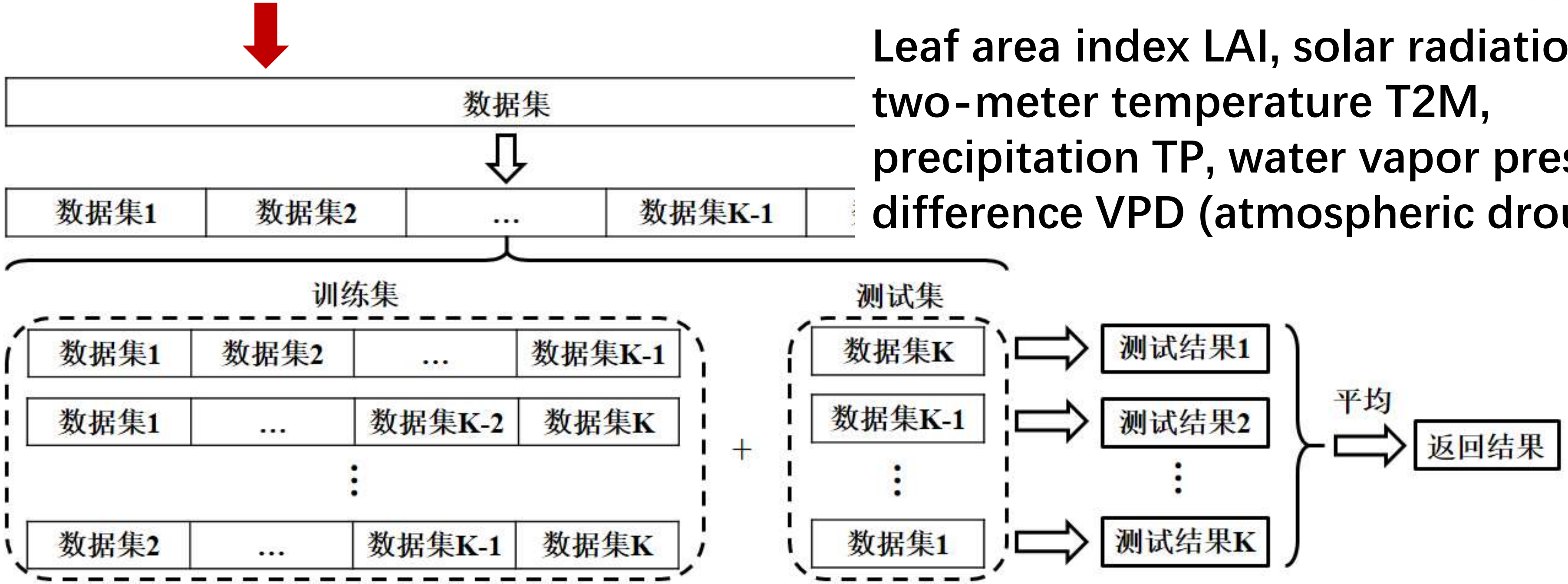
Built the random forest model





K=5
5折交叉验证法

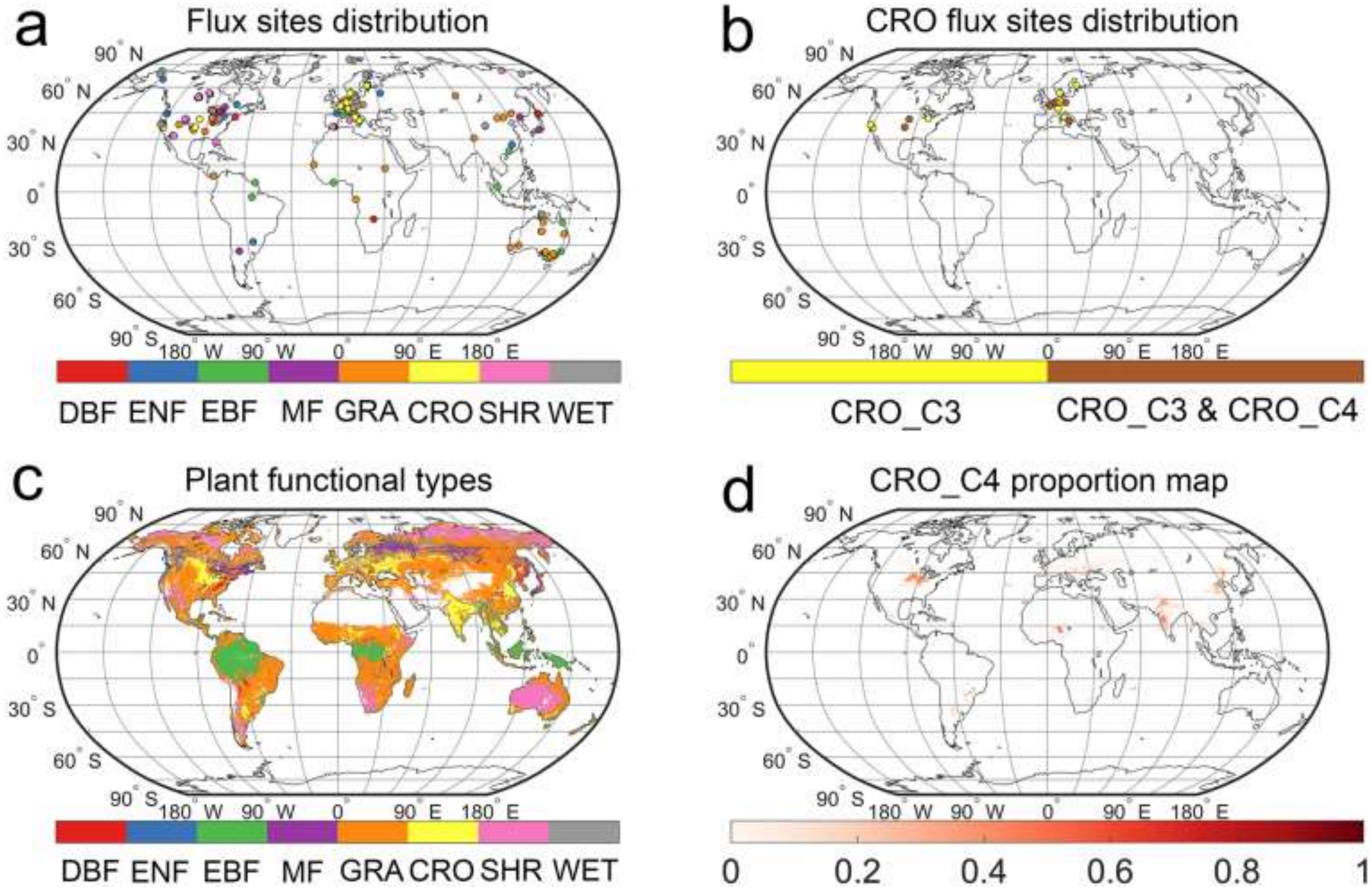
每次迭代过程中每个样本点只有一次被划入训练集或测试集的机会，避免过学习和欠学习状态。



Leaf area index LAI, solar radiation SSRD, two-meter temperature T2M, precipitation TP, water vapor pressure difference VPD (atmospheric drought)

PFT maps

NUIST



The spatial distribution map of major global crops released by EarthStat in 2000 was used (Monfreda et al., 2008), which provides the percentage of planted area for 175 crop types per 10-km grid cell. The CRO_C4 percentage was calculated by summing the six CRO_C4 types (corn, corn forage, sorghum, sorghum forage, millet, sugarcane)

We used nine PFTs here, including deciduous broad leaved forest (DBF), evergreen needle leaved forest (ENF), evergreen broad leaved forest (EBF), mixed forest (MF), grassland (GRA), CRO_C3, CRO_C4, shrub (SHR), and wetland (WET). The 206 flux sites in FLUXNET2015 were characterized by corresponding PFT colors under different PFTs.

Site-Scale Validation: PFT vs. nonPFT

NUIST

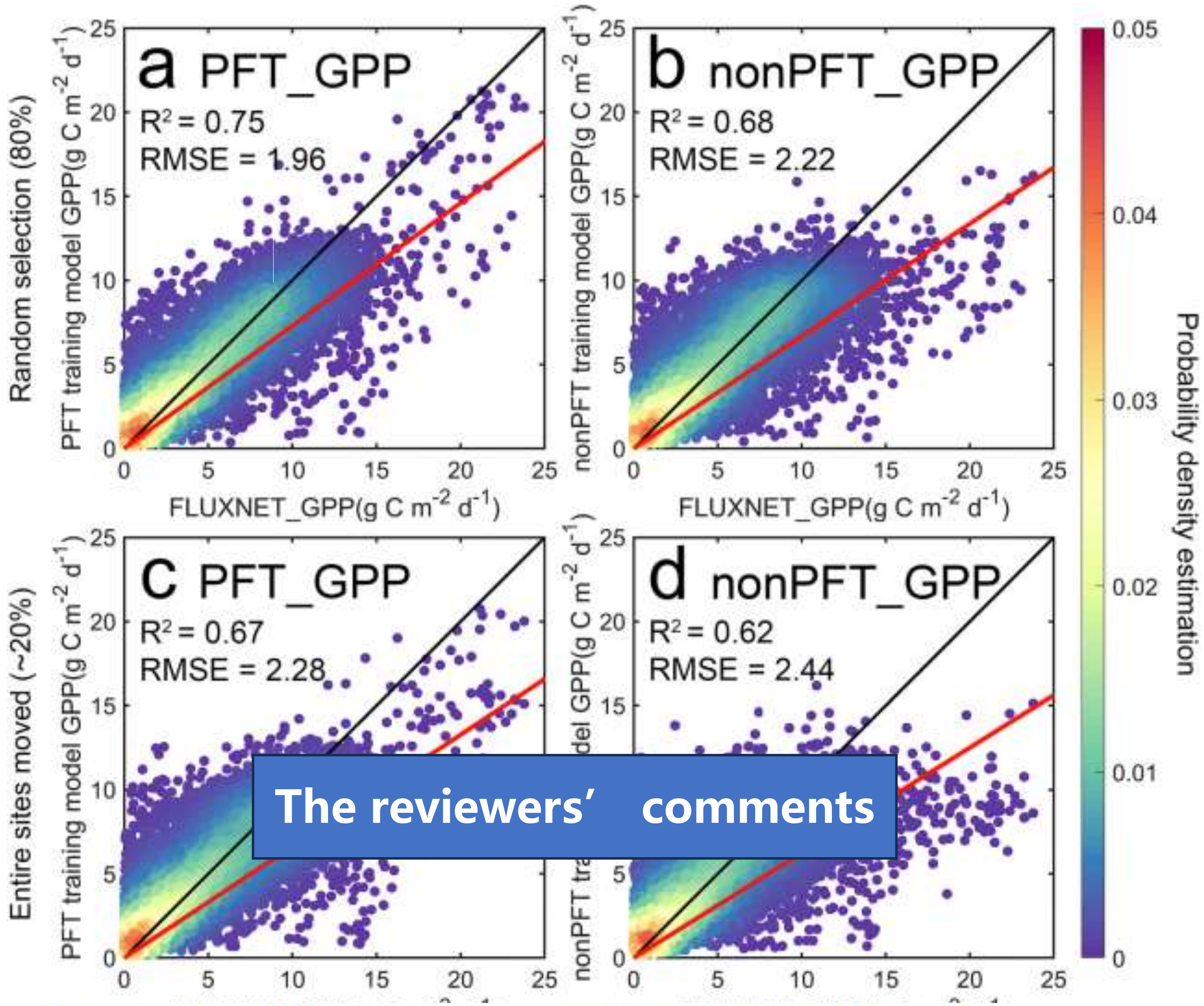


Figure 3. Scatter density diagram of the overall accuracy in plant functional type (PFT) training models and nonPFT training models. (a) PFT training model accuracy with 80% random selection and 20% testing sets. (b) NonPFT training model accuracy with 80% random selection and 20% testing sets. (c) PFT training model accuracy with entire sites moved (~20%) into testing sets. (d) NonPFT training model accuracy with entire sites moved (~20%) into testing sets. The red lines indicate the best linear fits determined by ordinary linear regression, and the black lines represent the 1:1 line. RMSE unit is g C m^{-2} d^{-1}.

By distinguishing the CRO_C3 and CRO_C4 in the CRO and training, different RF model based on each PFT could avoid systematic errors caused by differences in vegetation and growth environments effectively. In addition, the improvement of the overall training accuracy of the model, it also provided an opportunity for optimizing the prediction of GPP at the global scale.

Site-Scale Validation: PFT vs. nonPFT

NUIST

	R^2_PFT	RMSE_PFT	R^2_nonPFT	RMSE_nonPFT	R^2_diff	RMSE_diff
DBF	0.82 ± 0.01	1.92 ± 0.07	0.79 ± 0.01	2.16 ± 0.07	0.03	-0.24
ENF	0.68 ± 0.03	1.75 ± 0.08	0.65 ± 0.03	1.87 ± 0.09	0.03	-0.12
EBF	0.69 ± 0.07	1.86 ± 0.36	0.58 ± 0.09	2.16 ± 0.40	0.11	-0.30
MF	0.79 ± 0.03	1.51 ± 0.08	0.67 ± 0.04	2.02 ± 0.17	0.12	-0.51
GRA	0.78 ± 0.02	1.63 ± 0.09	0.75 ± 0.03	1.73 ± 0.12	0.03	-0.10
CRO_C3	0.51 ± 0.04	3.14 ± 0.15	0.47 ± 0.05	3.24 ± 0.20	0.04	-0.10
CRO_C4	0.84 ± 0.05	2.81 ± 0.32	0.78 ± 0.05	4.27 ± 0.58	0.06	-1.46
SHR	0.84 ± 0.04	0.71 ± 0.05	0.80 ± 0.06	1.29 ± 0.10	0.04	-0.58
WET	0.58 ± 0.04	2.33 ± 0.18	0.54 ± 0.08	2.57 ± 0.25	0.04	-0.24

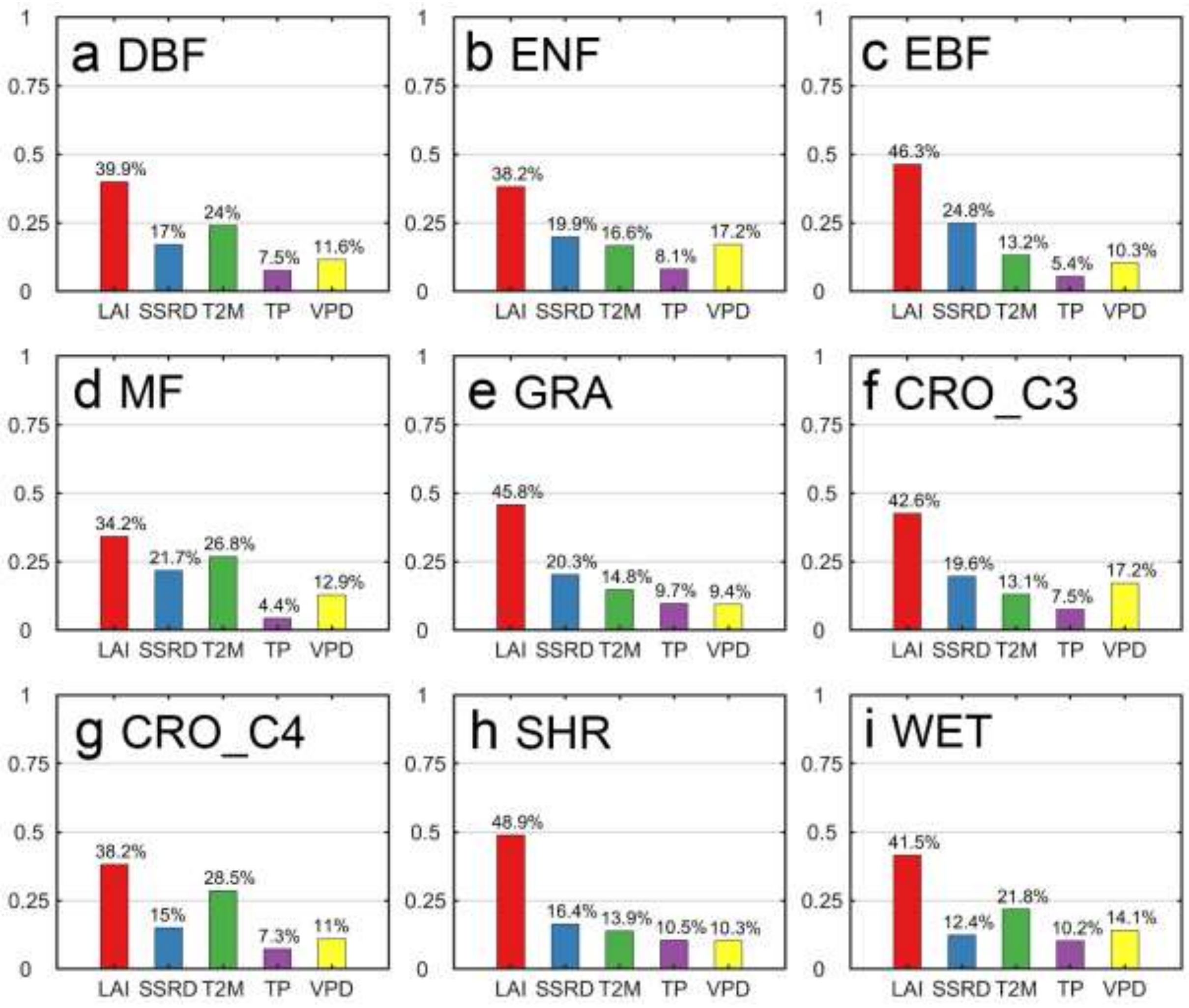
随机选择实验

- EBF、MF和CRO_C4的R^2均提升0.06以上。
- CRO_C3的R^2提升了0.04。
- CRO_C4的RMSE平均值降低了1.46 g C m^{-2} d^{-1}，达到2.81 g C m^{-2} d^{-1}。

全站点去除

- EBF、MF和CRO_C4的R^2均增加大于0.06。
- CRO_C4的RMSE减小量最大，达到1.32 g C m^{-2} d^{-1} 。

	R^2_PFT	RMSE_PFT	R^2_nonPFT	RMSE_nonPFT	R^2_diff	RMSE_diff
DBF	0.79 ± 0.07	2.12 ± 0.31	0.74 ± 0.12	2.28 ± 0.38	0.05	-0.16
ENF	0.62 ± 0.10	2.04 ± 0.39	0.61 ± 0.11	2.06 ± 0.34	0.01	-0.02
EBF	0.61 ± 0.09	2.20 ± 1.36	0.55 ± 0.12	2.26 ± 1.15	0.06	-0.06
MF	0.79 ± 0.07	1.88 ± 0.29	0.72 ± 0.21	2.09 ± 0.61	0.07	-0.21
GRA	0.72 ± 0.06	1.87 ± 0.51	0.71 ± 0.07	1.89 ± 0.54	0.01	-0.02
CRO_C3	0.48 ± 0.09	3.32 ± 0.74	0.47 ± 0.12	3.40 ± 0.74	0.01	-0.08
CRO_C4	0.76 ± 0.13	3.22 ± 0.46	0.65 ± 0.26	4.54 ± 1.24	0.11	-1.32
SHR	0.75 ± 0.14	0.90 ± 0.32	0.72 ± 0.18	1.37 ± 0.41	0.03	-0.47
WET	0.53 ± 0.22	2.67 ± 0.50	0.52 ± 0.22	2.68 ± 0.64	0.01	-0.01

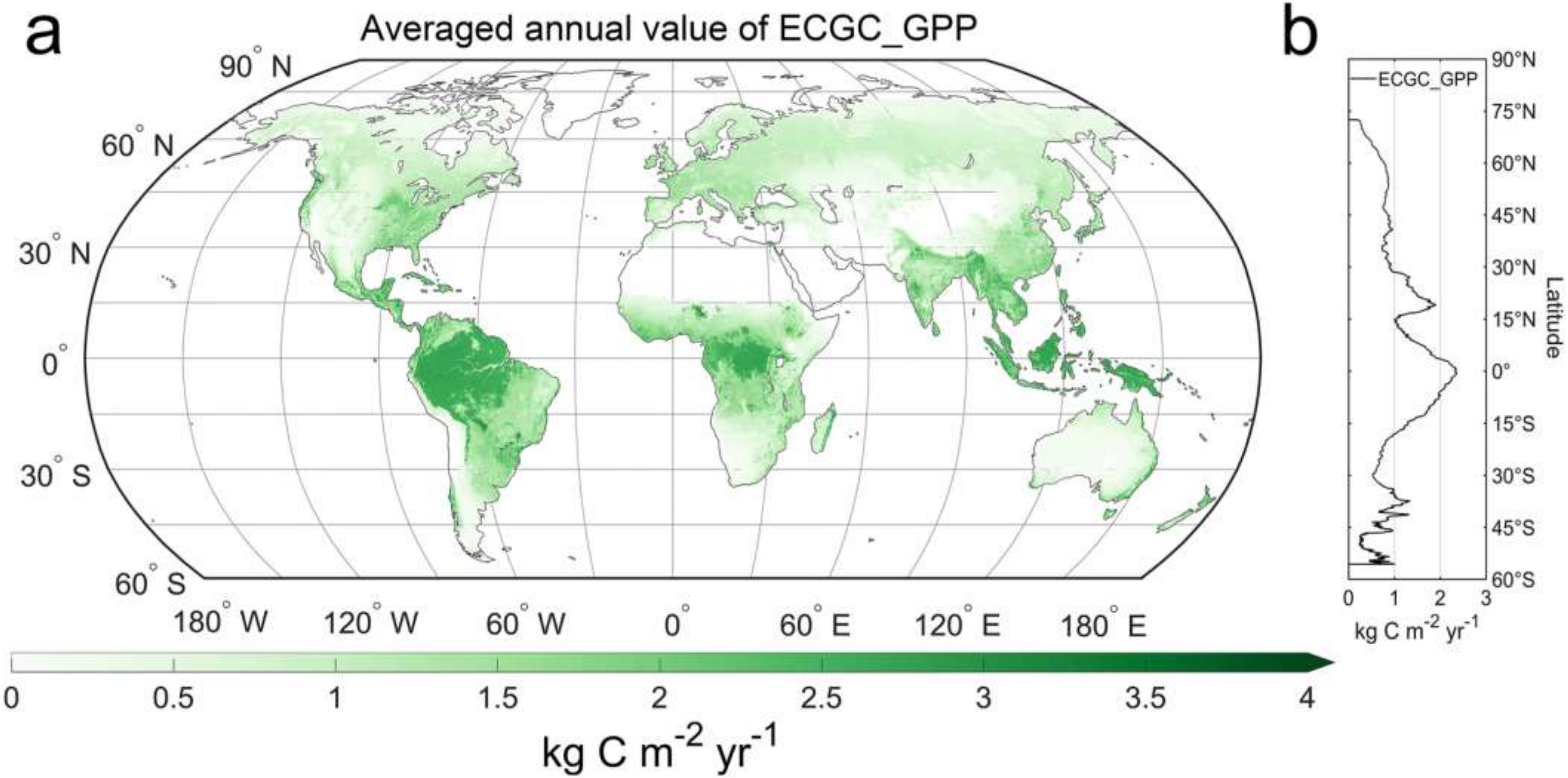


Comparison of feature contributions of nine plant functional type (PFT) training models. (a)–(i) Contribution of meteorological and remote sensing variables to the corresponding PFT.

The contribution of LAI is larger than other meteorological features among all PFTs, with an average contribution of 41.73%.

Details of input datasets used to drive RF model and GPP datasets.

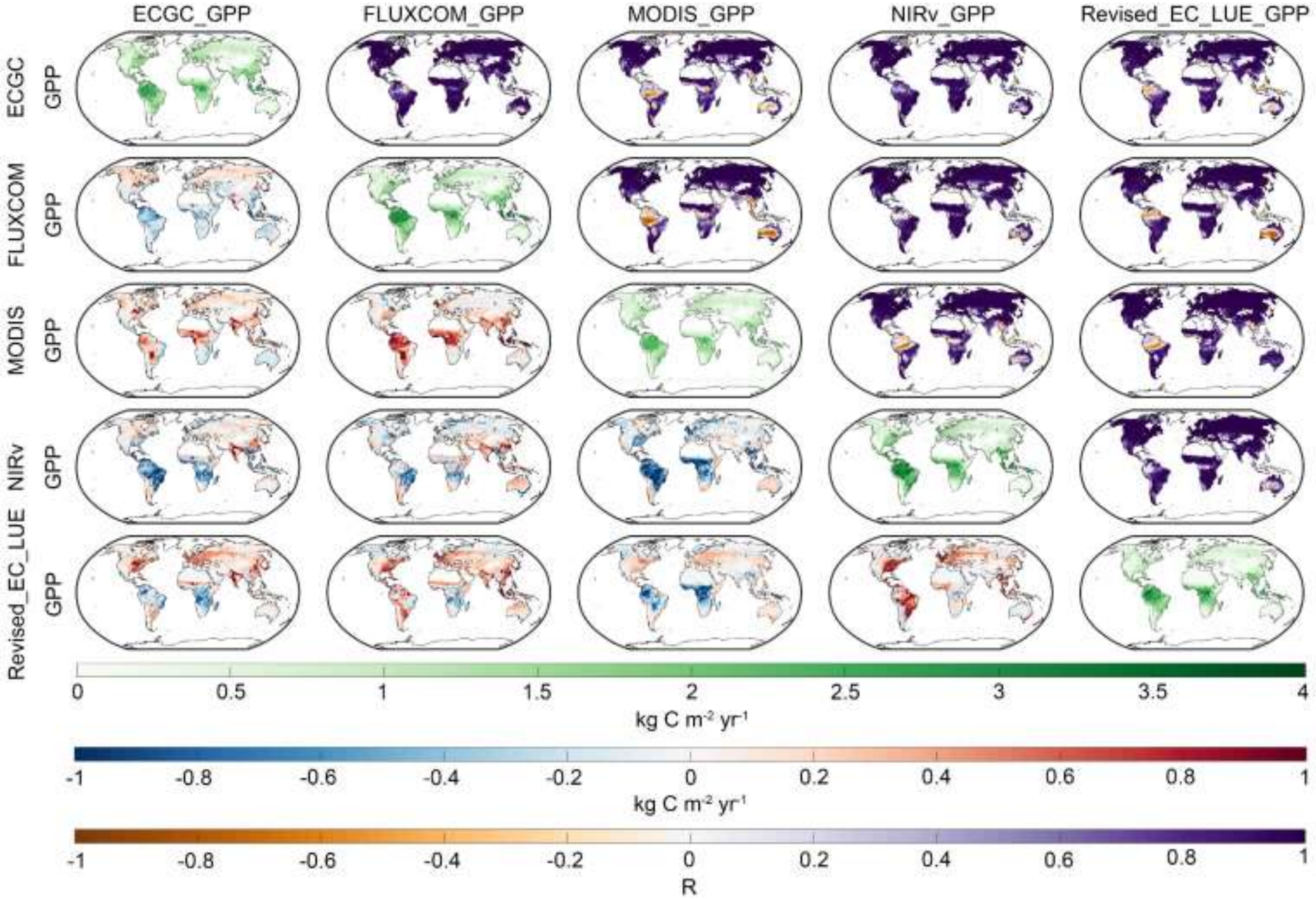
Datasets	Variables	Spatial resolutions	Temporal resolutions	Temporal coverage
FLUXNET2015	GPP, SSRD, T2M, TP, VPD	site scale	monthly	site-specific
MCD12C1	land cover	0.05 degree	yearly	2001 – 2019
Harvested Area and Yield for 175 Crops year 2000	corn, corn forage, sorghum, sorghum forage, millet, sugarcane	10 kilometers	yearly	2000
ERA5_Land	SSRD, T2M, TP, VPD	0.1 degree	monthly	1999 – 2019
GEOV2	LAI	1 kilometer	10 days	1999 – 2019
MCD15A2H	LAI	500 meters	8 days	2003 – 2019
MOD17A2H	GPP	500 meters	8 days	2001 – 2020
FLUXCOM_GPP	GPP	0.5 degree	monthly	1999 – 2013
Revised_EC_LUE_GPP	GPP	0.05 degree	8 days	1999 – 2018
NIRv_GPP	GPP	0.05 degree	monthly	1999 – 2018



Averaged annual value of ECGC_GPP from 1999 to 2019.

Compare with other GPP datasets

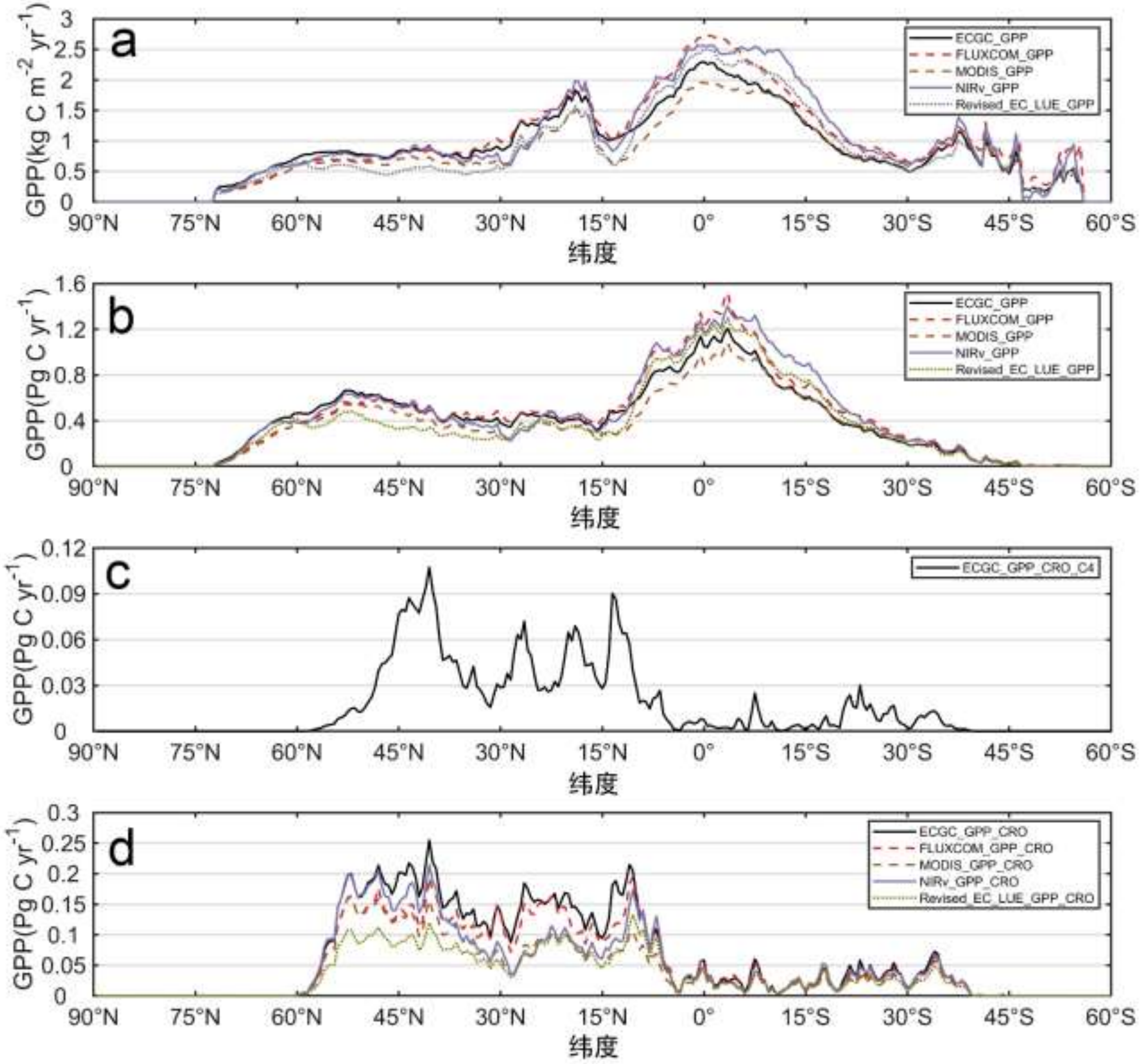
NUIST



ECGC_GPP has the highest correlation with FLUXCOM_GPP

Spatial comparison of 5 sets of GPP datasets from 2001-2013. The diagonal line is the mean distribution of each data set. Below the diagonal is the difference between the two data sets, and the horizontal minus vertical is used. On the diagonal is the spatial distribution of the correlation coefficient of the pairwise data sets.

Compare with other GPP datasets

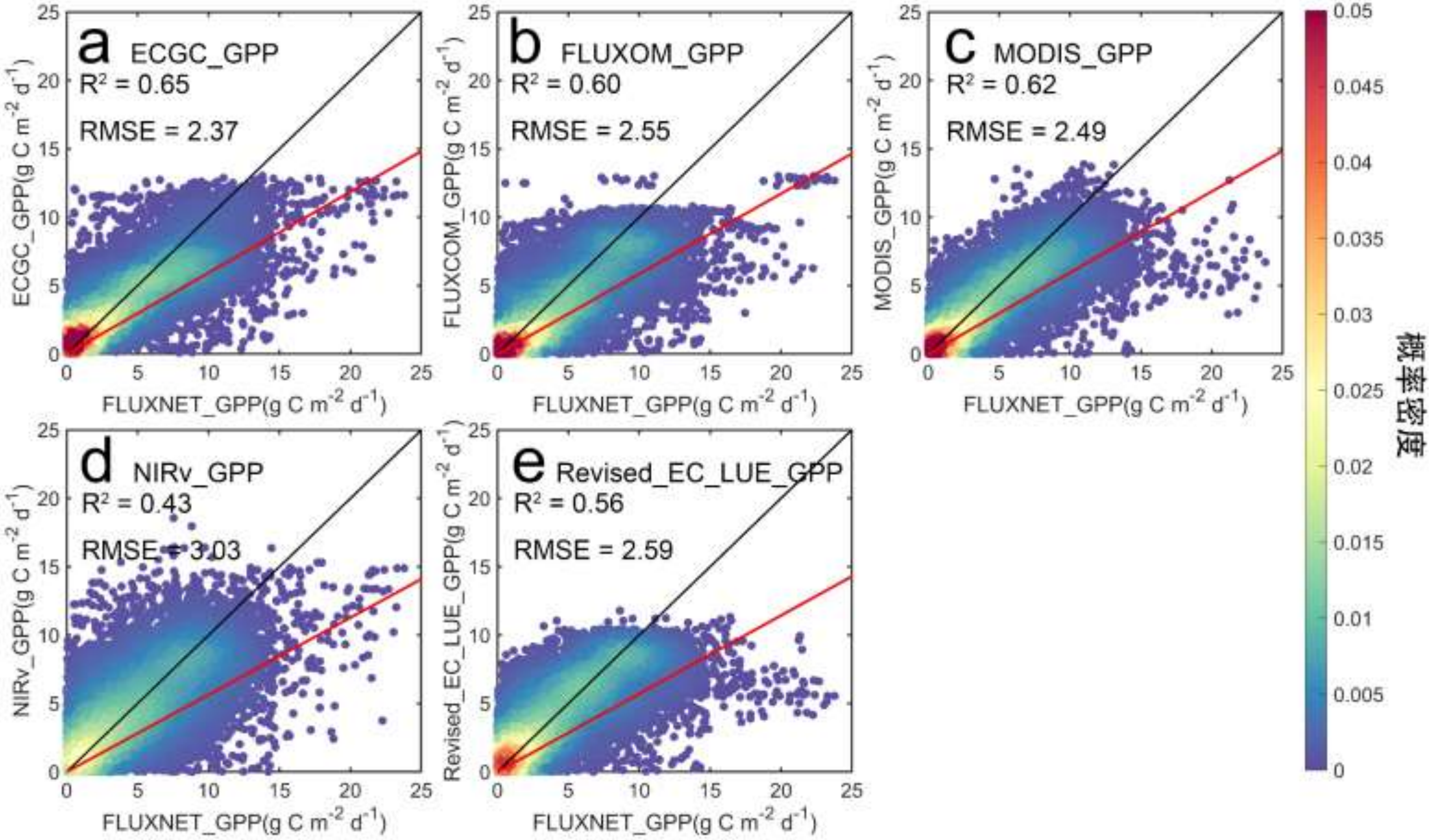


Global zonal comparison of five GPP datasets from 2001 to 2013

GPP gradually increases from dry and cold biomes (deserts and tundra) to warm and wet biomes (temperate and tropical forests). Equatorial and temperate regions show peaks. Hot and dry tropical regions show low GPP values.

After distinguishing C3 and C4 farmland, there is a significant increase in CRO_GPP at 10-45 degrees north latitude

Compare with other GPP datasets

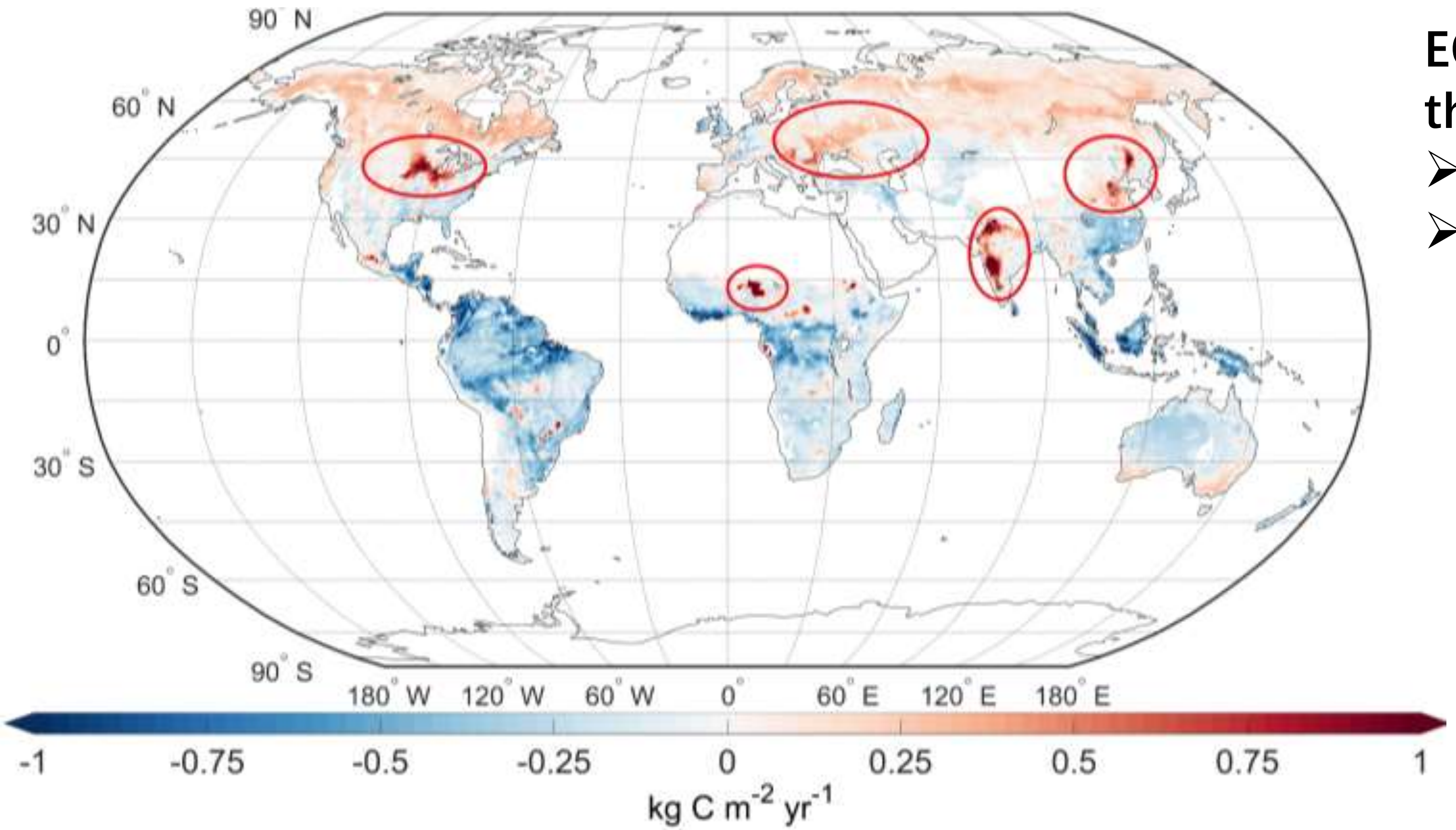


Site-level gross primary production (GPP) value comparisons between FLUXNET_GPP and five GPP data sets.

Compared with FLUXCOM_GPP, the R2 of ECGC_GPP has increased by 0.05, reaching the highest value of 0.65 among the five sets of GPP data.
RMSE has decreased by 0.18 g C m^{-2} d^{-1} reaching the lowest value of 2.37 g C m^{-2} d^{-1}。

Compare with other GPP datasets

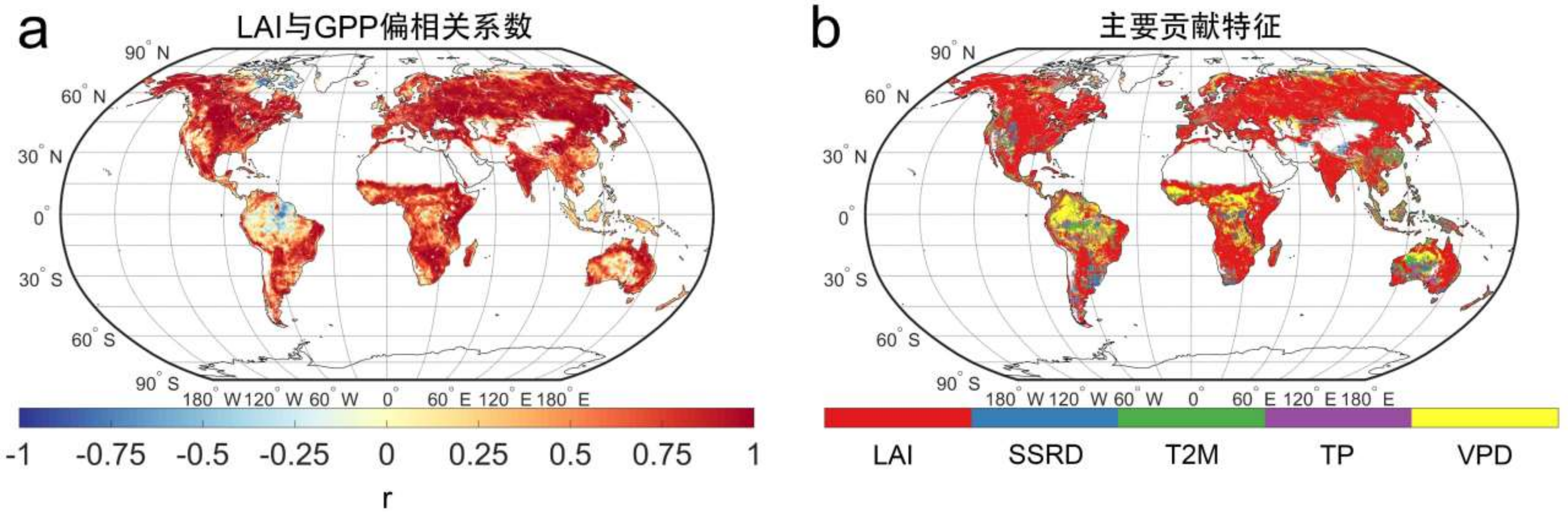
NUIST



The spatial difference maps for ECGC_GPP and FLUXCOM_GPP data sets from 2001 to 2013 (ECGC_GPP minus FLUXCOM_GPP).

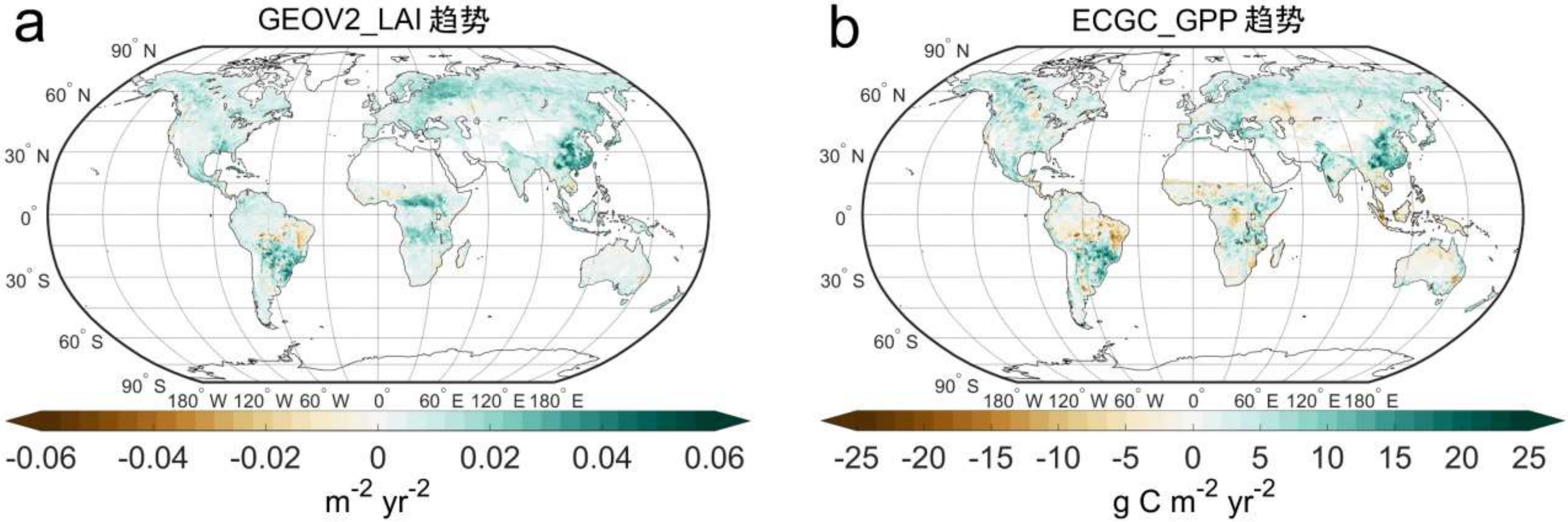
ECGC_GPP shows a higher GPP in the croplands than FLUXCOM.

- 76.38% cropland grids increased
- total cropland GPP increased by 18.68%



By calculating the spatial partial correlation coefficient of each feature and GPP, and get the main contribution space map. It can be seen that LAI estimates the main driving variables in the GPP model for the global PFT.

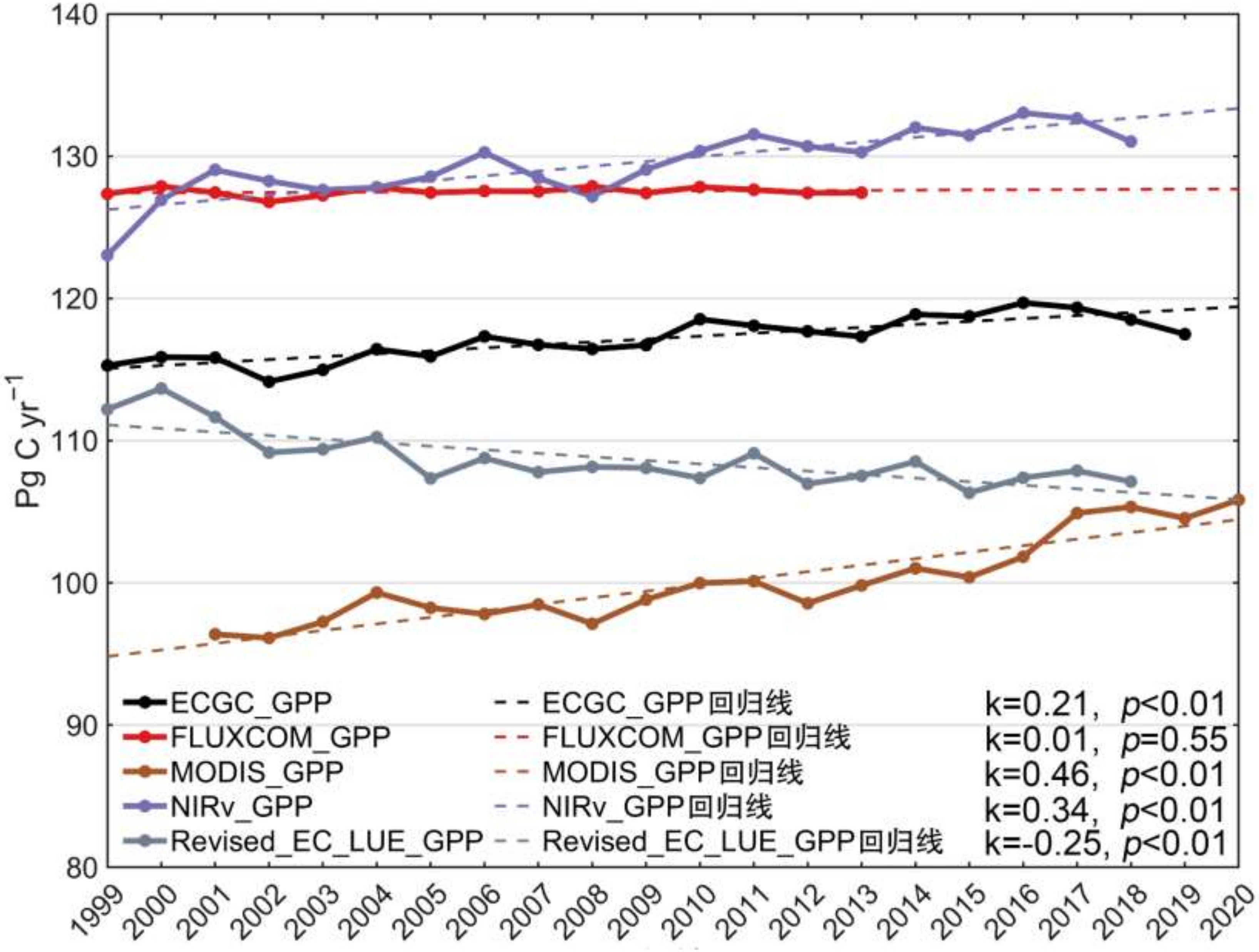
LAI contributions



In the PFT training model, LAI is the feature with the largest relative contribution, and the CO2 fertilization effect has a certain representation in LAI. Therefore, by comparing the global annual change trends of LAI and GPP, it can be found that the global annual change trends of LAI and GPP are highly consistent.

Long-term trends of GPP

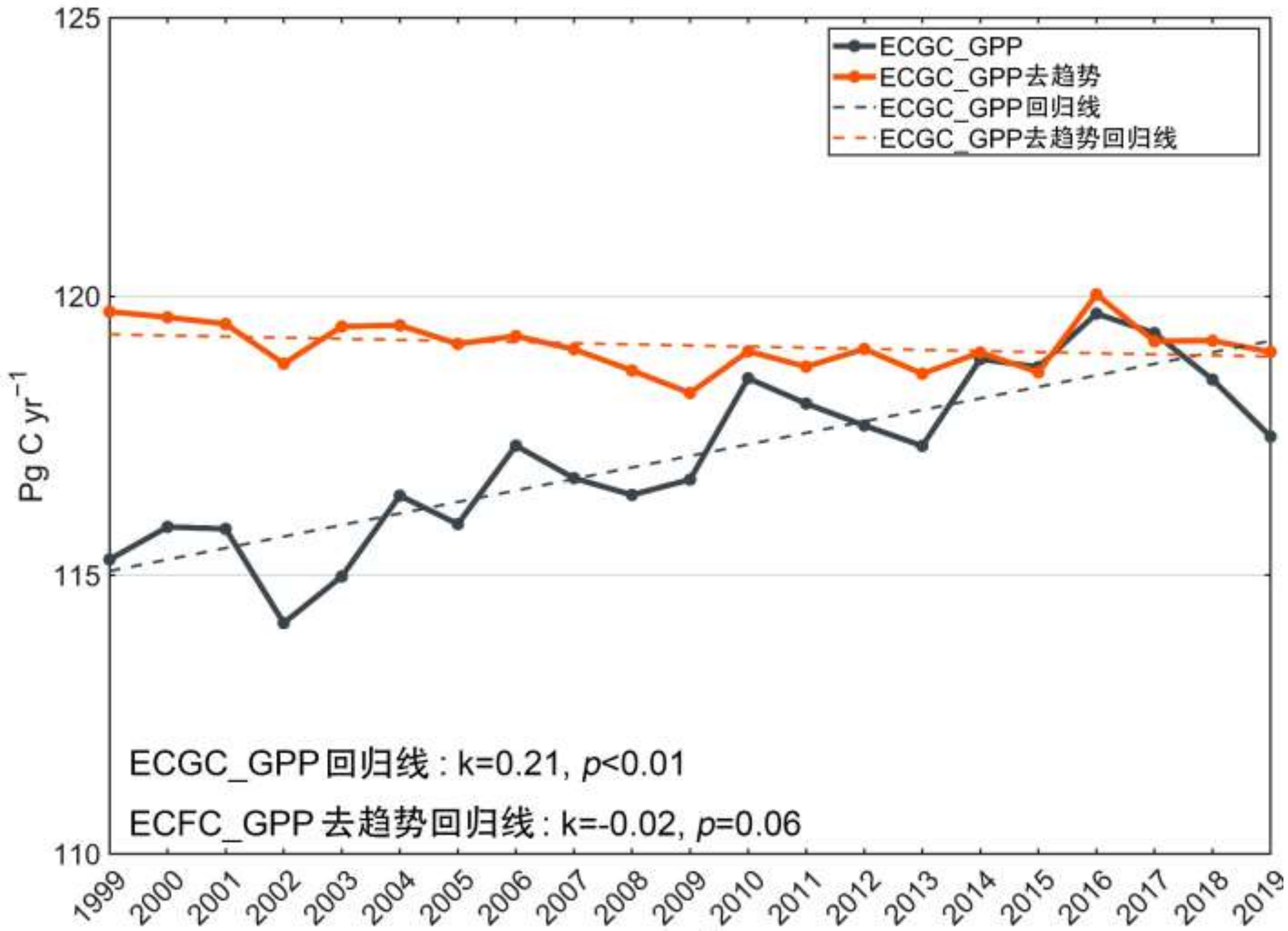
NUIST



The annual total of ECGC_GPP is 117.14 ± 1.51 Pg C yr^{-1}

The annual total of ECGC_GPP shows an upward trend 0.21 Pg C yr^{-2}

Compared to FLUXCOM_GPP, ECGC_GPP increased by 0.20 Pg C yr^{-2}



If LAI were detrended, the trend of global GPP was reduced from 0.21 Pg C yr^{-2} to -0.02 Pg C yr^{-2}.

Biogeosciences
An Interactive Open Access Journal of the European Geosciences Union
| EGU.eu |

Biogeosciences, 11, 3871-3880, 2014
www.biogeosciences.net/11/3871/2014/
doi:10.5194/bg-11-3871-2014

Article Metrics

Global cropland monthly gross primary production in the year 2000

T. Chen[1,2,3], G. R. van der Werf[1], N. Gobron[4], E. J. Moors[5], and A. J. Dolman[1]
[1]Department of Earth Sciences, Faculty of Earth and Life Sciences, VU University Amsterdam, the Netherlands
[2]International Center for Ecology, Meteorology and Environment (IceMe), School of Applied Meteorology, Nanjing Information Science and Technology, Nanjing, China

AGU ADVANCING EARTH AND SPACE SCIENC

JGR Biogeosciences

RESEARCH ARTICLE
10.1029/2022JG007100

Special Section:
Understanding carbon-climate feedbacks

Key Points:

- The accuracy of gross primary production (GPP) estimation can be improved by distinguishing plant functional types, especially for C3 and C4 crops
- Significant increasing trend is found in this random forest-based data set
- Leaf area index plays a leading role in both the average state and long-term trend of GPP

Supporting Information:
Supporting Information may be found in the online version of this article.

Correspondence to:
T. Chen,
txchen@nuist.edu.cn

Estimating Global GPP From the Plant Functional Type Perspective Using a Machine Learning Approach

Renjie Guo[1], Tiexi Chen[1,2,3], Xin Chen[1], Wenping Yuan[4], Shuci Liu[5], Bin He[6], Lin Li[7,8], Shengzhen Wang[2,3], Ting Hu[9], Qingyun Yan[9], Xueqiong Wei[1], and Jie Dai[1]

[1]School of Geographical Sciences, Nanjing University of Information Science and Technology, Nanjing, China, [2]School of Geographical Sciences, Qinghai Normal University, Xining, China, [3]School of Ecology and Environmental Science, Xining University, Xining, China, [4]School of Atmospheric Sciences, Sun Yat-sen University, Zhuhai, China, [5]Department of Environment and Science, Queensland Government, Brisbane, Australia, [6]College of Global Change and Earth System Science, Beijing Normal University, Beijing, China, [7]Computer Department, Qinghai Normal University, Xining, China, [8]Academy of Plateau Science and Sustainability, Qinghai Normal University, Xining, China, [9]School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing, China

Abstract The long-term monitoring of gross primary production (GPP) is crucial to the assessment of the carbon cycle of terrestrial ecosystems. In this study, a well-known machine learning model (random forest, RF) is established to reconstruct the global GPP data set named ECGC_GPP. The model distinguished nine functional plant types, including C3 and C4 crops, using eddy fluxes, meteorological variables, and leaf area index (LAI) as training data of RF model. Based on ERA5_Land and the corrected GEOV2 data, global monthly GPP data set at a 0.05° resolution from 1999 to 2019 was estimated. The results showed that the RF model could explain 74.81% of the monthly variation of GPP in the testing data set, of which the average contribution of LAI reached 41.73%. The average annual and standard deviation of GPP during 1999–2019 were 117.14 ± 1.51 Pg C yr^{-1}, with an upward trend of 0.21 Pg C yr^{-2} ($p < 0.01$). By using the plant functional type classification, the underestimation of cropland is improved. Therefore, ECGC_GPP provides reasonable

DRYAD

Who we a

Estimating global GPP from the plant functional type perspective using a machine learning approach

Guo, Renjie, Nanjing University of Information Science and Technology

Chen, Tiexi, Nanjing University of Information Science and Technology, https://orcid.org/0000-0002-2761-8703

Chen, Xin, Nanjing University of Information Science and Technology

Yuan, Wenping, Sun Yat-sen University, https://orcid.org/0000-0002-1469-4395

Liu, Shuci, Queensland Department of Environment and Science, https://orcid.org/0000-0002-5629-1703

He, Bin, Beijing Normal University

Li, Lin, Qinghai Normal University

Wang, Shengzhen, Qinghai Normal University

Hu, Ting, Nanjing University of Information Science and Technology

Yan, Qingyun, Nanjing University of Information Science and Technology

Wei, Xueqiong, Nanjing University of Information Science and Technology

Dai, Jie, Nanjing University of Information Science and Technology

rjguo1998@163.com, txchen@nuist.edu.cn

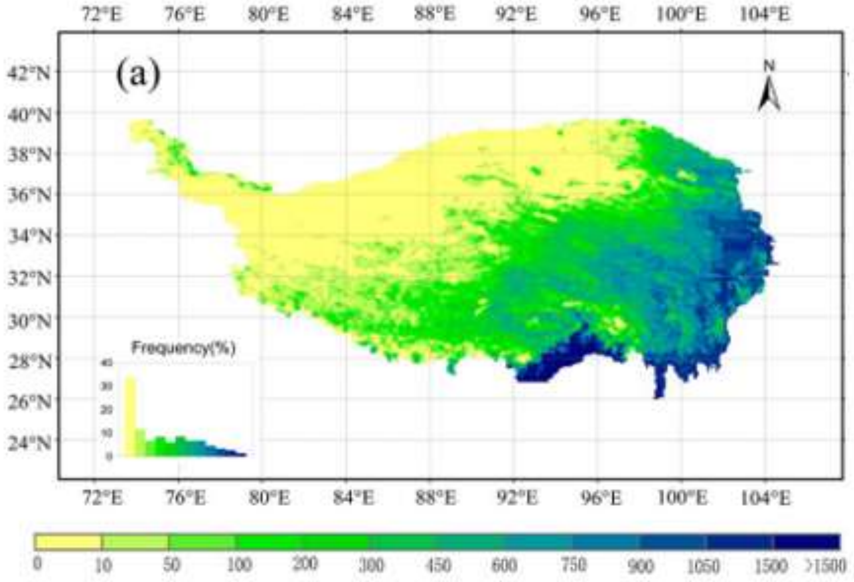
Publication date: April 18, 2023

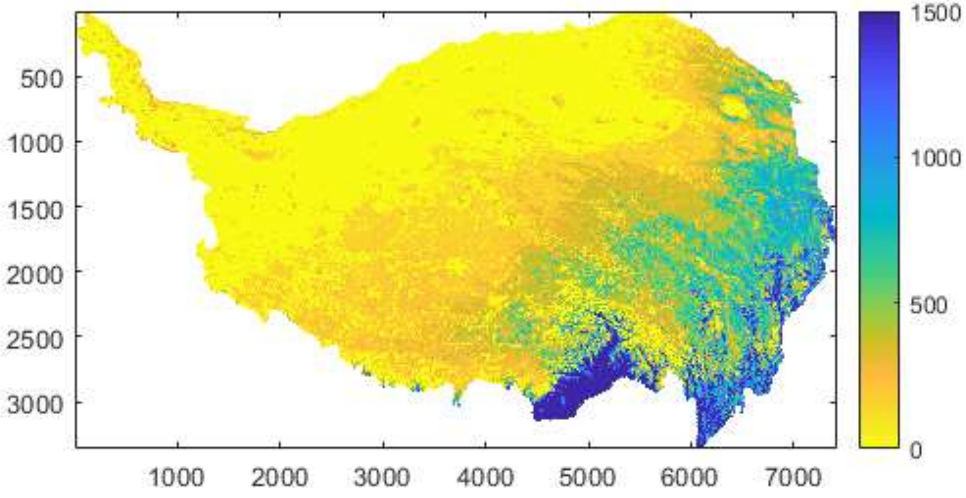
Publisher: Dryad

https://doi.org/10.5061/dryad.dncjsxm2v

https://datadryad.org/stash/dataset/doi:10.5061/dryad.dncjsxm2v

Idea: **Improve global estimates by continuously optimizing regional-scale estimates**





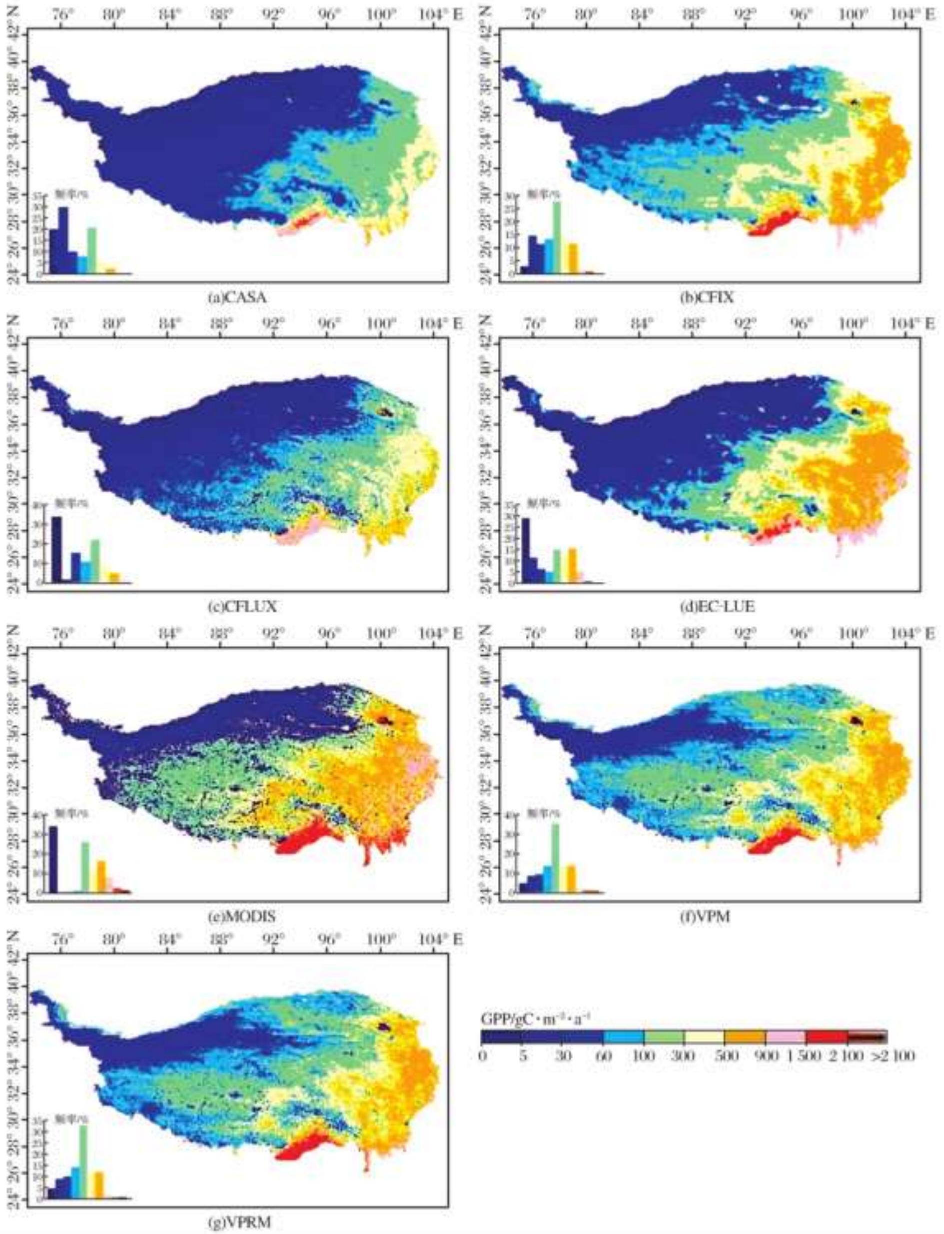


图 4 光能利用率模型模拟的青藏高原 2000～2010 年 GPP 年均值的空间分布图

感谢各位垂听

敬请各位专家批评指正~