

Demography, dynamics and data: building confidence for simulating changes in the world's forests

Annemarie H. Eckes-Shephard¹ , Arthur P. K. Argles² , Bogdan Brzeziecki³ , Peter M. Cox⁴ ,
 Martin G. De Kauwe⁵ , Adriane Esquivel-Muelbert^{6,7} , Rosie A. Fisher⁸ , George C. Hurtt⁹ ,
 Jürgen Knauer^{10,11} , Charles D. Koven¹² , Aleksi Lehtonen¹³ , Sebastiaan Luyssaert¹⁴ ,
 Laura Marqués^{15,16} , Lei Ma⁹ , Guillaume Marie¹⁷, Jonathan R. Moore⁴ , Jessica F. Needham^{8,12} ,
 Stefan Olin¹ , Mikko Peltoniemi¹³ , Karl Piltz¹, Hisashi Sato¹⁸ , Stephen Sitch¹⁹ ,
 Benjamin D. Stocker^{15,16} , Ensheng Weng²⁰ , Daniel Zuleta^{21,22}  and Thomas A. M. Pugh^{1,6,7} 

¹Department of Physical Geography and Ecosystem Science, Lund University, Lund, 22362, Sweden; ²Met Office Hadley Centre, Exeter, Devon, EX1 1PB, UK; ³Department of Silviculture, Institute of Forest Sciences, Warsaw University of Life Sciences, Warsaw, 02-776, Poland; ⁴Department of Mathematics and Statistics, Faculty of Environment, Science and Economy, University of Exeter, Exeter, EX4 4QF, UK; ⁵School of Biological Sciences, University of Bristol, Bristol, BS8 1TQ, UK; ⁶School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, B15 2TT, UK; ⁷Birmingham Institute of Forest Research, University of Birmingham, Birmingham, B15 2TT, UK; ⁸CICERO Center for International Climate Research, Oslo, 0349, Norway; ⁹Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA; ¹⁰Hawkesbury Institute for the Environment, Western Sydney University, Penrith, NSW, 2751, Australia; ¹¹School of Life Sciences, Faculty of Science, University of Technology Sydney, Ultimo, NSW, 2007, Australia; ¹²Climate and Ecosystem Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA; ¹³Natural Resources Institute Finland (Luke), Helsinki, FI-00790, Finland; ¹⁴Amsterdam Institute for Life and Environment, Vrije Universiteit Amsterdam, Amsterdam, 1081, the Netherlands; ¹⁵Institute of Geography, University of Bern, Bern, 3012, Switzerland; ¹⁶Oeschger Centre for Climate Change Research, University of Bern, Bern, 3012, Switzerland; ¹⁷Science Partners, Paris, 75010, France; ¹⁸Research Institute for Global Change (RIGC), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 3173-25 Showamachi, Kanazawa-ku, Yokohama, 236-0001, Japan; ¹⁹Department of Geography, Faculty of Environment, Science and Economy, University of Exeter, Exeter, EX4 4QF, UK; ²⁰Center for Climate Systems Research, Columbia University and NASA Goddard Institute for Space Studies, New York, NY 10025, USA; ²¹Forest Global Earth Observatory, Smithsonian Tropical Research Institute, Washington, DC, 20013-7012, USA; ²²Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, 413 90, Sweden

Summary

Author for correspondence:
 Annemarie H. Eckes-Shephard
 Email: annemarie.eckes-shephard@nateko.lu.se

Received: 16 March 2025
 Accepted: 1 August 2025

New Phytologist (2025)
 doi: 10.1111/nph.70643

Key words: demographic vegetation model benchmarking, forest demography, growth–mortality dynamics, land–surface modelling, model intercomparison, postdisturbance recovery, self–thinning, vegetation carbon.

- Vegetation demographic models (VDMs) are advanced tools for simulating forest responses to climate and land-use changes, and are essential for projecting carbon cycling and large-scale forest management strategies. Despite their increasing incorporation into Earth System Models, VDMs differ in their demographic assumptions, with no prior quantitative comparison of their performance.
- We benchmarked nine VDMs against observational data from boreal, temperate and tropical sites, assessing their accuracy in predicting tree growth, carbon turnover, biomass stocks and size distributions. Models were simulated under consistent climate conditions with post-disturbance recovery monitored for at least 420 yr.
- Postdisturbance carbon recovery trajectories showed significant variability while remaining within observational ranges. Initial regrowth rates varied substantially (0.03–0.60, 0.18–0.70 and 0.35–1.10 kgCm^{−2}yr^{−1} for boreal, temperate and tropical sites, respectively), influenced by each model's initial forest state. Models captured mature forest carbon content but showed compensating effects between overestimated growth and underestimated mortality rates.
- This first multi-model benchmarking identifies growth and mortality rates as critical calibration targets and highlights the need to refine postdisturbance establishment conditions for model development. We outline specific benchmarking variables needed to improve predictions of forest responses to environmental change.

Introduction

Forests play a key role in the climate system, via storage of carbon, interactions with the water cycle, and provision of numerous other ecosystem services. Large-scale assessments of future forest

dynamics, state and distribution in response to climatic and land-use changes, are commonly simulated using Dynamic Global Vegetation Models (DGVMs; Cramer *et al.*, 2001; Sitch *et al.*, 2008; Friend *et al.*, 2014). The first generation of such models that emerged in the late 1990s and early 2000s is still

widely applied, but is limited by their omission of forest structural heterogeneity and demographic process representation (Friend *et al.*, 2014). Recent studies indicate that a substantial portion of global forests is currently recovering from historical land-use change and disturbances (Shevliakova *et al.*, 2009; Hurtt *et al.*, 2011, 2020; Pugh *et al.*, 2019, 2024; Besnard *et al.*, 2021; Dalmonech *et al.*, 2022), a phenomenon that cannot be captured without model representation of forest size structure. These recovering forests display varying strength of carbon uptake and storage over their regrowth period (Pretzsch *et al.*, 2014). They further influence the extent to which forests reflect incoming radiation and exchange water vapour with the atmosphere (Bonan, 2008). Meanwhile, ongoing environmental change influences individual-tree physiology and hence demographic rates and turnover times. For example, environmental conditions affect tree growth and mortality rates (cf. Pretzsch *et al.*, 2023), both at the level of the individual (Bugmann & Bigler, 2011; Needham *et al.*, 2020) and manifesting at the stand level (Yu *et al.*, 2024). These are dynamics that can most realistically be captured by models representing the interplay of tree growth and mortality rates with forest structure. Therefore, realistically simulating current and future vegetation demographics is necessary for forecasting both carbon cycling and local water and energy dynamics. Furthermore, anticipating the impacts of forest management strategies as a means to reach net zero commitments requires a representation of forest demography.

The first generation of DGVMs largely approximates vegetation heterogeneity through distinguishing plant functional types (PFTs), while each PFT is represented by a population of geometrically and functionally identical individuals. This approach simplifies the behaviour of a population to that of an 'average individual', meaning that a change in environmental conditions triggers a single PFT-specific average physiological response. This response manifests on all associated carbon pools (e.g. leaves, stems and roots) in the same proportional way, rather than allowing for variability among individuals. The size of the individual depends on productivity and mortality rates of the PFT in a given gridcell and PFT-specific allometric relationships. Even in models where the vegetation distribution is dynamic, these vegetation dynamics are limited to simplified competition between PFTs, based on the resource use efficiency and productivity of a given PFT's average-individual response to climatic conditions. Well-performing individuals would ultimately inhabit a larger fraction of a gridcell. Notably, demography is highly simplified. Mortality is considered by reducing the number of (even-sized) individuals. Recruitment is considered by increasing the number of trees but (artificially) mixing the new saplings' mass with existing stock, thus reducing the size of the average individual to conserve carbon (C) mass. This necessary artefact precludes any meaningful comparison with the large number of individual-tree-level observations that exist (e.g. tree rings or dendrometer data, or repeated forest inventories tracking individuals), foregoing a powerful potential constraint on the modelled behaviour. Overall, while computationally efficient, the average-individual approach obscures the impact of demographic processes, with substantial implications

for the accurate simulation of vegetation recovery and responses to environmental change.

In reality, demographic processes are fundamental to forest compositional and structural dynamics. For example, growth and mortality processes of individual trees change the self-organisation of tree crowns and gap dynamics, thus affecting the light profile within the forest and driving successional changes in forest structure and composition (Brown & Parker, 1994; Canham *et al.*, 1994; Montgomery & Chazdon, 2001; Matsuo *et al.*, 2021). In turn, the light profile and other micro-environmental conditions caused by forest structure influence establishment rates (Gleason, 1917; Klopčič & Boncina, 2012). More specifically, tree growth and establishment rates are driven by species-specific strategies (Wright *et al.*, 2010; Reich, 2014), modified by climatic conditions (Lempereur *et al.*, 2015; Fridley & Wright, 2018; Babst *et al.*, 2019) and forest light conditions (Klopčič & Boncina, 2012). For example, both within regrowing natural (Chave *et al.*, 2020; Pretzsch *et al.*, 2023) and plantation forests (West *et al.*, 1981), growth rates can vary driven by competitive tree–tree interactions alone. As competition for resources (e.g. light, soil water and nutrients) intensifies with plant growth, this can lead to mortality. In even-aged monospecific forests, this competition-based mortality process has been described as the 'self-thinning law' (Reineke, 1933; Yoda *et al.*, 1963; Enquist *et al.*, 1998), and is used extensively in plantation forest management (Reineke, 1933; West *et al.*, 1997). Other mechanisms of mortality are dependent on tree size (Metcalf *et al.*, 2009; Johnson *et al.*, 2018), as many physiological processes are size-dependent. Likewise, tree size is a critical agent for light competition: tall trees can overtop the shorter trees and gain advantage. Size is a major determinant of the resilience and vulnerability of trees to drought (Nepstad *et al.*, 2007; da Costa *et al.*, 2010; Bennett *et al.*, 2015; McDowell & Allen, 2015), wind damage (Yap *et al.*, 2016), fire (Pausas & Keeley, 2021) and insect outbreaks (Carnicer *et al.*, 2011; Oliva *et al.*, 2014). The interplay of these demographic processes in a forest collectively determines the structure, resilience and vegetation response to environmental changes.

Overall, the change in tree growth and mortality rates throughout the course of forest regrowth changes the trajectory of carbon uptake and storage through time (Pretzsch *et al.*, 2014). Likewise, albedo and exchange of water with the atmosphere also vary in response to successional dynamics (Bonan, 2008), influencing carbon and local water and energy dynamics. Demographic processes influence and are influenced by forest functional diversity (Muscarella *et al.*, 2017; Ruiz-Benito *et al.*, 2017) and impact mammal habitat structure (Grelle, 2003; Sukma *et al.*, 2019). Therefore, forest demography ultimately impacts many aspects of ecosystem biogeochemical cycles, making the realistic modelling of these processes crucial (Bonan, 2008; Levine *et al.*, 2016) even for LSMs.

Vegetation demographic models (VDMs), in contrast to DGVMs, more closely approximate real forest structure and demographic processes, allowing them to realistically simulate vegetation in a transient state (Friend *et al.*, 1997; Moorcroft

et al., 2001; Smith *et al.*, 2001; Hurtt *et al.*, 2002; Sato *et al.*, 2007; Strigul *et al.*, 2008; Fisher *et al.*, 2015; Naudts *et al.*, 2015; Haverd *et al.*, 2018; Weng *et al.*, 2019; Pugh *et al.*, 2018; Argles *et al.*, 2020; Ma *et al.*, 2022; Weng *et al.*, 2022). VDMs represent size and/or age-structured ecosystem demographics, and in doing so resolve the processes of recruitment, growth and mortality. VDMs track either individual trees (Sato *et al.*, 2007) or, more commonly, ‘cohorts’ (groups of individuals with identical properties, such as size, age and functional type, simulated as a single representative individual). Size-based competition between these units (for light, water and nutrients) leads to an emergent forest structure. This structure is a property of the boundary conditions (climate and soil forcing variables, disturbances and human management), and plant physiological responses, determined by the tree’s life history (e.g. small–large tree) and PFT traits that collectively determine growth rates and mortality. Overall, the unique feature in contrast to first-generation DGVMs is that vegetation carbon emerges from these demographic processes and competitive interactions.

An increasing number of land-surface modelling groups are working towards incorporating demographic vegetation into their models, with the aim of more realistically representing vegetation dynamics in the land component of Earth System Models (ESMs; Bonan & Doney, 2018; Bonan *et al.*, 2024). These are not yet standard components of ESMs (e.g. Friend *et al.*, 2014; Fisher *et al.*, 2018; Fisher & Koven, 2020); however, with the exception of EC-Earth (Döscher *et al.*, 2022) and GFDL (Shevliakova *et al.*, 2024). Furthermore, while international benchmarking systems for land-surface models exist (Eyring *et al.*, 2016; Collier *et al.*, 2018; Melton *et al.*, 2020; Abramowitz *et al.*, 2024), they are not yet prepared for the types of outputs delivered by demographic models (e.g. forest structural variables, such as tree size distribution, and dynamic variables, such as growth and mortality rates).

So far, a multi-model benchmarking exercise on VDMs has been lacking, possibly due to several challenges we faced during the course of this work: the large variety of model philosophies, the need for identical definitions of output variables, (data available for multi-variable constraints (to avoid compensating effects during calibration), and benchmark definitions that are both ecologically plausible and model-inclusive. Here, we make the first combined assessment of how nine VDMs used in this study perform against observations and explore the pathway to developing effective benchmarking. Simulations cover the regrowth trajectory and mature forest state at each of a boreal, temperate and tropical site. We compare models against: a biome-wide chronosequence of carbon content during forest regrowth, plot-level mature forest carbon content, plot-level forest structure, plot-level growth and mortality rates and biome-wide self-thinning relationships. We discuss: the similarities and differences between models and between models and observations. Finally, we highlight process uncertainties and make recommendations on critical benchmarking variables and observations going forward that could substantially reduce uncertainty in modelling of forest futures.

Materials and Methods

Models

Nine DVMS (BiomeE, BiomeEP, CABLE-POP, EDv3, ELM-FATES, JULES-RED, LPJ-GUESS, ORCHIDEE and SEIB-DGVM), capable of global simulations were used to simulate forest regrowth and mature forest dynamics, emergent from forest structures under constant repeated 30-yr climate cycles.

Model differences relevant to this study are highlighted in Table 1. More model-specific information can be found in Supporting Information Methods S1. Model-specific PFT-species mapping is found in Table S1, and model-specific modes of biomass reduction are found in Table S2.

Comparisons of all these contrasting VDM structures are in many ways nontrivial, as there are many definitional and philosophical differences between their approaches. In order for these models to become a more mainstream feature of, for example, operational model intercomparison studies, it is necessary to find ways to consistently compare them both with each other and with data. For the purposes of this comparison study, for example, many models created bespoke output variables, such as gross woody biomass increment. We defined and performed a woody carbon budget balance check for each model to test for ‘demographic’ woody carbon balance (Methods S2; Eqn S1; Fig. S1). Not all modelling groups were able to implement all variables according to the ‘D-BEN’ (Demographic benchmarking) format (see table S3.1 in Notes S1 for the full list of variables simulated); where variables deviate, it is reported in Table S4. This study provides a platform for the first comparison of this type, and we anticipate that it will pave the way for subsequent efforts using a wider range of data products and emerging models.

Model simulations

Models were run following a common simulation protocol (Notes S1). Simulations were performed at stand level where possible, in order to be comparable against the observations. Simulations were performed for three locations representing three major forest types (boreal, Finland (FIN): 62.25 N, 23.25 E; temperate, Bialowieza (BIA): 52.75 N, 23.75 E; tropical, Barro Colorado Island (BCI): 9.25 N, –79.75 E). Simulations involved a spin-up period (according to each model’s own procedure), followed by 30 yr of equilibrium forest conditions. After this, all trees were killed, removed or reset (Table S2), and the forest was allowed to regrow for at least 420 yr or until equilibrium was reached (Table S5, column ‘lower’). PFTs were chosen to reflect the biome’s/site’s vegetation composition as closely as possible (Table S1). Observations of woody biomass regrowth dynamics, mature state woody biomass and forest structure were made available *a priori* for potential calibration. Some models were calibrated (by hand-tuning or direct parameterisation from observations), but modified parameters stayed within plausible ranges. Which models were recalibrated and their updated parameters are reported in Table S3. The meteorological forcing data used were a repeated 30-yr-long randomised climate time series

Table 1 Key differences in model structure and assumptions that are relevant in this project.

Process/model	BiomeE	BiomeEP	CABLE-POP	EDV3	ELM-FATES	JULES-RED	LPI-GUESS	ORCHIDEE	SEIB-DGVM
Self-thinning prescribed or emergent	Emergent	Emergent	Prescribed	Emergent	Emergent	Emergent	Emergent	Prescribed	Emergent
Initial recruits postdisturbance	Restocking	Restocking	C from atmosphere. function of existing biomass (growth suppression)	Restocking	C from seedbank with seedlings, restocked from NPP-turnover with decay-rate	C from extern npp. Number of saplings emerges	Sapling C: from estimate of first year's productivity. Sapling size depends on first year's productivity+ PFT allometry. Sapling count: stochastic	Saplings, from C and N from atmosphere	C from litter; initial size always 0.01 m dbh; sapling count stochastic
NPP (productivity) – level	Cohort	Cohort	Whole canopy	Cohort	Cohort	Whole canopy	Cohort	Cohort	Individual
Canopy structure (based on Fisher <i>et al.</i> , 2018)	Flat top crown with prescribed maximum crown LAI; layering according to tree height and crown area (PPA)	Flat top crown with prescribed maximum crown LAI; layering according to tree height and crown area (PPA)	External	Flat top crown with prescribed maximum canopy height; layering according to tree height	PPA with leaf layers	External	Flat top crown, but cohorts/PFTs are vertically overlapping/multilayer	Aggregated individual crowns; layering according to tree height; shade depends on solar angle	Individual crowns represented by cylinder shape
Light competition	Within and between PFT and cohort; explicit	Within and between PFT and cohort; explicit	Within PFT; implicit	Within and between PFT; explicit	Within and between PFT and cohort; explicit	Within and between PFT; implicit	Within and between PFT and cohort; explicit	Within PFT; explicit	Within and between PFT; explicit
Water competition	Between PFTs and cohorts	Between PFTs and cohorts	Extern	Between PFTs and cohorts	Between PFTs and cohorts	Extern	Between PFTs	Between PFTs	Between PFTs
Nutrient competition	Between PFTs and cohorts	Between PFTs and cohorts	Extern	Between PFTs and cohorts	Optional (not simulated in these simulations)	Extern	Between PFTs	Between PFTs	Not available
Active mortality mechanisms	Density-dependent, density-independent	Density-dependent, density-independent	Density-dependent, background	Density-dependent, density-independent	Density-dependent, background, stress-based	Background	Density-dependent, stochastic, density-independent, stochastic	Density-dependent, density-independent, background mortality	Density-dependent, density-independent, background mortality
Vegetation unit	Cohort	Cohort	Cohort	Cohort	Cohort	Cohort	Cohort	Cohort	Individual trees for woody PFTs, cohort for grass PFTs

For example, while some models simulate harvest, this is not listed here. Initial recruits postdisturbance: Model-specific settings and assumptions determine the initial size and relative composition postdisturbance and hence influence carbon regrowth trajectory. Here, carbon for new recruits is normally taken from different sources to ensure carbon budget closure (see restocking). Recruit size is either prescribed or emerges. Recruit number normally emerges. Species composition (i.e. what was allowed to grow) was prescribed in this project. Where the term *restocking* was used: recruit number and size were forced, by resetting to the initial cohort structure, so these recruits were not dependent on previous conditions, such as PFT productivity/seedbank strength/litter size. Self-thinning: *prescribed* self-thinning can be thought of as a function that relates mortality to the number density and size of individuals. *Emergent* can be thought of as a mortality rate that increases when light conditions/productivity decline, likely caused by the presence of too many trees. NPP (productivity)-level: *Whole canopy structure* indicates that productivity is calculated for the entire canopy as an integrated unit, which may include vertical light gradients and sunlit/shaded leaf distinctions, but does not separately track productivity for individual cohorts or size/age classes within PFTs. *Cohort*: often a 'tree-level productivity', upscaled to cohort with stem density. Canopy structure: uses expressions from Fisher *et al.* (2018), to distinguish between the ways vertical light distribution and thus light resource acquisition/competition/crown organisation. *Individual*: light profile on every individual differs, shading can also occur from an angle. *PPA*: Crowns self-organize into discrete canopy layers, within which all plants receive the same incoming radiation. The lower layer trees' light is determined by the leaves in upper layers. *Flat-Top*: the total leaf area of each cohort is conceptually distributed evenly across the whole patch and can be very thin. Problem: marginally taller cohorts outcompete their neighbours in terms of light availability. Nuances of this approach to address this are described in the table. Light competition: *Implicit* – competitive effects approximated through empirical or other relationships without mechanistic calculation of light interception. *Explicit* – competitive advantage determined by directly modelling light absorption and carbon gain differences between cohorts. Active mortality mechanisms: density-dependent (e.g. competition for resources, e.g. light (or carbon), water and nutrients); density-independent (related to the individual: e.g. size or age); stress-based mechanisms that are not density-dependent (e.g. frost, heat or cavitation mortality); and background (e.g. a constant rate). Some demographic models do not treat all processes themselves, which is marked by 'external'. For more model-descriptions and references, see Supporting Information Methods S1; Tables S1, S2.

(Notes S1) for 1991–2020 from CRUJRA v.2.2 (University of East Anglia Climatic Research Unit, 2021), and carbon dioxide (CO₂) levels were kept at 2020 levels at 412 ppm. Where applicable, models set nitrogen deposition values at constant 2015 levels specific to the modelled site: 5.01 kgNha⁻¹ yr⁻¹ (FIN), 10.07 kgNha⁻¹ yr⁻¹ (BIA) and 3.57 kgNha⁻¹ yr⁻¹ (BCI), based on Lamarque *et al.* (2010).

Observations

We investigate a set of six benchmarks specific to VDMs and assess their contrasting structural and process assumptions. Forest size structure: this integrates assumptions on allometry (how carbon is partitioned among individuals of different sizes) and demographic processes representation (establishment, growth and mortality). Woody growth fluxes: these interact with tree allocation and can highlight biases in productivity assumptions and are themselves influenced by vertical forest structure and individual-tree size. Mortality fluxes: these influence carbon turnover and shape forest dynamics. For example, the tree size at the time of death can have benign (small tree) or profound (large tree) effects on subsequent forest structure, dynamics and thus the overall carbon pools. Self-thinning: this is an emergent density-dependent mortality process that can serve as a multi-constraint on productivity-mortality compensatory effects in conjunction with ‘tree/carbon packing density’ in the simulated forest: The self-thinning rule describes the relationship between mean individual biomass (M) and the number density (N) of trees in an even-aged forest with: $M = k N^p$, with k being a species or environment-specific constant. The exponent, p , has been found to vary around the originally proposed value of $-3/2$, due to factors like species, climate, provenance and stand history (Pretzsch, 2006; Comeau *et al.*, 2010) or depending on the theoretical model applied (Mrad *et al.*, 2020). We work on the assumption that self-thinning operates as a general ecological process also in mixed-species forests (Pillet *et al.*, 2018). Carbon stock changes during regrowth: biomass (carbon) regrowth curves capture the net effect of forest dynamics during recovery after a disturbance, testing the ability of the VDMs to capture a phase in which many forests globally currently are. How the interplay of forest structure, mortality and growth differs during forest recovery and the mature forest state. During recovery after a strong disturbance, structure is initially homogeneous, but growth and mortality fluxes are overall very unequal and variable in time, whereas in mature forests, growth rates and mortality fluxes are relatively equal overall. A demographic model has to be able to replicate these diverse emergent properties and fluxes over the period of forest recovery.

Together, these benchmarks capture key interactions between growth, mortality and allometry, forming a robust multi-constraint dataset for demographic model evaluation. For model calibration and comparison, we used multiple types of datasets, biome-level regrowth curves of aboveground woody carbon ($C_{\text{wood,AG}}$) and mass-number density data as well as site-level mature forest $C_{\text{wood,AG}}$ dynamics and stand structure information. Besides the benchmarking datasets used in this study, VDMs uniquely enable the use of specific additional data constraints

(Table 3) which are highly abundant but incompatible with non-demographic models.

Site descriptions FIN: data were obtained from 57 conserved or low-intensity managed (last management > 30 yr ago) sites in central and southern FIN (Peltoniemi & Mäkipää, 2011).

BIA: five permanent monitoring plots, spread across Białowieża National Park, situated in northeast Poland (at 52.500 to 53.000 N, 23.833 to 24.250 E), with a combined area of 15.44 ha (Brzeziecki *et al.*, 2016).

BCI: BCI is a 50-ha ForestGEO research plot in Panama (9.1543 N, -79.8461 E; Condit *et al.*, 2019; Davies *et al.*, 2021). It is located in a lowland tropical moist forest and contains *c.* 240 000 stems of 300 species of trees and shrubs with diameter at breast height (DBH) ≥ 1 cm.

Mature forest $C_{\text{wood,AG}}$ dynamics The nature of the three $C_{\text{wood,AG}}$ (kgCm⁻²) datasets used for site simulations was diverse, ranging from a single site (BCI) to over five sites within a national park (BIA) to a multitude of largely unmanaged sites across southern and central FIN. As the plot sizes, number of plots and frequency of measurement differed greatly at the sites, the approach to obtaining variability ranges also differed by site. We report the number of data points used to derive each benchmark in Table S6 and all data processing in Methods S3. Mature forest benchmarking observations are shown in Fig. S2 and are used in a more aggregated form in Fig. 1.

Mature forest woody turnover time (τ) and woody biomass growth (WB_{growth}) The observations were aggregated from individual-tree measurements to stand level and based on the above-described census data from the sites. Calculations are described in Methods S3; we made use of all datapoints for the comparison.

Stand structure DBH data from the same three sites as above were binned into dbh classes of < 1 cm, < 5, < 10, < 15, < 20, < 30, < 40, < 50, < 60, < 70, < 80, < 90, < 100, < 150, < 200 and ≥ 200 , corresponding to simulation output. BCI, FIN and BIA report saplings starting from 1, 5 and 10 cm DBH, respectively, which map onto our dbh-class categories (‘< 5’, ‘< 10’ and ‘< 15’).

Stand structure data for number of stems per dbh class ($n_{\text{stem_size}}$), as well as aboveground woody carbon biomass per size class ($c_{\text{wood_size}}$) are reported. To obtain an upper and lower observational range for BCI, we performed bootstrapping (Table S6) and report the 95% confidence interval and median for each dbh class for the latest year recorded (2015). For BIA, we report the median of five sites, alongside the minimum and maximum across all sites for each dbh class for the latest year recorded (2012). For FIN, we report median, maximum and minimum of 57 sites, from the census conducted in years 2006–2007. Stand structure benchmarking observations are shown in Fig. S3.

Biome-level $C_{\text{wood,AG}}$ regrowth curves Regrowth curves were not available at the site level and were instead derived from

chronosequence data. For the tropics and temperate regions, we used precompiled data from Teobaldelli *et al.* (2009) and Poorter *et al.* (2016), respectively. For boreal biome-level data, we used biomass chronosequence data from conserved low-intensity

management (last management > 30 yr ago) stands in FIN (Central and South FIN regions; Korhonen *et al.*, 2021). More detailed postprocessing is described in Methods S3. The data's spatial extent is shown in Fig. S4. For all biome datasets,

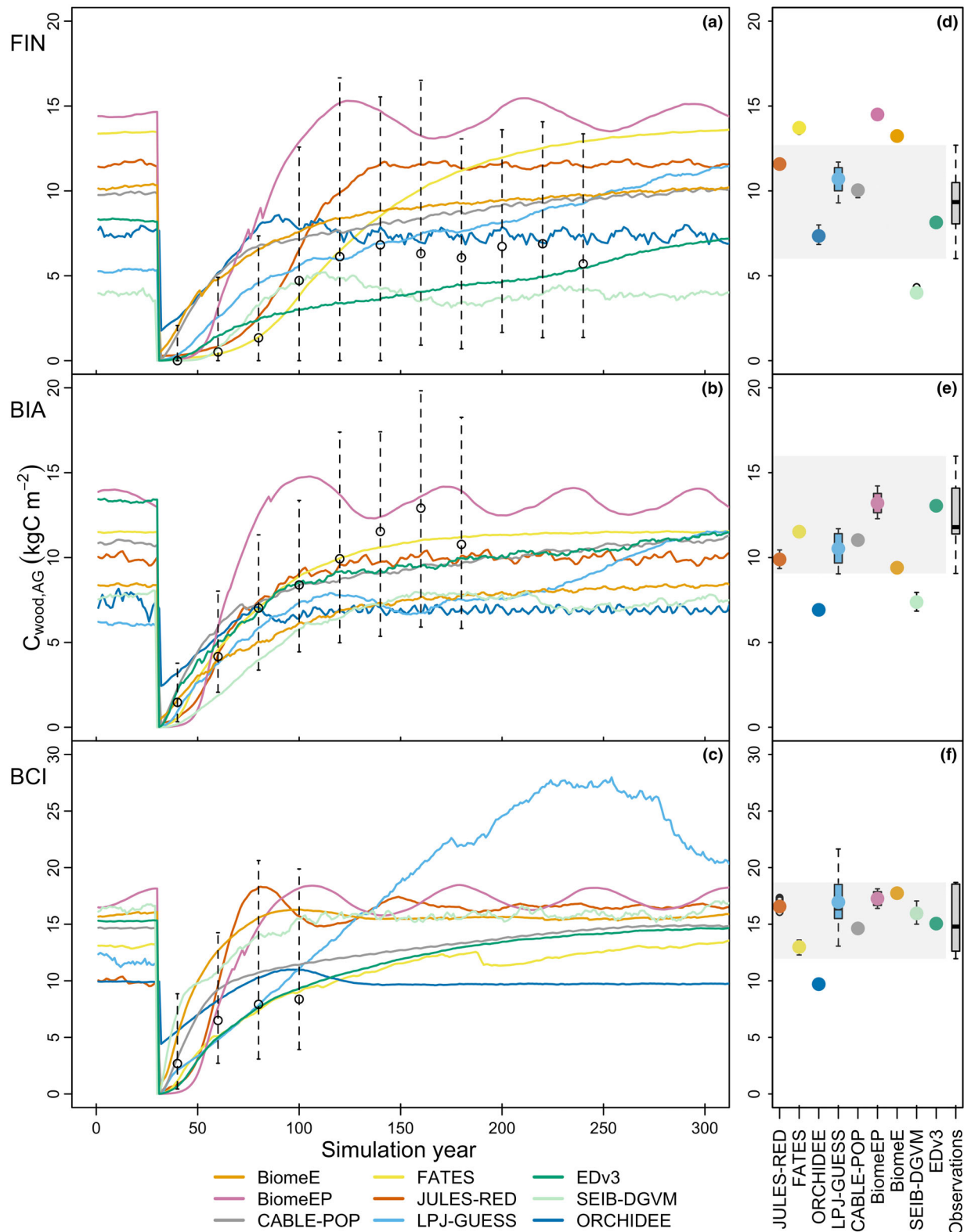


Fig. 1 Comparison between aboveground woody carbon ($C_{\text{wood,AG}}$) recovery (a–c) and aboveground woody carbon at mature forest equilibrium (Supporting Information Table S5) (d–f) after stand-replacing disturbance of the forest at a boreal (FIN), temperate (BIA) and tropical site (BCI). (a–c) Observations (black) of age bins of $C_{\text{wood,AG}}$ during regrowth and FIN mature forest $C_{\text{wood,AG}}$ are reported as median, and the whiskers end at the 10th and 90th percentile boundaries ($n(\text{age bins})$ or $n(\text{mature forest plot}) > 20$). (d–f) Mature forest $C_{\text{wood,AG}}$ observations at BCI are taken from median and upper and lower confidence interval values from spatial bootstrapping (6.25 ha) of a 50 ha plot with 95% confidence interval, across four repeated censuses. For mature forest at BIA, median, maximum and minimum observation values of five plots across six censuses were used. Mature forest at FIN observational range is derived from median, 10th and 90th percentile ranges from five census years. Model output at mature forest equilibrium was taken from a varying length of time series, depending on the model (Table S5). The time to equilibrium differed between models, and we also report individual model output graphs for their full time series in Notes S4. BCI, Barro Colorado Island; BIA, Bialowieza; FIN, Finland.

$C_{\text{wood,AG}}$ values were organised in 20-yr age bins, and we report the median, lower 10th percentile and upper 90th percentile boundary for these bins. At least 20 observations are in each bin, following Pugh *et al.* (2019). The resulting $C_{\text{wood,AG}}$ regrowth chronosequences cover 220, 160 and 80 yr for FIN, BIA and BCI, respectively (Figs 1, S2).

Mass-number density data for self-thinning-benchmarking

Observations on $C_{\text{wood,AG}}$ and number of stems (n_{stem}) were extracted from chronosequence data in Teobaldelli *et al.* (2009). We selected for broadleaf groups and conifer groups for comparison against the BIA and FIN sites, respectively, and removed dbh classes < 10 cm DBH. Observations from Poorter *et al.* (2016) do not contain the information to create a mass–density scatter of their regrowing forests data. To visualise mass-number density relationships of the observations, we calculated the logarithm of mean individual biomass log ($C_{\text{wood}}/n_{\text{stem}}$) and logarithm of n_{stem} (Fig. 3a,b).

Model data postprocessing

Identification of self-thinning in models To determine the self-thinning slope in the models, we fitted a linear model to the natural logarithm of tree number (n_{stem}) and mean tree biomass ($C_{\text{wood}}/n_{\text{stem}}$) for the subset of years during regrowth identified to be within the self-thinning period, fitting a linear regression with the `lm()` function with default settings from the `STATS` package in R Core Team (2023). Similar to the observations, we omitted dbh classes below 10 cm DBH, except for CABLE-POP, where this dbh-class distinction cannot be made in the current output. To determine the slope of the self-thinning curve and compare to observations, it was necessary to determine which period of the simulation was relevant to the process of self-thinning defined as even-sized competition. The period was extracted based on a combination of criteria, depending on the model (Table S7). In models that provided diagnostic outputs for self-thinning-related mortality (e.g. CABLE-POP, LPJ-GUESS and ELM-FATES), the selected period corresponds to when the self-thinning mortality rate exceeds the 95th percentile of annual self-thinning mortality rates observed during the simulation (Method 1, Figs S5, S6). This approach ensures focus on periods when self-thinning is the dominant mortality mechanism. For models with no explicit thinning-related mortality mechanism, the thinning period was either (Method 2, Fig. S7) chosen as the period with the top 5% of total mortality rate (e.g. BiomeE, EDv3 for BIA), or (Method 3, Fig. S8) the consecutive

points between the minimum and maximum densities, chosen semi-automatically as the consecutive points between the furthest to the bottom-left and the top-right points in the time series in self-thinning space (e.g. BiomeEP). Lastly, (Method 4, Figs S9, S10) the period was adjusted manually so that the trajectory of the simulation reflects what looks like a thinning trajectory (e.g. Enquist *et al.*, 1998) but could not be achieved with the semi-automatic approach of Method 3 (e.g. used for SEIB-DGVM). More in-depth explanations on the methods are found in Methods S4. Visualisations of the self-thinning period as part of the whole simulation period for each model and each site are found in Fig. S11.

Forest phase classification One notable feature of our comparison exercise is that the demographic models differed in their initial establishment conditions after the stand had been killed and therefore started their regrowth trajectory at different regrowth phases. We here defined additional qualitative ecological benchmarks, ‘forest phases’, which are distinct periods during forest regrowth describing the progression from open canopy to closed canopy. We use the forest phases benchmarks to meaningfully interpret the models’ behaviour alongside each other during regrowth. In a given forest phase, a collection of variables, such as canopy area (CA), number of stems (n_{stems}), mortality rate ($\text{cmort}_{\text{rate}}$) and total vegetation carbon (C_{veg}), must act together in a characteristic pattern that reflects the defining dynamics of that phase. For example, either small trees die in large numbers, detectable in the mortality flux, or they grow into larger trees, visible by them shifting over time from one size-class category to another. The former indicates that self-thinning is ongoing, whereas the latter indicates that there are still enough resources available for all trees to increase in size.

The characterisation framework we present in Table 2 was used to first determine which phase each model commences regrowth from (Table S9), followed by aligning the model output of growth and mortality rates along this common trajectory (Fig. 2). Interpretation of commonalities in growth and mortality rate dynamics was carried out once the model trajectories were aligned.

Modelled mortality and growth rate postprocessing The fractional mortality rate, $\text{cmort}_{\text{rate}}$ ($\% \text{ yr}^{-1}$) was postprocessed from the mortality flux, C_{mort} ($\text{kgC m}^{-2} \text{ yr}^{-1}$) and woody carbon, C_{wood} (kgC m^{-2}):

$$\text{cmort}_{\text{rate}} = \frac{C_{\text{mort}}}{C_{\text{wood}}} 100 \quad \text{Eqn 1}$$

Table 2 Forest recovery-phase classification necessary for forest phase alignment.

Phase	Variables involved/dominant	Description of dynamics in concert with auxiliary variables
Open-canopy phase – presence of grasses	%C _{veg} , <i>n</i> stems (<i>n</i> ha ⁻¹), CA (m ² ha ⁻¹), <i>cmort</i> _{rate} (% yr ⁻¹)	Caused by delayed onset in woody establishment, an open canopy (low CA) prevails with low <i>n</i> stems, higher fraction of Grasses in %C _{veg} , and no tree–tree competition (low <i>cmort</i> _{rate})
Open-canopy phase – growth	<i>n</i> stems (<i>n</i> ha ⁻¹), %C _{veg} , <i>cmort</i> _{rate} (% yr ⁻¹), CA (m ² ha ⁻¹)	Rapid increase in CA and fraction of trees in %C _{veg} ; <i>n</i> stems are relatively stable, but will decrease, once <i>cmort</i> _{rate} increases (a gradual rise in <i>cmort</i> _{rate} may begin towards the end of this phase)
Closed-canopy self-thinning	<i>cmort</i> _{rate} (% yr ⁻¹), <i>n</i> stems (<i>n</i> ha ⁻¹), CA (m ² ha ⁻¹), %C _{veg} , CAI (–)	A mortality (<i>cmort</i> _{rate}) spike and steep decline <i>n</i> stems (<i>n</i> ha ⁻¹), where not all established individuals grow into a larger size class, and self-thinning mortality occurs. Also indicated by canopy closure. CA c. 10 000 m ² ha ⁻¹ (but see Table S8)
Closed-canopy late-successional phase	%C _{veg} , <i>cmort</i> _{rate} (% yr ⁻¹)	A shift in composition is evident in %C _{veg} of the PFTs present (where models simulate this) and <i>cmort</i> _{rate} can show a slight directional change in the trajectories, caused by both the PFT composition and/or the forest structure now shift into a secondary succession stage, moving towards a dynamic equilibrium

Description of variables that are indicative of certain phases in forest development, as found in the model ensemble. These definitions were used to re-align the models to a similar forest state postdisturbance. The output time series of all variables per model and site are found in Supporting Information Notes S2 where modelling groups provided evidence and comments on the potential transition and starting points. CAI, crown area index. A more detailed description of forest phase classification, complementing this table, is provided in the Table S8.

Stand growth rate was directly obtained from the model output as the growth increment flux, *WBgrowth* (kgC m⁻² yr⁻¹). For this time series analysis, we applied a 30-yr smoothing on the mortality and growth increment flux to remove climate features from the repeated 30-yr scrambled sequence. This approach also reveals the dominant patterns in the model dynamics. To better capture postdisturbance dynamics, we excluded the 30 yr immediately following spin-up from the time series before applying the smoothing. We applied the function *rollmeans* (*k* = 30, *align* = 'left'), with otherwise default settings from the *zoo* package (Zeileis & Grothendieck, 2005) in R.

Turnover time τ is given by:

$$\tau = \frac{C_{\text{wood}}}{C_{\text{mort}}} \quad \text{Eqn 2}$$

For turnover time (Eqn 2) and mature forest woody growth flux comparison, we used annual output values from each model's respective equilibrium time period.

For comparison against observed values of turnover time and *WBgrowth* rate (Fig. 4), we did not smooth the output in this analysis, as smoothing the model time series over 30 yr would eliminate the temporal variability that serves as an approximation of the spatial variability captured in the observations. This would effectively remove the range depicted in the boxplots for many models in Fig. 4.

Results

Forest recovery

Carbon pools Ecosystem carbon recovery of all VDMs was evaluated against observations across three biomes (Fig. 1). All models, with one or two temporary exceptions, fall within the observed range of variability of aboveground woody carbon

(*C*_{cwood,AG}) recovery dynamics (Fig. 1a–c). At FIN, modelled *C*_{cwood,AG} recovery is predominantly above the median observed recovery during all time intervals (seven out of nine models). At BIA, *C*_{cwood,AG} recovery is initially spanning the whole range (first 20 yr) and progresses to be generally lower than the observed median in the following 100 yr (seven to eight out of nine models). Similarly, at BCI models span the full range of variability in the first 20 yr postdisturbance, but then continue to *C*_{cwood,AG} values at or above the median (seven of nine models). Consistent model-specific biases across all three sites, for example one model recovering consistently faster than the others, are not visible. Site conditions and PFT-physiological (i.e. parametric) differences may cause some models to react more strongly than others (for instance), SEIB-DGVM has one of the fastest *C*_{cwood,AG} recoveries in the first 20 yr postrecovery at BCI, and one of the slowest at FIN, whereas EDv3 recovers fast in the first 20 yr at BIA, but is in the mean to low model range after 20 yr at BCI. Regrowth speed does not seem to be linked to final equilibrium biomass values (Methods S7, Fig. S12). Initial *C*_{cwood,AG} wood levels postdisturbance differ across models, with ORCHIDEE being highest earliest at all sites. While the models fall within the range of the observations of *C*_{cwood,AG} at any given time point, there is significant variation both during recovery after disturbances and in mature forests among models. For example, after 30 yr of recovery, aboveground woody carbon ranges from 0.45 kgC m⁻² to 5.3 kgC m⁻² for FIN, 1.84 kgC m⁻² to 6.10 kgC m⁻² for BIA and 4.90 kgC m⁻² to 12.84 kgC m⁻² for BCI (see Table S10 for model-specific values). In mature forests, equilibrium values range from 4.0 to 14.50 kgC m⁻², 6.92 to 13.21 kgC m⁻² and 9.71 to 17.74 kgC m⁻², depending on the model (Table S10).

Woody growth and mortality rates during recovery Dynamics of woody biomass increment rates and woody mortality rates are highly variable between VDMs and across the first 250 yr

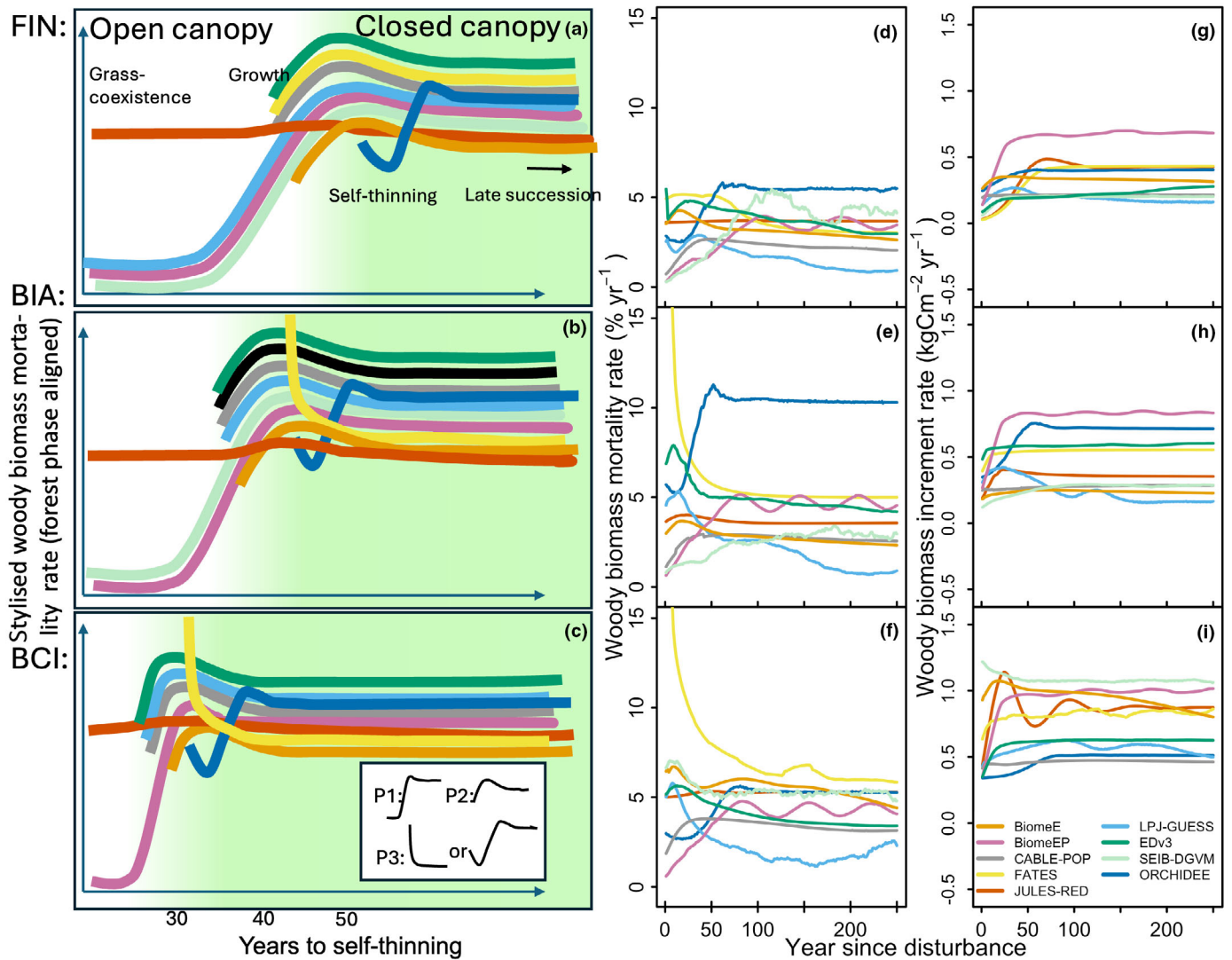


Fig. 2 Stylised woody biomass mortality rates (a–c) aligned on the forest recovery-phase timeline for FIN (a), BIA (b) and BCI (c). The years to self-thinning shown on the x-axis are an approximate ensemble mean for each site. Unaligned, 30-yr-smoothed woody biomass mortality rates (d–f) and unaligned, 30-yr-smoothed woody biomass increment rates (g–i). The different vegetation demographic models (VDMs) start their recovery after disturbance at different forest phases, but some mortality rate commonalities exist, expressed by the different shapes, representing each a different starting point or initialisation of regrowth postdisturbance: P1: mortality rate starts from low, nearly 0-values, peaks and then either decreases or remains flat, caused by a model initialising in open-canopy grass-coexistence phase. P2: mortality rate starts from higher values and then either peaks or remains flat, caused by a model initialising in open-canopy-growth phase. P3: mortality is immediately initiated, which causes a decline towards an equilibrium or which causes a decline and relief-period with less competition, until this peaks and stabilises (more detail in Supporting Information Methods S5), caused by a model initialised in closed-canopy self-thinning. The forest recovery phases are marked on the plots (a–c) as ‘grass-coexistence’ and ‘growth’, during open-canopy phase (white background) and a self-thinning and a transition towards secondary successional forest and a dynamic equilibrium is shown through the arrow (green background) during closed canopy. The green background is in a soft transition not only to indicate the gradualness of the process but also to highlight a difference between gap-models and PPA-based models in the transition towards self-thinning: where PPA-models can allow for canopy closure, but growth to continue until canopy packing is optimised, whereas canopy closure in gap-models initiates self-thinning. Note that FATES mortality rates are derived from total biomass and total biomass loss. FIN, Finland; BIA, Bialowieza; BCI, Barro Colorado Island.

after disturbance (with the least variability observed in JULES-RED).

In terms of mortality dynamics (Fig. 2d–f), the VDMs investigated here show a change in mortality rate over the course of stand development (largest change 23% yr⁻¹, mean change 4% yr⁻¹, smallest change 0.8% yr⁻¹). Modelled mortality rates are highest at different timings of regrowth, but

more commonly highest during early regrowth phases. These higher mortality rates during establishment and regrowth differ also in their magnitude relative to a model’s dynamic equilibrium that follows (not shown in the stylised graphs in Fig. 2a–c, but visible in Figs 2d–f, S13–S15), and the rate of decline of mortality rate towards the equilibrium is also different (Fig. 2d–f).

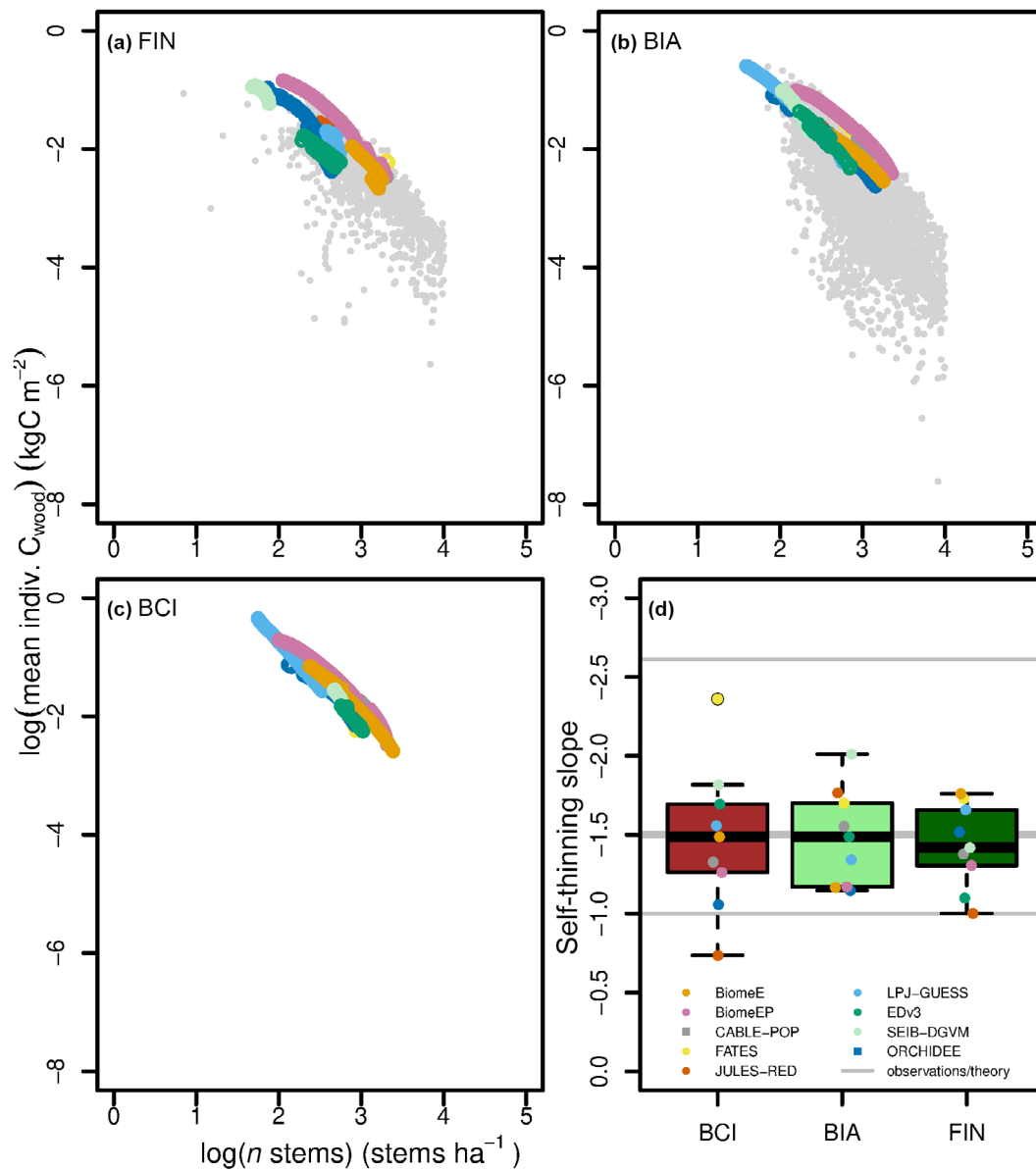


Fig. 3 Model output on average-individual woody biomass and stem numbers in ‘self-thinning space’ (a–c) and model self-thinning slopes (d) for all three sites. Grey dots in (a) and (b) are observed mass–density relationships of regrowing temperate and boreal forests, taken from Teobaldelli *et al.* (2009), filtered by tree size and stand age (> 10 dbh and > 10 yr). The coloured dots are the model-specific subset of the period of regrowth where the trajectory coincides with thinning behaviour (see also Supporting Information Fig. S11). These points were then used to fit the self-thinning line/slopes of each model, which are reported in the boxplot in (d) in context with the self-thinning slope $-3/2$, and observed lower (-2.612 ; Pretzsch, 2006) and theoretical higher (-1 ; Mrad *et al.*, 2020) thresholds, shown as grey lines. Boxplot whiskers extend to $1.5\times$ interquartile range. Self-thinning slope values are also reported in Table S7. Models with square instead of circle dots in (d) have prescribed or implicit self-thinning mortality.

In terms of growth dynamics (Fig. 2g–i), a similar picture emerges, with initially increasing woody growth increment rates for all VDMs, which normally peak or flatten early during regrowth. Models differ a lot, however, in the magnitude of change in growth rate during the first 250 yr, the amplitude of which ranges between 0.05 and 0.6 $\text{kgC m}^{-2} \text{yr}^{-1}$ depending on the VDMs. Even in shorter periods, such as 30 yr directly after disturbance, model annual growth rates can span 0.03–0.60, 0.18–0.80 and 0.35–1.10 $\text{kgC m}^{-2} \text{yr}^{-1}$ for FIN, BIA and BCI, respectively (Table S10).

To meaningfully assess the differences in simulated dynamics during forest regrowth, we realign mortality rates (% total $C_{\text{wood}} \text{yr}^{-1}$) of the VDMs during the first 250 yr of disturbance recovery (Fig. 2) and put them in context with prevailing forest recovery phases (Tables 2, S8, S9, S11; Methods S5).

Models begin their forest recovery at different phases of the regrowth trajectory and can be roughly divided into three groups (Fig. S16): Group 1: ‘open canopy – grass coexistence’ (BiomeEP, SEIB-DGVM, JULES-RED and LPJ-GUESS at FIN): relatively low mortality rates prevail initially (but can also be high

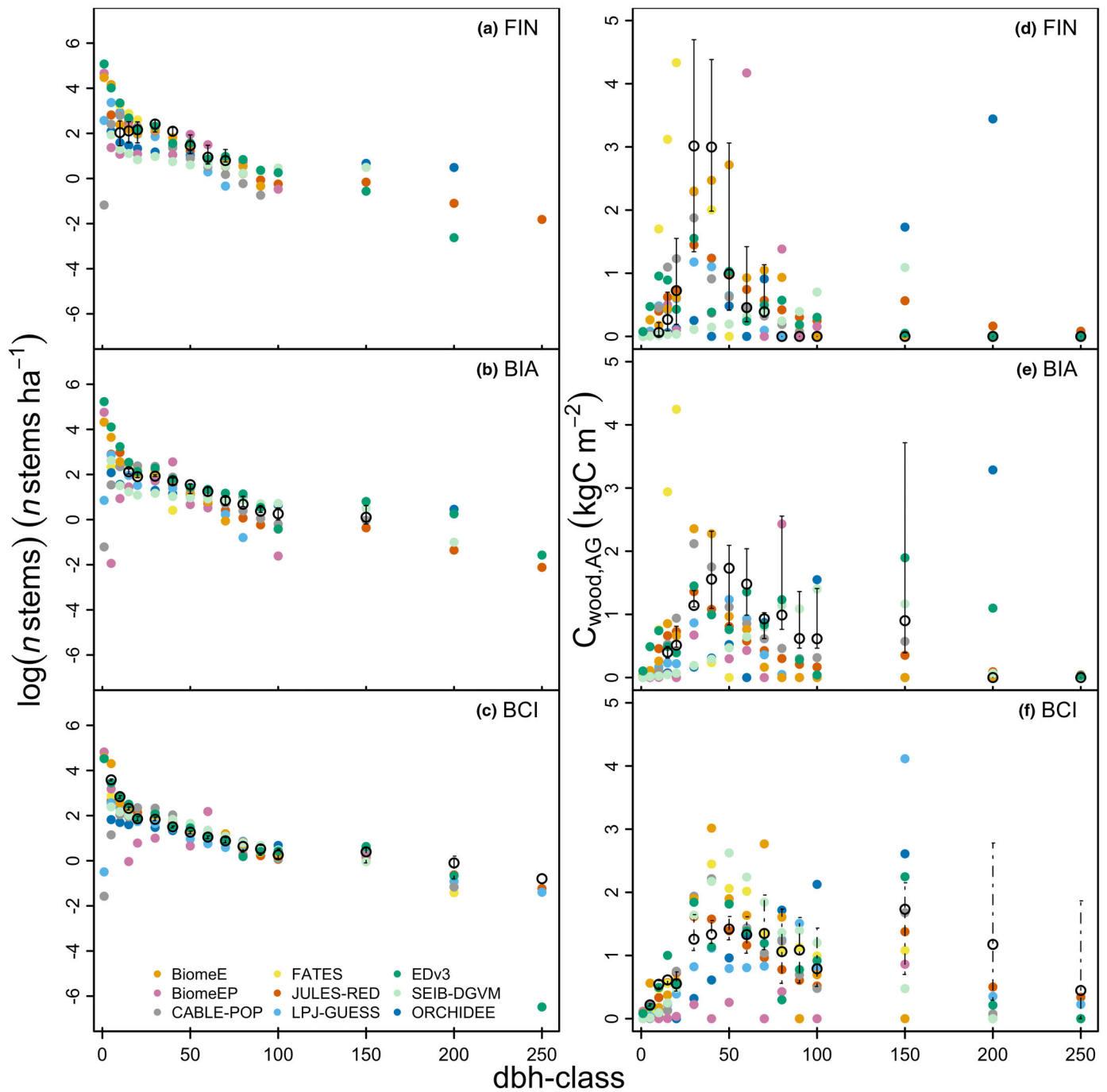


Fig. 4 Mature forest structure in terms of number of stems across multiple diameter at breast height (DBH) size-class categories ('dbh-classes') (a–c) and aboveground woody carbon ($C_{\text{wood,AG}}$) across multiple DBH across multiple size classes (d–f) against a boreal (FIN), temperate (BIA) and tropical (BCI) site. Observations (black): the median (open black circles) and whisker boundaries are derived using different methods: BCI is derived from spatial bootstrapping of a 50 ha plot and reported as with a 95% confidence interval (Supporting Information Table S6). For FIN, the range shown is the 10th and 90th percentile of binned size classes across multiple plots, and for BIA, we use the minimum and maximum available plot ($n = 5$) values. Note that LPJ-GUESS output origin deviates from the ensemble (Methods S6; Fig. S17). Model-specific individual plots can be found in Notes S3. BCI, Barro Colorado Island; BIA, Bialowieza; FIN, Finland.

if a model's saplings are allowed to compete with grasses or have a high mortality rate), and a significant presence of grasses is allowed due to either the low density and/or size of the first saplings. Group 2: 'open canopy/growth' (CABLE-POP, EDv3,

SEIB-DGVM at BCI, BiomeE, LPJ-GUESS at BIA and BCI): seedlings are established, and grass may be present but not as a significant fraction of vegetation biomass (% C_{veg} ; note again that there is no between-PFT competition in CABLE-POP and

ORCHIDEE, so the mechanism that reduces the % C_{veg} of grasses is purely the increase in woody PFT biomass, rather than the competition between woody PFTs and grasses). The length of this phase varies substantially between models. Group 3: 'closed canopy/self-thinning' (FATES at BIA and BCI, ORCHIDEE): while FATES forest phase identification (Table S9; Notes S2) shows a clear open-canopy grass-coexistence period, followed by an open-canopy-growth phase at BIA and BCI, their combined time is 20 to 6 yr respectively, and the dominant feature visible in the smoothing is the self-thinning mortality; therefore, FATES is here reported in group 3.

Self-thinning Self-thinning trajectories for most models are aligned along the upper boundary of the observations of observed mass–density relationships for BIA, and more spread at FIN (Fig. 3a,b). BiomeEP, CABLE-POP and FATES' self-thinning trajectories are the closest to the observed edge of mass–number density scatter. The number of points per model varies (Fig. 3a–c), representing the years the regrowing forest spent in a period of thinning. Thinning period duration varies across models and sites (BCI: ensemble mean: 44.75, min: 12, max: 98 yr; BIA mean: 70, min: 30, max: 128 yr; FIN mean: 80, min: 38, max: 175 yr; Table S7), but there is a general trend to fewer years of thinning at BCI. There is a mass–number density space in the observations with a high number of small trees and low mean individual mass which models do not occupy. Self-thinning slopes that emerge from the self-thinning lines are reported in Fig. 3d and fall within observed/theoretical ranges. There is no systematic behaviour visible for any model for only three sites in terms of self-thinning slope, but the ensemble spread is largest in the tropics and the ensemble median closely resembles the theoretical self-thinning slope.

Mature forest

Carbon pools Simulated mature forest $C_{cwood,AG}$ (Fig. 1d–f) either displays a very stable equilibrium (dots cover the boxplot for FATES, BiomeE, BiomeEP, CABLE-POP, EDv3, JULES-RED at FIN) while other models exhibit greater variability over time (whiskers are visible for LPJ-GUESS, SEIB, BiomeEP at BIA, JULES-RED at BIA and BCI). Models' mature forest $C_{cwood,AG}$ lies largely within the range of observations, with the highest number of models ($n=3$) deviating at FIN and BIA (where in the latter case all models underestimate mature forest carbon).

Forest structure The character of the stem number distribution across DBH size-class categories ('dbh classes') is generally captured in the models (Fig. 4a–c; Notes S3), although some models tend towards too low densities at FIN and BIA. The exception is the smaller size classes, the densities of which are commonly overestimated, apart from the tropical site BCI, where they are underestimated by the majority of models. Among the three sites, the general trends in tropical stem number distribution are captured best, while the highest number of models falls within the observational range at FIN, where variability in the observations is

largest. Models tend to predict larger trees than are observed at FIN and BIA (0 trees observed at dbh class 200 and 250), and underestimate tree numbers at BCI at dbh class 200 and 250, which manifests in high $C_{cwood,AG}$ in trees at FIN and BIA, but less than is observed at BCI. The general trend in forest structure (i.e. a high peak of $C_{cwood,AG}$ in the 40–50 DBH class, followed by a marked decline at FIN, a gradual decline after a peak at 50 DBH at BIA, and a gentler decline at 50 DBH at BCI) is generally replicated by the ensemble, if not by the individual models. ORCHIDEE and SEIB underestimate $C_{cwood,AG}$ in small-to-medium tree size classes while overestimating it in large classes. BiomeEP 'oscillates' in the number of stems at size classes below 100 DBH (i.e. some DBH classes can have 0 trees), and FATES has a small-tree bias at FIN and BIA, both in terms of stem numbers and $C_{cwood,AG}$ in stem.

Growth fluxes and carbon turnover time Observed ecosystem woody growth fluxes and turnover times from three mature forest sites were compared against model output (Fig. 5). Most deviation from the observations is visible at the tropical site BCI, where models overestimate the growth flux and in turn underestimate the turnover time. For BIA, most VDMs are within the observational variability of the turnover time, but this is more a feature of the high variability of turnover time recorded across the five plots than of a well-constrained model response. Generally, models at BIA cluster below the median turnover time of the observations and above the observational range for woody growth rate, as at BCI. Lastly, at FIN, the wide observational range allows more models to fall within growth increment flux variability, though similar to BIA, the majority of the models are below the median turnover time, and all models are above the median growth rate (Fig. 5d). VDM woody growth fluxes are generally increasing from the boreal to the tropical site, similar to the observations. BiomeE, LPJ-GUESS and CABLE-POP most often fall within the range of observations. Overall, gross woody growth increment ($\text{kgCm}^{-2} \text{yr}^{-1}$) is overestimated by the majority of the models at all sites (up to $0.6 \text{ kgCm}^{-2} \text{yr}^{-1}$), and turnover time is within the range of the variability or underestimated.

Discussion

Benchmarking forest recovery

While within the observational variability, VDMs differed significantly in the timing and speed of recovery following disturbance, especially in the first $c. 30$ yr (Fig. 1). We were not able to benchmark growth rates during the recovery phase due to a lack of suitably resolved observations. However, the large variability and overestimation in mature forest growth rates (Fig. 5d–f) are highly likely to be representative of the behaviour of the VDMs throughout the recovery phase, leading to the divergent forest carbon recovery trajectories in Fig. 1.

Furthermore, early recovery differences are caused by VDMs starting their regrowth trajectory from different forest 'initial establishment conditions' (open-canopy grass coexistence, open-canopy growth or immediate thinning conditions; Fig. 2). Once

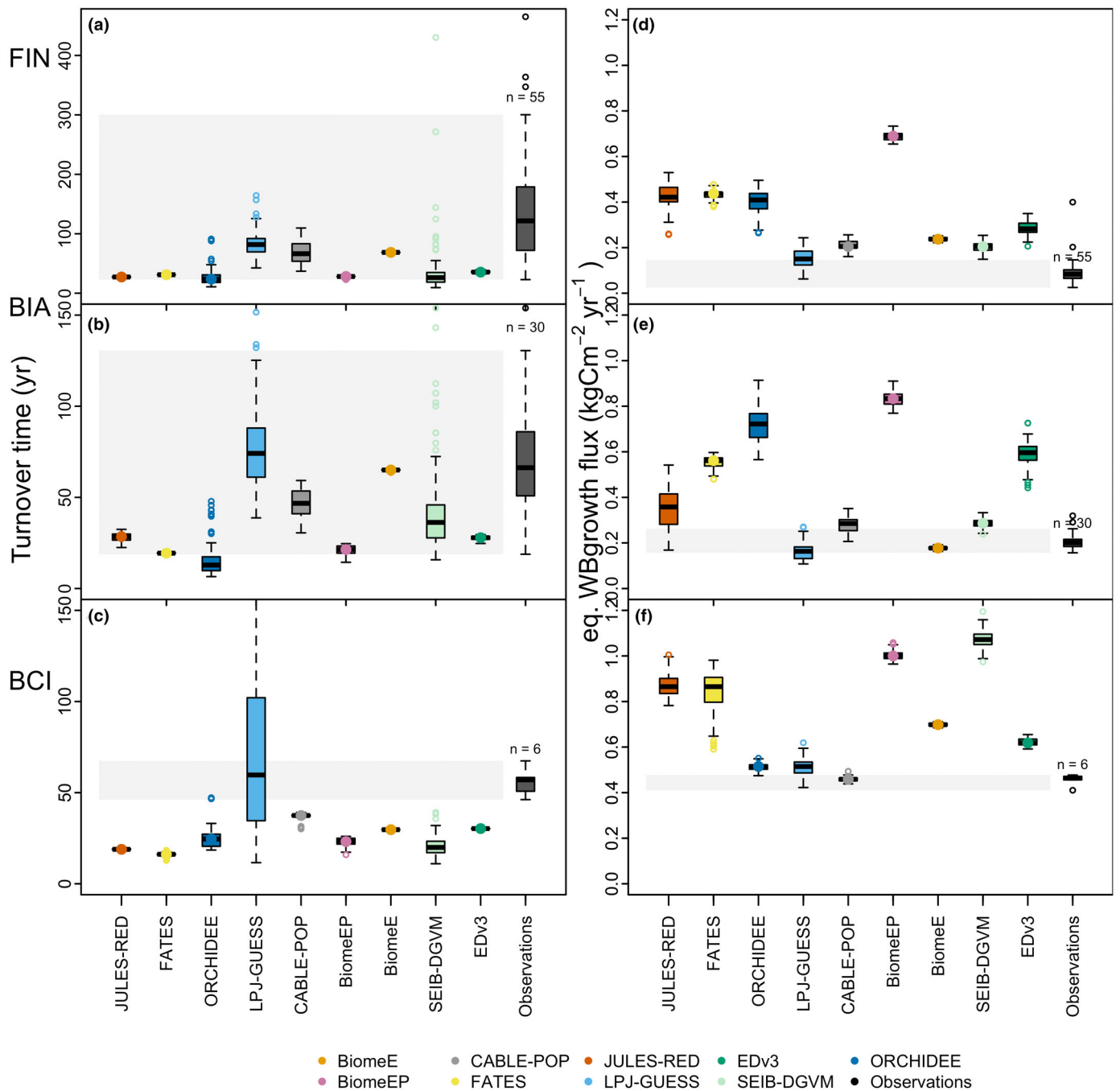


Fig. 5 Simulated and observed annual mature forest turnover time (a–d) and woody biomass increment flux (e, f) at a boreal (FIN), temperate (BIA) and tropical (BCI) site. Benchmark boxplots consist of observations from each year available and each plot available (e.g. five plots at BIA × 6 census years; 1 plot at BCI × 6 census years; and a total of 55 plot-year combination at FIN). Note that FATES includes nonwoody growth and mortality in its turnover variables plotted here. For BCI, the grey polygon shows the full data range. The upper whisker (1.5 × interquartile range) value for LPJ-GUESS boxplot in (c) is 193.34. BCI, Barro Colorado Island; BIA, Bialowieza; FIN, Finland.

we aligned the forest recovery phases, the general dynamics (i.e. high mortality rates during self-thinning, high growth rates before canopy closure) of the models are broadly similar, suggesting that important emergent ecological behaviours are captured. However, the underlying processes governing the timing of the

start and end of the phases and the magnitude of growth and mortality rates are in need of further development (Fig. 2).

We note that successional dynamics are probably a further contributor to the differences in forest recovery dynamics, but due to a lack of data, we did not address this within this study.

Plausible succession behaviour from shade-intolerant to shade-tolerant is visible in several models (Fig. S18), but the duration of their presence and degree of coexistence with late-successional PFTs varies, which can greatly contribute to different forest structure or turnover time, while the overall $C_{\text{wood,AG}}$ may be similar. This is a source of uncertainty in VDMs that requires benchmarking attention.

Growth and mortality rates during recovery show plausible dynamics in all models. In terms of growth rates during recovery, VDMs replicated some or all stages of 'boom-and-bust' growth dynamics described for tropical forests by Chave *et al.* (2020). After canopy closure, woody productivity decreases or remains at a stable peak, depending on the model. Similarly, observed stand-level production increases in Chave *et al.* (2020) throughout the 40 yr of record but declines sharply in the smaller size classes following canopy closure after 10 yr. In the models, the reductions in simulated woody productivity at the stand level could generally be attributed to some combination of increased respiration, changes in allocation of carbon to woody tissues as opposed to leaves, fine roots or reproduction, and ultimately changes to GPP due to community-weighted trait changes or due to height increases and succession in PFTs over time during development. For example, the two-phase pattern in growth rates in LPJ-GUESS and BiomeE is caused by a change in PFT-species composition with different physiologies (more clearly seen in Fig. S19).

Similarly, and in concert with growth rates, simulated mortality rates change over time in the stand as a result of competition. This is supported by observations of both temperate (Pretzsch *et al.*, 2023) and tropical forests (Rozendaal & Chazdon, 2015). The VDMs herein, which are capable of simulating self-thinning, are, in principle, able to account for these mortality dynamics and thus provide a crucial connection between forest recovery dynamics, the evolution of deadwood stocks, and hence forest soil dynamics. However, the large deviations in the timing of the different phases simulated by the models highlight the need for observational benchmarks that clearly differentiate between different forest phases in recovery. With forests in flux, and an increasing fraction of the landscape being in such a state (Shevliakova *et al.*, 2009; Pugh *et al.*, 2019), accurate representation and calibration of these regeneration and regrowth phases are clearly of high importance.

Overall, while we think that the novel qualitative benchmark (Table 2) used for the forest recovery-phase alignment is a cheap, pragmatic way of monitoring model-useful demographic transitions quickly. Nevertheless, the variables (e.g. CA, n_{stems} , % C_{veg}) and definitions will need refinement in the future. For example, the definition and monitoring of the closed-canopy self-thinning phase may need to accommodate the PPA multi-storey canopy organisation approach more appropriately, as there, understory can form and self-thinning is not immediately initiated once the canopy is closed. We faced challenges when attempting to balance ecological realism with model inclusivity. We view this forest phase framework as an important first attempt at aligning models within our community and with observed ecological processes, and are committed to improving these benchmarks in our community.

The self-thinning rule (Yoda *et al.*, 1963) is an ecological principle fundamental to understanding plant population dynamics and resource allocation. VDMs were able to replicate the self-thinning slope, demonstrating their capability to simulate key ecological processes, such as intertree competition for light based on size. However, the models varied substantially in the number of years they spent in self-thinning. This variation is likely determined by (a) the combination of growth rate and allometric assumptions that determine how quickly a recovering stand enters self-thinning and (b) the point at which other causes of mortality start to kill off the dominant trees. The general trend towards fewer years of thinning at BCI across models possibly reflects the higher growth rates at that site. As assumptions on canopy light environments impact the individual cohort's productivity, the differences in how VDMs treat canopy layers and light partitioning in the vertical canopy profile highlight an opportunity for future in-depth comparisons, particularly in the context of self-thinning and succession, but also coexistence.

The self-thinning period (time period a model spent in self-thinning) can also be influenced by the method of selecting self-thinning duration, which was kept as objective as possible, but varied depending on the model (see Fig. S11). We note that while the self-thinning period selected using selection Methods 1–4 was independently confirmed by the modeller in the forest recovery-phase analysis, it does not always align (Table S11). Self-thinning mortality in the models is conceptualised through carbon starvation mortality (EDv3, FATES, BiomeE, BiomeEP and SEIB-DGVM), growth-efficiency mortality (LPJ-GUESS and SEIB-DGVM) or explicit rules based on biomass–stem number relationships (ORCHIDEE, crowding mortality in CABLE-POP). Specifically, for ORCHIDEE, the biomass–stem number relationship is the main mechanism for triggering resource competition mortality more generally and therefore differs from the strict definition of self-thinning for even-aged stands in this work. ORCHIDEE's biomass–stem number relationship emerges well from the output but is a prescribed feature and cannot be used to determine the (even-sized) self-thinning period length. JULES-RED's self-thinning slope emerges as a combination of metabolic scaling theory, mortality and canopy area limiting seedling establishment (Argles *et al.*, 2020). Realistically replicating this well-established ecological behaviour (Yoda *et al.*, 1963; Enquist *et al.*, 1998) is crucial for more realistically simulating forest recovery and ecosystem carbon residence times.

Benchmarking mature forest

Simulations of mature forest carbon and structure largely align with observations across all three biome sites (Fig. 1). This is consistent with previous studies which have shown that DGVMs, even when nondemographic, are generally able to simulate realistic levels of biomass (Forkel *et al.*, 2016; Pugh *et al.*, 2020), although regional biases exist (Forkel *et al.*, 2019).

Yet, analysis of mature forest growth and mortality rates (Fig. 5) shows that similar biomass pools across models result from compensatory interactions between gross growth fluxes and turnover times, which also diverge significantly between models:

Most VDMs overestimated growth rates (six to seven models above the observed median, site-dependent) and, consistent with Pugh *et al.* (2020), underestimated carbon turnover times (six to eight models below the observed median turnover, though often within the observed variability). This accelerated cycling of carbon through wood in the models, compared with the real world, is potentially a result of not properly representing environmental constraints on woody growth (Delpierre *et al.*, 2016; Friend *et al.*, 2019), assumptions on allometry and allocation (Wolf *et al.*, 2011; Massoud *et al.*, 2019), or other poorly constrained demands on reproductive processes or root exudates, the latter found to make up 9–14% of global gross primary productivity (Chari *et al.*, 2024).

The general character of the forest structure (carbon and stem number distribution across size classes) is captured by the model ensemble with some individual model scatter (Fig. 4), although the log scale may obscure discrepancies. Some of the differences likely result from a lack of information on site history, as forest structure at the plot scale is strongly dependent on the management and gap–disturbance history at the observational sites. Here, the absence of forest management more than 200 yr ago cannot be guaranteed in the observations at FIN. If some selective logging did occur in the past, this could explain why we found the presence of large trees in the models but not in the observations. Averaging across larger areas typically dilutes the impact of any given historical disturbance on the overall statistics. However, it is possible that the observed total areas at FIN (56 plots, 8.88 ha) and BIA (5 plots, 44 ha) are too small relative to the typical size and frequency of disturbances to provide a general characterisation of the ecosystem. This is difficult to assess as the disturbance rates that would exist in the absence of management in European forests are not well known (e.g. Pugh *et al.*, 2024).

Notably, most models were able to simulate large trees (with the exception of BiomeE, BiomeEP and FATES in FIN and BIA), which have a very large influence on the total biomass simulated. Given the lack of available observations on the dynamic rates of large trees, particularly for mortality (Coomes & Allen, 2007; Gora & Esquivel-Muelbert, 2021), there is a particularly high potential for error in large tree mortality rates in VDMs.

Summary of results and outlook Overall, the benchmarking highlights that VDMs generally replicate carbon pools and forest structure in mature forests reasonably well. They also replicate (often emergent) processes, such as self-thinning and demonstrate plausible interactions between forest structural variables and carbon fluxes during regrowth phases. However, there are uncertainties in the duration of different stand development phases and divergence of growth rates and turnover time from observations. In this study, growth and mortality rate data were withheld from the modellers; these rates, if properly constrained, could propagate into addressing other issues, such as the length of the self-thinning period and the rate of $C_{\text{wood,AG}}$ recovery. We suggest that focusing on leveraging growth and mortality rate data to calibrate the models is a ‘low-hanging fruit’ that is likely to provide major improvements in model performance.

Assessment of available Benchmarks

In the benchmarks used in this study, there is variation in data suitability across variables from different biomes and stages of forest development, which consequently impacts the assessment of model performance (Seiler *et al.*, 2022). Similarly, model output outside the observed variability can result from a misalignment between the scale of the model and the scale at which observations were made (Fritsch *et al.*, 2020; Knapp *et al.*, 2022). Therefore, the quality of the benchmarks themselves and the scales at which they are compared must be treated with care. Below, we further address several aspects to which attention needs to be paid in the formulation of such benchmarks, both in terms of sampling and in terms of modelling. Generally, the benchmark data have to be collected, aggregated, down- or upscaled, and documented such that they can be meaningfully used for process validation.

Similarly, models have a range of different conceptual scale assumptions, meaning that individual VDMs or groups of VDMs each may need their own tailored benchmarks. At the same time, there are pragmatic trade-offs to be made between the time invested in complicated simulation setups and output procedures and realistically running and obtaining output for benchmarking efforts like this. Benchmarks need to be fit for purpose, but not necessarily perfect.

Trajectories through time The chronosequence data used for regrowth benchmarking in this study were only able to provide a very loose constraint on stand regrowth rates. Continental (BCI + BIA) to regional (FIN) scale data were used to construct chronosequences (Fig. S2), which integrates not only over stand regrowth stages but also over substantial environmental variability. Datasets for the temporal benchmarking of forest carbon pools are relatively numerous, but are normally derived from chronosequence data from multiple sites, which do not allow for an accurate determination of the time to recovery (Walker *et al.*, 2010). The effects of environmental variability could be mitigated against by simulating each plot individually using site-specific environmental driving data. This was, however, not practical to realise across many models in this study. Our sites’ mean annual temperatures are, however, generally representative of the regrowth benchmark population, with BIA and BCI within typical ranges and FIN slightly above the benchmark median, suggesting above-median regrowth dynamics (Methods S8; Fig. S20).

An alternative to chronosequences is comparisons made against longitudinal data from individual sites. Although we are aware of relatively few such plot-level datasets and they rarely extend beyond a few decades of observations, there are enough to create a benchmark at least in the tropics (Longworth *et al.*, 2014; Mesquita *et al.*, 2015; Rozendaal *et al.*, 2017); 2ndFOR collaborative research network, Tropical Managed Forest Observatory. These datasets come with the substantial benefit that they can often also provide growth and mortality rates, as well as overall net biomass increment, further constraining dynamics. Satellite-derived chronosequences (Heinrich *et al.*, 2021) can support but not

replace these plot-based benchmarks because of the importance of resolving growth and mortality dynamics and composition change.

Spatial scale The range in the mature forest aboveground woody carbon content at BIA (Fig. 1b) is derived from a limited number of plots ($n = 5$, with a total combined size of 44 ha), which are area-based averages of aboveground woody carbon content. The variability in biomass is a function of the area over which the individual data points are summed, with larger areas dampening the variability resulting from what are essentially stochastic processes. If model results representing one scale are compared with observational results derived from another scale, then the comparison is likely to be flawed. This will particularly be the case when variances (Knapp *et al.*, 2022), as opposed to the means, are compared. For future benchmarks, the scaling of observed forest variables will be both useful and necessary for making more appropriate use of small-scale sites, but transfer functions (Knapp *et al.*, 2022) at a meaningful sampling frequency level must be derived for more ecosystems than the tropics. Ideally, the combination of spatial scale and the time period observed (Urban *et al.*, 1987) needs to be large enough to provide a representative sample of the dynamics. For example, if a large fraction of the overall mortality derives from relatively rare events that kill many trees at the same location, such as blowdowns, then the observations and model outputs need to have sufficient spatial and temporal extent to characterise the return periods of these events. Because mortality regimes are regionally specific, the necessary temporal and spatial extent will therefore vary by forest region and vegetation type (Reilly *et al.*, 2016).

Suggestions for additional benchmarks

While this study has highlighted that process uncertainty is highest during establishment and regrowth, we cannot provide enough data constraints to address these deficiencies with confidence. To better validate demographic model process behaviour requires site-level data that are higher resolved (in time), diverse (in the variables measured, covering carbon pools and fluxes and structure), and at a process-relevant scale to adequately validate the models. We have identified a set of observations (Table 3) which would be desirable for process validation and present the state of data availability (to our knowledge).

Initial establishment conditions Benchmarks on seedling survival and regeneration (increase in a size class vs death) rates are useful approaches to evaluating initial conditions of the models, one of the processes that would benefit from further developments. Díaz-Yáñez *et al.* (2024) conducted a comprehensive intercomparison study on modelled forest recovery, revealing that models which overestimate recruitment compared with observations tend to show the highest mortality in small size classes, exceeding values seen in data. This suggests compensatory mechanisms at play, and while eventually stabilising, this would lead to artificial simulated turnover during the early stages of forest recovery. However, while short-term dynamics are critical, the

initial seedling establishment conditions also have an impact on long-term processes. Fisher *et al.* (2010) found that, among five demography-related parameters, sapling mortality had the greatest impact on community structure and ecosystem properties, highlighting the long-term influence of seedling survival.

Recruitment data along an environmental gradient across Europe are now available through the EuFoRia network (Díaz-Yáñez *et al.*, 2024), but such data are typically scattered across individual studies and often not compiled for benchmarking. Compilations of observations on seedling mortality, regrowth and small tree allometry (particularly crown area and bole height) across diverse ecosystems are urgently needed to benchmark initial conditions and their impact on later stages of forest recovery. This data would not only aid benchmarking, it would also be highly valuable to inform process development of seedling survival, especially at larger scales.

Forest recovery In the absence of regrowth observations for this study, we generated qualitative benchmarks ('forest recovery phases') containing a collection of plausibility criteria to reflect forest recovery-phase behaviour. While this may be a useful method to overcome data scarcity, it would be preferable to enhance these benchmarks with observations on ecosystem-specific boundary-values between the phases (if they exist). Two options for improving both our definitions of the forest phase benchmarks and reducing data scarcity are to engage satellite sequences and on-the-ground ecological field surveys. These two approaches could be pursued independently or, ideally, in tandem. Improved satellite spectral analysis on recovering forests (Mandl *et al.*, 2024) can provide information on canopy openness, and quick questionnaires for local ecologists can provide further information to help with determining a given forest's phase.

Forest structure While forest structure data from inventory plots with stand age exists, it is often not organised to specifically capture early recovery dynamics. Efforts to collate and structure such data would greatly enhance our ability to benchmark and constrain recovery-phase model processes with accuracy. Regrowth is especially crucial to resolve correctly in a time when many forests are in disequilibrium and our understanding of the future capacity of these young forests to gain and retain carbon needs improving.

Organising other benchmark variables into size classes could allow for deducing or constraining general mortality behaviour patterns across ecosystems (e.g. Pioniot *et al.*, 2022) or, for example, size-specific mortality reactions to water availability (Bennett *et al.*, 2015). Such data are altogether absent (with some few exceptions, see Table 3) for many types of forests.

Self-thinning lines integrate forest structure by showing the interplay between mass and number density within a forest and are therefore useful observations for constraining demographic models. Beyond self-thinning, general forest structure features, such as number and mass–density relationships exist between ecosystems, with the variability linked in part to rainfall seasonality (Yu *et al.*, 2024). These findings suggest that these

Table 3 Benchmarking variables useful for demographic model benchmarking.

Forest state/time period	Benchmark (1 = priority; 2 = desirable)	Process/state benchmarked	Benchmark variable (units)	Status	Requirements	Example studies/projects found (Ecosystem)	Process to be constrained
Recruitment	1	Recruitment	Seedling survival (% survival) seedling allometry: dbh (cm), height distribution (m)	Available	Multiple environmental gradients and ecosystems, Land-use history	EuFoRia (Díaz-Yáñez <i>et al.</i> , 2024) ForestGEO, https://hubbardbrook.org/data-catalog/	To determine trajectory of the ecosystem recovery speed, necessary, especially with increased disturbance/land-use change and at policy-relevant timescale.
Regrowth	2	Structure	Carbon content per size classes (kgC m ⁻²)	Exists (as growth increment rates exist) but not processed suitably	Multiple environmental gradients and ecosystems, Land-use history	Chave <i>et al.</i> (2020)	To constrain competitive interactions as forest regrows. Necessary to get correct, as it determines forest resilience, too
Regrowth	1	Dynamics	Growth increment rates (kgC m ⁻² yr ⁻¹)	Available: Few plot-level studies exist. Exists: Many satellite products exist, but the observations would need postprocessing from Benchmark 'Regrowth Pool Forest carbon content'	10–20yr intervals, finer at the beginning, depending on ecosystem productivity (e.g. Chave <i>et al.</i> , 2020), different units need conversion. Multiple environmental gradients and ecosystems, Land-use history. Need to standardise: Different definitions of 'biomass' for growth rates. Some used volume, others biomass	Plot level: Pretzsch <i>et al.</i> (2023) (temperate forests), Chave <i>et al.</i> (2020) (tropics, size class), Rozendaal & Chazdon (2015) (tropics), Chazdon (2021) (tropics)	To assess recovery sensitivity to climate and resource availability, as growth increment rates integrate forest structure and competition, they can to some extent replace these datasets. To constrain physiological processes related to woody biomass increment, using multi-site data
Regrowth	1	Structure	Number of stems per size class (n ha ⁻¹)	Exists	Described in the previous section	Mesquita <i>et al.</i> (2015) (Total number of stems only)	To constrain productivity, recruitment, competition and allometric assumptions
Regrowth	2	Trait dominance	Progression of dominant traits over time	Exists, numerous ecological observational studies.	Described in the previous section	Mesquita <i>et al.</i> (2015); Rozendaal <i>et al.</i> (2017)	This information could be leveraged to benchmark against number or diversity of functional traits present during the course of recovery, and how that relates to simulated forest resilience. To constrain the presence and (relative) number of early and late-successional traits, also in different environmental conditions

Table 3 (Continued)

Forest state/time period	Benchmark (1 = priority; 2 = desirable)	Process/state benchmarked	Benchmark variable (units)	Status	Requirements	Example studies/projects found (Ecosystem)	Process to be constrained
Regrowth	1	Pool	Forest carbon content (kgC m^{-2})	Available	Limited time (Nonchronosequence) series length available < 30 yr stand age	(1) 2ndFor (2) Cook-Patton <i>et al.</i> (2020), (3) Palviainen <i>et al.</i> (2020), this study. Need clear method/equation of how carbon content was estimated. Need age of the forest. (4) Satellite data: Heinrich <i>et al.</i> (2021)	A high-level variable to constrain forest recovery speed.
Regrowth	1	Dynamics	Mortality rates ($\text{kgC m}^{-2} \text{yr}^{-1}$, %/yr)	Available	10–20 yr intervals, finer at the beginning, depending on ecosystem productivity Desirable: by size class	Rozendaal & Chazdon (2015); Chazdon (2021)	To assess recovery sensitivity to climate and resource availability, as mortality rates integrate forest structure and competition, they can to some extent replace these datasets
Regrowth	1	Dynamics	Open-canopy phase – grass competition – duration (yr)	Required: Ecological survey design and data collection. Exists: satellite data	Created from a multitude of variables thought to act together to classify phases, qualitative combined use of: %Cveg, $n\text{stems}_{\text{CA}}$	Ecological surveys and satellite data for example Mandl <i>et al.</i> (2024) (satellite data for Europe)	All four phases are used to constrain carbon and ecological dynamics (growth rates, self-thinning rates) and successional trajectories, where simulated, and using multiple variables together, even if qualitatively, will get a link to different types of resilience. Vital data in times of more frequent disturbances
Regrowth	1	Dynamics	Open-canopy phase – growth duration (yr)	Available	Qualitative combined use of $n\text{stems}$, %Cveg, $\text{cmort}_{\text{rate}}$ (% yr^{-1}), CA	Described in the previous section	Described in the previous section
Regrowth	1	Dynamics	Closed-canopy self-thinning duration (yr)	Available	Qualitative combined use of $\text{cmort}_{\text{rate}}$ (% yr^{-1}), $n\text{stems}_{\text{CA}}$	Described in the previous section	Described in the previous section
Regrowth	1	Dynamics	Closed-canopy self-thinning duration (yr)	Exists to some extent in the tropics, but not postprocessed	Quantitative: postprocessed from mass and number density and growth rates.	Chazdon (2021)	To benchmark mortality processes and rates. A critical duration of carbon loss which cannot be explicitly simulated by nondemographic models

Table 3 (Continued)

Forest state/time period	Benchmark (1 = priority; 2 = desirable)	Process/state benchmarked	Benchmark variable (units)	Status	Requirements	Example studies/projects found (Ecosystem)	Process to be constrained
Regrowth	1	Dynamics	Closed-canopy late-successional phase (yr)	Required: Ecological survey design and data collection. Exists: satellite data Available	Qualitative combined use of %Cveg, cmort _{late} (% yr ⁻¹)	See qualitative phases above	See qualitative phases above
Mature forest	1	Pool, structure	Carbon content, per size classes (kgC m ⁻²)	Available	Need clear methods+ allometric equations for how carbon content was estimated	This study; NFI-data; https://www.remoteforests.org/ (Bin <i>et al.</i> , 2012)	A high-level variable to constrain forest carbon boundary, but most meaningful when used in context with forest structure (cwood_size / nstem_size) To constrain biomass dynamics, age-related productivity changes, ecosystem stability in context with the below mortality rates
Mature forest	2	Dynamics	Growth rates (kgC m ⁻² yr ⁻¹)	Available	Ideally also by dbh-size class	This study, https://forestsplots.net/ NFI-data example: https://hubbardbrook.org/data-catalog/	To constrain age-related and competition-related mortality and overall forest turnover.
Mature forest	1	Dynamics	Mortality rates (kgC m ⁻² yr ⁻¹)	Available	Ideally also by dbh-size class	This study, Bennett <i>et al.</i> (2015); Píponiot <i>et al.</i> (2022)	To constrain life-trait and successional trait parameterisation
Mature forest	2	Trait dominance, coexistence	Dominant traits	Available	Stand age information or metadata information on the forest having reached closed-canopy late-successional phase (further above in this table)	Kattge <i>et al.</i> (2020); Rüger <i>et al.</i> (2020)	To constrain life-trait and successional trait parameterisation
Regrowth / recovery after different disturbance agents	1	Combination of all	Collection of many of the above	Available	Different types, scales and severities of disturbance agents will impact forest structure and subsequent dynamics differently, temperate forest only	Defoliation (FoRTE experiment; Atkins <i>et al.</i> , 2023) Girdling (accelerated succession experiment; Gough <i>et al.</i> , 2021) Fire/harvest chronosequences (Gough <i>et al.</i> , 2007)	With increasing disturbance severity and frequency, these disturbance manipulation experiments and datasets are unique, and give us insights into DVM's realism in recovery behaviour. Useful for comparing against a variety of variables, structural, fluxes and traits

Table 3 (Continued)

Forest state/time period	Benchmark (1 = priority; 2 = desirable)	Process/state benchmark	Benchmark variable (units)	Status	Requirements	Example studies/projects found (Ecosystem)	Process to be constrained
Responses to Environmental change, environmental limits, canopy status.	1	Dynamics	Normalised ring width index	Available	Some postprocessing and method-development on how it can be compared against model output will be necessary.	International Tree-Ring Database: https://www.frames.gov/catalog/56 Dendroecological network: https://www.uvm.edu/femc/dendro African Tree Ring network: https://pastglobalchanges.org/science/wg/atrn/intro Tropical tree-ring network: https://tropicaltreeringnetwork.org	To benchmark demographic models' physiological growth response to environmental change. Potentially also useful to distinguish canopy and subcanopy growth responses to environmental change
Actual forest state and dynamics	1	Combination of all	Collection of many of the above	Exists	Summaries at a grid level for comparison with large-scale VDM simulations. Requires that the same of observations is representative of forest in that grid cell	National Forest Inventory plots. Combinations of research plots and systematic forest age maps	The overall ability of the model to capture the actual forest state and dynamics, combining across all processes and input datasets is a fundamental characteristic to assess

Benchmarks can be categorised as necessary or desirable/optional for demographic models. The benchmarks can guide future development and can therefore further be divided into (1) priority for this generation of VDMs, and (2) desirable to further develop VDMs. Status column specifications: Exists: studies have been done, but time investment into collecting and cataloguing the studies and collective formatting is required. Available: where networks have been creating a common data standard, which modellers could benefit from, or studies have collected multiple studies' data. Some of these networks are not openly accessible and data agreements are needed, this is not distinguished here. Required: In need of collection.

relationships can be used to constrain not only competition for light but also for water. Self-thinning lines or older-growth mass–density relationships would be valuable tools for model verification. Longitudinal self-thinning (while also keeping track of other mortality causes) throughout the whole forest recovery trajectory would be a useful dataset for model constraint, as it would implicitly integrate forest structure observations with forest growth rates. Active remote sensing also offers new opportunities to benchmark forest structure in VDMs (Drake *et al.*, 2002; Dubayah *et al.*, 2010, 2020). Airborne lidar data on vegetation 3-D structure have been powerfully used together with the Ecosystem Demography model in a range of studies from local to regional scales (Thomas *et al.*, 2008; Dubayah *et al.*, 2010; Antonarakis *et al.*, 2011; Hurtt *et al.*, 2016, 2019, 2024; Ma *et al.*, 2021). Recently, these same benefits have been demonstrated globally using spaceborne lidar (Dubayah *et al.*, 2022; Ma *et al.*, 2022, 2023).

Growth interannual variability Tree-ring data, though abundant, has not yet been systematically used to benchmark VDM woody biomass growth responses to environmental conditions, with few site-specific, single-model exceptions (e.g. Jeong *et al.*, 2021; Xu *et al.*, 2024). Tree-ring data reflect both individual-tree growth and collective biomass changes across plots in response to the environment. Forest structure and carbon result from the balance between biomass loss through mortality and accumulation through growth, with the latter being both a physiological response and demographic response that serves as a key driver in ecosystem dynamics. While reusing samples that have been taken for climate reconstruction is feasible, it requires working around their sample biases (Jeong *et al.*, 2021), and networks are emerging that target the use of tree rings in ecological studies (DendroEcological Network (DEN), Tropical tree-ring network (TTR), African Tree Ring Network (ATRN), REMOTE forests).

Other benchmarks Besides structural information, species data, while not directly comparable to large-scale simulations and PFT abstraction in global VDMs (Table 3, Regrowth – ‘Trait dominance’), are also required to determine whether models can correctly represent succession dynamics, something this study has not covered. Achieving successional dynamics and coexistence in VDMs is still challenging and would be a valuable next step in benchmarking. Individual studies on species abundance (which could be abstracted through PFT or functional trait abundance in VDMs) and resilience on abandoned farmland exist (e.g. in the Tropics; Mesquita *et al.*, 2015). However, knowledge of typical successional dynamics, including the timescales on which forests typically transition from one species to another, is often well established regionally, at least for northern-hemisphere temperate and boreal forests (Van Cleve & Viereck, 1981; Shorohova *et al.*, 2009; Taylor *et al.*, 2020). This could be assembled through the literature review or surveys of local experts.

In addition to land-use and forest management, forest structure is also influenced by a diversity of disturbance events (Table 3, ‘Regrowth/recovery after different disturbance agents’),

expected to increase in frequency and severity in the future (Seidl *et al.*, 2017; McDowell *et al.*, 2020), where different agents (bark beetle, wind, fire) leave behind different forest structures, and therefore recovery trajectories and levels of resilience. Large-scale disturbance manipulation studies, with a comprehensive suite of benchmark variables measured, would provide an additional valuable benchmark for this (e.g. FoRTE project; Atkins *et al.*, 2023).

Benchmarking against the overall state of global forests The combination of all variables (biomass, forest structure, mortality and harvest rates, growth rates, as realised with all legacies, in response to environmental change) into a coherent constraining system should be the ultimate goal of a benchmarking effort for VDMs.

High-quality constraints of forest dynamics and structure require that the benchmark suite not only isolates behaviour during particular phases of stand development and management, but also covers the actual state of the forest. This requires benchmarking data that are representative of the current state of the forest. Available forest plot data broadly fall into two categories: research plots for which there is often fairly detailed information on management or its absence, and National Forest Inventory (NFI) plots for which management information is often very limited or absent.

While the limited meta-information on NFI plots makes them challenging to use to isolate particular aspects of forest behaviour, their systematic sampling across the forest makes them very well suited to benchmark the actual forest structure and dynamics that have emerged from the combination of ecology, management and land-use history. Making use of this constraint, however, requires that VDMs can simulate not only demography but also land-use change and forest management, and is naturally not only a test of VDMs themselves but also the input datasets used. Research plots are well suited to isolate behaviour, although those plots which are suitable to derive dynamics tend to focus on mature or old-growth forests. It is crucial that observational data are appropriately selected to benchmark the feature of interest. To gain the spatial representativeness needed to benchmark VDMs effectively, large compilations of plots are needed. These are challenging to assemble both for reasons of diverse data ownership and also because of the diverse sampling protocols used. Emerging efforts to analyse forest dynamics at continental/biome scales are making benchmark assembly increasingly feasible (Hubau *et al.*, 2020; Astigarraga *et al.*, 2024; Bialic-Murphy *et al.*, 2024; EuForIA network). To make maximal leverage of these efforts to improve VDMs, modellers must be closely involved in their analysis, requiring the development of close collaborations across disciplines.

Ultimately, the choice of benchmarks depends on the processes studied. However, establishing a standardised minimum set of benchmarks to validate fundamental demographic and ecological principles, as well as carbon stocks and fluxes, would provide a crucial baseline for ensuring the robustness and credibility of VDM predictions. Implementing such benchmarks as routine checks should become standard practice in the field.

Standardisation of benchmark definitions (including metadata, units and data-formatting rules) is a work in progress (see Table S6 for observations), and needs to be paired conceptually and technically with the correct output (table 3.1 in Notes S1 for model output), and requires further refinement in the future.

Call for benchmarking and implications

We have compared the performance, assumptions and outputs of nine VDMs. The models differ in their fundamental assumptions on how forests are conceptualised both in time and vertical space. Yet, all replicate observed forest dynamics and structure reasonably well, both during regrowth and mature forest stages at three sites, each representing a distinct biome. However, we find that development attention needs to be focused on (1) disturbance regeneration and recovery/forest regrowth, where models diverge the most, and (2) calibration to improve woody growth and mortality rates to determine turnover time. Finally, the performance of the models against a limited set of observations showcases the need for more in-depth benchmarking studies, standardised model evaluation, routine demographic carbon budget checks and harmonisation of definitions.

Models differed the most in the initial conditions and subsequent years, but since all models behave ecologically plausibly along a sequence of forest recovery phases, we believe that, once calibrated, these models have the structural capabilities to represent forest dynamics throughout the full regrowth trajectory directly from postestablishment through to old growth. By resolving forest structure, growth and mortality at the individual or cohort-level, VDMs typically introduce more parameters compared with nondemographic models, but because those parameters have direct 'points of contact' with real-world observations of individual trees and stands, there is substantial potential to tightly constrain their ranges. Our expectation is that with process representation at this fundamental level of process detail adequately constrained/calibrated, we can increase the realism in the simulated future vegetation dynamics. This advancement could help contribute to decreasing the persistent problem of vegetation carbon stock divergence among models in simulations under future climate change (Arora *et al.*, 2013, 2020; Friedlingstein *et al.*, 2014; Lovenduski & Bonan, 2017), provide assessments of the potential and permanence of nature-based solutions (e.g. Cook-Patton *et al.*, 2020), as well as our ability to resolve the current forest carbon sink (O'Sullivan *et al.*, 2024).

We proposed a list of benchmarking variables specific to demographic models that are necessary or desirable to comprehensively challenge our demographic models at all stages of forest development (Table 3). Future challenges will be to collate and format existing data to the processes and scales VDMs are designed to represent. Comprehensive datasets covering forest dynamics, pools and structure at all stages of forest development and biomes are needed. It is much easier to call for more data collection than it is to carry it out. We have identified some cases where additional collection would be valuable. But for the most part, suitable observations exist but are not always available to model developers, whether for reasons of being collated and

analysed in the form needed or because of data ownership. Close collaboration between groups making the observations and those developing VDMs will be required to bridge this gap.

Technical and conceptual modelling and postprocessing challenges pertained to the output variables themselves. During the course of the initiative, evolving definitions and outputs of variables and a need for defining a 'demographic' woody carbon balance (Methods S2) underscore the need for standards for variable definitions, formats and model testing. Enhanced formatting is likewise required if the benchmarking variables become components of land-surface suites, such as iLAMB. For example, frameworks and outputs will need to be modified/extended to accommodate the specific structure and dimensions of demographic data (e.g. the 'self-thinning' space). Our simulation protocol (Notes S1) can provide a basis for further efforts in this space.

This study is a first step towards a currently lacking demographic model benchmarking suite. Even a subset of these variables from our benchmarking suite (Table 3) will be an indispensable tool to evaluate and provide credibility to these increasingly policy-relevant models at the level of detail relevant for policy makers and for the simulation of future vegetation responses to climate and land-use change. Our combined efforts of the demographic modelling community in this study are a first step towards this. With improved calibration and postrecovery initialisation, VDMs are ready to support realistic assessments of afforestation and reforestation efforts, enhancing the credibility of policy-relevant short-term and long-term forest carbon and structure projections.

Acknowledgements

The Barro Colorado Island (BCI) forest dynamics research project was made possible by the National Science Foundation grants to Stephen P. Hubbell: DEB-0640386, DEB-0425651, DEB-0346488, DEB-0129874, DEB-00753102, DEB-9909347, DEB-9615226, DEB9615226, DEB-9405933, DEB-9221033, DEB-9100058, DEB8906869, DEB-8605042, DEB-8206992 and DEB-7922197, support from the Forest Global Earth Observatory, the Smithsonian Tropical Research Institute, the John D. and Catherine T. MacArthur Foundation, the Mellon Foundation, the Small World Institute Fund and numerous private individuals, and through the hard work of over 100 people from 10 countries over the past three decades. We also value the work by the field teams of the National Forest Inventory for collecting forest inventory data across Finnish forests for decades. The plot project is part of the Forest Global Earth Observatory (ForestGEO), a global network of large-scale demographic tree plots. The research presented in this paper is a contribution to the Strategic Research Area 'Biodiversity and Ecosystem Services in a Changing Climate', BECC, 'Modelling the Regional and Global Earth system', MERGE and the Nature-based Future Solutions profile area. EW acknowledges funding from NASA Modeling, Analysis and Prediction (MAP) Program (80NSSC21K1496). This research was supported by the Forest-Value programme, the European Commission, Vinnova, the

Swedish Energy Agency and Formas for the project FORECO. TAMP, AHES, MP, AL and SO acknowledge funding from the European Union's Horizon Europe Research and the Innovation Programme projects ForestPaths (101056755). TP, SO and KP acknowledge funding from the European Research Council project Tree2Globe (101141836), CLIMB-Forest (101059888) and AI4PEX (101137682). The ORCHIDEE simulations were performed using HPC resources from GENCI-TGCC on grants 2024-06328. SL and GM were funded by Horizon 2020, Holi-Soils (SEP-210673589) and Horizon Europe INFORMA (101060309). DZ was supported by the Swedish Research Council (VR; grant no.: 2024-05057) and the Swedish Strategic Research Area 'Biodiversity and Ecosystem services in a Changing Climate' (BECC; <https://www.lu.se>). The authors thank Jonna Åhlin for editing work on the manuscript. GCH and Lei Ma acknowledge the support of the NASA Carbon Monitoring System (80NSSC25K7221 and 80NSSC21K1059) and the NASA GEDI Competed Science Team (80NSSC24K0599). RF acknowledges the European Union's Horizon 2020 research and innovation programme (grant agreement no.: 101003687) (project PROVIDE, Paris Agreement Overshooting) and 101003536 (ESM2025 – Earth System Models for the Future). MGDK acknowledges funding from the UK Natural Environment Research Council (NE/W010003/1). JN and CDK were supported as part of the Next Generation Ecosystem Experiments – Tropics, funded by the US Department of Energy, Office of Science, Office of Biological and Environmental Research. PMC and RF acknowledge funding from the European Union's Horizon Europe Research and Innovation Programme, project NextGenCarbon (101184989). BB acknowledges funding from the General Directorate of State Forests in Poland (grant no.: EO.271.3.1.2019). APKA was funded by the Met Office Climate Science for Service Partnership (CSSP) Brazil project under the International Science Partnerships Fund (ISPF). BDS and Laura Marqués were funded by the Swiss National Science Foundation (grant no.: PCEFP2_181115). The authors thank two anonymous reviewers for their useful feedback on the manuscript.

Competing interests

None declared.

Author Contributions

TAMP and AHES conceived the initiative and initial simulation protocol and coordinated the intercomparison work. RAF, CDK, PMC, JN, APKA, SL, AEM, SO, JK, SS, DZ and MGDK refined the simulation protocol. JK (CABLE-POP), Lei Ma and GCH (EDv3), AHES and SO (LPJ-GUESS), Laura Marqués and BS (BiomeEP), EW (BiomeE), JN (FATES), APKA, PMC and JRM (JULES-RED), HS (SEIB-DGVM), SL and GM (ORCHIDEE) done model runs and output postprocessing into demographic benchmarking ('D-BEN') format. KP, DZ, MP, AL, BB and AEM performed the observational data preprocessing and advice. Laura Marqués performed the analysis code review. AHES coordinated the simulation output analysis and

wrote the first draft. All authors commented on the manuscript and accepted the final version.

ORCID

Arthur P. K. Argles  <https://orcid.org/0000-0002-1346-6518>
 Bogdan Brzeziecki  <https://orcid.org/0000-0001-5786-3075>
 Peter M. Cox  <https://orcid.org/0000-0002-0679-2219>
 Martin G. De Kauwe  <https://orcid.org/0000-0002-3399-9098>
 Annemarie H. Eckes-Shephard  <https://orcid.org/0000-0002-2453-3843>
 Adriane Esquivel-Muelbert  <https://orcid.org/0000-0001-5335-1259>
 Rosie A. Fisher  <https://orcid.org/0000-0003-3260-9227>
 George C. Hurtt  <https://orcid.org/0000-0001-7278-202X>
 Jürgen Knauer  <https://orcid.org/0000-0002-4947-7067>
 Charles D. Koven  <https://orcid.org/0000-0002-3367-0065>
 Aleksi Lehtonen  <https://orcid.org/0000-0003-1388-0388>
 Sebastiaan Luyssaert  <https://orcid.org/0000-0003-1121-1869>
 Lei Ma  <https://orcid.org/0000-0002-3959-4155>
 Laura Marqués  <https://orcid.org/0000-0002-3593-5557>
 Jonathan R. Moore  <https://orcid.org/0000-0002-7725-3343>
 Jessica F. Needham  <https://orcid.org/0000-0003-3653-3848>
 Stefan Olin  <https://orcid.org/0000-0002-8621-3300>
 Mikko Peltoniemi  <https://orcid.org/0000-0003-2028-6969>
 Thomas A. M. Pugh  <https://orcid.org/0000-0002-6242-7371>
 Hisashi Sato  <https://orcid.org/0000-0002-6510-4914>
 Stephen Sitch  <https://orcid.org/0000-0003-1821-8561>
 Benjamin D. Stocker  <https://orcid.org/0000-0003-2697-9096>
 Ensheng Weng  <https://orcid.org/0000-0002-1858-4847>
 Daniel Zuleta  <https://orcid.org/0000-0001-9832-6188>

Data availability

All analysis scripts, benchmarking data and model output data specific to this manuscript are deposited on [10.5281/zenodo.14415143](https://zenodo.org/record/14415143). Model versions and availability: LPJ-GUESS: v.4.1, branch mechanistic_mortality_dben_r11972 <https://zenodo.org/records/15837787> and postprocessing from https://github.com/teatree1212/DBEN_postprocess_LPJGUESstoDBEN, also deposited on <https://zenodo.org/records/15837636>. ORCHIDEE: This study made use of ORCHIDEE r8696 [10.5281/zenodo.15805716](https://zenodo.org/record/15805716). EDv3.0 is from <https://zenodo.org/records/6901510>. CABLE-POP (rev 9526): [10.5281/zenodo.17050747](https://zenodo.org/record/17050747). BiomeEP: model version from <https://github.com/lauramarques/rsfun>; model runs and postprocessing scripts from <https://github.com/geco-bern/DBEN>. FATES: model version from <https://github.com/JessicaNeedham/fates/tree/jfn-dben-api35-seedhist> and postprocessing scripts: <https://github.com/JessicaNeedham/dben-archive>. BiomeE: <https://github.com/wengensheng/BiomeESS>. SEIB-DGVM: 3.10 is from <https://seib-dgvm.com/en/>. JULES-RED: both the model branch, r24142_add_red_sci_vn1.1, and the rose suite used in the simulation and postprocessing, u-cq124, are available from

the Met Office Science Repository Service: access can be applied for via the link https://jules-lsm.github.io/access_req/JULES_access.html.

References

- Abramowitz G, Ukkola A, Hobeichi S, Cranko Page J, Lipson M, De Kauwe M, Green S, Brenner C, Frame J, Nearing G *et al.* 2024. On the predictability of turbulent fluxes from land: PLUMBER2 MIP experimental description and preliminary results. *EGU Sphere* 1: 1–47.
- Antonarakis AS, Saatchi SS, Chazdon RL, Moorcroft PR. 2011. Using Lidar and Radar measurements to constrain predictions of forest ecosystem structure and function. *Ecological Applications: A Publication of the Ecological Society of America* 21: 1120–1137.
- Argles APK, Moore JR, Huntingford C, Wiltshire AJ, Harper AB, Jones CD, Cox PM. 2020. Robust ecosystem demography (RED v.1.0): a parsimonious approach to modelling vegetation dynamics in Earth system models. *Geoscientific Model Development* 13: 4067–4089.
- Arora VK, Boer GJ, Friedlingstein P, Eby M, Jones CD, Christian JR, Bonan G, Bopp L, Brovkin V, Cadule P *et al.* 2013. Carbon–concentration and carbon–climate feedbacks in CMIP5 earth system models. *Journal of Climate* 26: 5289–5314.
- Arora VK, Katavouta A, Williams RG, Jones CD, Brovkin V, Friedlingstein P, Schwinger J, Bopp L, Boucher O, Cadule P *et al.* 2020. Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models. *Biogeosciences* 17: 4173–4222.
- Astigarraga J, Esquivel-Muelbert A, Ruiz-Benito P, Rodríguez-Sánchez F, Zavala MA, Vilà-Cabrera A, Schelhaas MJ, Kunstler G, Woodall CW, Cienciala E *et al.* 2024. Relative decline in density of Northern Hemisphere tree species in warm and arid regions of their climate niches. *Proceedings of the National Academy of Sciences of the United States of America* 121: e2314899121.
- Atkins JW, Shiklomanov A, Mathes KC, Bond-Lamberty B, Gough CM. 2023. Effects of forest structural and compositional change on forest microclimates across a gradient of disturbance severity. *Agricultural and Forest Meteorology* 339: 109566.
- Babst F, Bouriaud O, Poulter B, Trouet V, Girardin MP, Frank DC. 2019. Twentieth century redistribution in climatic drivers of global tree growth. *Science Advances* 5: eaat4313.
- Bennett AC, McDowell NG, Allen CD, Anderson-Teixeira KJ. 2015. Larger trees suffer most during drought in forests worldwide. *Nature Plants* 1: 1–5.
- Besnard S, Koirala S, Santoro M, Weber U, Nelson J, Günter J, Hérault B, Kassi J, N'Guessan A, Neigh C *et al.* 2021. Mapping global forest age from forest inventories, biomass and climate data. *Earth System Science Data* 13: 4881–4896.
- Bialic-Murphy L, McElderry RM, Esquivel-Muelbert A, van den Hoogen J, Zuidema PA, Phillips OL, de Oliveira EA, Loayza PA, Alvarez-Davila E, Alves LF *et al.* 2024. The pace of life for forest trees. *Science* 386: 92–98.
- Bin Y, Ye W, Muller-Landau HC, Wu L, Lian J, Cao H. 2012. Unimodal tree size distributions possibly result from relatively strong conservatism in intermediate size classes. *PLoS ONE* 7: e52596.
- Bonan GB. 2008. Forests and climate change: forcings, feedbacks, and the climate benefits of forests. *Science* 320: 1444–1449.
- Bonan GB, Doney SC. 2018. Climate, ecosystems, and planetary futures: the challenge to predict life in Earth system models. *Science* 359: eaam8328.
- Bonan GB, Lucier O, Coen DR, Foster AC, Shuman JK, Laguë MM, Swann ALS, Lombardozzi DL, Wieder WR, Dahlin KM *et al.* 2024. Reimagining earth in the earth system. *Journal of Advances in Modeling Earth Systems* 16: e2023MS004017.
- Brown MJ, Parker GG. 1994. Canopy light transmittance in a chronosequence of mixed-species deciduous forests. *Canadian Journal of Forest Research* 24: 1694–1703.
- Brzeziecki B, Pommerening A, Miścicki S, Drozdowski S, Żyburka H. 2016. A common lack of demographic equilibrium among tree species in Białowieża National Park (NE Poland): evidence from long-term plots. *Journal of Vegetation Science* 27: 460–469.
- Bugmann H, Bigler C. 2011. Will the CO₂ fertilization effect in forests be offset by reduced tree longevity? *Oecologia* 165: 533–544.
- Canham CD, McAninch JB, Wood DM. 1994. Effects of the frequency, timing, and intensity of simulated browsing on growth and mortality of tree seedlings. *Canadian Journal of Forest Research* 24: 817–825.
- Carnicer J, Coll M, Ninyerola M, Pons X, Sánchez G, Peñuelas J. 2011. Widespread crown condition decline, food web disruption, and amplified tree mortality with increased climate change-type drought. *Proceedings of the National Academy of Sciences, USA* 108: 1474–1478.
- Chari NR, Tumber-Dávila SJ, Phillips RP, Bauerle TL, Brunn M, Hafner BD, Klein T, Obersteiner S, Reay MK, Ullah S *et al.* 2024. Estimating the global root exudate carbon flux. *Biogeochemistry* 167: 895–908.
- Chave J, Piponiot C, Maréchaux I, De Foresta H, Larpin D, Fischer FJ, Derroire G, Vincent G, Hérault B. 2020. Slow rate of secondary forest carbon accumulation in the Guianas compared with the rest of the Neotropics. *Ecological Applications* 30: e20004.
- Chazdon R. 2021. Tree abundance in eight 1-ha tropical forest plots in northeastern Costa Rica from 1997–2017. *Dryad*. doi: 10.5061/dryad.ncjxksvr.
- Collier N, Hoffman FM, Lawrence DM, Keppel-Aleks G, Koven CD, Riley WJ, Mu M, Randerson JT. 2018. The international land model benchmarking (ILAMB) system: design, theory, and implementation. *Journal of Advances in Modeling Earth Systems* 10: 2731–2754.
- Condit R, Pérez R, Aguilar S, Lao S, Foster R, Hubbell S. 2019. Complete data from the Barro Colorado 50-ha plot: 423617 trees, 35 years: 573292682 bytes.
- Comeau PG, White M, Kerr G, Hale SE. 2010. Maximum density–size relationships for Sitka spruce and coastal Douglas-fir in Britain and Canada. *Forestry* 83: 461–468.
- Cook-Patton SC, Leavitt SM, Gibbs D, Harris NL, Lister K, Anderson-Teixeira KJ, Briggs RD, Chazdon RL, Crowther TW, Ellis PW *et al.* 2020. Mapping carbon accumulation potential from global natural forest regrowth. *Nature* 585: 545–550.
- Coomes DA, Allen RB. 2007. Mortality and tree-size distributions in natural mixed-age forests. *Journal of Ecology* 95: 27–40.
- da Costa ACL, Galbraith D, Almeida S, Portela BTT, da Costa M, de Athaydes Silva Junior J, Braga AP, de Gonçalves PHL, de Oliveira AA, Fisher R *et al.* 2010. Effect of 7 yr of experimental drought on vegetation dynamics and biomass storage of an eastern Amazonian rainforest. *New Phytologist* 187: 579–591.
- Cramer W, Bondeau A, Woodward FI, Prentice IC, Betts RA, Brovkin V, Cox PM, Fisher V, Foley JA, Friend AD *et al.* 2001. Global response of terrestrial ecosystem structure and function to CO₂ and climate change: results from six dynamic global vegetation models: ECOSYSTEM DYNAMICS, CO₂ and CLIMATE CHANGE. *Global Change Biology* 7: 357–373.
- Dalmonech D, Marano G, Amthor JS, Cescatti A, Lindner M, Trotta C, Collalti A. 2022. Feasibility of enhancing carbon sequestration and stock capacity in temperate and boreal European forests via changes to management regimes. *Agricultural and Forest Meteorology* 327: 109203.
- Davies SJ, Abiem I, Abu Salim K, Aguilar S, Allen D, Alonso A, Anderson-Teixeira K, Andrade A, Arellano G, Ashton PS *et al.* 2021. ForestGEO: Understanding forest diversity and dynamics through a global observatory network. *Biological Conservation* 253: 108907.
- Delpierre N, Berveiller D, Granda E, Dufrene E. 2016. Wood phenology, not carbon input, controls the interannual variability of wood growth in a temperate oak forest. *New Phytologist* 210: 459–470.
- Díaz-Yáñez O, Käber Y, Anders T, Bohn F, Braziunas KH, Brūna J, Fischer R, Fischer SM, Hetzer J, Hickler T *et al.* 2024. Tree regeneration in models of forest dynamics: a key priority for further research. *Ecosphere* 15: e4807.
- Döscher R, Acosta M, Alessandri A, Anthoni P, Arsouze T, Bergman T, Bernardello R, Boussetta S, Caron L-P, Carver G *et al.* 2022. The EC-Earth3 Earth system model for the coupled model intercomparison project 6. *Geoscientific Model Development* 15: 2973–3020.
- Drake JB, Dubayah RO, Clark DB, Knox RG, Blair JB, Hofton MA, Chazdon RL, Weishampel JF, Prince S. 2002. Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sensing of Environment* 79: 305–319.

- Dubayah R, Armston J, Healey SP, Bruening JM, Patterson PL, Kellner JR, Duncanson L, Saarela S, Ståhl G, Yang Z *et al.* 2022. GEDI launches a new era of biomass inference from space. *Environmental Research Letters* 17: 095001.
- Dubayah R, Blair JB, Goetz S, Fatoyinbo L, Hansen M, Healey S, Hofton M, Hurtt G, Kellner J, Luthcke S *et al.* 2020. The global ecosystem dynamics investigation: high-resolution laser ranging of the earth's forests and topography. *Science of Remote Sensing* 1: 100002.
- Dubayah RO, Sheldon SL, Clark DB, Hofton MA, Blair JB, Hurtt GC, Chazdon RL. 2010. Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. *Journal of Geophysical Research – Biogeosciences* 115: 1306.
- Enquist BJ, Brown JH, West GB. 1998. Allometric scaling of plant energetics and population density. *Nature* 395: 163–165.
- Eyring V, Righi M, Lauer A, Evaldsson M, Wenzel S, Jones C, Anav A, Andrews O, Cionni I, Davin EL *et al.* 2016. ESMVALTOOL (v.1.0) – a community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP. *Geoscientific Model Development* 9: 1747–1802.
- Fisher R, McDowell N, Purves D, Moorcroft P, Sitch S, Cox P, Huntingford C, Meir P, Woodward FI. 2010. Assessing uncertainties in a second-generation dynamic vegetation model caused by ecological scale limitations. *New Phytologist* 187: 666–681.
- Fisher RA, Koven CD. 2020. Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. *Journal of Advances in Modeling Earth Systems* 12: e2018MS001453.
- Fisher RA, Koven CD, Anderegg WRL, Christoffersen BO, Dietze MC, Farrior CE, Holm JA, Hurtt GC, Knox RG, Lawrence PJ *et al.* 2018. Vegetation demographics in earth system models: a review of progress and priorities. *Global Change Biology* 24: 35–54.
- Fisher RA, Muszala S, Versteinstein M, Lawrence P, Xu C, McDowell NG, Knox RG, Koven C, Holm J, Rogers BM *et al.* 2015. Taking off the training wheels: the properties of a dynamic vegetation model without climate envelopes, CLM4.5(ED). *Geoscientific Model Development* 8: 3593–3619.
- Forkel M, Carvalhais N, Rödenbeck C, Keeling R, Heimann M, Thonicke K, Zaehle S, Reichstein M. 2016. Enhanced seasonal CO₂ exchange caused by amplified plant productivity in northern ecosystems. *Science* 351: 696–699.
- Forkel M, Drüke M, Thurner M, Dorigo W, Schaphoff S, Thonicke K, von Bloh W, Carvalhais N. 2019. Constraining modelled global vegetation dynamics and carbon turnover using multiple satellite observations. *Scientific Reports* 9: 18757.
- Fridley JD, Wright JP. 2018. Temperature accelerates the rate fields become forests. *Proceedings of the National Academy of Sciences, USA* 115: 4702–4706.
- Friedlingstein P, Meinshausen M, Arora VK, Jones CD, Anav A, Liddicoat SK, Knutti R. 2014. Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *Journal of Climate* 27: 511–526.
- Friend AD, Eckes-Shepherd AH, Fonti P, Rademacher TT, Rathgeber CBK, Richardson AD, Turton RH. 2019. On the need to consider wood formation processes in global vegetation models and a suggested approach. *Annals of Forest Science* 76: 14.
- Friend AD, Lucht W, Rademacher TT, Keribin R, Betts R, Cadule P, Ciais P, Clark DB, Dankers R, Falloon PD *et al.* 2014. Carbon residence time dominates uncertainty in terrestrial vegetation responses to future climate and atmospheric CO₂. *Proceedings of the National Academy of Sciences, USA* 111: 3280–3285.
- Friend AD, Stevens AK, Knox RG, Cannell MGR. 1997. A process-based, terrestrial biosphere model of ecosystem dynamics (hybrid v.3.0). *Ecological Modelling* 95: 249–287.
- Fritsch M, Lischke H, Meyer KM. 2020. Scaling methods in ecological modelling. *Methods in Ecology and Evolution* 11: 1368–1378.
- Gleason HA. 1917. The structure and development of the plant association. *Bulletin of the Torrey Botanical Club* 44: 463–481.
- Gora EM, Esquivel-Muelbert A. 2021. Implications of size-dependent tree mortality for tropical forest carbon dynamics. *Nature Plants* 7: 384–391.
- Gough CM, Bohrer G, Hardiman BS, Nave LE, Vogel CS, Atkins JW, Bond-Lamberty B, Fahey RT, Fotis AT, Grigri MS *et al.* 2021. Disturbance-accelerated succession increases the production of a temperate forest. *Ecological Applications* 31: e02417.
- Gough CM, Vogel CS, Harrold KH, George K, Curtis PS. 2007. The legacy of harvest and fire on ecosystem carbon storage in a north temperate forest. *Global Change Biology* 13: 1935–1949.
- Grelle CEV. 2003. Forest structure and vertical stratification of small mammals in a Secondary Atlantic Forest, Southeastern Brazil. *Studies on Neotropical Fauna and Environment* 38: 81–85.
- Haverd V, Smith B, Nieradzik L, Briggs PR, Woodgate W, Trudinger CM, Canadell JG, Cuntz M. 2018. A new version of the CABLE land surface model (Subversion revision r4601) incorporating land use and land cover change, woody vegetation demography, and a novel optimisation-based approach to plant coordination of photosynthesis. *Geoscientific Model Development* 11: 2995–3026.
- Heinrich VHA, Dalagnol R, Cassol HLG, Rosan TM, de Almeida CT, Silva Junior CHL, Campanharo WA, House JI, Sitch S, Hales TC *et al.* 2021. Large carbon sink potential of secondary forests in the Brazilian Amazon to mitigate climate change. *Nature Communications* 12: 1785.
- Hubau W, Lewis SL, Phillips OL, Affum-Baffoe K, Beekman H, Cuní-Sanchez A, Daniels AK, Ewango CEN, Fauser S, Mukinzi JM *et al.* 2020. Asynchronous carbon sink saturation in African and Amazonian tropical forests. *Nature* 579: 80–87.
- Hurtt G, Ma L, Lamb R, Campbell E, Dubayah R, Hansen M, Huang C, Leslie-Bole H, Lister A, Lu J *et al.* 2024. Beyond MRV: combining remote sensing and ecosystem modeling for geospatial monitoring and attribution of forest carbon fluxes over Maryland, USA. *Environmental Research Letters* 2: 19.
- Hurtt G, Zhao M, Sahajpal R, Armstrong A, Birdsey R, Campbell E, Dolan K, Dubayah R, Fisk JP, Flanagan S *et al.* 2019. Beyond MRV: high-resolution forest carbon modeling for climate mitigation planning over Maryland, USA. *Environmental Research Letters* 14: 45013.
- Hurtt GC, Chini L, Sahajpal R, Frolking S, Bodirsky BL, Calvin K, Doelman JC, Fisk J, Fujimori S, Klein Goldewijk K *et al.* 2020. Harmonization of global land use change and management for the period 850–2100 (LUH2) for CMIP6. *Geoscientific Model Development* 13: 5425–5464.
- Hurtt GC, Chini LP, Frolking S, Betts RA, Fischer G, Hibbard K, Houghton RA, Kindermann G, Riahi K. 2011. Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Climatic Change* 109: 117–161.
- Hurtt GC, Pacala SW, Moorcroft PR, Caspersen J, Shevliakova E, Houghton RA, Moore B. 2002. Projecting the future of the U.S. carbon sink. *Proceedings of the National Academy of Sciences, USA* 99: 1389–1394.
- Hurtt GC, Thomas RQ, Fisk JP, Dubayah RO, Sheldon SL. 2016. The impact of fine-scale disturbances on the predictability of vegetation dynamics and carbon flux. *PLoS ONE* 11: e0152883.
- Jeong J, Barichivsky J, Peylin P, Haverd V, McGrath M, Vuichard N, Evans M, Babst F, Luyssaert S. 2021. Using the International Tree-Ring Data Bank (ITRDB) records as century-long benchmarks for global land-surface models. *Geoscientific Model Development* 14: 5891–5913.
- Johnson DJ, Needham J, Xu C, Massoud EC, Davies SJ, Anderson-Teixeira KJ, Bunyavechewin S, Chambers JQ, Chang-Yang C-H, Chiang J-M *et al.* 2018. Climate sensitive size-dependent survival in tropical trees. *Nature Ecology & Evolution* 2: 1436–1442.
- Kattge J, Bönsch G, Díaz S, Lavorel S, Prentice IC, Leadley P, Tautenhahn S, Werner GDA, Aakala T, Abedi M *et al.* 2020. TRY plant trait database – enhanced coverage and open access. *Global Change Biology* 26: 119–188.
- Klopčič M, Boncina A. 2012. Recruitment of tree species in mixed selection and irregular shelterwood forest stands. *Annals of Forest Science* 69: 915–925.
- Knapp N, Attinger S, Huth A. 2022. A question of scale: modeling biomass, gain and mortality distributions of a tropical forest. *Biogeosciences* 19: 4929–4944.
- Korhonen K, Ahola A, Heikkinen J, Henttonen H, Hotanen J-P, Ihalainen A, Melin M, Pitkänen J, Rätty M, Sirviö M *et al.* 2021. Forests of Finland 2014–2018 and their development 1921–2018. *Silva Fennica* 55.
- Lamarque J-F, Bond TC, Eyring V, Granier C, Heil A, Klimont Z, Lee D, Liousse C, Mieville A, Owen B *et al.* 2010. Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application. *Atmospheric Chemistry and Physics* 10: 7017–7039.
- Lempereur M, Martin-StPaul NK, Damesin C, Joffre R, Ourcival J-M, Rocheteau A, Rambal S. 2015. Growth duration is a better predictor of stem

- increment than carbon supply in a Mediterranean oak forest: implications for assessing forest productivity under climate change. *New Phytologist* 207: 579–590.
- Levine NM, Zhang K, Longo M, Baccini A, Phillips OL, Lewis SL, Alvarez-Dávila E, Segalin de Andrade AC, Brienen RJW, Erwin TL *et al.* 2016. Ecosystem heterogeneity determines the ecological resilience of the Amazon to climate change. *Proceedings of the National Academy of Sciences, USA* 113: 793–797.
- Longworth JB, Mesquita RC, Bentos TV, Moreira MP, Massoca PE, Williamson GB. 2014. Shifts in dominance and species assemblages over two decades in alternative successions in central Amazonia. *Biotropica* 46: 529–537.
- Lovenduski NS, Bonan GB. 2017. Reducing uncertainty in projections of terrestrial carbon uptake. *Environmental Research Letters* 12: 44020.
- Ma L, Hurtt G, Ott L, Sahajpal R, Fisk J, Lamb R, Tang H, Flanagan S, Chini L, Chatterjee A *et al.* 2022. Global evaluation of the Ecosystem Demography model (ED v.3.0). *Geoscientific Model Development* 15: 1971–1994.
- Ma L, Hurtt G, Tang H, Lamb R, Campbell E, Dubayah R, Guy M, Huang W, Lister A, Lu J *et al.* 2021. High-resolution forest carbon modelling for climate mitigation planning over the RGGI region, USA. *Environmental Research Letters* 16: 045014.
- Ma L, Hurtt G, Tang H, Lamb R, Lister A, Chini L, Dubayah R, Armston J, Campbell E, Duncanson L *et al.* 2023. Spatial heterogeneity of global forest aboveground carbon stocks and fluxes constrained by spaceborne lidar data and mechanistic modeling. *Global Change Biology* 29: 3378–3394.
- Mandl L, Viana-Soto A, Seidl R, Stritih A, Senf C. 2024. Unmixing-based forest recovery indicators for predicting long-term recovery success. *Remote Sensing of Environment* 308: 114194.
- Massoud EC, Xu C, Fisher RA, Knox RG, Walker AP, Serbin SP, Christoffersen BO, Holm JA, Kueppers LM, Ricciuto DM *et al.* 2019. Identification of key parameters controlling demographically structured vegetation dynamics in a land surface model: CLM4.5(FATES). *Geoscientific Model Development* 12: 4133–4164.
- Matsuo T, Martínez-Ramos M, Bongers F, van der Sande MT, Poorter L. 2021. Forest structure drives changes in light heterogeneity during tropical secondary forest succession. *Journal of Ecology* 109: 2871–2884.
- McDowell NG, Allen CD. 2015. Darcy's law predicts widespread forest mortality under climate warming. *Nature Climate Change* 5: 669–672.
- McDowell NG, Allen CD, Anderson-Teixeira K, Aukema BH, Bond-Lamberty B, Chini L, Clark JS, Dietze M, Grossiord C, Hanbury-Brown A *et al.* 2020. Pervasive shifts in forest dynamics in a changing world. *Science* 368: eaaz9463.
- Melton JR, Arora VK, Wisernig-Cojoc E, Seiler C, Fortier M, Chan E, Teckentrup L. 2020. CLASSIC v.1.0: the open-source community successor to the Canadian Land Surface Scheme (CLASS) and the Canadian Terrestrial Ecosystem Model (CTEM) – part 1: model framework and site-level performance. *Geoscientific Model Development* 13: 2825–2850.
- Mesquita R, Massoca PE, Jakovac CC, Bentos TV, Williamson GB. 2015. Amazon rain forest succession: stochasticity or land-use legacy? *Bioscience* 65: 849–861.
- Metcalfe CJE, Horvitz CC, Tuljapurkar S, Clark DA. 2009. A time to grow and a time to die: a new way to analyze the dynamics of size, light, age, and death of tropical trees. *Ecology* 90: 2766–2778.
- Montgomery RA, Chazdon RL. 2001. Forest structure, canopy architecture, and light transmittance in tropical wet forests. *Ecology* 82: 2707–2718.
- Moorcroft PR, Hurtt GC, Pacala SW. 2001. A method for scaling vegetation dynamics: the ecosystem demography model. *Ecological Monographs* 71: 557–586.
- Mrad A, Manzoni S, Oren R, Vico G, Lindh M, Katul G. 2020. Recovering the metabolic, self-thinning, and constant final yield rules in mono-specific stands. *Frontiers in Forests and Global Change* 3: 1240.
- Muscarella R, Lohbeck M, Martínez-Ramos M, Poorter L, Rodríguez-Velázquez JE, van Breugel M, Bongers F. 2017. Demographic drivers of functional composition dynamics. *Ecology* 98: 2743–2750.
- Naudts K, Ryder J, McGrath MJ, Otto J, Chen Y, Valade A, Bellasen V, Berhongaray G, Bönisch G, Campioli M *et al.* 2015. A vertically discretised canopy description for ORCHIDEE (SVN r2290) and the modifications to the energy, water and carbon fluxes. *Geoscientific Model Development* 8: 2035–2065.
- Needham JF, Chambers J, Fisher R, Knox R, Koven CD. 2020. Forest responses to simulated elevated CO₂ under alternate hypotheses of size- and age-dependent mortality. *Global Change Biology* 26: 5734–5753.
- Nepstad DC, Tohver IM, Ray D, Moutinho P, Cardinot G. 2007. Mortality of large trees and lianas following experimental drought in an Amazon forest. *Ecology* 88: 2259–2269.
- Oliva J, Stenlid J, Martínez-Vilalta J. 2014. The effect of fungal pathogens on the water and carbon economy of trees: implications for drought-induced mortality. *New Phytologist* 203: 1028–1035.
- O'Sullivan M, Sitoh S, Friedlingstein P, Luijckx IT, Peters W, Rosan TM, Arneeth A, Arora VK, Chandra N, Chevallier F *et al.* 2024. The key role of forest disturbance in reconciling estimates of the northern carbon sink. *Communications Earth & Environment* 5: 1–10.
- Palviainen M, Laurén A, Pumpanen J, Bergeron Y, Bond-Lamberty B, Larjavaara M, Kashian DM, Köster K, Prokushkin A, Chen HYH *et al.* 2020. Decadal-scale recovery of carbon stocks after wildfires throughout the boreal forests. *Global Biogeochemical Cycles* 34: e2020GB006612.
- Pausas JG, Keeley JE. 2021. Wildfires and global change. *Frontiers in Ecology and the Environment* 19: 387–395.
- Peltoniemi M, Mäkipää R. 2011. Quantifying distance-independent tree competition for predicting Norway spruce mortality in unmanaged forests. *Forest Ecology and Management* 261: 30–42.
- Pillet M, Joetzer E, Belmin C, Chave J, Ciais P, Dourdain A, Evans M, Hérault B, Luysaert S, Poulter B. 2018. Disentangling competitive vs. climatic drivers of tropical forest mortality. *Journal of Ecology* 106: 1165–1179.
- Pipionti C, Anderson-Teixeira KJ, Davies SJ, Allen D, Bourg NA, Burslem DFRP, Cárdenas D, Chang-Yang C, Chuyong G, Cordell S *et al.* 2022. Distribution of biomass dynamics in relation to tree size in forests across the world. *New Phytologist* 234: 1664–1677.
- Poorter L, Bongers F, Aide TM, Almeyda Zambrano AM, Balvanera P, Becknell JM, Boukili V, Brancalion PHS, Broadbent EN, Chazdon RL *et al.* 2016. Biomass resilience of Neotropical secondary forests. *Nature* 530: 211–214.
- Pretzsch H. 2006. Species-specific allometric scaling under self-thinning: evidence from long-term plots in forest stands. *Oecologia* 146: 572–583.
- Pretzsch H, Biber P, Schütze G, Uhl E, Rötzer T. 2014. Forest stand growth dynamics in Central Europe have accelerated since 1870. *Nature Communications* 5: 4967.
- Pretzsch H, del Río M, Arcangeli C, Bielak K, Dudzinska M, Ian Forrester D, Kohnle U, Ledermann T, Matthews R, Nagel R *et al.* 2023. Competition-based mortality and tree losses. An essential component of net primary productivity. *Forest Ecology and Management* 544: 121204.
- Pugh TAM, Arneeth A, Kautz M, Poulter B, Smith B. 2019. Important role of forest disturbances in the global biomass turnover and carbon sinks. *Nature Geoscience* 12: 730–735.
- Pugh TAM, Jones CD, Huntingford C, Burton C, Arneeth A, Brovkin V, Ciais P, Lomas M, Robertson E, Piao SL *et al.* 2018. A large committed long-term sink of carbon due to vegetation dynamics. *Earth's Future* 6: 1413–1432.
- Pugh TAM, Rademacher T, Shafer SL, Steinkamp J, Barichivich J, Beckage B, Haverd V, Harper A, Heinke J, Nishina K *et al.* 2020. Understanding the uncertainty in global forest carbon turnover. *Biogeosciences* 17: 3961–3989.
- Pugh TAM, Seidl R, Liu D, Lindeskog M, Chini LP, Senf C. 2024. The anthropogenic imprint on temperate and boreal forest demography and carbon turnover. *Global Ecology and Biogeography* 33: 100–115.
- R Core Team. 2023. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Reich PB. 2014. The world-wide 'fast-slow' plant economics spectrum: a traits manifesto. *Journal of Ecology* 102: 275–301.
- Reilly MJ, Spies TA. 2016. Disturbance, tree mortality, and implications for contemporary regional forest change in the Pacific Northwest. *Forest Ecology and Management* 374: 102–110.
- Reineke LH. 1933. Perfecting a stand-density index for even-aged forests. *Journal of Agricultural Research* 20: 627–638.
- Rozendaal DMA, Chazdon RL. 2015. Demographic drivers of tree biomass change during secondary succession in northeastern Costa Rica. *Ecological Applications* 25: 506–516.
- Rozendaal DMA, Chazdon RL, Arreola-Villa F, Balvanera P, Bentos TV, Dupuy JM, Hernández-Stefanoni JL, Jakovac CC, Lebrija-Trejos EE,

- Lohbeck M *et al.* 2017. Demographic drivers of aboveground biomass dynamics during secondary succession in neotropical dry and wet forests. *Ecosystems* 20: 340–353.
- Rüger N, Condit R, Dent DH, DeWalt SJ, Hubbell SP, Lichstein JW, Lopez OR, Wirth C, Farrior CE. 2020. Demographic trade-offs predict tropical forest dynamics. *Science* 368: 165–168.
- Ruiz-Benito P, Ratcliffe S, Jump AS, Gómez-Aparicio L, Madrigal-González J, Wirth C, Kändler G, Lehtonen A, Dahlgren J, Kattge J *et al.* 2017. Functional diversity underlies demographic responses to environmental variation in European forests. *Global Ecology and Biogeography* 26: 128–141.
- Sato H, Itoh A, Kohyama T. 2007. SEIB–DGVM: a new dynamic global vegetation model using a spatially explicit individual-based approach. *Ecological Modelling* 200: 279–307.
- Seidl R, Thom D, Kautz M, Martin-Benito D, Peltoniemi M, Vacchiano G, Wild J, Ascoli D, Petr M, Honkaniemi J *et al.* 2017. Forest disturbances under climate change. *Nature Climate Change* 7: 395–402.
- Seiler C, Melton JR, Arora VK, Sitch S, Friedlingstein P, Anthoni P, Goll D, Jain AK, Joetzier E, Lienert S *et al.* 2022. Are terrestrial biosphere models fit for simulating the global land carbon sink? *Journal of Advances in Modeling Earth Systems* 14: e2021MS002946.
- Shevliakova E, Malyshev S, Martinez-Cano I, Milly PCD, Pacala SW, Ginoux P, Dunne KA, Dunne JP, Dupuis C, Findell KL *et al.* 2024. The Land Component LM4.1 of the GFDL Earth System Model ESM4.1: model description and characteristics of land surface climate and carbon cycling in the historical simulation. *Journal of Advances in Modeling Earth Systems* 16: e2023MS003922.
- Shevliakova E, Pacala SW, Malyshev S, Hurtt GC, Milly PCD, Caspersen JP, Sentman LT, Fisk JP, Wirth C, Crevoisier C. 2009. Carbon cycling under 300 years of land use change: importance of the secondary vegetation sink. *Global Biogeochemical Cycles* 23: 156.
- Shorohova E, Kuuluvainen T, Kangur A, Jõgiste K. 2009. Natural stand structures, disturbance regimes and successional dynamics in the Eurasian boreal forests: a review with special reference to Russian studies. *Annals of Forest Science* 66: 201.
- Sitch S, Huntingford C, Gedney N, Levy PE, Lomas M, Piao SL, Betts R, Ciais P, Cox P, Friedlingstein P *et al.* 2008. Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five Dynamic Global Vegetation Models (DGVMs): uncertainty in Land Carbon Cycle Feedbacks. *Global Change Biology* 14: 2015–2039.
- Smith B, Prentice IC, Sykes MT. 2001. Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space. *Global Ecology* 25: 248–469.
- Strigul N, Pristinski D, Purves D, Dushoff J, Pacala S. 2008. Scaling from trees to forests: tractable macroscopic equations for forest dynamics. *Ecological Monographs* 78: 523–545.
- Sukma HT, Di Stefano J, Swan M, Sitters H. 2019. Mammal functional diversity increases with vegetation structural complexity in two forest types. *Forest Ecology and Management* 433: 85–92.
- Taylor AR, Gao B, Chen HYH. 2020. The effect of species diversity on tree growth varies during forest succession in the boreal forest of central Canada. *Forest Ecology and Management* 455: 117641.
- Teobaldelli M, Somogyi Z, Migliavacca M, Usoltsev VA. 2009. Generalized functions of biomass expansion factors for conifers and broadleaved by stand age, growing stock and site index. *Forest Ecology and Management* 257: 1004–1013.
- Thomas RQ, Hurtt GC, Dubayah R, Schilz MH. 2008. Using lidar data and a height-structured ecosystem model to estimate forest carbon stocks and fluxes over mountainous terrain. *Canadian Journal of Remote Sensing* 34: S351–S363.
- University of East Anglia Climatic Research Unit, Harris. 2021. *CRU JRA v2.2: a forcings dataset of gridded land surface blend of Climatic Research Unit (CRU) and Japanese reanalysis (JRA) data*; Jan 1901–Dec 2020.
- Urban DL, O'Neill RV, Shugart HH. 1987. Landscape ecology. *BioScience* 37: 119–127.
- Van Cleve K, Viereck LA. 1981. Forest succession in relation to nutrient cycling in the Boreal Forest of Alaska. In: West DC, Shugart HH, Botkin DB, eds. *Forest succession: concepts and application*. New York, NY, USA: Springer, 185–211.
- Walker LR, Wardle DA, Bardgett RD, Clarkson BD. 2010. The use of chronosequences in studies of ecological succession and soil development. *Journal of Ecology* 98: 725–736.
- Weng E, Aleinov I, Singh R, Puma MJ, McDermid SS, Kiang NY, Kelley M, Wilcox K, Dybzinski R, Farrior CE *et al.* 2022. Modeling demographic-driven vegetation dynamics and ecosystem biogeochemical cycling in NASA GISS's Earth system model (MODELE-BIOME v.1.0). *Geoscientific Model Development* 15: 8153–8180.
- Weng E, Dybzinski R, Farrior CE, Pacala SW. 2019. Competition alters predicted forest carbon cycle responses to nitrogen availability and elevated CO₂: simulations using an explicitly competitive, game-theoretic vegetation demographic model. *Biogeosciences* 16: 4577–4599.
- West DC, Shugart HH, Botkin DB. 1981. *Forest succession: concepts and application*. New York, NY, USA: Springer.
- West GB, Brown JH, Enquist BJ. 1997. A general model for the origin of allometric scaling laws in biology. *Science* 276: 122–126.
- Wolf A, Ciais P, Bellassen V, Delbart N, Field CB, Berry JA. 2011. Forest biomass allometry in global land surface models. *Global Biogeochemical Cycles* 25: 123.
- Wright SJ, Kitajima K, Kraft NJB, Reich PB, Wright IJ, Bunker DE, Condit R, Dalling JW, Davies SJ, Díaz S *et al.* 2010. Functional traits and the growth-mortality trade-off in tropical trees. *Ecology* 91: 3664–3674.
- Xu X, van der Sleen P, Groenendijk P, Vlam M, Medvigy D, Moorcroft P, Peticord D, Ma Y, Zuidema PA. 2024. Constraining long-term model predictions for woody growth using tropical tree rings. *Global Change Biology* 30: e17075.
- Yap SL, Davies SJ, Condit R. 2016. Dynamic response of a Philippine dipterocarp forest to typhoon disturbance. *Journal of Vegetation Science* 27: 133–143.
- Yoda K, Kira T, Ogawa H, Hozumi K. 1963. Self-thinning in overcrowded pure stands under cultivated and natural conditions. *Journal of Biology, Osaka City University* 14: 107–129.
- Yu K, Chen HYH, Gessler A, Pugh TAM, Searle EB, Allen RB, Pretzsch H, Ciais P, Phillips OL, Brienen RJW *et al.* 2024. Forest demography and biomass accumulation rates are associated with transient mean tree size vs. density scaling relations. *PNAS Nexus* 3: pgae008.
- Zeileis A, Grothendieck G. 2005. zoo: S3 infrastructure for regular and irregular time series. *Journal of Statistical Software* 14: 1–27.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Demographic (woody) carbon budget test results, and visualisation of the temporal trajectory of the outputted carbon pool variable C_{wood} , and the flux-derived carbon pool.

Fig. S2 Observed chronosequence regrowth and mature forest dynamics benchmarking data.

Fig. S3 Benchmarking data of stand structure, number of stems, and woody biomass by size class.

Fig. S4 Distribution of data used for this study.

Fig. S5 Example of the plots that can visualise the self-thinning period and length.

Fig. S6 An example of using Method 1.

Fig. S7 An example of using Method 2.

Fig. S8 An example of using Method 3.

Fig. S9 An example of using Method 4.

Fig. S10 Two-step process for manual self-thinning selection Method 4.

Fig. S11 Self-thinning method selection for each individual model and site.

Fig. S12 Mean equilibrium biomass plotted against mean annual regrowth rate over the first 50 yr postdisturbance.

Fig. S13 Woody mortality rates at FIN, for all models, smoothed using a 30-yr running mean.

Fig. S14 Woody mortality rates at BIA, for all models, for all models, smoothed using a 30-yr running mean.

Fig. S15 Woody mortality rates at BCI, for all models, for all models, smoothed using a 30-yr running mean.

Fig. S16 Example of a mortality rate trajectory within the forest recovery phases.

Fig. S17 Stand structure from LPJ-GUESS run output at simulation year 450 with no patch-destroying disturbance, and a 100-yr disturbance interval (default) turned on.

Fig. S18 Succession patterns in models that simulate between-PFT competition as part of demographic dynamics.

Fig. S19 Growth rates at Bialowieza for LPJ-GUESS and BiomeE, showcasing the impact of PFT succession on the growth rate during recovery (30-yr-smoothed).

Fig. S20 Frequency distribution of mean temperature for the years 1900–2023 for all sites for which regrowth datapoints exist, and frequency distribution of mean temperature for years 1991–2020, for the three simulation sites.

Methods S1 Model descriptions and model-specific setups for this study.

Methods S2 Demographic carbon balance.

Methods S3 Observational data.

Methods S4 Determination of the self-thinning period and slope.

Methods S5 Forest recovery phases and phase alignment.

Methods S6 Stand structure benchmarks.

Methods S7 'Naïve' – unaligned regrowth rate analysis.

Methods S8 Mean temperature analysis for regrowth sites.

Notes S1 Simulation protocol used to perform the benchmarking simulations.

Notes S2 Forest phase alignment, additional variables used for evidence for identifying the forest Phases.

Notes S3 Stand structure benchmark per model and site, with the observational data ranges reported in Table S6.

Notes S4 Regrowth benchmark per model and site, with the observational data ranges reported in Table S6.

Table S1 Individual model-PFTs mapped onto site-specific species (plant categories for BCI).

Table S2 Description of model-specific modes of biomass reductions that was invoked to enable the complete forest removal after 30 yr of simulation.

Table S3 Parameters adjusted from default parameter set to obtain P0 output.

Table S4 Mapping of model-output variables to D-BEN variables, where these deviate from the 'D-BEN' definition (table S3.1 in Notes S1).

Table S5 Upper and lower simulation year that make up the period from which each model is considered in equilibrium.

Table S6 Benchmark data overview in terms of range, type and justification for the method to derive observed ranges, alongside with some comments/observations on the behaviour of the data, and filtering criteria.

Table S7 Self-thinning selection method per model and site and resulting duration and self-thinning slope identification Methods 1–4 are briefly explained.

Table S8 Detailed forest phase classification.

Table S9 Forest Recovery Phase Classification. Initial phase and self-thinning timing.

Table S10 Forest Carbon and dynamics variables summarised across the first 30 yr of regrowth and equilibrium.

Table S11 Independent identification of self-thinning phase onset from two methods, self-thinning onset determination and forest phase determination.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

Disclaimer: The New Phytologist Foundation remains neutral with regard to jurisdictional claims in maps and in any institutional affiliations.