Reviewer #3: Natalie Douglas

General Comments:

Comment #1

The authors have presented the improvements in calibration of the DALEC terrestrial carbon cycle model (through DifferLand) when assimilating long term biomass growth trajectories over fast flux observations alone or with a single year of biomass stock.

While the manuscript is well written, there are areas in the manuscript that require significant improvement before publication.

Response #1

Thank you for your constructive comments and positive assessment of our manuscript. We have revised the manuscript in response to your suggestions, as detailed in our point-by-point responses below.

Comment #2

The novelty of the presented study needs to be made clearer in the manuscript with emphasis on what is the added value to what to exists in the current literature. Ostensibly, the novelty is in assimilating a recent dataset of biomass growth curves for model calibration. The manuscript makes reference to the baseline experiment following the original DifferLand configuration that is constrained by high-frequency data. What is different about what is presented here for this experiment? There is also a comparison against the CARDAMOM product in the present study. What does this offer beyond what is offered by the CARDAMOM product?

Response #2

Thank you. We expanded the Introduction to acknowledge earlier studies using repeated biomass observations and chronosequences. The revised text is reproduced in our response to **Reviewer #1 Comment #2** and states (**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."

The baseline follows the existing DifferLand configuration. It provides a reference for comparison with the two biomass-constrained experiments. We clarified this role in the Methods (**Line 272–273**):

"The baseline experiment follows the existing DifferLand configuration and provides a reference for comparison with the two biomass-constrained experiments."

Following **Reviewer #1 Comment #5,** we replaced the Bloom et al. (2016) comparison with CARDAMOM v13. We used this comparison to assess how the magnitudes and spatial patterns of our allocation fractions and residence times differ from those obtained with CARDAMOM v13. CARDAMOM v13 assimilated woody-biomass observations for calendar years 2017 and 2018, whereas our biomass curves describe variation with stand age. Details of the revised comparison are provided in Section 3.3 and our response to **Reviewer #1 Comment #5**.

Comment #3

The second major area for improvement is in the clarity of the methodology. Despite having read the document several times, I found it difficult to comprehend the setup of the experiments. The specific comments give detail but I highlight the main sources of confusion as follows. The cost function used in the calibration is not given (despite normalisation of one of the terms given in the supplementary material) without which it is impossible to identify how uncertainty information is implemented, if at all, given the lack of detail on this matter. It is difficult to decipher how the initial conditions from the biomass growth curve experiment are calculated and used in the other experiments and for when. The manuscript would benefit substantially from having some time series figures for a specific site to show observational data against prior and post-calibration trajectories. Finally, it is unclear how the central hypothesis of the paper is validated. From what I can tell, independent validation is performed for the fast-process observations through a temporal hold-out period, and the retrieved allocation parameters are compared against independent field measurements. However, the biomass-growth trajectories themselves appear to be used primarily as calibration constraints.

Response #3

Thank you. We have revised the Methods, Fig. 1, and the Supplementary Information to clarify the cost function, initial conditions, biomass time-series comparisons, and the distinction between calibration fit and independent evaluation.

1. Calibration objective and observational uncertainty

Supplementary Text S1 now provides the complete composite objective:

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

Here, R_{fast} contains the NNSE terms for GPP, ET, RECO, and LAI. R_B is the biomass term, which differs among calibration experiments. The remaining terms represent the direct parameter references, leaf-economics relationship, and ecological and dynamical constraints. The complete definitions are reproduced in our response to **Reviewer #1 Comment #11.** The revised Methods explicitly distinguish numerical normalization from uncertainty weighting (**Line 354–358**):

"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. NNSE scales the fit terms for different datasets but does not weight observations by their uncertainties."

The objective did not use product-specific observational uncertainties as weights. Table S3 documents the available uncertainty information and how it was used. Text S6, Fig. S10, and Fig. S11 report full-domain sensitivity tests using higher and lower GPP based on the reported FluxSat standard errors, biomass–age curves refitted at the reported observational-error limits, and biomass-term weights of 0.5 and 2. The methods and results are detailed in our response to **Reviewer #1 Comment #3**.

2. Initial conditions and experimental stages

Each calibration experiment jointly optimized its own model parameters and initial pools. The Methods define the starting state of the growth-curve-constrained experiment and refer to Text S5 for details (**Line 269–270**):

"The optimized initial pools in the growth-curve-constrained experiment correspond to the beginning of the 100-year simulation rather than to a specific calendar year (Text S5)."

For the baseline and single-year biomass-constrained experiments, the revised Methods section states (**Line 273–276**):

"During calibration, these two experiments independently optimized their own initial pools together with the model parameters. For both experiments, the optimized initial pools defined the model state at the beginning of the 2001–2012 calibration simulation and were not taken from the growth-curve-constrained experiment."

The growth-curve calibration covered stand ages 1–100. The 2001–2012 observations of GPP, ET, RECO, and LAI were aligned with simulation years 89–100. The forcing sequence and alignment are explained in our response to **Reviewer #2 Comment #23.**

The growth-curve initial pools were shared only in the subsequent 100-year forward comparison. We clarified both the purpose and limitation of that comparison (**Line 307–314**):

"For each grid cell and optimization run, we performed 100-year forward simulations using all three calibrated parameter sets. Each simulation used the initial pools from the corresponding growth-curve-constrained experiment. These initial pools define the starting state of the biomass–age simulation, rather than a state in a specific calendar year. This common initialization standardized the starting pools across the forward simulations. With the same forcing, this comparison shows how the three calibrated parameter sets produce different biomass accumulation from the same starting state. However, those initial pools were jointly optimized with the growth-curve parameter set, not with the baseline or single-year parameter sets. The comparison therefore remains dependent on the chosen initial pools and their compatibility with each parameter set."

The revised Fig. 1 distinguishes calibration, temporal evaluation, and the 100-year forward simulations using the same starting pools. It identifies the constraints and time periods for each experiment.

3. Comparison of biomass time series from the three experiments

Fig. S5 compares the original biomass–age constraint with DALEC biomass time series from all three experiments at three 0.5° grid cells. The baseline and single-year biomass-constrained experiments show biomass accumulation without assimilating the full biomass–age curve. We did not plot random optimization starting values as pre-calibration predictions because they were numerical starting points, not prior ecological estimates.

In Fig. S5, all three parameter sets used the corresponding initial pools from the growth-curve-constrained experiment as the same starting state. The figure shows the median simulated biomass at each stand age across five optimization runs for each experiment and labels the locations of the three cells. The figure and caption are reproduced as **Fig. R3** in our response to **Reviewer #1 Comment #4.** Agreement with the parameter reference values is reported separately in Table S7 and Text S7.

4. Interpretation of the evaluation

The reviewer is correct that the biomass–age curves were used as calibration constraints. Their close reproduction therefore cannot independently validate the benefit of assimilating those same curves. We revised the Results accordingly (**Line 430–432**):

"As expected, the growth-curve-constrained simulation closely reproduced the age-100 biomass values from the biomass–age curves used in calibration (Fig. 2c, $R^2 = 0.99$, RMSE = 11.3 Mg ha^{-1}). These statistics measure calibration fit, rather than independent validation."

The reported R^2 and RMSE describe agreement across grid cells at stand age 100, rather than fit to the full biomass–age curve.

The 2013–2019 evaluation continued from each experiment's calibrated end state without refitting to the withheld observations, as explained in our response to **Reviewer #2 Comment #31.** It tests performance for GPP, RECO, ET, and LAI. Field-derived allocation estimates provide an external comparison of the retrieved allocation fractions. However, uncertainties for individual observations were unavailable, and sampled plots may not represent the full grid cell, as discussed in our response to **Reviewer #1 Comment #3.** Neither comparison independently validates biomass accumulation over 100 years.

Specific Comments and Technical Corrections:

Comment #4

Abstract: there should be some mention of the DALEC model (in relation to DifferLand) somewhere in the abstract

Response #4

Thank you for this suggestion. We now identify DALEC in relation to DifferLand. The revised text reads:

"Here, we employ DifferLand, a JAX-based differentiable framework implementing the DALEC model, to jointly assimilate high-frequency observations with either a single year of biomass stock data or satellite-derived biomass–age curves for boreal forests."

Comment #5

Ln 23-24: should also mention the single-year biomass stock as Ln 28-29 are a little confusing without previously mentioning it.

Response #5

We introduced the single-year biomass stock experiment before presenting its results, as shown in our response to **Reviewer #3 Comment #4**.

Comment #6

Ln 27: change 'reduce' to 'reduces'

Response #6

We changed "reduce" to "reduces".

Comment #7

Ln 28: ' a single year OF biomass stock'

Response #7

We added "of" in "a single year of biomass stock".

Comment #8

Ln 43-49: unclear which is the primary source of uncertainty: model process representation or parameterisation within models

Response #8

We clarified that both model structure and parameterization contribute to uncertainty. Their relative importance depends on the model, available data, and prediction considered. The revised text is provided in our response to **Reviewer #1 Comment #6.**

Comment #9

Ln 55-56: does this mean that DALEC is not PFT-based?

Response #9

DALEC itself is not PFT-based. The sentence has also been removed in accordance with **Reviewer #2, Comment #7.**

Comment #10

Ln 57-58: this opening sentence is a little out of place and paraphrased in Ln 76-77. It can be removed.

Response #10

We removed the opening sentence, as noted in our response to **Reviewer #2 Comment #7.**

Comment #11

Ln 71: I don't think there is a need to mention VOD here.

Response #11

We removed the reference to VOD.

Comment #12

Ln 76-85: I initially thought that a new paragraph should start here with 'A key challenge…' but there is some repetition here and some redundant sentences could be removed. Also, 'reducing equifinality' should be 'mitigating equifinality'.

Response #12

We removed the repeated definition and the claim about reducing equifinality. The revised text describes the information provided by a single biomass stock and by biomass–age curves, as shown in our response to **Reviewer #2 Comment #10**.

Comment #13

Ln 88: 'these regions'?

Response #13

We replaced “these regions” with explicit references to forest biomes and boreal forests.

Comment #14

Ln 91: I don’t think a paragraph break is needed here. Perhaps remove ‘To achieve this’ as it seems as though you’re saying ‘To achieve focusing on boreal forests’.

Response #14

We removed “To achieve this” and joined the DifferLand description to the preceding paragraph.

Comment #15

Ln 92: It is unclear here and throughout why calculation of gradients is required and why the ability to do so is beneficial.

Response #15

We clarified that the Adam optimizer uses gradients to update model parameters and initial conditions. The gradients describe how the calibration objective changes when these optimized quantities change. The Introduction revision is provided in our response to **Reviewer #2 Comment #11.**

Comment #16

Ln 101: I think the title needs work, since the bulk of this section discusses the DALEC model. Additionally, there is a section called ‘Optimisation and Cost Function’ which is part of the calibration framework

Response #16

We renamed Section 2.1 “DifferLand Framework and DALEC Model”.

Comment #17

Ln 103: and multi-scale?

Response #17

We added “multi-source and multi-scale”.

Comment #18

Ln 107: ‘along with data assimilation’. Isn’t data assimilation what you’re using TO calibrate the parameters?

Response #18

We removed “along with data assimilation”.

Comment #19

Ln 110: ‘to be back propogated’. I think this is worth a little discussion to draw out the use of gradients and why DifferLand has advantages here.

Response #19

We clarified the role of backpropagation. Biomass at a given time depends on earlier simulated fluxes and pool sizes. Automatic differentiation computes how the objective changes with model parameters and initial conditions through those earlier calculations. Adam then uses these gradients to update both. The revised Methods passage is provided in our response to **Reviewer #2 Comment #12.**

Comment #20

Ln 128: Here, and later, ‘CEDA’ should be ‘CDEA’.

Response #20

We corrected “CEDA” to “CDEA” throughout.

Comment #21

Ln 154: Define RECO earlier – Ln 66?

Response #21

We defined ecosystem respiration (RECO) at its first occurrence in the Introduction.

Comment #22

Ln 159: ‘All observation datasets were aggregated to …’, what were they originally. This information could be added to Table S3.

Response #22

Table S3 now lists the native spatial and temporal resolutions of the products before aggregation to the 0.5° model grid.

Comment #23

Ln 160: Is 2020 the only year available? Is it lacking any temporal variability?

Ln 172-173: Which one? ‘Reference for comparison’ or ‘static constraint’?

Response #23

ESA CCI Biomass v6 provides ten AGB maps for 2007, 2010, and 2015–2022. We used the 2020 map to select mature-forest pixels and as the biomass constraint in the single-year experiment. The ESA CCI Biomass v6 Product User Guide warns that differences in satellite observations among years can bias apparent AGB changes, particularly between 2007 or 2010 and 2015–2022. It recommends using change uncertainty and quality flags and validating changes before further analysis. We therefore did not interpret differences between annual maps as direct measurements of age-related forest growth. The reasons for using biomass–age curves

are explained in our response to **Reviewer #1 Comment #4.**

The selected 2020 map provides a static constraint in the single-year biomass-constrained experiment. We removed "reference for comparison". The revised Methods section states (**Line 209–210**):

"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."

Reference

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.

Comment #24

Ln 174-181: It is a little misleading to say 'satellite-derived curves'. From the previous paragraph, it seems that multiple processes (and fitting) have been applied with multiple sources of uncertainty. How do you deal with the uncertainty in your observational sets?

Response #24

In "Forest Biomass Constraints for Slow Processes", we replaced "satellite-derived curves" with "ESA CCI-based biomass–age curves constructed by Xu et al." and specified age grouping and Chapman–Richards fitting. Their construction is described in our response to **Reviewer #1 Comment #4.** We tested biomass–age curves refitted separately at the lower and upper limits of the reported biomass observational errors from ESA CCI Biomass. These tests did not use the AGB_{sd} field of the v6 2020 biomass map. Text S6 and Fig. S11 report the full-domain sensitivity results. Our responses to **Reviewer #1 Comments #3 and #8** explain the methods and remaining uncertainty in curve fitting, forest age, disturbance history, and model structure.

Comment #25

Ln 182: 'Data for parameter priors'. Are the parameters observable/measurable?

Response #25

LCMA is a measurable leaf trait, but not all 26 model parameters have directly measured reference values.

Our LCMA reference was derived from an SLA product. CUE was a data-driven estimate linked to f_{auto} by $CUE = 1 - f_{auto}$, and ce_{opt} was a separately pre-calibrated, model-specific reference rather than a direct observation. We renamed the subsection "Empirical Reference Constraints".

Comment #26

Ln183-191: As far as I can tell, only one of the 26 parameters is detailed here. Do you not need priors for the others? Without an explicit cost function, it is impossible to tell how this prior information is incorporated.

Response #26

Only ce, LCMA, and f_{auto} received direct soft reference penalties. The other 23 physical parameters were constrained by their bounds, assimilated observations, and applicable EDCs. We clarified this in the Methods

(**Line 233–238**):

"Of the 26 optimized physical parameters, only ce, LCMA, and f_{auto} received direct soft reference penalties. CUE was linked to f_{auto} through CUE = 1 − f_{auto}. The leaf lifespan–LCMA relationship was not an additional parameter-specific prior but constrained the realized annual foliar turnover generated by τ_{foliar}, c_{rfall}, and d_{fall} relative to the optimized LCMA. No direct empirical references were assigned to the other 23 physical parameters, which were optimized within the bounds listed in Table S1 and jointly constrained by the assimilated data streams and applicable EDCs."

Finally, we added the complete objective to Supplementary Text S1:

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

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 reference penalties were soft constraints: ce, LCMA, and f_{auto} remained optimized rather than fixed at their reference values. Their coefficients were fixed regularization strengths, not inverse observational variances.

Comment #27

Ln 211: Retrieved simultaneously from the observational data? Or parameters from the observational data?

Ln 212: I don't think it is the differentiable framework specifically that allows initial conditions to be treated as parameters.

Response #27

We clarified that the 26 model parameters and eight initial pools were jointly optimized by minimizing the calibration objective (**Line 259–262**):

"The initial sizes of the eight ecological pools (labile, foliar, wood, root, litter, soil organic matter, plant available water, plant unavailable water) and the 26 model parameters were retrieved simultaneously from the observational data. Automatic differentiation was used to calculate the gradients of the objective function with respect to the model parameters."

We removed the claim that treating initial conditions as optimized quantities is specific to a differentiable framework. DifferLand supplies the gradients used by Adam to update model parameters and initial pools.

Comment #28

Ln 216: I may have misunderstood but there are some conflicting messages here (note my later comments on Ln 521-522 and the closing sentence of the Discussion). There is inconsistency/incompatibility in terms of climate when cycling over meteorological drivers for simulations and using full biomass trajectories.

Response #28

We clarified that the curves represent a space-for-time relationship, not a historical time series for an individual

stand. The revised Methods section states (**Line 266–269**):

"The biomass–age curves combine biomass observations with stand ages derived from the 1985–2020 fire record using a space-for-time method. We thus cycled the 1985–2020 meteorological and atmospheric CO_2 forcing to simulate biomass accumulation, rather than to reconstruct the historical development of individual stands."

We did not separately test the effects of stand age and climate. This limitation is discussed in our response to **Reviewer #2 Comment #24**.

Comment #29

Ln 218: are the initial conditions set to 0? If this is the case, what is the calibration of them for?

Response #29

Initial pools were optimized and were not fixed to zero. The explanation is provided in Text S5, as quoted in our response to **Reviewer #2 Comment #22:**

"The initial live biomass was not fixed to zero because the current DALEC configuration does not explicitly represent seedling establishment or recruitment. The initial litter, soil organic carbon, and water pools were also optimized and need not be zero."

Comment #30

Ln 222: has the baseline experiment been performed before? If so, it should be clear what is new here.

Response #30

The baseline follows the existing DifferLand configuration. We reran it for the present domain and datasets as a reference for the two biomass-constrained experiments. The new contribution is the regional assimilation of grid-specific biomass–age curves across stand ages 1–100, as explained in our response to **Reviewer #1 Comment #2**.

Comment #31

Ln 225: Would this be a single point in a time series extracted from a single grid-point?

Response #31

The single-year constraint is one biomass value per 0.5° grid cell, calculated as the average over the selected mature-forest pixels. It is neither a time series nor a value from one native pixel. The revised Methods section states (**Line 286–289**):

"In the single-year biomass-constrained experiment, the 2020 biomass map provided a single mature-forest stock constraint rather than a biomass time series. This stock estimate was compared with the modeled total biomass at the end of the 12-year calibration simulation and was not interpreted as a biomass observation for calendar year 2012."

Comment #32

Ln 237: Five independent initialisations seems small. Are prior distributions known? Is there any uncertainty information incorporated here?

Response #32

The five optimization runs assessed optimization stability and cannot characterize the full parameter uncertainty. We clarified this in the Methods (**Line 295–297**):

"The five optimization runs assessed optimization stability. They were not samples from a probabilistic prior or posterior distribution and do not quantify parameter uncertainty."

We did not specify a joint probabilistic prior. Only ce, LCMA, and f_{auto} received soft reference penalties. The remaining parameters and initial pools were constrained by bounds, observations, and EDCs. Observational uncertainties were not used to weight the objective. The sensitivity tests and their limitations are described in our response to **Reviewer #1 Comment #3.**

Comment #33

Ln 252-259: Seems like a good place to illustrate with some time series.

Response #33

We expanded Fig. S5 to show biomass time series from all three experiments alongside the biomass–age curves. It is reproduced as **Fig. R3** in our response to **Reviewer #1 Comment #4**.

Comment #34

Ln 282: 'Prior constraint' – do you penalise movements away or do you hold it fixed?

Ln 285: I find the discussion of the cost function to be quite late in the document, especially since gradients have been discussed multiple times already. I also find it strange that the cost function never appears anywhere in the document (including the supplementary material) while normalising the observational term is.

Ln 293: What are the independent estimates of all the priors?

Ln 296: It would be good to see the cost function to see how this penalisation is quantified and how prior information is incorporated.

Ln 298-311: it is not clear to me which of these constraints is the canopy efficiency resulting from pre-calibration.

Response #34

The ce, LCMA, and fauto references entered as soft penalties. These parameters remained optimized and were not fixed at their reference values. Table S7 reports their optimized-to-reference ratios.

We introduced the objective where gradients are first described in Section 2.1 (**Line 131–133**):

"The calibration objective combines model–data fit terms, soft empirical reference penalties, and ecological and dynamical constraints. Its full formulation is provided in "Optimization and Cost Function" and Supplementary Text S1."

LCMA was derived from the SLA product, CUE from a data-driven estimate, and ce_{opt} from a separate pre-

calibration. The other 23 physical parameters had no direct empirical references. The sources and roles of these constraints are explained in our responses to **Reviewer #3 Comments #25 and #26**.

Supplementary Text S1 gives the complete objective and the equations and coefficients for its penalties. It is reproduced in our response to **Reviewer #1 Comment #11**.

The pre-calibrated canopy-efficiency reference is ce_{opt}. It enters through the soft P_{CE} penalty, separately from the pre- and post-integration EDC terms. The canopy-efficiency parameter ce remained optimized.

Comment #35

Ln 313-327: If I have understood this section correctly, a machine learning algorithm was used to identify the explainability of the calibrated parameters in terms of importance features. This section could do with some more detail, for example, training/testing etc and a little more explanation of the SHAP values.

Response #35

We expanded the Methods to specify the model configuration, training and validation procedure, and calculation and interpretation of SHAP values.

We added the gradient boosting model settings (**Line 401–402**):

"We used scikit-learn's GradientBoostingRegressor with 200 trees, a learning rate of 0.05, maximum tree depth of 5, a subsampling fraction of 0.8, and max_features set to 'sqrt'."

We clarified the implementation of the random five-fold cross-validation and the calculation of predictive performance (**Line 404–408**):

"Each fold was used once for validation, with the remaining four folds used for training. We repeated model fitting with 100 random seeds while keeping the fold assignments fixed and summarized each grid cell's held-out predictions by their median across runs. The goodness of fit was quantified using the coefficient of determination (R^2) and root mean square error (RMSE) calculated from the pooled held-out median predictions across the five folds."

We also expanded the SHAP explanation to identify what the values represent and how predictor importance was calculated (**Line 409–412**):

"SHAP quantifies each predictor's contribution to a model prediction relative to its expected value, allowing us to describe how these contributions vary along environmental gradients. We calculated interventional SHAP values for held-out median predictions. Relative predictor importance was calculated from the mean absolute values of these contributions across grid cells and normalized to sum to 100%."

Comment #36

Ln 340: and northern?

Response #36

We revised the spatial description to include localized biomass overestimation in northern grid cells.

Comment #37

Ln 341: Sentence beginning 'Notably,..' doesn't follow from the previous sentence.

Response #37

We removed "Notably".

Comment #38

Ln 348: I have understood correctly, the biomass growth curves have been used to calibrate the a 107 year run from the DALEC model. Given the complexity of the derivation of the growth curves, and the cycling over meteorological data in the simulation, R^2 seems very high.

Response #38

We corrected the simulation length to 100 years, covering stand ages 1–100. $R^2 = 0.99$ and RMSE = 11.3 Mg ha^{-1} describe agreement across grid cells at age 100 with the assimilated biomass–age curves. They do not measure fit across all ages or independent validation. This distinction is explained in our response to **Reviewer #2 Comment #31.**

Comment #39

Ln 353-361: Fig S6j-l don't appear to be mentioned.

Response #39

Fig. S7 (formerly Fig. S6) now has four columns, including the calibrated parameter values from the growth-curve-constrained experiment. The revised text cites fine-root allocation in Fig. S7k and fine-root residence time in Fig. S7o. We added "Fine-root residence times differed little from the baseline (Fig. S7o)."

Comment #40

Ln 355: parameter or parameters?

Response #40

We replaced "the parameter" with "the wood turnover parameter".

Comment #41

Ln 366: r, R and R^2 have been used, there should be consistency.

Response #41

We use Pearson's r for the allocation comparison because it retains the direction of the association. R^2 is used for the biomass comparison. We keep the notation consistent within each comparison, as explained in our response to **Reviewer #1 Comment #26.**

Comment #42

Ln 380: It would be preferable to move the Figures up so that captions do not break onto the following page.

Response #42

We adjusted the figure placement to keep Fig. 2 and its caption together.

Comment #43

Ln 396: Is this suggesting that the relationship is linear? It doesn't appear so in Fig 3c. While 3e is more obviously nonlinear, there is some nonlinearity in 3c.

Response #43

We revised the description to acknowledge the nonlinearity in Fig. 3c (**Line 496–497**):

"Solar radiation was found to be the most important predictor and showed a generally negative but nonlinear association with predicted wood residence time (Fig. 3c)."

Comment #44

Ln 402: should be Fig 3h.

Response #44

The revised text cites Fig. 3f, which shows annual mean temperature. Fig. 3h shows bulk density.

Comment #45

Ln 405-409: am I right in thinking that while DALEC doesn't distinguish between PFTs, the fraction of ENF and DBF emerge as important in determining leaf lifespan?

Response #45

Yes. ENF and DBF fractions emerged as important predictors in the post-calibration analysis, although DALEC was not parameterized separately by PFT. We clarified (**Line 505–508**):

"DALEC parameters were calibrated separately for each grid cell without PFT-specific parameter sets, while ENF and DBF fractions were used in the subsequent spatial analysis. Their associations with retrieved leaf lifespan therefore describe variation among grid cells with different vegetation composition."

Comment #46

Ln 421: and drier?

Response #46

Yes. We have clarified that this refers to drier atmospheric conditions (higher VPD).

Comment #47

Ln 436: DNF for consistency.

Response #47

We replaced “deciduous needleleaf forests” with “DNF” consistently in this paragraph.

Comment #48

Ln 440: is this different to how you might expect other PFTs to behave?

Response #48

The proposed interpretation concerns a general relationship with atmospheric water demand and is not intended to imply a response unique to DNF. We did not estimate separate VPD responses for individual PFTs. To clarify the scope, we revised the transition (**Line 538–539**):

"Across the study grid cells, leaf allocation also varied nonlinearly with atmospheric water demand."

Comment #49

Ln 445: I don’t see much of a plateau, more of a turning point

Response #49

We agree that Fig. 5g shows a turning point rather than a plateau. We replaced “but plateaued thereafter” with “but declined at higher temperatures.”

Comment #50

Ln 450: shift from wood allocation

Response #50

We clarified that the shift is from wood towards fine-root allocation and explicitly linked it to the contrasting temperature-related patterns:

"The contrasting allocation patterns along the temperature gradient suggest a shift from wood towards fine roots in colder grid cells (Fig. 4f and Fig. S15d)."

Comment #51

Ln 460: is the CARDAMOM product considered a benchmark? If so, what is the value added/novelty here?

Response #51

CARDAMOM provides an external comparison of allocation fractions and residence times. CARDAMOM values are also model results, so this comparison does not independently establish the accuracy of our retrievals. Section 3.3 reports regional magnitudes, relative differences, and comparisons with the published CARDAMOM intervals. The contribution of using biomass–age curves for calibration is explained in our response to **Reviewer #1 Comment #2,** and the updated CARDAMOM comparison is detailed in our response

to **Reviewer #1 Comment #5.**

Comment #52

Ln 491-516: more of a summary than a discussion of the results. There is also a lot of focus on equifinality as opposed to incorporating data that constrains the slow processes.

Response #52

We revised the opening two Discussion paragraphs, as described in our response to **Reviewer #2 Comment #35.** We shortened the results summary and removed the claim that the study directly tested equifinality. The revised opening states (**Line 588–590**):

"At the boreal-domain scale, our comparison of the three experiments indicates that grid-specific biomass–age curves add information on the rate and shape of long-term live-biomass accumulation. This information goes beyond that provided by high-frequency observations or a single contemporary biomass stock."

We also revised the interpretation of the slow-state constraint without claiming that equifinality was directly tested:

"Adding the biomass–age curve to the calibration objective penalizes parameter sets that reproduce the high-frequency observations but are inconsistent with that curve."

Comment #53

Ln 521-522: This statement appears to contradict the closing statement of the Discussion. What are the implications of cycling over meteorological drivers when assimilating against full biomass trajectories?

Response #53

We compared modeled biomass with the curve at every stand age from 1 to 100. The forcing sequence was 1985–2020, 1985–2020, and 1985–2012. Cycling it repeats its climate and atmospheric CO_2 trends. Thus, biomass at each simulated stand age is influenced by the weather and CO_2 level of a particular year in that repeated record. This design does not separately test the effects of stand age and climate.

We have clarified this limitation in the Discussion (**Line 632–636**):

"The biomass–age curves derived from the space-for-time method may not fully capture future growth under climate change and increasing atmospheric CO_2. Cycling the 1985–2020 drivers repeats their climate and CO_2 trends. Each simulated stand age is exposed to a particular year of that repeated record. This design does not separately test the effects of stand age and climate. The retrieved parameters therefore depend on the chosen forcing design."

The rationale for selecting the forcing period is provided in our response to **Reviewer #2 Comment #24.**

Comment #54

Ln 526-528: While the uncertainty in the observations is touched upon here, there should be a more extensive treatment of the observational and other uncertainties throughout.

Response #54

We clarified uncertainty treatment and the sensitivity tests in the Methods. The Results summarize the findings, and the Discussion explains the remaining limitations.

The Methods section now distinguishes numerical normalization from uncertainty weighting:

"NNSE scales the fit terms for different datasets but does not weight observations by their uncertainties."

Table S3 lists the uncertainty information available for each product and how it was used. Text S6, Fig. S10, and Fig. S11 report full-domain sensitivity tests for GPP, biomass–age curves, and the weight of the biomass term. The growth-curve-constrained experiment retained lower spatial mean wood NPP allocation and longer spatial mean wood residence time than the baseline experiment across these tests. The allocation difference narrowed under lower GPP and the upper biomass curve. The Methods and Results additions are quoted in our response to **Reviewer #1 Comment #3**.

The Discussion addresses uncertainty in the GPP, ET, RECO, and LAI products, including LAI seasonality. It also discusses the space-for-time and disturbance assumptions of the biomass curves, the dependence of retrieved model parameters on the jointly optimized initial pools, conversion from aboveground to total biomass, differences between field sites and grid-cell means, and model structure. It also states (**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."

Comment #55

Ln 533: Wouldn't any method incorporating accumulation trajectories achieve the same result? What was specific to DifferLand?

Response #55

Other calibration methods could in principle use the same biomass–age curves. In this study, DifferLand provided the differentiable DALEC implementation used to compute gradients through the full simulation and jointly optimize model parameters and initial conditions, as described in our response to **Reviewer #2 Comment #12.** We revised the relevant sentence accordingly:

"However, the growth-curve-constrained calibration retrieved effective mean parameters by optimizing against the biomass–age curve. DifferLand provided the differentiable DALEC implementation used to compute gradients through the full simulation and jointly optimize model parameters and initial conditions."

Comment #56

Ln 545: 'mitigate equifinality'

Ln 562: How is being more equifinal defined and how so in this context?

Response #56

Thank you. We removed the statement that additional long-term constraints could "mitigate" or "further reduce" equifinality, as described in our response to **Reviewer #2 Comment #37.** The revised sentence states (**Line**

675–676):

"Beyond forest biomass growth curves, additional long-term observations could provide complementary information on slow soil carbon pools which were not directly constrained in this study."

The phrase “highly equifinal” was also imprecise because equifinality concerns non-unique model solutions rather than the short-term fluxes themselves. We therefore replaced it with a direct description of the relevant information limitation (**Line 691–693**):

"In semi-arid and savanna systems, long-term constraints on woody cover and carbon stocks may be particularly valuable because episodic rainfall strongly influences short-term fluxes, which may therefore provide limited information on long-term woody carbon accumulation (Brandt et al., 2016; Poulter et al., 2014)."

Reference

Brandt, M., Hiernaux, P., Tagesson, T., Verger, A., Rasmussen, K., Diouf, A. A., Mbow, C., Mougin, E., and Fensholt, R.: Woody plant cover estimation in drylands from Earth Observation based seasonal metrics, Remote Sensing of Environment, 172, 28-38, 10.1016/j.rse.2015.10.036, 2016.

Poulter, B., Frank, D., Ciais, P., Myneni, R. B., Andela, N., Bi, J., Broquet, G., Canadell, J. G., Chevallier, F., Liu, Y. Y., Running, S. W., Sitch, S., and van der Werf, G. R.: Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle, Nature, 509, 600-603, 10.1038/nature13376, 2014.

Comment #57

Ln 566-568: This sentence needs rewording. The carbon sink is vulnerable to disturbances that shorten carbon residency times regardless of the study showing slower carbon turnover rates when incorporating biomass growth curves.

Response #57

We revised the sentence as suggested (**Line 696–698**):

"Our results show that assimilating the biomass–age curves yields longer effective wood residence times than calibration with high-frequency observations alone. Regardless of this difference between calibration experiments, boreal biomass carbon storage remains vulnerable to disturbances that shorten carbon residence time."

Comment #58

Ln 580: See my earlier comments regarding the use of biomass curves in a changing climate and cycling other meteorological drivers for simulations.

Response #58

The limitations associated with applying space-for-time biomass curves under a changing climate and with cycling the 1985–2020 forcing are now acknowledged in the revised Discussion. Please see our responses to **Reviewer #2 Comment #24** and **Reviewer #3 Comment #53** for the corresponding revisions and their implications.

Comment #59

Fig 1: some clarification is required here. The caption should make it clear that the single-year biomass stock was not included in the biomass growth curve experiment. I'm unclear on the initial pool set-up. Is it that the growth curves experiment provides the initial states for 2001 to the baseline and single year experiments? This is what I initially thought but Ln 246-251 suggest that the initial conditions are for the 100 year run. Why are the initial conditions being calibrated if simulating from stand disturbance when they should be almost zero? Doesn't that mean everything is disturbed at the same time. I'm quite confused about this aspect. The assimilation years for the single year experiment and the biomass growth curve experiment would be handy in the figure.

Response #59

We redrew Fig. 1 to distinguish the initial pools optimized by each experiment during calibration from the same starting pools used in the 100-year forward simulations. The figure separates calibration, temporal evaluation, and those forward simulations, and labels their constraints and time periods.

Each experiment optimized its own initial pools during calibration. For the baseline and single-year experiments, these pools define the start of the 2001–2012 simulation. For the growth-curve-constrained experiment, they define the initial state at the beginning of the 100-year simulation. The initial pools from the growth-curve-constrained experiment were used with all three parameter sets only for the subsequent 100-year forward simulations.

The revised figure also shows that the 2020 mature-forest biomass stock was included only in the single-year biomass-constrained experiment, while the biomass–age curve was included only in the growth-curve-constrained experiment. The calibration and evaluation periods are explicitly labeled.

As explained in Text S5, the initial live biomass was not fixed to zero because the current DALEC configuration does not explicitly represent seedling establishment or recruitment. The initial litter, soil organic carbon, and water pools were also optimized and need not be zero. Stand age describes the biomass–age curve, not a common calendar date of disturbance for all stands.

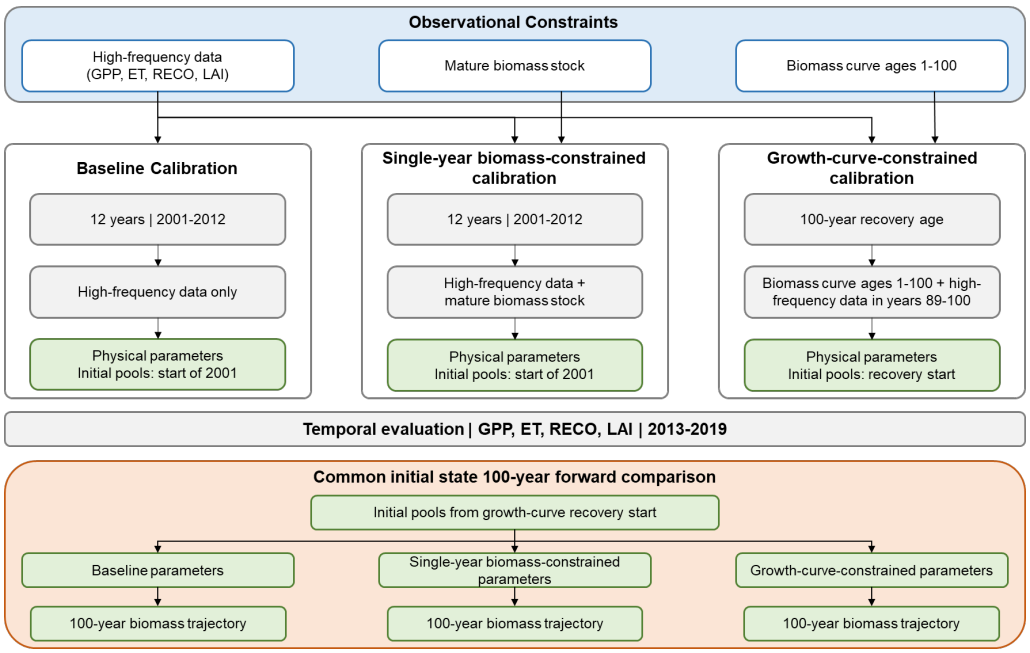


Fig. R9 (A copy of Fig. 1) Schematic illustration of the experimental design. The framework comprises three calibration experiments, each of which independently retrieved its own model parameters and initial pools. The baseline and single-year biomass-constrained experiments used 12-year calibration simulations corresponding to 2001–2012. The baseline experiment assimilated GPP, ET, RECO, and LAI, whereas the single-year experiment additionally assimilated a mature-forest biomass stock derived from the 2020 biomass map. This biomass stock was not assimilated in the growth-curve-constrained experiment. The growth-curve-constrained experiment used a 100-year calibration, assimilating the biomass–age curve over stand ages 1–100 and aligning the 2001–2012 GPP, ET, RECO, and LAI observations with simulation years 89–100. Observations from 2013–2019 were used only for temporal evaluation of GPP, ET, RECO, and LAI. Following calibration, the parameter sets from all three experiments were evaluated in 100-year forward simulations using the corresponding initial pools from the growth-curve-constrained experiment as a common prescribed starting state. These forward simulations were separate from the calibration simulations and compared biomass accumulation from the same starting pools.

Comment #60

Fig 6: What are the units? What range of raw values would be expected to determine if these differences are substantial?

Response #60

Fig. 6 now labels allocation differences as dimensionless differences between fractions on the 0–1 scale, and residence-time differences in years. Section 3.3 and Table S5 report the corresponding spatial means and 5th–95th percentile ranges. Mean wood allocation is 0.534 (0.232–0.758) in this study and 0.341 (0.157–0.600) in CARDAMOM v13. Mean wood residence time is 43.9 years (25.4–70.3) and 50.6 years (14.0–118.9), respectively. These ranges describe spatial variation. Comparisons with CARDAMOM v13 posterior intervals are reported separately in Fig. S17 and Table S6, as explained in our response to **Reviewer #1 Comment #5.**
