Emergent climate and CO2 sensitivities of net primary productivity in ecosystem models do not agree with empirical data in temperate forests of eastern North America

Rollinson, C. R., Liu, Y., Raiho, A., Moore, D. J. P., McLachlan, J., Bishop, D. A., Dye, A., Matthes, J. H., Hessl, A., Hickler, T., Pederson, N., Poulter, B., Quaife, T. ORCID: https://orcid.org/0000-0001-6896-4613, Schaefer, K., Steinkamp, J. and Dietze, M. C. (2017) Emergent climate and CO2 sensitivities of net primary productivity in ecosystem models do not agree with empirical data in temperate forests of eastern North America. Global Change Biology, 23 (7). pp. 2755-2767. ISSN 1354-1013 doi: https://doi.org/10.1111/gcb.13626 Available at https://centaur.reading.ac.uk/68717/

To link to this article DOI: http://dx.doi.org/10.1111/gcb.13626

Publisher: Blackwell Publishing

Title: 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

Running Head: Model and data NPP sensitivity to climate

Authors

Rollinson, Christine R.[1,2]*, Yao Liu[3], Ann Raiho[4], David J.P. Moore[3], Jason McLachlan[4], Daniel A. Bishop[5], Alex Dye[6], Jaclyn Hatala Matthes[7], Amy Hessl[6], Thomas Hickler[8,9], Neil Pederson[5], Benjamin Poulter[10], Tristan Quaife[11], Kevin Schaefer[12], Jörg Steinkamp[8], Michael C. Dietze[1]

Author Affiliations

1. Department of Earth & Environment, Boston University, Boston, MA USA
2. Morton Arboretum, Lisle, IL, USA
3. School of Natural Resources and the Environment, University of Arizona, AZ, USA
4. Department of Biological Sciences, University of Notre Dame, South Bend, IN, USA
5. Harvard Forest, Harvard University, Petersham, MA, USA
6. Department of Geology and Geography, West Virginia University, Morgantown, WV, USA
7. Department of Biological Sciences, Wellesley College, Wellesley, MA USA
8. Senckenberg Biodiversity and Climate Research Centre (BiK-F), Frankfurt/Main, Germany

9. Department of Physical Geography, Geosciences, Goethe University, Frankfurt/Main, Germany
10. Institute on Ecosystems and Department of Ecology, Montana State University, Bozeman, MT 59717, USA
11. Department of Meteorology, University of Reading, Reading, BERKS, UK.
12. National Snow and Ice Data Center, Cooperative Institute for Research in Environmental Sciences, University of Colorado at Boulder, Boulder, CO, USA

*** Corresponding Author:** crollinson@mortonarb.org, +1 630-719-2422

Keywords: climate sensitivity, emergent response, climate change, net primary productivity, ecosystem model, model-data comparison, paleoecology, tree rings

Paper Type: Primary research article

Abstract

Ecosystem models show divergent responses of the terrestrial carbon cycle to global change over the next century. Individual model evaluation and multi-model comparisons with data have largely focused on individual processes at sub-annual to decadal scales. Thus far, data-based evaluations of emergent ecosystem responses to climate and CO_2 at multi-decadal and centennial time scales have been rare. We compared the sensitivity of net primary productivity (NPP) to temperature, precipitation, and CO_2 in ten ecosystem models with the sensitivities found in tree-ring reconstructions of NPP and raw ring-width series at six temperate forest sites. These model-data comparisons were evaluated at three temporal extents to determine whether the rapid, directional changes in temperature and CO_2 in the recent past skew our observed responses to multiple drivers of change. All models tested here were more sensitive to low growing season precipitation than tree-ring NPP and ring widths in the past 30 years, although some model precipitation responses were more consistent with tree rings when evaluated over a full century. Similarly, all models had negative or no response to warm growing season temperatures while tree-ring data showed consistently positive effects of temperature. Although precipitation responses were least consistent among models, differences among models to CO_2 drive divergence and ensemble uncertainty in relative change in NPP over the past century. Changes in forest composition within models had no effect on climate or CO_2 sensitivity. Fire in model simulations reduced model sensitivity to climate and CO_2, but only over the course of multiple centuries. Formal evaluation of emergent model behavior at multi-decadal and multi-centennial time scales is essential to reconciling model projections with observed ecosystem responses to past climate change. Future evaluation should focus on improved representation of

disturbance and biomass change as well as the feedbacks with moisture balance and CO_2 in individual models.

Introduction

Changes in temperature and precipitation regimes over the past millennium have been associated with shifts in ecosystem composition and structure in the paleoecological record (Prentice *et al.*, 1991; Davis & Shaw, 2001; Shuman *et al.*, 2002; Clifford & Booth, 2015). Modern empirical data from experiment- and observation-based studies provide evidence that rapidly increasing temperatures, altered precipitation regimes, and rising atmospheric CO_2 concentrations are causing changes in ecosystem dynamics today (Boisvenue & Running, 2006; Morin *et al.*, 2009; Fisichelli *et al.*, 2013; Peñuelas *et al.*, 2013). However, the effects of climate change in long-lived ecosystems such as the temperate forests of the eastern and midwestern United States remain contested because changes in forest composition and structure take decades to centuries to occur (Renwick & Rocca, 2014). These observation-based studies capture emergent, long-term ecosystem responses to climate and CO_2 variability that are the product of feedbacks and interactions among physiological and biogeochemical processes.

Terrestrial ecosystem models are used to make mechanistic, process-based projections of ecosystem response to changing climate and CO_2 in the past, present, and future. However, the mechanistic, bottom-up approach used to build and evaluate ecosystem models is mismatched in spatial and temporal scale from observations of ecosystem response to climate change. Differences in model parameterization or structural representation of physiological process can cause major divergences in the resulting ecosystem dynamics through time (De Kauwe *et al.*, 2013; Walker *et al.*, 2015). Model intercomparisons of ecosystem response to climate change over the next century typically show increased divergence of ecosystem dynamics at the end of the century (e.g. Friedlingstein *et al.*, 2014). Although most policy and management decisions occur at sub-centennial temporal scales, life spans of common temperate tree species in the northeastern United States can range from 120 years for *Betula papyrifera* to over 800 years for *Tsuga canadensis* (Burns & Honkala, 1990). This means that in the absence of widespread disturbance or management, even the model simulations of ecosystem response to climate change are far shorter than the multi-generational scales at which changes in forest composition or structure in response to shifts in climate occur. These slow, gradual shifts in forest composition or structure may mediate ecosystem sensitivity to climate and CO_2 in both models and reality. For example, mortality and recruitment are inherent processes underlying the types of ecosystem-scale responses to climate change observed in the past, and occur at temporal and spatial scales beyond that which we can observe with current available ecological data. This can then cause the observed responses of ecosystems to past climate change over multiple centuries to be different from that observed in shorter studies focused on a few decades.

Tree rings provide annually resolved records of individual and forest response to environmental variation over the temporal scales of decades and centuries across temperate forests. Tree rings have been used to understand forest responses to climate variability at individual sites and entire continents (Williams *et al.*, 2013; Charney *et al.*, 2016; D’Orangeville *et al.*, 2016), but these approaches have been largely disconnected from assessments of how climate change impacts ecosystems in models. Tree rings are increasingly used to quantify and constrain components of the terrestrial carbon cycle in forests as well as individual- and forest-level responses to climate (Graumlich *et al.*, 1989; Davis *et al.*, 2009; Babst *et al.*, 2013; Dye *et al.*, 2016). These efforts are critical for understanding long-term forest responses to climate variability as multiple field experiments have indicated that long-term individual and ecosystem responses to warming and elevated CO_2 diverge from initial responses found in the first few years of manipulation (Melillo *et al.*, 2002; Hollister *et al.*, 2005; Matesanz *et al.*, 2009; Norby *et* *al.*, 2010). This long-term evaluation of climate impacts on forest ecosystems is particularly important for disentangling the simultaneous directional shifts in temperature, precipitation, CO_2, and disturbance that complicate much of modern ecological research (Foster *et al.*, 1998; Turner *et al.*, 2003; Gómez-Aparicio *et al.*, 2011).

This paper compares emergent sensitivity of net primary productivity (NPP) to climate and CO_2 in ecosystem models to those found in tree-ring data at multiple temporal scales. The goals of this paper are: 1) determine whether annual ecosystem NPP sensitivity to climate and CO_2 in ecosystem models matches those found in individual tree- and forest-level tree-ring data; 2) compare the climate and CO_2 sensitivities in models and data from short temporal extents where patterns are dominated by inter-annual climate variability with those from centennial-scale records that contain low-frequency climatic shifts; and 3) quantify the contribution of differences in model responses to climate and CO_2 to model ensemble uncertainty in NPP dynamics through time. To achieve these goals, we analyze the temporal trends of NPP in ten ecosystem models that have been run from 850 to 2010 A.D. at six temperate forest sites in the Upper Midwestern and Northeastern United States. The influences of growing season temperature, precipitation, and CO_2 on change in model NPP are then compared to those found in plot-level NPP reconstructions from tree rings at two sites as well as raw ring widths from plot-based sampling and the International Tree Ring Databank (ITRDB) at four sites. Climate and CO_2 responses are analyzed for three temporal extents: 1) the scale of past paleoecological responses to climate change (850-2010 A.D.); 2) the period of historical temperature and precipitation records (1901-2010 A.D.); and 3) the scale of robust modern ecological data (1980-2010 A.D.). We use these results to explain model-data discrepancies in regional drivers of NPP through time as well as quantify the drivers of uncertainty within the model ensemble over the past millennium.

Materials and Methods

Model experimental overview

We modeled ecosystem carbon flux and composition change in response to climate variation from 850-2010 A.D. at six sites in the Northeastern and Upper Midwestern United States. These sites are located on an east-west gradient that represent present-day differences in temperature, precipitation, and forest types (Table 1, S1). Models were presented with common, continuous 6-hourly meteorological climate and monthly CO_2 forcing data. CO_2 was taken from the Law Dome time series (850-2000) and the NOAA Mauna Loa record (2001-2010) (Keeling *et al.*, 2005; Wei *et al.*, 2014). Climate drivers included air temperature, water-equivalent precipitation rate, total incoming shortwave radiation, incoming longwave radiation, surface air pressure, specific humidity, and wind speed. The continuous, 6-hourly 850-2010 meteorological driver set was developed by using an artificial neural network to spatially and temporally downscale output from CCSM4 output from the Paleoclimate Modeling Intercomparison Project, Phase III (PMIP3) past millennium simulations (850-1849, Crucifix *et al.*, 2012) and Coupled Model Intercomparison Project, Phase 5 (CMIP5) simulations (1850-1900, Taylor *et al.*, 2012) using 6-hourly, 0.5-degree CRUNCEP data (1901-2010, Wei *et al.*, 2014) according to Kumar *et al.*, (2012). Due to mismatches at transitional periods between CCSM4 and CRUNCEP products, all variables except wind were bias-corrected to avoid sharp jumps in climatology in the time series. Temperature was corrected as an additive bias whereas short- and long-wave radiation, and precipitation were corrected using a ratio bias to conserve the hydrological sensitivity among meteorological variables (Hempel *et al.*, 2013).

Ten models representing variants of five independent ecosystem models completed simulations of ecosystem dynamics at all six sites from 850-2010 A.D. and provided monthly- or annual-resolution output on composition and the carbon cycle (Table 2). Full model protocol can be found in Supporting Information 2. Dynamic vegetation models allow plant communities to self-assemble and change through time while static vegetation models prescribe plant functional types (PFTs) by fractional area based on potential vegetation distribution from Ramankutty and Foley (1999). All models were spun-up to steady-state at 850 A.D. by cycling the first 20 years of forcing data (850-869 A.D.) with a constant CO_2 concentration of 277 ppm. Not all models used all meteorological variables as drivers. LINKAGES was the only model to not include CO_2 as a driver and only includes temperature and precipitation drivers whereas all other models included at a minimum, temperature, precipitation, shortwave radiation, and CO_2. Even though LINKAGES does not include CO_2 as a driver of ecosystem dynamics, it was included in our analyses to help isolate potential confounding effects of increasing CO_2 with other drivers of ecosystem change.

Our analyses focus on two model characteristics (dynamic vegetation and fire) and the dynamics of two ecosystem properties (composition, aboveground biomass) that are linked to slow processes associated with ecosystem responses to climate and CO_2 change (Table 2). However, because PFTs in all models except for SiBCASA, a biome-based model, respond to climate independently, they may still experience shifts in the relative community composition if defined by fluctuations in PFT biomass. Composition variability through time was quantified as the standard deviation of percent evergreen biomass through time within each site for each model. Although fire was possible in ED, CLM, and LPJ model variants, the conditions for fire were not triggered in ED2 and ED2-LU, so the effects of fire were only present in four models

(CLM-BGC, CLM-CN, LPJ-WSL, LPJ-GUESS). Fire in ecosystem models primarily affects ecosystems by reducing biomass, but mortality or changes in growth can also cause similar biomass fluctuations in all models. To calculate biomass variability, biomass was first normalized to the site mean so that biomass through time was expressed as a percent of mean. Biomass variability was then calculated as the standard deviation of this normalized biomass through time at each site for each model. All models except JULES-STATIC used aboveground biomass (AGB) as the measure of biomass variability. Leaf area index was used as a proxy for biomass in JULES-STATIC because there is no biomass or vegetation dynamics beyond leaf area simulated in JULES without coupling to the TRIFFID model. In other models LAI had a mean correlation with AGB of 0.73 ± 0.18 and ranged from 0.55 in CLM-BGC to 0.99 in LINKAGES. In JULES-TRIFFID, the correlation between LAI and AGB was 0.93.

Empirical Data

We used two types of tree-ring data to provide an empirical estimate of climate sensitivity for trees and forests from our study region. First, stand-level aboveground NPP was reconstructed from sampling mapped fixed-area plots according to Dye *et al.*, (2016), which enable aboveground NPP to be calculated on a per area basis that is comparable to ecosystem model output (Table 1). NPP was reconstructed from five plots in two separate stands at Harvard and three plots in a single stand at Howland. A nested sampling scheme was used at both locations where two to three increment cores were taken from all trees greater than 10 cm diameter at breast height (DBH) in a 13-m radius plot and trees greater than 20 cm DBH were cored in a 20-m radius plot. Growth measurements from all cores for each were averaged and used to reconstruct DBH through time that was then converted to aboveground biomass and annual aboveground biomass increment (proportional to NPP) for each tree and plot using species-specific equations that can be found in the appendix of Dye *et al.*, (2016). Due to the potential decline in sample replication, and the potential for unmeasured mortality, analyses of tree-ring NPP sensitivity to climate were restricted to 1980-2010 (Clark *et al.*, 2001; Foster *et al.*, 2014).

In addition to tree-ring estimates of NPP, we also analyzed the sensitivity of raw ring width (RW) to climate from the trees used to generate the NPP estimates as well as from nine datasets from the International Tree Ring Databank (ITRDB) (Grissino-Mayer & Fritts, 1997). Records from the ITRDB were found in close proximity to the Demming Lake, Howland, and UNDERC sites, allowing more robust comparisons of climate sensitivity across the modeled study sites (S3). In cases where multiple cores existed for a tree, the mean of ring width for each year was used. Although we include supplemental sensitivity analyses using *a priori* detrended ring width index (RWI, S4), we chose raw ring widths as our response variable rather than RWI or basal area increment (BAI) for two reasons. First, we chose not to detrend rings widths *a priori* for our primary analyses in order to preserve potential low-frequency effects of changing climate or CO_2 in the modern era that might be confounded with increasing stem size. Second, BAI was not used because measurements of actual diameters were not available for ITRDB trees. The CRUNCEP climate records used to drive the models were matched to the tree-ring data for each site.

Sensitivity Analysis

Model and tree-ring sensitivities to climate were analyzed using generalized additive models (GAMs) where NPP or RW is the sum of non-linear temperature, precipitation and CO_2 effects. In these analyses, we used site-level annual NPP as the response variable for ecosystem models, plot-level annual aboveground woody increment for tree-ring NPP, and mean raw ring width for each individual in the RW analysis. Temperature, precipitation, and CO_2 effects were estimated with three-knot thin plate regression splines while an additional size effect was fit with three-knot cubic smoothing spline using the *gam* function of the *mgcv* package in R 3.2.3 (Wood, 2012). Even though LINKAGES lacks CO_2 as a model driver, this effect was included in our statistical analyses as a test of the statistical model's ability to attribute variation in NPP to climate and CO_2. The size term was included to account for effects of biomass (models, tree-ring NPP) or individual size (raw ring width) that could alter the maximum potential NPP or ring width independent of climate. Total site or plot aboveground biomass was used for the size effect in models and tree-ring NPP GAMs, while reconstructed DBH was used for size in the RW GAMs. DBH was reconstructed by subtracting ring widths from the DBH at the time of sampling at Harvard and Howland, but was reconstructed by summing ring widths from the ITRDB records, where no DBH information was available. Because ring width sums may not accurately reflect the true DBH of ITRDB trees, in RW models the size effect was fit separately for each ITRDB core. This effectively individually detrends ITRDB cores by removing effects of increasing size through time while simultaneously assessing the climate and CO_2 sensitivity of both ITRDB and plot-based samples. We also performed supplemental analyses with *a priori* detrended RWI using a two-thirds spline that removes long-term trends in tree-ring series (S5). Because *a priori* detrending removes age- and size-based trends in the data, the size factor was removed from our climate and CO_2 GAM in RWI analyses. We also include an analysis with all datasets restricted to Harvard and Howland sites where tree-ring NPP records are available (S6).

We normalized the sensitivities of model NPP and tree-ring width to climate and CO_2 for all analyses to facilitate comparison of model agreement of change in NPP through time and relative influences of temperature, precipitation, and CO_2 as drivers of that change. In all cases, the response variable of NPP or RW was converted to a percent deviation from the model or tree-ring dataset mean. We assessed model agreement of both change in NPP and drivers of change as the standard deviation around the ensemble mean before and after 1901, where CO_2 begins to sharply increase. To prescribe the effects of individual drivers on change in NPP or RW through time, we quantified the relative effects of temperature, precipitation, and CO_2 on NPP or RW in each year. The CO_2 effect was offset from the GAM-estimated effect so that the beginning of simulation CO_2 concentration in 850-855 A.D. received a weight of zero.

Models were analyzed at three temporal scales to determine whether observations from short periods in the modern era were able to capture the relationships between climate and CO_2 change and ecosystem responses over periods with low-frequency, directional shifts in climate. These three temporal extents are: 1) scale of robust, modern empirical records (1980-2010), 2) the scale of observational climate data (1901-2010), and 3) the full model simulation extent (850-2010). This model-centric analysis was complemented with comparisons to the shifts in sensitivity seen in the RW models at the two shorter scales. Because GAMs center the spline-based effects on the means of the given data (i.e. the mean of observed temperature, precipitation, and CO_2), all three temporal scales were re-centered on their respective 1980-2010 means to facilitate cross-scale comparisons. Differences in climate and CO_2 sensitivities within and across temporal scales both within individual models and at the ensemble level were assessed by comparing the mean slopes and 95% confidence intervals around each effect. We used an ANOVA to compare variability in the model ensemble across temporal scales. In this ANOVA we compared the absolute deviation of mean climate and CO_2 sensitivity for each model to that of the ensemble mean at each scale.

We also quantified the effects of slow ecosystem processes among models on NPP sensitivity to climate and CO_2 using linear regression. Here, we considered two categorical characteristics of models as well as two related continuous variables to identify trends in model dynamics and sensitivity to climate and CO_2. Categorical model characteristics included vegetation scheme (static or dynamic) and the presence of fire (yes or no) while composition (fraction evergreen) and biomass variability described similar model dynamics as continuous effects. Each characteristic of slow ecosystem processes was correlated with the mean slopes of model sensitivities to temperature, precipitation and CO_2. We performed this correlation analysis at all three temporal scales to determine if the effects of these slow processes were significant at particular, characteristic scales.

Results

Spatial and temporal patterns of NPP in ecosystem models

Terrestrial ecosystem models disagreed about both the spatial patterns of NPP in the Northeastern and Upper Midwestern United States as well as the change in NPP through time (Fig. 1). Midwestern sites tended towards lower NPP than those further east with ensemble-mean NPP ranging from 3.60 ± 2.14 MgC m^{-2} yr^{-1} (mean ± SD) at Billy's Lake to 6.0 ± 2.5 MgC m^{-2} yr^{-1} at Harvard Forest. ED-LU had the highest mean NPP of the ensemble (7.12 ± 2.80 MgC m^{-2} yr^{-1}) and JULES-STATIC had the lowest (1.51 ± 1.17 MgC m^{-2} yr^{-1}), but this pattern varied across individual sites and times. Tree-ring estimates of NPP (NPP_{TR}) at both Harvard and Howland were much lower than predicted by models (Fig. 1). From 1980-2010, NPP_{TR} was 1.73 ± 0.27 MgC ha^{-1} yr^{-1} at Harvard and 1.17 ± 0.13 MgC ha^{-1} yr^{-1} at Howland. The mean NPP of the model ensemble from 1980-2010 was 7.47 ± 3.73 MgC ha^{-1} yr^{-1} at Harvard and 7.07 ± 3.93 MgC ha^{-1} yr^{-1} at Howland. NPP_{TR} was lower than the lowest model NPP at both sites (2.90 ± 0.38 MgC ha^{-1} yr^{-1} at Harvard, 3.48 ± 0.30 at Howland MgC ha^{-1} yr^{-1}).

To facilitate comparisons of the impacts of climate and CO_2 on NPP, through time, we analyzed percent change in NPP relative to each model's mean. Once relativized, models and tree rings displayed similar levels of variability in NPP and ring width: model NPP variability was 16 ± 8%, NPP_{TR} was 12%, and RW was 20%. Even though ED2-LU had the highest NPP variability in absolute terms, with a standard deviation through time of 1.83 MgC m^{-2} yr^{-1}, but JULES-STATIC displayed higher temporal NPP variability relative to its mean (29%). In contrast, SiBCASA showed the least absolute and relative NPP variability through time: 0.28 MgC m^{-2} yr^{-1}, 5% of its mean. The higher relative NPP variability of JULES-STATIC was accompanied by the most stable biomass (temporal standard deviations of 3%) while JULES-TRIFFID showed the least temporal variability of composition with only a mean 2% variability through time (Table 2). LPJ-GUESS, whose mean NPP and NPP variability were close to the ensemble mean, had the highest variability of both composition and biomass (35% and 24%, respectively).

NPP and RW sensitivity to climate and CO_2

Comparisons between models and tree-ring data at sub-centennial temporal scales reveal large disparities between the effects of climate and CO_2 on NPP and tree growth (Fig. 2). At the

1980-2010 temporal extent, the ensemble of models and tree-ring data agreed on a positive relationship between NPP or growth and precipitation, but mean model sensitivity (0.09 ± 0.05% mm^{-1} yr^{-1}) was nine times higher than both tree-ring NPP and ring widths (both 0.01% mm^{-1} yr^{-1}). The overall signs of sensitivity were opposite between models and tree-ring data for temperature (model ensemble = -0.99 ± 4.03% $˚C^{-1}$, RW = 11.0% $˚C^{-1}$, NPP_{TR} = 14.2% $˚C^{-1}$) and CO_2 (model ensemble = 0.20 ± 0.16% ppm^{-1}, RW = -0.7% ppm^{-1}, NPP_{TR} = 0.01% ppm^{-1}). However, the 95% confidence interval for NPP_{TR} does encompass 0 and is consistent with relatively CO_2-insensitive models such as LPJ-WSL. *A priori* detrending of ring width to ring width index dramatically reduced temperature, precipitation, and CO_2 sensitivity in the individual-based tree-ring analysis, causing temperature and CO_2 sensitivities to be more consistent with model sensitivities, but less consistent with NPP_{TR} (S5). Trends within and among the sensitivities of models and tree-ring data were similar between analyses including all sites and when analyses were restricted to just Harvard and Howland (S6). Increasing temporal scale had relatively minor effects on climate sensitivity in model NPP and increased sensitivity in RW (Fig. 2, S7). This resulted in greater consistency among precipitation responses in RW and models, but continued model-data disagreement in temperature and CO_2 at the 1901-2010 scale. Temperature sensitivity of NPP in LINKAGES was a notable exception to this trend, and showed similar positive effects of lower growing season temperatures as the tree-ring datasets, but then diverged and displayed negative effects of warm temperatures.

Spatial and temporal variability in NPP among models corresponded to differences among models in emergent ecosystem sensitivities to temperature, precipitation, and CO_2. At the full temporal extent of our model simulations (850-2010), models showed consistent positive effects of precipitation and CO_2 on NPP, although with varying magnitudes of the effect (Fig. 2).

For precipitation, the ensemble of models had an average 0.13 ± 0.11% increase in NPP per millimeter precipitation, but individual models showed sensitivities ranging from 0.02 ± 0.03% mm^{-1} yr^{-1} in SiBCASA to 0.32 ± 0.16% mm^{-1} yr^{-1} in JULES-STATIC. Ensemble NPP sensitivity to CO_2 was 0.36 ± 0.34% ppm^{-1} with individual effects in models with CO_2 ranging from 0.06 ± 0.04% ppm^{-1} in CLM-BGC to 1.03 ± 0.20% ppm^{-1} in JULES-STATIC. LINKAGES, which lacks CO_2 effects in its model structure, supports the ability of the GAM approach to correctly attribute impacts of climate and CO_2 on NPP by displaying a small NPP response of -0.01 ± 0.00% ppm^{-1}. Temperature was the only effect included in models to show differences in the direction of NPP sensitivity. Most models had a generally negative temperature effect and the ensemble mean effect was a decrease of 2.37 ± 9.14% in NPP per ˚C. JULES-STATIC showed the strongest effect with a 23.48 ± 0.89% decrease per ˚C. Overall, LINKAGES had a positive temperature effect (14.77 ± 19.43% per ˚C), although this effect became negative at higher temperatures. Over the full course of model simulations, the presence of fire was correlated with decreased sensitivity to temperature ($t=-2.3$, $p=0.03$) and CO_2 ($t=-2.8$, $p<0.01$). Decreased model NPP sensitivity to temperature was also associated with increased biomass variability through time ($t=-2.7$, $p=0.01$). Models with more variable composition over multiple centuries also tended to be less sensitive to CO_2, but this trend was not significant ($t=-2.0$, $p=0.06$).

An analysis of variance indicated there was greater agreement among model sensitivities to precipitation and CO_2 at the two shorter temporal scales than at the multi-centennial (850-2010) extent (Fig. 2; precipitation: $t=4.6$, $p<0.01$, CO_2: $t=4.0$, $p<0.01$), but temperature showed similar ensemble variability across temporal scales ($p>0.05$). The presence of fire in ecosystem models correlated with reduced sensitivity to temperature, precipitation, and CO_2, but only at the multi-centennial temporal scale (Table 3). Slow ecosystem processes had weakened effects on model NPP sensitivity to climate and CO_2 at the shorter temporal scales of empirical data and were not significantly associated with patterns in model sensitivities at these multi-centennial scales (Table 3). Similarly, models with high temporal variability of biomass were also less sensitive to temperature, but only over the course of multiple centuries. Factors associated with changes in composition including composition stability and whether a model had dynamic or static vegetation had no correlation with climate sensitivity in our model ensemble.

Drivers of model ensemble patterns agreement through time

Differences in model sensitivity to increasing CO_2 explained increasing model ensemble uncertainty in change in NPP since 1900 (Fig. 3). Model ensemble variability of normalized NPP prior to 1900 was 11.0 ± 5.3% (Fig. 3b). After 1900, this variability more than doubled to 20.6 ± 7.4%. This shift in model agreement occurred as NPP in most models shifted from being primarily influenced by precipitation to showing varying degrees of CO_2 enhancement (Fig. 3b, c). In the 1980-2010 extent, the effect of CO_2 on NPP in models that include dynamic CO_2 ranged from +2% to +110% with model variability in CO_2 response proportionate to that of inter-model NPP deviation during this time period (CO_2 variability = 29 ± 3%; NPP variability = 28 ± 5%, Fig. 3c). LINKAGES was the only model to not include a CO_2 driver and our statistical method misattributed a slight 0.6% decrease in NPP to increasing CO_2.

Discussion

Analysis of emergent responses to temperature, precipitation, and CO_2 at multi-decadal and multi-centennial time scales reveals many inconsistencies among models and data. The ten terrestrial ecosystem models included in our study showed variability in both mean NPP across space and relative changes in NPP through time (Fig. 1). Similar discrepancies have been widely observed in other multi-model comparisons, which have attributed differences among models to numerous causes including ecosystem feedbacks and uncertainties in model processes and parameterization (Piao *et al.*, 2013; Friedlingstein *et al.*, 2014; Walker *et al.*, 2014). We found widely varying patterns of climate and CO_2 effects among models and data across multiple temporal scales (Fig. 2). However, the ensemble of models displayed less variation in climate and CO_2 responses at short temporal scales. Small differences in model structure and parameterization compound over time and drive model divergence, skewing model projections at ecology- and policy-relevant timescales. Fire and changes in biomass are only associated with differences in precipitation and CO_2 sensitivity among models at the multi-centennial scale, indicating that feedbacks involving disturbance and biomass only have discernible impacts on climate and CO_2 over multiple forest generations (Table 3). Although model responses to precipitation are the most variable, responses to CO_2 are the greatest source of ensemble divergence of the past 100 years.

Models and empirical datasets have widely varying estimates of baseline NPP, so standardization is important to compare the impacts of climate and CO_2 through time across datasets. To be consistent across all models in the ensemble, we analyzed total ecosystem NPP, which includes the total carbon allocated to both above- and belowground tissues. Most models in our ensemble simulated NPP values within the range of what has been observed for forests common in the Northeastern and Upper Midwestern United states. Representative field-based estimates of NPP for our study region range from 1.3 MgC ha^{-1} yr^{-1} in mature boreal forests to as high as 8.7 MgC ha^{-1} yr^{-1} in the oak-hickory forests that dominate further south when a 50%

carbon content of biomass is assumed (Jenkins *et al.*, 2001; Luyssaert *et al.*, 2007; Goulden *et al.*, 2011). Even if the models were perfectly accurate, their values would still be greater than the aboveground-only NPP quantified through tree rings in our study, which is estimated to be between 45 and 65% of the total carbon. Furthermore, our tree-ring productivity estimates only include the aboveground biomass increment from one year to the next, which will not capture any biomass that is produced and turns over within a year. A recent comparison of tree-ring estimates of NPP at Harvard Forest indicated good agreement with repeat forest censuses (Eisen & Plotkin, 2015; Dye *et al.*, 2016), but are substantially lower than values estimated through remote sensing for coniferous and deciduous forests at the same site (5.5 and 6.8 MgC ha^{-1} yr^{-1}, respectively; Turner *et al.*, 2005). In order to reconcile total and aboveground NPP estimates, better quantification of amounts and controls of allocation and turnover among tissues is essential (Jenkins *et al.*, 2001; Litton *et al.*, 2007; Luyssaert *et al.*, 2007). Some empirical studies suggest that the fraction of NPP allocated to aboveground wood could vary from year to year due to tree maturation, increasing CO_2, or interannual climate variability (DeLucia *et al.*, 2005; Norby *et al.*, 2005; Doughty *et al.*, 2014). Although some ecosystem models allocate carbon in response to changing resource limitations (De Kauwe *et al.*, 2014), most do not and the models used in this analysis are based on simple allometric rules. Despite these inconsistencies in NPP values among models and empirical datasets, it is informative to compare the emergent sensitivities of productivity to climate and CO_2 across datasets.

Emergent model NPP sensitivities to temperature and precipitation are not supported by the responses observed in plot or individual tree-ring records. Specifically, most models were over-sensitive to precipitation relative to tree rings and had predominantly negative or no response to temperature whereas both tree-ring NPP and raw rings widths showed consistently positive responses (Fig. 2). The strongly negative effect of temperature coupled with a strongly positive effect of precipitation in the models suggests that the ecosystem models in our ensemble may be over-sensitive to both temperature- and precipitation-driven droughts despite rapid post-drought recovery seen in many ecosystem models (Anderegg *et al.*, 2015). Although the tree rings in our study show positive correlations with temperature in both raw ring width and *a priori* detrended ring width index, other studies provide support for negative effects of high temperatures on tree growth in the eastern United States (e.g. Rollinson *et al.*, 2016). Nonetheless, models appear to be lacking moderating feedbacks that can cause positive temperature effects at low temperatures or reduce precipitation sensitivity. Indeed, the model in our ensemble that most closely matches the empirical relationship between temperature and NPP is LINKAGES, which calculates NPP directly from growing degree days rather than from separate GPP and autotrophic respiration functions (Post & Pastor, 1996).

Models were also over-sensitive to precipitation relative to tree rings. Traditional tree-ring analyses have focused on growth responses to drought indices such as the Palmer Drought Severity Index that combines temperature, precipitation, and soil moisture into a single drought metric (e.g. Speer *et al.*, 2009; McEwan *et al.*, 2011). However, process-based ecosystem models drive productivity from independent forcing of temperature and precipitation and temperature- versus precipitation-based droughts may impact ecosystems differently (Anderegg *et al.*, 2013). Furthermore, models can vary in their representation of soil and hydrology so that the same temperature and precipitation drivers may result in different droughts both among models and compared to empirical system (Cook *et al.*, 2015). Tree-ring studies that have used mixed-modeling frameworks to look at the simultaneous influences of temperature and precipitation on growth have revealed stronger influences of temperature on growth than precipitation (Rollinson *et al.*, 2016). In order to improve the ecosystem-scale model responses to both temperature and precipitation, further empirical research is needed to separate temperature and precipitation effects on tree growth.

Model responses to precipitation were more variable than temperature or CO2 and consequently there was little consensus in our model ensemble about when and where ecosystem productivity was driven by temperature versus precipitation (Figs. 2 & 3). Despite sensitivity disparities with tree-ring data, the strong control of precipitation in many models at the multi-centennial scale supports hypotheses that drought may have been instrumental in driving ecosystem dynamics and species shifts in New England 500-600 years before present (Clifford & Booth, 2015). However, the relative insensitivity of productivity to temperature is difficult to reconcile with apparent temperature-driven millennial-scale compositional shifts in the paleoecological record (Prentice *et al.*, 1991; Blois *et al.*, 2013). In the model ensemble, fire and changes in biomass variability reduce model sensitivity to temperature rather than changes in relative composition, which suggests this lack of temperature sensitivity is not due to stabilizing shifts in plant functional types. Challenges of recreating spatial and temporal patterns of observed composition is a pervasive problem in ecosystem modeling and is known to have cascading impacts that bias other aspects of ecosystem dynamics including NPP and transpiration (Matthes *et al.*, 2016). Improved plant functional type parameterization within and across models through careful data collection will undoubtedly help address some of the discrepancies in magnitude and relative importance of temperature and precipitation seen between models and data in our study.

Differences in model responses to increasing CO_2 drive model ensemble uncertainty in NPP since 1900 (Fig. 3). The discrepancy of CO_2 enhancement effects in the model ensemble reflects the long-running debate in empirical studies about whether or not increased CO_2 has sustained effects on tree growth and ecosystem productivity. One synthesis from Free-Air CO_2 Enrichment (FACE) studies, where conditions were raised to ~170 ppm above modern, increased productivity by 23%, or 13% ppm^{-1} (Norby *et al.*, 2005). This is slightly less than half of the 43% increase in NPP attributed to CO_2 in our model ensemble over a similar increase in CO_2 over the past 100 years (Fig. 3). However, a 30-year study that found 12% increased stem width from increased CO_2 attributed most of this difference to early growth increases that diminished as the trees matured (Hättenschwiler *et al.*, 1997). The effect of increasing competition for resources such as nitrogen and light during forest maturation has been used to explain a lack of CO_2 enhancement on tree growth in older forests quantified through tree rings (van der Sleen *et al.*, 2014; Fernández-de-Uña *et al.*, 2016). In tree rings, rising CO_2 has been associated with increased water use efficiency, but there has been little evidence that this translates into increased stem growth (Gedalof & Berg, 2010; Andreu-Hayles *et al.*, 2011; Peñuelas *et al.*, 2011; van der Sleen *et al.*, 2014; Frank *et al.*, 2015).

Many of the explanations for a lack of CO_2-induced growth increases in tree-ring widths center around the challenges of separating out the effects of increasing CO_2 from increasing tree size and stand-level forest dynamics (Jacoby & D’Arrigo, 1997; Andreu-Hayles *et al.*, 2011). In our results, stand-level NPP from tree rings showed no CO_2 effect, but individual ring widths displayed a negative effect that arises from the aforementioned challenges even though we explicitly accounted for increasing individual size in our analyses (Fig. 2, S5). This model- and data-based uncertainty in the long-term effect of increasing CO_2 on ecosystems is the greatest source of uncertainty in current and future forest carbon cycle modeling (Sitch *et al.*, 2008; IPCC, 2013). Both fire and changes in biomass are correlated with CO_2 sensitivity in our ensemble of models across temporal scales (Table 3), indicating that disturbance and recovery processes may have strong influence on how CO_2 impacts ecosystem NPP. Consequently, accurate representation of causes of forest disturbances and process of recovery at local scales is necessary to reconcile models with empirical data and to make improved predictions of future forest dynamics under continued global change (Seidl *et al.*, 2011; Dietze *et al.*, 2014; Steinkamp & Hickler, 2015).

Conclusions

Models and data display conflicting responses of NPP to climate and CO_2 and these differences are more pronounced at the multi-centennial time scales of past climate-driven ecosystem change. Models are generally over-sensitive to growing season precipitation relative to tree-ring datasets and display negative responses to temperature while tree rings show consistent positive effects. Further research on independent temperature and precipitation effects is necessary to diagnose systematic weaknesses in ecosystem models. Similarly, differences in model responses to CO_2 drive divergence of model ensemble NPP over the past century as the effects compound through time while CO_2 concentration consistently increases. Despite empirical evidence for positive CO_2 effects on individual processes included in ecosystem models, the net effects on tree growth and forest-level NPP remain unclear and poorly constrained. Differences in disturbance as measured through the presence of fire in model simulations only impacts the sensitivity of productivity to climate and CO_2 at multi-centennial times scales that capture multiple tree generations. The paths toward model improvement and reconciling discrepancies with data will be model-specific as each model has unique structural and parameterization requirements. However, accurate representation of fire and other disturbance processes and ecosystem recovery in models will require improved synthesis of short-term ecophysiological processes with gradual shifts in forest composition and structure seen in historical and paleoecological records. We advocate that this process of refinement include consideration and formal evaluation of emergent, ecosystem-level behavior at the multi-decadal and multi-centennial temporal scales at which changes in forest composition and biomass occur. This approach would better align model evaluation with the organizational and temporal scales of forest dynamics and policy decisions.

Acknowledgements

This work was supported by NSF Macrosystems grant DEB #1241891 (Dietze, Matthes, Rollinson), 1241851 (Moore, Liu, Quaife, Schaefer, Hickler, Steinkamp), 1241930 (Pederson, Hessl, Dye, Poulter), and 1241874 (McLachlan, Raiho). Model, and data contributions are as follows (in model alphabetical order, * indicates contribution sufficient to warrant authorship): CLM – Liu*, Moore*, F. Montane, A. Arellano, D. Ricciuto; ED, ED-LU – Rollinson*, Matthes*, and Dietze*; JULES, JULES-TRIFFID - Quaife*; LINKAGES – Raiho*, Dietze*; LPJ-GUESS – Steinkamp*, Hickler*, M. Forrest, C. Werner; LPJ-WSL – Poulter*; SiBCASA – Schaefer*, E. Jafarov; Tree-Ring Data – Dye*, Bishop*, Pederson*, Hessl*. Analyses and writing were performed by Rollinson. Rollinson, Dietze, Liu, McLachlan, Moore and Raiho contributed to conceptual framework of the manuscript. All authors contributed to manuscript preparation. All code used for analyses and figures in this manuscript is publically available on Github: https://github.com/PalEON-Project/Temporal-Scaling-MS.

References

Anderegg LDL, Anderegg WRL, Berry JA (2013) Not all droughts are created equal: Translating meteorological drought into woody plant mortality. *Tree Physiology*, **33**, 701–712.

Anderegg WRL, Schwalm C, Biondi F et al. (2015) Pervasive drought legacies in forest ecosystems and their implications for carbon cycle models. *Science*, **349**, 528–532.

Andreu-Hayles L, Planells O, Gutiérrez E, Muntan E, Helle G, Anchukaitis KJ, Schleser GH (2011) Long tree-ring chronologies reveal 20th century increases in water-use efficiency but no enhancement of tree growth at five Iberian pine forests. *Global Change Biology*, **17**, 2095–2112.

Babst F, Poulter B, Trouet V et al. (2013) Site- and species-specific responses of forest growth to climate across the European continent. *Global Ecology and Biogeography*, **22**, 706–717.

Best MJ, Pryor M, Clark DB et al. (2011) The Joint UK Land Environment Simulator (JULES), model description. Part 1: Energy and water fluxes. *Geoscientific Model Development*, **4**, 677–699.

Blois JL, Williams JW, Fitzpatrick MC et al. (2013) Modeling the climatic drivers of spatial patterns in vegetation composition since the Last Glacial Maximum. *Ecography*, **36**, 460–473.

Boisvenue C, Running SW (2006) Impacts of climate change on natural forest productivity - evidence since the middle of the 20th century. *Global Change Biology*, **12**, 862–882.

Burns RM, Honkala BH (1990) Silvics of North America: 1. Conifers; 2. Hardwoods. *Agricultural Handbook*, **654**.

Charney ND, Babst F, Poulter B et al. (2016) Observed forest sensitivity to climate implies large changes in 21st century North American forest growth. *Ecology Letters*, **19**, 1119–1128.

Clark D a, Brown S, Kicklighter DW, Chambers JQ, Thomlinson JR, Ni J (2001) Measuring net primary production in forest concepts and field methods. *Ecological Applications*, **11**, 356–370.

Clark DB, Mercado LM, Sitch S et al. (2011) The Joint UK Land Environment Simulator (JULES), model description. Part 2: Carbon fluxes and vegetation dynamics. *Geoscientific Model Development*, **4**, 701–722.

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

Cook BI, Ault TR, Smerdon JE (2015) Unprecedented 21st century drought risk in the American Southwest and Central Plains. *Science Advances*, **1**, 1–7.

Crucifix M, Harrison S, Brierley C (2012) Recent and deep pasts in paleoclimate model intercomparison project. *Eos, Transactions American Geophysical Union*, **93**, 539–539.

D’Orangeville L, Duchesne L, Houle D, Kneeshaw D, Côté B, Pederson N (2016) Northeastern North America as a potential refugium for boreal forests in a warming climate. *Science*, **352**, 1452–1455.

Davis MB, Shaw RG (2001) Range shifts and adaptive responses to Quaternary climate change. *Science (New York, N.Y.)*, **292**, 673–9.

Davis SC, Hessl AE, Scott CJ, Adams MB, Thomas RB (2009) Forest carbon sequestration changes in response to timber harvest. *Forest Ecology and Management*, **258**, 2101–2109.

DeLucia EH, Moore DJ, Norby RJ (2005) Contrasting responses of forest ecosystems to rising atmospheric CO2: Implications for the global C cycle. *Global Biogeochemical Cycles*, **19**, 1–9.

Dietze M, Serbin S, Davidson C et al. (2014) A quantitative assessment of a terrestrial biosphere model’s data needs across North American biomes. *Journal of Geophysical Research:*

Biogeosciences, **119**, 286–300.

Doughty CE, Malhi Y, Araujo-Murakami A et al. (2014) Allocation trade-offs dominate the response of tropical forest growth to seasonal and interannual drought. *Ecology*, **95**, 2192–2201.

Dye A, Barker-Plotkin A, Bishop D, Pederson N, Poulter B, Hessl A (2016) Comparing tree-ring and permanent plot estimates of aboveground net primary production in three Eastern U.S. forests. *Ecosphere*, in press.

Eisen K, Plotkin AB (2015) Forty years of forest measurements support steadily increasing aboveground biomass in a maturing , Quercus -dominant northeastern forest. *Journal of the Torrey Botanical Society*, **142**, 97–112.

Fernández-de-Uña L, McDowell NG, Cañellas I, Gea-Izquierdo G (2016) Disentangling the effect of competition, CO2 and climate on intrinsic water-use efficiency and tree growth. *Journal of Ecology*, n/a-n/a.

Fisichelli NA, Frelich LE, Reich PB (2013) Temperate tree expansion into adjacent boreal forest patches facilitated by warmer temperatures. *Ecography*, **36**, 001–010.

Foster DR, Motzkin G, Slater B (1998) Land-Use History as Long-Term Broad-Scale Disturbance: Regional Forest Dynamics in Central New England. *Ecosystems*, **1**, 96–119.

Foster JR, D’Amato AW, Bradford JB (2014) Looking for age-related growth decline in natural forests: Unexpected biomass patterns from tree rings and simulated mortality. *Oecologia*, **175**, 363–374.

Frank DC, Poulter B, Saurer M et al. (2015) Water-use efficiency and transpiration across European forests during the Anthropocene. *Nature Climate Change*, **5**, 579–583.

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.

Gedalof Z, Berg AA (2010) Tree ring evidence for limited direct CO2 fertilization of forests over the 20th century. *Global Biogeochemical Cycles*, **24**, 2–7.

Gerten D, Schaphoff S, Haberlandt U, Lucht W, Sitch S (2004) Terrestrial vegetation and water balance - Hydrological evaluation of a dynamic global vegetation model. *Journal of Hydrology*, **286**, 249–270.

Gómez-Aparicio L, García-Valdés R, Ruíz-Benito P, Zavala MA (2011) Disentangling the relative importance of climate, size and competition on tree growth in Iberian forests: implications for forest management under global change. *Global Change Biology*, **17**, 2400–2414.

Goulden ML, Mcmillan AMS, Winston GC, Rocha A V, Manies KL, Harden JW, Bond-Lamberty BP (2011) Patterns of NPP, GPP, respiration, and NEP during boreal forest succession. *Global Change Biology*, **17**, 855–871.

Graumlich LJ, Brubaker LB, Grier CC (1989) Long-term trends in forest net primary productivity: Casade Mountains, Washington. *Ecology*, **70**, 405–410.

Grissino-Mayer HD, Fritts HC (1997) The International Tree-Ring Data Bank: an enhanced global database serving the global scientific community. *The Holocene*, **7**, 235–238.

Hättenschwiler S, Miglietta F, Raschi A, Körner C (1997) Thirty years of in situ tree growth under elevated CO2: a model for future forest responses? *Global Change Biology*, **3**, 463–471.

Hempel S, Frieler K, Warszawski L, Schewe J, Piontek F (2013) A trend-preserving bias correction – The ISI-MIP approach. *Earth System Dynamics*, **4**, 219–236.

Hollister RD, Webber PJ, Tweedie CE (2005) The response of Alaskan arctic tundra to experimental warming: differences between short- and long-term responses. *Global Change Biology*, **11**, 525–536.

IPCC (2013) Summary for Policy Makers. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM). Camberidge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jacoby GC, D'Arrigo RD (1997) Tree rings, carbon dioxide, and climatic change. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 8350–3.

Jenkins JC, Birdsey RA, Pan Y (2001) Biomass and NPP estimation for the mid-Atlantic region (USA) using plot-level forest inventory data. *Ecological Applications*, **11**, 1174–1193.

De Kauwe MG, Medlyn BE, Zaehle S et al. (2013) Forest water use and water use efficiency at elevated CO2: a model-data intercomparison at two contrasting temperate forest FACE sites. *Global Change Biology*, **19**, 1759–1779.

De Kauwe MG, Medlyn BE, Zaehle S et al. (2014) Where does the carbon go? A model-data intercomparison of vegetation carbon allocation and turnover processes at two temperate forest free-air CO2 enrichment sites. *New Phytologist*, **203**, 883–899.

Keeling CD, Whorf TP, (Carbon Dioxide Research Group) (2005) Atmospheric CO2 concentrations (ppmv) derived from in situ air samples collected at Mauna Loa Observatory, Hawaii. *Scripps Institude of Oceanography*.

Kumar J, Brooks B-GJ, Thornton PE, Dietze MC (2012) Sub-daily Statistical Downscaling of Meteorological Variables Using Neural Networks. *Procedia Computer Science*, **9**, 887–896.

Litton CM, Raich JW, Ryan MG (2007) Carbon allocation in forest ecosystems. *Global Change Biology*, **13**, 2089–2109.

Luyssaert S, Inglima I, Jung M et al. (2007) CO2 balance of boreal, temperate, and tropical forests derived from a global database. *Global Change Biology,* **13**, 2509–2537.

Matesanz S, Brooker RW, Valladares F, Klotz S (2009) Temporal dynamics of marginal steppic vegetation over a 26-year period of substantial environmental change. *Journal of Vegetation Science*, **20**, 299–310.

Matthes JH, Goring S, Williams JW, Dietze MC (2016) Benchmarking historical CMIP5 land-climate feedbacks across the Upper Midwest and Northeastern United States. *Journal of Geophysical Research: Biogeosciences*.

McEwan RW, Dyer JM, Pederson N (2011) Multiple interacting ecosystem drivers: toward an encompassing hypothesis of oak forest dynamics across eastern North America. *Ecography*, **34**, 244–256.

Medvigy D, Wofsy SC, Munger JW, Hollinger DY, Moorcroft PR (2009) Mechanistic scaling of ecosystem function and dynamics in space and time: Ecosystem Demography model version 2. *Journal of Geophysical Research*, **114**, 1–21.

Melillo JM, Steudler PA, Aber JD et al. (2002) Soil Warming Feedbacks the Carbon-Cycle Climate. *Science*, **298**, 2173–2176.

Morin X, Lechowicz MJ, Augspurger C, O’Keefe J, Viner D, Chuine I (2009) Leaf phenology in 22 North American tree species during the 21st century. *Global Change Biology*, **15**, 961–975.

Norby RJ, Delucia EH, Gielen B et al. (2005) Forest response to elevated CO2 is conserved across a broad range of productivity. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 18052–6.

Norby RJ, Warren JM, Iversen CM, Medlyn BE, McMurtrie RE (2010) CO2 enhancement of forest productivity constrained by limited nitrogen availability. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 19368–73.

Oleson KW, Lawrence DM, Gordon B et al. (2010) Technical description of version 4.0 of the Community Land Model (CLM). *NCAR/TN-478+STR NCAR Technical Note*, 266.

Peñuelas J, Canadell JG, Ogaya R (2011) Increased water-use efficiency during the 20th century did not translate into enhanced tree growth. *Global Ecology and Biogeography*, **20**, 597–608.

Peñuelas J, Sardans J, Estiarte M et al. (2013) Evidence of current impact of climate change on life: a walk from genes to the biosphere. *Global change biology*, **19**, 2303–38.

Piao S, Sitch S, Ciais P et al. (2013) Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO2 trends. *Global Change Biology*, **19**, 2117–2132.

Post WM, Pastor J (1996) Linkages - an individual-based forest ecosystem model. *Climatic Change*, **34**, 253–261.

Prentice IC, Bartlein PJ, Webb T (1991) Vegetation and climate change in eastern North America since the last glacial maximum. *Ecology*, **72**, 2038–2056.

Ramankutty N, Foley JA (1999) Estimating historical changes in global land cover: croplands from 1700 to 1992. *Global Biogeochemical Cycles*, **13**, 997–1027.

Renwick KM, Rocca ME (2014) Temporal context affects the observed rate of climate-driven range shifts in tree species. *Global Ecology and Biogeography*, **24**, 44–51.

Rollinson CR, Kaye MW, Canham CD (2016) Interspecific variation in growth responses to climate and competition of five eastern tree species. *Ecology*, **97**, 1003–1011.

Schaefer K, Collatz GJ, Tans P et al. (2008) Combined Simple Biosphere/Carnegie-Ames-Stanford Approach terrestrial carbon cycle model. *Journal of Geophysical Research*, **113**, G03034.

Seidl R, Fernandes PM, Fonseca TF et al. (2011) Modelling natural disturbances in forest ecosystems: a review. *Ecological Modelling*, **222**, 903–924.

Shuman B, Webb T, Bartlein P, Williams JW (2002) The anatomy of a climatic oscillation: Vegetation change in eastern North America during the Younger Dryas chronozone. *Quaternary Science Reviews*, **21**, 1777–1791.

Sitch S, Smith B, Prentice IC et al. (2003) Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology*, **9**, 161–185.

Sitch S, Huntingford C, Gedney N et al. (2008) Evaluation of the terrestrial carbon cycle, future plant geography and climate-carbon cycle feedbacks using five Dynamic Global Vegetation Models (DGVMs). *Global Change Biology*, **14**, 2015–2039.

van der Sleen P, Groenendijk P, Vlam M et al. (2014) No growth stimulation of tropical trees by 150 years of CO2 fertilization but water-use efficiency increased. *Nature Geoscience*, **8**, 24–28.

Smith B, Warlind D, Arneth A, Hickler T, Leadley P, Siltberg J, Zaehle S (2014) Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model. *Biogeosciences*, **11**, 2027–2054.

Speer JH, Grissino-Mayer HD, Orvis KH, Greenberg CH (2009) Climate response of five oak species in the eastern deciduous forest of the southern Appalachian Mountains, USA.

Canadian Journal of Forest Research, **39**, 507–518.

Steinkamp J, Hickler T (2015) Is drought-induced forest dieback globally increasing? *Journal of Ecology*, **103**, 31–43.

Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, **93**, 485–498.

Turner MG, Collins SL, Lugo AL, Magnuson JJ, Rupp TS, Swanson FJ (2003) Disturbance Dynamics and Ecological Response: The Contribution of Long-Term Ecological Research. *BioScience*, **53**, 46–56.

Turner DP, Ritts WD, Cohen WB et al. (2005) Site-level evaluation of satellite-based global terrestrial gross primary production and net primary production monitoring. *Global Change Biology*, **11**, 666–684.

Walker AP, Hanson PJ, De Kauwe MG et al. (2014) Comprehensive ecosystem model-data synthesis using multiple data sets at two temperate forest free-air CO2 enrichment experiments: Model performance at ambient CO2 concentration. *Journal of Geophysical Research: Biogeosciences*, **119**, 937–964.

Walker AP, Zaehle S, Medlyn BE et al. (2015) Predicting long-term carbon sequestration in response to CO2 enrichment: How and why do current ecosystem models differ? *Global Biogeochemical Cycles*, **29**, 476–495.

Wei Y, Liu S, Huntzinger DN et al. (2014) NACP MsTMIP: Global and North American Driver Data for Multi-Model Intercomparison. *Data Set*.

Williams AP, Allen CD, Macalady AK et al. (2013) Temperature as a potent driver of regional forest drought stress and tree mortality. *Nature Climate Change*, **3**, 292–297.

Wood S (2012) mgcv: GAMs with GCV/AIC/REML smoothness estimation and GAMMs by PQL.

Supporting Information Captions

Supporting Information 1: Growing season temperature and precipitation time series for 850-2010 A.D. for all six sites that were used as model drivers and in for sensitivity analysis.

Supporting Information 2: Model simulation protocol.

Supporting Information 3: Location and record information for tree-ring width information from the International Tree Ring Databank (Grissino-Mayer & Fritts, 1997).

Supporting Information 4: Mean and 95% confidence interval of spine-detrended ring width index (RWI) for all tree ring records used in analyses.

Supporting Information 5: Sensitivity of relativized NPP and tree-ring width index (RWI) to growing season temperature, growing season precipitation, and CO_2 across three temporal scales.

Supporting Information 6: Sensitivity of relativized NPP and tree-ring width to growing season temperature, growing season precipitation, and CO_2 across three temporal scales restricted to Harvard and Howland sites.

Supporting Information 7: Relativized sensitivity of model NPP or ring width to growing season (May – September) temperature, growing season precipitation, and CO_2 by temporal extent.

Table 1: Location, modern forest type, and mean model driver growing season (May through September) temperature and precipitation for the six modeling locations and which data types were available at each site for sensitivity analysis. Temperature and precipitation values presented are mean ± standard deviation from 850-2010 A.D. Data type codes are as follows: M = ecosystem models, RW = raw tree-ring widths, NPP_{TR} = tree-ring NPP reconstruction.

Name	Longitude	Latitude	Forest Type	Temp (˚C)	Precip (mm)	Data Types
Demming Lake	-95.17	47.17	Mixed	15.6 ± 1.0	375 ± 74	M, RW
Billy's Lake	-94.58	46.28	Mixed	16.4 ± 1.0	398 ± 81	M
UNDERC	-89.53	46.22	Mixed	14.3 ± 0.9	411 ± 74	M, RW
Minden Bog	-82.83	43.61	Evergreen	16.4 ± 0.8	375 ± 65	M
Harvard Forest	-72.18	42.54	Deciduous	15.6 ± 0.7	520 ± 86	M, RW, NPP_{TR}
Howland Forest	-68.73	45.25	Mixed	13.9 ± 0.8	492 ± 80	M, RW, NPP_{TR}

Table 2: List of models and key model characteristics used in analyses. Static vegetation models were prescribed site composition based on modern forest type information (Table 1). However, because composition was assessed as fraction biomass or leaf area (LPJ & JULES models), relative composition could shift through time. Composition and biomass variability was quantified as the standard deviation of the relative fraction evergreen or biomass through time over the full modeling temporal extent. For fire occurrence, "No" indicates no fire occurred in the model simulations at any locations. Version numbers are not tracked in SiBCASA.

Model	Vers.	Vegetation Scheme	Composition Variability (%)	Fire Occurrence	Biomass Variability (%)	Citation
CLM-BGC	4.5	Static	32%	Yes	22%	Oleson *et al.*, 2010
CLM-CN	4.5	Static	29%	Yes	17%	Oleson *et al.*, 2010
ED2	2.1	Dynamic	20%	No	20%	Medvigy *et al.*, 2009
ED2-LU	2.1	Dynamic	21%	No	19%	Medvigy *et al.*, 2009
JULES-STATIC	4.1	Static	25%	No	3%	Best *et al.*, 2011; Clark *et al.*, 2011
JULES-TRIFFID	4.1	Dynamic	02%	No	6%	Best *et al.*, 2011; Clark *et al.*, 2011
LINKAGES	1 .0	Dynamic	22%	No	19%	Post & Pastor, 1996
LPJ-GUESS	3.1	Dynamic	35%	Yes	24%	Sitch *et al.*, 2003; Gerten *et al.*, 2004; Smith *et al.*, 2014
LPJ-WSL	1.0	Dynamic	28%	Yes	12%	Sitch *et al.*, 2003; Gerten *et al.*, 2004
SiBCASA		Static	00%	No	4%	Schaefer *et al.*, 2008

Table 3: ANOVA effect sizes of changes in climate and CO_2 sensitivity for four key characteristics of ecosystem models and ecosystem dynamics. Effects are expressed as mean change in normalized NPP (%) per unit climate effect ± standard error. Static vegetation and fire effects are relative to dynamic vegetation scheme and absence of fire, respectively. * and bold indicate significance at $p<0.05$.

Effect	**Character**	**1980-2010**	**1901-2010**	**850-2010**
Temperature	Static Vegetation	3.8 ± 3.68	4.13 ± 3.68	4.17 ± 3.68
	Composition Var.	12.67 ± 14.92	7.67 ± 17.3	-8.16 ± 20.47
	Fire Occurs	0.99 ± 3.31	-1.4 ± 3.31	**-8.18 ± 3.31 ***
	Biomass Var.	-0.67 ± 7.2	-9.89 ± 8.82	**-20.03 ± 9.15 ***
Precipitation	Static Vegetation	-0.01 ± 0.2	0.15 ± 0.2	0.06 ± 0.2
	Composition Var.	0.03 ± 0.2	-0.06 ± 0.23	-0.25 ± 0.27
	Fire Occurs	-0.01 ± 0.04	-0.02 ± 0.04	**-0.11 ± 0.04 ***
	Biomass Var.	0.06 ± 0.1	0.01 ± 0.12	-0.2 ± 0.13
CO_2	Static Vegetation	-0.01 ± 0.2	0.15 ± 0.2	0.06 ± 0.2
	Composition Var.	-0.34 +/- 0.7	-0.33 +/- 0.81	-1.81 ± 0.96
	Fire Occurs	-0.17 +/- 0.15	-0.15 +/- 0.15	**-0.45 ± 0.15 ***
	Biomass Var.	0.09 +/- 0.39	-0.13 +/- 0.48	-0.61 ± 0.5

Figure Captions

Figure 1: Net primary production (NPP) across all sites for the full modeling temporal extent (850-2010) for all ten ecosystem models. Sites are as follows: a) Demming Lake, b) Billy's Lake, c) UNDERC, d) Minden Bog, e) Harvard Forest, f) Howland Forest. Lines indicate the 10-year running means. Black dots at Harvard and Howland indicate tree-ring estimates of NPP.

Figure 2: Normalized sensitivity of NPP and tree ring width to growing season temperature, growing season precipitation, and CO_2 across three temporal scales. Climate and CO_2 effects have been normalized to the mean NPP or ring width for each model to facilitate comparison across models and data of different scales. Colored lines indicate the mean estimated response and color shaded area indicate a 95% confidence interval around that response. Gray shaded regions indicate the ranges of climate or CO_2 not observed in the models at that temporal scale and are presented only for aiding visualization and as such differences among models in these regions should not be interpreted.

Figure 3: *a)* 10-year running mean net primary productivity for ten terrestrial ecosystem models (lines) with 95% confidence intervals from among-site variability (shaded areas). *b)* 10-year running mean of the ensemble mean of normalized model NPP (colored line) with shaded 95% confidence interval (from spread among models) showing relative change in NPP through time in the model ensemble. Color of each line indicate the relative control of temperature (red), precipitation (blue) and CO_2 (green) on change in relative NPP, dashed line indicates the normalized mean of 100%. *c)* Mean effect of temperature (red), precipitation (blue), and CO_2 (green) on normalized NPP in the model ensemble. Lines indicate the ensemble mean and the shaded area indicates the 95% confidence intervals.