Various responses of global heterotrophic respiration to variations in soil moisture and temperature enhance the positive feedback on atmospheric warming

Jingyu Zeng[1, 2], Tao Zhou[1,2*], Li Cao[1,2], Ying Yu[1,2], E Tan[1,2], Yajie Zhang[1,2], Xuemei Wu[1,2], Jingzhou Zhang[1,2], Qi Zhang[1,2], Yancheng Qu[1,2], Jian Liu[1,2], Peixia Liu[1,2], Xia Liu[1,2], Hui Luo[1,2]

1 Key Laboratory of Environmental Change and Natural Disaster of Ministry of Education, Beijing Normal University, Beijing 100875, China

2 State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing, 100875, China

Correspondence: Tao Zhou (tzhou@bnu.edu.cn)

Table. S1: Input variables of R_s and R_h random forest models.

Model	Input variables
RF_R_s	MAP, MAT, DSR, STL2, SWVL2, VHI, GPP
RF_R_h	MAT, ST, SM, VHI, GPP, RICHNESS, R_s

Among them, MAP is the mean annual precipitation, MAT is the mean annual temperature, DSR is the downward shortwave radiation, ST is the soil temperature, and SM is the soil moisture. VHI is the vegetation health index, GPP is the gross primary productivity, RICHNESS is the species richness, and R_s is the soil respiration obtained from random forests. The relationship between the R_s and R_h and independent variables is estimated through the random forest model established in R.

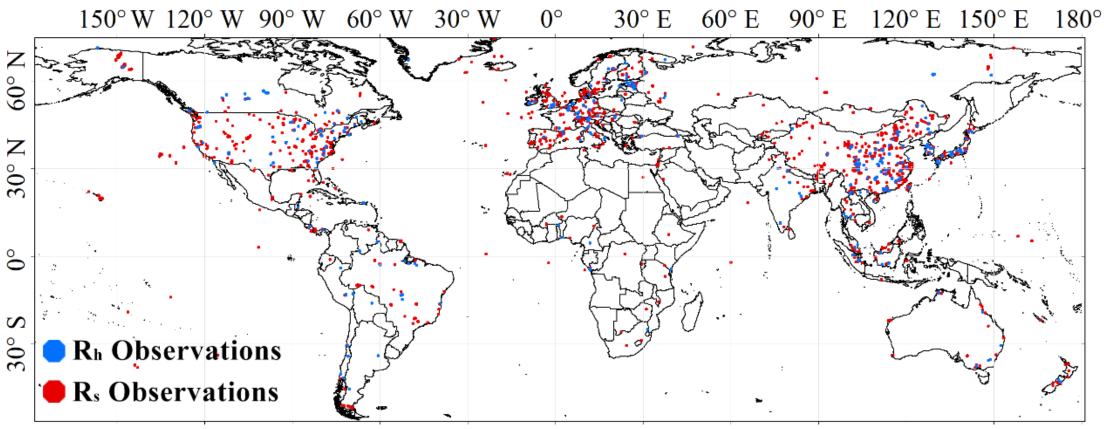

Fig. S1. Distribution of the soil respiration and heterotrophic respiration observation.

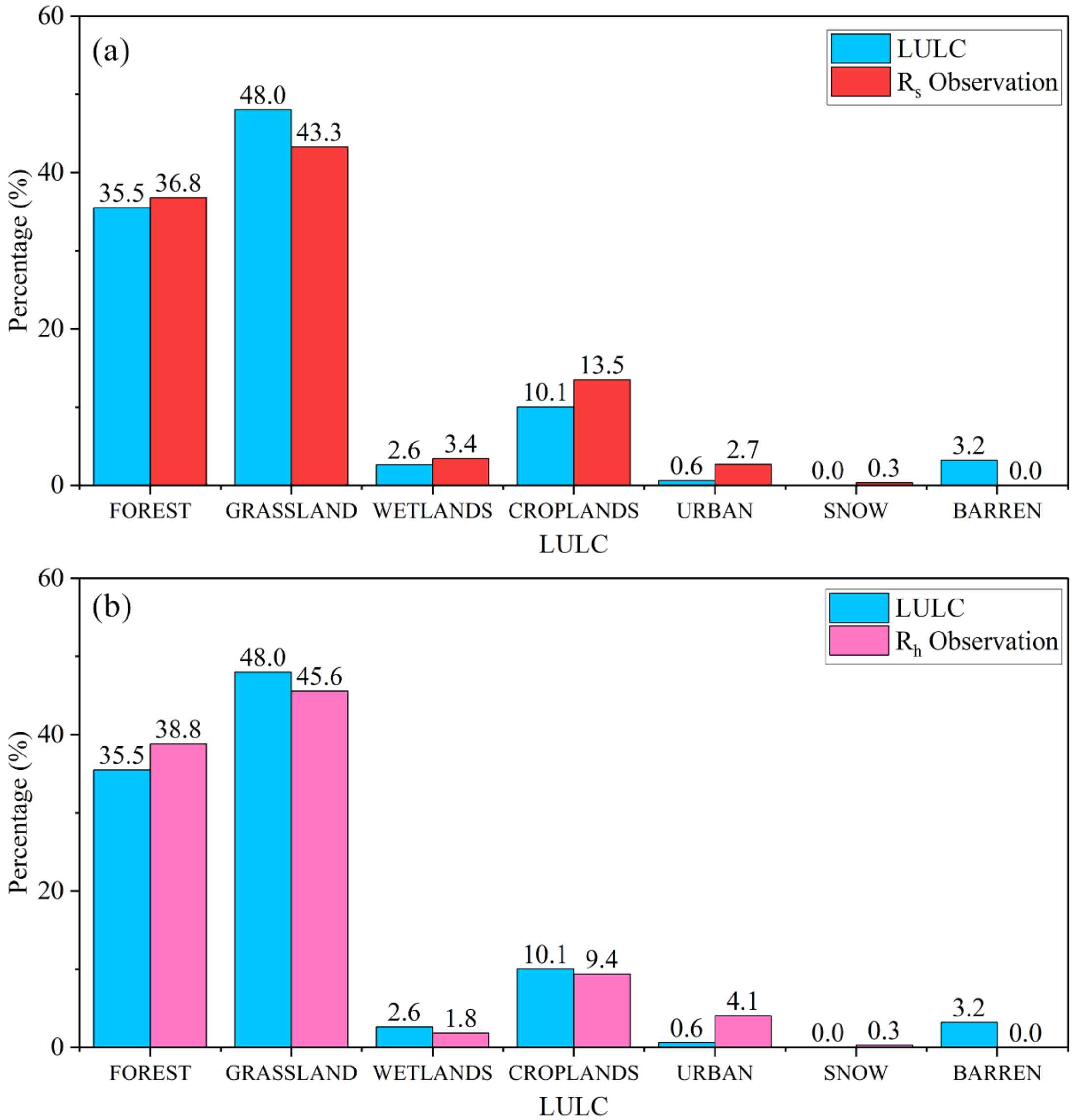

Fig. S2. Evaluation of spatial representativeness of observational data. Comparing the proportion of observed data in different ecosystem types using MODIS land cover type products. (a)R_s, (b)R_h, LULC: Land Use and Land Cover.

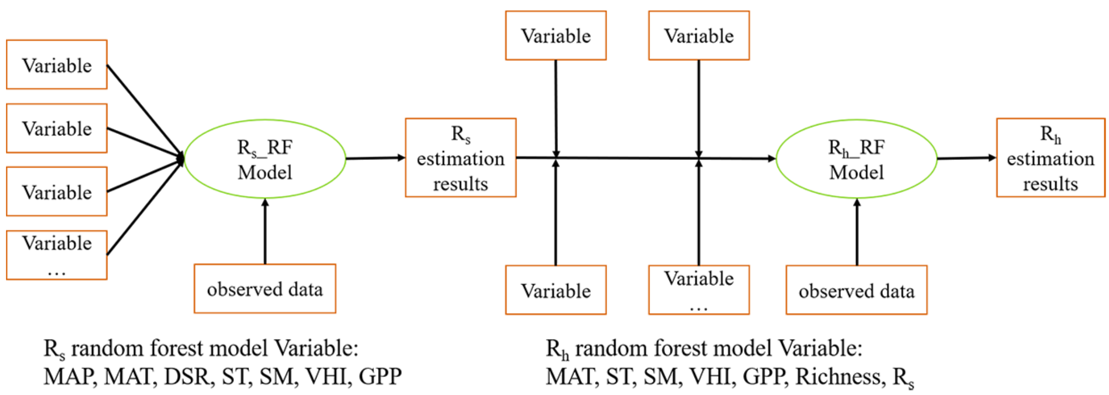

Fig. S3. Estimation of R_h using a two-stage random forest modeling framework.

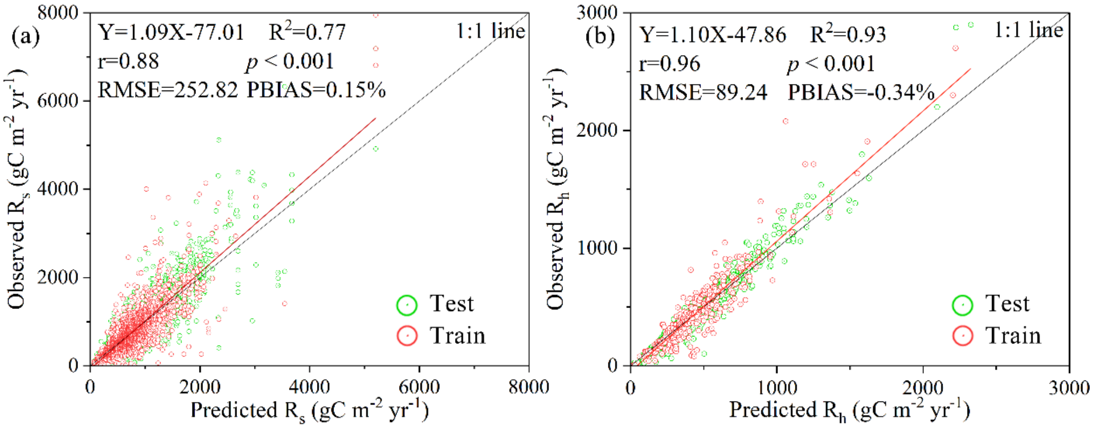

Fig. S4. Performance of the soil respiration and heterotrophic respiration random forest model. (a) R_s random forest model, (b) R_h random forest model.

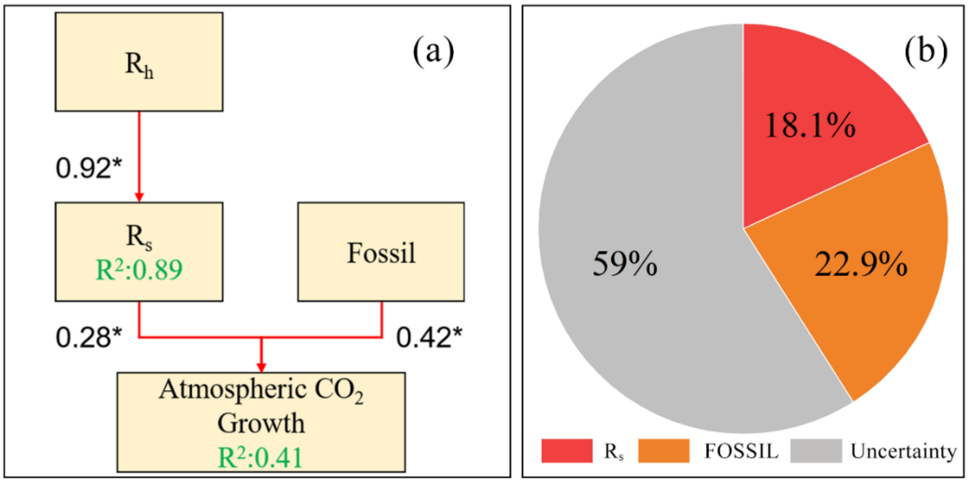

Fig. S5. Quantification of the impact of soil respiration and energy emissions on atmospheric CO_2 growth. (a) Structural equation modeling evaluates the relationship between R_h, R_s, Fossil and atmospheric CO_2 growth, Chisq/df=0.172; CFI=0.996; RMSEA=0.003; SRMR=0.009; *: p<0.05; (b) Variance decomposition quantifies the contributions of Rs, Fossil, and other factors to atmospheric CO_2 growth variation.

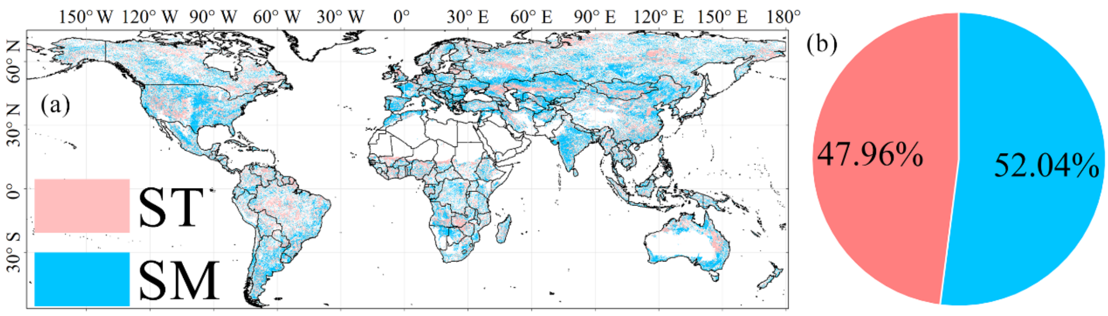

Fig. S6. The main driving factors of global terrestrial R_h/R_s (soil temperature and soil moisture). Exclude grid points that did not pass the significance test through partial correlation analysis.

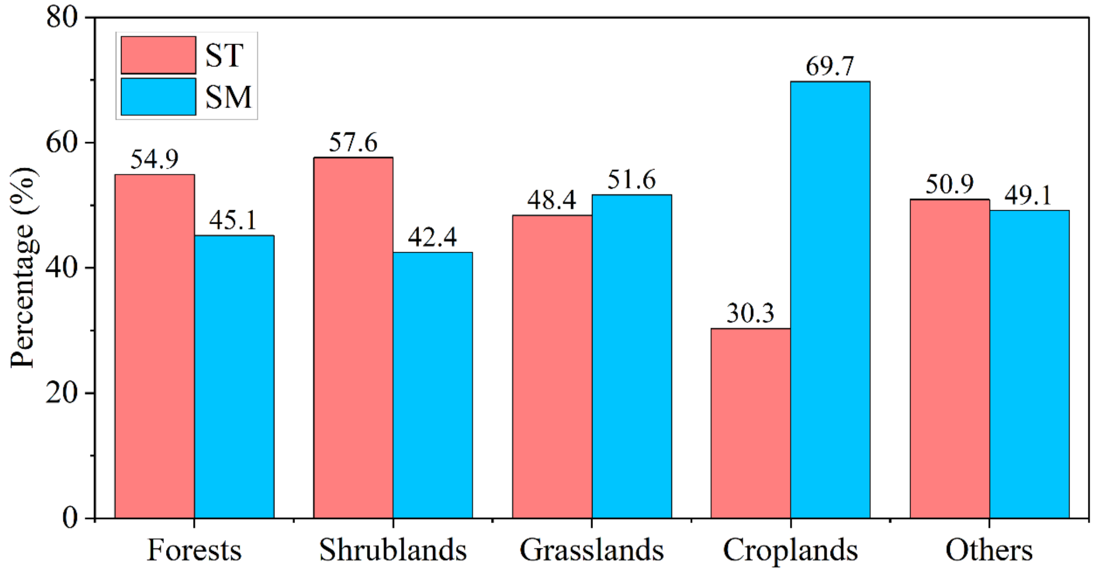

Fig. S7. The main driving factors of global terrestrial R_h/R_s (soil temperature and soil moisture) in various ecosystem types.

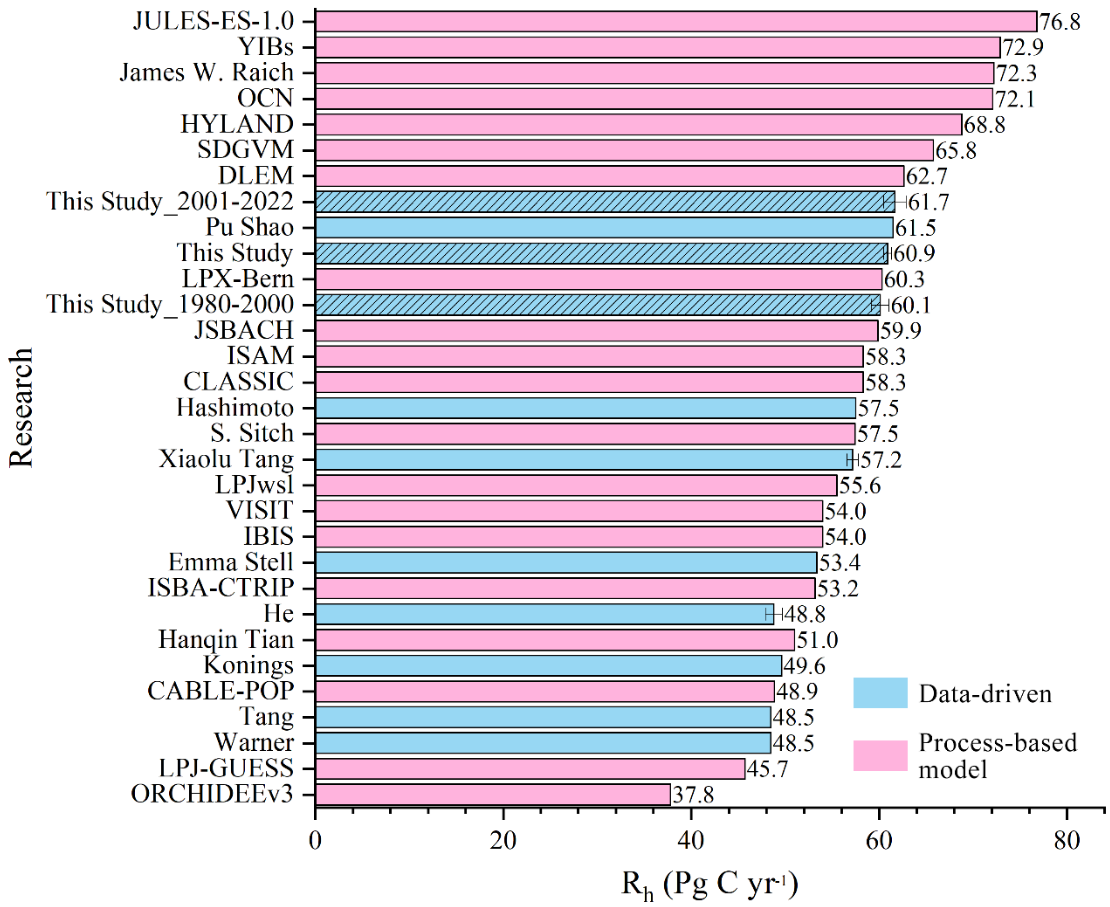

Fig. S8. Comparison of global terrestrial R_h estimated by this study and different methods and models ([1-28]).

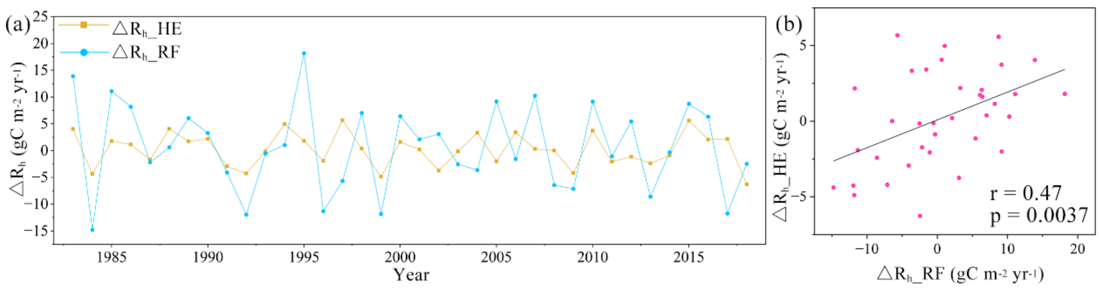

Fig. S9. Consistency in temporal variation between the R_h data from He et al [5]. and the R_h estimated in this study. (a) The interannual variation of the two; (b) The scatter relationship between the two. He et al. estimated the global terrestrial ecosystem R_h directly on the basis of 761 observational data points combined with environmental variables.

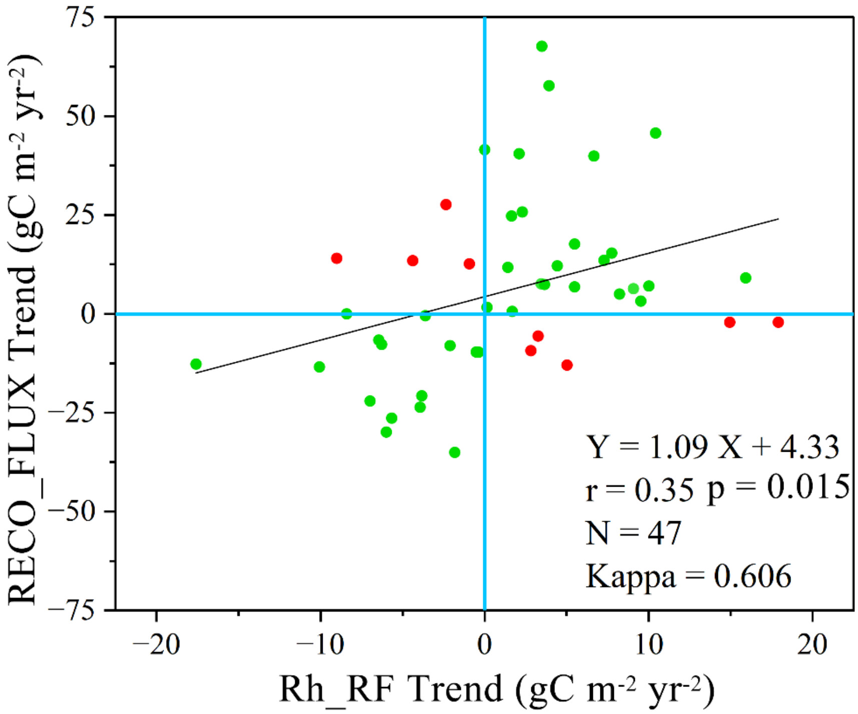

Fig. S10. Consistency in trends between the observed RECO from 47 flux stations and the estimated R_h in this study.

Supplementary Methods

Observation data collection standards and quality control methods

1. The collected R_s and R_h data must include accurate latitude and longitude information as well as specific or time range information. If the location description of the data is too vague (e.g., located in a certain region or county), it will not be included in the database of this study.

2. When the latitude, longitude, and sampling time of site data are completely consistent, we will merge the data using the arithmetic mean method.

3. For controlled experimental studies (e.g., warming, humidification, and nitrogen fertilizer application), we only extract and retain data from the control group (i.e., under natural conditions without human interference).

4. For observational data providing environmental information, we retain the environmental information (e.g., temperature, precipitation). When the study only provides carbon flux and its spatiotemporal information, we match the data with corresponding spatial grid data based on latitude, longitude, and sampling year.

5. To eliminate the influence of outliers, we exclude the top and bottom 2.5% of the data when building the model, retaining 95% of the data for model construction.

6. One of our criteria for data collection is that the data sources must be inventory data published by government units, databases published by scientific research institutions, and data from peer-reviewed research papers.

7. If the data from a paper does not provide a specific sampling time, the sampling year is considered as the publication year of the paper. If the data sampling year is a certain period, the environmental information (temperature, precipitation, etc.) of the data is taken as the average value of that period.

8. For papers that do not provide specific data but contain data values in images, we extract the numerical values through digital image scanning.

9. For R_s and R_h data, if the provided value is in mol units, it is converted to the same mass unit by the molar mass fraction.

10. For data that only provide R_s and the ratio of autotrophic respiration to heterotrophic respiration, heterotrophic respiration is calculated based on the ratio provided in the

literature.

11. Multi-year continuous measurements at the same location are considered independent, i.e., each year's R_h data is treated as an observational data.

12. For observation data from the same database (e.g., SRDB V5.0), we will record the name of the database and the author of the data source. For observation data from different sources of literature, we will record the names of the literature and their authors.

FLUXNET2015

For over three decades, the eddy covariance technique has been used to measure land-atmosphere exchanges of greenhouse gases and energy at sites around the world to study and determine the function and trajectories of both ecosystems and the climate system. The technique allows nondestructive measurement of these fluxes at a high temporal resolution and ecosystem level, making it a unique tool. Based on high frequency (10–20 Hz) measurements of vertical wind velocity and a scalar (CO_2, H_2O, temperature, etc.), it provides estimates of the net exchange of the scalar over a source footprint area that extends up to hundreds of meters around the point of measurement. The FLUXNET2015 dataset provides ecosystem-scale data on CO_2, water, and energy exchange between the biosphere and the atmosphere, and other meteorological and biological measurements, from 212 sites around the globe (over 1500 site-years, up to and including year 2014) (https://fluxnet.org/data/download-data/). Data were quality controlled and processed using uniform methods, to improve consistency and intercomparability across sites. The dataset is already being used in a number of applications, including ecophysiology studies, remote sensing studies, and development of ecosystem and Earth system models. This study will select flux observation data from FLUXNET2015 that have been continuously observed for more than 10 points and match our estimated spatial position of global terrestrial ecosystem R_h, in order to validate and evaluate the reliability of our estimated temporal variation of R_h (Table S2). Table S2 can be obtained from the following link: https://figshare.com/s/871963d1e851846931d8

we combined the RECO data from flux tower observations (FLUXNET 2015) to compare our estimated global terrestrial ecosystem R_h. Specifically, we matched the flux tower's latitude and longitude and the observation data's time (we chose annual data) to our estimated R_h data, ensuring spatial location and temporal matching during the comparison.

R_h products produced by He

This study will use the global terrestrial ecosystem heterotrophic respiration dataset published by He et al [5]. to evaluate the consistency and reliability of our estimated R_h from a temporal perspective. This dataset provides a global gridded product at 0.5-degree resolution of predicted annual soil heterotrophic respiration during 1982–2018 (https://datadryad.org/stash/dataset/doi:10.5061/dryad.b2rbnzsj9). A Random Forest approach was used to derive the predicted R_h trained with 761 observations with 19 predictors. To improve the RF model accuracy, a stratified 10-fold cross-validation was used, by grouping 761 observations into three climate zone classes (i.e., tropical, temperate and boreal zones) and ensuring each class was approximately equally represented across each fold. The average predicted map across the RF model ensemble was used as the final product.

Atmospheric CO_2 concentration from NOAA Global Monitoring Laboratory

The Global Monitoring Laboratory has measured carbon dioxide and other greenhouse gases for several decades at a globally distributed network of air sampling sites (https://gml.noaa.gov/ccgg/trends/global.html). A global average is constructed by first fitting a smoothed curve as a function of time to each site, and then the smoothed value for each site is plotted as a function of latitude for 48 equal time steps per year. A global average is calculated from the latitude plot at each time step. This study will select the global atmospheric CO_2 concentration observation data released by NOAA Global Monitoring Laboratory to compare and evaluate the reliability of our estimated global terrestrial ecosystem R_s and R_h from a temporal perspective (Table S3).

For the atmospheric observations of CO_2 concentration changes, we treat the globe as a complete system. The carbon changes in terrestrial ecosystems show good consistency with those observed in the atmosphere on an annual scale. Therefore, we compared the estimated annual R_s, R_h and R_a of the global terrestrial ecosystem with the atmospheric observations of carbon content changes from 1980 to 2022, ensuring temporal matching while avoiding spatial mismatch issues since the comparison is conducted at a global scale.

Table. S3: Global Annual Atmospheric Carbon Dioxide Content Observed by NOAA Global Monitoring Laboratory

year	mean	unc	year	mean	unc	year	mean	unc
1979	336.85	0.11	1994	358.33	0.08	2009	386.5	0.04
1980	338.91	0.07	1995	360.17	0.05	2010	388.75	0.06
1981	340.11	0.09	1996	361.93	0.04	2011	390.62	0.05
1982	340.86	0.03	1997	363.05	0.05	2012	392.65	0.06
1983	342.53	0.06	1998	365.7	0.04	2013	395.4	0.06
1984	344.07	0.08	1999	367.8	0.05	2014	397.34	0.05
1985	345.54	0.07	2000	368.96	0.06	2015	399.65	0.05
1986	346.97	0.07	2001	370.57	0.05	2016	403.06	0.06
1987	348.68	0.1	2002	372.58	0.04	2017	405.22	0.07
1988	351.16	0.07	2003	375.15	0.04	2018	407.61	0.07
1989	352.79	0.07	2004	376.95	0.06	2019	410.07	0.07
1990	354.05	0.07	2005	378.98	0.05	2020	412.44	0.06
1991	355.39	0.07	2006	381.15	0.05	2021	414.7	0.07
1992	356.09	0.06	2007	382.9	0.04	2022	417.08	0.07
1993	356.83	0.07	2008	385.02	0.05	2023	419.32	0.15

The specific calculation process of Mann–Kendell method

The null hypothesis (H_0) shows the data series x_k (k = 1, 2, 3..., n) as a sample of n independently and identically distributed; whereas the alternative hypothesis (H_1) is that there has been an increasing or decreasing trend in the data series. The Mann–Kendall test statistics are calculated as follows:

$$S = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \mathrm{sgn}(x_j - x_i), \tag{1}$$

where xj is the sequential data value, n is the length of the data set, and sgn is calculated as follows:

$$\mathrm{sgn}(x_j - x_i) = \begin{cases} 1 & if x_j > x_i \\ 0 & if x_j = x_i \\ -1 & if x_j < x_i \end{cases} \tag{2}$$

According to Mann and Kendall, when $n \geq 8$, the test statistics (S) are typically distributed with the mean and variance as follows:

$$E(S) = 0, \tag{3}$$

$$V(S) = [n(n-1)(2n+5)] - \sum_{m=1}^{n} m(m-1)(2m+5)/18, \tag{4}$$

where m is the total number of variables with continuous and identical values. The standardized test statistic (Z) is calculated using the following equation:

$$Z = \begin{cases} \frac{S-1}{\sqrt{V(S)}} & S > 0 \\ 0 & S = 0 \\ \frac{S+1}{\sqrt{V(S)}} & S < 0 \end{cases} \tag{5}$$

$|Z\alpha|$ = 1.65, 1.96, and 2.58, which correspond with the critical values at the significance level p = 0.1, 0.05, and 0.01, respectively. If $|Z| > |Z\alpha|$, H_0 is rejected and p = 0.05 and 0.01 are considered in our research.

To analyze the trend of the time series variable, we used the robust estimator for the amplitude of trend slopes:

$$\mathrm{Slope} = \mathrm{Median}(\mathrm{x_j} - \mathrm{x_i})/(\mathrm{j} - \mathrm{i})(1 \leq \mathrm{i} < j \leq n), \tag{6}$$

where slope is the rate of change of the entire data series x_k (k = 1, 2, 3..., n). If it is positive, the series increases monotonically; otherwise, the series decreases monotonically. Median represents the function that takes the median value.

Supplementary References:

1 Delire, C. *et al.* The Global Land Carbon Cycle Simulated With ISBA-CTRIP: Improvements Over the Last Decade. *Journal of Advances in Modeling Earth Systems* **12**, e2019MS001886 doi:10.1029/2019ms001886 (2020).

2 Friend, A. D., Stevens, A. K., Knox, R. G. & Cannell, M. G. R. A process-based, terrestrial biosphere model of ecosystem dynamics (Hybrid v3.0). *Ecological Modelling* **95**, 249-287, doi:10.1016/S0304-3800(96)00034-8 (1997).

3 Hashimoto, S. *et al.* Global spatiotemporal distribution of soil respiration modeled using a global database. *Biogeosciences* **12**, 4121-4132, doi:10.5194/bg-12-4121-2015 (2015).

4 Haverd, V. *et al.* A new version of the CABLE land surface model (Subversion revision r4601) incorporating land use and land cover change, woody vegetation demography, and a novel optimisation-based approach to plant coordination of photosynthesis. *Geoscientific Model Development* **11**, 2995-3026, doi:10.5194/gmd-11-2995-2018 (2018).

5 He, Y. *et al.* Observation-based global soil heterotrophic respiration indicates underestimated turnover and sequestration of soil carbon by terrestrial ecosystem models. *Global Change Biology* **28**, 5547-5559, doi:10.1111/gcb.16286 (2022).

6 Kato, E., Kawamiya, M., Kinoshita, T. & Ito, A. in *Conference on Earth System Science (ESS) - Global Change, Climate and People.* 146-152.

7 Kato, E., Tsuguki, K., Akihiko, I., Michio, K. & and Yamagata, Y. Evaluation of spatially explicit emission scenario of land-use change and biomass burning using a process-based biogeochemical model. *Journal of Land Use Science* **8**, 104-122, doi:10.1080/1747423X.2011.628705 (2013).

8 Konings, A. G. *et al.* Global satellite-driven estimates of heterotrophic respiration. *Biogeosciences* **16**, 2269-2284, doi:10.5194/bg-16-2269-2019 (2019).

9 Lienert, S. & Joos, F. A Bayesian ensemble data assimilation to constrain model parameters and land-use carbon emissions. *Biogeosciences* **15**, 2909-2930, doi:10.5194/bg-15-2909-2018 (2018).

10 Mauritsen, T. *et al.* Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO2. *Journal of Advances in Modeling Earth Systems* **11**, 998-1038, doi:10.1029/2018ms001400 (2019).

11 Meiyappan, P., Jain, A. K. & House, J. I. Increased influence of nitrogen limitation on CO2 emissions from future land use and land use change. *Global Biogeochemical Cycles* **29**, 1524-1548, doi:10.1002/2015gb005086 (2015).

12 Melton, J. R. *et al.* CLASSIC v1.0: the open-source community successor to the Canadian Land Surface Scheme (CLASS) and the Canadian Terrestrial Ecosystem Model (CTEM) - Part 1: Model framework and site-level performance. *Geoscientific Model Development* **13**, 2825-2850, doi:10.5194/gmd-13-2825-2020 (2020).

13 Poulter, B., Frank, D. C., Hodson, E. L. & Zimmermann, N. E. Impacts of land cover and climate data selection on understanding terrestrial carbon dynamics and the CO2 airborne fraction. *Biogeosciences* **8**, 2027-2036, doi:10.5194/bg-8-2027-2011 (2011).

14 Raich, J. W. & Potter, C. S. GLOBAL PATTERNS OF CARBON-DIOXIDE EMISSIONS FROM SOILS. *Global Biogeochemical Cycles* **9**, 23-36, doi:10.1029/94GB02723 (1995).

15 Sellar, A. A. *et al.* UKESM1: Description and Evaluation of the UK Earth System Model. *Journal of Advances in Modeling Earth Systems* **11**, 4513-4558, doi:10.1029/2019ms001739 (2019).
16 Shao, P., Zeng, X. B., Moore, D. J. P. & Zeng, X. D. Soil microbial respiration from observations and Earth System Models. *Environmental Research Letters* **8**, 034034 doi:10.1088/1748-9326/8/3/034034 (2013).
17 Sitch, S. *et al.* Recent trends and drivers of regional sources and sinks of carbon dioxide. *Biogeosciences* **12**, 653-679, doi:10.5194/bg-12-653-2015 (2015).
18 Smith, B. *et al.* Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model. *Biogeosciences* **11**, 2027-2054, doi:10.5194/bg-11-2027-2014 (2014).
19 Stell, E., Warner, D., Jian, J. S., Bond-Lamberty, B. & Vargas, R. Spatial biases of information influence global estimates of soil respiration: How can we improve global predictions? *Global Change Biology* **27**, 3923-3938, doi:10.1111/gcb.15666 (2021).
20 Tang, X. L. *et al.* Spatial and temporal patterns of global soil heterotrophic respiration in terrestrial ecosystems. *Earth System Science Data* **12**, 1037-1051, doi:10.5194/essd-12-1037-2020 (2020).
21 Tian, H. Q. *et al.* North American terrestrial CO2 uptake largely offset by CH4 and N2O emissions: toward a full accounting of the greenhouse gas budget. *Climatic Change* **129**, 413-426, doi:10.1007/s10584-014-1072-9 (2015).
22 Tian, H. Q. *et al.* Global patterns and controls of soil organic carbon dynamics as simulated by multiple terrestrial biosphere models: Current status and future directions. *Global Biogeochemical Cycles* **29**, 775-792, doi:10.1002/2014gb005021 (2015).
23 Vuichard, N. *et al.* Accounting for carbon and nitrogen interactions in the global terrestrial ecosystem model ORCHIDEE (trunk version, rev 4999): multi-scale evaluation of gross primary production. *Geoscientific Model Development* **12**, 4751-4779, doi:10.5194/gmd-12-4751-2019 (2019).
24 Walker, A. P. *et al.* The impact of alternative trait-scaling hypotheses for the maximum photosynthetic carboxylation rate (Vcmax) on global gross primary production. *New Phytologist* **215**, 1370-1386, doi:10.1111/nph.14623 (2017).
25 Warner, D. L., Bond-Lamberty, B., Jian, J., Stell, E. & Vargas, R. Spatial Predictions and Associated Uncertainty of Annual Soil Respiration at the Global Scale. *Global Biogeochemical Cycles* **33**, 1733-1745, doi:10.1029/2019gb006264 (2019).
26 Yuan, W. P. *et al.* Multiyear precipitation reduction strongly decreases carbon uptake over northern China. *Journal of Geophysical Research-Biogeosciences* **119**, 881-896, doi:10.1002/2014jg002608 (2014).
27 Yue, X. & Unger, N. The Yale Interactive terrestrial Biosphere model version 1.0: description, evaluation and implementation into NASA GISS ModelE2. *Geoscientific Model Development* **8**, 2399-2417, doi:10.5194/gmd-8-2399-2015 (2015).
28 Zaehle, S. & Friend, A. D. Carbon and nitrogen cycle dynamics in the O-CN land surface model: 1. Model description, site-scale evaluation, and sensitivity to parameter estimates. *Global Biogeochemical Cycles* **24**, Gb1005. doi:10.1029/2009gb003521 (2010).