Compensatory water effects link yearly global land CO_2 sink changes to temperature

Postprint version

Martin Jung, Markus Reichstein, Christopher R. Schwalm, Chris Huntingford, Stephen Sitch, Anders Ahlström, Almut Arneth, Gustau Camps-Valls, Philippe Ciais, Pierre Friedlingstein, Fabian Gans, Kazuhito Ichii, Atul K. Jain, Etsushi Kato, Dario Papale, Ben Poulter, Botond Raduly, Christian Rödenbeck, Gianluca Tramontana, Nicolas Viovy, Ying-Ping Wang, Ulrich Weber, Sönke Zaehle & Ning Zeng

Published in: **Nature**

Reference: Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P., Gans, F., Ichii, K., Jain, A. K., Kato, E., Papale, D., Poulter, B., Raduly, B., Rödenbeck, C., Tramontana, G., Viovy, N., Wang, Y.-P., Weber, U., Zaehle, S., Zeng, N. (2017). Compensatory water effects link yearly global land CO2 sink changes to temperature. Nature, 541(7638), 516-520. doi:10.1038/nature20780.

Web link: https://www.nature.com/nature/journal/v541/n7638/full/nature20780.html

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 640176

Compensatory water effects link yearly global land CO_2 sink changes to temperature

Martin Jung[1], Markus Reichstein[1,2], Christopher R. Schwalm[3], Chris Huntingford[4], Stephen Sitch[5], Anders Ahlström[6,7], Almut Arneth[8], Gustau Camps-Valls[9], Philippe Ciais[10], Pierre Friedlingstein[11], Fabian Gans[1], Kazuhito Ichii[12,13], Atul K. Jain[14], Etsushi Kato[15], Dario Papale[16], Ben Poulter[17], Botond Raduly[16,18], Christian Rödenbeck[19], Gianluca Tramontana[16], Nicolas Viovy[10], Ying-Ping Wang[20], Ulrich Weber[1], Sönke Zaehle[1,2] & Ning Zeng[21,22]

1 Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, 07745 Jena, Germany.
2 Michael-Stifel-Center Jena for Data-driven and Simulation Science, Friedrich-Schiller-Universität Jena, 07743 Jena, Germany.
3 Woods Hole Research Center, 149 Woods Hole Road, Falmouth, Massachusetts 02540, USA.
4 Centre for Ecology and Hydrology, Wallingford, Oxfordshire OX10 8BB, UK.
5 College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4QF, UK.
6 Department of Earth System Science, School of Earth, Energy and Environmental Sciences, Stanford University, Stanford, California 94305, USA.
7 Department of Physical Geography and Ecosystem Science, Lund University, 223 62 Lund, Sweden.
8 Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research, 82467 Garmisch-Partenkirchen, Germany.
9 Image Processing Laboratory, Universitat de València, Catedrático José Beltrán, Paterna 46980, València, Spain.
10 Laboratoire des Sciences du Climat et de l'Environnement, CEA CNRS UVSQ, 91191 Gif-sur-Yvette, France.
11 College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter EX4 4QE, UK.
12 Department of Environment Geochemical Cycle Research, Japan Agency for Marine-Earth Science and Technology, 3173-25, Showa-machi, Kanazawa-ku, Yokohama 236-0001, Japan.
13 Center for Global Environmental Research, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, 305-8506, Japan.
14 Department of Atmospheric Sciences, University of Illinois, Urbana, Illinois 61801, USA.
15 Global Environment Program, The Institute of Applied Energy, Tokyo 105-0003, Japan.
16 Department for Innovation in Biological, Agro-food and Forest systems, University of Tuscia, 01100 Viterbo, Italy.
17 NASA Goddard Space Flight Center, Biospheric Science Laboratory, Greenbelt, Maryland 20771, USA.
18 Department of Bioengineering, Sapientia Hungarian University of Transylvania, 530104 M-Ciuc, Romania.
19 Max Planck Institute for Biogeochemistry, Department of Biogeochemical Systems, 07745 Jena, Germany.
20 CSIRO Oceans and Atmosphere, PMB #1, Aspendale, Victoria 3195, Australia.
21 Institute of Atmospheric Physics, Chinese Academy of Science, Beijing 100029, China.
22 Department of Atmospheric and Oceanic Science, University of Maryland, College Park, Maryland 20742, USA.

Large interannual variations in the measured growth rate of atmospheric carbon dioxide (CO_2) originate primarily from fluctuations in carbon uptake by land ecosystems[1, 2, 3]. It remains uncertain, however, to what extent temperature and water availability control the carbon balance of land ecosystems across spatial and temporal scales[3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14]. Here we use empirical models based on eddy covariance data[15] and process-based models[16, 17] to investigate the effect of changes in temperature and water availability on gross primary productivity (GPP), terrestrial ecosystem respiration (TER) and net ecosystem exchange (NEE) at local and global scales. We find that water availability is the dominant driver of the local interannual variability in GPP and TER. To a lesser extent

this is true also for NEE at the local scale, but when integrated globally, temporal NEE variability is mostly driven by temperature fluctuations. We suggest that this apparent paradox can be explained by two compensatory water effects. Temporal water-driven GPP and TER variations compensate locally, dampening water-driven NEE variability. Spatial water availability anomalies also compensate, leaving a dominant temperature signal in the year-to-year fluctuations of the land carbon sink. These findings help to reconcile seemingly contradictory reports regarding the importance of temperature and water in controlling the interannual variability of the terrestrial carbon balance[3, 4, 5, 6, 9, 11, 12, 14]. Our study indicates that spatial climate covariation drives the global carbon cycle response.

Large interannual variations in recently measured atmospheric CO_2 growth rates originate primarily from fluctuations in carbon uptake by land ecosystems, rather than from the oceans or variations in anthropogenic emissions[1, 2, 3]. There is a general consensus that the tropical regions contribute the most to terrestrial carbon variability[1, 8, 18, 19]. The observed positive correlation between mean tropical land temperature and CO_2 growth rate[3, 5, 6, 12, 13] implies smaller land carbon uptake and enhanced atmospheric CO_2 growth during warmer years, with a sensitivity of about 5 gigatonnes of carbon per year per K. There is a tight relationship between this sensitivity on interannual timescales and long-term changes in terrestrial carbon per degree of warming across multiple climate carbon-cycle models[6].

Despite this strong emergent relationship with mean tropical land temperature, several studies suggest that variations in water availability have an important[8, 10, 11, 14], even a dominant role[4, 9], in shaping the interannual variability (IAV) of the carbon balance of extensive semi-arid and sub-tropical systems. Furthermore, the recent doubling of the tropical carbon cycle sensitivity to interannual temperature variability has been linked to interactions with changing moisture regimes[13]. A full understanding of the processes governing the climatic controls of terrestrial carbon cycling on interannual timescales and across spatial scales is therefore still lacking. Here we show that the ‘temperature versus water’ debate can be resolved by simultaneously assessing the carbon-cycle response to fluctuations in both temperature and water availability at both local and global scales.

Using both machine learning algorithms and process-based global land models, we derived spatial and temporal patterns of the IAV of CO_2 uptake by plants via photosynthesis (GPP) and of CO_2 loss through respiration (TER). NEE equals TER minus GPP, thus allowing analysis of the IAV of NEE for CO_2.

Machine learning algorithms were used to translate gridded inputs of daily air temperature, water availability and radiation, among others[15], into time-varying 0.5° grids of TER and GPP for the 1980–2013 period (FLUXCOM; see Methods). Three machine learning algorithms were trained on FLUXNET[20]-based *in situ* TER and GPP flux estimates from two flux partitioning methods[21, 22]. These three fitting algorithms combined with two partitioning methods provided six sets of GPP and TER estimates each, which combined yield 36 FLUXCOM NEE ensemble members.

In a complementary approach, we examined simulations of GPP and TER from an ensemble of seven global land surface or dynamic vegetation models[16, 17] (TRENDYv3, see Methods). These process-based model simulations follow a common protocol and used the same climate-forcing data set as the observation-based FLUXCOM models. Both sets of results are expected to be more uncertain in the tropics owing to the less reliable climate- and satellite-based inputs and a sparse coverage of flux measurements[23].

We analysed FLUXCOM and TRENDYv3 simulations independently, but in a consistent manner. We derived NEE as the difference between TER and GPP, that is, a positive value of NEE indicates a flux of carbon from the land to the atmosphere. To isolate IAV we detrended GPP and TER for each grid cell and month (see Methods). We find that global patterns of NEE IAV are consistent between FLUXCOM and

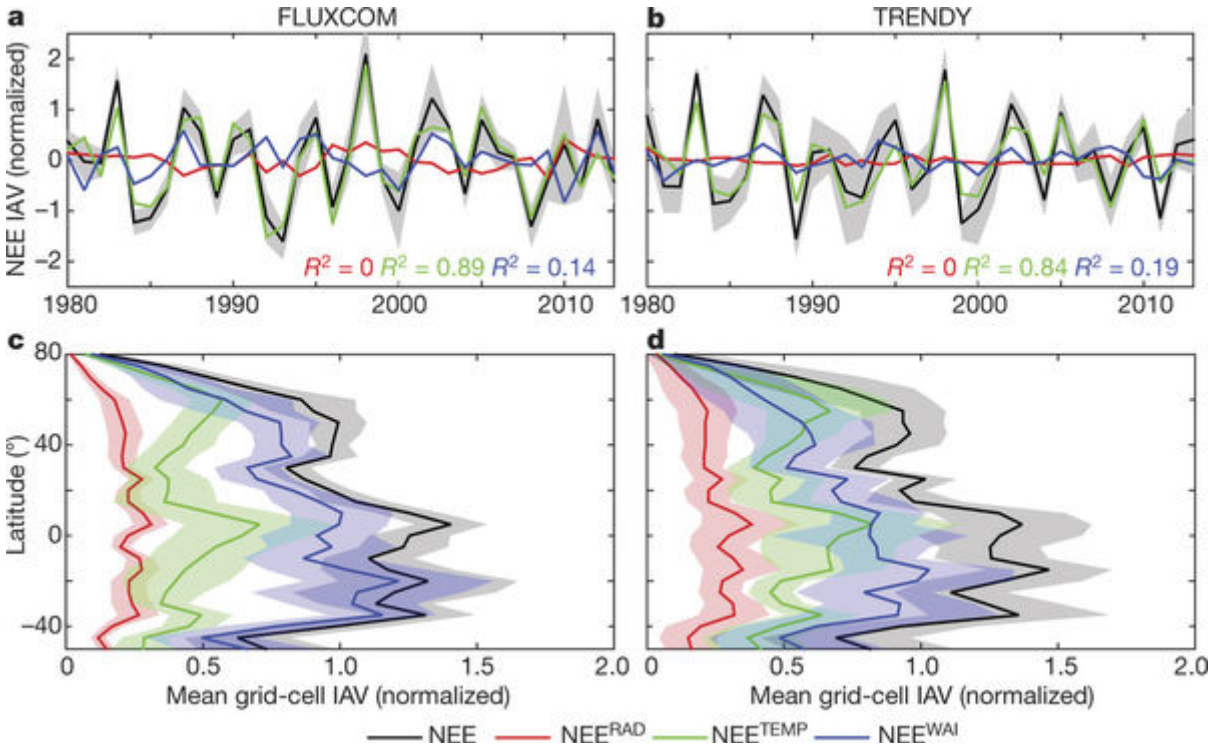
TRENDYv3 (Extended Data Fig. 1 and Supplementary Information section 1). Both approaches reproduce ($r \approx 0.8$) the globally integrated NEE IAV derived from atmospheric CO_2 concentration measurements and transport[24]. Both approaches also show the largest IAV in the tropics (Extended Data Fig. 1). To obtain the contributions of different environmental variables to IAV, we decomposed carbon flux anomalies (ΔFLUX) of each year, month, and grid cell (subscripts y, m and s) into their additive components forced by detrended anomalies of temperature (ΔTEMP), shortwave incoming radiation (ΔRAD), and soil-moisture-related water availability (ΔWAI; see Methods):

$$\begin{aligned}\Delta\mathrm{FLUX}_{s,m,y} &= a_{s,m}^{\mathrm{TEMP}} \times \Delta\mathrm{TEMP}_{s,m,y} + a_{s,m}^{\mathrm{RAD}} \times \Delta\mathrm{RAD}_{s,m,y} \\ &\quad + a_{s,m}^{\mathrm{WAI}} \times \Delta\mathrm{WAI}_{s,m,y} + \varepsilon_{s,m,y} \qquad (1) \\ \Delta\mathrm{FLUX}_{s,m,y} &\approx \Delta\mathrm{FLUX}_{s,m,y}^{\mathrm{TEMP}} + \Delta\mathrm{FLUX}_{s,m,y}^{\mathrm{RAD}} + \Delta\mathrm{FLUX}_{s,m,y}^{\mathrm{WAI}}\end{aligned}$$

Here $a_{s,m}$ represents the estimated sensitivity of the flux anomaly, $\Delta\mathrm{FLUX}_{s,m,y}$ (GPP or TER) to each respective climate-forcing anomaly (ΔTEMP, ΔRAD, ΔWAI) for a given grid cell and month, and $\varepsilon_{s,m,y}$ is the residual error term. The product of a given sensitivity (for example, a^{TEMP}) and corresponding climate-forcing anomaly (for example, ΔTEMP) constitutes the flux anomaly component driven by this climate factor (for example, $\mathrm{GPP}^{\mathrm{TEMP}}$). Thus, equation (1) estimates the contributions of temperature, radiation, and water availability anomalies to the carbon flux anomalies (see Supplementary Information section 2 for verification).

Our analysis reveals a contrasting pattern of NEE IAV controlled by temperature or moisture, depending on spatial scale. At the global scale, temperature drives spatially integrated NEE IAV (Fig. 1a and b, compare green and black curves), in line with previous findings based on correlations between anomalies in temperature and CO_2 growth rate[3, 5, 6, 12, 13]. Globally integrated NEE anomalies due to variations in radiation ($\mathrm{NEE}^{\mathrm{RAD}}$) and water availability ($\mathrm{NEE}^{\mathrm{WAI}}$) play only a minor part (compare red, blue and black curves in Fig. 1a and b). The dominant global influence of temperature is in contrast to the dominant local influence of water availability when analysing all grid cells individually (Fig. 1c and d; compare blue and black curves of zonal mean of grid-cell IAV).

Figure 1: Climatic controls on NEE IAV at global and local scales for the period 1980–2013 derived from machine-learning-based (FLUXCOM) and process-based (TRENDY) models.

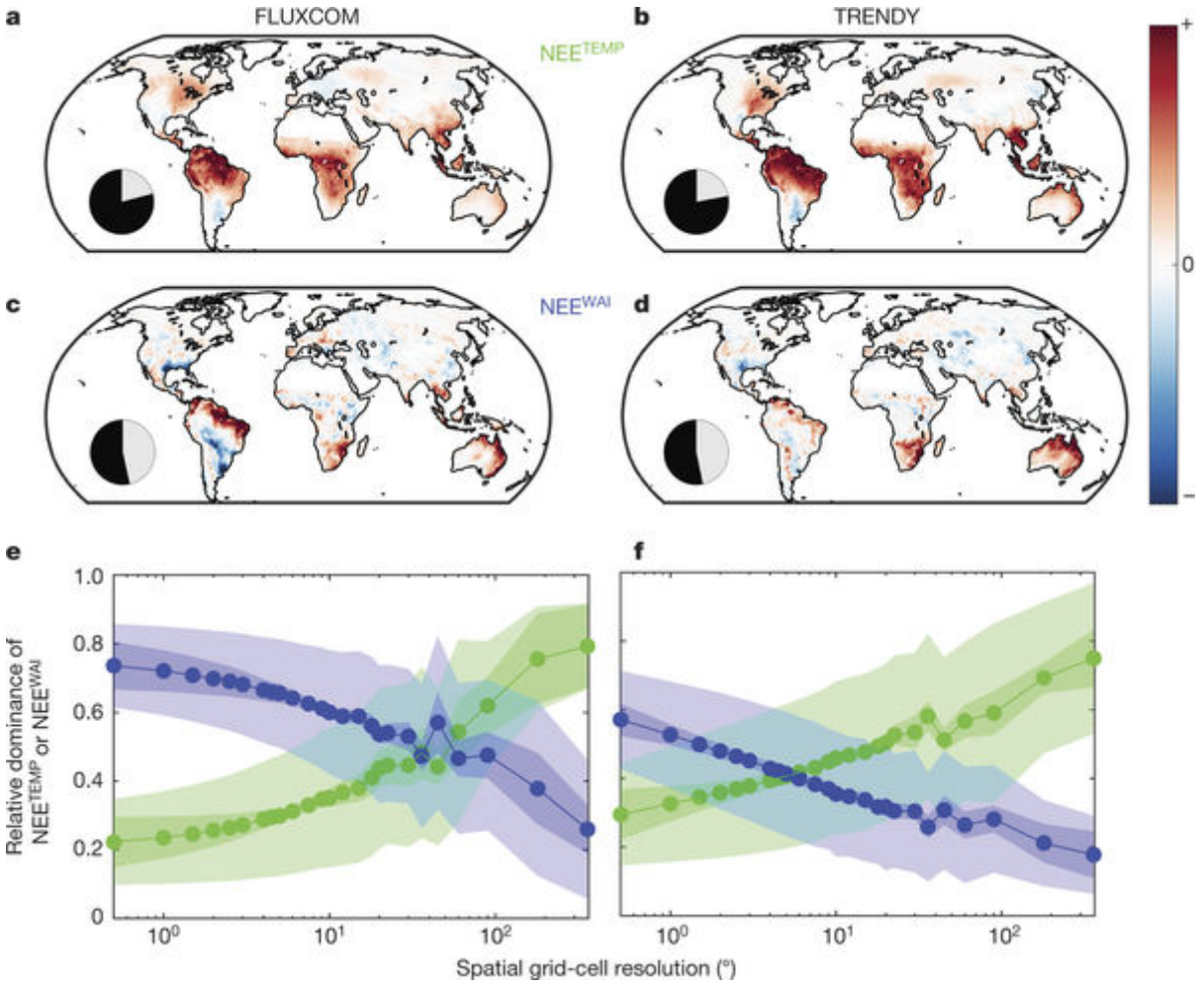


a, **b**, The comparison of globally integrated annual NEE anomalies with NEE anomalies driven only by temperature, water availability and radiation (all normalized by the standard deviation (s.d.) of globally integrated NEE) reveals temperature to be the dominant global control. R^2 values between the climatic NEE components and total NEE are given. **c**, **d**, Mean grid-cell IAV magnitude (see equation (3) in Methods) of NEE components for latitudinal bands reveals water to be the dominant local control. Uncertainty bounds where given as shaded area reflect the spread among FLUXCOM or TRENDY ensemble members (±1 s.d.).

Radiation causes the smallest NEE IAV at the grid-cell level (red curve in Fig. 1c and d) but there are indications based on other climate-forcing data that radiation could have a more important role than temperature locally (Supplementary Information section 3). Temperature variations are important for NEE IAV (green curve in Fig. 1c and d) in high latitudes and the inner tropics, but in general, the grid-cell average water-related NEE variability (NEE^{WAI}, blue curve) is larger. Water-related NEE variability peaks at subtropical latitudes where semi-arid ecosystems dominate. This finding is consistent with studies emphasizing the role of water-limited semi-arid ecosystems on global NEE IAV[4, 9]. We now assess how this can be reconciled with the emergent temperature control of globally integrated NEE IAV. Going from grid-cell to global scale shifts the emerging controls on NEE IAV from water availability (local) towards temperature (global).

We hypothesized that the dominance of temperature in globally integrated NEE IAV results from a stronger compensation of positive and negative NEE^{WAI} anomalies between different grid cells compared to NEE^{TEMP} when going from the local to the global scale. To test this, we first illustrate the dominant spatial patterns of temperature versus water compensation using empirical orthogonal functions (EOF) of the annual NEE^{TEMP} and NEE^{WAI} anomalies (Fig. 2a–d). Here, the leading EOF of NEE^{WAI} (about 10% variance explained) has strong anti-correlated spatial patterns of positive and negative values (Fig. 2c and d), which correspond to El Niño/Southern Oscillation imprints on moisture effects (R^2 with Niño 3.4 SST index[25] of 0.75).

Figure 2: Effects of spatial covariation and scale on temperature versus water control of NEE IAV for FLUXCOM and TRENDY models.



Spatial patterns of the first EOF of annual NEE^{TEMP} (**a**, **b**), and NEE^{WAI} (**c**, **d**) anomalies (see Methods) show large spatial coherence for NEE^{TEMP} (dominant positive values) and anti-correlated patterns for NEE^{WAI} (positive and negative values are shown on the colour scale; magnitudes are not informative and were omitted for clarity). This is underpinned in the inset pie charts which show the proportion of total positive (black) and negative (grey) covariances among grid cells for NEE^{TEMP} and NEE^{WAI} anomalies (see equations (4) and (5) in Methods). **e**, **f**, The relative dominance (see equation (6) in Methods) of NEE^{TEMP} (green) increases with successive spatial aggregation, while the relative dominance of NEE^{WAI} (blue) decreases. Outer uncertainty bounds in **e** and **f** given as shaded area refer to the spread among respective ensemble members (±1 s.d.); inner uncertainty bounds refer to ±1 s.d. with respect to the change of relative dominance with spatial aggregation (see equation (7) in Methods).

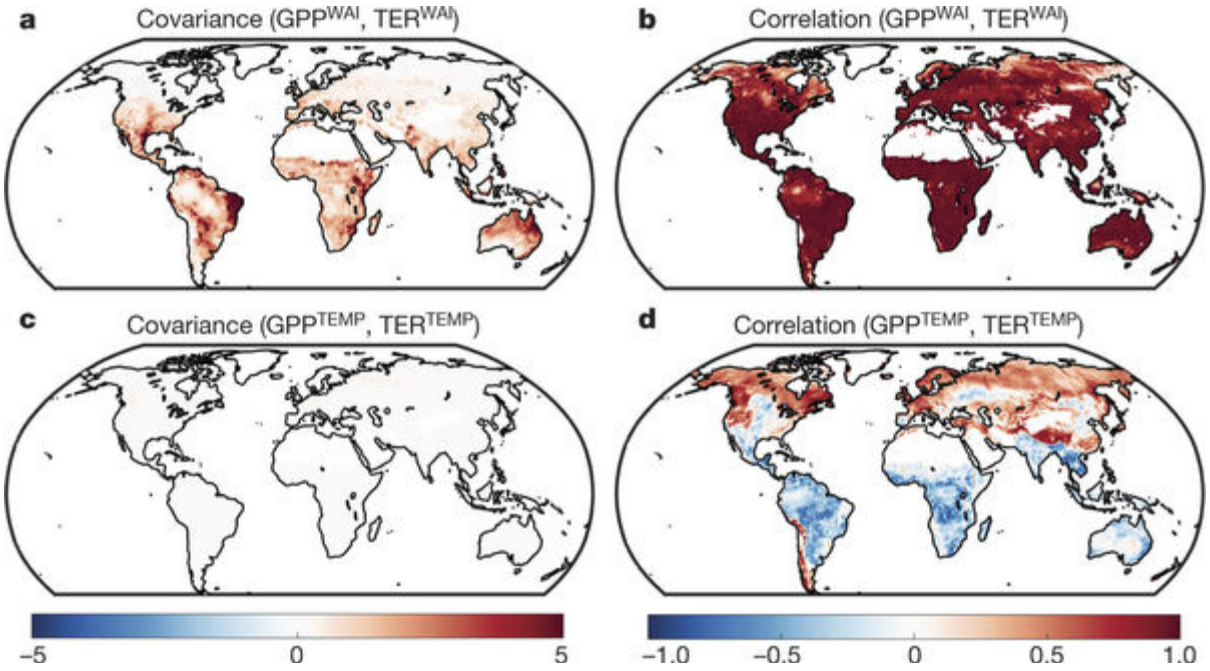
In comparison, the leading EOF of NEE^{TEMP} (about 22% variance explained) shows a more spatially uniform response, in particular across the tropics (Fig. 2a and b). This pattern of much larger spatial coherence of NEE^{TEMP} anomalies, compared to NEE^{WAI} anomalies, is also evident in their respective sums of positive and negative covariances among all grid cells (see the inset pie charts in Fig. 2a–d). For NEE^{TEMP} the sum of positive covariances is far larger than the negative ones (79% versus 21%), whereas positive and negative covariances are almost in balance (53% versus 47%) for NEE^{WAI}. As a consequence of the larger spatial coherence of NEE^{TEMP} anomalies, as compared to NEE^{WAI} anomalies, we observe a shift of the dominant NEE IAV control from water at the local scale to temperature at the global scale. We illustrate this change in Fig. 2e and f by presenting the relative dominance of water- and temperature-related NEE IAV for increasing levels of spatial aggregation. This is a robust feature within and among FLUXCOM and TRENDY approaches (Extended Data Fig. 2). We also find that the rise and decay of NEE^{TEMP} and NEE^{WAI} dominance respectively with spatial scale occurs in all major biomes (Supplementary Information section 4). This pattern is probably related to the different climatic characteristics of precipitation and air temperatures, with the

former, but not the latter, being associated with moisture conservation and offsetting spatial anomaly patterns.

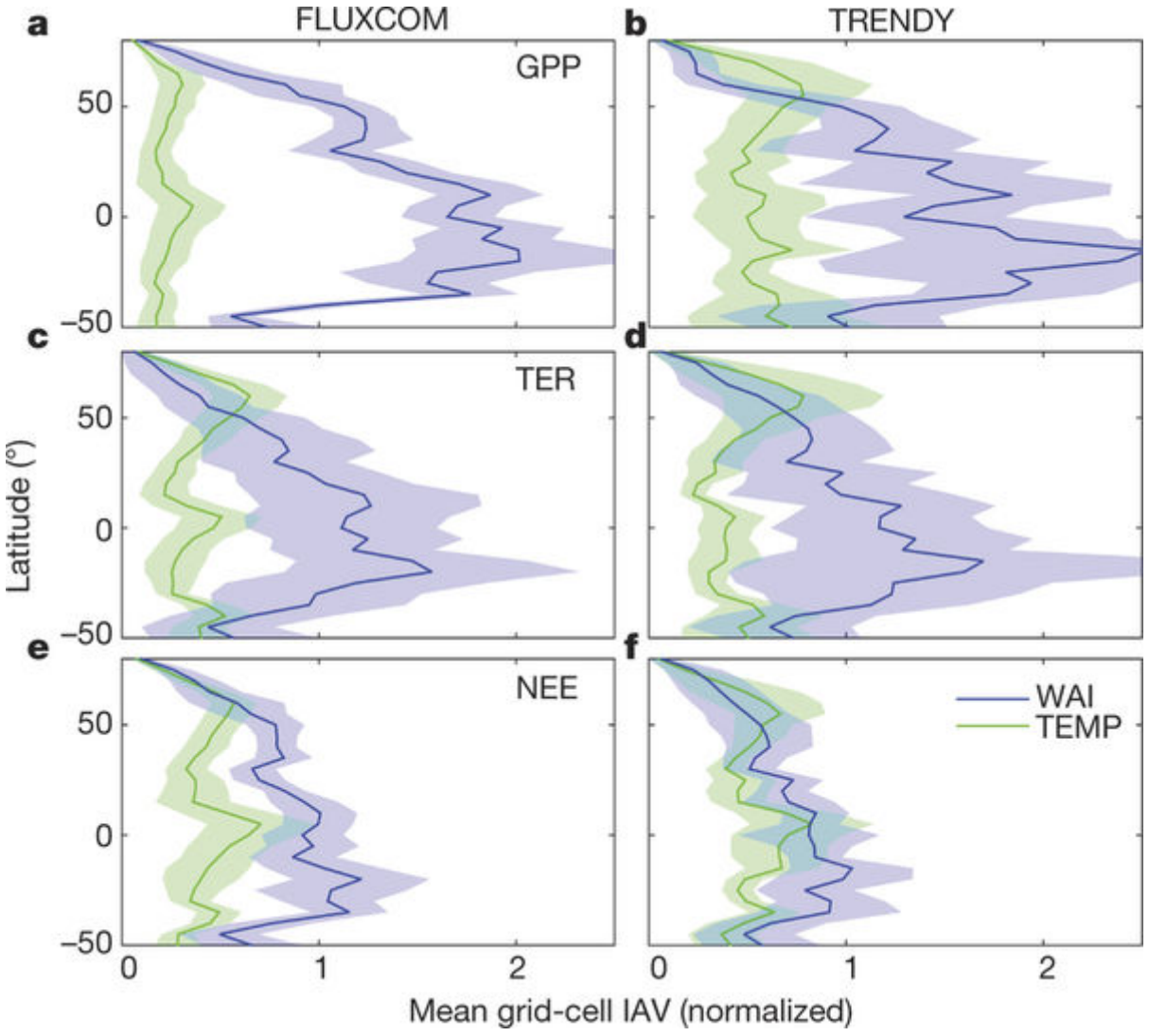
We next assess how local water- and temperature related NEE IAV emerges from the interaction of photosynthesis (GPP) and respiration (TER) processes. We compare the magnitudes of water- versus temperature-driven GPP and TER variability and find that WAI is overall the most important factor controlling the local IAV of both gross fluxes (Fig. 3a–d), with particularly large variability in both fluxes in semi-arid regions (Supplementary Information section 4 and 5). However, the local IAV of NEE related to WAI (NEE^{WAI}, Fig. 3e, f) is reduced compared to the components GPP^{WAI} and TER^{WAI}.

Our results indicate that, in addition to the spatial compensation of NEE^{WAI} discussed above, there is also a local compensation mechanism, whereby GPP^{WAI} and TER^{WAI} covary and thus locally counterbalance each other (Fig. 4a, b). This is probably due to the concomitant positive relationship of soil moisture with productivity and with respiration. The combined effect is a smaller net effect of WAI on NEE. Specifically, two-thirds of the WAI effect on GPP is offset by the WAI effect on TER (0.67 ± 0.33 for FLUXCOM, 0.69 ± 0.14 for TRENDY; mean slope ± s.d. across ensemble members of global TER^{WAI} versus GPP^{WAI}). These patterns are qualitatively consistent between the data-driven FLUXCOM (Fig. 4) and process-based TRENDY models (Extended Data Fig. 3) and agree with previous observations of simultaneous declines of GPP and TER during droughts[26, 27, 28, 29, 30]. However, magnitudes of TER^{WAI} versus GPP^{WAI} covariances differ substantially among model ensemble members (Extended Data Fig. 4). This probably reflects the large uncertainty of respiration processes to moisture variations, whereas flux partitioning uncertainties seem negligible (Supplementary Information section 6).

Figure 4: Spatial patterns of covariance and correlation of WAI- and TEMP-driven GPP and TER IAV for FLUXCOM models.



Maps of the covariance of annual anomalies (see equation (8) in Methods) of GPP and TER climatic components show large compensation effects (positive covariance) for WAI (**a**) but nearly no covariance for TEMP (**c**). Correlations between GPP^{WAI} and TER^{WAI} are large and ubiquitously positive (**b**) while correlations among GPP^{TEMP} and TER^{TEMP} are weaker, with a distinct spatial pattern of negative correlations in hot regions (**d**). All results refer to the mean of all FLUXCOM ensemble members. See Extended Data Fig. 3 for equivalent TRENDY results, and Extended Data Fig. 4 for uncertainties.



The IAV magnitude (see equation (3) in Methods) of the WAI component is much larger than the IAV of the TEMP component for gross fluxes (**a**–**d**), while this difference is smaller for NEE (**e**, **f**) owing to compensation. Uncertainty bounds as shaded area reflect the spread among FLUXCOM or TRENDY ensemble members (±1 s.d.).

In contrast to offsetting NEE water effects, our analysis indicates a weak local temperature amplification effect of GPP and TER IAV in the tropics. Local temperature effects on GPP and TER IAV are inversely correlated over the tropics (Fig. 4d). This is because GPP decreases with increasing temperature, probably because the thermal optimum of photosynthesis has been exceeded, whereas respiration increases with temperature. Thus increasing temperatures in the tropics increase NEE by reducing GPP and increasing TER. However, owing to lower variances of the temperature components of GPP and TER (Fig. 3a–d), this local temperature amplification effect in the tropics is quantitatively negligible (Fig. 4c) compared to the local water compensation effect (Fig. 4d). Overall, this causes the difference of temperature-forced versus water-forced variability of NEE to be smaller compared to the influence of these drivers on the gross fluxes (compare distance between blue and green curves in Fig. 3a–d versus Fig. 3e and f).

Our analysis shows water availability as the overall dominant driver of the IAV of photosynthesis and respiration at local scales, even though this water signal is effectively absent in the globally integrated NEE IAV. This pattern is driven by: (1) the local compensatory effects of water availability on GPP and TER, and (2) the spatial anti-correlation of water-controlled NEE anomalies; and thus a compensation in space. These

two compensatory water effects leave temperature as the dominant factor globally, which resolves why there have been conflicting conclusions surrounding whether NEE IAV is forced thermally or hydrologically.

These findings suggest that climate does not only force the carbon cycle locally, but that, perhaps more importantly, the spatial covariation of climate variables drives the integrated global carbon-cycle response. Consequently, any analysis conducted on integrated signals over larger regions precludes inferences on the driving mechanisms at the ecosystem scale. Likewise, the apparent temperature-dominated IAV of the residual land sink, a traditional target of global carbon-cycle modellers, contains little information on local carbon-cycle processes. Our findings suggest that potential changes in spatial covariations among climate variables associated with global change may drive apparent changes of carbon-cycle sensitivities and perhaps even the strength of climate-carbon cycle feedbacks.

References

1. Bousquet, P. et al. Regional changes in carbon dioxide fluxes of land and oceans since 1980. Science 290, 1342–1346 (2000).

2. Keeling, C. D., Piper, S. C. & Heimann, M. in Aspects of Climate Variability in the Pacific and Western Americas (ed. Peterson, D. H.) 305–363 (Washington, D.C., 1989).

3. Kindermann, J., Wurth, G., Kohlmaier, G. H. & Badeck, F. W. Interannual variation of carbon exchange fluxes in terrestrial ecosystems. Glob. Biogeochem. Cycles 10, 737–755 (1996).

4. Ahlstrom, A. et al. The dominant role of semi-arid ecosystems in the trend and variability of the land CO2 sink. Science 348, 895–899 (2015).

5. Anderegg, W. R. L. et al. Tropical nighttime warming as a dominant driver of variability in the terrestrial carbon sink. Proc. Natl Acad. Sci. USA 112, 15591–15596 (2015).

6. Cox, P. M. et al. Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. Nature 494, 341–344 (2013).

7. Nemani, R. R. et al. Climate-driven increases in global terrestrial net primary production from 1982 to 1999. Science 300, 1560–1563 (2003).

8. Peylin, P. et al. Multiple constraints on regional CO2 flux variations over land and oceans. Glob. Biogeochem. Cycles 19, http://dx.doi.org/10.1029/2003GB002214 (2005).

9. Poulter, B. et al. Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle. Nature 509, 600–603 (2014).

10. Tian, H. Q. et al. Effect of interannual climate variability on carbon storage in Amazonian ecosystems. Nature 396, 664–667 (1998).

11. Wang, J., Zeng, N. & Wang, M. Interannual variability of the atmospheric CO2 growth rate: roles of precipitation and temperature. Biogeosciences 13, 2339–2352 (2016).

12. Wang, W. et al. Variations in atmospheric CO2 growth rates coupled with tropical temperature. Proc. Natl Acad. Sci. USA 110, 13061–13066 (2013).

13. Wang, X. et al. A two-fold increase of carbon cycle sensitivity to tropical

temperature variations. Nature 506, 212–215 (2014).

14. Zeng, N., Mariotti, A. & Wetzel, P. Terrestrial mechanisms of interannual CO2 variability. Glob. Biogeochem. Cycles 19, http://dx.doi.org/10.1029/2004GB002273 (2005).

15. Tramontana, G. et al. Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. Biogeosciences 13, 4291–4313 (2016).

16. Le Quéré, C. et al. Global carbon budget 2014. Earth Syst. Sci. Data Discuss. 7, 521–610 (2014).

17. Sitch, S. et al. Recent trends and drivers of regional sources and sinks of carbon dioxide. Biogeosciences 12, 653–679 (2015).

18. Gurney, K. R., Baker, D., Rayner, P. & Denning, S. Interannual variations in continental-scale net carbon exchange and sensitivity to observing networks estimated from atmospheric CO2 inversions for the period 1980 to 2005. Glob. Biogeochem. Cycles 22, http://dx.doi.org/10.1029/2007gb003082 (2008).

19. Peylin, P. et al. Global atmospheric carbon budget: results from an ensemble of atmospheric CO2 inversions. Biogeosciences 10, 6699–6720 (2013).

20. Baldocchi, D. Turner Review No. 15. 'Breathing' of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. Aust. J. Bot. 56, 1–26 (2008).

21. Lasslop, G. et al. Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation. Glob. Change Biol. 16, 187–208 (2010).

22. Reichstein, M. et al. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. Glob. Change Biol. 11, 1424–1439 (2005).

23. Schimel, D. et al. Observing terrestrial ecosystems and the carbon cycle from space. Glob. Change Biol. 21, 1762–1776 (2015).

24. Rödenbeck, C., Houweling, S., Gloor, M. & Heimann, M. CO2 flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport. Atmos. Chem. Phys. 3, 1919–1964 (2003).

25. Rayner, N. A. et al. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J. Geophys. Res. D 108, http://dx.doi.org/10.1029/2002JD002670 (2003).

26. Biederman, J. A. et al. Terrestrial carbon balance in a drier world: the effects of water availability in southwestern North America. Glob. Change Biol. 22, 1867–1879 (2016).

27. Bonal, D., Burban, B., Stahl, C., Wagner, F. & Hérault, B. The response of tropical rainforests to drought—lessons from recent research and future prospects. Ann. For. Sci. 73, 27–44 (2016).

28. Ciais, P. et al. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. Nature 437, 529–533 (2005).

29. Schwalm, C. R. et al. Assimilation exceeds respiration sensitivity to drought: a FLUXNET synthesis. Glob. Change Biol. 16, 657–670 (2010).

30. van der Molen, M. K. et al. Drought and ecosystem carbon cycling. Agric. For. Meteorol. 151, 765–773 (2011).

Methods

The FLUXCOM global carbon flux data set

Three machine learning methods were trained on daily carbon flux estimates from 224 flux tower sites using meteorological measurements and satellite data as inputs[15]: Random Forests[31], Artificial Neural Networks[32], Multivariate Adaptive Regression Splines[33]. Models were trained separately for two variants of GPP and TER, derived from the flux partitioning methods of refs 22 and 21. Each method used the same 11 input driver data listed in Supplementary Information section 7. This set of driver data was obtained from an extensive variable selection analysis[15, 34]. Details, along with extensive model evaluation based on cross-validation are given in ref. 15.

To produce spatio-temporal grids of carbon fluxes, the trained machine learning algorithms require only spatio-temporal grids of its input driver data[35]. We forced the models with grids of 0.5° spatial resolution and daily time step for the period 1980–2013[36]. High-resolution satellite-based predictor variables (see Supplementary Information section 7) were tiled by plant functional type, that is, grids for each plant functional type containing the mean value per plant functional type and time step at 0.5° were created. The distribution of plant functional type originates from the majority class of annually resolved MODIS land cover product (collection 5)[37] for each high-resolution pixel. Climatic predictor variables are based on CRUNCEPv6 (http://esgf.extra.cea.fr/thredds/catalog/store/p529viov/cruncep/V6_1901_2014/catalog.html) to be consistent with the TRENDY ensemble. CRUNCEPv6 is based on a merged product of Climate Research Unit observation-based monthly 0.5° climate variables[38] (1901–2013) and the high temporal (6-hourly) resolution National Centers for Environmental Prediction (NCEP) reanalysis. The variables affected by the climate forcing data set are marked in Supplementary Information section 7. Of the 11 predictor variables, only temperature, radiation and water availability can generate IAV. The water availability index (WAI, see supplement 3 in ref. 15) is based on a simple dynamic soil water balance model, which was driven with daily precipitation and potential evapotranspiration by CRUNCEPv6 (see Supplementary Information section 8 for cross-consistency with TRENDY-based soil moisture). The machine learning models were run for each plant functional type separately, and a weighted mean over the fractions of plant functional type was obtained for each grid cell. The distribution of plant functional type is representative of the period 2001–2012; no land cover change was considered. Empirical models were run to spatially estimate GPP and TER. Then NEE was derived by the carbon mass balance approach (NEE = TER minus GPP), which allows us to decompose precisely how NEE IAV emerges from (co-)variations of TER and GPP. We verify that NEE IAV derived as ‘TER minus GPP’ is consistent with upscaling NEE directly (Supplementary Information section 6). Overall, 36 combinations of NEE were derived by considering all possible combinations of ‘TER minus GPP’ realizations, resulting from different machine learning approaches and flux partitioning variants. The individual model runs were finally aggregated to monthly means.

The TRENDY global carbon flux data set

We used simulations of seven dynamic global vegetation models from the TRENDY v3 ensemble[16, 17] for the period 1980–2013, which have a spatial resolution of 0.5° (model simulations with coarser resolution were

omitted): CABLE[39], ISAM[40], LPJ[41], LPJ-GUESS[42], ORCHIDEE[43], VEGAS[14], VISIT[44]. These models were forced by a common set of input data sets and experimental protocol (experiment 'S2')[16, 17]. Climate forcing (CRUNCEPv6) is the same as for FLUXCOM. Global atmospheric CO_2 was derived from ice core and NOAA monitoring station data, and provided at annual resolution over the period 1860–2013[16]. The dynamic global vegetation models were run from preindustrial steady state (NEE = 0) with changing fields of climate and atmospheric CO_2 concentration over the twentieth century. Land use and land cover changes were not considered. For consistency with FLUXCOM, NEE was derived as the difference between TER and GPP, that is, fire emissions available from some models were not included. TER was calculated as the sum of simulated autotrophic and heterotrophic respiration.

Anomalies and decomposition

Detrended monthly anomalies were obtained by removing the linear trend over years for each pixel and month (least-squares fitting), which also centres the mean to zero for a given pixel and month. This procedure was applied consistently to GPP, TER, shortwave radiation (RAD), air temperature (TEMP) and water availability (WAI), in both FLUXCOM and TRENDY simulations. For TRENDY models the simulated soil moisture was used instead of WAI. The resulting IAV of GPP and TER was decomposed into the contributions forced by TEMP, RAD and WAI following equation (1) using a multiple linear (ordinary least-squares) regression with zero intercept for each pixel and month. NEE sensitivities and NEE components were derived from GPP and TER results, which is equivalent to decomposing NEE (=TER − GPP) directly. We validate and discuss the approximation of IAV contributions using equation (1) in Supplementary Information section 2.

Notation

All analysis is based on detrended monthly anomalies (equation (1) aggregated to annual means). For simplicity, we omit the Δ notation for 'anomaly' in the following. Superscripts TEMP, WAI and RAD refer to surface air temperature, water availability, and incoming shortwave radiation of a respective carbon flux anomaly, respectively. Subscripts s, y and e refer to indexes of grid cell, year and ensemble member, respectively. The mean and standard deviation are denoted μ and σ respectively, where the subscripts of these operators tell whether the operation is done over grid cells (for example, μ_s is an average over all grid cells), years (for example, σ_y is the standard deviation over the years), or ensemble members. All main results refer to the mean of FLUXCOM or TRENDY ensemble members (μ_e) and the standard deviation (σ_e) is used as uncertainty estimate. Whenever we calculated a mean over 0.5° grid cells (μ_s) we accounted for different grid cell areas (area-weighted mean) and used a consistent mask of valid values between FLUXCOM and TRENDY. Because several analyses are referenced with respect to the sum of climatic components of NEE we denote NEE* as:

$$\mathrm{NEE}^{*}_{s,y} = \mathrm{NEE}^{\mathrm{TEMP}}_{s,y} + \mathrm{NEE}^{\mathrm{WAI}}_{s,y} + \mathrm{NEE}^{\mathrm{RAD}}_{s,y} \qquad (2)$$

Spatial patterns of IAV magnitude

(See, for example, Fig. 1c and d, and 3.) To describe spatial patterns of IAV magnitude M of climatic components of carbon fluxes (for example, $\mathrm{GPP}^{\mathrm{WAI}}$) we computed the standard deviation of its annual values (σ_y) for each grid cell 's'. This standard deviation was then normalized by the mean (μ_s) temporal standard deviation (σ_y) of NEE* to provide a relative metric of IAV magnitude, where values above 1 indicate IAV magnitudes larger than average NEE* IAV. This scaling accounts for the known underestimation of IAV magnitude in the upscaling approach[35] but does not change any patterns.

$$M_s = \frac{\sigma_y\left(\mathrm{FLUX}^{\mathrm{COMP}}_{s,y}\right)}{\mu_s\left(\sigma_y\left(\mathrm{NEE}^{*}_{s,y}\right)\right)} \qquad (3)$$

Figure 1c and d shows mean and standard deviations across ensemble members (μ_e and σ_e) for NEE components for latitudinal bins of 5°. The same holds for Fig. 3, which also shows GPP and TER components.

EOFs and spatial covariances

(See Fig. 2a–d.) We first calculated mean spatio-temporal grids of NEE climatic components across ensemble members $\left(\mu_e\left(\mathrm{NEE}^{\mathrm{COMP}}_{s,y,e}\right)\right)$. We then multiplied those with grid-cell areas to convert flux densities into fluxes per grid cell, and normalized them by the standard deviation of NEE* across time and space, $\sigma_{s,y}\left(\mu_e\left(\mathrm{NEE}^{*}_{s,y,e}\right)\right)$. EOFs were then computed for each climatic component without additional scaling in MATLAB using the 'pca' function. The spatial pattern of first principal components (leading EOFs) of $\mathrm{NEE}^{\mathrm{TEMP}}$ and $\mathrm{NEE}^{\mathrm{WAI}}$ was plotted with the same colour scale. The values on the colour bar themselves are not informative and were therefore omitted for clarity. The leading EOF explains about 22% of spatial $\mathrm{NEE}^{\mathrm{TEMP}}$ variance and about 10% of spatial $\mathrm{NEE}^{\mathrm{WAI}}$ variance in both FLUXCOM and TRENDY ensemble means.

To quantify the degree of spatial covariance of NEE climatic components (inset pie charts in Fig. 2a–d) we calculated a large covariance matrix of all grid cells versus all grid cells for each NEE climatic component (annual anomalies multiplied with grid cell area), where the elements of this covariance matrix $\left(c^{\mathrm{COMP}}_{i,j}\right)$ were calculated according to equation (4):

$$c^{\mathrm{COMP}}_{i,j} = \mathrm{cov}_y\left(\mathrm{NEE}^{\mathrm{COMP}}_{si,y}, \mathrm{NEE}^{\mathrm{COMP}}_{sj,y}\right) \qquad (4)$$

Here i and j index the two grid cells for which the covariance is calculated. By definition the variance of the globally integrated anomalies equals the sum of all terms in the covariance matrix. To determine the share of positive versus negative spatial covariance of the total variance, we summed positive and negative covariance terms respectively (equation (5)). The sum of variances (the diagonal of the covariance matrix where $i = j$) was omitted in the pie charts because they accounted for less than 1% of the total covariance (tcov) budget.

$$\begin{aligned} \mathrm{tcov}^{\mathrm{COMP}}_{+} &= \sum_{i=1}\sum_{j\neq i} c^{\mathrm{COMP}}_{i,j} \,\Big|\, c^{\mathrm{COMP}}_{i,j} > 0 \\ \mathrm{tcov}^{\mathrm{COMP}}_{-} &= \sum_{i=1}\sum_{j\neq i} c^{\mathrm{COMP}}_{i,j} \,\Big|\, c^{\mathrm{COMP}}_{i,j} < 0 \end{aligned} \qquad (5)$$

Scale dependence of relative dominance of $\mathrm{NEE}^{\mathrm{TEMP}}$ and $\mathrm{NEE}^{\mathrm{WAI}}$

(See Fig. 2e and f.) We defined relative dominance D of a climatic component (COMP) of NEE (for example, $\mathrm{NEE}^{\mathrm{TEMP}}$) as the mean variance ($\mu_s$) of annual anomalies (σ^2_y) of this component divided by the mean variance of NEE*:

$$D^{\mathrm{COMP}} = \frac{\mu_s\left(\sigma_y^2\left(\mathrm{NEE}_{s,y}^{\mathrm{COMP}}\right)\right)}{\mu_s\left(\sigma_y^2\left(\mathrm{NEE}_{s,y}^{*}\right)\right)} \qquad (6)$$

To illustrate how this relative dominance changes systematically with spatial scale we aggregated NEE components successively to coarser spatial resolutions starting at 0.5° (around 54,000 grid cells) and ending with 'global'(one grid cell at 360° resolution) and recomputed relative dominance for each spatial resolution. In total 24 levels of spatial resolution were used: 0.5°, 1°, 1.5°, 2.5°, 3°, 4°, 4.5°, 5°, 6°, 7.5°, 9°, 10°, 12°, 15°, 18°, 20°, 22.5°, 30°, 36°, 45°, 60°, 90°, 180° and 360°.

These computations were carried out for each ensemble member separately and the mean across ensemble members (μ_e) was plotted for each spatial resolution as dots connected with a line. The uncertainty reflected by the spread of ensemble members (σ_e) was plotted as light shaded area. This uncertainty is dominated by the uncertainty of the mean relative dominance and not by the uncertainty on the systematic change with spatial aggregation. To visualize that, we provided a dark-shaded area in the plots, representing the uncertainty on the shape of the curve ('U' in equation (7)). This is based on the standard deviation across ensemble members after subtracting the mean relative dominance over all spatial resolutions (l in equation (7) for each ensemble member (equation (7)). Figure 2e and f shows the effect of shifting the relative dominance of $\mathrm{NEE}^{\mathrm{WAI}}$ versus $\mathrm{NEE}^{\mathrm{TEMP}}$ with spatial resolution considering the entire global vegetated area, but we repeated this analysis for different biomes (see Supplementary Information section 4) by considering only grid cells belonging to a specific biome.

$$U_l = \sigma_e(D_{l,e} - \mu_l(D_{l,e})) \qquad (7)$$

Covariance of the TEMP and WAI components of GPP and TER

(See Fig. 4.) We computed the correlation coefficient and covariance between GPP and TER components (for example, $\mathrm{GPP}^{\mathrm{TEMP}}$ versus $\mathrm{TER}^{\mathrm{TEMP}}$) for each grid cell and ensemble member. The covariance terms were normalized to the mean variance of NEE* (equation (8)). Figure 4 shows the mean across the ensemble members (μ_e) for FLUXCOM, and Extended Data Fig. 3 shows the mean for the TRENDY ensemble. Extended Data Fig.4 shows latitudinal patterns of the spread among ensemble members (σ_e) for FLUXCOM and TRENDY. The robustness of FLUXCOM results with respect to different NEE flux partitioning methods is assessed in Supplementary Information section 6. The normalized covariance of the temperature- and water-availability components of GPP and TER is as follows:

$$\mathrm{COV}_{s,\mathrm{norm}}\left(\mathrm{GPP}_{s,y}^{\mathrm{COMP}}, \mathrm{TER}_{s,y}^{\mathrm{COMP}}\right) = \frac{\mathrm{COV}_y\left(\mathrm{GPP}_{s,y}^{\mathrm{COMP}}, \mathrm{TER}_{s,y}^{\mathrm{COMP}}\right)}{\mu_s\left(\sigma_y^2(\mathrm{NEE}_{s,y}^{*})\right)} \qquad (8)$$

Comparison with atmospherically based data

(See Extended Data Fig. 1.) We used three data sources of atmospherically based net CO_2 flux exchange. The first is based on the annually resolved global carbon budget (GCP)[13], which uses measurements of atmospheric CO_2 growth rate and estimates of fossil-fuel emissions, ocean uptake, and land-use-change emissions to derive the global land flux as a residual. The second is based on the Jena CarboScope atmospheric transport inversion[24] (Jena Inversion, version s81_3.7) covering the full time period of the study. The third is an ensemble of ten atmospheric inversions[19] used for the REgional Carbon Cycle Assessment and Processes (RECCAP) activity covering the period 1990–2012, with each inversion covering a different time period. Four versions of the Jena Inversion have been removed from the original 14 member RECCAP ensemble to make it an independent assessment. We used globally integrated net land CO_2 flux estimates

from the three data sources to assess globally integrated NEE IAV of FLUXCOM and TRENDY. For the Jena and RECCAP inversions, we additionally calculated the integrated net land CO_2 flux for areas north and south of 30° N. All time series were detrended. For RECCAP inversions we calculated the median estimate of the available inversion estimates per year. All time series were normalized by the standard deviation of the respective globally integrated annual net land CO_2 flux.

Data availability

The FLUXCOM data that support the findings of this study are available from the Data Portal of the Max Planck Institute for Biogeochemistry (https://www.bgc-jena.mpg.de/geodb/projects/Home.php) with the identifier doi:10.17871/FLUXCOM_RS_METEO_CRUNCEPv6_1980_2013_v1. The TRENDY v3 data that support the findings of this study are available from S.S. (s.a.sitch@exeter.ac.uk) upon reasonable request. The source data for Fig. 1a–d, Fig. 2e and f, and Fig. 3a–f are additionally provided as Excel spreadsheets with the online version of the paper.

References

31. Breiman, L. Random forests. Mach. Learn. 45, 5–32 (2001).

32. Haykin, S. Neural Networks: A Comprehensive Foundation (Prentice Hall, 1998).

33. Friedman, J. H. Multivariate adaptive regression splines. Ann. Stat. 19, 1–67 (1991).

34. Jung, M. & Zscheischler, J. A guided hybrid genetic algorithm for feature selection with expensive cost functions. Procedia Comput. Sci. 18, 2337–2346 (2013).

35. Jung, M. et al. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. J. Geophys. Res. Biogeosci. 116, G00J07 (2011).

36. Jung, M. & FLUXCOM team FLUXCOM (RS+ METEO) global land carbon fluxes using CRUNCEP climate data. FLUXCOM Data Portal http://dx.doi.org/10.17871/FLUXCOM_RS_METEO_CRUNCEPv6_1980_2013_v1 (2016).

37. Friedl, M. A. et al. MODIS collection 5 global land cover: algorithm refinements and characterization of new datasets. Remote Sens. Environ. 114, 168–182 (2010).

38. New, M., Hulme, M. & Jones, P. Representing twentieth-century space-time climate variability. Part II: Development of 1901-96 monthly grids of terrestrial surface climate. J. Clim. 13, 2217–2238 (2000).

39. Wang, Y. P., Law, R. M. & Pak, B. A global model of carbon, nitrogen and phosphorus cycles for the terrestrial biosphere. Biogeosciences 7, 2261–2282 (2010).

40. Jain, A. K., Meiyappan, P., Song, Y. & House, J. I. CO2 emissions from land-use change affected more by nitrogen cycle, than by the choice of land-cover data. Glob. Change Biol. 19, 2893–2906 (2013).

41. Sitch, S. et al. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. Glob. Change Biol. 9, 161–185 (2003).

42. Smith, B., Prentice, I. C. & Sykes, M. T. Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space. Glob. Ecol. Biogeogr. 10, 621–637 (2001).

43. Krinner, G. et al. A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. Glob. Biogeochem. Cycles 19, GB1015, http://dx.doi.org/1010.1029/2003GB002199 (2005).

44. Kato, E., Kinoshita, T., Ito, A., Kawamiya, M. & Yamagata, Y. Evaluation of spatially explicit emission scenario of land-use change and biomass burning using a process-based biogeochemical model. J. Land Use Sci. 8, 104–122 (2013).

SUPPLEMENTARY INFORMATION

SI-1 Cross-consistency analysis of global GPP, TER, and NEE from FLUXCOM and TRENDY

We assessed the consistency of global integrated flux anomaly patterns (GPP, TER, and NEE) both between FLUXCOM and TRENDY means, as well as among individual ensemble members. Overall, the interannual variability (IAV) between the FLUXCOM and TRENDY means of GPP, TER, and NEE are largely consistent (Fig. SI-1.1) with correlations between 0.82 (for GPP) and 0.89 (for TER, see Fig. SI-1.2).

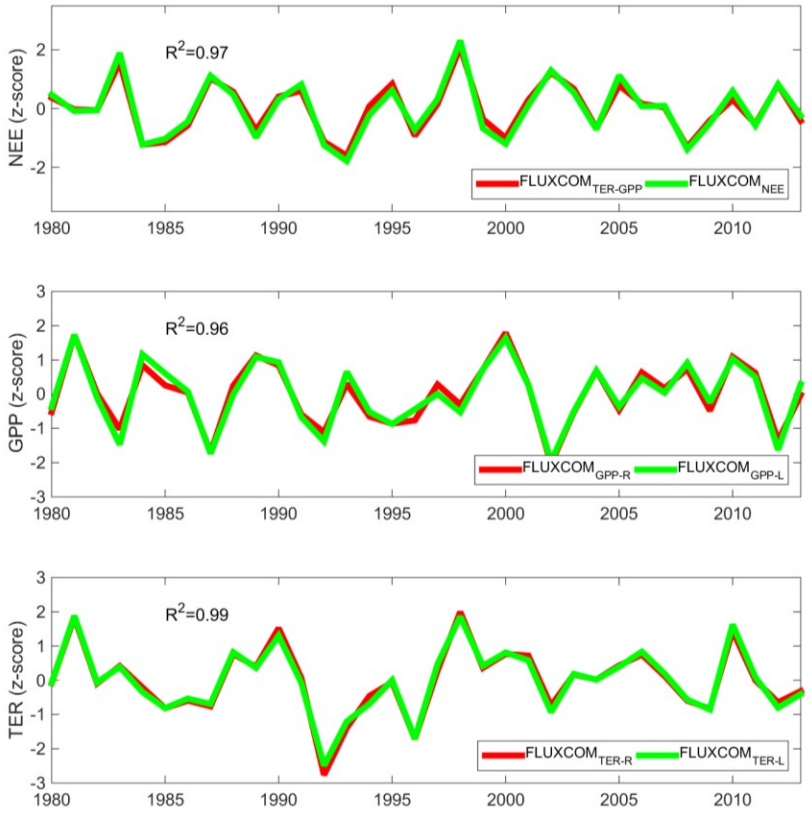


Figure SI-1.1: Time series of globally integrated flux anomalies (z-scores) of NEE, GPP, and TER. Thick solid lines refer to the ensemble mean for FLUXCOM and TRENDY; thin dotted lines show all individual ensemble members. A detailed assessment of correlations is presented in subsequent figures.

Global integrated NEE anomalies of FLUXCOM and TRENDY are more consistent with results from the Jena Inversion[24] (r=0.8-0.81) than with those from the residual land sink provided by the Global Carbon Project[16] (r=0.68-0.75). The correlation between the Jena Inversion results and residual land sink is the same (0.87) as between FLUXCOM and TRENDY NEE, highlighting that there remains uncertainty in the top-down derived global land sink IAV as well. The better correlation with the Jena Inversion is likely due to a consistent integration of atmospheric and ocean carbon measurements with atmospheric transport in a sophisticated inversion framework. The fact that both FLUXCOM and TRENDY do not account for land use change emissions may contribute to the remaining discrepancy with the atmospherically based estimates. TRENDY shows a slightly better correlation with the residual land sink (r=0.75) than FLUXCOM (r=0.68), perhaps because it has been a traditional target by modelers, which might have tuned their models intentionally or not.

By inspecting the spread of correlations among ensemble members within FLUXCOM or TRENDY, we found that ensemble members of FLUXCOM are more consistent with each other than among TRENDY models (Fig. SI-1.3). More importantly, we observed that the consistency of FLUXCOM vs TRENDY is comparable or better than the consistency among individual TRENDY models. All these results are encouraging and provide confidence in the patterns since both approaches are conceptually independent, with FLUXCOM being based on machine learning based extrapolations of FLUXNET measurements, and with TRENDY being based on process representations.

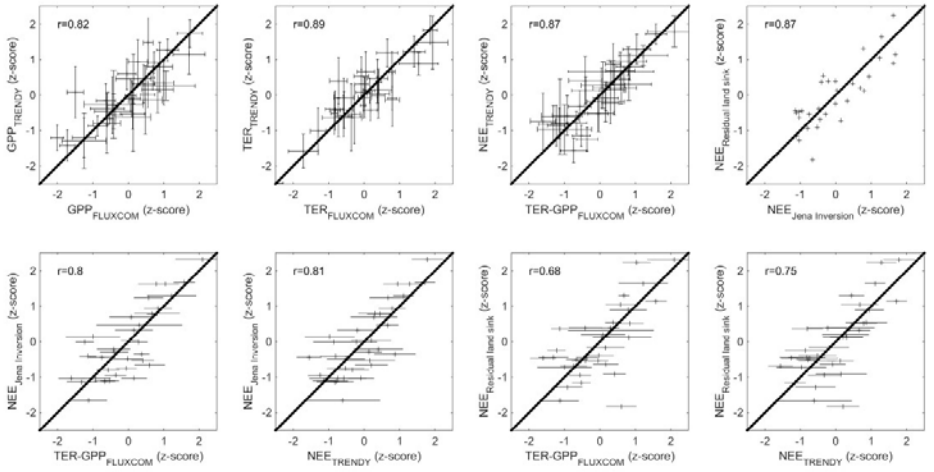


Figure SI 1.2: Correlations of globally integrated flux anomalies between FLUXCOM and TRENDY with independent, top-down estimates of global carbon balance IAV (residual land sink and Jena Inversion). Shown is the mean of the ensemble with error bars marking one standard deviation of the ensemble.

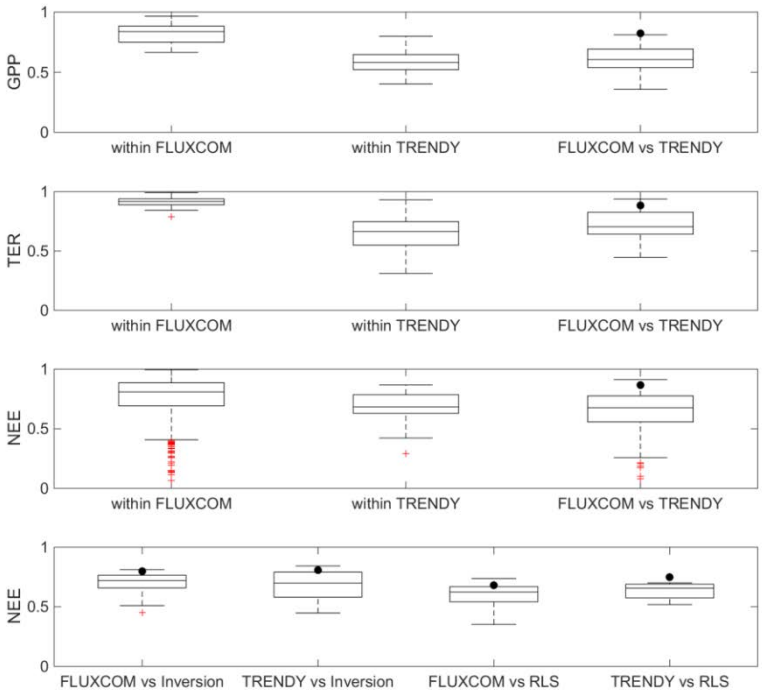


Figure SI-1.3: Box plots showing correlations between the global flux IAV of ensemble members. Black dots mark the correlation of the ensemble means.

SI-2 Evaluation of the decomposition of flux anomalies into their climatically driven components

The conclusions of this paper are substantiated by the robustness of the decomposition of the flux anomalies into components driven by radiation (RAD), temperature (TEMP), and water availability (WAI). The simple decomposition model (see Eq.1 in main article) assumes: a) the three selected climatic variables are sufficient to describe IAV of the fluxes; b) higher-order interaction terms among the three climatic variables are negligible; c) the function between climatic anomalies and carbon flux anomalies is linear. With respect to a) the assumption is valid for FLUXCOM, while for TRENDY models, additional model inputs (e.g. relative humidity, wind speed, etc) could generate IAV in the carbon fluxes which would not be explained by the decomposition model. In addition, dynamic effects could play a role in IAV, in particular for TER due to lagged effects from explicitly simulated carbon pools. With respect to b) and c), it is important to remember that the sensitivities are estimated for each month of a grid cell, and highly localized in time and space. This means that interactions with seasonally varying drivers are implicitly taken into account, and linearity between climatic anomalies and carbon flux anomalies is only assumed for a given month and location.

To test the above mentioned assumptions, we evaluated how well we can approximate the IAV of the carbon fluxes with the sum of the three components. We found that we can precisely reconstruct the flux anomalies from FLUXCOM with the three climatic components (Fig SI-2.1). The results for TRENDY (Fig. SI-2.2) are not as precise in comparison, but still very good, which allows at least for qualitative cross-consistency checks among FLUXCOM and TRENDY. We conclude that the simple linear, but highly localized, decomposition model provides acceptable results. Potentially, more accurate results could be obtained for TRENDY with more complex approaches, which, however, would be at the expense of less transparency and inconsistency with FLUXCOM, and is therefore not suitable here.

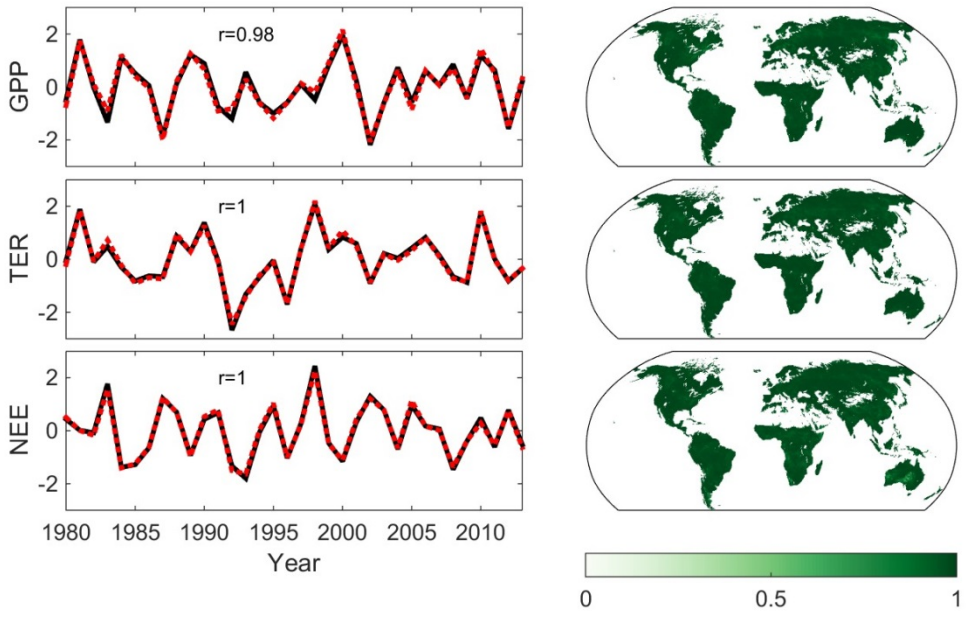


Figure SI-2.1: Left: Comparison of the original time series of global flux anomalies (z-scores, solid black line) from FLUXCOM with the sum of TEMP, WAI, and RAD components (dotted red line). Right: maps of the correlation coefficient between the original flux IAV and the sum of TEMP, WAI, RAD components for each pixel. Plotted is the mean across all ensemble members.

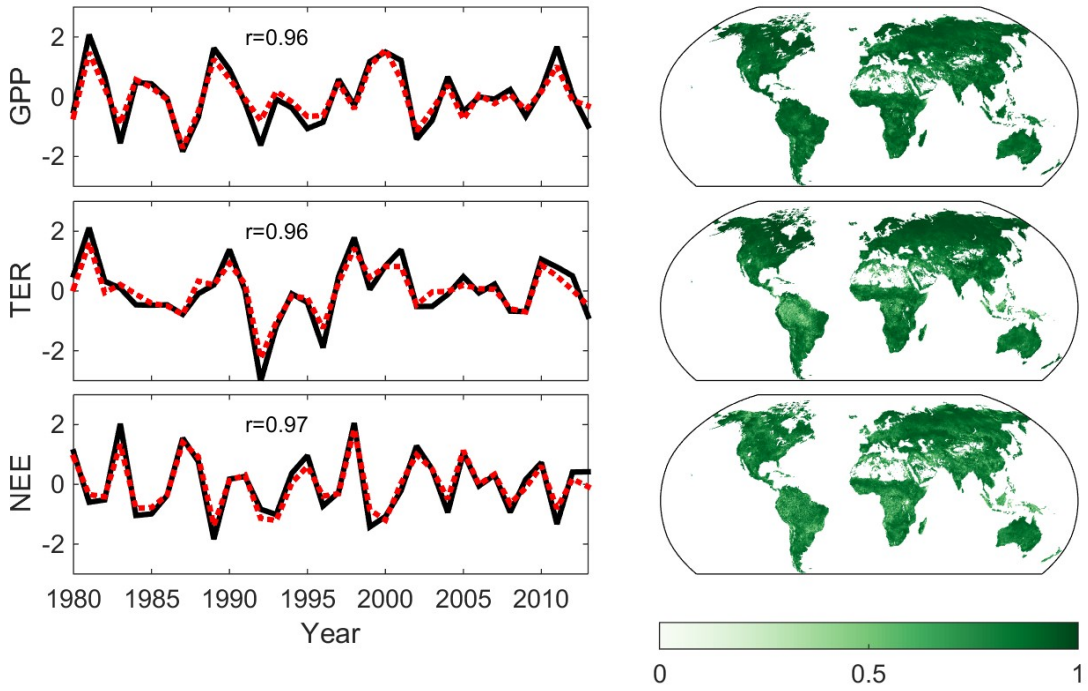


Figure SI-2.2: Same as Figure SI-2.1 but for TRENDY.

Another potential concern of the decomposition model is related to the identifiability of the individual coefficients, given that some degree of co-linearity is expected among anomalies of the three climatic variables. This plays an important role in some cases, e.g., where one is dealing with observational data of different origins and each variable is associated with uncertainty, where some inconsistencies among the different variables exist, or where only proxies of the proximate drivers are available (e.g. soil moisture vs precipitation). In such cases, sensitivities can become biased high for the larger apparent sensitivities associated with better observed variables, or biased low for 'proxy' variables. None of these concerns apply here, because we are dealing with a consistent set of model inputs and model outputs. Potential errors in model inputs would propagate to model outputs (see SI-3 for an assessment) but do not affect the models inherent sensitivities. We can obtain confidence in the derived sensitivities by confronting them against our general understanding of carbon cycle processes, and by using a qualitative comparison between FLUXCOM (Fig. SI-2.3) and TRENDY (Fig. SI-2.4) based sensitivities. The following paragraphs present and discuss mean patterns of various carbon flux sensitivities to temperature, water availability, and moisture for FLUXCOM and TRENDY ensembles. Please keep in mind, that the spread among individual ensemble members can be large.

Radiation. FLUXCOM GPP shows an ubiquitous positive sensitivity to radiation, with larger sensitivities in the tropics compared to the northern high latitudes, and with low sensitivities in subtropical regions. TRENDY GPP is qualitatively consistent with that pattern, except for weakly negative sensitivities in subtropical regions of the southern hemisphere. Because photosynthesis is driven by sun-light, a positive sensitivity of GPP to radiation was expected. The low sensitivity in subtropical regions is expected because in these areas photosynthesis should operate at light saturation and be primarily limited by water availability. The weak negative sensitivities in the southern subtropics in TRENDY might represent the effect of high leaf temperatures under dry and high radiation conditions, which decreases photosynthesis in the models. While FLUXCOM TER shows

almost no sensitivity to radiation, we observe some positive sensitivity in the inner tropics in TRENDY and weak negative sensitivities in the southern subtropics. We would expect that TER is either insensitive to radiation (FLUXCOM), or, that we observe similar, but weaker, patterns as those seen in the sensitivity of GPP to radiation (TRENDY), due to the coupling of TER and GPP. Because the sensitivity of GPP to radiation is much larger than that of TER to radiation, the sensitivity of NEE to radiation is dominated by that of GPP, but with a reversed sign. This results in a net negative sensitivity of NEE to radiation, i.e. increasing carbon uptake with increasing radiation (particularly in the tropics), which is qualitatively consistent among FLUXCOM and TRENDY.

Temperature. The patterns of the sensitivity of GPP to temperature is dominated by negative values in the tropics in both FLUXCOM and TRENDY, i.e. decreasing photosynthesis with increasing temperature, indicating the optimal temperatures for photosynthesis are exceeded in these areas. FLUXCOM GPP shows only a weak positive sensitivity to temperature during the shoulder seasons in the northern hemisphere, when photosynthesis is often assumed to be limited by cold temperatures. This pattern of positive sensitivities of GPP to temperature in high-latitudes is stronger in the TRENDY models, perhaps because these models do not account for temperature acclimation of photosynthesis. As expected, we observe a ubiquitous positive sensitivity of TER to temperature in FLUXCOM and TRENDY, with lower sensitivities in the subtropics where respiration processes are primarily driven by moisture effects. Consequently, NEE shows a positive sensitivity to temperature in both FLUXCOM and TRENDY (increasing carbon loss with increasing temperature). In tropical regions, both FLUXCOM and TRENDY show an amplified sensitivity of NEE to temperature because the sensitivities of GPP and TER to temperature are of opposite sign, and therefore add up.

Water availability. As expected, both FLUXCOM and TRENDY show ubiquitous positive sensitivities of GPP and TER to water availability, with the largest sensitivities in subtropical (semi-arid) regions, and with weaker sensitivities of TER to WAI than GPP to WAI. FLUXCOM also shows a sizeable sensitivity of both GPP and TER to WAI in northern mid-latitudes in summer, while in TRENDY models this pattern is weaker for TER. The sensitivity of NEE to WAI is negative (i.e. increasing carbon uptake with increasing water availability) in those regions where GPP also shows large sensitivity to water availability, but the effect is less pronounced because the sensitivities of GPP and TER to WAI are in the same direction, and thus compensate. TRENDY even shows a weak positive sensitivity of NEE to WAI in norther high latitudes in summer, because TER is more sensitive to WAI than GPP in these models.

These results suggest that the extracted sensitivities are very plausible, with some interesting differences between FLUXCOM (data-driven) and TRENDY (process-models). The robustness of the identifiability of the sensitivities can be further illustrated by inspecting the sensitivities of GPP to temperature and radiation. Temperature and radiation are generally positively correlated, yet the extracted sensitivities of GPP to each are of opposite sign in the tropics (consistent with our understanding).

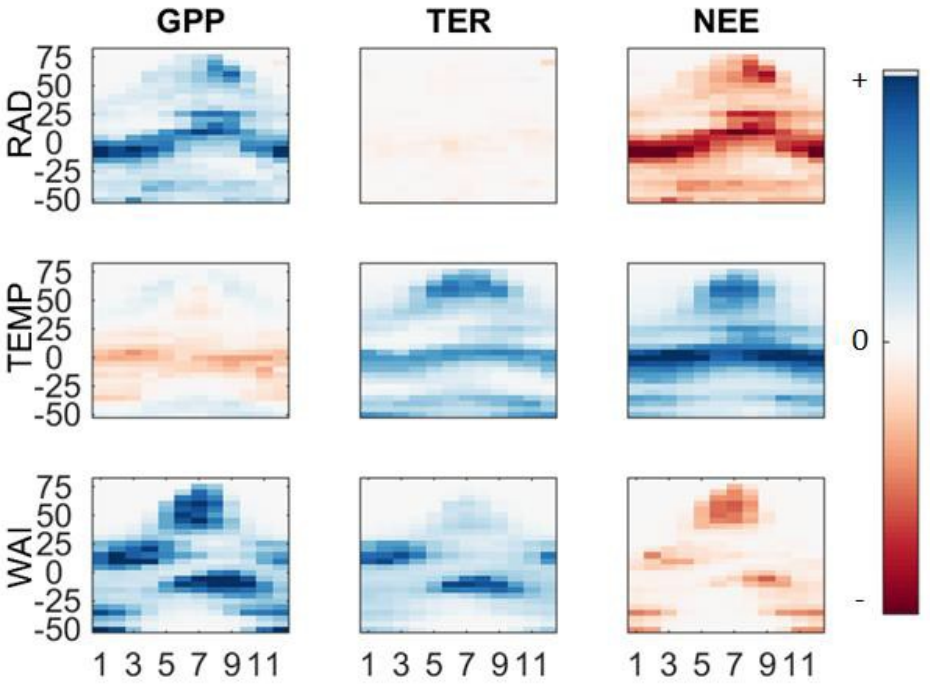


Figure SI 2.3: Sensitivities of GPP, TER, and NEE, to RAD, TEMP, and WAI as a function of latitude (y-axis) and month (x-axis) for FLUXCOM. We first obtained the median sensitivity for each 5 degree latitudinal band and month for each ensemble member. The sensitivities of each ensemble member are then scaled with the maximum absolute sensitivities of all three carbon fluxes for each climatic driver. We then took the median of ensemble members for each bin. The scaling of the sensitivities allows for a comparison of sensitivities between GPP, TER, and NEE for the same climatic driver, and was necessary because TRENDY models used different definitions and units of soil moisture.

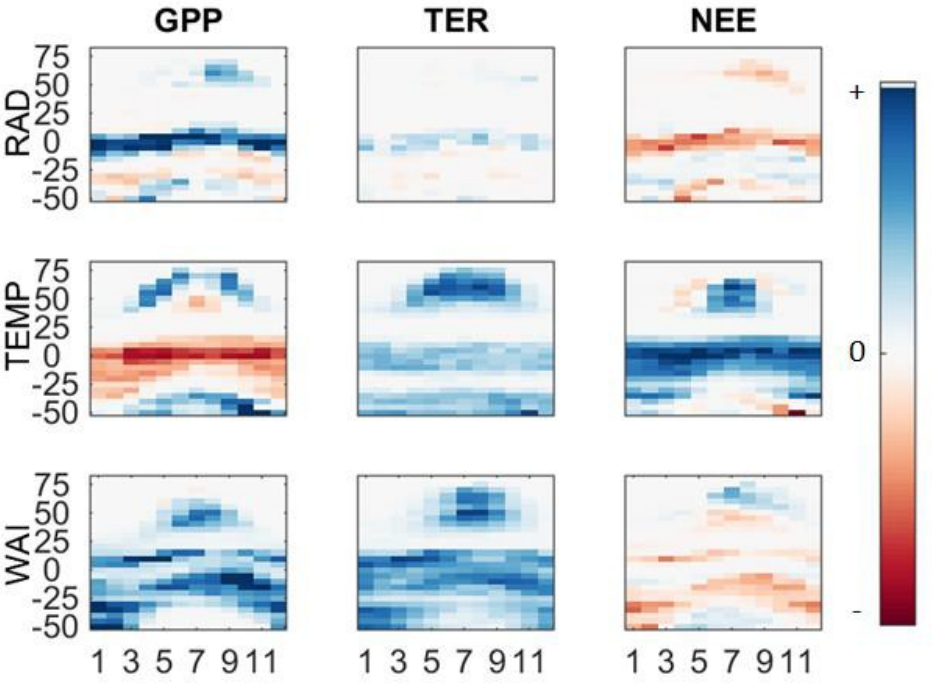


Figure SI 2.4: Same as Figure SI 2.3 but for TRENDY.

SI-3 Effect of different climate forcing data sets on FLUXCOM results

There is no perfect global gridded climate forcing data set. The CRUNCEPv6 product has a regionally highly variable station data support and the uncertainties of the climatic variables themselves are not well known. In general, uncertainty of these model input data are a widely known source of uncertainty of model outputs. Therefore, we conducted additional FLUXCOM runs with two alternative daily climate forcing data sets: 1) based on the widely used WATCH FORCING DATA ERA-INTERIM (WFDEI[45]) for the period 1980-2013, where biases of the reanalysis were corrected based on gridded meteorological station data, and 2) a combined product of satellite based precipitation from GPCP[46], satellite based radiation from CERES[47] (SYN1deg), and air temperature from CRUNCEP for the period 2001-2013 (limited in length by CERES products). We refer to the latter forcing (2) as 'CERES' in the following. We processed and analysed these additional runs in exactly the same way as the CRUNCEP runs presented in the main article (see Methods).

Confronting globally integrated NEE (TER-GPP, mean across ensemble) of WFDEI and CERES based forcing with the top down atmospheric constraint from the Jena Inversion (cf EDF1), we find that the results of both forcing data sets are inferior when compared to NEE based on CRUNCEP. The correlation between the Jena Inversion and the WFDEI based NEE is only 0.65 at global scale for the period 1980-2013, compared to 0.8 for the CRUNCEP based NEE over the same period. For the period 2001-2013, the correlation with the CERES based NEE is 0.76 compared to 0.87 for CRUNCEP based forcing.

The qualitative picture that NEE^{WAI} dominates NEE IAV across grid cells ("locally") on the one hand, and that the strongest correlations of globally integrated NEEs are obtained with globally integrated NEE^{TEMP} on the other hand, remains consistent with CRUNCEP based forcing (compare Fig SI-3.1 with Fig.1 in main article). This consistency is perhaps even more striking, as NEE^{TEMP} shows the least variability locally in WFDEI and CERES based results. The IAV of NEE^{RAD} is found to exceed that of NEE^{TEMP} locally for WFDEI and CERES forcing, which was not the case for CRUNCEP based runs. This might suggest that the role of radiation might be underestimated in CRUNCEP based results because of lower radiation interannual variability in the CRUNCEP data. It is, however, not conclusive, because CRUNCEP based globally integrated NEE compares better with the independent atmospherically based NEE.

Importantly, the local dominance of water availability for both GPP and TER IAV is consistent between results obtained from all three forcing data sets (Fig. SI-3.2). The same holds for the compensation effect between GPP^{WAI} and TER^{WAI}. There are some differences among results from different forcing data sets with respect to latitudinal shapes of IAV magnitude. For example, CERES based results show a double-peak at subtropical bands which is not evident in CRUNCEP and WFDEI based results, perhaps also because CERES results are based on a shorter time period (2001-2013). Overall, uncertainty of climate input data sets, in particular with respect to radiation and precipitation, still constitute an important source of uncertainty in carbon cycle modelling. Nevertheless, the key findings of the paper and the qualitative picture are robust to multiple climate input data sets.

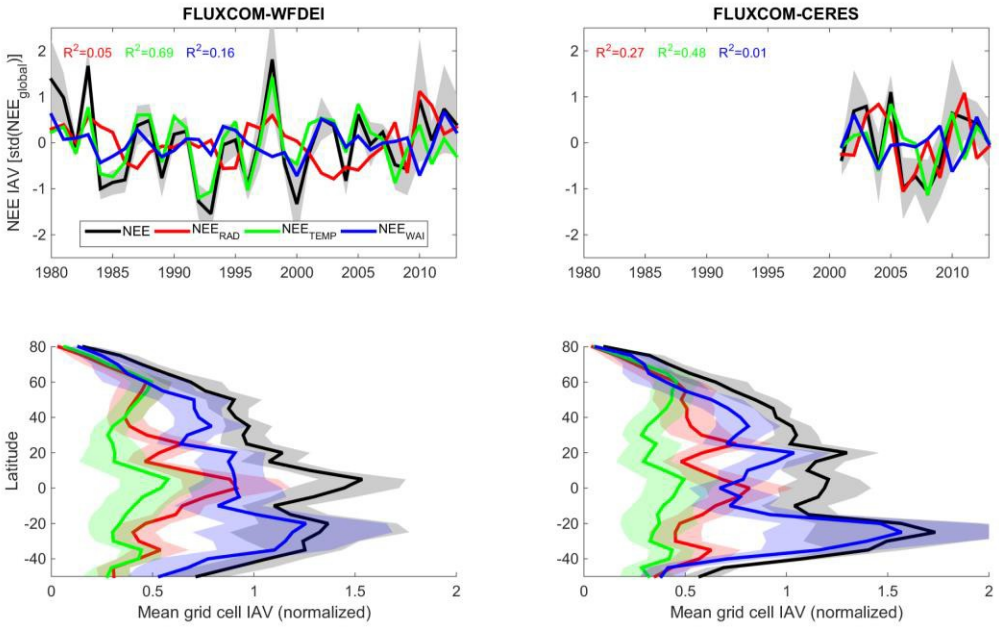


Fig. SI-3.1: same as Fig.1 in main article but for FLUXCOM forced by WFDEI climate forcing for the period 1908-2013 (left) and FLUXCOM forced with CERES radiation and GPCP precipitation for the period 2001-2013 (right).

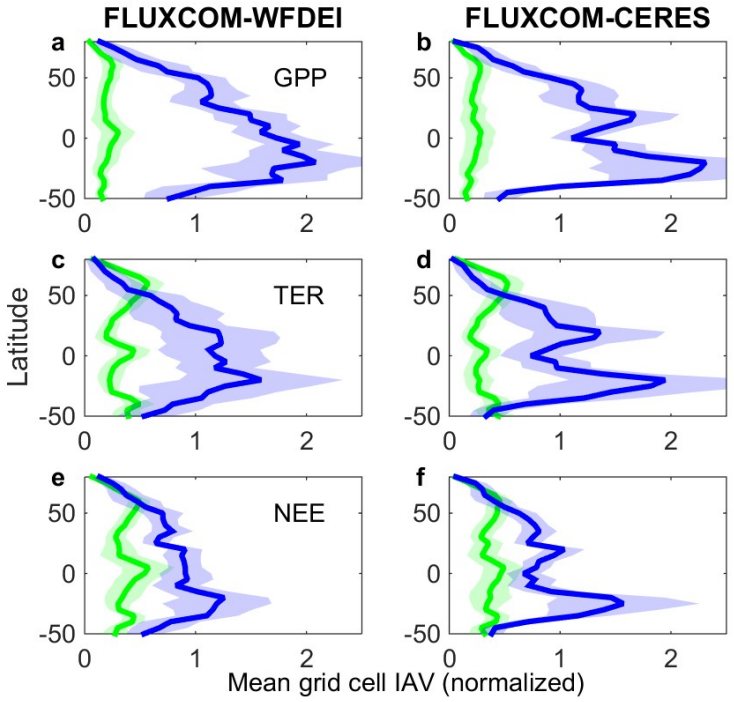


Fig. SI-3.2: same as Fig.3 in main article but for FLUXCOM forced by WFDEI climate forcing for the period 1908-2013 (left) and FLUXCOM forced with CERES radiation and GPCP precipitation for the period 2001-2013 (right).

SI-4 Regional analysis

We summarize the main findings of our study for the major biomes defined by Ahlström et al.[4] but omitted the class 'sparsely vegetated' due to its small relevance for global carbon flux IAV. The key finding that the dominance of NEE^{WAI} decays with increasing spatial aggregation, and that the dominance of NEE^{TEMP} rises with increasing spatial aggregation is evident in all the biomes for both FLUXCOM and TRENDY (Fig. SI-4.1). Differences among biomes are mainly evident as "offsets" of the NEE^{WAI} and NEE^{TEMP} curves. For example, in tundra & arctic shrublands as well as in extra-tropical (mainly boreal) forests NEE^{TEMP} and NEE^{WAI} are roughly equally dominant at local scale (leftmost data point), while for example in the semi-arid, and grass- & cropland regions, NEE^{WAI} is more dominant at local scale. The largest differences between FLUXCOM and TRENDY are related to the larger dominance of NEE^{WAI} in semi-arid, grass- and cropland regions for FLUXCOM compared to TRENDY, but the individual uncertainties (spread among ensemble members) is also large.

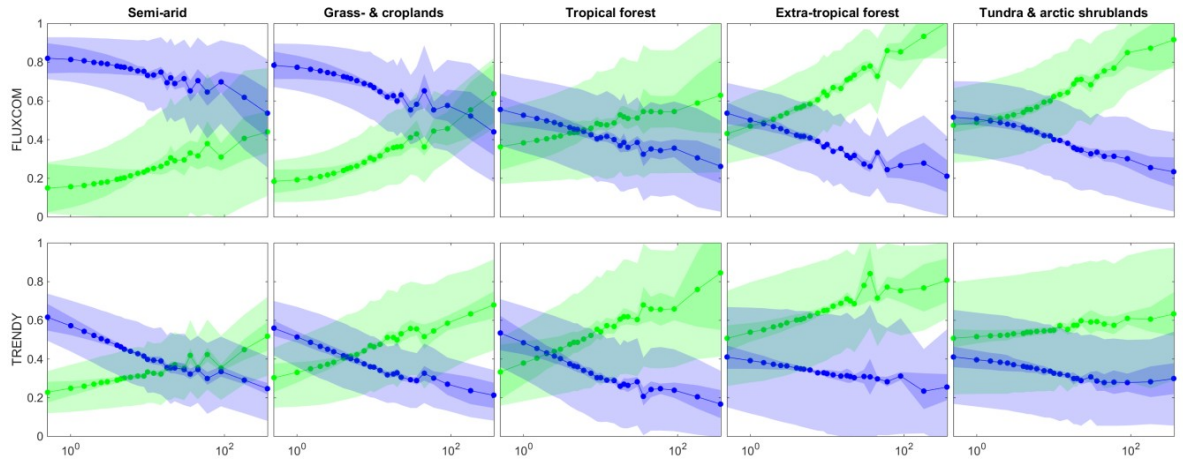


Fig. SI-4.1: Scale dependence (spatial resolution of grid cells (°), x-axis) of NEE^{WAI} and NEE^{TEMP} dominance (y-axis) for different biomes (cf. Fig.2 e,f in main article).

WAI stands out as the most important 'local' factor causing GPP IAV in all biomes for FLUXCOM (Fig. SI-4.2a), while TRENDY shows TEMP as the dominant local control for extra-tropical forests and tundra (Fig. SI-4.3). The latter two biomes also show TEMP as the dominant source of TER IAV, while all other biomes show WAI as the dominant source of TER IAV. This is consistent between FLUXCOM and TRENDY (compare Fig. SI-4.2b and Fig. SI-4.3b).The compensation effect between GPP^{WAI} and TER^{WAI} is also evident in all biomes and can be seen as a decrease in area of the blue polygon in Fig.SI-4.2c for NEE compared to those in Fig.SI-4.2a,b for the gross fluxes. The effect of different spatial compensation of NEE^{WAI} and NEE^{TEMP} anomalies can be seen by comparing the top and bottom rows in Fig. SI-4.2 and SI-4.3, where the green polygon (NEE^{TEMP}) is expanded with respect to the blue polygon (NEE^{WAI}) in the bottom row, showing an increase in IAV magnitude for the spatially integrated NEE^{TEMP} component (e.g. compare SI-4.2c and SI-4.2f). One major difference between FLUXCOM and TRENDY is that TEMP tends to be more important in the TRENDY spatially integrated GPP IAV (compare SI-4.2d and SI-4.3d) due to a larger temperature sensitivity of GPP for these regions in TRENDY compared to FLUXCOM (see also SI-2). Whether this is due to the missing process of temperature acclimation of photosynthesis in TRENDY models remains speculative.

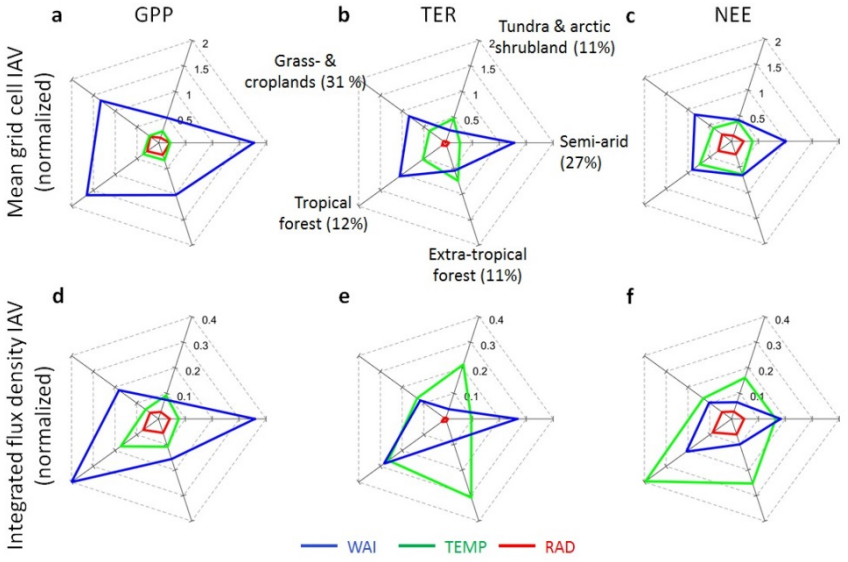


Fig. SI-4.2: Interannual variability of WAI, TEMP, and RAD components of GPP, TER, and NEE for different biomes based on FLLUXCOM. a-c: Mean grid cell IAV magnitude (see Equation 3 in Methods). d-f: IAV of the integrated flux component over that region defined as standard deviation of annual sums (normalized by the same denominator as in Equation 3 in Methods). Percentages refer to the fractional cover of the global vegetated area. Uncertainties were omitted for clarity.

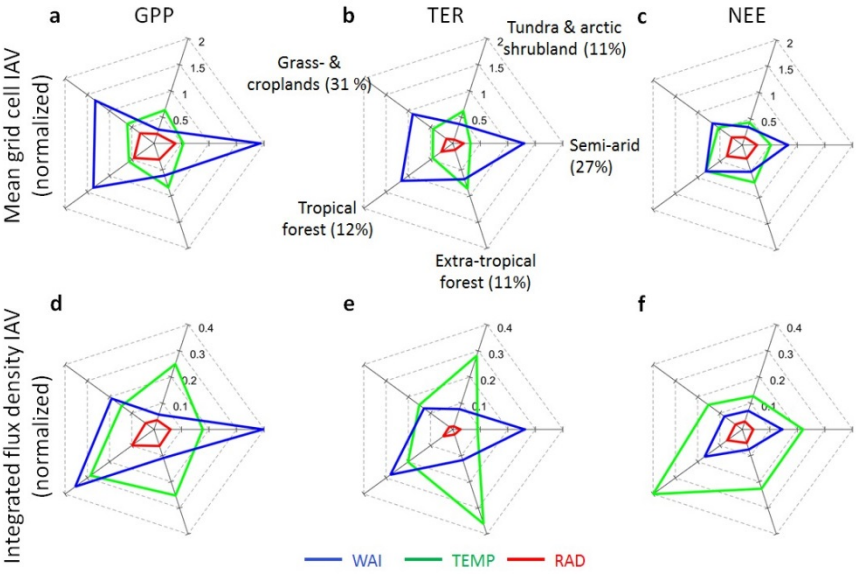


Fig. SI-4.3: Same as Fig.SI-4.2 but based on TRENDY.

SI-5 Spatial patterns and uncertainty maps of temperature and water availability driven components of GPP, TER, and NEE

Here we provide spatial patterns of IAV magnitude of the TEMP and WAI component of GPP, TER, and NEE. These form the basis for Fig. 1c,d and Fig. 3 in the main article, which summarize the means across latitudinal bands.

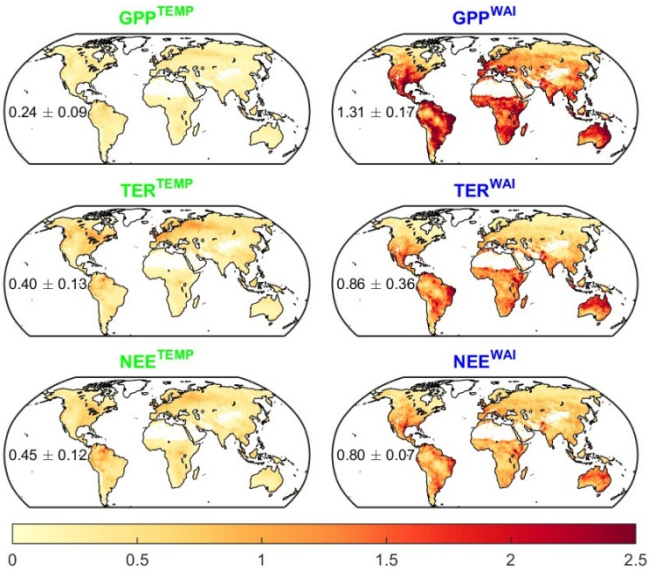


Fig. SI-5.1: Spatial patterns of IAV magnitude of the TEMP and WAI components of GPP, TER, and NEE for FLUXCOM. IAV magnitude is normalized (see Equation 3 in Methods). Numbers in the panels refer to the mean across all grid cells ± 1 s.d. of ensemble members (i.e. using the notations in the Methods section: $\mu_e(\mu_s(M_{s,e}))$ and σ_e $(\mu_s(M_{s,e}))$).

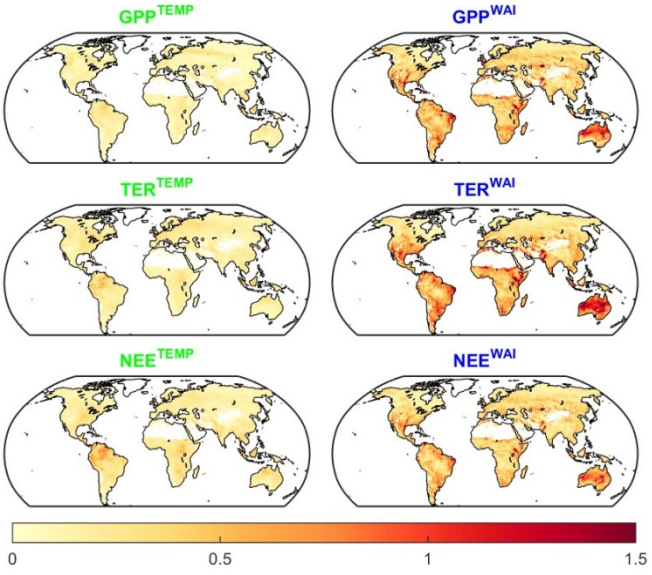


Fig. SI-5.2: Corresponding map of uncertainty (1 standard deviation across ensemble members) for FLUXCOM. Please note the different scales between this figure and Fig.SI-5.1.

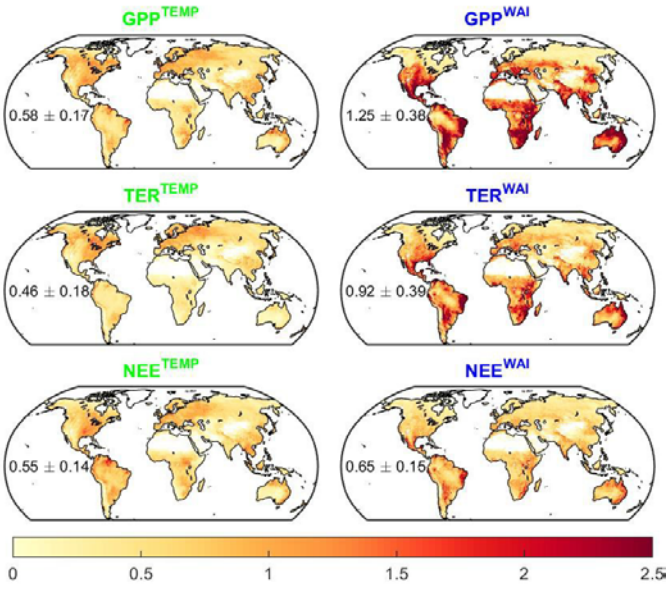


Fig. SI-5.3: Same as Fig. SI-5.1 but for TRENDY.

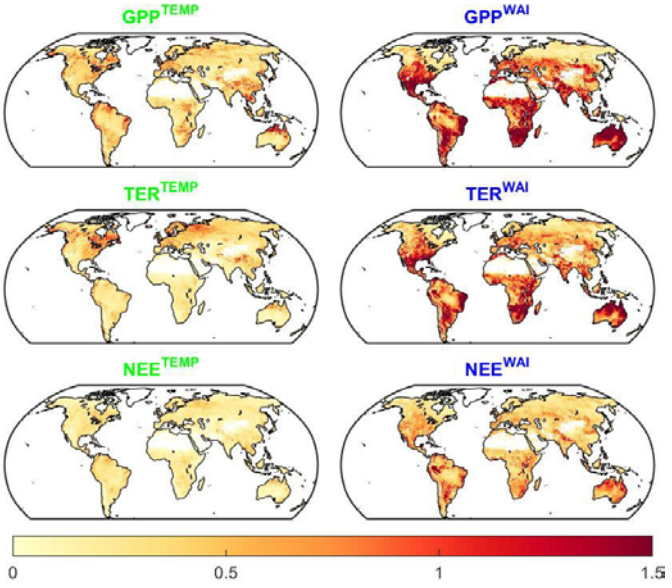


Fig. SI-5.4: Same as Fig. SI-5.2 but for TRENDY.

SI-6 Robustness of FLUXCOM results when using different NEE flux partitioning methods

One potential concern about the validity of FLUXCOM based IAV patterns of TER, GPP, and NEE computed as the difference of the two is related to the fact that GPP and TER are not directly measured. GPP and TER are derived from half-hourly NEE measurements using a model with parameters which are estimated in moving windows of a few days (i.e. highly localized). To test whether this ”flux partitioning” concern is justified, we compare NEE IAV derived from upscaling NEE directly with NEE derived as the difference between TER and GPP. Additionally, we check the consistency of GPP and TER variants based on independent flux partitioning methods (Fig. SI-6.1, SI-6.2). The two flux partitioning approaches by Reichstein et al.[22] and Lasslop et al.[21] use different underlying models, emphasizing either nighttime data to derive TER, or daytime data to derive GPP respectively. Clearly, the results based on independently derived GPP and TER values are highly consistent, and, most importantly, the IAV of upscaling NEE directly is consistent with the IAV derived from first upscaling TER and GPP independently and then taking the difference.

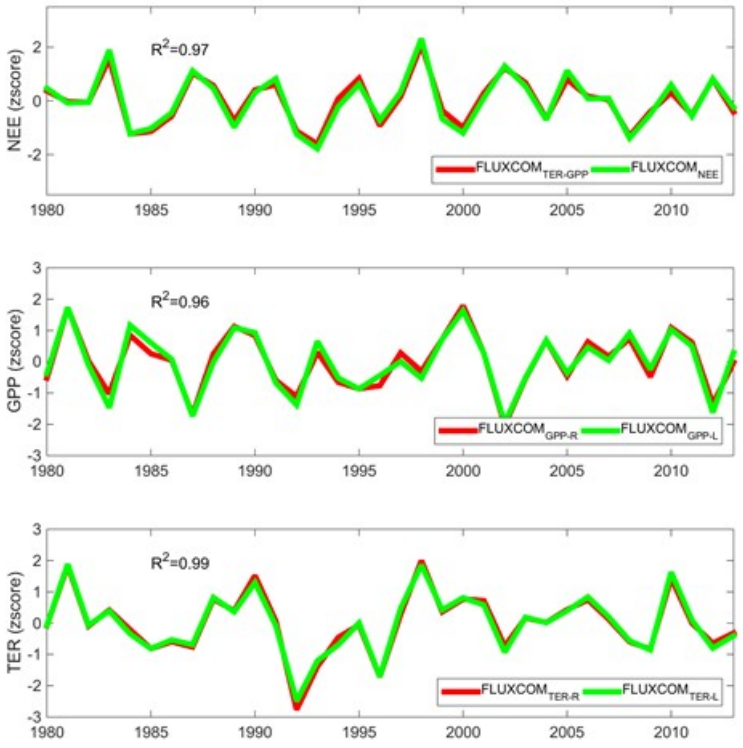


Fig. SI-6.1: Comparison of globally integrated fluxes (FLUXCOM) between different methods to partition measured NEE into GPP and TER at flux towers. The top panel shows a comparison between NEE being directly upscaled with NEE computed as the difference between TER and GPP (used in this paper). Lower panels show comparison between flux partitioning of Reichstein et al (‘R’) and Lasslop et al (‘L’). Presented are z-scores of mean values of different machine learning algorithms.

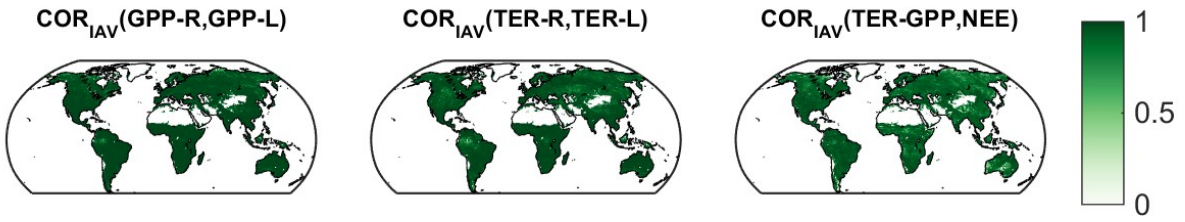


Fig. SI-6.2: Correlation maps for interannual variations of GPP, TER, and NEE originating from different flux partitioning methods. For GPP and TER, flux partitioning of Reichstein et al is labelled as 'R' and Lasslop et al as 'L). The correlation map for NEE refers to the comparison of computing NEE as TER-GPP vs the direct upscaling of NEE.

In addition, we investigated if seasonal variations of GPP based on the two flux partitioning methods show differences consistent with values based on sun induced fluorescence (SIF) from GOME-2. SIF is a space-based measurement closely related to photosynthesis. We observe a striking consistency between FLUXCOM based GPP and SIF (Fig. SI-6.3), with no differences among flux partitioning methods. We find substantially better agreement between FLUXCOM GPP and SIF than between TRENDY GPP and SIF, in particular for semi-arid regions. This is particularly encouraging because these regions are relatively poorly covered by the FLUXNET measurements used to train the machine learning models, which lends confidence to the extrapolation capacity of the FLUXCOM approach. Low correlations between GPP and SIF in the inner tropics are primarily due to low seasonal variability there. Furthermore, frequent cloud cover and associated contamination of SIF likely contributes to these low correlations as well.

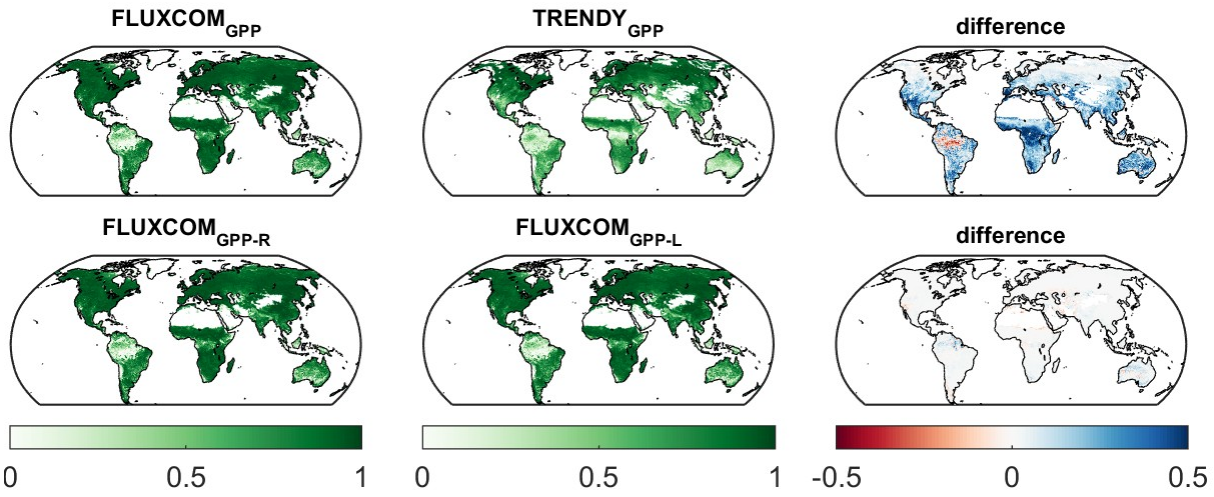


Fig. SI-6.3: Maps of R^2 between GPP and sun-induced fluorescence from GOME-2[48]. Top panel compares mean R2 values between FLUXCOM GPP estimates and GPP simulated by TRENDY models. Bottom panel compares FLUXCOM GPP based on two different flux partitioning methods (Reichstein et al ('R') and Lasslop et al ('L')). R^2 values are based on a monthly climatology of GPP and SIF between 2008-2013.

For completeness, we verified that the covariance patterns of the temperature and water availability components of GPP and TER (see Fig.4 in main article) are robust with respect to flux partitioning. Clearly, the results are highly consistent among different combinations of GPP and TER variants (Fig. SI-6.4). All results presented here show that the uncertainty of IAV patterns related to flux

partitioning seem negligible, which in turn suggests that the spread among FLUXCOM ensemble members is driven by different machine learning methods.

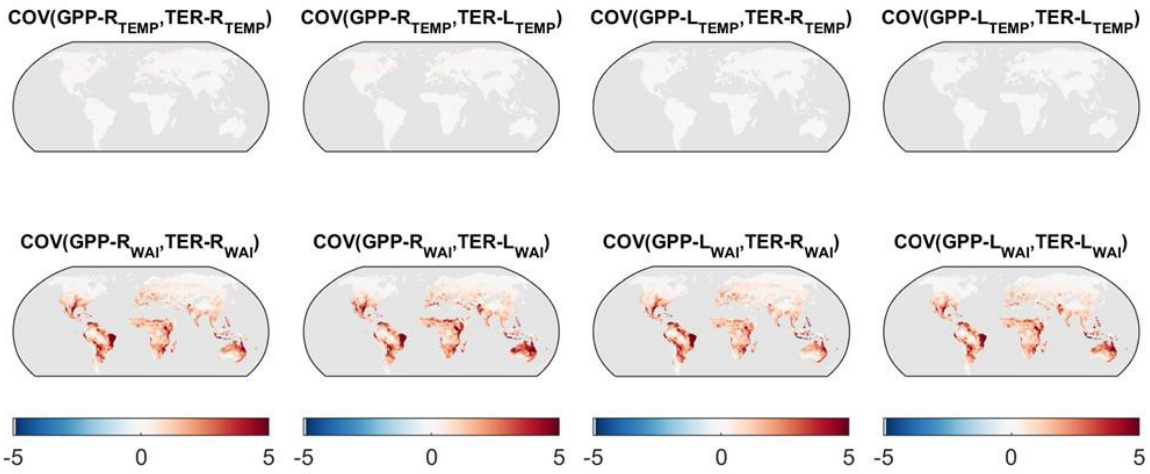


Fig. SI-6.4.: Maps of interannual covariance between GPP^{TEMP} and TER^{TEMP} (top) and between GPP^{WAI} and TER^{WAI} (bottom) of FLUXCOM (see Fig. 4) for different combinations of GPP and TER estimates due to different flux partitioning methods (Reichstein et al ('R') and Lasslop et al ('L')).

SI-7 Predictor variables of FLUXCOM

Type of variability	Predictor variable	Predictor type
Spatial, seasonal, interannual	Air temperature	Climate
	Water availability of lower layer	Climate
	Product of global shortwave radiation and normalized difference vegetation index*	Climate & satellite
Spatial, seasonal	Product of potential shortwave radiation and enhanced vegetation index*	Climate & Satellite
	Product of fraction of absorbed photosynthetic active radiation* and daytime land surface temperature*	Satellite
	Night-time land surface temperature*	Satellite
Spatial	Minimum of normalized difference water index*	Satellite
	Amplitude of red reflectance*	Satellite
	Amplitude of normalized difference vegetation index*	Satellite
	Amplitude of water availability of lower layer*	Climate
	Plant functional type	Satellite

Table SI-7: Predictor variables of machine learning based FLUXCOM models. All variables marked with an asterisk (*) are based on a mean seasonal cycle for the period 2001-2010; Satellite based predictors are derived from MODIS land products (see Tramontana et al.[15] for details).

SI-8: Cross-consistency between soil moisture simulated by TRENDY models and the water availability index

To verify the appropriateness of the simple water availability index (WAI) to represent soil moisture dynamics, we calculated pixel wise correlations between annual anomalies of WAI with annual anomalies of simulated soil moisture from TRENDY models. To put these correlations in perspective, we calculate the same type of correlations between the soil moisture values from the different TRENDY models. Fig. SI-8.1 shows that WAI IAV is as well correlated with simulated soil moisture IAV as the simulated soil moisture is correlated among different TRENDY models. This suggests that WAI IAV is a suitable surrogate for soil moisture IAV.

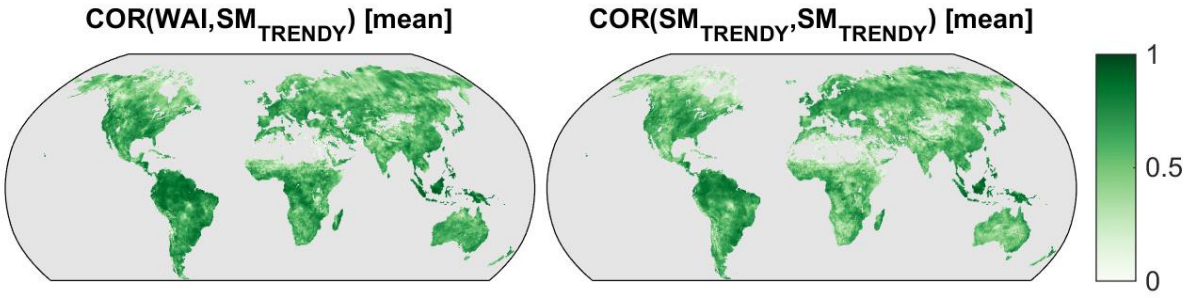


Fig. SI-8.1: Maps of correlation between annual WAI anomalies and annual soil moisture anomalies by TRENDY models (left). Shown is the mean correlation for all TRENDY models. The right map shows the mean correlation of TRENDY based soil moisture (mean of all possible combinations of TRENDY models).

References

45 Weedon, G. P. *et al.* The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resources Research* **50**, 7505-7514, doi:10.1002/2014wr015638 (2014).

46 Huffman, G. J. *et al.* Global Precipitation at One-Degree Daily Resolution from Multisatellite Observations. *Journal of Hydrometeorology* **2**, 36-50, doi:doi:10.1175/1525-7541(2001)002<0036:GPAODD>2.0.CO;2 (2001).

47 Doelling, D. R. *et al.* Geostationary Enhanced Temporal Interpolation for CERES Flux Products. *J. Atmos. Ocean. Technol.* **30**, 1072-1090, doi:doi:10.1175/JTECH-D-12-00136.1 (2013).

48 Köhler, P., Guanter, L. & Joiner, J. A linear method for the retrieval of sun-induced chlorophyll fluorescence from GOME-2 and SCIAMACHY data. *Atmos. Meas. Tech.* **8**, 2589-2608, doi:10.5194/amt-8-2589-2015 (2015).