Response to comments

Manuscript ID: egusphere-2026-2241

Title: Constraining boreal carbon allocation and turnover by assimilating forest growth dynamics in a differentiable framework

Journal: Geoscientific Model Development

Reviewer #1: T. Luke Smallman, David T. Milodowski, Tim Green

General Comments:

Comment #1

The manuscript is a study using an intermediate complexity model of the terrestrial carbon cycle and constraining it at pixel-scale with a variety of data covering different components of the carbon cycle. The text is well written, with a good flow. However, we do have significant comments to make on this manuscript – and hope the authors recognise them as intended to be constructive.

The key conclusion appears to be that time series information on wood stocks are needed to constrain carbon allocation to, and then residence time of, the wood pool. However, this conclusion is not new. The need for time series information on wood stocks has been proposed and demonstrated at site level more than a decade ago with EO time series of woody biomass used in regional and global scale analyses (further details later). Possibly we have misunderstood and the novelty the specific method being used? In which case, the choice of using a constructed chronosequence is an interesting one but not justified in the existing text.

The authors' goal is valuable – namely constraining the slow(er) carbon pools in order to understand current carbon cycling. This understanding would enhance estimates of future responses. We strongly support the use of data assimilation (DA) approaches which provide local calibration, reflecting a realistic spatial variability of biodiversity / functional traits and avoids simplistic plant functional type approaches.

Overall, the paper has significant weaknesses and needs to be revised. We have several significant concerns.

Response #1

Thank you for your constructive comments. We have revised the manuscript accordingly, with the specific changes detailed in our point-by-point responses below.

Comment #2

Firstly, related to the novelty of the analysis presented. There is a significant body of published literature on DA for carbon cycle modelling that has been ignored by the current study. This literature includes recognition and demonstration of the importance of constraining the allocation of productivity and turnover rates for the long-lived carbon pools.

Fox et al. (2009) presented results from a series of data assimilation experiments using a version of DALEC and multiple different DA algorithms, highlighting that you gain little constraint on allocation and turnover of the long-lived wood C pool when reliant on constraints from observations of "fast" C-cycle processes (NEE, LAI) regardless of retrieval algorithm. Previously, Williams et al., (2005) identified the need for long-term measurements of woody biomass to improve constraints of turnover rates and residence times.

Smallman et al. (2017) presented results from a series of DA experiments using the CARDAMOM DA framework, clearly showing that assimilation of repeated woody biomass observations constrains decadal ecosystem carbon cycle uncertainty in aggrading forests. This study compared high information content analyses using local observations, including repeat observations of woody biomass over time, with retrievals using progressively sparser remotely sensed information (repeated, single, and no woody biomass observations). The results demonstrated how the inverse model representation of C cycling varies with different levels of assimilated biomass information. Assimilation of repeated biomass observations reduced uncertainty and/or bias in all ecosystem C pools not just wood, compared to analyses using single or no stock information.

Considering larger spatial domains, Smallman et al. (2021) explicitly quantified the covariance structure between uncertainty in wood dynamics and fast / slow observables, also demonstrating clearly that fast processes are weakly connected to the turnover dynamics of wood and soil.

George-Chacon et al. (2023) assimilated biomass data collected across forest regrowth chronosequences to help constrain the calibration of wood C dynamics and thus improve confidence in model forecasts. A chronosequence approach was used here due to the complex fine spatial scale disturbance patterns inherent in the local management regime.

Moreover, the use of repeat biomass observations is now commonplace in DA studies for C cycle modelling (e.g. Ge et al., 2018; Bilir et al., 2025; Williams et al., 2025; Friedlingstein et al., 2025).

We do not dispute the finding that assimilating observations on biomass accumulation leads to improved constraints on woody turnover rates and residence times, which are otherwise poorly characterised by observations of fast processes. However, given the abundance of existing literature it is unclear what new knowledge we have gained from the experiments presented.

Response #2

Thank you for providing a great summary of previous relevant studies. We agree that the original Introduction did not adequately acknowledge the relevant model-data fusion literature and could be read as presenting an established finding as new. We have therefore added the following paragraph (**Line 71–81**):

"Previous model-data fusion studies have established that short-term observations of net ecosystem exchange (NEE) and LAI can constrain fast canopy processes while carrying little information about woody carbon allocation and turnover. This limitation persists across different inversion algorithms and motivates the use of long-term stem inventories or other observations directly linked to slow carbon pools (Fox et al., 2009; Williams et al., 2009; Williams et al., 2005). Experiments comparing calibration with repeated woody biomass observations, a single observation, or no biomass observations in aggrading forests showed that repeated woody information reduced uncertainty and bias across ecosystem carbon pools (Smallman et al., 2017). Landscape-scale assimilation of biomass chronosequences has also been used to constrain biomass dynamics in tropical dry forests (George-Chacon et al., 2023). More recent large-domain CARDAMOM applications have expanded spatial coverage using a single biomass stock, biomass maps from several years, or annual biomass-change information (Bilir et al., 2025; Smallman et al., 2021; Williams et al., 2025), but these biomass constraints are tied to calendar years rather than stand age."

Accordingly, we no longer present the limited constraint of woody processes by fast observations, the use of repeated or chronosequence biomass information, or DifferLand itself as new. We clarified the specific new element evaluated in this study as follows (**Line 108–112**):

"Building on these earlier site- and landscape-scale studies, the new element evaluated here is the regional-

scale assimilation of grid-specific, century-scale biomass–age curves across the boreal domain. We compare the baseline experiment without biomass constraints, single-year biomass-constrained experiment, and growth-curve-constrained experiment, and conduct a spatial analysis of the resulting effective allocation and turnover parameters."

We agree that the general value of biomass observations for constraining slow carbon dynamics has been established and have revised the manuscript to acknowledge the relevant literature. However, we also wish to clarify that model development also requires evaluating established concepts across different model structures, datasets, spatial scales, and calibration designs. Here, our contribution is the regional-scale implementation of grid-specific, century-scale biomass–age constraints in a differentiable DALEC framework. To our knowledge, biomass–age relationships of this form have not previously been assimilated in model–data fusion studies. We have revised the manuscript to distinguish this specific model-development contribution from the broader ecological principle established by earlier studies.

Reference

Bilir, T. E., Bloom, A. A., Konings, A. G., Liu, J., Parazoo, N. C., Quetin, G. R., Norton, A. J., Worden, M. A., Levine, P. A., Ma, S., Braghiere, R. K., Longo, M., Bowman, K., Saatchi, S., Schimel, D. S., Miller, C. E., O'Sullivan, M., Kang, Y., Pandey, S., Patton, A. J., Yang, Y., and Liu, Y.: Satellite-Constrained Reanalysis Reveals CO2 Versus Climate Process Compensation Across the Global Land Carbon Sink, AGU Advances, 6, e2025AV001689, 10.1029/2025AV001689, 2025.

Fox, A., Williams, M., Richardson, A. D., Cameron, D., Gove, J. H., Quaife, T., Ricciuto, D., Reichstein, M., Tomelleri, E., Trudinger, C. M., and Van Wijk, M. T.: The REFLEX project: Comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data, Agricultural and Forest Meteorology, 149, 1597-1615, 10.1016/j.agrformet.2009.05.002, 2009.

George-Chacon, S. P., Smallman, T. L., Dupuy, J. M., Hernández-Stefanoni, J. L., Milodowski, D. T., and Williams, M.: Isolating the effects of land use and functional variation on Yucatán's forest biomass under global change, Frontiers in Forests and Global Change, Volume 6 - 2023, 10.3389/ffgc.2023.1204596, 2023.

Smallman, T. L., Exbrayat, J. F., Mencuccini, M., Bloom, A. A., and Williams, M.: Assimilation of repeated woody biomass observations constrains decadal ecosystem carbon cycle uncertainty in aggrading forests, Journal of Geophysical Research: Biogeosciences, 122, 528-545, 10.1002/2016JG003520, 2017.

Smallman, T. L., Milodowski, D. T., Neto, E. S., Koren, G., Ometto, J., and Williams, M.: Parameter uncertainty dominates C-cycle forecast errors over most of Brazil for the 21st century, Earth Syst. Dynam., 12, 1191-1237, 10.5194/esd-12-1191-2021, 2021.

Williams, M., Schwarz, P. A., Law, B. E., Irvine, J., and Kurpius, M. R.: An improved analysis of forest carbon dynamics using data assimilation, Global Change Biology, 11, 89-105, 10.1111/j.1365-2486.2004.00891.x, 2005.

Williams, M., Milodowski, D. T., Smallman, T. L., Dexter, K. G., Hegerl, G. C., McNicol, I. M., O'Sullivan, M., Roesch, C. M., Ryan, C. M., Sitch, S., and Valade, A.: Precipitation–fire functional interactions control biomass stocks and carbon exchanges across the world's largest savanna, Biogeosciences, 22, 1597-1614, 10.5194/bg-22-1597-2025, 2025.

Williams, M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D. Y., Kattge, J., Leuning, R., Luo, Y., Tomelleri, E., Trudinger, C. M., and Wang, Y. P.: Improving land surface models with FLUXNET data, Biogeosciences, 6, 1341-1359, 10.5194/bg-6-1341-2009, 2009.

Comment #3

Secondly, the experimental design choices have not been adequately justified. As a result, we are not clear that the analysis is robust or making the best use of the observations – though we recognise this manuscript is submitted to GMD. It would help to make clearer which element is the 'new development', the framework

itself or the choice of data constraints? Or both? The methodological choices have not been adequately justified. The chosen cost-function lacks accounting of observation uncertainty in the calibration; neglecting uncertainty is the most serious concern regarding the robustness of the study. Likewise, their validation uses highly uncertain data, with potentially poor scale matching between the analysis and the data – the implications of which are not discussed. The largest challenge for the validation is also the neglecting of their respective uncertainties.

Response #3

Thank you for raising these concerns. As detailed in our response to **Reviewer #1 Comment #2**, the new element evaluated here is the regional-scale assimilation of grid-specific, century-scale biomass–age curves across the boreal domain. Their construction, applicability, and comparison with multi-year ESA CCI AGB are addressed in our response to **Reviewer #1 Comment #4**.

Observational uncertainties were not included in the calibration objective because comparable uncertainty estimates were not available for all calibration inputs. We therefore revised the manuscript in two respects. First, we conducted sensitivity analyses across the boreal domain to assess how the retrieved parameters and differences between calibration experiments respond to changes in observations and the weight of the biomass term. Second, we clarified the observational uncertainty and differences in spatial representativeness between field sites and model grid cells that affect the comparison with field-derived allocation estimates.

1 Sensitivity to observational uncertainty

To assess how observational uncertainty affects the retrieved parameters and differences between experiments, we tested changes in GPP and biomass inputs across the boreal domain. We also tested the effect of halving and doubling the weight of the biomass term in the objective function. Text S6 now describes these tests and their results:

"Text S6 Sensitivity to observational errors and biomass-term weight

For the growth-curve-constrained experiment, we used additional biomass–age curves fitted separately to the lower and upper error limits reported in ESA CCI Biomass. We recalibrated the model with each curve and compared the results with those obtained using the original curve. We also halved and doubled the weight of the biomass term while retaining the original curve. For all three calibration experiments, we tested lower and upper GPP time series defined as max(GPP − SE, 0) and GPP + SE, where SE is the reported FluxSat standard error.

For the curves fitted to the lower and upper biomass error limits, median changes in wood residence time were −10.13 and +13.34 years, respectively (Fig. S11). The corresponding median changes in wood NPP allocation were −0.060 and +0.017. Halving and doubling the biomass weight produced much smaller median changes in wood residence time (+0.41 and −0.39 years) and allocation (−0.0019 and +0.0026).

The GPP tests also affected the retrieved parameters (Fig. S10). In the growth-curve-constrained experiment, lower and upper GPP produced median changes of +10.77 and −4.63 years in wood residence time, respectively. The corresponding median changes in wood NPP allocation were −0.023 and +0.008. The baseline and single-year biomass-constrained experiments showed larger median changes in wood allocation.

We compared the growth-curve-constrained and baseline experiments using the spatial means of wood NPP allocation and wood residence time. The growth-curve-constrained experiment retained lower mean wood allocation and longer mean wood residence time under all tested changes in GPP, biomass curves and biomass weight. Under lower GPP, the mean allocation difference (growth-curve-constrained minus baseline) narrowed

from −0.112 to −0.024. With the upper biomass curve, it narrowed from −0.085 to −0.008. Mean wood residence time remained 16.42–38.02 years longer in the growth-curve-constrained experiment across the sensitivity tests. In the tests with lower GPP and with the upper biomass curve, most grid cells had higher wood allocation in the growth-curve-constrained experiment, although its spatial mean remained lower."

We have also summarized the sensitivity tests in the Methods (**Line 391–393**):

"To assess sensitivity to observational uncertainty, we repeated calibration using lower and upper GPP estimates based on the reported standard errors. We also used biomass–age curves fitted separately to the lower and upper biomass error limits. Additional tests halved and doubled the weight of the biomass term (Text S6)."

The Results summarize the main findings (**Line 480–482**):

"Under all tested changes, the growth-curve-constrained experiment had lower mean wood NPP allocation and longer mean wood residence time than the baseline experiment. However, the allocation difference became much smaller under lower GPP and the upper biomass curve (Text S6, Fig. S10, and Fig. S11)."

2 Uncertainty and representativeness of the field comparison

The field-derived allocation estimates carry their own measurement and methodological uncertainties. Although spatial aggregation harmonizes the nominal resolutions of the observational data and model outputs, it does not eliminate differences in spatial representativeness between sampled sites and grid-cell means. We therefore clarified the limitations in Supplementary Text S4:

"Aggregating the plot data and high-resolution reference curves to 0.5° harmonized their nominal spatial resolutions with that of the model grid cells, but it did not eliminate differences in spatial representativeness between field sites and grid-cell means. The field-derived allocation dataset did not provide observation-level standard deviations, standard errors, confidence intervals, or other quantitative uncertainty estimates (Xia et al., 2019). The comparison was therefore not uncertainty-weighted, and observational error bars could not be shown."

The Discussion also states these limitations (**Line 651–653**):

"The comparison with field-derived allocation estimates remains limited by differences between sampled sites and model grid-cell means. Uncertainty estimates were also unavailable for individual field observations."

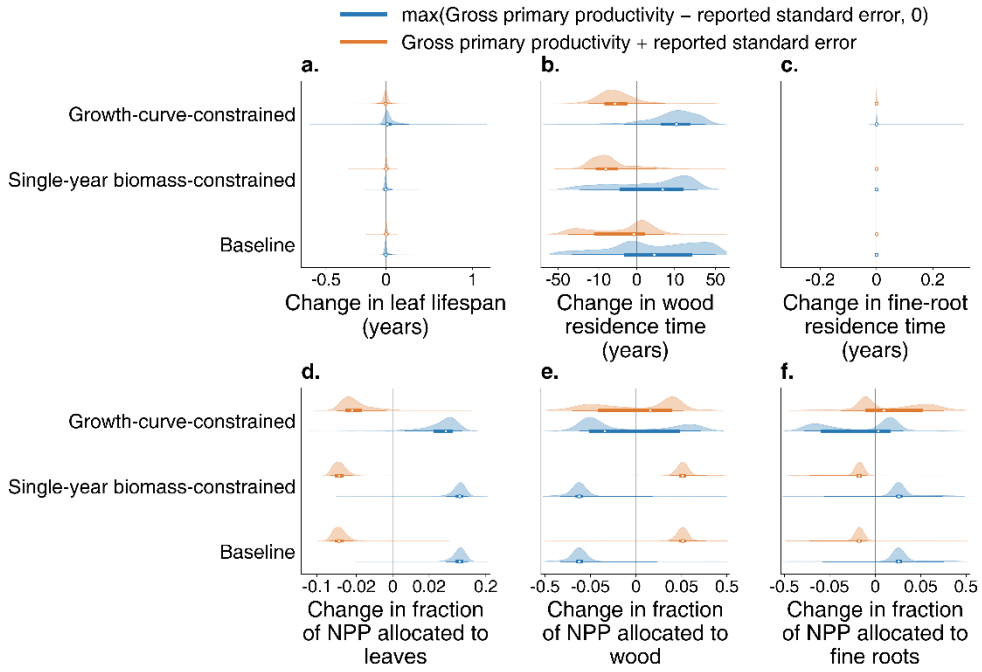


Fig. R1 (A copy of Fig. S10) Sensitivity of six retrieved parameters to lower and upper FluxSat GPP estimates. Changes are relative to the original calibration in each experiment. Half-violins show distributions across grid cells. Circles mark medians, thick bars show interquartile ranges, and thin bars show the 5th–95th percentiles. Horizontal axes use an inverse hyperbolic sine (asinh) scale to compress long tails, with tick labels in the original units.

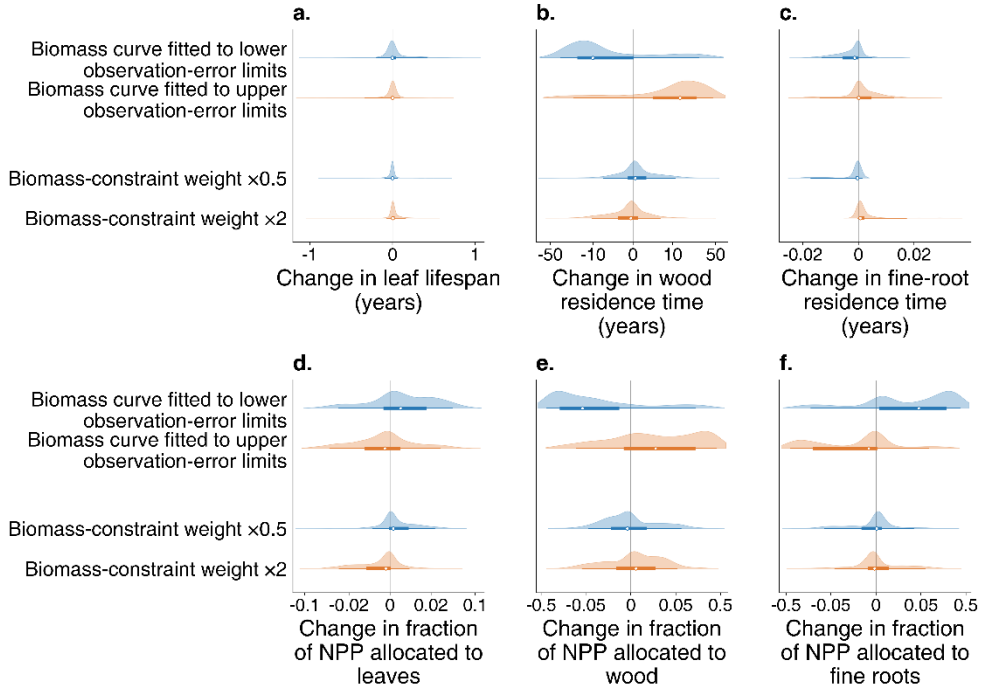


Fig. R2 (A copy of Fig. S11) Sensitivity of six retrieved parameters to biomass inputs and biomass-term weight in the growth-curve-constrained experiment. Tests used biomass–age curves fitted separately to the lower and upper observational error limits, or halved and doubled the biomass-term weight. Changes are relative to the original calibration. Half-violins show distributions across grid cells. Circles mark medians, thick bars show interquartile ranges, and thin bars show the 5th–95th percentiles. Horizontal axes use an inverse hyperbolic sine (asinh) scale to compress long tails, with tick labels in the original units.

Reference

Xia, J., Yuan, W., Lienert, S., Joos, F., Ciais, P., Viovy, N., Wang, Y.-p., Wang, X., Zhang, H., Chen, Y., and Tian, X.: Global Patterns in Net Primary Production Allocation Regulated by Environmental Conditions and Forest Stand Age: A Model-Data Comparison, Journal of Geophysical Research: Biogeosciences, 124, 2039-2059, doi.org/10.1029/2018JG004777, 2019.

Comment #4

A key aspect of the methodology presented is the construction of chronosequences to provide information on long-term dynamics of the wood C pool. At present it is very difficult to evaluate the approach with the information provided. However, in general there is a need for a much clearer description of the methods used to generate the local chronosequences, and how model fit to the "observed" chronosequence is assessed in the cost function. There needs to be a clear assessment of the uncertainty associated with the methods, and a detailed interrogation of the underlying assumptions.

There are a number of challenges to constructing a representative chronosequence from 100m data that is representative of each 0.5 degree pixel. The use of recovery from detected fire disturbances seems challenging for many boreal forests, particularly in Europe, where the forests are intensively managed, and fire activity is rare due to rapid fire suppression when fires do occur. How does the approach suggested account for multi-modal disturbance and the impact of management on woody turnover and biomass stocks?

The approach to estimate "old-growth" biomass based on the local 90th percentile does not allow for environmental gradients, which may result in variation in old-growth forest biomass, and may lead to biased estimates based on a percentile when aggregated to the model pixel level. It is also unclear how such an approach applies in parts of the boreal domain where a large majority of the forest stands are in managed systems.

Relatedly, it would also be beneficial to understand how the information content of the constructed chronosequences compares to assimilating the time series of AGB data that are available from the full ESA-CCI dataset. This might be a more informative experiment than the ones provided, but it would be dependent on robust uncertainty propagation through the chronosequences into the analysis. This is needed to justify why chronosequence is needed over existing datasets, given the additional layers of assumptions and complexity needed to derive them.

Response #4

Thank you for raising this important concern. We agree that the manuscript needs a clearer description of how the biomass–age curves were constructed and used in calibration. We also clarify the assumptions involved in applying these curves to managed forests and selecting mature-forest pixels, and explain our reasons for using the curves rather than the available multi-year AGB maps. We address these points separately below.

1 Construction of the biomass–age curves and their use in calibration

We clarified the construction of the curves from Xu et al. (2026a) in the Methods (**Line 195–198**):

"To constrain long-term carbon turnover and allocation, we used the ESA CCI-based biomass–age curves (Xu et al., 2026). In the source methodology, a 100 m ESA CCI aboveground biomass map is spatially matched with Landsat-derived time since the last detected fire (Long et al., 2019). Within each 0.5° grid cell, biomass pixels are grouped by stand age, and the median aboveground biomass for each stand age is fitted with a Chapman–Richards function."

The Methods describe the conversion from aboveground to total biomass (**Line 217–221**):

"To ensure structural consistency between the observational constraint and the model's state variables, we converted both the satellite AGB-age trajectories fitted by Xu et al. (2026) and the separate 2020 ESA CCI AGB into total biomass constraints (sum of above- and belowground biomass). This conversion was performed using spatially explicit Root-to-Shoot (R:S) ratios derived from a global map of root biomass, which was generated by a random forest model trained on extensive field measurements (Huang et al., 2021)."

The Methods also explain how the curves were used in calibration (**Line 358–361**):

"For the growth-curve constraint, modeled total live biomass was calculated as the sum of the labile, foliar, root, and wood carbon pools at the end of each simulation year. The 100 annual values were compared with the biomass–age curve at stand ages 1–100 using NNSE in the observational objective."

Xu et al. compared satellite-derived biomass curves with 2,548 field AGB–age observations from 751 sites grouped into 159 0.5° grid cells, obtaining a mean grid-level R^2 of 0.56. In Canada, age-related recovery patterns derived from the shorter fire record were broadly similar to those derived from forest-age records spanning 150 years (**Fig. R4**). These checks support the construction method. Moreover, a comparable space-for-time approach combining satellite biomass, disturbance history, and Chapman–Richards fitting has also been applied to tropical forests (Xu et al., 2026b). **Fig. R3** shows examples of the biomass-age curves used in our calibration and the corresponding biomass simulated by DALEC.

We have added a brief description of these comparisons to the Methods (**Line 206–208**):

"Xu et al. (2026a) evaluated this method against field biomass–age observations, obtaining a mean grid-level R^2 of 0.56. They also found similar age-related biomass patterns using a Canadian forest-age map that included stands up to 150 years old."

The sensitivity tests using biomass–age curves fitted to the lower and upper biomass error limits are described in our response to **Reviewer #1 Comment #3**.

2 Applicability of the curves and the mature-forest mask

We have clarified the use of the curves as general biomass–age relationships and the implications of forest management in the Discussion (**Line 625–630**):

"Although stand age was estimated from fire history and the curves were fitted using post-fire observations, we use them as general empirical biomass–age relationships. These relationships may provide a useful approximation of biomass accumulation following other disturbances, including harvesting, drought-induced mortality, insect outbreaks and windthrow. Where management has affected the sampled stands, its effects may be reflected in the observed biomass. However, intensive management during stand development may alter biomass accumulation in ways that are not well represented by these curves."

For the mature-forest mask, the 90th-percentile threshold is calculated separately within each 0.5° grid cell and therefore reflects local biomass differences associated with environmental conditions and management. It identifies relatively high-biomass stands under local conditions, including in managed forests, although selection can still favor more productive sites within a cell.

3 Reasons for using biomass–age curves rather than multi-year AGB maps

We have added this explanation to the Methods section (**Line 209–214**):

"In the single-year biomass-constrained experiment, we used the 2020 ESA CCI Biomass v6 AGB map (Santoro and Cartus, 2025) to provide a static constraint for mature forest states. The ESA CCI-based Xu

curves describe fitted biomass–age relationships, whereas the 2020 map provides the contemporary biomass information used for the mature-forest mask and the single-year biomass constraint. The available ESA CCI v6 maps span 2007–2022 and do not by themselves provide the same age-dependent information. The product user guide cautions that biases can affect estimated AGB changes, particularly between 2007 or 2010 and 2015–2022 (Santoro and Cartus, 2025)."

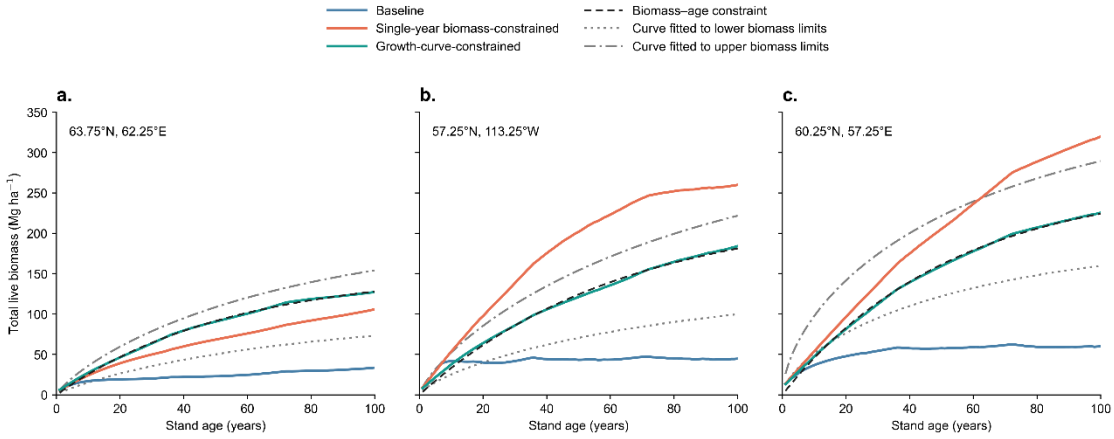


Fig. R3 (A copy of Fig. S5) Examples of the biomass–age constraints (black dashed lines) and biomass simulated under the three calibration experiments.

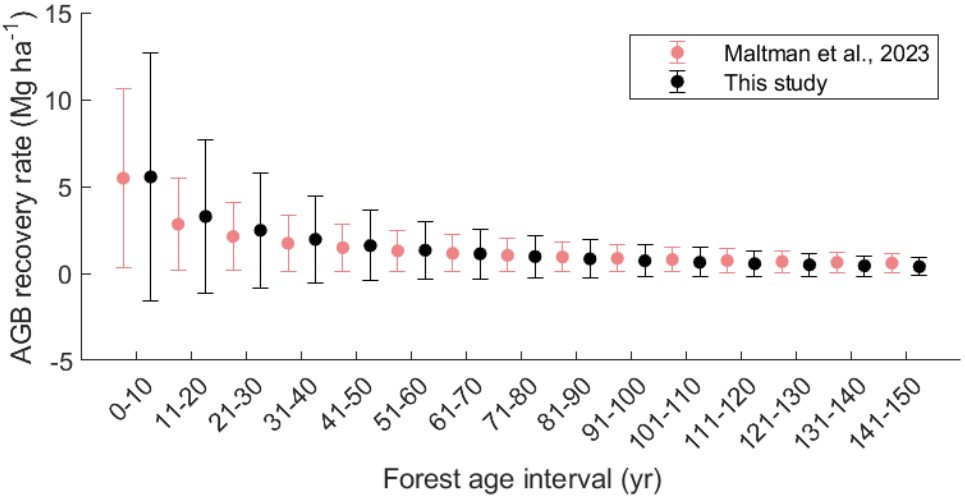


Fig. R4 Comparison of AGB recovery rates derived using short and long forest-age records in Canadian boreal forests. Reproduced from Supplementary Fig. S11 of Xu et al. (2026a), "Enduring carbon sinks in boreal forests despite recent acceleration of fires" (manuscript under review). Recovery curves were fitted using the same biomass dataset with either the shorter fire record used by Xu et al. (2026a) or the forest-age map of Maltman et al. (2023), spanning 150 years. Points and error bars show the mean and standard deviation of grid-cell recovery rates within 10-year age intervals. "This study" in the legend refers to Xu et al. (2026a).

Reference

Maltman, J. C., Hermosilla, T., Wulder, M. A., Coops, N. C., and White, J. C.: Estimating and mapping forest age across Canada's forested ecosystems, Remote Sensing of Environment, 290, 113529, 10.1016/j.rse.2023.113529, 2023.

Santoro, M. and Cartus, O.: ESA Biomass Climate Change Initiative (Biomass_cci): Global datasets of forest above-ground biomass for the years 2007, 2010, 2015, 2016, 2017, 2018, 2019, 2020, 2021 and 2022, v6.0, NERC EDS Centre for Environmental Data Analysis [dataset], 10.5285/95913FFB6467447CA72C4E9D8CF30501, 2025.

Xu, Y., Ciais, P., Li, W., Saatchi, S., Santoro, M., Cescatti, A., Schepaschenko, D., He, G., Ceccherini, G., He, J., Fan, L., Brandt, M., Fensholt, R., Wigneron, J.-P., Kay, H., Sitch, S., Bastos, A., Bowing, S., Ritter, F., Fayad, I., and Schwartz, M.: Enduring carbon sinks in boreal forests despite recent acceleration of fires, Submitted, 2026a.

Xu, Y., Ciais, P., Santoro, M., Bourgoin, C., Ritter, F., Pellissier-Tanon, A., Feng, Y., Zhou, C., He, G., Heinrich, V., Besnard, S., Robinson, N., Cook-Patton, S. C., Chave, J., Aragao, L. E. O. C., Ometto, J. P., Bowring, S. P. K., Fayad, I., Zhu, L., Su, Y., Wigneron, J.-P., and Li, W.: Small persistent humid forest clearings drive tropical forest biomass losses, Nature, 649, 375-380, 10.1038/s41586-025-09870-7, 2026b.

Comment #5

The authors' scale –appropriate corroboration is with Bloom et al., (2016), but its value is unclear. Firstly, for the global analysis presented by Bloom et al., there was no biomass observation constraint for the Boreal zone, as global aboveground biomass data at the time were limited to the tropics. The lack of observation constraint on woody biomass in the Bloom et al., 2016 analysis means that the parameters relevant to allocation and turnover carry high uncertainties, which propagate into estimates for residence time and other aspects of the C cycle. Secondly, this analysis is no longer the state-of-the-art representation of the state of the C cycle produced with the CARDAMOM framework. There have been three more recent Boreal / Arctic focussed CARDAMOM studies which include at least 1 (López-blanco et al., 2019) and sometimes repeat (Hugelius et al., 2024; Varney et al., 2026, from Trendy v13 and v14 respectively) biomass information. CARDAMOM has been contributed to the trendy model intercomparison across several years providing analyses using different combinations of repeat biomass constraint. These are all freely available to download (e.g. Trendy v14, https://datashare.ed.ac.uk/handle/10283/9174), and would provide a much more meaningful comparison against your current analysis.

Finally, the comparison with Bloom et al., (2016) lacks a comparison of the uncertainties in either analysis. CARDAMOM produces ensemble analyses for the carbon cycle, with the spread of the ensemble for a given state variable characterising the uncertainty propagated through the calibration into the diagnostic analysis. When benchmarking, it is important to account for these uncertainties to understand whether differences between approaches are significant.

Response #5

Thank you for this great suggestion.

(1) Comparison with CARDAMOM v13

We have replaced the comparison with Bloom et al. (2016) in Section 3.3 (**Line 559–580**) and Fig. 6 with the CARDAMOM v13 analysis associated with Hugelius et al. (2024), using the archive distributed by Smallman and Williams (2023). The associated study reported woody-biomass constraints for 2017 and 2018.

The publicly distributed outputs from López-Blanco et al. (2019) provide transit times for total vegetation, dead organic matter, and total carbon, but not for wood separately. Their allocation fractions are expressed relative to GPP, and the separately published marginal quantiles do not allow reconstruction of posterior intervals for NPP allocation. For CARDAMOM v14 (Smallman, 2027) the required tissue-allocation and residence-time outputs were not yet available to us at the time of analysis. We therefore used v13 for the revised quantitative comparison.

Based on this new comparison, we have rewritten Section 3.3 to report the original values, relative differences, and whether our allocation fractions, leaf lifespan, and residence times fall within the corresponding

CARDAMOM v13 5th–95th percentile intervals. The full revised section reads (**Line 559–580**):

3.3 Comparison with CARDAMOM v13

"We compared the allocation fractions, leaf lifespan, and residence times from our growth-curve-constrained experiment with the archive associated with Hugelius et al. (2024), hereafter CARDAMOM v13. CARDAMOM v13 assimilated woody-biomass observations for 2017 and 2018. We averaged the results from our 0.5° grid cells within each 1° CARDAMOM v13 grid cell. Details of the comparison methods are provided in Text S8.

The spatial mean fraction of NPP allocated to wood was 0.534 in our growth-curve-constrained experiment and 0.341 in CARDAMOM v13, with a PBIAS of +56.3% (Fig. 6a). The fraction of NPP allocated to wood in our experiment fell within the CARDAMOM v13 5th–95th percentile interval in 29.7% of the compared grid cells. Mean wood residence time was 43.9 years in our experiment and 50.6 years in CARDAMOM v13, with a PBIAS of −13.2% (Fig. 6b). Our wood residence times fell within the annual CARDAMOM v13 5th–95th percentile intervals in 53.4% of comparisons across grid cells and years (2001–2019).

The spatial mean fraction of NPP allocated to leaves was 0.228 in this study and 0.397 in CARDAMOM v13, with a PBIAS of −42.5% (Fig. S16a). Fine-root allocation had similar spatial means of 0.247 and 0.250, respectively, with a PBIAS of −1.3% (Fig. S16c). The spatial mean leaf lifespan in this study was 0.800 years, compared with 0.246 years for the CARDAMOM v13 foliar stock-to-outflux timescale, with a PBIAS of +225.0% (Fig. S16b). Fine-root residence times had spatial means of 0.314 years in this study and 0.401 years in CARDAMOM v13, with a PBIAS of −21.7% (Fig. S16d). Spatial ranges, paired differences, and comparisons with the CARDAMOM v13 intervals are provided in Fig. S17 and Tables S5 and S6.

Differences between the two studies may arise from model assumptions, input data, and calibration methods, and cannot be attributed solely to the biomass constraints. Comparisons with the published CARDAMOM v13 intervals provide context for these differences, but do not account for uncertainty in both studies and therefore do not establish statistical significance."

(2) Comparison with CARDAMOM v13 uncertainty intervals

Fig. S17 shows the widths of the CARDAMOM v13 5th–95th percentile intervals and compares our wood allocation fractions and wood residence times with those intervals.

The comparison methods are described in Text S8:

"Text S8 Comparison with CARDAMOM v13

Data source

We used the CARDAMOM v13 archive associated with Hugelius et al. (2024) and distributed by Smallman and Williams (2023). The distributed model outputs cover 2001–2019 at 1° resolution. The publicly documented DALEC4 implementation jointly optimizes initial carbon pools and process parameters.

Definitions of allocation fractions and residence times

Our allocation fractions describe NPP partitioning, with leaf allocation combining the fractions assigned to the leaf and labile pools. CARDAMOM v13 allocation fractions describe realized tissue growth relative to total tissue production, including labile-to-leaf release.

Our leaf lifespan is derived from the integrated annual fractional leaf-loss rate (Text S2), and wood and fine-root residence times are the reciprocals of calibrated annual turnover rates. CARDAMOM v13 reports annual carbon stock-to-total-outflux ratios, including explicitly represented disturbance losses.

Comparison with CARDAMOM v13 intervals

For allocation fractions, the CARDAMOM v13 archive provides one median and one 5th–95th percentile interval per grid cell. For residence times, it provides annual values, and we used the mean of the annual medians for 2001–2019. We calculated the percentages of our allocation fractions below, within, or above the corresponding intervals across grid cells. For leaf lifespan and residence times, these percentages were calculated across grid cells and years by comparing our value in each grid cell with each annual CARDAMOM v13 interval.

We compared only grid cells with values from both datasets and gave each 1° grid cell equal weight in the spatial summaries. PBIAS was calculated as 100 times the sum of paired differences divided by the sum of the corresponding CARDAMOM v13 values. Positive PBIAS indicates higher values in our experiment."

The revised Fig. 6, new Fig. S17, and Tables S5 and S6 are reproduced below. Fig. S16 provides the leaf and fine-root comparisons described in the revised Section 3.3 above.

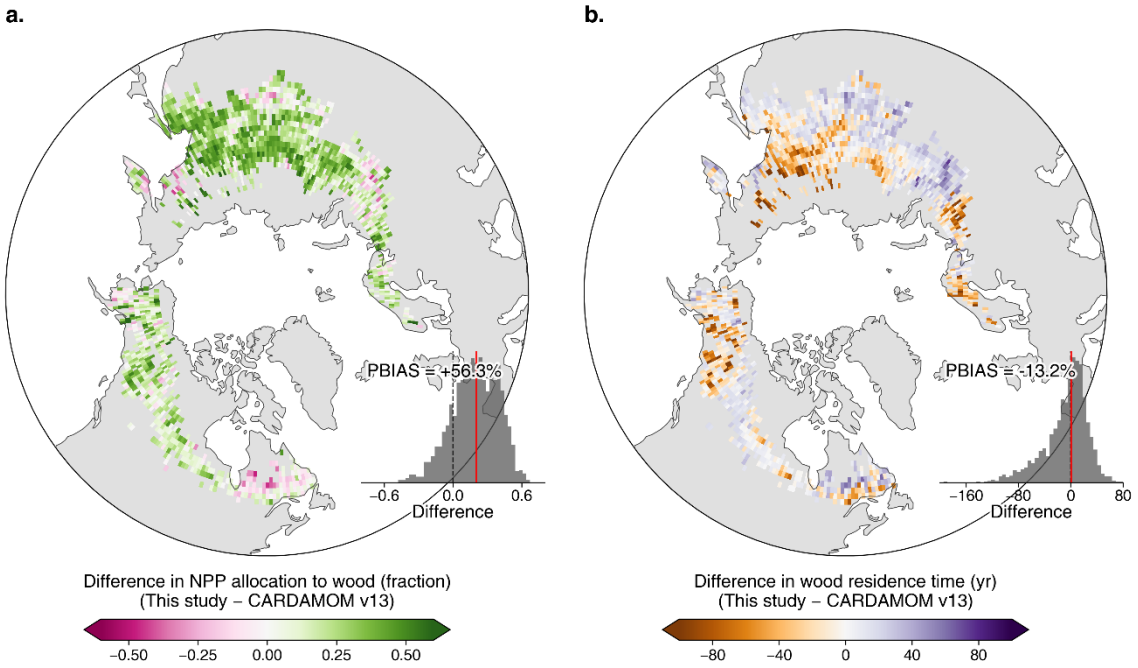


Fig. R5 (A copy of Fig. 6) Spatial differences in (a) the fraction of NPP allocated to wood and (b) wood carbon residence time between this study and CARDAMOM v13 (Hugelius et al., 2024). Differences are calculated as this study minus CARDAMOM v13 and are expressed as dimensionless fractions in (a) and years in (b). Insets show the distributions of paired differences, with solid red lines indicating the medians and dashed black lines indicating zero. PBIAS denotes the percent bias relative to CARDAMOM v13.

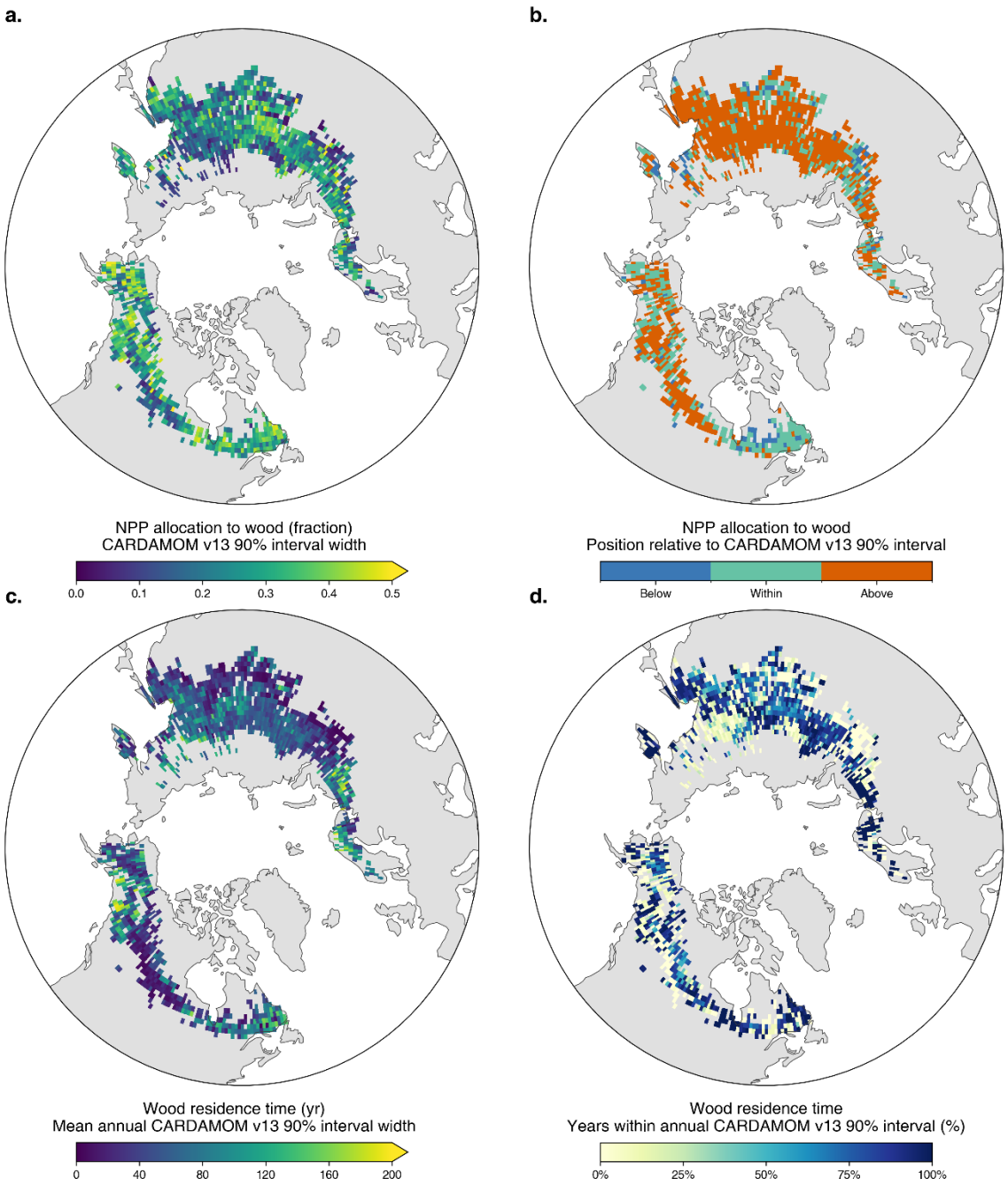


Fig. R6 (A copy of Fig. S17) Wood allocation and residence time compared with CARDAMOM v13 5th–95th percentile intervals. Panels show (a) the interval width for wood allocation, as a dimensionless fraction, (b) whether our wood allocation fraction falls below, within, or above that interval, (c) the mean width of annual wood residence-time intervals over 2001–2019, in years, and (d) the percentage of years in which our wood residence time falls within the corresponding annual interval.

Table R1 (from Table S5) Allocation fractions and residence times from our growth-curve-constrained experiment and CARDAMOM v13 in grid cells with values from both datasets.

Quantity	This study, mean (spatial P5–P95)	CARDAMOM v13, mean (spatial P5–P95)
Wood allocation (fraction)	0.534 (0.232–0.758)	0.341 (0.157–0.600)
Leaf allocation (fraction)	0.228 (0.151–0.310)	0.397 (0.261–0.535)
Fine-root allocation (fraction)	0.247 (0.040–0.591)	0.250 (0.076–0.541)
Wood residence time (year)	43.9 (25.4–70.3)	50.6 (14.0–118.9)
Leaf lifespan (year)	0.800 (0.500–1.264)	0.246 (0.164–0.385)
Fine-root residence time (year)	0.314 (0.311–0.318)	0.401 (0.203–0.879)

Table R2 (from Table S6) Paired differences and comparisons with CARDAMOM v13 intervals. Paired differences are calculated as our values minus the corresponding CARDAMOM v13 values

Quantity	Mean paired difference	PBIAS (%)	Spearman ρ	Below / within / above interval (%)
Wood allocation (fraction)	+0.192	+56.3	0.135	6.3 / 29.7 / 64.0
Leaf allocation (fraction)	-0.169	-42.5	-0.140	79.6 / 20.1 / 0.3
Fine-root allocation (fraction)	-0.00313	-1.3	0.057	45.2 / 24.1 / 30.7
Wood residence time (year)	-6.66	-13.2	0.176	21.5 / 53.4 / 25.0
Leaf lifespan (year)	+0.554	+225.0	-0.049	0.0 / 0.3 / 99.7
Fine-root residence time (year)	-0.0869	-21.7	-0.012	14.9 / 72.8 / 12.3

Reference

Bloom, A. A., Exbrayat, J.-F., van der Velde, I. R., Feng, L., and Williams, M.: The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times, Proceedings of the National Academy of Sciences, 113, 1285-1290, 10.1073/pnas.1515160113, 2016.

Hugelius, G., Ramage, J., Burke, E., Chatterjee, A., Smallman, T. L., Aalto, T., Bastos, A., Biasi, C., Canadell, J. G., Chandra, N., Chevallier, F., Ciais, P., Chang, J., Feng, L., Jones, M. W., Kleinen, T., Kuhn, M., Lauerwald, R., Liu, J., López-Blanco, E., Luijkx, I. T., Marushchak, M. E., Natali, S. M., Niwa, Y., Olefeldt, D., Palmer, P. I., Patra, P. K., Peters, W., Potter, S., Poulter, B., Rogers, B. M., Riley, W. J., Saunois, M., Schuur, E. A. G., Thompson, R. L., Treat, C., Tsuruta, A., Turetsky, M. R., Virkkala, A. M., Voigt, C., Watts, J., Zhu, Q., and Zheng, B.: Permafrost Region Greenhouse Gas Budgets Suggest a Weak CO2 Sink and CH4 and N2O Sources, But Magnitudes Differ Between Top-Down and Bottom-Up Methods, Global Biogeochemical Cycles, 38, e2023GB007969, 10.1029/2023GB007969, 2024.

López-Blanco, E., Exbrayat, J. F., Lund, M., Christensen, T. R., Tamstorf, M. P., Slevin, D., Hugelius, G., Bloom, A. A., and Williams, M.: Evaluation of terrestrial pan-Arctic carbon cycling using a data-assimilation system, Earth Syst. Dynam., 10, 233-255, 10.5194/esd-10-233-2019, 2019.

Smallman, T. L.: The contemporary global terrestrial carbon cycle - a systemic model-data fusion analysis, University of Edinburgh. School of GeoSciences. National Centre for Earth Observation [dataset], 2027.

Smallman, T. L. and Williams, M.: Carbon and water cycle analysis across global permafrost domain, University of Edinburgh. School of GeoSciences. Global Change Institute [dataset], 2023.

Introduction:

Comment #6

L44-45: "These uncertainties primarily stem from differences in how models represent key biogeochemical and biophysical processes…". This statement directly contradicts the conclusions Famiglietti et al., (2021) which is cited in the preceding sentence, which shows that parameter uncertainty was a dominant source of uncertainty – it is the underpinning argument for why predictive skill has not improved with added complexity. Likewise, Smallman et al., (2021), showed that parametric uncertainty alone across the Brazil was sufficient to generate a forecast spread equal in size to Earth System Model ensembles. This section could be framed better around the contrasting narratives, e.g. the traditional land surface model community which argues for increasing complexity verses the data assimilation community, which stresses a need for process complexity to be linked to and balanced by observational constraints.

Response #6

Thank you for this suggestion. We have revised the paragraph accordingly. The revised text (**Line 45–52**):

"Uncertainty in terrestrial carbon-cycle projections arises from interacting sources, including process representation and model structure, parameterization, meteorological and atmospheric forcing, initial states, and observational constraints. Their relative contributions depend on the model, data availability, spatial and temporal scale, and forecast experiment. Studies using data-assimilation ensembles have shown that parameter uncertainty alone can be large (Famiglietti et al., 2021; Smallman et al., 2021), while inter-model comparisons also demonstrate substantial effects of structural and process-representation choices (Canadell et al., 2021). Thus, increasing process complexity alone does not guarantee improved predictive skill unless the added processes and their parameters are adequately constrained by observations."

Reference

Canadell, J. G., Monteiro, P. M. S., Costa, M. H., Cotrim da Cunha, L., Cox, P. M., Eliseev, A. V., Henson, S., Ishii, M., Jaccard, S., Koven, C., Lohila, A., Patra, P. K., Piao, S., Rogelj, J., Syampungani, S., Zaehle, S., and Zickfeld, K.: Global Carbon and Other Biogeochemical Cycles and Feedbacks, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 673-816, 10.1017/9781009157896.007, 2021.

Famiglietti, C. A., Smallman, T. L., Levine, P. A., Flack-Prain, S., Quetin, G. R., Meyer, V., Parazoo, N. C., Stettz, S. G., Yang, Y., Bonal, D., Bloom, A. A., Williams, M., and Konings, A. G.: Optimal model complexity for terrestrial carbon cycle prediction, Biogeosciences, 18, 2727-2754, 10.5194/bg-18-2727-2021, 2021.

Smallman, T. L., Milodowski, D. T., Neto, E. S., Koren, G., Ometto, J., and Williams, M.: Parameter uncertainty dominates C-cycle forecast errors over most of Brazil for the 21st century, Earth Syst. Dynam., 12, 1191-1237, 10.5194/esd-12-1191-2021, 2021.

Comment #7

L62-64: Given this statement, isn't your later result statement that high frequency observations are not able to inform slow pools is invalid / redundant? It's already been established, and you don't present a hypothesis as to why this might not be the case. See Fox et al., (2009), Williams et al., (2009), Smallman et al., (2017) which made the same arguments earlier based on in-situ observations. Also, Smallman et al., (2021) which explicitly quantified the covariance structure between uncertainty in wood dynamics and fast / slow observables also showing that fast processes are weakly connected.

Response #7

We retained the statement about the limited information provided by high-frequency observations for slow carbon pools as background and expanded the supporting model-data fusion literature as described in our response to **Reviewer #1 Comment #2.** We deleted the redundant generalizing statement from the Results. The baseline experiment is retained only as a within-study reference for evaluating the additional information provided by a single-year biomass observation and the full biomass–age curve.

Comment #8

L69-71: The choice to create chronosequences of woody biomass from a single Earth Observation map of aboveground biomass has not been justified. Why not use the now more than decade long maps of woody biomass that exist? The EO maps are characterised by very large pixel-level uncertainties. How were these uncertainties from the source dataset propagated into the chronosequence, and how are the uncertainties in the chronosequence characterised and propagated into the calibration? It would be good to see some examples of the chronosequence data and their calibrated curves / statistical uncertainties.

Response #8

Thank you for this comment. We used the ESA CCI-based biomass–age curves provided by Xu et al. (2026). Their construction, underlying assumptions, and the reasons for using them instead of the multi-year ESA CCI maps are discussed in our response to **Reviewer #1 Comment #4.**

We assessed the effect of biomass observational uncertainty by calibrating with biomass–age curves fitted separately to the lower and upper limits of the reported biomass errors. These sensitivity tests are described in our response to **Reviewer #1 Comment #3** (Text S6 and Fig. S11). Examples of the biomass–age curves and corresponding biomass simulated by DALEC from all three experiments are shown in **Fig. R3**.

Reference

Xu, Y., Ciais, P., Li, W., Saatchi, S., Santoro, M., Cescatti, A., Schepaschenko, D., He, G., Ceccherini, G., He, J., Fan, L., Brandt, M., Fensholt, R., Wigneron, J.-P., Kay, H., Sitch, S., Bastos, A., Bowing, S., Ritter, F., Fayad, I., and Schwartz, M.: Enduring carbon sinks in boreal forests despite recent acceleration of fires, Submitted, 2026.

Comment #9

L83-85: Unclear whether this statement is correct. Constraint on woody pools and their dynamics is valuable across all forests. But arguably the biggest concern for Boreal forests is the carbon stored in their soils. Something which you later state in L87.

L87-89: Unclear, are you arguing that the soils are largely static, thus the net C balance is driven by biomass change? I think this is disputable, for example, recent RECCAP2 papers focused on the permafrost domain (e.g., Hugelius et al., 2024) where soil heterotrophic respiration is roughly half the size of photosynthesis suggesting a large role of the soil.

Response #9

We agree that constraints on woody pools and their dynamics are valuable across forest biomes rather than being uniquely important in boreal forests. We have revised the corresponding sentence to state (**Line 95–97**):

"Such slow-process constraints are valuable across forest biomes because they provide information on long-

lived biomass pools and their dynamics."

We also agree that our original wording overstated the role of biomass accumulation in the complete boreal carbon balance and could be interpreted as assuming that soil carbon is largely static. This was not our intention. We have revised the boreal framing to recognize the importance of soil and permafrost carbon and to define the scope of the present study more precisely (**Line 98–102**):

"Boreal forests store approximately 30% of global terrestrial carbon (Hansen et al., 2023), most of which resides in carbon-rich soils and permafrost (Hugelius et al., 2024; Schuur, 2019). Disturbance-induced, non-equilibrium live-biomass accumulation following stand-replacing events is an important component of boreal carbon dynamics (Jonsson and Wardle, 2009), making the region a useful testbed for evaluating biomass growth curves as model constraints. Accordingly, this study focuses on the biomass component of boreal forest carbon dynamics."

Reference

Hansen, W. D., Foster, A., Gaglioti, B., Seidl, R., and Rammer, W.: The Permafrost and Organic LayEr module for Forest Models (POLE-FM) 1.0, Geosci. Model Dev., 16, 2011-2036, 10.5194/gmd-16-2011-2023, 2023.

Hugelius, G., Ramage, J., Burke, E., Chatterjee, A., Smallman, T. L., Aalto, T., Bastos, A., Biasi, C., Canadell, J. G., Chandra, N., Chevallier, F., Ciais, P., Chang, J., Feng, L., Jones, M. W., Kleinen, T., Kuhn, M., Lauerwald, R., Liu, J., López-Blanco, E., Luijkx, I. T., Marushchak, M. E., Natali, S. M., Niwa, Y., Olefeldt, D., Palmer, P. I., Patra, P. K., Peters, W., Potter, S., Poulter, B., Rogers, B. M., Riley, W. J., Saunois, M., Schuur, E. A. G., Thompson, R. L., Treat, C., Tsuruta, A., Turetsky, M. R., Virkkala, A. M., Voigt, C., Watts, J., Zhu, Q., and Zheng, B.: Permafrost Region Greenhouse Gas Budgets Suggest a Weak CO2 Sink and CH4 and N2O Sources, But Magnitudes Differ Between Top-Down and Bottom-Up Methods, Global Biogeochemical Cycles, 38, e2023GB007969, 10.1029/2023GB007969, 2024.

Jonsson, M. and Wardle, D. A.: Structural equation modelling reveals plant-community drivers of carbon storage in boreal forest ecosystems, Biology Letters, 6, 116-119, 10.1098/rsbl.2009.0613, 2009.

Schuur, T.: Permafrost and the Global Carbon Cycle, in: Arctic Report Card 2019, edited by: Richter-Menge, J., Druckenmiller, M. L., and Jeffries, M., NOAA, 58-65, 2019.

Comment #10

L114-118: The carbon cycle core may be DALEC2 (Bloom & Williams, 2015) but I think you are missing this reference which includes the plant available and unavailable water concept (Yang et al., 2022).

Response #10

Done accordingly.

Methods:

Comment #11

The methods are insufficient to understand the analysis design and implementation. The analysis seems to neglect uncertainties in each dataset used in the calibration, validation and corroboration activities. This is also concerning because the uncertainties associated with these different datasets will vary substantially. The cost function (Normalised Nash-Sutcliffe Efficiency) used penalises observations away from the mean of the observations. This is not compatible with seasonal or long-term datasets as the seasonality or long-term trend in the chronosequence is informative. Relatedly, the construction of the chronosequences themselves are not adequately described to understand what assumptions have been made or their implications. Presumably biases

towards the mean will penalise extremes? How do you include temporally static priors such as LCMA in this cost function. Similarly, how is the leaf economic spectrum implemented? Can you justify neglecting the observations' uncertainties?

Response #11

Our response to **Reviewer #1 Comment #3** addresses observational uncertainty, and our response to **Reviewer #1 Comment #4** explains the construction of the biomass–age curves. Here, we clarify the cost function and how the static parameter references (Canopy efficiency, LCMA, and CUE) and leaf economics spectrum were included.

The observed mean in Nash–Sutcliffe efficiency (NSE) is used only to normalize the summed model–data mismatch and is not the calibration target. Each modeled value is compared with its corresponding observation. Squaring the residuals gives greater weight to larger model–data mismatches. If the model matches an observation, the error at that point is zero even when the observation is far from the mean. We clarified this distinction and the biomass terms used in each experiment in the Methods (**Line 349–357**):

"This function integrates model-data fit terms (J_{obs}), parameter priors (J_{prior}), and ecological constraints (J_{EDC}) into a single scalar metric. Normalized Nash–Sutcliffe efficiency (NNSE) transforms Nash–Sutcliffe efficiency (NSE) onto a bounded range of (0, 1] (Text S1). The observed mean occurs only in the NSE normalization denominator, while every modeled value is compared with its corresponding observation. Thus, seasonal cycles and the biomass-age trend remain explicit constraints. NNSE puts the fit scores for different variables on the same numerical range. All three experiments included NNSE terms for GPP, ET, RECO, and LAI. The baseline experiment included no biomass term. The single-year biomass-constrained experiment used a term based on the squared relative difference between modeled and observed mature biomass. The growth-curve-constrained experiment used NNSE to compare modeled biomass with the biomass–age curve."

Temporally static references such as LCMA were not evaluated using NNSE. We clarified their separate treatment in the Methods (**Line 369–372**):

"Static parameter references (Canopy efficiency, LCMA, and CUE) entered the objective J_{prior} as separate penalties based on squared differences between the logarithms of ce, LCMA, and CUE and the logarithms of their respective reference values. The parameters could depart from those values during calibration."

We also clarified how the leaf economics relationship entered the objective (**Line 389–390**):

"For the leaf economics spectrum constraint, we used the squared difference between the logarithms of the realized annual foliar turnover and the turnover predicted from LCMA."

To make the implementation explicit, Supplementary Text S1 now provides the complete objective and defines its components:

"During optimization, we minimized an objective function that included the NNSE terms with negative signs. Valid-data masks selected available observations only and were not inverse-variance weights. Defining R_{fast} as the sum of the NNSE values for GPP, ET, RECO, and LAI, the complete objective was:

$$J = -R_{fast} - R_B + P_{CE} + P_{LCMA} + P_{CUE} + P_{LES} + P_{preEDC} + P_{postEDC}$$

R_B was zero in the baseline experiment, $1 - \left[\left(\widehat{B} - B_{obs}\right)/B_{obs}\right]^2$ in the single-year biomass-constrained experiment, and the NNSE between the 100 modeled and observed annual biomass values in the growth-curve-constrained experiment. The static-reference penalties were $P_{CE} = 0.1\left[\ln(ce) - \ln\left(ce_{opt}\right)\right]^2$, $P_{LCMA} = [\ln(LCMA) - \ln(LCMA_{ref})]^2$, and $P_{CUE} = [\ln(CUE) - \ln(CUE_{ref})]^2$. These penalties were applied only where

valid reference values were available. P_{LES} was the squared log difference between realized annual foliar turnover and the annual turnover predicted from LCMA. The pre-integration term, P_{preEDC}, summed smooth negative-log-sigmoid penalties for violations of the parameter-based turnover and allocation inequalities. The post-integration term, $P_{postEDC}$, was evaluated from the model simulations and penalized the applicable carbon-pool input-output imbalances and pool-boundary violations. Both were non-negative soft penalties added to J rather than hard rejection criteria. The growth-curve-constrained experiment allowed vegetation and litter pools to accumulate over the 100-year simulation, whereas the baseline and single-year biomass-constrained experiments additionally applied the mature-state pool-balance penalties."

Text S1 further distinguishes the leaf-economics term from the other EDC penalties:

"Although referred to as EDC7 because it represents an ecological relationship, P_{LES} was evaluated separately from P_{preEDC} and $P_{postEDC}$ and added directly to the total objective. With LCMA in g C m^{-2}, the target leaf lifespan was $LL_{target} = \exp[1.0514 \ln(LCMA) - 1.7608]/12$ years, giving $\Phi_{target} = 1/LL_{target}$. The realized annual turnover was $\Phi_{model} = N_y^{-1} \sum_t \Phi_{fall,t}$, where $\Phi_{fall,t}$ is the daily leaf-fall factor and N_y is the number of simulated years. The unit-weighted penalty was $P_{LES} = \left[\ln(\Phi_{model} + 10^{-9}) - \ln(\Phi_{target} + 10^{-9})\right]^2$."

Comment #12

L128: Throughout use CEDA acronym. Should be CDEA

Response #12

Thank you for catching this. We have corrected “CEDA” to “CDEA” throughout the manuscript and Supplementary Information.

Comment #13

L136: There are several CARDAMOM-DALEC analyses conducted using the TRENDY intercomparison forcings and contributed to the Global Carbon Budget papers. These outputs are in the public domain and would match your analysis more closely. See references section below for information.

Response #13

We now use CARDAMOM v13 associated with Hugelius et al. (2024) in Section 3.3 and Fig. 6. The comparison methods and treatment of the published intervals are described in Text S8 and Tables S5 and S6. Please see details in our response to **Reviewer #1 Comment #5.**

Reference

Hugelius, G., Ramage, J., Burke, E., Chatterjee, A., Smallman, T. L., Aalto, T., Bastos, A., Biasi, C., Canadell, J. G., Chandra, N., Chevallier, F., Ciais, P., Chang, J., Feng, L., Jones, M. W., Kleinen, T., Kuhn, M., Lauerwald, R., Liu, J., López-Blanco, E., Luijkx, I. T., Marushchak, M. E., Natali, S. M., Niwa, Y., Olefeldt, D., Palmer, P. I., Patra, P. K., Peters, W., Potter, S., Poulter, B., Rogers, B. M., Riley, W. J., Saunois, M., Schuur, E. A. G., Thompson, R. L., Treat, C., Tsuruta, A., Turetsky, M. R., Virkkala, A. M., Voigt, C., Watts, J., Zhu, Q., and Zheng, B.: Permafrost Region Greenhouse Gas Budgets Suggest a Weak CO2 Sink and CH4 and N2O Sources, But Magnitudes Differ Between Top-Down and Bottom-Up Methods, Global Biogeochemical Cycles, 38, e2023GB007969, 10.1029/2023GB007969, 2024.

Comment #14

L145-155: GPP and Reco are from different datasets, have you checked whether their implied NEE is realistic / consistent with independent estimates? This is important as it seems you are neglecting the uncertainties. More questions / comments on uncertainty later.

L156: Have you evaluated the LAI dataset for appropriate LAI seasonality across the evergreen forest areas? Most EO LAI datasets show an unrealistically seasonal cycle (i.e. implying leaf mean residence time of < 1 year, in areas which should be 3-5 years) (Green et al., 2026).

Response #14

Thank you for these suggestions. We compared the NEE implied by the GPP and RECO products with FLUXCOM-X. The comparison showed agreement in seasonal variation but weaker agreement in annual variation and mean sink magnitude.

We also evaluated the seasonal variation of GLASS LAI in grid cells with an evergreen needleleaf forest (ENF) fraction of at least 0.5. The data showed low winter LAI and pronounced seasonal variation, which may affect the calibrated leaf lifespan.

We report these evaluations in Text S9 and Fig. S18 (reproduced below as **Fig. R7**):

"Text S9 Consistency of GPP and RECO products and seasonality of LAI products

To evaluate the carbon balance implied by the separately derived FluxSat GPP (Joiner et al., 2018) and Zeng et al. (2020) RECO products, we calculated NEE as RECO − GPP, with negative values indicating an ecosystem carbon sink, and compared it with FLUXCOM-X X-BASE v1.0 NEE (Nelson et al., 2024) over 2001–2019. Across the boreal study domain, the two estimates showed 96.0% agreement in sink or source sign, and their spatial patterns of 2001–2019 mean NEE were correlated (r = 0.57). Their area-weighted mean NEE values were −200.2 and −151.7 g C m^{-2} yr^{-1}, respectively. Agreement was closer in grid cells with an evergreen needleleaf forest (ENF) fraction $\geqslant$0.5, with mean values of −139.8 and −152.1 g C m^{-2} yr^{-1}, spatial r = 0.62, and 98.0% sign agreement. The monthly NEE time series for the study domain were strongly correlated (r = 0.97), whereas the annual correlation was weak (r = 0.21) and NEE calculated as Zeng RECO minus FluxSat GPP had greater interannual variability (14.2 versus 8.5 g C m^{-2} yr^{-1}; Fig. S18). This comparison shows agreement in seasonal variation but weaker agreement in annual variation and mean sink magnitude. FLUXCOM-X also uses FLUXNET data, so this comparison is not fully independent. Biases shared by the products may remain undetected even when their NEE values agree.

To assess the seasonality of GLASS LAI (Ma and Liang, 2022), we calculated the mean LAI for each calendar month over 2001–2019 in each 0.5° grid cell. We examined cells with an evergreen needleleaf forest (ENF) fraction of at least 0.5, based on mean cover over 1985–2020. For comparison, we also examined cells with a deciduous broadleaf forest (DBF) fraction of at least 0.5. These fractions describe the whole grid cell, whereas LAI was averaged over mature-forest pixels. In the ENF group, the spatial median LAI was 0.21 in January and 2.58 in July. The corresponding values in the DBF group were 0.19 and 4.32 (Fig. S18d)."

We revised the Discussion as follows (**Line 600–607**):

"Furthermore, the NEE implied by the separately derived FluxSat GPP (Joiner et al., 2018) and Zeng et al. (2020) RECO products agreed broadly with FLUXCOM-X (Nelson et al., 2024) in sink or source sign, spatial pattern, and seasonal cycle, but showed a stronger mean sink and weak agreement in interannual variability

(Text S9, Fig. S18). Our assessment of GLASS LAI (Ma and Liang, 2022) showed low winter values and strong seasonal variation in grid cells with an ENF fraction of at least 0.5 (Fig. S18d). This seasonality may influence the retrieved phenology and foliar-turnover parameters. Green et al. (2026) demonstrated at three northern evergreen needleleaf forest sites that assimilating spuriously seasonal satellite-derived LAI can bias inferred leaf traits and carbon-cycle dynamics, even when simulated carbon fluxes agree with observations."

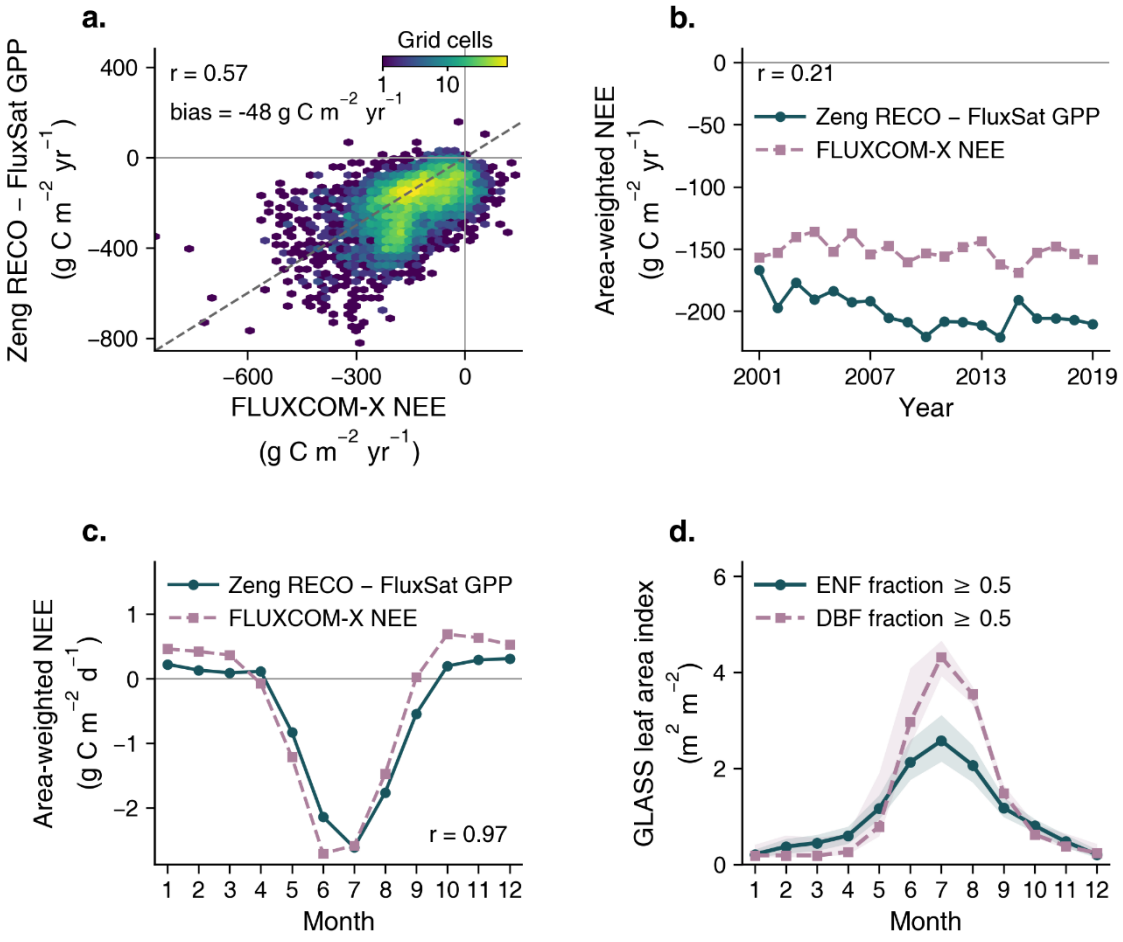


Fig. R7 (A copy of Fig. S18) Comparison of NEE derived from the GPP and RECO products with FLUXCOM-X, and seasonal variation of GLASS LAI. (a) Mean NEE over 2001–2019 calculated as Zeng RECO minus FluxSat GPP versus FLUXCOM-X X-BASE NEE for each 0.5° grid cell. The dashed line shows the 1:1 relationship. (b) Area-weighted annual means across the study domain. (c) Area-weighted mean seasonal cycles. Negative NEE indicates an ecosystem carbon sink. (d) Seasonal variation of GLASS LAI in grid cells with evergreen needleleaf forest (ENF) or deciduous broadleaf forest (DBF) fractions of at least 0.5. Monthly values are means over 2001–2019. Lines and shading show the spatial medians and interquartile ranges.

Reference

Green, T. J., Milodowski, D. T., Smallman, T. L., Mäkelä, A., and Williams, M.: Spurious seasonality of Earth observation LAI across three northern evergreen needleleaf forests: Implications for analyses of the carbon cycle, EGUsphere, 2026, 1-32, 10.5194/egusphere-2026-2222, 2026.

Joiner, J., Yoshida, Y., Zhang, Y., Duveiller, G., Jung, M., Lyapustin, A., Wang, Y., and Tucker, C. J.: Estimation of Terrestrial Global Gross Primary Production (GPP) with Satellite Data-Driven Models and Eddy Covariance Flux Data, 10.3390/rs10091346, 2018.

Ma, H. and Liang, S.: Development of the GLASS 250-m leaf area index product (version 6) from MODIS data

using the bidirectional LSTM deep learning model, Remote Sensing of Environment, 273, 112985, 10.1016/j.rse.2022.112985, 2022.

Nelson, J. A., Walther, S., Gans, F., Kraft, B., Weber, U., Novick, K., Buchmann, N., Migliavacca, M., Wohlfahrt, G., Šigut, L., Ibrom, A., Papale, D., Göckede, M., Duveiller, G., Knohl, A., Hörtnagl, L., Scott, R. L., Dušek, J., Zhang, W., Hamdi, Z. M., Reichstein, M., Aranda-Barranco, S., Ardö, J., Op de Beeck, M., Billesbach, D., Bowling, D., Bracho, R., Brümmer, C., Camps-Valls, G., Chen, S., Cleverly, J. R., Desai, A., Dong, G., El-Madany, T. S., Euskirchen, E. S., Feigenwinter, I., Galvagno, M., Gerosa, G. A., Gielen, B., Goded, I., Goslee, S., Gough, C. M., Heinesch, B., Ichii, K., Jackowicz-Korczynski, M. A., Klosterhalfen, A., Knox, S., Kobayashi, H., Kohonen, K. M., Korkiakoski, M., Mammarella, I., Gharun, M., Marzuoli, R., Matamala, R., Metzger, S., Montagnani, L., Nicolini, G., O'Halloran, T., Ourcival, J. M., Peichl, M., Pendall, E., Ruiz Reverter, B., Roland, M., Sabbatini, S., Sachs, T., Schmidt, M., Schwalm, C. R., Shekhar, A., Silberstein, R., Silveira, M. L., Spano, D., Tagesson, T., Tramontana, G., Trotta, C., Turco, F., Vesala, T., Vincke, C., Vitale, D., Vivoni, E. R., Wang, Y., Woodgate, W., Yepez, E. A., Zhang, J., Zona, D., and Jung, M.: X-BASE: the first terrestrial carbon and water flux products from an extended data-driven scaling framework, FLUXCOM-X, Biogeosciences, 21, 5079-5115, 10.5194/bg-21-5079-2024, 2024.

Zeng, J., Matsunaga, T., Tan, Z.-H., Saigusa, N., Shirai, T., Tang, Y., Peng, S., and Fukuda, Y.: Global terrestrial carbon fluxes of 1999–2019 estimated by upscaling eddy covariance data with a random forest, Scientific Data, 7, 313, 10.1038/s41597-020-00653-5, 2020.

Comment #15

L162-163: A 50 Mg/ha value seems a very low threshold (i.e., 25 MgC ha) for mature forests. Reference to justify? Can you also show examples of the chronosequences? Can you justify why you have not assimilated the actual EO biomass across the 20-year period?

L214-218: See earlier comment asking for justification over why chronosequence chosen over assimilating repeat biomass time series in the contemporary period.

Response #15

The 50 Mg ha^{-1} value serves as an absolute lower bound within the combined screening criterion described in Section 2.2.2, which also requires AGB to exceed the local 90th percentile within each 0.5° grid cell. The percentile criterion identifies the upper end of the local biomass distribution, while the absolute lower bound screens out low-biomass pixels that may nevertheless rank highly within a grid cell. A local reference is relevant because biomass in older boreal stands varies substantially with site conditions and disturbance history. For example, Wirth et al. (1999) reported AGB of only 60–100 Mg ha^{-1} in 200-year-old lichen-type Siberian Scots pine stands, compared with approximately 200 Mg ha^{-1} in Vaccinium-type stands.

We checked how this screening rule affected the availability of mature-forest pixels. Across the study grid cells, the median local 90th percentile was 95.8 Mg ha^{-1}, and 86.3% of cells had a local 90th percentile of at least 50 Mg ha^{-1}. Thus, in most cells, the local percentile already provides the more restrictive biomass criterion. Raising the absolute lower bound to 75 or 100 Mg ha^{-1} while retaining the percentile criterion would leave 19.4% or 37.4% of the study cells without eligible pixels, respectively. These results support retaining the lower bound of 50 Mg ha^{-1} alongside the local percentile criterion to include forests with lower biomass. Examples of the biomass–age curves and biomass simulated by DALEC are shown in **Fig. R3** in our response to **Reviewer #1 Comment #4**. That response also explains why we used the curves rather than directly assimilating the multi-year ESA CCI AGB maps. We added the following explanation in Supplementary Text S3:

Text S3 Selection of mature-forest pixels

"Mature-forest pixels were selected using two conditions: AGB above 50 Mg ha^{-1} and above the local 90th

percentile within each 0.5° grid cell. The local percentile allows the threshold to vary with biomass conditions among grid cells, while the absolute lower bound removes very low-biomass pixels. Older stands can have comparatively low biomass under some site conditions. For example, Wirth et al. (1999) reported 60–100 Mg ha^{-1} in 200-year-old lichen-type Siberian Scots pine stands, compared with approximately 200 Mg ha^{-1} in Vaccinium-type stands. The median local 90th percentile across our grid cells was 95.8 Mg ha^{-1}, and 86.3% of cells had a local 90th percentile of at least 50 Mg ha^{-1}. Raising the absolute lower bound to 75 or 100 Mg ha^{-1} would leave 19.4% or 37.4% of the study cells without eligible pixels, respectively. These criteria do not establish that the selected forests are old-growth or have never been disturbed."

Reference

Wirth, C., Schulze, E. D., Schulze, W., von Stünzner-Karbe, D., Ziegler, W., Miljukova, I. M., Sogatchev, A., Varlagin, A. B., Panvyorov, M., Grigoriev, S., Kusnetzova, W., Siry, M., Hardes, G., Zimmermann, R., and Vygodskaya, N. N.: Above-ground biomass and structure of pristine Siberian Scots pine forests as controlled by competition and fire, Oecologia, 121, 66-80, 10.1007/s004420050908, 1999.

Comment #16

L203-206: These independent data will carry their own, probably significant uncertainties. How are you accounting for these in your evaluation?

L240-242, L252-254: A 5-member ensemble is not a robust characterisation of uncertainty. This uncertainty is likely to be a gross underestimate as it has had no observational uncertainty propagated into it. Finally, this uncertainty is not presented anywhere in the paper that I can see.

Response #16

Thank you for raising these two issues. Observation-level uncertainty estimates were not available in the Xia et al. (2019) dataset used here, so the comparison was not uncertainty-weighted and observational error bars could not be shown. The uncertainty and site-to-grid representativeness limitations are documented in Text S4 and discussed in our response to **Reviewer #1 Comment #3.**

We also agree that five optimization runs do not provide a robust characterization of parameter uncertainty. Our original terminology might overstate their role. As clarified in the Methods, these runs were used to assess sensitivity to the initial parameter values. They were not samples from a probabilistic prior or posterior distribution. The variation among runs is reported in Fig. S4, but it does not quantify observational or posterior uncertainty.

The five optimization runs did not propagate observational uncertainty. The additional GPP and biomass sensitivity analyses and their limitations are presented in our response to **Reviewer #1 Comment #3.**

Reference

Xia, J., Yuan, W., Lienert, S., Joos, F., Ciais, P., Viovy, N., Wang, Y.-p., Wang, X., Zhang, H., Chen, Y., and Tian, X.: Global Patterns in Net Primary Production Allocation Regulated by Environmental Conditions and Forest Stand Age: A Model-Data Comparison, Journal of Geophysical Research: Biogeosciences, 124, 2039-2059, doi.org/10.1029/2018JG004777, 2019.

Comment #17

L268-282: Unclear why you need a pre-calibration on this if you are assimilating GPP and LAI information in the final contemporary period. Also, as the same datasets are being used are there consequences for the outputs?

Response #17

The pre-calibration follows the established procedure used in the ACM-based DifferLand configuration of Fang and Gentine (2024). It estimates canopy efficiency using relatively well-watered periods. The resulting ce_{opt} is used as a reference value during full calibration, while canopy efficiency can still vary. This reduces compensation between canopy efficiency and water-stress parameters. We agree that reusing FluxSat GPP and GLASS LAI does not provide independent information and may increase the influence of product-specific biases. We have clarified these points in the Methods:

Line 332–334:

"Following Fang and Gentine (2024), we first estimated a reference value for canopy efficiency (ce) during relatively well-watered periods. This pre-calibration aims to reduce compensation between canopy efficiency and water-stress parameters during the full calibration."

Line 343–346:

"The retrieved optimal value (ce_{opt}) was subsequently incorporated into the full model calibration as a reference value, while ce could still vary. Because the same FluxSat GPP and GLASS LAI products were used in both stages, the pre-calibration does not provide independent information and may increase the influence of product-specific biases."

Reference

Fang, J. and Gentine, P.: Exploring Optimal Complexity for Water Stress Representation in Terrestrial Carbon Models: A Hybrid-Machine Learning Model Approach, Journal of Advances in Modeling Earth Systems, 16, e2024MS004308, 10.1029/2024MS004308, 2024.

Comment #18

L306-308: As wood supplies carbon to soil, this soil constraint imposes an indirect constraint on the wood also.

L307-311: Can you provide more information on how within the EDCs you imposed the leaf economic constraint?

Response #18

We have clarified this coupling in the Methods. **Line 383–385**:

"Because inputs to the soil organic matter pool include wood litter and decomposed litter, EDC6 can nevertheless constrain wood turnover and allocation indirectly through the coupled carbon-pool dynamics."

We have also provided the explicit implementation of the leaf economics constraint in Text S1, as shown in our response to **Reviewer #1 Comment #11.** It is termed EDC7 because it represents an ecological relationship. It was evaluated separately from the pre- and post-integration EDC penalties and added directly to the total objective. Its contribution was a unit-weighted penalty based on the squared difference between the logarithms of realized annual foliar turnover and the turnover predicted from LCMA. It was therefore a soft relationship constraint rather than a hard constraint or an NNSE term.

Comment #19

L320: Does the 5-subsets refer to the 5-ensemble member or a resampling of the medians in space?

Response #19

The five subsets contain grid cells, not optimization runs. We clarified this in the Methods (**Line 403–404**):

"For each parameter, we first calculated a median across the five optimization runs within each grid cell. We then randomly divided these grid-cell median values into five spatial subsets."

Results:

Comment #20

How well do you fit to your assimilated time series and prior information? These statistics would provide valuable information on the success of the calibration.

Do your different calibrations imply differences in the net carbon balance over the contemporary period? i.e., does the information propagate away from wood alone to other pools in the C cycle?

Response #20

1 Fit to assimilated time series and prior information

The median calibration NNSE for GPP, RECO, LAI, and ET ranged from 0.806 to 0.959. The growth-curve-constrained experiment fitted the biomass–age curves with a median NNSE of 0.997. The single-year biomass-constrained experiment had a median absolute relative error of 0.009% against its biomass constraint. For the static parameter references, median optimized-to-reference ratios across the three experiments ranged from 0.923 to 0.956 for canopy efficiency, 0.837 to 0.844 for LCMA, and 1.019 to 1.025 for CUE. Table S7 reports these statistics over the grid cells retained for comparison. We first took the median across the five optimization runs within each grid cell. These statistics describe calibration fit and agreement with parameter reference values, not independent validation.

We added the following description to Text S7, with the results for each experiment reported in Table S7:

"Text S7 Calibration fit and agreement with parameter reference values

Median calibration NNSE values for GPP, RECO, LAI, and ET ranged from 0.806 to 0.959 (Table S7). The growth-curve-constrained experiment reproduced the assimilated biomass–age curves with a median NNSE of 0.997, while the single-year biomass-constrained experiment had a median absolute relative error of 0.009% against its assimilated mature-biomass constraint. Median optimized-to-reference ratios across the experiments ranged from 0.923 to 0.956 for canopy efficiency, from 0.837 to 0.844 for LCMA, and from 1.019 to 1.025 for CUE. These statistics quantify calibration fit or agreement with parameter reference values rather than independent validation."

2 Net carbon balance and pools beyond wood

Yes. Compared with the baseline experiment, the growth-curve-constrained experiment had lower mean NEP over both 2001–2012 and 2013–2019. It also had higher soil organic carbon and lower litter carbon stocks at the end of each period. The revised Results report these differences (**Line 472–479**):

"The three experiments differed in net carbon balance and pools beyond wood over 2001–2019 (Fig. S9). Relative to the baseline, the growth-curve-constrained experiment reduced mature-forest-area-weighted NEP by 17.56 g C m^{-2} yr^{-1} in 2001–2012 and 11.16 g C m^{-2} yr^{-1} in 2013–2019, mainly because of higher RECO, while NPP differed by only −0.15 and +1.21 g C m^{-2} yr^{-1}. Live carbon at the ends of 2012 and 2019 was higher by 50.54 and 49.64 Mg C ha^{-1}, litter was lower by 1.24 and 1.36 Mg C ha^{-1}, and SOM was higher by 95.43

and 95.67 Mg C ha^{-1}, respectively. The single-year biomass-constrained experiment primarily increased live carbon relative to the baseline (+37.39 and +40.03 Mg C ha^{-1}), with much smaller differences in NEP, litter, and SOM. These results show that the effects of the different calibrations extended beyond wood to ecosystem carbon uptake and the litter and soil organic matter pools."

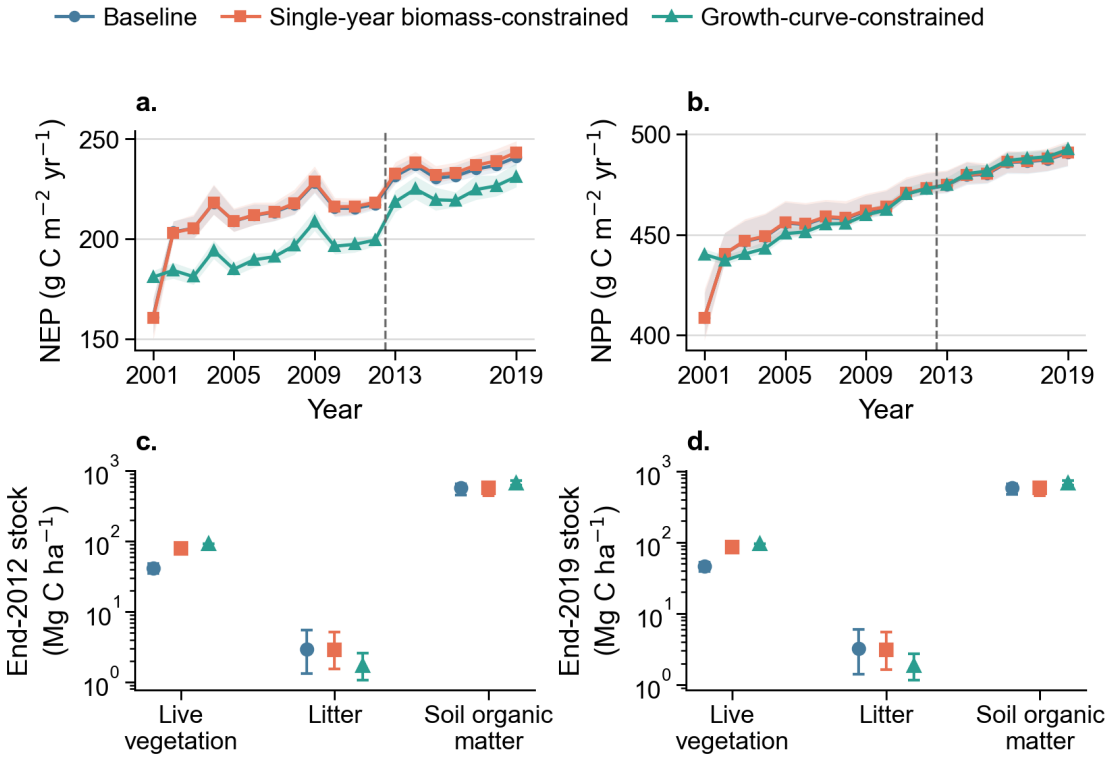


Fig. R8 (A copy of Fig. S9) Carbon balance and carbon pools across the three experiments over the 2001–2012 calibration period and the 2013–2019 temporal evaluation. (a) Annual net ecosystem production (NEP = GPP − RECO). (b) Annual net primary production (NPP). The vertical dashed line marks the boundary between calibration and temporal evaluation periods. (c, d) Live vegetation, litter, and soil organic matter stocks at the ends of 2012 and 2019, respectively. Lines and points are mature-forest-area-weighted means. Shading and whiskers show the complete range across the five optimization runs. This range shows sensitivity to optimization initialization and is not an observational or posterior uncertainty interval.

Table R3 (A copy of Table S7) Calibration fit to assimilated observations and agreement with static parameter reference values. Values are spatial medians [interquartile ranges] after taking the median across the five optimization runs within each grid cell. NNSE values closer to 1 indicate closer fits. The single-year biomass entry is the absolute relative error at the end of the calibration period. Smaller values indicate closer fits. Entries for the static parameter references are optimized/reference ratios. Values closer to 1 indicate closer agreement.

Metric	Baseline	Single-year biomass	Growth curve
GPP calibration NNSE	0.925 [0.909-0.938]	0.925 [0.909-0.938]	0.914 [0.890-0.932]
ET calibration NNSE	0.810 [0.753-0.843]	0.811 [0.754-0.843]	0.806 [0.746-0.843]
RECO calibration NNSE	0.959 [0.948-0.967]	0.959 [0.948-0.967]	0.955 [0.943-0.965]
LAI calibration NNSE	0.900 [0.841-0.933]	0.900 [0.840-0.933]	0.863 [0.772-0.918]
Single-year biomass relative error (%)	-	0.009 [0.005-0.014]	-

Metric	Baseline	Single-year biomass	Growth curve
Biomass calibration NNSE	-	-	0.997 [0.984-0.998]
Canopy efficiency optimized/reference ratio	0.923 [0.861-1.002]	0.924 [0.861-1.002]	0.956 [0.871-1.051]
LCMA optimized/reference ratio	0.837 [0.809-0.892]	0.837 [0.809-0.892]	0.844 [0.809-0.909]
CUE optimized/reference ratio	1.019 [1.010-1.031]	1.019 [1.010-1.031]	1.025 [1.011-1.049]

Comment #21

You refer to changes in e.g. NPP allocation or residence times, or their statistical fits but I cannot see the actual values in the text. It would help to quantify these changes. This information should then be included in the discussion with comparisons to the existing literature.

At various points you refer to NPP as fractions (i.e. 0-1) or percentages (0-100). This is confusing. In figure 2 you refer to NPP fractions but present numbers in 0-100.

Foliage, wood and root residence times are not consistently presented. Sometimes as turnover fraction per year, sometimes as mean residence times in years. Be consistent across all to improve clarity.

Response #21

We added absolute parameter values to the Results and a quantitative comparison with the existing literature to the Discussion.

In the Results (**Line 442–446**):

"Across the grid cells analyzed in each experiment, mean NPP allocation to wood was 0.683, 0.683, and 0.547 in the baseline, single-year biomass-constrained, and growth-curve-constrained experiments, respectively. The corresponding spatial mean wood residence times were 14.2, 34.2, and 42.8 years. In the growth-curve-constrained experiment, mean NPP allocation to leaf and fine roots was 0.234 and 0.219, respectively, and their spatial mean residence times were 0.85 and 0.31 years."

In the Discussion (**Line 618–621**):

"The increase in spatial mean wood residence time from 14.2 to 42.8 years was accompanied by a decrease in mean NPP allocation to wood from 0.683 to 0.547. Compared with the field-based allocation estimates compiled by Xia et al. (2019), the growth-curve-constrained wood allocation estimates had a mean positive bias of 0.0586 and a Pearson correlation of 0.57 (Fig. S8h)."

We standardized NPP allocation throughout the text, figures, and supplementary material as dimensionless fractions ranging from 0 to 1. Differences and absolute error metrics for allocation are now expressed on the same fractional scale, including the difference maps in Fig. 2. The corresponding numerical values, labels, captions, and allocation equations were revised consistently. We also standardized the presentation of leaf, wood, and fine-root turnover throughout the text and figures using mean carbon residence times in years. Wood and fine-root residence times were calculated as the inverse of their annual turnover rates. The associated figure labels, captions, and quantitative descriptions were revised accordingly. Relative-bias statistics such as PBIAS and proportions of grid cells or years remain expressed as percentages.

Reference

Xia, J., Yuan, W., Lienert, S., Joos, F., Ciais, P., Viovy, N., Wang, Y.-p., Wang, X., Zhang, H., Chen, Y., and Tian, X.: Global Patterns in Net Primary Production Allocation Regulated by Environmental Conditions and

Forest Stand Age: A Model-Data Comparison, Journal of Geophysical Research: Biogeosciences, 124, 2039-2059, doi.org/10.1029/2018JG004777, 2019.

Comment #22

L331-333: Are these two sentences compatible? They do not appear to be.

Response #22

We have clarified the distinction between the general tendency and regional differences:

"The baseline experiment generally underestimated mature forest biomass, particularly in high-latitude regions, although overestimation occurred in parts of the southern boreal ecotones (Fig. 2a)."

Comment #23

L333-338: Somewhere can you show the actual EO data, example chronosequences and / or domain averaged chronosequence?

Response #23

We show examples of the biomass–age curves used in calibration and the biomass simulated by DALEC at each stand age under all three experiments for three grid cells in **Fig. R3** in our response to **Reviewer #1 Comment #4.**

Comment #24

L342: A longer mean residence time?

Response #24

Yes. We now report this result directly as a longer mean wood residence time.

Comment #25

L349: This is not “Notable”, it is the expected result. That observations on fast fluxes do not inform slow pools is an underpinning hypothesis noted in your introduction and evidenced elsewhere in existing literature. The lack of coupling has already been reported numerous times in the literature using DALEC (e.g., Fox et al., 2009, Williams et al., 2009, Smallman et al., 2017, Smallman et al., 2021).

Response #25

We removed “Notably” and revised the wording to avoid presenting it as a new finding.

Comment #26

L362-369: Quantify the reported numbers and their variability not just statistical differences. Furthermore, elsewhere you report R2 but here you report R (R2 equivalent 0.33). Be consistent to improve clarity.

Response #26

We added numerical estimates and their spatial variability to the field-allocation comparison. The revised Results now state (**Line 464–467**):

"Across the 34 comparison grid cells, field-derived wood allocation averaged 0.435 ± 0.142, compared with 0.687 ± 0.124, 0.692 ± 0.118, and 0.494 ± 0.203 for the baseline, single-year biomass-constrained, and growth-curve-constrained experiments, respectively (mean ± standard deviation across grid cells)."

We retained Pearson's correlation coefficient, r, for the field-allocation comparison because it describes both the strength and direction of the association between calibrated and field-derived allocation estimates. Squaring this coefficient would remove the sign and change the information conveyed. We therefore retain r in Fig. S8 and the corresponding text, together with mean bias, while retaining R^2 for the spatial biomass comparison in Fig. 2.

Comment #27

Fig 2 (and most others). You need to show the actual estimates / evaluation data somewhere too.

Response #27

The absolute growth-curve-constrained estimates of wood mean carbon residence time and the fraction of NPP allocated to wood are shown in Fig. 3a and Fig. 4a, respectively, while Fig. 2 highlights differences among the calibration experiments. We have also added a fourth column to Fig. S7 showing the absolute growth-curve-constrained estimates of the fraction of NPP allocated to leaves, leaf lifespan, the fraction of NPP allocated to fine roots, and fine-root mean carbon residence time. These panels show the parameter values alongside the difference maps.

Comment #28

L388: Leaf life span, this has a prior provided with it. How well was it fitted? Do you return the prior or have added additional information from the analysis?

Response #28

The LCMA–leaf-lifespan relationship was implemented as a soft ecological constraint rather than fixing leaf lifespan to a prescribed value. Both LCMA and the phenology parameters controlling realized leaf lifespan were optimized during calibration. The revised Methods section states (**Line 389–390**):

"For the leaf economics spectrum constraint, we used the squared difference between the logarithms of the realized annual foliar turnover and the turnover predicted from LCMA."

As suggested, we added a quantitative assessment of this constraint to Text S7:

"The leaf economics relationship was evaluated as a soft constraint rather than as a fixed value. In the growth-curve-constrained experiment, the mean realized leaf lifespan was 0.85 years, compared with 0.99 years predicted from the optimized LCMA. In paired grid-cell comparisons, realized lifespans were typically 15% shorter than the lifespans predicted from LCMA (interquartile range: 9–18% shorter). Thus, the optimization remained influenced by the trait relationship but did not simply reproduce it."

Comment #29

L389: Wood and root dynamics. Not clear exactly what you mean here. The turnover or allocation parameters or both?

Response #29

Thank you for pointing out this ambiguity. We refer to both allocation fractions and mean carbon residence times for wood and fine roots and have revised the wording to state this explicitly.

Comment #30

L406: "The fraction of DBF…" is this not just the inverse of the ENF fraction? You are simulating forest areas only, therefore should ENF and DBF not sum to 1?

Response #30

ENF and DBF are not complementary in our analysis. These predictors describe grid-cell land-cover composition from the source PFT maps, rather than the relative proportions of forest types within the simulated forest area. The source maps also include DNF, other tree PFTs, and non-forest cover, so restricting the calibration to forests does not imply that ENF and DBF sum to one. We clarified this definition in the Methods (**Line 246–248**):

"For each grid cell, we averaged the area fractions of boreal needleleaf evergreen (ENF), boreal broadleaf summergreen (DBF), and boreal needleleaf deciduous (DNF) vegetation over 1985–2020. Each fraction was calculated relative to the entire grid-cell area."

Comment #31

L460-484: The authors corroborate / benchmark their analysis against a global reanalysis produced with CARDAMOM (Bloom et al., 2016). This is not an informative benchmark for several reasons outlined towards the start of this document.

Response #31

We replaced the main Bloom et al. (2016) comparison with CARDAMOM v13. Section 3.3 and Tables S5 and S6 report the allocation fractions, leaf lifespan, and residence times from both studies, their differences, and comparisons with the CARDAMOM v13 uncertainty intervals. The limitations are explained in our response to **Reviewer #1 Comment #5.**

Reference

Bloom, A. A., Exbrayat, J.-F., van der Velde, I. R., Feng, L., and Williams, M.: The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times, Proceedings of the National Academy of Sciences, 113, 1285-1290, 10.1073/pnas.1515160113, 2016.

Discussion and Supplementary Materials:

Comment #32

L512-513: Again, not "notable", it is expected.

Response #32

We removed "Notably" and clarified that this result represents calibration fit rather than independent validation. The revised Discussion states (**Line 613–617**):

"Assimilating the biomass–age curves substantially improved the fit to these curves, with the RMSE decreasing from 138.7 to 11.3 Mg ha^{-1}. The latter value represents calibration fit rather than independent validation. Together with the similar performance against the temporally withheld high-frequency observations, this result illustrates the value of using information across timescales to constrain long-term carbon storage."

Comment #33

L525: Uncertainties have not been reported.

Response #33

We acknowledge that the submitted analysis did not quantify observational or posterior uncertainty. We tested lower and upper GPP based on the reported standard errors, biomass–age curves fitted to the lower and upper biomass-error limits, and half or twice the original weight of the biomass term. The methods and results are reported in Text S6, Fig. S10, and Fig. S11, as explained in our response to **Reviewer #1 Comment #3**. These tests do not constitute full uncertainty propagation.

Comment #34

How can the Normalised Nash-Sutcliffe Efficiency be compatible with the growth curve assimilation? This surely bias' towards the mean and penalises extremes? This issue will be consistent for all the assimilated data.

Response #34

As explained in our response to **Reviewer #1 Comment #11**, NNSE compares each modeled value with its corresponding observation along the biomass–age curve. The observed mean is used only in the NSE normalization denominator, so observations are not penalized simply because they are far from the mean.

Comment #35

Fig S2. I think an example plot(s) showing an actual regrown curve with the uncertainties in the ESA CCI dataset, which are considerable should be given.

Response #35

Fig. S2 illustrates the growth function schematically. Examples of the ESA CCI-based biomass–age curves provided by Xu et al. and the corresponding biomass simulated by DALEC are now shown in Fig. S5, reproduced as **Fig. R3** in our response to **Reviewer #1 Comment #4**.

To assess sensitivity to biomass observational uncertainty, we calibrated using biomass–age curves fitted separately to the lower and upper biomass-error limits. The methods and results are described in our response to **Reviewer #1 Comment #3** (Text S6 and Fig. S11).

Comment #36

Fig S3. This plot proves my earlier point about the very large uncertainties inherent in ecological data and analyses. The uncertainty needs to be accounted for.

Response #36

Fig. S3 shows substantial residual variation around the fitted leaf-trait relationship. This relationship entered the calibration as a soft ecological penalty rather than an uncertainty-weighted likelihood. We clarified the Fig. S3 caption as follows:

"This relationship was implemented as a unit-weighted soft ecological constraint rather than an uncertainty-weighted likelihood. The residual scatter ($RMSE_{ln}$ = 0.70) was not propagated as an observation-error model."

Agreement with this soft constraint is quantified in Text S7, as described in our response to **Reviewer #1 Comment #28.** This comparison assesses agreement with the leaf lifespan predicted from LCMA but does not account for the residual uncertainty in the trait relationship, which remains a limitation of the calibration.

Comment #37

Fig 5S. why is there no equivalent plot for the slow datasets and or the prior information? Could this information be presented in the text if not as a map?

Response #37

We have added quantitative summaries of the fit to the slow constraints and agreement with the parameter reference values in Table S7 and Text S7. Table S7 reports NNSE for the fit to the biomass–age curves for the growth-curve-constrained experiment, absolute relative error for the single-year biomass constraint, and optimized-to-reference ratios for canopy efficiency, LCMA, and CUE. These statistics and the corresponding Text S7 passage are reproduced in our response to **Reviewer #1 Comment #20.**

Agreement with the leaf-economics soft constraint is quantified separately in Text S7, with the relevant passage quoted in our response to **Reviewer #1 Comment #28.** Examples of biomass–age constraints and simulated biomass from all three experiments are also shown in Fig. S5, reproduced as **Fig. R3** in our response to **Reviewer #1 Comment #4.** These additions provide the requested numerical and graphical descriptions of calibration fit and agreement with parameter reference values.

Comment #38

Fig S6. Add column with the actual values in at least one of the columns. E.g. the growth-curve constrained.

Response #38

Fig. S7 (formerly Fig. S6) now includes a fourth column showing the absolute growth-curve-constrained values, while retaining the three original difference columns.

Comment #39

Fig S7. This evaluation is of little value unless you consider the uncertainties in both the analysis and the observations – assuming they are available?

Response #39

For Fig. S8 (formerly Fig. S7), observation-level uncertainties were not available in Xia et al. (2019), so we could not add observational error bars. Our five optimization runs also do not provide posterior intervals. We retained the figure as an external comparison and state that it cannot establish statistical significance after accounting for uncertainty in both datasets. Its caption now reads:

"Observation-level uncertainties were not reported in Xia et al. (2019), so observational error bars are unavailable. The five optimization runs do not provide posterior samples or intervals."

Reference

Xia, J., Yuan, W., Lienert, S., Joos, F., Ciais, P., Viovy, N., Wang, Y.-p., Wang, X., Zhang, H., Chen, Y., and Tian, X.: Global Patterns in Net Primary Production Allocation Regulated by Environmental Conditions and Forest Stand Age: A Model-Data Comparison, Journal of Geophysical Research: Biogeosciences, 124, 2039-2059, doi.org/10.1029/2018JG004777, 2019.

Comment #40

Fig S8. Why is residence time not consistently presented? Present all in years, consistent with the wider literature.

Response #40

In Fig. S12 (formerly Fig. S8) we now present wood and fine-root mean carbon residence times in years. The calibrated residence times were obtained as the reciprocals of annual turnover rates, and the gradient boosting models were trained directly on these residence-time targets. The plotted R^2 and RMSE were recalculated from the median predictions for the pooled held-out grid cells.
