The decadal state of the terrestrial carbon cycle: global retrievals of terrestrial carbon allocation, pools and residence times

Citation for published version:

Digital Object Identifier (DOI):
10.1073/pnas.1515160113

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Proceedings of the National Academy of Sciences

Publisher Rights Statement:
http://www.sherpa.ac.uk/romeo/issn/1091-6490/

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
The decadal state of the terrestrial carbon cycle: global retrievals of terrestrial carbon allocation, pools and residence times

A. Anthony Bloom$^{1,2,3}$, Jean-François Exbrayat$^{2,3}$, Ivar van der Velde$^4$, Liang Feng$^{2,3}$, Mathew Williams$^{2,3}$

1. Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
2. School of GeoSciences, University of Edinburgh, Edinburgh, UK
3. National Centre for Earth Observation, Edinburgh, UK
4. Wageningen University, Wageningen, The Netherlands

Corresponding author: A. Anthony Bloom (abloom@jpl.nasa.gov)

Biological Sciences: Environmental Sciences

Abstract

The terrestrial carbon cycle is currently the least constrained component of the global carbon budget. Large uncertainties stem from a poor understanding of plant carbon allocation, stocks, residence times and carbon use efficiency. Imposing observational constraints on the terrestrial carbon cycle and its processes is therefore necessary to better understand its current state and to predict its future state. We combine a
diagnostic ecosystem carbon model with satellite observations of leaf area and biomass (where and when available) and soil carbon data to retrieve the first global estimates of carbon cycle state and process variables at a 1°×1° resolution; retrieved variables are independent from the plant functional type and steady-state paradigms. Our results reveal global emergent relationships in the spatial distribution of key carbon cycle states and processes. Live biomass and dead organic carbon residence times exhibit contrasting spatial features ($r=0.3$). Allocation to structural carbon is highest in the wet tropics (85–88%) in contrast to higher latitudes (73–82%), where allocation shifts towards photosynthetic carbon. Carbon use efficiency is lowest (0.42–0.44) in the wet tropics. We find an emergent global correlation between retrievals of leaf mass per leaf area and leaf lifespan ($r=0.64–0.80$) that matches independent trait studies. We show that conventional land-cover types cannot adequately describe the spatial variability of key carbon states and processes (multiple correlation median: 0.41). This mismatch has strong implications for the prediction of terrestrial carbon dynamics, which is currently based on globally applied parameters linked to land-cover or plant functional types.

**Significance**

Quantitative knowledge of terrestrial carbon pathways and processes is fundamental for understanding the biosphere’s response to a changing climate. Carbon allocation, stocks and residence times together define the dynamic state of the terrestrial carbon cycle. These quantities are difficult to measure and remain poorly quantified on a global scale.
Here we retrieve global 1°×1° carbon state and process variables by combining a carbon balance model with satellite observations of biomass and leaf area (where and when available) and global soil carbon data. Our results reveal emergent continental-scale patterns and relationships between carbon states and processes. We find conventional land-cover types cannot capture continental-scale variations of retrieved carbon variables: this mismatch has strong implications for terrestrial carbon cycle predictions.

© 2015. All rights reserved.

Keywords: carbon cycle, LAI, biomass, soil carbon, model-data fusion, allocation, residence time.

Introduction

The terrestrial carbon (C) cycle remains the least constrained component of the global C budget (1). In contrast to a relatively stable increase of the ocean CO$_2$ sink from 0.9 Pg C yr$^{-1}$ to 2.7 Pg C yr$^{-1}$ over the past 40 years, terrestrial CO$_2$ uptake has been found to vary between a net 4.1 Pg C yr$^{-1}$ sink to a 0.4 Pg C yr$^{-1}$ source, and accounts for a majority of the inter-annual variability in atmospheric CO$_2$ growth. The complex response of terrestrial ecosystem CO$_2$ exchanges to short- and long-term changes in temperature, water availability, nutrient availability and rising atmospheric CO$_2$ (2–6) remain highly uncertain in C cycle model projections (7). As a result, there are large gaps in our
understanding of terrestrial C dynamics, including the magnitude and residence times of
the major ecosystem C pools (8, 9) and rates of autotrophic respiration (10). Moreover,
the impact of climatic extremes on C cycling, such as recent Amazon droughts (11),
highlights the importance of understanding the terrestrial C cycle sensitivity to climate
variability. To understand terrestrial CO₂ exchanges in the past, present and future, we
need to better constrain current dynamics of ecosystem C cycling, from regional to
global scales.

C uptake, allocation, pool stocks, residence times, respiration and disturbance together
drive net CO₂ exchanges (12) on sub-daily to millennial timescales; these C state and
process variables also determine the temporal sensitivity of the net C balance to climatic
variability. For example, global changes in photosynthetic uptake could lead to a rapid
response from short-lived C pools (such as foliage, fine roots and litter), or to a
prolonged response from the long-lived C pools (such as woody biomass and soil C),
with very different outcomes on ecosystem source/sink behavior. Quantitative
knowledge of terrestrial C pathways is therefore central to understanding the temporal
responses of the major terrestrial C fluxes – including heterotrophic respiration (13),
fires (14, 15) and wetland CH₄ emissions (16, 17) – to inter-annual variations in C uptake.

While C dynamics have been extensively measured and analyzed at site-level (18– 21),
the respiration and allocation of fixed C, and its residence time within the major C pools,
are difficult and expensive to measure at site level, and remain poorly quantified on
global scales. As a result, global terrestrial C cycle models rely on land-cover type specific C cycling parameters – based on spatially pre-assigned plant functional types – to determine C fluxes and C pools (22). Globally spanning C cycle observations can provide a much-needed constraint on the spatial variability and associated dynamics of the terrestrial C cycle. Over the past decade a growing number of datasets have enhanced understanding of the terrestrial C cycle, including global scale canopy dynamics (NASA Moderate Resolution Imaging Spectroradiometer – MODIS leaf area index – LAI – and burned area products), empirically derived global soil C data (Harmonized World Soil Database – HWSD, 23), satellite-based above and below ground biomass maps for the tropics (ABGB, 24, 25), and Greenhouse Gases Observing Satellite (GOSAT) CO₂ and plant fluorescence (26, 27). These spatially and temporally explicit datasets provide an enhanced view of the terrestrial C cycle, and can be used together to retrieve consistent global C state and process variables. Significant efforts in data-driven estimates of the global C fluxes have been made over the past decade. These include estimates based on atmospheric CO₂ concentrations (1, 28, 29); high-resolution global primary production maps (30) based on FLUXNET eddy covariance tower datasets (18); the mean residence time of terrestrial C (31); ecosystem respiration dependence on temperature, based on FLUXNET data (32) and global C cycle data assimilation systems (33).

Given an increasing number of C cycle observations, what remains an outstanding challenge is to produce a data-consistent analysis of terrestrial C cycling – including
retrievals of C fluxes, C pools, autotrophic respiration, allocation fractions and residence times – based on multiple global-scale earth observations and datasets. Current global-scale terrestrial biosphere models, due to their complexity and structures, are ill-equipped to ingest an ever-increasing volume of earth observations to estimate (instead of prescribing) model parameters, based on the currently available observations. To overcome this challenge, we use a model-data fusion (MDF) approach to retrieve terrestrial C state and process variables during the period 2001–10, without invoking plant functional type or steady-state assumptions. We bring together global MODIS LAI, a tropical biomass map (24), a soil C dataset (23), MODIS burned area (34), and a diagnostic ecosystem C balance model (DALEC2, 19, 35) to retrieve C state and process variables by producing a novel data-consistent and spatially explicit analysis of terrestrial C cycling on a global 1°×1° grid (Fig. 1; we henceforth refer to this model-data fusion setup as the CARbon DAta MOdel framework, or CARDAMOM). Specifically, we address the following questions: how is C uptake partitioned between the live biomass pools and respiration? What is the residence time of C within the major ecosystem C pools? How do estimates of C cycle states and processes vary spatially and to what degree do emergent variable patterns match land-cover maps? We use a Markov Chain Monte Carlo MDF algorithm to retrieve C state and process variables – and their associated uncertainty – within each 1°×1° degree grid cell (see Materials and Methods). The MDF approach retrieves the state and process variables that minimize the model mismatch against any available C cycle observations. Therefore, in the absence of extra-
tropical biomass data or winter-time MODIS LAI observations, estimates of 2001–10 C cycle state and process variables are achievable, albeit more uncertain.

Results

Distinct C allocation patterns emerge from our terrestrial C analysis (Fig. 2). Net primary production (NPP) allocation to structural biomass (wood and fine roots) is largely ≥80% (area-weighted 25th – 75th %ile range = 85–88%) in the wet tropics (<23°N/S; annual precipitation > 1500mm), in contrast to the dry tropics (77–87%), and extra-tropical regions (73–82%). The highest NPP allocations to foliage (≥30%) spatially coincide with major grassland areas, including the North America prairies, Central Asia steppes and the Sahel region in Africa. The dry tropics exhibit relatively high NPP allocation to labile C (7–14%; Fig. S1); this reflects the increasing impact of seasonality on production as precipitation declines, requiring labile C stores for leaf flush. Carbon use efficiency (CUE = 1 - autotrophic respiration fraction) is overall lowest in within the wet tropics (0.42–0.44) in contrast to the dry tropics (0.45–0.50), temperate (23–55°N/S; 0.47–0.50) and high latitudes (>55°N/S; 0.49–0.50).

Live biomass and dead organic C residence times exhibit contrasting spatial features (r = 0.3; Fig. 3). Within the majority of wet tropical land area (56%) – especially across most of the Amazon (76%) and Congo (69%) river basins – the longest C residence time occurs within the woody pool (Fig. S1). In the dry tropics and extra-tropical latitudes, soil C
residence times exceed wood C residence time by a median factor of 2.6 (1.6–4.3).

Woody residence time is typically shorter in the dry tropics (8–19 yrs) compared to other biomes (wet tropics: 12–21 yrs; temperate: 21–29 yrs; high latitudes: 25–28 yrs).

Litter C residence time is typically longer in extra-tropical ecosystems (0.8–1.6 yrs) in comparison to tropical ecosystems (0.4–0.5 yrs). The longest foliar residence time (or leaf lifespan) occurs in the wet tropics and semi-arid regions (Fig. S1).

Overall, the wet tropics are characterized by relatively high structural C (>100 tC ha⁻¹) and photosynthetic C (>2.5 tC ha⁻¹)(Fig. 4): in contrast, the dry tropics and extra-tropical regions exhibit less structural and/or photosynthetic C. Foliar C stocks are typically larger in the wet tropics (2.8–4.7 tC ha⁻¹) relative to other biomes (0.2–0.6 tC ha⁻¹); similarly, fine root stocks are also greater in the wet tropics (4.0–5.3 tC ha⁻¹), compared to other biomes (0.8–2.7 tC ha⁻¹). Root:shoot (fine root C:leaf C) is lowest in the wet tropics (1.1–1.5), followed by the dry tropics (1.6–1.9) and extra-tropics (1.8–2.1). We find larger woody C uncertainties (1°×1° 90% confidence range / median) in the extra-tropics (1.8–4.6) in contrast to tropical woody C (1.4–1.6) due to the latitudinal limits of the total above- and below-ground biomass map (24). Litter C is greater in high latitudes (2.4–3.4 tC ha⁻¹) relative to temperate (0.6–2.4 tC ha⁻¹) and tropical (0.2–2.6 tC ha⁻¹) regions. High-latitude ecosystems have higher labile C stocks linked to seasonal leaf expansion (0.2–0.5 tC ha⁻¹) relative to temperate (0.1–0.3 tC ha⁻¹) and tropical (0.1–0.3 tC ha⁻¹) ecosystems.
We find high leaf C mass per leaf area (LCMA) values in the wet tropics (85-97 gC m$^{-2}$), and in semi-arid regions, such as the Sahel, South-western United States and the Australian continent (typically >100 gC m$^{-2}$; Fig. 5); LCMA estimates are lower (typically <80 gC m$^{-2}$) in high latitudes and the dry tropics. We find a positive correlation between leaf lifespan and LCMA in high-latitude ($r = 0.79$), temperate ($r = 0.80$), dry tropical ($r = 0.78$) and wet tropical ($r = 0.64$) areas.

Global GPP (global 25$^{th}$ – 75$^{th}$ percentile = 91–134 Pg C yr$^{-1}$), ecosystem respiration (91–137 Pg C yr$^{-1}$) and fires (1.3–2.0 Pg C yr$^{-1}$) are broadly consistent with the MsTMIP terrestrial carbon model ensemble (22), data-driven estimates (36) and bottom-up inventories (37) (Fig. S5). The Net Carbon Exchange uncertainty (-8 to +13 Pg C yr$^{-1}$) is an order of magnitude greater than mode NCE (-2 Pg C yr$^{-1}$); NCE latitudinal uncertainty is larger but comparable to the MsTMIP model range. Global atmospheric model CO$_2$ concentrations based on CARDAMOM mode NCE fluxes are seasonally consistent ($r^2 = 0.93$, RMSE = 0.53 ppm CO$_2$) with mean total column CO$_2$ measurements (38, Fig. S6). The mean integrated C residence time by (31) is within the range of individual pool residence times at locations B, T, D and W (Fig. 3). The 2001–10 CARDAMOM analysis spatial and temporal LAI variability is consistent with the MODIS LAI constraints ($r^2 = 0.8$; RMSE = 0.6 m$^2$/m$^2$). When alternative GPP (36), alternative model structure or biased data constraints ($\pm 20\%$) are imposed at locations B, T, D and W, 88% of median sensitivity analysis estimates are within $\pm 50\%$ of median C state and process variable retrievals (Fig. S2).
Retrieved C cycle variables are broadly consistent with a range of in-situ measurements (Table S3). Estimates of CUE within the Amazon river basin are comparable to the upper bound of recent measurements (0.32–0.47)(39). Recent estimates of extra-tropical forest C density (40) are on average 38% lower than CARDAMOM total biomass estimates within forested areas (although these are typically within the CARDAMOM 1°×1° uncertainty). Estimates of mean Amazon woody C residence times (15–21 yrs) are lower but comparable to above-ground woody C residence times derived from site-level measurements (~20–70 yrs; 20).

We find that 88-99% of C state and process variability is accounted for by 8 empirical orthogonal basis functions (EOFs, Fig. 6); in other words, retrieved C state and process variables are largely explained by eight modes of spatial variability (Fig. S4). On average, GLOBCOVER land-cover type classifications (41, e.g. deciduous forests, evergreen forests and grasslands) account for <50% of C state and process variability (median multiple correlation coefficient R = 0.41); GLOBCOVER land-cover types best describe spatial variations in C stocks (0.5≤R≤0.8), followed by LCMA (R = 0.4), residence times (0.3≤R≤0.5) and allocation fractions (0.1≤R≤0.4).

Discussion
Typically C allocation and residence time parameters are based on land-cover types in global-scale terrestrial C cycle studies (9, 22, amongst others); here, spatially broad allocation and residence patterns emerge instead, as a result of the model-data fusion approach. For example, high biomass ecosystems throughout the wet tropics display similar C allocation, residence time and LCMA configurations (Fig. 2–5). Similarly, we find that dead organic matter (DOM) C residence is generally longer in high latitudes (Fig. 3). In comparison to conventional land-cover types, EOFs 1-4 account for a larger degree of the spatial structures in retrieved C variables (Fig. 6); for most variables, the two dominant EOF modes – which together reflect first-order variations in latitude and global precipitation patterns (Fig. S4) – explain more spatial variability than GLOBCOVER land-cover types. The mismatch between land-cover types and retrieved variables has major implications for the estimation and prediction of terrestrial C cycling, which is currently based on small sets of globally applied parameters linked to land-cover types. The importance of climate, biodiversity, fire and anthropogenic disturbance in generating these mismatches needs to be explored in further research (42).

It also is clear that plant traits vary across biomes (Fig. 2-4, S1), not just at biome boundaries (43), and that there are continental-scale trade-offs and correlations among traits (44). Our analysis is consistent with these viewpoints: for example, the emergent relationship between LCMA (proportional to leaf mass per area) and leaf lifespan (Fig. 5) matches the positive correlation found in global plant trait datasets (45). Evaluating global plant-trait patterns emerging from CARDAMOM provides a novel opportunity for
connections to theoretical and functional biodiversity research, and a route to integrating this knowledge into predictive terrestrial C cycle modeling.

The residence times of major C stocks provide substantial insights into the sensitivity and potential future trajectories of the terrestrial C cycle. For example, land-cover changes in the wet tropics may result in rapid DOM C losses, given the relatively short DOM residence times (<30 yrs) (Fig. 3). In contrast, high-latitude C residence times are an order of magnitude higher (30–300 yrs), and therefore shifts in C allocation or turnover rates are likely to result in long-lived C flux responses. Overall, given the predominant role of C residence times in future terrestrial uptake responses (9), the derived residence times provide a first-order estimate of ecosystem response times as a result of changes in C cycling regimes. However, we note that model structure is likely to be a major source of uncertainty in long-lived (>10yr) C flux predictions. For example, while reduced complexity models can capture some of the principal long-term (>10yr) DOM dynamics represented in earth system models (46) systematic errors in DOM dynamics can arise due to the under-representation of processes controlling DOM residence times (47, 48). We also note that our decadal analysis is unlikely to be able to capture slow feedback processes acting on longer time-scales such as permafrost re-mobilization and priming (49). The large allocation and stocks and short residence time of wood in the wet tropics indicates the potentially rapid potential for regrowth and C accumulation post-disturbance (50). We note that fires are less frequent but major
events within boreal ecosystems (51), and therefore longer time-periods are required for retrievals to fully account for the effect of fires on high-latitude C residence times.

C state and process variable retrievals are sensitive to the uncertainty characteristics of C cycle observations (35) and the prior parameter ranges (Table S1). We highlight that the current coverage and accuracy of C cycle observations (24, 52) remains a major limiting factor in our approach. For example, temperate and high-latitude C stock and residence time uncertainties are higher due to the absence of biomass observations. Undoubtedly, future estimates of globally-spanning biomass density will provide a major constraint on CARDAMOM estimates of extra-tropical C state and process variables (53).

Land-to-atmosphere C flux estimates could be used to further constrain CARDAMOM C fluxes (Fig. S5) and C cycle variables associated to non-steady C states. For example, soil C residence time samples are negatively correlated with corresponding mean 2001–10 NCE samples at locations B (r = -0.3), T (r = -0.4), D (r = -0.5) and W (r = -0.3); therefore, regional- or grid-scale estimates of NCE could provide a much-needed additional constraint on soil C residence time. CARDAMOM flux magnitude and uncertainty can be used as prior information in global atmospheric CO₂ inversions; in turn, the assimilation of GOSAT and OCO-2 atmospheric CO₂ observations (54) should further constrain CARDAMOM NCE estimates and their associated uncertainties. In this manner, non-steady state C fluxes can ultimately be reconciled with ecosystem state and process variables, such as C stocks and residence times.
The CARDAMOM approach provides a framework to test alternative model structures (55): in this manner, combined C cycle model parametric and structural uncertainties can be characterized, while ensuring consistency between models and global-scale datasets. This assessment would amount to a major step forward from conventional C cycle model inter-comparison studies. Ultimately an ensemble of models can be used to determine the degree to which retrievals of key C state and process variables are model-dependent. Moreover, alternative model structures could be used in CARDAMOM to assimilate globally spanning plant traits related to C cycling (56) and satellite observations such as solar-induced fluorescence (27), vegetation optical depth (57), soil moisture (58, 59) and changes in above-ground biomass (25, 60, 61). We anticipate that the incorporation of additional datasets and alternative model structures into CARDAMOM will generate quantifiable reductions in retrieved C variable uncertainties and new ecological insights on the state of terrestrial C cycle.

Materials and Methods

We grid MODIS LAI, ABGB (24), and HWSD topsoil and subsoil (0-100cm) C density (23) at a 1°×1° resolution (section S1 of the SI). The Data Assimilation Linked Ecosystem Carbon model version two (DALEC2) is analytically described by (35); an overview of DALEC2 C fluxes and pools is shown in Fig. 1. The 17 DALEC2 parameters (controlling the processes of photosynthesis and phenology, allocation, and turnover rates) and six
initial C pools robustly characterize terrestrial ecosystem C balance (19). DALEC2 is a
generic representation of C-cycling, where plant-functional types (PFTs) are not explicit:
instead, model parameters are treated as unknown and independent quantities for each
1°×1° grid cell (Table S1). We incorporate a fire C loss parameterization to account for
seasonal and inter-annual variations in fire C fluxes from DALEC2 (section S2 of the SI).
The model drivers consist of monthly time-step ERA interim meteorology and MODIS
burned area (34) at a 1°×1° resolution.

For each 1° x 1° grid cell, we use Bayesian inference to retrieve the probability of
DALEC2 model parameters $x_i$ (Table S1) given observational constraints $O_i$, henceforth
$p(x_i | O_i)$, where

$$p(x_i | O_i) \propto p(x_i) p(O_i | x_i) \quad (1)$$

$p(x_i)$ is the prior parameter information and $p(O_i | x_i)$ is the likelihood of $x_i$ with respect to
$O_i$. We use a Markov Chain Monte Carlo algorithm to sample $x_i$ from $p(x_i | O_i)$: we
henceforth refer to the retrieved DALEC2 parameter values at pixel $i$ as $y_i$. Within each
grid cell, C allocation fractions, residence times within each C pool, stocks, LCMA, and
associated C fluxes are derived from 4000 samples of $y_i$ (section S3 of the SI). We hence
obtain a probability density function (PDF) for all C cycle variables within each 1°×1° grid
cell.
We do not impose PFT specific prior parameter distributions, or steady state assumptions: \( p(x_i) \) consists of ecologically viable parameter ranges (Table S1) and ecological and dynamical constraints (35). Together these guarantee ecologically consistent parameter retrievals within a globally prescribed parameter space, without imposing spatially explicit prior parameter information.

From the C state and process variable estimates within each 1°×1° grid cell we use 4000 samples of \( y \), to determine the mean, median, mode, and \( \% \)ile ranges for each C state and process variable ranges. In Fig. 2-4, we present C allocation, residence time and C stock \( 5^{th}, 25^{th}, 75^{th} \) and \( 95^{th} \) \%iles at four selected locations: \( B \): boreal [62.5°N, 81.5°E]; \( T \): temperate [40.5°N, 120.5°W]; \( D \): dry tropics [12.5°N, 20.5°E] and \( W \): wet tropics [7.5°S, 60.5°W]. We chose \( B, T, D, W \) as representative examples for C state and process variable values within each area (the full 1°×1° C state and process variable maps are shown in Fig. S1). To determine the robustness of our C state and process variable estimates, we perform dedicated sensitivity tests to characterize the role of systematic errors in data constraints and model structure: we repeat our C variable retrievals using ±20% LAI, ±20% ABGB, ±20% HWSD, ±20% combustion coefficients, alternative GPP (36) and limited heterotrophic respiration at <0°C (section S4 of the SI).

We compare our results against in-situ and regional observations of C allocation, pools and residence times (section S5 of the SI), and we evaluate the resulting fluxes against atmospheric CO₂ observations across 12 Total Carbon Column Observing Network sites.
(38) by incorporating NCE results in a 4D atmospheric transport model (29). To determine whether global land-cover types can predict the spatial variability of our results, we conduct a multiple correlation coefficient analysis between C state and process variables and 18 GLOBCOVER land-cover fractions at 1°×1°. We also employed a principal component analysis on C state and process variables to retrieve the primary 1°×1° empirical orthogonal functions (EOFs). The details of the CARDAMOM results evaluation and analyses are fully described in sections S5-8 of the SI. Statistical abbreviations throughout the text include $r$ (Pearson correlation coefficient), $RMSE$ (root-mean-square error). All spatially derived $r$ and $RMSE$ values reported in the text are area-weighted. Retrieved C variable ranges – reported as area-weighted 25th – 75th %iles – are derived from 1°×1° mean allocation and C stocks, log-based mean C residence times (Fig. S1) and median LCMA values (Fig. 6). All CARDAMOM datasets presented in this study can be downloaded from datashare.is.ed.ac.uk/handle/10283/875.

Acknowledgements

AAB, JE, LF and MW were funded by the NERC National Centre for Earth Observation (NCEO), UK. This work made use of the Edinburgh Compute and Data Facility resources (ECDF, ecdf.ed.ac.uk). The research leading to these results has received funding from the European Union’s FP7 (2007-2013) under grant agreement no. 283080, project GEOCARBON. IV was financially supported under a Netherlands Organization for
Scientific Research project (VIDI: 864.08.012). The TCCON Network is supported by NASA’s Carbon Cycle Science Program through a grant to the California Institute of Technology. Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

Contributions

AAB and MW conceived the research. AAB undertook the research with assistance from JE, IV, MW and LF. AAB and MW wrote the paper with assistance from the other authors.

References


55. Keenan TF, Carbone MS, Reichstein M, Richardson A (2011) The model–data fusion pitfall: assuming certainty in


57. Lawrence H, et al. (2014) Comparison between SMOS Vegetation Optical Depth products and MODIS vegetation

the IEEE* 98,(5): 666-687.


60. Le Toan T, et al. (2011) The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial


64. Huntzinger DN, et al. (2015, in press). NACP MsTMIP: Global 0.5-deg Terrestrial Biosphere Model Outputs (version


**Figure captions**

**Fig. 1:** Diagnostic ecosystem carbon (C) balance model DALEC2 (19,35) and datasets used to retrieve 1°×1° C state and process variables. Gross primary production (GPP), a function of climate and foliar C, is partitioned into autotrophic respiration (Ra) and net primary production (NPP). NPP is partitioned into the live biomass pools. Plant mortality provides input to the dead organic matter (DOM) pools. Heterotrophic respiration (Rh) is derived from decomposing DOM pools. Fire fluxes are derived from burned area data (35) and all C pools (see section S3). Within each 1° x 1° grid cell, we use a Bayesian model-data fusion algorithm to retrieve C state/process variables and uncertainties; variables are retrieved without prior land-cover type or steady state assumptions. Data constraints consist of MODIS leaf area, total biomass (24, tropics only) and soil C (23).
Details on the Bayesian fusion approach are provided in the Materials and Methods section.

**Fig. 2:** Retrievals of net primary production (NPP) allocation to structural (wood and fine roots) and photosynthetic (labile and foliage) C pools. Allocation fractions were retrieved at 1° × 1° using a Bayesian model-data fusion approach (Fig. 1). The gross primary production allocation fraction retrievals at locations B, T, D and W are shown in the three right-hand panels (median = black dot, 50% confidence range = box and 90% confidence range = line).

**Fig. 3:** Retrievals of C residence time (RT) in live biomass and dead organic matter C pools; residence times are retrieved at 1° × 1° using a Bayesian model-data fusion approach (Fig. 1). Brown colors denote ecosystems with high residence times for all C pools, green areas denote ecosystems with long live biomass C residence times and orange areas denote ecosystems with low live biomass residence time. The residence times for individual C pools at locations B, T, D and W are shown in the three right-hand panels (median = black dot, 50% confidence range = box and 90% confidence range = line). Mean C residence times by (31) are shown as grey boxes (50% confidence interval) and black dots (median).

**Fig. 4:** Retrieved mean photosynthetic (foliar and labile) and structural (wood and fine roots) C pool stocks; C stocks are retrieved at 1° × 1° using a Bayesian model-data fusion approach.
approach (Fig. 1). Retrieved mean C stocks for each pool at locations B, T, D and W are shown in the four right-hand panels (median = black dot, 50% confidence range = box and 90% confidence range = line). Dark colors denote high structural C/high photosynthetic C ecosystems, green colors denote low structural C/high photosynthetic C ecosystems, red colors denote low photosynthetic C/high structural C ecosystems, and yellow colors denote low photosynthetic C/low structural C ecosystems.

Fig. 5. Top: Retrieved median 1° x 1° leaf C mass per leaf area (LCMA, gC m⁻²). Top right: zonal mean of median LCMA and 50% confidence range. Bottom: LCMA against leaf lifespan for high latitudes (>55°N/S), temperate regions (23°–55°N/S), dry tropics (precip.<1500mm, <23°N/S) and wet tropics (precip.>1500mm, <23°N/S).

Fig. 6. Multiple correlation coefficients (R, x-axis) of retrieved C state and process variables (allocation fractions, residence times, mean C pools and LCMA; y-axis) against 18 GLOBCOVER land-cover fractions and C variable primary empirical orthogonal functions (EOFs). R denotes the ability of GLOBCOVER land-cover types and primary EOFs to predict 1° x 1° state and process variables (R would equal 1 if all C state and process variables could be expressed as a linear sum of land-cover fractions or EOFs).
**Data constraints**

- **C1** MODIS leaf area
- **C2** Biomass (24)*
- **C3** Soil C (23)
  *Tropics only*

**Model drivers**

- **D1** Meteorology
- **D2** MODIS burned area

**Carbon (C) fluxes**

- **GPP** = Gross primary production
- **Ra** = Autotrophic respiration
- **Rh** = Heterotrophic respiration
- **NPP** = Net primary production \((GPP - Ra)\)
- **Re** = Ecosystem respiration \((Ra + Rh)\)
- **NCE** = Net C exchange \((Re + Fires - GPP)\)
Supporting information

“The decadal state of the terrestrial carbon cycle: global retrievals of terrestrial carbon allocation, pools and residence times”.

A. Anthony Bloom$^{1,2,3}$, Jean-François Exbrayat$^{2,3}$, Ivar van der Velde$^4$, Liang Feng$^{2,3}$, Mathew Williams$^{2,3}$

1. Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
2. School of GeoSciences, University of Edinburgh, Edinburgh, UK
3. National Centre for Earth Observation, Edinburgh, UK
4. Wageningen University, Wageningen, The Netherlands

S1. Global 1° × 1° grid model-data fusion

**Global datasets.** We grid the 30-second Harmonized World Soil Database soil carbon (C) density (HWSD, based on national inventories of top 1 m soil bulk density and organic C content) (23) and a ~1km × 1km above- and below-ground pan-tropical biomass map (24) at 1° × 1°. We grid the MOD15A2 MODIS LAI product (1km × 1km) and a MODIS Burned Area product (0.25° × 0.25°) (34) at a 1° × 1° monthly resolution for each month within the period 2001-2010. While finer spatial/temporal resolutions can potentially be implemented, we chose a 1° × 1° monthly resolution as a consequence of the computational cost of our approach. We use ERA-interim 1° × 1° monthly re-analysis products as DALEC2 drivers; see (35) for MODIS quality flag and ERA-interim driver details. We exclude 1° × 1° grid-cells where desert and ice-covered areas account for more than 90% of the grid-cell land-cover (based on GLOBCOVER 2009 Global land-cover maps, 41), as their role in the terrestrial C cycle is negligible.

**Model-data fusion.** Within each 1° × 1° degree grid cell $i$, we use the 1° × 1° aggregated biomass (tropics only), soil C and MODIS LAI datasets (observations $O_i$) to constrain DALEC2 parameters $\mathbf{x}_i$ (for a complete description of the DALEC2 model and C pools, we refer the reader to (35) and references therein). We implement a Metropolis-Hastings Markov Chain Monte Carlo (MHMCMC, 33, 35) to determine the probability of $\mathbf{x}_i$ given observational constraints $O_i$ (see equation 1 in main text).

The prior ranges of DALEC2 parameters $\mathbf{x}_i$ are shown in Table S1. We also imposed a prior log-normal distribution on autotrophic respiration fraction $x_{i,o}$ (autotrophic respiration $= 0.5 \times 1.2^{\pm 1}$) and a prior log-normal distribution on canopy efficiency $x_{i,c}$ (canopy efficiency parameter $= 17.5 \times 1.2^{\pm 1}$), where $\pm 1$ represents a normal distribution with mean 0 and variance 1. These constraints yield a range of results that are broadly consistent with the global GPP range reported by (30) and represent the range of autotrophic respiration estimates reported by (62). The prior parameter probability, $p(\mathbf{x}_i)$ is therefore expressed as:

$$p(\mathbf{x}_i) = p_{BW}(\mathbf{x}_i) e^{-0.5 \left( \frac{\log(x_{i,o}) - \log(0.5)}{\log(1.2)} \right)^2} e^{-0.5 \left( \frac{\log(x_{i,c}) - \log(17.5)}{\log(1.2)} \right)^2}, \quad (1)$$
where $p_{bw}(x_i)$ is the prior parameter probability described by (35). Within each $1^\circ \times 1^\circ$ grid cell, we prescribe an uncertainty factor of 1.5 to mean 2001-10 HWSD soil $C$ and total above and below-ground biomass density (i.e. mean labile + foliar + fine roots + wood), and an uncertainty factor of 2 to mean monthly MODIS LAI observations. For total biomass, given that the maximum entropy algorithm employed by (24) was based on bins of 12.5 tC ha$^{-1}$, we anticipate that low biomass density values (such as the edges of the Sahel and Kalahari deserts) exhibit comparable uncertainty. We therefore prescribe an uncertainty factor of $max(1.5, 12.5/B_i)$, where $B_i$ is the total biomass density and the $max(\cdot)$ function denotes the maximum of the two values. The likelihood function $p(O_i|x_i)$ is therefore expressed as:

$$p(O_i|x_i) = \exp\left(-0.5 \sum_{j=1}^{N} \left(\frac{M_{ij} - O_{ij}}{U_{ij}}\right)^2\right), \quad (2)$$

where $O_{ij}$ and $U_{ij}$ are the $j^{th}$ observations and uncertainty factors at location $i$, and $M_{ij}$ is the equivalent DALEC2 model output based on parameter vector $x_i$, (we note that $O_{ij}, U_{ij}$ and $M_{ij}$ are log-transformed, e.g. for a soil $C$ value of 100 tC ha$^{-1}$, $O_{ij} = \log(100)$ and $U_{ij} = \log(1.5)$). For each LAI observation, $M_{ij}$ is the DALEC2 foliar $C$ (on the corresponding month) divided by leaf $C$ mass per leaf area. For biomass and soil $C$, $M_{ij}$ is the DALEC2 soil $C$ stock and mean live biomass (labile + foliar + fine roots + wood) on Jan 1$^{st}$ 2001. For the analytical description of DALEC2, the MHMCMC algorithm and $p_{bw}(x_i)$, we refer the reader to (35) and references therein; the DALEC2 fire module is described in section S3.

**Ecological and Dynamical Constraints.** The twelve Ecological and Dynamical Constraints (EDCs 1-12, 35) are a component of the prior parameter probability – $p_{bw}(x_i)$ – and consist of relative constraints on allocation parameters, turnover rates, growth rates, exponential decays and steady state proximity. When steady state is not assumed, steady state proximity conditions are necessary to distinguish between real and nonsensical $C$ pool trajectories (35). We developed a simpler numeric equivalent of the steady state proximity EDCs (EDCs 9-12), to account for the stochastic $C$ losses from fires. For each pool, we derive the steady state proximity factor ($S_{prox}$) as follows:

$$S_{prox} = \frac{\overline{C_{input}}}{\overline{C_{output}}} \quad (3)$$

where $\overline{C_{input}}$ and $\overline{C_{output}}$ are the mean inputs and outputs from each pool. We impose a steady state proximity condition of $0.5 > S_{prox} > 2$ for each pool.

We found that EDC 8 – the ecological and dynamic constraint limiting rapid exponential pool trajectories – was excessively rigid for relatively small amounts of exponential pool trajectories (which can occur naturally and/or as a model artifact). Here we use a simpler approach to minimize the rapid exponential decay of $C$ pools: we ensure that the steady state proximity of each $C$ pool at time zero – $S_{prox(jan2001)}$ – is within 0.05 of $S_{prox}$, i.e.

$$|S_{prox} - S_{prox(jan2001)}| < 0.05. \quad (4)$$

$S_{prox(jan2001)}$ can be derived as:
\[ S_{\text{prox}(\text{Jan}2001)} = S_{\text{prox}} \times \frac{C_{\text{Jan}2001-10}}{C_{\text{Jan}2001}}, \quad (5) \]

where \(C_{\text{Jan}2001-10}\) is the mean January C pool stock and \(C_{\text{Jan}2001}\) is the C pool stock in January 2001. The CARDAMOM code used in this manuscript (DALEC2 model, EDCs and adaptive Metropolis-Hastings Markov Chain Monte Carlo) is available upon request.

**S2. DALEC2 fire module**

To determine the monthly C losses from fires at time \(t\), we determine the monthly fraction of each grid cell burned, \(B_{\text{area}(t)}\), based on the MODIS-derived burned area product (34). At each monthly time step, the fire losses within each \(1^\circ \times 1^\circ\) grid cell are derived as follows:

\[ F_e(t) = B_{\text{area}(t)} \times \sum_{p=1}^{6} k_{\text{factor}(p)} C_{(p,t)}, \quad (6) \]

where \(F_e(t)\) are the total fire C emissions at time \(t\), \(k_{\text{factor}(p)}\) is the combustion factor for pool \(p\), and \(C_{(p,t)}\) is the C in pool \(p\) at time \(t\). We also impose a resilience factor \(r\) to the remaining pools within the burned area: from live biomass pool, a fire-mortality flux is derived from the un-combusted C pools as follows:

\[ F_m(t,p) = B_{\text{area}(t)} \times (1 - k_{\text{factor}(p)}) (1 - r) C_{(p,t)}, \quad (7) \]

The fire-mortality C flux from foliage, roots and labile is deposited into the litter pool, and fire mortality C flux from wood is transferred to the soil C pool. Equally, \((1 - r)\) of un-combusted litter C is transferred to the soil C pool. The \(k_{\text{factor}}\) values for labile (0.1), foliar (0.9), root (0.1), wood (0.1), litter (0.5) and soil C (0.01) are broadly equivalent to the \(k_{\text{factor}}\) values used by the Global Fire Emission Database (37). We apply a resilience factor of \(r = 0.5\). The sensitivity calculations associated with \(k_{\text{factor}(p)}\) and \(r\) are described in section S4.

**S3. Global state and process variables**

The spatial distributions of individual C pool allocation fractions, residence times and stocks are shown in Fig. S1. The residence time for each C pool at grid cell \(i\) is derived as follows:

\[ RT_{\text{pool}(j)} = \frac{C_{\text{pool}(j)}}{F_{\text{in}(j)} - \Delta C_{\text{pool}(j)}} \times 365.25, \quad (8) \]

where \(C_{\text{pool}(j)}\) is the mean pool size, \(F_{\text{in}(j)}\) is the mean daily C pool input and \(\Delta C_{\text{pool}(j)}\) is the mean daily change in pool size throughout 2001-10 for the \(j^{\text{th}}\) parameter vector sample of \(\mathbf{y}\) (i.e. \(C_{\text{pool}(j)}, F_{\text{in}(j)}\) and \(\Delta C_{\text{pool}(j)}\) are calculated from DALEC2 output driven with \(j^{\text{th}}\) parameter vector sample of \(\mathbf{y}\)). Mean live biomass (dead organic matter) pool residence times are derived based on equation 8, where \(C_{\text{pool}(j)}, F_{\text{in}(j)}\) and \(\Delta C_{\text{pool}(j)}\) are the total live biomass (dead organic matter) C corresponding to \(j^{\text{th}}\) parameter vector sample of \(\mathbf{y}\). Leaf lifespan is equivalent to \(RT_{\text{foliar}}\). Reported global and zonal 25\textsuperscript{th} – 75\textsuperscript{th} %ile ranges of total annual fluxes were derived from the sum of monthly 1\textsuperscript{st} × 1\textsuperscript{st} 25\textsuperscript{th} and 75\textsuperscript{th} %iles for each flux multiplied by the 1\textsuperscript{st} × 1\textsuperscript{st} grid cell area; the same approach was used to derive median fluxes and mode net C exchange (NCE).
Monthly mode NCE within each 1° × 1° grid-cell was derived by binning NCE samples into 0.01 gC m² day⁻¹ intervals.

S4. Sensitivity Tests

We determine the sensitivity of C allocation, residence times and C pool size estimates at locations B, T, D and W (see Materials and Methods in main text for B, T, D and W coordinates) to LAI, biomass and soil C data constraints (sensitivity tests S1-S6), fire combustion and resilience factor coefficients (sensitivity test S7-S10), the use of MPI GPP (36) instead of the default DALEC (19) GPP (sensitivity test S11), and the suppression of heterotrophic respiration under -10°C (sensitivity test S12). The sensitivity experiments are summarized in Table S2. The results of the sensitivity tests are shown in Fig. S2.

S5 Comparison against in-situ and regional observations

CARDAMOM results are compared against a range of in-situ measurements in Table S3. We compare each in-situ measurement against the 50% and 90% confidence range of the mean 1° × 1° values within the stated region. Comparison details and footnotes are included in Table S3. We also compare CARDAMOM total biomass against a boreal forest biomass dataset derived from synthetic aperture radar data (BIOMASAR map, 40) aggregated to 1° × 1°. The CARDAMOM-to-BIOMASAR comparison is conducted for the total biomass across all 1° × 1° areas with at least 95% BIOMASAR map coverage; total BIOMASAR biomass within those areas is 38% lower than CARDAMOM biomass. We note that lower-than-expected LCMA estimates in boreal ecosystems (Fig. 6) could be explained by (i) understory plant traits (linked to deciduous shrubs); or (ii) seasonal MODIS LAI biases (53). In particular, the significant correlation between LCMA and leaf lifespan suggests that retrieved LCMA accuracy could be strongly linked to seasonal biases in MODIS LAI.

S6 Comparison to GLOBCOVER land-cover types and EOFs

For each 1° × 1° grid cell i we determine the fraction of each GLOBCOVER (41) land-cover type L, \( F_{L(i)} \). We then determine the Pearson’s correlation coefficients \( r_{Ls} \) between \( f_i \) (the vector of all 1° × 1° land-cover type \( L \) fractions) and each C state and process variable vector \( c_s \) (the vector of each 1° × 1° state and process variables (denoted by subscript \( S \)) consist of allocation fractions, C residence times, C pool sizes and leaf C mass per leaf area. The \( r_{Ls}^2 \) values between each GLOBCOVER land-cover type fraction \( L \) and each C state/process variable \( S \) are shown in Fig. S3. The land-cover categories are: irrigated croplands (CRI); Rainfed croplands (CRR); Mosaic cropland > vegetation (MCV); Mosaic vegetation > cropland (MVC); Closed to open broadleaved evergreen or semi-deciduous forest (BESDF); Closed broadleaved deciduous forest (BDF); Open broadleaved deciduous forest/woodland (BDFW); Closed (>40%) needleleafed evergreen forest (NEF); Open needleleafed deciduous or evergreen forest (NEDEF); Closed to open mixed forest (MF); Mosaic forest or shrubland > grassland (MFSG); Mosaic grassland > forest or shrubland (MGFS); Closed to open shrubland (SRB); Closed to open herbaceous vegetation (GRA); Sparse vegetation (SPA); Closed to open broadleaved forest regularly flooded (FWE); forest or shrubland, permanently flooded (SWE); Closed to open vegetation on flooded or waterlogged soil (WET) (54).
The multiple correlation coefficient $R_S$ between C state/process variable $S$ and 18 GLOBCOVER land-cover type fractions is derived as follows:

$$R_S = r_S^T R_{LL}^{-1} r_S,$$

(9)

where $r_S$ is the $1 \times 18$ vector of correlations coefficients between state/process variable vector $c_S$ and 18 $1^\circ \times 1^\circ$ land-cover type fraction vectors $f_i$, $r_S^T$ is the transpose of $r_S$, and $R_{LL}^{-1}$ is the inverse of the correlation matrix $R_{LL}$, which contains the inter-correlations between 18 land-cover type fraction vectors $f_i$. $R_S$ is equivalent to the maximum correlation (Pearson’s $r^2$) between the spatial variability of C state/process variable $c_S$ and the best-fitting linear combination of land-cover type fractions $f_i$. The resulting $R_S$ values are shown in Fig. 6 (main text).

We also employ a multiple correlation coefficient analysis on the empirical orthogonal functions (EOFs, or the “primary modes” of variability) of all $c_i$. We conducted a principal component analysis to derive the eight primary EOFs (EOFs were derived using “pca.m” function in Matlab; each $c_i$ vector is centered at zero and scaled to the standard deviation of $c_i$). Standardized EOFs (normalized by EOF standard deviation) and EOF coefficients are shown in Fig. S4. The EOF maps exhibit the primary modes of $c_i$ variability in space; for each $c_i$, the maximum spatial variability explained by EOFs 1 – N is the sum of standardized EOFs 1 – N multiplied by their associated coefficients. EOF multiple correlation coefficients $R_{S(EOF)}$ were derived for the primary two, four and eight EOFs based on equation 9, where $R_{LL}$ is the identity matrix (as EOFs are orthogonal). $R_{S(EOF)}$ results are shown in Fig. 6 in the main text.

**S7 Comparison against the MsTMIP terrestrial biosphere model ensemble**

We compare GPP, ecosystem respiration and NCE against the MsTMIP terrestrial biosphere model ensemble Version 1.0 (64) net C exchange (note: total C exchange reported as net ecosystem exchange, or ‘NEE’, by MsTMIP). The $0.5^\circ \times 0.5^\circ$ monthly GPP, total (heterotrophic and autotrophic) respiration and NCE values for 2001-10 – based on the BG1 simulation – were downloaded from [http://nacp.ornl.gov/mstimipdata/](http://nacp.ornl.gov/mstimipdata/), and were aggregated to a $1^\circ \times 1^\circ$ grid (the BG1 simulation includes time-varying nitrogen deposition, atmospheric CO$_2$ and land-use history (22)). The eight MsTMIP models shown in Fig. S5 are BIOME-BGC, CLASS-CTEM-N, CLM4VIC, CLM4, DLEM, ISAM, TEM6, TRIPLEX-GHG (for the sake of brevity, we did not label each individual MsTMIP model in Fig. S5);

**S8 Atmospheric CO$_2$ comparison**

We incorporated the 2009-2010 CARDAMOM monthly mode net C exchange (NCE) values into the GEOS-Chem atmospheric chemistry and transport model (29). The GEOS-Chem model simulations are based on GEOS-Chem version 8.2, driven by NASA GEOS-5 meteorological fields. In addition to NCE, fossil fuel emissions and oceanic surface CO$_2$ fluxes are prescribed (56). We compared the 2009-2010 GEOS-Chem model CO$_2$ concentrations against the monthly mean anomaly across 12 Total Carbon Column Observing Network sites (TCCON, 38): Bialystok, Poland; Darwin, Australia; Eureka, Canada; Garmisch, Germany; Karlsruhe, Germany; Lauder, New Zealand; Lauder, New Zealand; Lamont, Oklahoma; Orleans, France; Park Falls, Wisconsin; Sodankyla, Finland; Wollongong, Australia. Details of the GEOS-Chem TCCON comparison are reported by (63) and references therein. We note that the uncertainty in the GEOS-Chem trend
due to CARDAMOM flux uncertainty is substantial: global NCE 25th–75th percentile = -8 – +13Pg C yr⁻¹, which roughly corresponds to a ±5ppm growth rate (1). To evaluate the CARDAMOM seasonal NCE variability, we compare the linearly de-trended model and observations (Fig. S6).

Figures

Fig. S1: Left two columns: posterior GPP C allocation to autotrophic respiration (equivalent to 1 – C use efficiency), labile C, foliar C, fine roots, wood (mean, left column) and associated uncertainty (standard deviation, right column). Middle two columns: Posterior C residence time in foliar C, fine roots, wood, litter and soil C (log-based mean, left column) and associated uncertainty factors (based on logarithmic standard deviation, right column). Right two columns: Posterior mean 2001-10 C stocks in labile, foliar, fine roots, wood, litter C pools (mean, left column) and associated uncertainties (standard deviation, right column).

Fig. S2. Posterior median and 50% confidence ranges shown for 1° × 1° grid-cells B, T, D and W shown for the unperturbed results (S0) and sensitivity experiments S1-S12. The coordinates of B, T, D and W are reported in the Materials and Methods (locations shown in inset map). Across all locations, 88% of median sensitivity analysis estimates (sensitivity tests S1-S12) are within ±50% of unperturbed median C state and process variable retrievals.

Fig. S3. Pearson correlation coefficients ($r^2$, shown in color bar) between GLOBCOVER land-cover types fractions (x-axis) and C state and process variables (y-axis), based on their correlation across all 1° × 1° grid cells within the global study area. See section S5 for land-cover type acronyms.

Fig. S4: Maps: Eight primary 1° × 1° standardized empirical orthogonal functions (EOFs 1-8) derived from a principal component analysis of standardized C state and process variables (see section S6). The two dominant modes (EOF1 and EOF2) together reflect first-order global variations in C state/process variables ($c_i$) due to in latitude and precipitation, while higher order modes reflect increasingly complex spatial structures (however, EOFs 3-8 typically account for a smaller portion of $c_i$ spatial variability). Scatter plots: standardized EOF 1-8 coefficients corresponding to each C state/process variable (shown as symbol-color combinations). The linear sum of standardized EOFs 1-4 (1-8) and their associated coefficients reproduces 29-95% (88-99%) of C state/process variability (see Fig. 6 in the main text).

Fig. S5: CARDAMOM zonal profiles of median gross primary production, ecosystem respiration, fires and net C exchange (red). The 50% confidence range is depicted as a light-pink shaded area. The blue lines represent the 8 global MsTMIP models (64; see section S7 for details). The dashed black line denotes the flux-tower derived GPP (36). The continuous black line denotes the GFED version 3 total C emissions (36).

Fig. S6. 2009-2010 GEOS-Chem model - with CARDAMOM mode NCE - compared against mean monthly TCCON atmospheric column measurements across 12 TCCON sites: the left panel shows atmospheric CO₂ concentrations, and the right panel shows the linearly de-trended CO₂ anomalies. The de-trended comparison Pearson’s $r = 0.93$ and RMSE = 0.53 ppm.
Tables

Table S1: DALEC2 parameters, descriptions and prior ranges (the DALEC2 equations are fully described in (35)).

Table S2: Sensitivity tests for C allocation, residence times and C pool size estimates at locations B, T, D and W.

Table S3: In-situ observations and CARDAMOM posterior state and process variable estimates.
**CARDAMOM**

median (mode for NCE)

**CARDAMOM**

50% C.R.

**FLUXNET**

derived GPP

*Jung et al., 2009*

**MsTMIP models**

*Huntzinger et al., 2013.*

**GFEDv3**

*van der Werf et al., 2010*
Table S1: DALEC2 parameters, descriptions and prior ranges (the DALEC2 equations are fully described in (35)).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Prior Range</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Allocation fractions</strong></td>
<td></td>
</tr>
<tr>
<td>Autotrophic Respiration</td>
<td>0.2-0.8*</td>
</tr>
<tr>
<td>Labile</td>
<td>0.01-0.5</td>
</tr>
<tr>
<td>Foliage</td>
<td>0.01-0.5</td>
</tr>
<tr>
<td>Fine roots</td>
<td>0.01-0.5</td>
</tr>
<tr>
<td>Wood</td>
<td>0.01-0.5</td>
</tr>
<tr>
<td><strong>Turnover rates</strong></td>
<td></td>
</tr>
<tr>
<td>Woody C turnover rate</td>
<td>$2.5 \times 10^{-5} – 10^{-3}$ d$^{-1}$</td>
</tr>
<tr>
<td>Fine root turnover rate</td>
<td>$10^{-4} – 10^{-2}$ d$^{-1}$</td>
</tr>
<tr>
<td>Litter turnover rate</td>
<td>$10^{-4} – 10^{-2}$ d$^{-1}$</td>
</tr>
<tr>
<td>Soil organic C turnover rate</td>
<td>$10^{-7} – 10^{-3}$ d$^{-1}$</td>
</tr>
<tr>
<td>Litter mineralization rate</td>
<td>$10^{-2} – 10^{-5}$ d$^{-1}$</td>
</tr>
<tr>
<td>Exponential temperature dependence</td>
<td>0.018-0.08</td>
</tr>
<tr>
<td><strong>Canopy parameters</strong></td>
<td></td>
</tr>
<tr>
<td>Leaf onset day</td>
<td>1-365.25</td>
</tr>
<tr>
<td>Leaf fall day</td>
<td>1-365.25</td>
</tr>
<tr>
<td>Canopy efficiency</td>
<td>5 – 50*</td>
</tr>
<tr>
<td>Leaf C mass per leaf area (LCMA)</td>
<td>5 - 200 gC m$^{-2}$</td>
</tr>
<tr>
<td>Annual leaf loss fraction</td>
<td>1/8 – 1</td>
</tr>
<tr>
<td>Labile C release period</td>
<td>10 – 100 days</td>
</tr>
<tr>
<td>Leaf fall period</td>
<td>20 – 150 days</td>
</tr>
<tr>
<td><strong>Initial C stocks</strong></td>
<td></td>
</tr>
<tr>
<td>Labile C</td>
<td>1-2000gC m$^{-2}$</td>
</tr>
<tr>
<td>Foliar C</td>
<td>1-2000gC m$^{-2}$</td>
</tr>
<tr>
<td>Fine root C</td>
<td>1-2000gC m$^{-2}$</td>
</tr>
<tr>
<td>Litter C</td>
<td>1-2000gC m$^{-2}$</td>
</tr>
<tr>
<td>Above &amp; Below ground wood</td>
<td>1 - 100,000gC m$^{-2}$</td>
</tr>
<tr>
<td>Soil C (1m depth)</td>
<td>1 - 200,000gC m$^{-2}$</td>
</tr>
</tbody>
</table>

* Autotrophic Respiration and Canopy efficiency parameter log-normal prior distributions are described in section S1.
Table S2: Sensitivity tests for C allocation, residence times and C pool size estimates at locations B, T, D and W.

<table>
<thead>
<tr>
<th>Sensitivity Test(s)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1 &amp; S2</td>
<td>+20% &amp; -20% in LAI observations</td>
</tr>
<tr>
<td>S3 &amp; S4</td>
<td>+20% &amp; -20% increase in biomass observations</td>
</tr>
<tr>
<td>S5 &amp; S6</td>
<td>+20% &amp; -20% increase in HWSD Soil Carbon observations</td>
</tr>
<tr>
<td>S7 &amp; S8</td>
<td>+20% (^1) &amp; -20% increase in fire combustion factors</td>
</tr>
<tr>
<td>S9 &amp; S10</td>
<td>+20% &amp; -20% increase in fire resilience factor</td>
</tr>
<tr>
<td>S11</td>
<td>Use mean 1° × 1° aggregated MPI GPP (36) as driver</td>
</tr>
<tr>
<td>S12</td>
<td>No heterotrophic respiration(^2) under -10°C</td>
</tr>
</tbody>
</table>

\(^1\)Foliar combustion factor increase by 10% (from 0.9 to 0.99).
\(^2\)Respiration temperature dependence coefficient (19) set to zero at <-10°C, scaled by unity at >0°C, and scaled from 0 to 1 between -10°C and 0°C.
Table S3: In-situ observations and CARDAMOM posterior state and process variable estimates.

<table>
<thead>
<tr>
<th>Measurement (region)</th>
<th>CARDAMOM range*</th>
<th>In-situ observations (study)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fine roots (Amazon river basin)</td>
<td>9.2 – 10.8 tC ha(^{-1}) (2.8 – 11.5 tC ha(^{-1}))</td>
<td>5 – 8 tC ha(^{-1})(^{(a)}) (65)</td>
</tr>
<tr>
<td>Fine roots (North-East U.S.; &gt;30°N, &gt;100°W)</td>
<td>1.6 – 3.3 tC ha(^{-1}) (0.9 – 6.0 tC ha(^{-1}))</td>
<td>1.25 tC ha(^{-1})(^{(a)}) (66)</td>
</tr>
<tr>
<td>Fine root residence time (North-East U.S., &gt;30°N, &gt;100°W)</td>
<td>1.1 – 1.5 yrs (0.9 – 3.2 yrs)</td>
<td>0.83 – 1.25 yrs(^{(a)}) (66)</td>
</tr>
<tr>
<td>Fine root RT (global: where woody C &gt; 10tC ha(^{-1}))</td>
<td>1.2 – 2.6 yrs (0.9 – 4.7 yrs)</td>
<td>1.25 – 2.5 yrs(^{(b)}) (67)</td>
</tr>
<tr>
<td>Wood Carbon RT (Amazon river basin)</td>
<td>15 – 21 yrs (9 – 24 yrs)</td>
<td>~20 – 70 yrs(^{(b)}) (20, above-ground only)</td>
</tr>
<tr>
<td>Carbon Use Efficiency (CUE)(^{***}) (Amazon river basin)</td>
<td>0.42 – 0.43 (0.42 – 0.45)</td>
<td>Amazon field sites: 0.32 – 0.47(^{(b)}) (39)</td>
</tr>
<tr>
<td>Fine root C (Lat &gt; 66°N)</td>
<td>0.3 – 0.4 tC ha(^{-1}) (0.2 – 0.6 tC ha(^{-1}))</td>
<td>Arctic Ecosystems: 0.1 – 5 tC ha(^{-1})(^{(b)}) (68)</td>
</tr>
</tbody>
</table>

\(^*\) Area-weighted 25\(^{th}\) – 75\(^{th}\) %ile (5\(^{th}\) – 95\(^{th}\) %ile) \(1^\circ \times 1^\circ\) C state and process variables (see Materials and Methods).

\(^{**}\) Dry mass to C mass conversion factor = 0.5

\(^{***}\) CUE = 1 – autotrophic respiration fraction

\(^{(a)}\) Individual site range

\(^{(b)}\) Regional or global range