Forest responses to last-millennium hydroclimate variability are governed by spatial variations in ecosystem sensitivity

Article

Accepted Version

Rollinson, C. R. ORCID: https://orcid.org/0000-0003-0181-7293, Dawson, A., Raiho, A. M., Williams, J. W., Dietze, M. C. ORCID: https://orcid.org/0000-0002-2324-2518, Hickler, T., Jackson, S. T., McLachlan, J., Moore, D. J. P., Poulter, B., Quaife, T. ORCID: https://orcid.org/0000-0001-6896-4613, Steinkamp, J. ORCID: https://orcid.org/0000-0002-7861-8789, Trachsel, M. and Morin, X. (2021) Forest responses to last-millennium hydroclimate variability are governed by spatial variations in ecosystem sensitivity. Ecology Letters, 24 (3). pp. 498-508. ISSN 1461-023X doi: 10.1111/ele.13667 Available at https://centaur.reading.ac.uk/96165/

To link to this article DOI: http://dx.doi.org/10.1111/ele.13667

Publisher: Wiley

ECOLOGY LETTERS

Forest responses to last-millennium hydroclimate variability are governed by spatial variations in ecosystem sensitivity

Journal:	*Ecology Letters*
Manuscript ID	ELE-01100-2020.R1
Manuscript Type:	Letters
Date Submitted by the Author:	18-Nov-2020
Complete List of Authors:	Rollinson, Christine; Morton Arboretum, Center for Tree Science Dawson, Andria; Mount Royal University, Department of General Education Raiho, Ann; University of Notre Dame, Department of Biological Sciences Williams, John; University of Wisconsin, Geography; University of Wisconsin-Madison, Center for Climatic Research Dietze, Michael; Boston University, Earth and Environment Hickler, Thomas; Senckenberg Research Institutes and Natural History Museum, Biodiversity and Climate Research Centre; Goethe University Frankfurt, Department of Physical Geography Jackson, Stephen; DOI Southwest Climate Science Center, U.S. Geological Survey; University of Arizona, Dept. of Geosciences McLachlan, Jason; University of Notre Dame, Biological Sciences Moore, David; University of Arizona, School of Natural Resources and the Environment Poulter, Ben; NASA Goddard Space Flight Center, Quaife, Tristan; University of Reading, Department of Meteorology Steinkamp, Jörg; Senckenberg Society for Nature Research, Senckenberg Biodiversity and Climate Research Centre (SBiK-F); Johannes Gutenberg University Trachsel, Mathias; University of Wisconsin-Madison, Department of Geography

Forest responses to last-millennium hydroclimate variability are governed by spatial variations in ecosystem sensitivity

Authors:

1. **Christine R. Rollinson***, crollinson@mortonarb.org
    a. Center for Tree Science, The Morton Arboretum, 4100 Illinois Route 53, Lisle, IL, 60532
2. **Andria Dawson,** andria.dawson@gmail.com
    a. Department of General Education, Mount Royal University, Calgary, Alberta, T3E 6K6, Canada
3. **Ann M. Raiho**, ann.raiho@gmail.com
    a. Department of Biological Sciences, University of Notre Dame, 100 Galvin Life Science Center, Notre Dame, IN, 46556,
4. **John W. Williams**, jwwilliams1@wisc.edu, jww@geography.wisc.edu
    a. Department of Geography and Center for Climatic Research, University of Wisconsin-Madison, Madison, WI 53704
5. **Michael C. Dietze**, dietze@bu.edu orcid:0000-0002-2324-2518
    a. Department of Earth and Environment, Boston University, 685 Commonwealth Ave, Boston, MA 02215
6. **Thomas Hickler**, thomas.hickler@senckenberg.de
    a. Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Senckenberganlage 25, 60325 Frankfurt/Main, Germany
    b. Department of Physical Geography, Goethe University, Frankfurt/Main, Germany
7. **Stephen T. Jackson**, jackson@uwyo.edu; stjackson@usgs.gov
    a. US Geological Survey, Southwest and South Central Climate Adaptation Centers
    b. Department of Geosciences, University of Arizona, Tucson, AZ 85721
8. **Jason McLachlan,** Jason.S.McLachlan.2@nd.edu, jmclachl@nd.edu
    a. Department of Biological Sciences, University of Notre Dame, 100 Galvin Life Science Center, Notre Dame, IN, 46556
9. **David JP Moore**, davidjpmoore@email.arizona.edu
    a. School of Natural Resources, University of Arizona, Tucson, AZ, 85721
10. **Benjamin Poulter**, benjamin.poulter@nasa.gov
    a. NASA GSFC, Biospheric Sciences Lab., Greenbelt, MD 20771
11. **Tristan Quaife**, t.l.quaife@reading.ac.uk
    a. Department of Meteorology, University of Reading, Reading, RG6 6BB UK
12. **Jörg Steinkamp**, steinkamp.joerg@gmail.com
    a. Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Frankfurt/Main, Germany
    b. Johannes Gutenberg University, Mainz, Germany
13. **Mathias Trachsel**, mtrachs@gmail.com
    a. Department of Geography, University of Wisconsin-Madison, Madison, WI 53704

Running Title

Past forest variability and sensitivity

Keywords

stability, vulnerability, drought, ecosystem modeling, paleoecology, climate change,

Statement of Authorship:

CRR, AD, AMR, and JWW designed the study. AD, AMR, JWW, STJ, JM, and MT created pollen reconstructions and aided in interpretation (STEPPS, ReFAB). AD and AMR wrote the pollen methods. CRR, AMR, MCD, JM, DJPM, BP, TQ, and JS performed ecosystem model simulations and aided in interpretation. CRR and JWW wrote the manuscript with additional input from AD, AMR, and all authors.

Data Accessibility Statement:

Should the manuscript be accepted, the data supporting the results will be archived in two public repositories and the DOIs will be included at the end of this article. Pollen data is already available or will be made available upon acceptance on the EDI data portal as an msb-paleon product. The Environmental Data Initiative is an NSF-funded program tailored towards environmental data and works closely with the US Long-Term Ecological Research (LTER) Network, NSF Macrosystems Biology program (which funded our work), and DataONE. Terrestrial ecosystem model drivers are being archived on the ORNL DAAC and will be available at the following DOI: https://doi.org/10.3334/ORNLDAAC/1779. The Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) is managed by NASA's Earth Science Data and Information Systems program and is well suited to archive ecosystem model output, which is often large and has converged on netcdf as a standard file format. These repositories have been approved by *Ecology Letters* editorial staff. All code for analyses is publicly available on Github: https://github.com/PalEON-Project/EcosystemVariability

Manuscript Metadata:

Abstract Word Count: 147

Manuscript Word Count: 4,991

Text Box Word Count: N/A

Number of References: 67

Number of Figures: 4

Number of Tables: 1

Number of Text Boxes: 0

Corresponding Author: Christine R. Rollinson, Center for Tree Science, The Morton Arboretum, 4100 Illinois Route 53, Lisle, IL 60532; phone: 630-719-2422; email: crollinson@mortonarb.org

Abstract

Forecasts of future forest change are governed by ecosystem sensitivity to climate change, but ecosystem model projections are under-constrained by data at multidecadal and longer timescales. Here, we quantify ecosystem sensitivity to centennial-scale hydroclimate variability, by comparing dendroclimatic and pollen-inferred reconstructions of drought, forest composition and biomass for the last millennium with five ecosystem model simulations. In both observations and models, spatial patterns in ecosystem responses to hydroclimate variability are strongly governed by ecosystem sensitivity rather than climate exposure. Ecosystem sensitivity was higher in models than observations and highest in simpler models. Model-data comparisons suggest that interactions among biodiversity, demography, and ecophysiology processes dampen the sensitivity of forest composition and biomass to climate variability and change. Integrating ecosystem models with observations from timescales extending beyond the instrumental record can better understand and forecast the mechanisms regulating forest sensitivity to climate variability in a complex and changing world.

Introduction

Exposure to 21st-century climate change is expected to profoundly impact global forest composition, diversity, and structure (Dawson *et al.* 2011; Keeley *et al.* 2019), but the sensitivity of ecosystems to climate variability at multi-decadal to centennial time scales is poorly constrained by instrumental observations. Multiple observational studies that employ subcontinental- to continental-scale data networks across a broad range of timescales have sought to empirically estimate the sensitivity of forest ecosystems to climate variability. The sensitivity of tree growth rates, biomass accumulation, and ecophysiological processes to interannual climate variability is well-documented by dendroecological data, with compelling evidence that forest sensitivity to climate depends on forest age and is non-stationary across space and time (Charney *et al.* 2016; Klesse *et al.* 2018; Thom *et al.* 2019; Peltier & Ogle 2020). On glacial-interglacial timescales, networks of fossil pollen records show that temperature variations are the primary driver of forest composition and species distributions (Shuman *et al.* 2004; Nolan *et al.* 2018), while over the last several thousand years, hydroclimate variability has strongly affected forest composition and structure in temperate forests of the northeastern and upper midwestern United States (Booth *et al.* 2012; Shuman *et al.* 2019).

Terrestrial ecosystem models used to forecast responses to climate change often have difficulty reproducing broad-scale and long-term responses to environmental variability, despite being well-grounded in empirical evidence and ecological theory (Friedlingstein *et al.* 2006, 2014; Matthes *et al.* 2016). These models mechanistically connect ecophysiological processes and climate variability to past and present changes in forest composition and structure but are subject to uncertainty in external forcings (e.g., drivers), process representation, and parametrization that complicates data-model comparisons (Figure 1) (LeBauer *et al.* 2013; Matthes *et al.* 2016; Dietze 2017; McLachlan & PalEON Project 2018). Each model includes hypotheses about the primary processes and ecosystem characteristics governing forest change, various simplifying assumptions, and tradeoffs between computational tractability and process complexity (De Kauwe *et al.* 2013; Walker *et al.* 2014; Medlyn *et al.* 2015). Previous data-model comparisons have returned mixed evidence about whether models underestimate or overestimate the sensitivity of forest processes such as net primary productivity (NPP) and mortality to climate change (Schimel *et al.* 2015; Walker *et al.* 2015; Rollinson *et al.* 2017). As a result, projections of forest compositional and structural responses to climate change have high uncertainty, which propagates to increased uncertainty in science-based adaptation planning (Friedlingstein *et al.* 2014).

Several challenges have traditionally hindered the joint analysis and integration of terrestrial ecosystem models and paleoecological data to better constrain modeled responses to climate variations at multi-decadal and longer timescales. First, the raw observations collected from fossil pollen records (counts of individual pollen taxa) have no direct counterparts in ecosystem models. Bayesian hierarchical models are providing new process-based approaches to infer emergent ecosystem properties from fossil pollen records, such as forest composition, diversity, percent cover, and biomass (Raiho *et al.* in prep; Blarquez & Aleman 2016; Dawson *et al.* 2016), but the number of state variables that can be estimated from paleoecological data remains small relative to the number of latent (i.e., unobservable) variables simulated by ecosystem models (Fig. 1). Second, pre-instrumental model-data comparisons are complicated by reliance on driver datasets derived from general circulation models (GCMs). GCMs generally capture macroscale spatial patterns and low-frequency trends in climate but are unable to fully capture the complexity and stochasticity of local to regional-scale weather phenomena at the subdaily resolution needed to drive ecosystem models, resulting in systematic spatial and temporal biases in model simulations (Anav *et al.* 2013; Matthes *et al.* 2016; Dietze *et al.* 2018). Third, the native temporal resolution varies between paleodata and models and requires a temporal standardization. Due to these challenges, the predicted sensitivity of ecosystem model state variables such as forest composition and biomass to climate change is largely unvalidated by observations at multidecadal and longer timescales, resulting in wide divergence among terrestrial ecosystem models in their 21st-century projections (Friedlingstein *et al.* 2006, 2014). Fourth, terrestrial ecosystem models vary widely in represented processes, which can challenge intermodel comparisons but also provide insight into key governing ecological processes when data-model discrepancies emerge.

Here, we seek to establish the patterns of forest ecosystem and climate variability in the north-central and northeastern US for the last millennium (850-1850 C.E.) and identify the mechanisms underpinning both forest ecosystem sensitivity and observed data-model discrepancies. In these analyses, we test hypotheses about the relative importance of hydroclimate exposure, defined as the magnitude of drought variability, and ecosystem sensitivity as determinants of the variability seen in forest ecosystems. We also hypothesize that ecosystem models will be overly sensitive to hydroclimate variability due to insufficient representation of ecophysiological and demographic processes that can dampen climate responses. To this end, we present a novel series of data-model and model-model comparisons that are designed to overcome traditional barriers to data-model intercomparison for pre-instrumental times. Our analyses combine dendroclimatic indices of drought, recently published Bayesian spatiotemporal estimates of forest composition and biomass derived from pollen that provide independent checks on last-millennium simulations from five terrestrial ecosystem models for the northeastern and upper midwestern United States. The data-model comparisons discriminate among differing representations of forest processes such as productivity and demography, while the model-model comparisons help diagnose causal relationships among ecological processes, changes in forest states, and climate variability (Fig. 1). To test hypotheses while also overcoming known geographic biases in the model simulations of ecosystem state such as forest composition that source back to biases in the climate model drivers (Matthes *et al.* 2016), we develop a new variability metric that we apply to the data and model-derived products that focuses on comparisons among variability of hydroclimate, composition, and biomass (Fig. 1). Our results indicate that at centennial timescales, spatial patterns in the variability of forest composition and biomass are regulated by ecological factors such as ecotonal position and complexity rather than climate exposure as defined by the local magnitude of climate variability.

Materials & Methods

Overview

We employ a combination of data-model and model-model comparisons (Fig. 1) in which we combine paleoclimatic and paleoecological datasets to draw inferences about past variations in hydroclimate and forest composition and biomass. The temporal domain of this study is 850-1850 AD and is bounded by the temporal extent of the climate drivers available for our model simulations (850 AD) and time of EuroAmerican settlement-era tree surveys (ca. 1850 AD). In our study, ‘data’ refers to observation-based statistical models of past drought, forest composition, and biomass, reconstructed from tree rings, historical tree surveys, and networks of fossil pollen records. These data-based inferences are fully independent of the ecosystem model simulations. Model-based comparisons are from the PalEON Ecosystem Model Intercomparison Project (PEMIP) (Rollinson *et al.* 2017), which used spatially and temporally downscaled past climate simulations from the Fifth Coupled Model Intercomparison Project (CMIP5) as drivers.

Comparisons among ecosystem model simulations and empirical data rely on normalized values compared in environmental space, rather than geographic space, in order to reduce the effects of any bias in the climate drivers in our analyses and to focus on sensitivity of ecosystems to climate variability (Supplemental Figure 1).

Observational Datasets

The empirically inferred datasets leverage recent advances in pollen-vegetation modeling (Dawson *et al.* 2016), a form of proxy system modeling (Evans *et al.* 2013) in which ecosystem state variables such as composition and biomass are estimated along with associated observational uncertainties. Of the three inferred datasets used here, two were derived from networks of fossil pollen records provided by individual data contributors and the Neotoma Paleoecology Database and were calibrated against historical surveys of forest composition and structure from the early stages of EuroAmerican settlement (Liu *et al.* 2011; Dawson *et al.* 2016; Goring *et al.* 2016; Kujawa *et al.* 2016; Paciorek *et al.* 2016). Pollen-based inferences are based on statistical pollen-vegetation models (PVMs) called STEPPS and ReFAB, and represent fractional vegetation composition and total woody biomass, respectively, for 12 tree genera that are common elements of upper Midwest forests. STEPPS is a Bayesian hierarchical spatio-temporal model that infers fractional forest composition from networks of fossil pollen records (Paciorek & McLachlan 2009; Dawson *et al.* 2016, 2019b; Trachsel *et al.* 2020). STEPPS employs a process-based representation of pollen dispersal and production, with taxon-specific parameterizations. STEPPS is calibrated using spatial datasets of pollen samples and forest composition data, here from the settlement era (Paciorek & McLachlan 2009; Dawson *et al.* 2016), then run for fossil pollen assemblages for other time intervals to produce posterior estimates of past forest composition. Using this framework, STEPPS: (i) explicitly characterizes uncertainty in data and processes, with posterior distributions of process parameters and state variables such as forest composition, and (ii) borrows information across space and time, allowing for spatially comprehensive estimates of composition. For both the upper Midwestern USA (UMW; Minnesota, Wisconsin, Michigan) (Dawson *et al.* 2019a) and the northeastern USA (NEUS) (Trachsel *et al.* 2020), STEPPS has been used to estimate centennially resolved forest composition for the late Holocene (250 B.C. to 1750 A.D) at a 24 km grid; here we use the results from 850 to 1750 AD.

ReFAB also employs a similar approach to STEPPS but focuses specifically on estimating total aboveground woody biomass (Raiho *et al.* in prep). ReFAB is calibrated using the relationship between settlement-era multivariate pollen counts and biomass from PLS surveys (Paciorek *et al.* 2019). Parameter estimates from calibration are then used to reconstruct centennially resolved biomass for 77 sites in the UMW for the last 10,000 years (Raiho *et al.* in prep). ReFAB can characterize the uncertainty in sediment pollen age estimates, calibration parameters, the relationship between species composition and total aboveground woody biomass, and species-level allometries.

The Living Blended Drought Atlas (LBDA) provides yearly estimates of summer (mean June, July, August) Palmer Severity Drought Index (PDSI) for North America, based on networks of tree-growth chronologies (Cook *et al.* 2010; Woodhouse *et al.* 2010). We used PDSI as our measure of hydroclimate variability because it is an important predictor of forest dynamics in this domain and can also be calculated directly from the meteorological forcings used for the ecosystem model simulations (Clifford & Booth 2015; Cook *et al.* 2015). LBDA PDSIs are provided at 0.5-degree spatial grid resolution. Due to varying temporal extent of tree-growth chronologies, the temporal extent of the LBDA varies. The earliest years in this spatial domain ranged from 0 to 1671 AD, while the latest year was 2005 (Supplemental Figure 1).

Modeling Datasets

PEMIP model simulations here are composed of five ecosystem models with dynamic vegetation (ED2; LINKAGES; LPG-WSL; LPJ-GUESS; and JULES-TRIFFID) run at 254 locations across the eastern and midwestern US at 0.5-degree spatial resolution (Rollinson *et al.* 2020). These models vary in how they characterize forest composition and carbon dynamics and range from species-based with little ecophysiological process representation (e.g., LINKAGES) to detailed ecophysiology and cohort representation, but reliance on plant functional types (PFTs; e.g. ED2, Table 1). LPJ-GUESS and LPJ-WSL both included stochastic fire disturbances in their simulations, while other models such as ED and LINKAGES include processes of tree mortality that assume landscape-scale equilibrium (Rollinson *et al.* 2017).

PEMIP climate drivers were temporally downscaled and bias-corrected from existing past climate simulations to meet the external forcing needs of the ecosystem model ensemble (Supplemental Figure 1) (Kumar *et al.* 2012; Rollinson *et al.* 2017). CCSM4 output from the Paleoclimate Modeling Intercomparison Project, Phase III (PMIP3) past millennium simulations and the Coupled Model Intercomparison Project, Phase 5 (CMIP5) historical simulations were downscaled to 0.5-degree spatial resolution and 6-hourly temporal resolution using standard protocols (Kumar *et al.* 2012; Rollinson *et al.* 2017). After the 6-hourly PEMIP climate driver datasets were created, they were then temporally averaged to meet the specific driver requirements of individual ecosystem models, which vary in temporal resolution. ED2 and JULES-TRIFFID use the full suite of 6-hourly drivers for temperature, precipitation, shortwave radiation, longwave radiation, surface pressure, specific humidity, wind speed, and carbon dioxide concentration. Meteorological drivers for the two LPJ variants include daily temperature, precipitation, and shortwave radiation plus longwave radiation for LPJ-WSL. LINKAGES only requires monthly average temperature and precipitation. Soil texture used to parameterize locations in model simulations was extracted from the Harmonized World Soil Database (Wei *et al.* 2014). Monthly temperature and precipitation were combined with soil water holding capacity computed from model driver soil texture and depth to calculate PDSI, following (Cook *et al.* 2015), but using the Thornthwaite equation for evapotranspiration (Thornthwaite & Mather 1957; Pelton *et al.* 1960). We used the Thornthwaite equation so that the calculation of PDSI was independent of internal model dynamics, including evapotranspiration, which can vary widely among ecosystem models, even when given the same temperature and precipitation drivers, due to differences in model structure and parameterization. From the ecosystem models, we extracted fractional forest composition and total aboveground biomass, which can be directly compared to paleoecological observations, and four variables that are latent, i.e., unobservable in the paleoecological record (Fig. 1): gross primary productivity (GPP), net primary productivity (NPP), net ecosystem exchange (NEE), and leaf area index (LAI).

Analyses

Analyses focused on the comparison of empirical data and ecosystem model outputs of centennial-scale variability in forest composition and biomass driven by drought variability over the last 1,000 years. Our analytical approach involved three key stages to maximize commensurability between observations and model output: 1) temporal homogenization of all variables to a common centennial resolution; 2) development of a common normalized variability metric for ecosystem and drought variability to facilitate comparison across different variables, and 3) use of hydroclimate sensitivity as the basis for all model-data and model-model comparisons to minimize the potential effects of biases in the climate model drivers.

i. Temporal Homogenization

For annually resolved datasets in our study, including the LBDA and all model output and drivers, a generalized additive model (GAM) was used to generate time series with the similar centennial-scale smoothing as the pollen inferred observational datasets. In this process, the response variable for analysis (e.g. drought, biomass, GPP) was modeled as a function of time (year) using a thin-plate regression spline with one knot per 100 years (e.g. 10 knots for a 1,000 year window) using the *gam* function in the *mgcv* package in R (Wood 2017; Simpson 2018). To capture the temporal uncertainty similar to that generated in the PVMs, we generated a 1000-member posterior distribution of each predicted variable through time using the error and covariance of the intercept and spline parameters. We then extracted the predicted values at 100-year intervals corresponding to the windows captured by the STEPPS and ReFAB output.

ii. Variability Metric

To facilitate comparisons among variables with different units such as composition and biomass, we developed a base metric for all analyses, consisting of the normalized mean temporal variability of each dataset (eq. 1).

$$\textbf{equation 1: } variability_i = \ln \frac{\overline{d_i}}{\bar{x}}$$

$$\textbf{equation 2: } d_{i,t} = \left| x_{i,t} - x_{i,t-1} \right|$$

Mean temporal variability at each location ($\overline{d_i}$) for each variable (e.g., composition, biomass, PDSI) was calculated as the mean of the absolute first differences between adjacent time points

$(t, t-1)$ extracted from centennially resolved time series for each location (i) (eq. 2). The use of first differences is a discretization of the first derivative and describes the rate of change at each timestep. Each first-difference calculation was based on the mean of the posterior draws from the STEPPS or ReFAB PVM or to the GAMs fitted to the LBDA data and ecosystem model variables. We normalized variability by dividing the mean first differences for each location ($\overline{d_i}$) by the variable mean for that dataset across the entire spatiotemporal domain ($\overline{x}$). For forest compositional data, the variability metric was calculated using the taxon or plant functional type (PFT) with the highest fractional composition at each location, with the choice of taxon or PFT allowed to vary among sites. For all analyses and presented results, normalized variability is log-transformed to meet standard statistical assumptions of Gaussian distributions and homoscedasticity (eq. 1).

iii. Hydroclimate Sensitivity

After the normalized temporal variability was calculated for PDSI and all ecosystem variables, sensitivity to hydroclimate variability was defined as the slope of a linear regression between variability as the independent variable and variability of the ecosystem response variable such as composition or biomass. These analyses always used the appropriate observational or modeled PDSI variability (i.e., LBDA for the pollen-inferred compositional variability; calculated PEMIP driver PDSI variability for the model-simulated compositional variability) to ensure internal consistency between climatic forcing and ecosystem response.

Results

In the observational data, variability in forest composition or biomass in the northeastern US (NEUS) and upper midwestern US (UMW), did not correlate to drought variability (Table 1, Figs. 2, 3) in contrast with the hypothesis that high exposure to climate variability should lead to increased compositional variability. Neither the full spatiotemporal domain (Table 1) nor the UMW (Fig. 3, sensitivity slope = 0.010 SE 0.018) showed a significant relationship between reconstructed drought and composition variability, although the NEUS showed weak sensitivity (Fig. 3, sensitivity slope = 0.065 SE 0.027). Reconstructed biomass variability (Fig 2., biomass reconstructions not available for the NEUS, (Paciorek *et al.* 2019)) also was uncorrelated to drought variability (Table 1) and instead showed the highest variability at the historic prairie-forest ecotone (Fig. 2) (Goring & Williams 2017). In pollen-based reconstructions, composition and biomass variability were weakly but positively related (Fig. 3c, R^2=0.09, slope=0.479 SE 0.187) and locations with higher taxonomic richness tended to have higher variability (Supplemental Fig. 2).

Modeled ecosystem sensitivity to drought variability was generally similar to or higher than observations, with less-complex models tending to have a too-high predicted sensitivity relative to the empirical reconstructions (Fig. 3). Composition variability was more sensitive to drought variability than in reconstructions for three of five ecosystem models (ED2, LPJ-WSL, and TRIFFID), with the data-model discrepancy most pronounced in models with fewer plant types or taxa (Fig. 3a, Table 1). JULES-TRIFFID, which had only two tree PFTs (deciduous and evergreen), had the highest drought sensitivity (composition slope = -8.633 SE = 1.075, composition sensitivity slope 0.411 SE = 0.022). LPJ-WSL and ED2, with respectively six and five PFTs, had similar mean compositional variability (LPJ-WSL slope = -7.829 SE = 0.943, ED2 slope = -7.156 SE = 0.514), although LPJ-WSL was approximately twice as sensitive to hydroclimate variability as ED2 (Fig. 3a, Table 1, LPJ-WSL slope = 0.252 SE =0.018, ED2 slope = 0.118 SE = 0.018). LINKAGES, which simulated 15 individual species, had among the lowest sensitivity to drought variability (Fig. 3a, Table 1, composition slope = -6.598 SE = 0.478, composition sensitivity slope 0.074 SE = 0.018).

Ecosystem models with simpler representation of vegetation ecophysiology (LINKAGES, JULES-TRIFFID) also had a too-high sensitivity of biomass to drought variability relative to empirical reconstructions (Table 1, Fig. 3b). Both LINKAGES and JULES-TRIFFID showed a tight positive coupling of biomass sensitivity to drought variability, which corresponded to strong correlations between biomass and composition variability (Fig. 3c). LINKAGES showed a one-to-one relationship between composition and biomass variability, which is much stronger than reconstructions (Fig. 3c). Of all the models, only LPJ-WSL was consistent with the data in showing a weakly negative relationship between biomass and PDSI variability (Fig. 3b) while also showing a positive correlation between biomass and composition variability (Fig. 3c).

Further analysis of latent variables in the ecosystem models confirmed that variations in modeled ecosystem sensitivity to hydroclimate variability is linked to model complexity of ecosystem composition and processes (Fig. 4). There is a cascading series of linkages in physiological variables within and among taxa (Figs. 1, 4), in which gross primary productivity (GPP) is directly influenced by temperature and moisture availability, while other state variables such as net primary productivity (NPP), leaf area index (LAI), and aboveground biomass (AGB) are regulated by additional downstream processes that may decouple their variability from climate variability (Fig. 1). Hence, in most models, GPP variability is the most sensitive to drought variability (Fig. 4, Supplemental Table 1). In all models, sensitivity of forest composition to drought variability seems to be most closely linked to sensitivity of NPP. NPP sensitivity tended to be higher in low-diversity models such as JULES-TRIFFID (Figure 4, Supplemental Table 1). Higher diversity through more tree types or taxa was associated with higher compositional variability and reduced sensitivity to drought (Figure 3, Table 1, Supplemental Figure 2).

Models with more detailed representation of plant ecophysiology and either demography or disturbance (e.g., ED2, LPJ-GUESS, LPJ-WSL) also tended to have lower biomass sensitivity to hydroclimate variability (Fig. 4) and agree more closely with observations (Fig. 3). Biomass sensitivity to drought variability in our model ensemble was similar to NEE sensitivity in all models except LPJ-GUESS (Fig. 4, Supplemental Table 1). LINKAGES and JULES-TRIFFID may be overly sensitive to hydroclimate variability for entirely different reasons. LINKAGES has a fairly simple representation of ecophysiological processes while being able to represent species-level demographic dynamics (Table 1). In contrast, JULES-TRIFFID contains a sophisticated representation of ecophysiology but for only two tree PFTs and five PFTs total (Table 1). The other models tend to be more intermediate cases, with intermediate to more sophisticated representations of both ecophysiology and vegetation dynamics.

Discussion

Over the last millennium (850-1850 A.D.), both paleodata networks and model simulations suggest that spatial patterns in forest composition and biomass variability in northeastern and upper midwestern United States are governed more by spatial variations in ecosystem sensitivity and less by spatial variations in exposure to climate variability. Ecotonal regions such as the prairie-forest border have higher variability in composition and structure than areas of high PDSI variability (Fig. 2). The intermodel comparisons suggest that added complexity allows slow-to-change variables such as composition and biomass to be insensitive to climate variability at centennial scales despite sensitivity of fast-changing ecophysiological processes such as gross and net primary productivity (Fig. 4). Incorporation of ecological processes and characteristics such as diversity and demography all tend to reduce simulated climate sensitivity and better align simulations with observations (Figs. 3, 4).

These analyses represent a milestone towards the goal of more comprehensive and rigorous data-model comparisons for timescales and time periods extending beyond the instrumental record. Common challenges for multi-centennial data-model comparisons include 1) a need for process-informed statistical models of inference for paleoecological data, 2) generally lower temporal resolution in paleoecological data than in model simulations and with more latent variables than for the instrumental period, 3) biases in paleoclimatic simulations leading to biases in ecosystem model simulations, and 4) differences among models in driver datasets and represented processes. The pollen-vegetation models used in our study include processes for pollen productivity and dispersal that translates relative pollen abundances into metrics of forest composition and biomass that can be directly compared to those produced by ecosystem models (Paciorek & McLachlan 2009; Dawson *et al.* 2016). We further increased the commensurability between centennially resolved pollen-based quantifications of forest change and higher-frequency information from tree rings and ecosystem models by using GAMs to achieve time series with similarly temporally smoothed properties (Simpson 2018). By focusing on time series variability rather than directly comparing magnitude and timing of change in specific geographic locations or taxonomic groupings we were able to overcome documented ecosystem model biases arising from driver, process, and parameter limitations (Matthes *et al.* 2016; Dietze 2017). Finally, we leveraged differences in process representation among models as a means of evaluating the importance of specific ecosystem processes for producing emergent patterns of climate sensitivity that are consistent with paleoecological data (Medlyn *et al.* 2015; McLachlan & PalEON Project 2018).

Prior studies have indicated that forest composition and growth is sensitive to climate variability at annual to centennial scales (Shuman *et al.* 2004; Allen *et al.* 2010; Thom *et al.* 2019), yet there is also increasingly strong evidence that tree-climate relationships are non-stationary and subject to multiple interacting factors, leading to spatially complex forest responses to climate change (Girardin *et al.* 2016) and variations in climatic sensitivity across space and time (Rollinson *et al.* in press; Thom *et al.* 2019; Peltier & Ogle 2020; Wilmking *et al.* 2020). Several possible explanations exist for the reporting here of generally low sensitivity of forest composition and biomass to hydroclimate in reconstructions (Fig. 2). First, this apparent insensitivity may be due to the temporal grain of this study. The centennially resolved temporal grain of our analyses limits detection of annual-scale growth variations, the effects of stochastic or short-lived extreme events such as sub-decadal to decadal drought (Breshears *et al.* 2005; Allen *et al.* 2010; Seidl *et al.* 2011), or disturbance events such as fire and pest outbreaks, unless these are large enough to cause stand-replacing mortality events. Disturbance processes are often unrepresented in ecosystem models or treated as purely stochastic and with implicit assumptions of landscape-scale equilibria (Seidl *et al.* 2011; Fisher *et al.* 2018; McCabe & Dietze 2019). Of the ecosystem models used here, LPJ-WSL and LPJ-GUESS included fire in their simulations as a semi-mechanistic process following GLOBFIRM (Thonicke *et al.* 2001), which estimates burned area as a function of daily fire probabilities that are a function of fuel moisture and fuel load threshold. These models showed dampened biomass sensitivity to hydroclimate variability that was more closely aligned with observations (Fig. 4), but so did ED2, which lacked fire. Hence, process representation of fire or similar semi-stochastic disturbances is not a clear differentiator among modelled estimates of ecosystem climate sensitivity.

Second, apparent climate sensitivity might increase if the temporal extent was increased to include larger climate variations during the Holocene and last deglaciation. Although the last millennium includes climatic events such as the Medieval Climate Anomaly and Little Ice Age (PAGES 2k Consortium 2013), these climate variations appear to have been muted relative to earlier hydroclimate and temperature variations (Fischer *et al.* 2018). During the Holocene, hydroclimatic variability around the North Atlantic appears to have been an important driver of forest compositional changes and the collapses of individual tree species (Shuman *et al.* 2019). Large vegetation changes associated with the abrupt temperature variations of the Younger Dryas and last deglaciation are well documented (Williams *et al.* 2011), but the temporal extent of this study was constrained by the temporal extent of the last-millennium PMIP3/CMIP5 simulations used to drive ecosystem models (Braconnot *et al.* 2011; Taylor *et al.* 2012). As the next generation of transient Holocene simulations become available, the conclusions reached here about low apparent sensitivity can be revisited.

Third, this paper focuses on spatial patterns of climate and ecosystem variability, whereas most prior paleoecological studies have tended to focus on temporal variations (Shuman *et al.* 2004; Booth *et al.* 2012). Our analyses of low sensitivity are consistent with recent dendroecological studies of climate-driven rates of tree growth, which are quickly shifting from assumptions of stationary tree-climate relationships to demonstrations of spatially complex forest responses (Girardin *et al.* 2016) and variations in climatic sensitivity varies across space and time (Rollinson *et al.* in press; Thom *et al.* 2019; Peltier & Ogle 2020; Wilmking *et al.* 2020). By focusing on spatial variations in ecosystem variability over the last millennium, our analyses suggest spatial variation in ecosystem properties are a more important regulator than spatial variations in climate exposure. Finally, uncertainties in the proxy-based reconstructions may lower correlations as detrending techniques used to remove non-climatic signals such as age effects may dampen estimates of centennial-scale variability (Allen *et al.* 2018; Esper *et al.* 2018). Despite lower PDSI variability in the LBDA than model drivers, we do not think that spatial variability in hydroclimate variability in the empirical dataset is too low to detect effects on ecosystem variability. For example, hydroclimate data syntheses for the last 2000 years suggest opposite patterns of hydroclimate variations between Minnesota/Wisconsin and New England, which explain 30% of variance in the hydroclimate records (Shuman *et al.* 2019).

Process-based ecosystem models are the main vehicle for forecasting climate-driven ecosystem dynamics across a range of timescales and in principle are better able to accommodate past and future no-analog climates (Williams & Jackson 2007; Veloz *et al.* 2012). However, all ecosystem models face tradeoffs in their ability to represent taxonomic or functional diversity versus detailed ecophysiological processes that drive ecosystem change (Fisher *et al.* 2018). Process-based ecosystem models will never be able to capture the full complexity of ecosystems nor perfectly reproduce the patterns of climatological or ecological variability observed in the past due to observational uncertainties and incomplete constraints of many processes and parameterizations (Dietze 2017). This paper has shown how multiple paleoecological data streams can be combined with harmonized paleoclimatic simulations and multiple terrestrial ecosystem models to gain new insight into a) how diversity and biological processes can dampen ecosystem sensitivity to drought variability at broad spatial scales and b) the importance of complex representations of these aspects of ecosystems to achieve better agreement with the data. Nevertheless, these analyses followed a traditional approach in which past ecosystem reconstructions and simulations were run independently and compared at the final stage of analysis. The next major step forward is to move to a full data-assimilation framework, in which paleoecological observations and simulations are combined to overcome systematic biases in model drivers, parameterization, and output to better evaluate paleoecological change using mechanistic process-based frameworks (McLachlan & PalEON Project 2018). Through this iterative process that draws upon an ever-growing and diversifying suite of observational data streams (Farley *et al.* 2018), we can better understand the mechanisms regulating forest sensitivity to climate variability across a broad range of timescales and thereby better forecast future forest dynamics in a complex and rapidly changing world.

Acknowledgements

This work reflects the efforts of the Paleoecological Observatory Network (PalEON Project), funded by the National Science Foundation MacroSystems Biology under grants DEB-1241891, DEB-1241868, DEB-1241874, and DEB-1241851 and special thanks to Jody Peters, PalEON Project coordinator. PDSI calculations from ecosystem model drivers were derived from code graciously provided by Ben Cook. Fossil pollen data were obtained from the Neotoma Paleoecology Database (http://www.neotomadb.org) and its constituent database the North American Pollen Database. The work of the data contributors, data stewards, and the Neotoma community is gratefully acknowledged. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Allen, C.D., Macalady, A.K., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M., *et al.* (2010). A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *For. Ecol. Manag.*, 259, 660–684.

Allen, K.J., Villalba, R., Lavergne, A., Palmer, J.G., Cook, E.C., Fenwick, P., *et al.* (2018). A comparison of some simple methods used to detect unstable temperature responses in tree-ring chronologies. *Dendrochronologia*, 48, 52–73.

Anav, A., Friedlingstein, P., Kidston, M., Bopp, L., Ciais, P., Cox, P., *et al.* (2013). Evaluating the Land and Ocean Components of the Global Carbon Cycle in the CMIP5 Earth System Models. *J. Clim.*, 26, 6801–6843.

Blarquez, O. & Aleman, J.C. (2016). Tree biomass reconstruction shows no lag in postglacial afforestation of eastern Canada. *Can. J. For. Res.*, 46, 485–498.

Booth, R.K., Jackson, S.T., Sousa, V.A., Sullivan, M.E., Minckley, T.A. & Clifford, M.J. (2012). Multi-decadal drought and amplified moisture variability drove rapid forest community change in a humid region. *Ecology*, 93, 219–226.

Braconnot, P., Otto-Bliesner, B., Jungclaus, J. & Peterschmitt, J.-Y. (2011). The Paleoclimate Modeling Intercomparison Project contribution to CMIP5. *CLIVAR Exch.*, 16, 15–19.

Breshears, D.D., Cobb, N.S., Rich, P.M., Price, K.P., Allen, C.D., Balice, R.G., *et al.* (2005). Regional vegetation die-off in response to global-change-type drought. *Proc. Natl. Acad. Sci.*, 102, 15144–15148.

Charney, N.D., Babst, F., Poulter, B., Record, S., Trouet, V.M., Frank, D., *et al.* (2016). Observed forest sensitivity to climate implies large changes in 21st century North American forest growth. *Ecol. Lett.*, 19, 1119–1128.

Clifford, M.J. & Booth, R.K. (2015). Late-Holocene drought and fire drove a widespread change in forest community composition in eastern North America. *The Holocene*, 25, 1102–1110.

Cook, B.I., Ault, T.R. & Smerdon, J.E. (2015). Unprecedented 21st century drought risk in the American Southwest and Central Plains. *Sci. Adv.*, 1, e1400082.

Cook, E.R., Seager, R., Heim, R.R., Vose, R.S., Herweijer, C. & Woodhouse, C. (2010). Megadroughts in North America: placing IPCC projections of hydroclimatic change in a long-term palaeoclimate context. *J. Quat. Sci.*, 25, 48–61.

Dawson, A., Goring, S., Paciorek, C.J., Jackson, S.T., McLachlan, J.S. & Williams, J.W. (2019a). *STEPPS 2000 Year Forest Composition Estimates, Upper Midwest US, Level 2*. Environmental Data Initiative.

Dawson, A., Paciorek, C.J., Goring, S.J., Jackson, S.T., McLachlan, J.S. & Williams, J.W. (2019b). Quantifying trends and uncertainty in prehistoric forest composition in the upper Midwestern United States. *Ecology*, 100.

Dawson, A., Paciorek, C.J., McLachlan, J.S., Goring, S., Williams, J.W. & Jackson, S.T. (2016). Quantifying pollen-vegetation relationships to reconstruct ancient forests using 19th-century forest composition and pollen data. *Quat. Sci. Rev.*, 137, 156–175.

Dawson, T.P., Jackson, S.T., House, J.I., Prentice, I.C. & Mace, G.M. (2011). Beyond predictions: biodiversity conservation in a changing climate. *Science*, 332.

De Kauwe, M.G., Medlyn, B.E., Zaehle, S., Walker, A.P., Dietze, M.C., Hickler, T., *et al.* (2013). Forest water use and water use efficiency at elevated CO_2 : a model-data intercomparison at two contrasting temperate forest FACE sites. *Glob. Change Biol.*, 19, 1759–1779.

Dietze, M.C. (2017). Prediction in ecology: a first-principles framework. *Ecol. Appl.*, 27, 2048–2060.

Dietze, M.C., Fox, A., Beck-Johnson, L.M., Betancourt, J.L., Hooten, M.B., Jarnevich, C.S., *et al.* (2018). Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proc. Natl. Acad. Sci.*, 115, 1424–1432.

Esper, J., George, S.St., Anchukaitis, K., D'Arrigo, R., Ljungqvist, F.C., Luterbacher, J., *et al.* (2018). Large-scale, millennial-length temperature reconstructions from tree-rings. *Dendrochronologia*, 50, 81–90.

Evans, M.N., Tolwinski-Ward, S.E., Thompson, D.M. & Anchukaitis, K.J. (2013). Applications of proxy system modeling in high resolution paleoclimatology. *Quat. Sci. Rev.*, 76, 16–28.

Farley, S.S., Dawson, A., Goring, S.J. & Williams, J.W. (2018). Situating Ecology as a Big-Data Science: Current Advances, Challenges, and Solutions. *BioScience*, 68, 563–576.

Fischer, H., Meissner, K.J., Mix, A.C., Abram, N.J., Austermann, J., Brovkin, V., *et al.* (2018). Palaeoclimate constraints on the impact of 2 °C anthropogenic warming and beyond. *Nat. Geosci.*, 11, 474–485.

Fisher, R.A., Koven, C.D., Anderegg, W.R.L., Christoffersen, B.O., Dietze, M.C., Farrior, C.E., *et al.* (2018). Vegetation demographics in Earth System Models: A review of progress and priorities. *Glob. Change Biol.*, 24, 35–54.

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., *et al.* (2006). Climate–Carbon Cycle Feedback Analysis: Results from the C^{4}MIP Model Intercomparison. *J. Clim.*, 19, 3337–3353.

Friedlingstein, P., Meinshausen, M., Arora, V.K., Jones, C.D., Anav, A., Liddicoat, S.K., *et al.* (2014). Uncertainties in CMIP5 Climate Projections due to Carbon Cycle Feedbacks. *J. Clim.*, 27, 511–526.

Girardin, M.P., Bouriaud, O., Hogg, E.H., Kurz, W., Zimmermann, N.E., Metsaranta, J.M., *et al.* (2016). No growth stimulation of Canada's boreal forest under half-century of combined warming and CO_2 fertilization. *Proc. Natl. Acad. Sci.*, 113, E8406–E8414.

Goring, S.J., Mladenoff, D.J., Cogbill, C.V., Record, S., Paciorek, C.J., Jackson, S.T., *et al.* (2016). Novel and Lost Forests in the Upper Midwestern United States, from New Estimates of Settlement-Era Composition, Stem Density, and Biomass. *PLOS ONE*, 11, e0151935.

Keeley, J.E., van Mantgem, P. & Falk, D.A. (2019). Fire, climate and changing forests. *Nat. Plants*, 5, 774–775.

Klesse, S., Babst, F., Lienert, S., Spahni, R., Joos, F., Bouriaud, O., *et al.* (2018). A Combined Tree Ring and Vegetation Model Assessment of European Forest Growth Sensitivity to Interannual Climate Variability. *Glob. Biogeochem. Cycles*.

Kujawa, E.R., Goring, S., Dawson, A., Calcote, R., Grimm, E.C., Hotchkiss, S.C., *et al.* (2016). The effects of anthropogenic land cover change on pollen-vegetation relationships in the American Midwest. *Anthropocene*, 15, 60–71.

Kumar, J., Brooks, B.-G.J., Thornton, P.E. & Dietze, M.C. (2012). Sub-daily Statistical Downscaling of Meteorological Variables Using Neural Networks. *Procedia Comput. Sci.*, 9, 887–896.

LeBauer, D.S., Wang, D., Richter, K.T., Davidson, C.C. & Dietze, M.C. (2013). Facilitating feedbacks between field measurements and ecosystem models. *Ecol. Monogr.*, 83, 133–154.

Liu, F., Mladenoff, D.J., Keuler, N.S. & Moore, L.S. (2011). Broadscale variability in tree data of the historical Public Land Survey and its consequences for ecological studies. *Ecol. Monogr.*, 81, 259–275.

Matthes, J.H., Goring, S., Williams, J.W. & Dietze, M.C. (2016). Benchmarking historical CMIP5 plant functional types across the Upper Midwest and Northeastern United States. *J. Geophys. Res. Biogeosciences*, 121, 523–535.

McCabe, T.D. & Dietze, M.C. (2019). Scaling Contagious Disturbance: A Spatially-Implicit Dynamic Model. *Front. Ecol. Evol.*, 7.

McLachlan, J.S. & PalEON Project. (2018). Forecasting long-term ecological dynamics using open paleodata. *Past Glob. Change Mag.*, 26, 76–76.

Medlyn, B.E., Zaehle, S., De Kauwe, M.G., Walker, A.P., Dietze, M.C., Hanson, P.J., *et al.* (2015). Using ecosystem experiments to improve vegetation models. *Nat. Clim. Change*, 5, 528–534.

Nolan, C., Overpeck, J.T., Allen, J.R.M., Anderson, P.M., Betancourt, J.L., Binney, H.A., *et al.* (2018). Past and future global transformation of terrestrial ecosystems under climate change. *Science*, 361, 920–923.

Paciorek, C.J., Cogbill, C.V., Peters, J.A., Goring, S.J., Williams, J.W., Mladenoff, D.J., *et al.* (2019). *Statistically-estimated tree biomass, stem density, and basal area for the upper Midwestern United States at the time of Euro-American settlement* (preprint). Ecology.

Paciorek, C.J., Goring, S.J., Thurman, A.L., Cogbill, C.V., Williams, J.W., Mladenoff, D.J., *et al.* (2016). Statistically-Estimated Tree Composition for the Northeastern United States at Euro-American Settlement. *PLOS ONE*, 11, e0150087.

Paciorek, C.J. & McLachlan, J.S. (2009). Mapping Ancient Forests: Bayesian Inference for Spatio-Temporal Trends in Forest Composition Using the Fossil Pollen Proxy Record. *J. Am. Stat. Assoc.*, 104, 608–622.

PAGES 2k Consortium. (2013). Continental-scale temperature variability during the past two millennia. *Nat. Geosci.*, 6, 339–346.

Peltier, D.M.P. & Ogle, K. (2020). Tree growth sensitivity to climate is temporally variable. *Ecol. Lett.*, 1–20.

Pelton, W.L., King, K.M. & Tanner, C.B. (1960). An evaluation of the Thornthwaite and mean temperature methods for determining potential evapotranspiration. *Agron. J.*, 52, 387–395.

Raiho, A.M., Paciorek, C.J. & McLachlan, J.S. (in prep). Species migrations created a 5,000 year regional carbon sink past.

Rollinson, C., Dietze, M., Mathes, J.H., Steinkamp, J., Hickler, T., Poulter, B., *et al.* (2020). *PalEON: Terrestrial Ecosystem Model Ensemble for the Northeastern U.S., 850-2010*. ORNL Distributed Active Archive Center.

Rollinson, C.R., Finley, A.O., Alexander, M.R., Banerjee, S., Dixon Hamil, K.-A., Koenig, L., *et al.* (in press). Working across space and time: nonstationarity in ecological research and application. *Front. Ecol. Environ.*

Rollinson, C.R., Liu, Y., Raiho, A., Moore, D.J.P., McLachlan, J., Bishop, D.A., *et al.* (2017). Emergent climate and CO $_2$ sensitivities of net primary productivity in ecosystem models do not agree with empirical data in temperate forests of eastern North America. *Glob. Change Biol.*, 23, 2755–2767.

Schimel, D., Stephens, B.B. & Fisher, J.B. (2015). Effect of increasing CO $_2$ on the terrestrial carbon cycle. *Proc. Natl. Acad. Sci.*, 112, 436–441.

Seidl, R., Fernandes, P.M., Fonseca, T.F., Gillet, F., Jönsson, A.M., Merganičová, K., *et al.* (2011). Modelling natural disturbances in forest ecosystems: a review. *Ecol. Model.*, 222, 903–924.

Shuman, B., Newby, P., Huang, Y. & Webb, T. (2004). Evidence for the close climatic control of New England vegetation history. *Ecology*, 85, 1297–1310.

Shuman, B.N., Marsicek, J., Oswald, W.W. & Foster, D.R. (2019). Predictable hydrological and ecological responses to Holocene North Atlantic variability. *Proc. Natl. Acad. Sci.*, 116, 5985–5990.

Simpson, G.L. (2018). Modelling Palaeoecological Time Series Using Generalised Additive Models. *Front. Ecol. Evol.*, 6.

Taylor, K.E., Stouffer, R.J. & Meehl, G.A. (2012). An Overview of CMIP5 and the Experiment Design. *Bull. Am. Meteorol. Soc.*, 93, 485–498.

Thom, D., Golivets, M., Edling, L., Meigs, G.W., Gourevitch, J.D., Sonter, L.J., *et al.* (2019). The climate sensitivity of carbon, timber, and species richness covaries with forest age in boreal–temperate North America. *Glob. Change Biol.*, 25, 1446–2458.

Thonicke, K., Venevsky, S., Sitch, S. & Cramer, W. (2001). The role of fire disturbance for global vegetation dynamics: coupling fire into a Dynamic Global Vegetation Model. *Glob. Ecol. Biogeogr.*, 10, 661–677.

Thornthwaite, C.W. & Mather, J.R. (1957). Instructions and tables for computing potential evapotranspiration and the water balance. *Publ. Climatol.*, 10, 185–311.

Trachsel, M., Dawson, A., Paciorek, C.J., Williams, J.W., McLachlan, J.S., Cogbill, C.V., *et al.* (2020). Comparison of settlement-era vegetation reconstructions for STEPPS and REVEALS pollen–vegetation models in the northeastern United States. *Quat. Res.*, 95, 23–42.

Veloz, S.D., Williams, J.W., Blois, J.L., He, F., Otto-Bliesner, B. & Liu, Z. (2012). No-analog climates and shifting realized niches during the late quaternary: implications for 21st-century predictions by species distribution models. *Glob. Change Biol.*, 18, 1698–1713.

Walker, A.P., Hanson, P.J., De Kauwe, M.G., Medlyn, B.E., Zaehle, S., Asao, S., *et al.* (2014). Comprehensive ecosystem model-data synthesis using multiple data sets at two temperate forest free-air CO $_2$ enrichment experiments: Model performance at ambient CO $_2$ concentration: FACE MODEL-DATA SYNTHESIS. *J. Geophys. Res. Biogeosciences*, 119, 937–964.

Walker, A.P., Zaehle, S., Medlyn, B.E., De Kauwe, M.G., Asao, S., Hickler, T., *et al.* (2015). Predicting long-term carbon sequestration in response to CO $_2$ enrichment: How and why do current ecosystem models differ? *Glob. Biogeochem. Cycles*, 29, 476–495.

Wei, Y., Liu, S., Huntzinger, D.N., Michalak, A.M., Viovy, N., Post, W.M., *et al.* (2014). The North American Carbon Program Multi-scale Synthesis and Terrestrial Model Intercomparison Project – Part 2: Environmental driver data. *Geosci. Model Dev.*, 7, 2875–2893.

Williams, J.W., Blois, J.L. & Shuman, B.N. (2011). Extrinsic and intrinsic forcing of abrupt ecological change: case studies from the late Quaternary: Extrinsic and intrinsic abrupt ecological change. *J. Ecol.*, 99, 664–677.

Williams, J.W. & Jackson, S.T. (2007). Novel Climates, No-Analog Communities, and Ecological Surprises. *Front. Ecol. Environ.*, 5, 475–482.

Wilmking, M., Maaten-Theunissen, M., Maaten, E., Scharnweber, T., Buras, A., Biermann, C., *et al.* (2020). Global assessment of relationships between climate and tree growth. *Glob. Change Biol.*, 26, 3212–3220.

Wood, S.N. (2017). *Generalized Additive Models: An Introduction with R*. 2nd edn. Chapman and Hall/CRC, New York.

Woodhouse, C.A., Meko, D.M., MacDonald, G.M., Stahle, D.W. & Cook, E.R. (2010). A 1,200-year perspective of 21st century drought in southwestern North America. *Proc. Natl. Acad. Sci.*, 107, 21283–21288.

Manuscript Tables

Table 1: Comparison of 1) ecosystem model complexity, based on representation of diversity, demographic, and ecophysiological processes with 2) variability in forest composition (Comp) and biomass (Biom) and sensitivity to hydroclimate variability. Variability is a normalized metric of total change in the centennially resolved time series. Sensitivity is presented as the slope and standard error of linear regression between composition or biomass variability and hydroclimate variability. PFT = plant functional types. For sensitivity columns, * indicates slopes significantly different from zero ($p<0.05$); † indicates model slope significantly different from pollen ($p<0.05$).

Data Source & Model Name	**Tree Diversity Representation**	**Demographic Representation**	**Vegetation Processes**	**Comp. Var. (SD)**	**Comp. Sens. (SE)**	**Biom. Var. (SD)**	**Biom. Sens. (SE)**
Pollen: STEPPS, ReFAB	Genera: 12 trees	relative abundance	[implicit]	-2.032 (0.617)	0.026 (0.019)	-7.798 (0.770)	-0.156 (0.119)
ED2	PFTs: 5 tree	cohort	photosynthesis, allocation, cross-PFT competition, cross-cohort competition	-7.156 (0.514)	0.118 (0.018)*†	-7.505 (0.446)	-0.079 (0.027)*
LINK-AGES	Species: 15 tree	individual	cross-PFT competition, cross-cohort competition	-6.598 (0.478)	0.074 (0.018)*	-6.741 (0.999)	0.230 (0.028)*†
LPJ-GUESS	PFTs: 6 tree, 1 grass	cohort	photosynthesis, allocation, cross-PFT competition, cross-cohort competition	-7.290 (0.452)	0.056 (0.018)*	-7.379 (0.597)	-0.069 (0.027)*
LPJ-WSL	PFTs: 5 tree, 1 grass	PFT	photosynthesis, allocation, cross-PFT competition, cross-PFT competition	-7.829 (0.943)	0.252 (0.018)*†	-7.106 (0.964)	-0.020 (0.027)
JULES-TRIFFID	PFTs: 2 Tree, 2 grass, 1 shrub	PFT	Photosynthesis, allocation, cross-PFT competition	-8.633 (1.075)	0.411 (0.022)*†	-8.639 (0.952)	0.203 (0.033)*†

Manuscript Figures

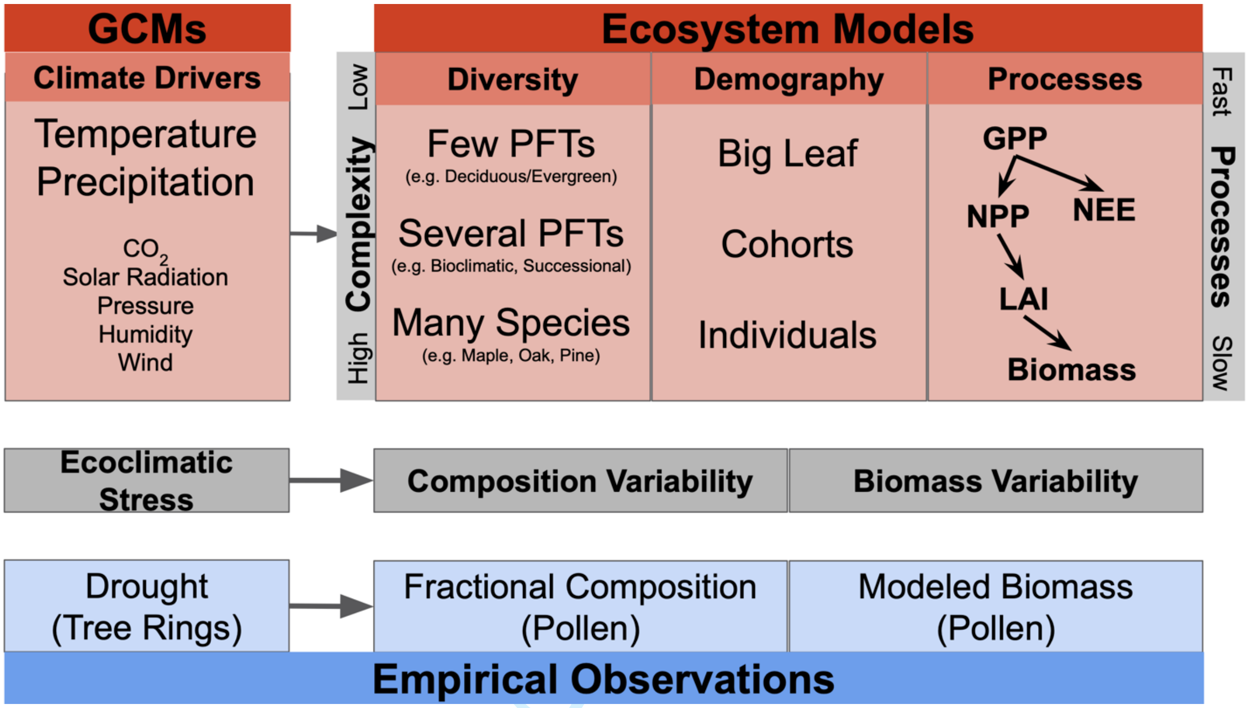


Figure 1: Overview of the unified conceptual framework (gray boxes) for parallel analysis of empirical data (blue boxes) and model output (red boxes). For ecosystem models, we describe the latent climatic and ecosystem processes that are unobservable in paleoecological data and differences among models in complexity. Complexity here is organized into three categories: 1) diversity, ranging from a few plant functional types (PFTs) to many species; 2) demography, ranging from 'big leaf' models with no explicit treatment of forest demography to models with individual trees; and 3) ecophysiological processes. Changes in forest biomass emerge from latent ecophysiological processes including gross primary productivity (GPP), net primary productivity (NPP), net ecosystem exchange (NEE), and leaf area index (LAI). Ecophysiological processes are controlled by model representation of higher-level vegetation processes (Table 1). Latent model drivers, processes, and states (red boxes) result in estimates of forest composition and biomass that can be compared to paleoecological data products (blue boxes). Models vary in complexity due to design philosophy and tradeoffs between model complexity and computational speed.

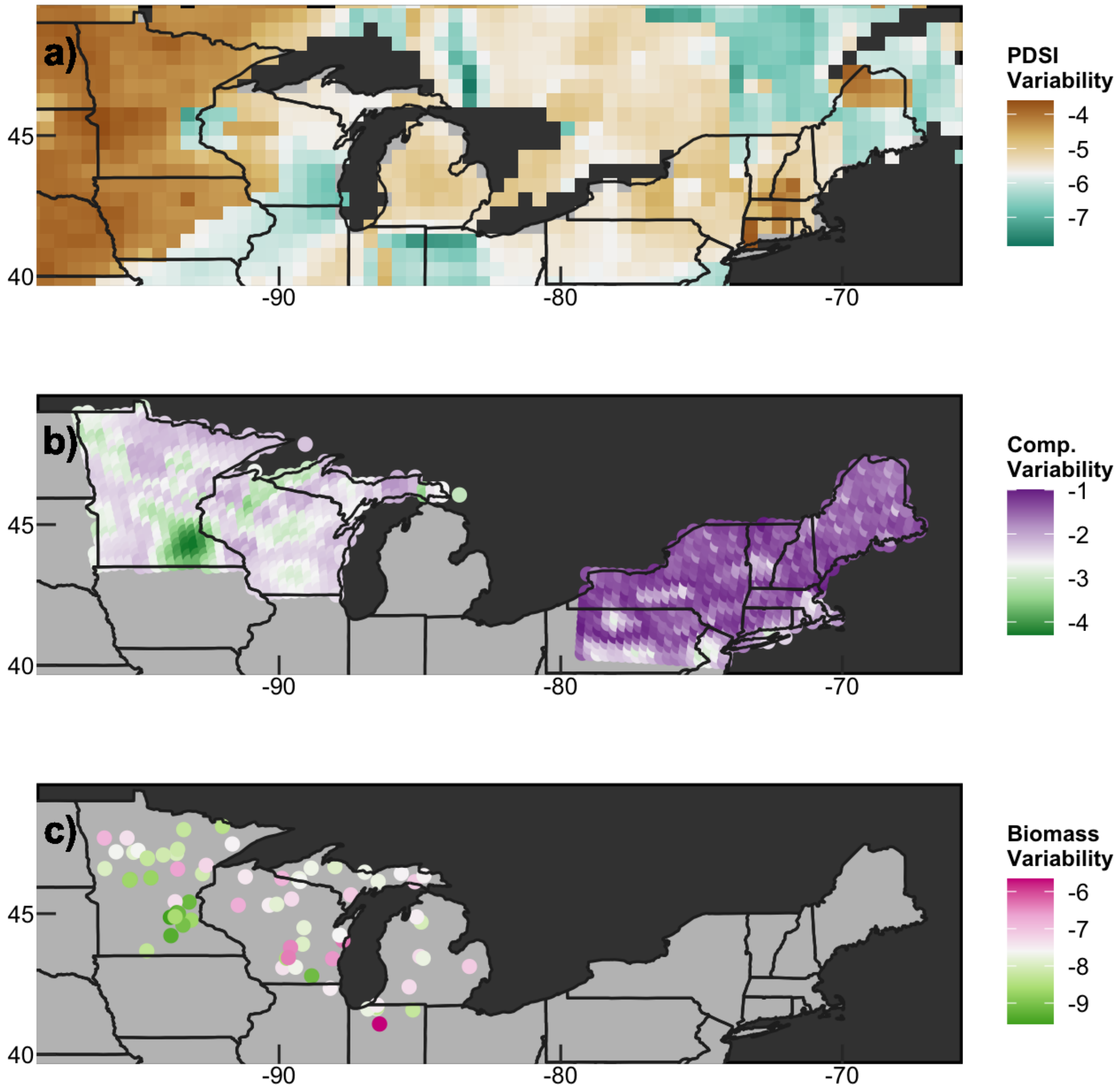


Figure 2: Spatial distribution of inferred temporal variability for 850 to 1850 AD for a) drought (PDSI) from the Living Blended Drought Atlas (44), b) forest composition from the STEPPS pollen-vegetation model (8, 24), and c) forest aboveground biomass from the ReFab pollen-biomass model (7). All variability estimates were divided by mean to facilitate inter-variable comparison (*Methods*). Spatial extents of compositional and biomass reconstructions are uneven across the study domain, as is the temporal extent of reconstructed drought variability (Supplemental Figure 1). Empirical comparisons of composition or biomass variability with drought variability are restricted to the common temporal extents for each location. In the log scale, negative values indicate locations with low variability whereas more positive values indicate high variability.

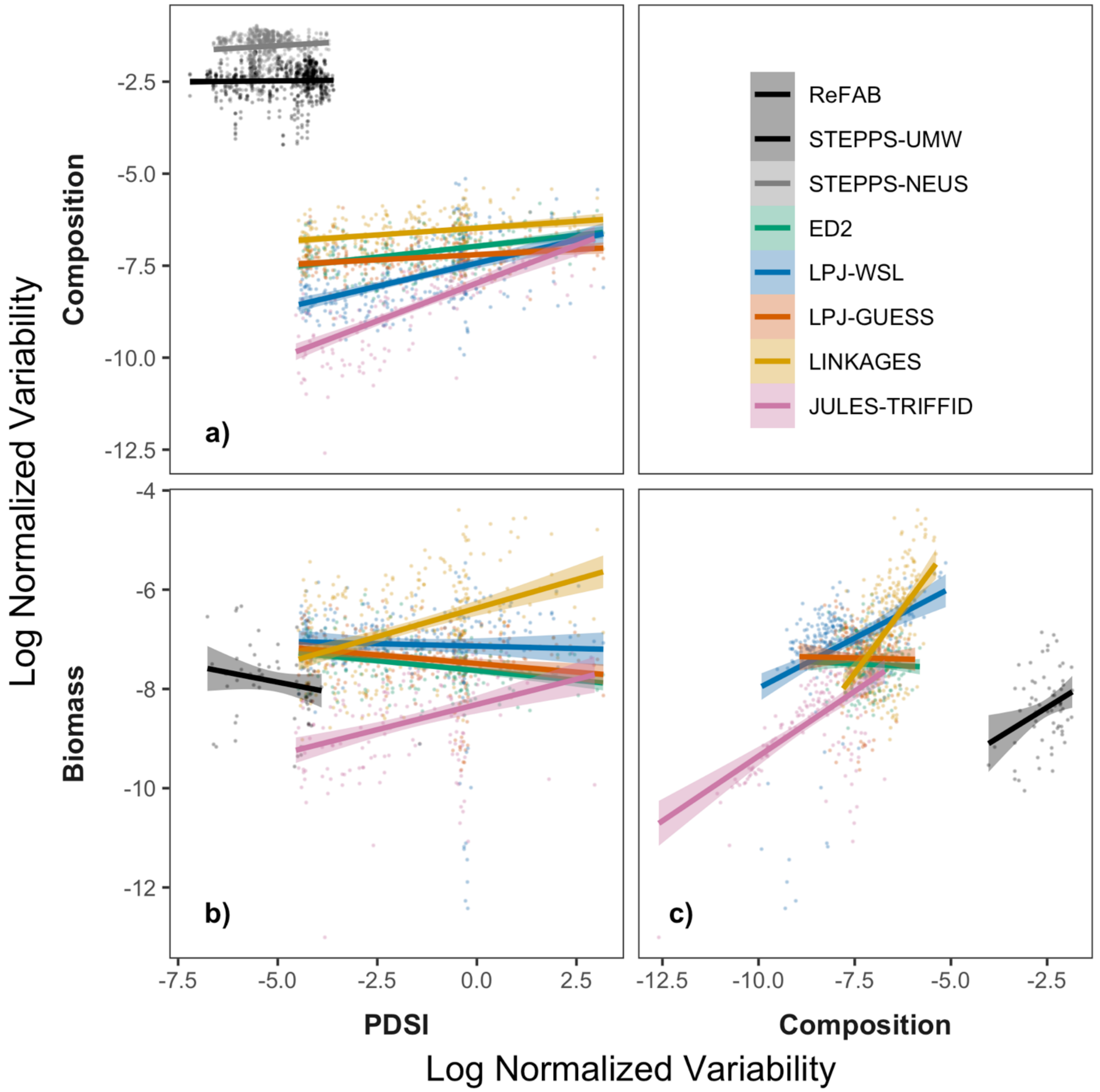


Figure 3: Inferred (black, gray) and simulated (colors) sensitivity of variability of forest composition and biomass to ecohydrological variability (PDSI) (a,b) and of biomass variability to compositional variability (c). Inferred variables suggest weak to no correlation (low sensitivity) between climate variability and ecosystem variability (composition and biomass). In contrast, ecosystem models generally simulate higher sensitivity of ecosystems to climate variability. Inferred compositional (STEPPS) and biomass (ReFAB) variability are positively correlated, while this relationship varied among models. In the log scale, negative values indicate locations with low variability whereas more positive values indicate high variability.

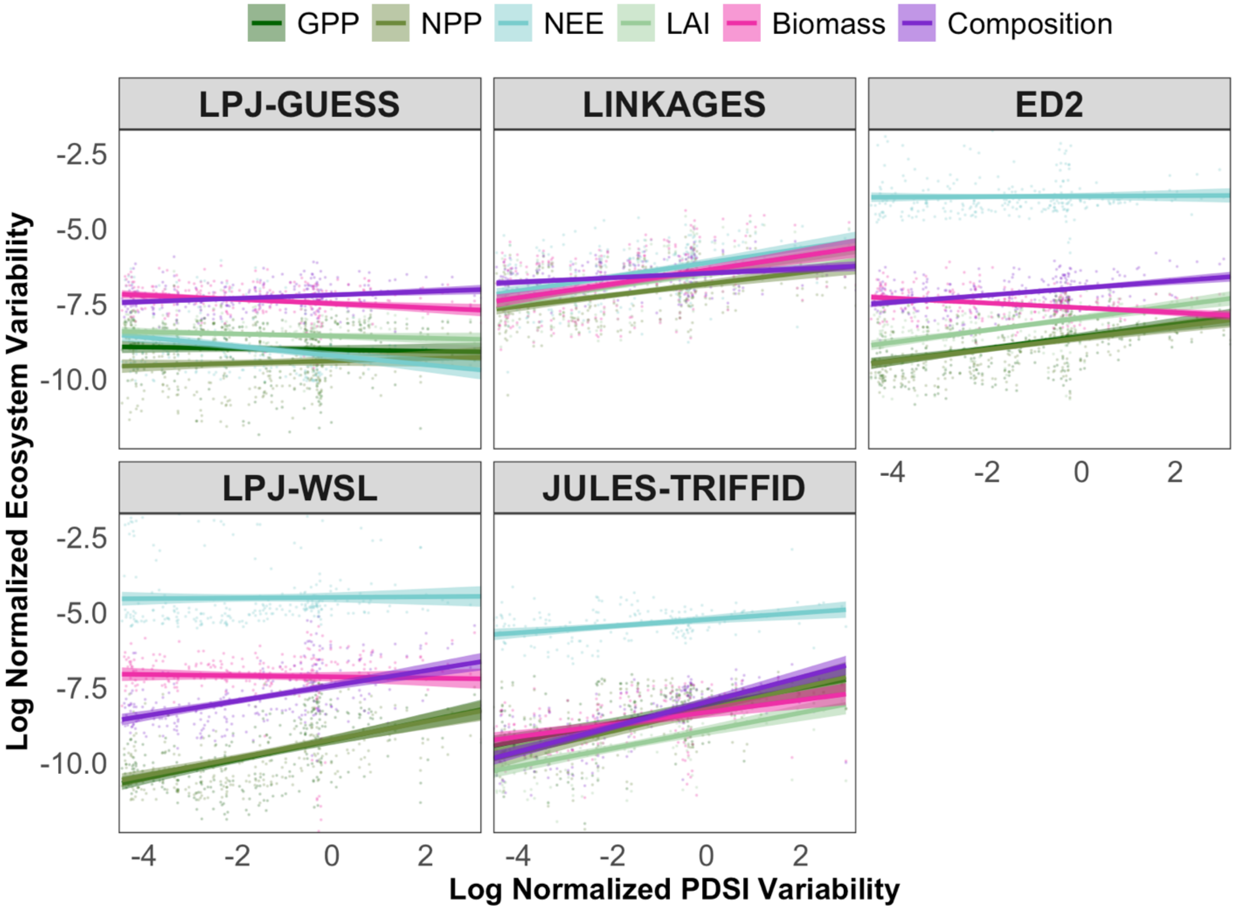


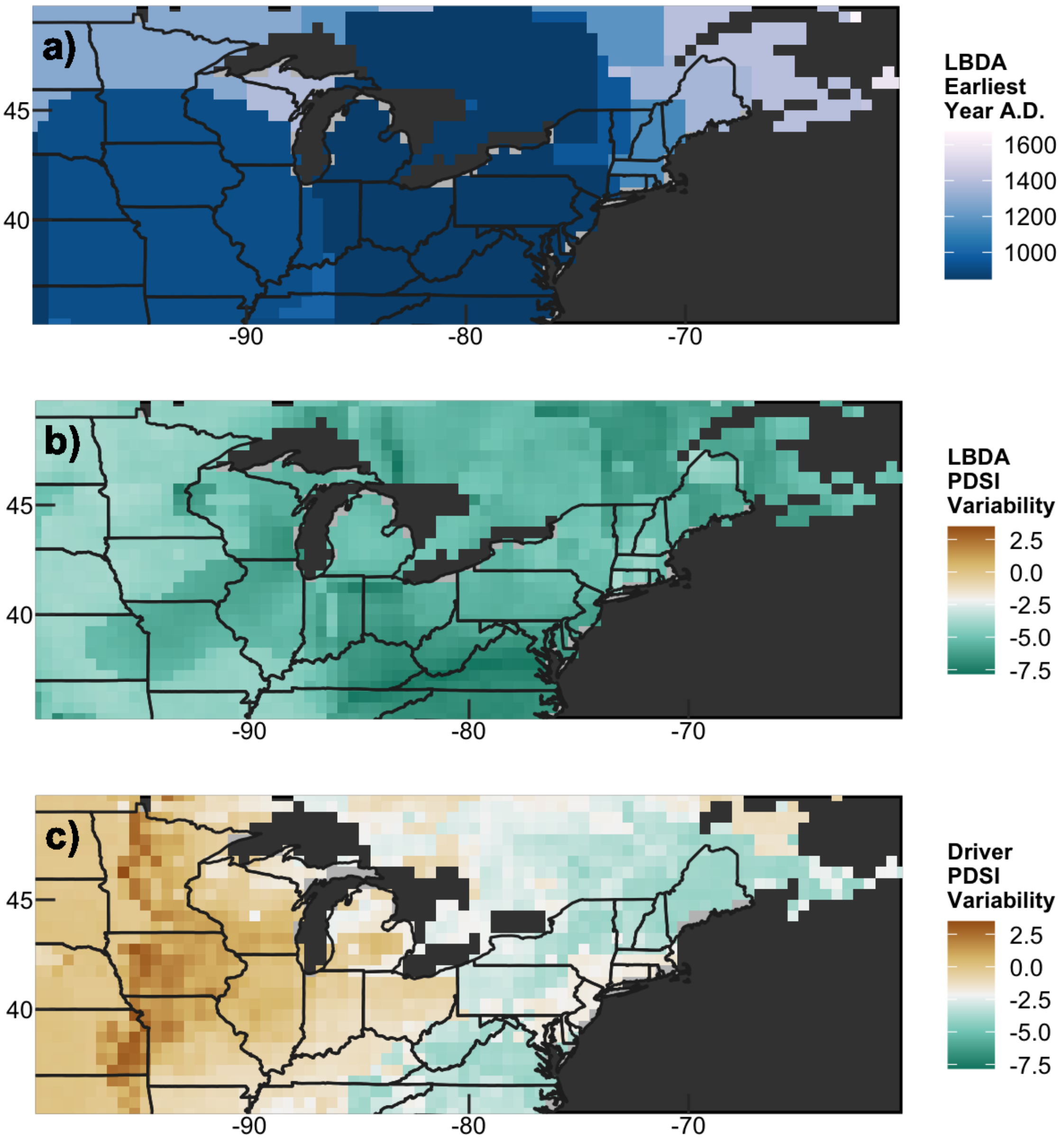
Figure 4: Diagnosing the observed and latent relationships among ecohydrological variability and variability in forest composition, structure, and function in five terrestrial ecosystem models. Models in the top row (LPJ-GUESS, LINKAGES, ED2) have individual- or cohort-level representation of demography whereas those in the bottom row (LPJ-WSL, JULES-TRIFFID) do not simulate demography. All models showed positive correlations between composition and drought variability, but some models showed positive biomass sensitivities (LINKAGES, JULES-TRIFFID) while others were negative (ED2, LPJ-WSL, LPJ-GUESS). In all models, composition sensitivity to hydroclimate variability was most similar to NPP whereas biomass sensitivity tended to mirror NEE. In the log scale, negative values indicate locations with low variability whereas more positive values indicate high variability.

Supplemental Tables

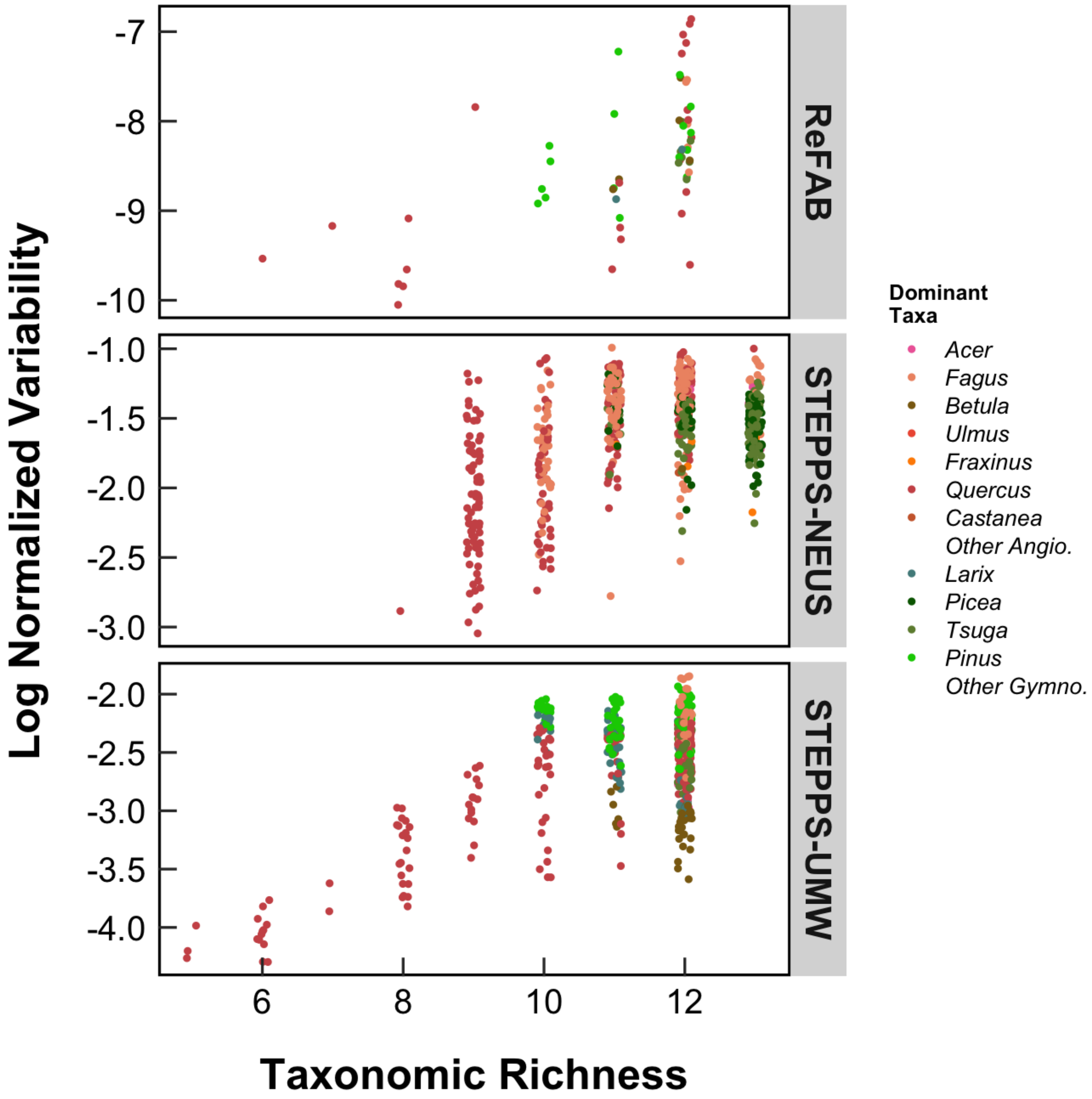
Supplemental Table 1: Sensitivity of latent state variability to hydroclimate (PDSI) variability in ecosystem models and pollen data products. Sensitivity is presented as the mean and standard error slope from log-log regression; * indicates slopes significantly different from zero ($p<0.05$). LINKAGES does not simulate GPP. LAI output was not available for LPJ-WSL.

Model	**GPP**	**NPP**	**NEE**	**LAI**	**Biomass**	**Composition**
Pollen					-0.156 (0.119)	0.026 (0.019)
ED2	0.201 (0.028)*	0.190 (0.025)*	0.008 (0.024)	0.203 (0.024)*	-0.079 (0.015)*	0.118 (0.017)*
LPJ-WSL	0.320 (0.033)*	0.301 (0.033)*	0.010 (0.034)		-0.020 (0.034)	0.252 (0.029)*
LPJ-GUESS	-0.022 (0.031)	0.038 (0.034)	-0.152 (0.031)*	-0.034 (0.022)	-0.069 (0.020)*	0.056 (0.015)*
LINKAGES		0.186 (0.027)*	0.232 (0.030)*	0.222 (0.031)*	0.230 (0.033)*	0.074 (0.016)*
JULES-TRIFFID	0.294 (0.051)*	0.365 (0.051)*	0.110 (0.028)*	0.295 (0.035)*	0.203 (0.038)*	0.411 (0.033)*

Supplemental Figures



Supplemental Figure 1: Comparison of log normalized PDSI variability in empirically-inferred reconstructions from the Living Blended Drought Atlas (LBDA, 41, a, b) and model drivers (c). Due to the regional differences in the length of tree-ring chronologies available for PDSI reconstruction, the temporal extent of analyses involving LBDA drought is uneven across space. Overall, model drivers had greater PDSI variability than seen in the LBDA, but both datasets show greater variability in the western region of the study domain.



Supplemental Figure 2: Relationship between taxonomic richness and log normalized biomass (ReFAB) and composition (STEPPS) variability in pollen-inferred datasets.