Ecological Applications, 24(4), 2014, pp. 699–715

Confronting terrestrial biosphere models with forest inventory data

JEREMY W. LICHSTEIN,[1,6] NI-ZHANG GOLAZ,[2] SERGEY MALYSHEV,[3] ELENA SHEVLIAKOVA,[3] TAO ZHANG,[1] JUSTIN SHEFFIELD,[4] RICHARD A. BIRDSEY,[5] JORGE L. SARMIENTO,[2] AND STEPHEN W. PACALA[3]

[1]*Department of Biology, University of Florida, Gainesville, Florida 32611 USA*
[2]*Program in Atmospheric and Ocean Sciences, Princeton University, Princeton, New Jersey 08544 USA*
[3]*Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544 USA*
[4]*Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey 08544 USA*
[5]*Newtown Square Corporate Campus, USDA Forest Service, Newtown Square, Pennsylvania 19073 USA*

Abstract. Efforts to test and improve terrestrial biosphere models (TBMs) using a variety of data sources have become increasingly common. Yet, geographically extensive forest inventories have been under-exploited in previous model–data fusion efforts. Inventory observations of forest growth, mortality, and biomass integrate processes across a range of timescales, including slow timescale processes such as species turnover, that are likely to have important effects on ecosystem responses to environmental variation. However, the large number (thousands) of inventory plots precludes detailed measurements at each location, so that uncertainty in climate, soil properties, and other environmental drivers may be large. Errors in driver variables, if ignored, introduce bias into model–data fusion. We estimated errors in climate and soil drivers at U.S. Forest Inventory and Analysis (FIA) plots, and we explored the effects of these errors on model–data fusion with the Geophysical Fluid Dynamics Laboratory LM3V dynamic global vegetation model. When driver errors were ignored or assumed small at FIA plots, responses of biomass production in LM3V to precipitation and soil available water capacity appeared steeper than the corresponding responses estimated from FIA data. These differences became nonsignificant if driver errors at FIA plots were assumed to be large. Ignoring driver errors when optimizing LM3V parameter values yielded estimates for fine-root allocation that were larger than biometric estimates, which is consistent with the expected direction of bias. To explore whether complications posed by driver errors could be circumvented by relying on intensive study sites where driver errors are small, we performed a power analysis. To accurately quantify the response of biomass production to spatial variation in mean annual precipitation within the eastern United States would require at least 40 intensive study sites, which is larger than the number of sites typically available for individual biomes in existing plot networks. Driver errors may be accommodated by several existing model–data fusion approaches, including hierarchical Bayesian methods and ensemble filtering methods; however, these methods are computationally expensive. We propose a new approach, in which the TBM functional response is fit directly to the driver-error-corrected functional response estimated from data, rather than to the raw observations.

Key words: carbon cycle model; data assimilation; errors in explanatory variables; global ecosystem model; land surface model; measurement error models.

INTRODUCTION

Terrestrial ecosystems play a key role in climate–carbon-cycle feedbacks, and vegetation response to climate change is one of the largest sources of uncertainty in projecting the future state of the Earth system (Friedlingstein et al. 2006, Denman et al. 2007). In terrestrial biosphere models (TBMs) designed to study climate change (e.g., Foley et al. 1996, Sitch et al. 2003, Krinner et al. 2005), vegetation responds to environmental variation largely through short timescale physiological processes, e.g., stomatal conductance and biochemical temperature sensitivities of photosynthesis and respiration parameterized from measurements of individual leaves over minutes to hours. Some longer timescale processes are also represented in some TBMs, including shifts in the distribution of broadly defined plant functional types (PFTs; e.g., Foley et al. 1996, Sitch et al. 2003) and feedbacks between the C and N cycles (Thornton et al. 2007, Gerber et al. 2010, Zaehle and Friend 2010). However, other slow timescale processes that are likely to have strong impacts on vegetation response to global change (e.g., within-PFT shifts in plant species composition and local adaptation within species) have been largely ignored in TBMs (Rastetter 1996, Leuzinger et al. 2011, Luo et al. 2011*b*).

Geographically extensive forest inventories, such as the U.S. Forest Inventory and Analysis (FIA) program,

Manuscript received 30 March 2013; revised 3 September 2013; accepted 12 September 2013. Corresponding Editor: A. D. McGuire.

[6] E-mail: jlichstein@ufl.edu

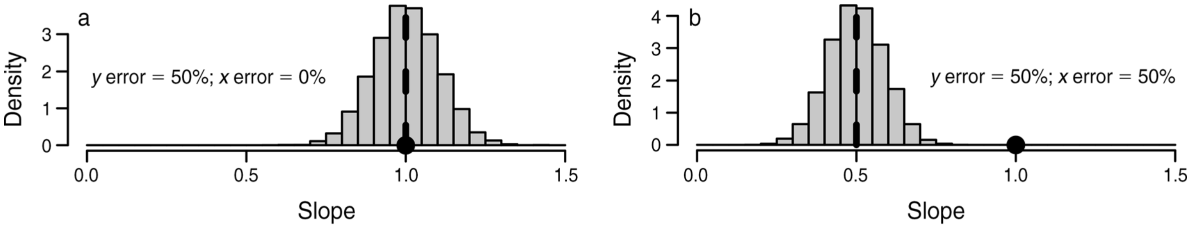

FIG. 1. Errors in explanatory variables (x) bias statistical analyses that ignore these errors, whereas errors in response variables (y) cause uncertainty but no bias. Histograms show the distribution of estimated slopes from 1000 ordinary least squares (OLS) regressions of randomly generated data with a true slope of 1.0 (indicated by the black circle on the x-axis). The y-axis is scaled so that the distribution sums to 1. Dashed vertical lines show the expected OLS slopes when measurement errors in x account for (a) 0% and (b) 50% of the observed variance in x. Measurement errors in y, which cause uncertainty but no bias in OLS regression, account for 50% of the observed variance in y in both cases. The expected OLS slope is equal to $b\sigma_x^2/(\sigma_x^2 + \sigma_{\mathrm{err}}^2)$, where b is the true slope, and σ_x^2 and σ_{err}^2 are, respectively, the variances of the true (unobserved) x values and the errors in x (Fuller 1987). The variance σ_{err}^2 is equal to 0 and σ_x^2 in panels a and b, respectively.

provide vast amounts of data that could be used to improve the performance of TBMs across spatial gradients in climate and other environmental drivers (see Plate 1). Patterns of forest biomass and productivity along these gradients integrate biological mechanisms across a range of timescales, from fast timescale physiological processes to the much slower timescales of species replacement and evolutionary adaptation. Despite a growing movement in the modeling community to rigorously evaluate and improve TBMs using a variety of data sources (e.g., Kucharik et al. 2000, Barrett 2002, Luo et al. 2003, Raupach et al. 2005, Zhou and Luo 2008, Medvigy et al. 2009, Randerson et al. 2009), the geographic gradients sampled by systematic forest inventories have not been exploited by most model–data syntheses.

Geographically extensive forest inventories present both opportunities and challenges for model–data fusion (or, synonymously, "data assimilation"), in which model state variables and/or parameters are estimated using quantitative methods (Raupach et al. 2005, Williams et al. 2009, Luo et al. 2011*a*). Previous model–data fusion efforts (e.g., Luo et al. 2003, Medvigy et al. 2009, Richardson et al. 2010) have focused largely on assimilating data from intensive study sites where errors in driver variables (e.g., meteorology and soil physical properties) may be small enough that they contribute little to model error (Spadavecchia et al. 2011). Accordingly, specification of observation errors has focused on errors in C fluxes or other variables whose values are predicted by TBMs (Hollinger and Richardson 2005, Raupach et al. 2005, Williams et al. 2009, Keenan et al. 2011). However, driver errors may be large at forest inventory plots, due to the impracticality of collecting detailed environmental data at a large number of sites (e.g., ~100 000 FIA plots).

The "errors-in-x" problem is well-known among statisticians (Griliches and Ringstad 1970, Fuller 1987, Chesher 1991, Schennach 2004), but its implications for model–data fusion with TBMs are not widely appreciated. Simply put, the problem is that errors in explanatory variables (x) lead to biased parameter estimates in statistical models that fail to account for these errors. For example, ordinary least squares (OLS) regression slopes are biased toward zero in the presence of errors in x (Fig. 1). Although the OLS case is simplistic, it provides qualitative insights into how driver error can bias model–data fusion. Suppose we wish to adjust vegetation parameter values in a TBM (e.g., C allocation to fine roots, leaves, or other tissues) so that modeled growth (biomass production) matches, as closely as possible, observed growth in forest inventory plots across geographic gradients of temperature and precipitation. The meteorological data used to drive the TBM will only approximate the actual conditions at inventory plots. If we ignore these errors while optimizing the model–data fit, we will obtain TBM parameter estimates that produce the wrong relationship between growth and meteorology and climate. To see this, consider the extreme case in which meteorological errors are so large that the TBM drivers are uncorrelated with the actual conditions at inventory plots, and therefore uncorrelated with the growth observations. In this case, an optimization algorithm that minimized the deviations between modeled and observed growth would erroneously eliminate from the TBM much or all of the growth response to meteorology and climate. Thus, we expect that ignoring driver errors in model–data fusion will bias TBMs toward having weaker responses to driver variables than the true responses of real ecosystems.

In this paper, we explore how driver errors can bias model–data fusion using FIA data and the Geophysical Fluid Dynamics Laboratory (GFDL) LM3V land model (Shevliakova et al. 2009). We focus on the United States north of 35° N and east of 100° W (hereafter "eastern United States") because (1) it is dominated by a single PFT (temperate deciduous trees comprise >80% of live biomass of non-plantation forests in the region according to FIA data), allowing us to address the largely ignored problem of within-PFT variation in TBMs; (2) a large number of FIA plots in this region have recently been measured and remeasured using a well-documented, standardized protocol (Bechtold and Patterson

2005); and (3) the data are publicly available. We ask (1) Does LM3V's functional response (i.e., the dependence of forest growth on climate and soil physical properties) appear similar to the functional response of real forests? (2) Does accounting for driver errors at FIA plots affect the perceived similarity between model and data? (3) How do driver errors affect TBM parameter optimization? (4) Is the existing network of intensive study sites (where driver errors are presumably small) adequate to constrain the within-PFT response of TBMs across environmental gradients? In *Discussion*, we outline several alternative approaches to model–data fusion that account for driver errors, with the goal of avoiding the biases exposed in this paper.

METHODS

Overview

We compared aboveground wood production (kg $C \cdot m^{-2} \cdot yr^{-1}$; hereafter, "growth") in LM3V (Shevliakova et al. 2009) to FIA data in the eastern United States. LM3V explicitly represents age-dependent ecosystem dynamics, and we controlled for forest age in our model–data comparisons. We quantified the response of growth in FIA data and LM3V output to climate (mean annual temperature and precipitation) and soil available water capacity (AWC) using a variety of linear and nonlinear regression models. It is straightforward to quantify TBM functional responses, because the values of driver variables imposed on the model are known without error. In contrast, estimation of functional responses from FIA data are complicated by driver errors (see *Introduction*). We estimated climate and AWC errors at FIA plots, and then incorporated these estimates into measurement error models (MEMs; Fuller 1987) to obtain unbiased estimates of FIA growth responses to climate and AWC. To explore potential biases in model–data fusion when driver errors are ignored, we fit LM3V vegetation parameters to FIA data using a standard optimization approach. Finally, we performed a power analysis to determine the number of intensive study sites (i.e., sites with negligible driver errors) needed to characterize the growth–precipitation response within the eastern United States.

Summary of the LM3V global land model

LM3V (Shevliakova et al. 2009) is the terrestrial component of the GFDL Earth System Model ESM2.1 and simulates the dynamics of broadly defined PFTs. Gerber et al. (2010) developed a coupled C–N version of LM3V, but here we focus on the C-only version described in Shevliakova et al. (2009). LM3V tracks the age of secondary vegetation (e.g., forests recovering from harvest or non-forest land use), which facilitates comparisons with forest inventory plots of known stand age. Net primary production (NPP) in LM3V responds to meteorology and soil moisture on a half-hourly time step. The NPP formulation is based on the Farquhar photosynthesis model (Farquhar et al. 1980, Collatz et al. 1991, Leuning 1995) and is similar to that in the IBIS model (Foley et al. 1996). Plant respiration is assumed proportional to maximum photosynthetic rate (V_{Cmax}), and both share the same Arrhenius-type temperature sensitivity. Allocation to vegetation C pools shifts with vegetation height to preserve constant ratios between leaf, sapwood, and fine-root areas according to the pipe model (Shinozaki et al. 1964). Leaf and fine-root allocational fractions approach PFT-specific values as vegetation height approaches a prescribed PFT-specific maximum height; the remainder of NPP is allocated to sapwood, which is converted to heartwood at a constant rate. Vegetation C pools turn over at PFT-specific rates.

Hydrology in LM3V is based on the Land Dynamics (LaD) model (Milly and Shmakin 2002). LaD tracks three water stores: snow pack, root-zone soil water, and ground water. Root-zone inputs are from snowmelt and precipitation, and losses are due to drainage and evaporation (modified in LM3V to separate soil evaporation and plant transpiration). Drainage prevents the root-zone store from exceeding its capacity, which is the product of vegetation rooting depth (based on values reported in Jackson et al. 1996) and plant-available soil water capacity derived from a global map (Dunne and Willmott 1996). Hereafter, we refer to root-zone water capacity as soil "available water capacity" (AWC; kg/m^2). Soil AWC (Fig. 2) is a static map in LM3V, with rooting depths prescribed based on the static distribution of PFTs in LaD (~1.0 m depth in the eastern United States [Milly and Shmakin 2002]).

LM3V configuration

LM3V was run at 1° spatial resolution and forced with the Sheffield et al. (2006) 1° spatial, three-hourly meteorology (1948–2006; Fig. 2), which is constrained to match monthly statistics from the 0.5° Climatic Research Unit (CRU) climate product (New et al. 1999, 2000, Mitchell and Jones 2005), and thus corrects for known temperature and precipitation biases in the NCEP-NCAR reanalysis (Kalnay et al. 1996, Kistler et al. 2001, see Sheffield et al. 2004). To minimize model–data discrepancies due to disturbance history, we implemented LM3V without fire or harvest, and we simulated different-aged LM3V vegetation for comparison with similar-aged FIA plots with no record of recent harvest. Specifically, vegetation in each LM3V grid cell was initiated with 1 kg C/m^2 of live biomass in the years 1952, 1932, and 1912 to yield 50-, 70-, and 90-year-old vegetation, respectively, in the year 2002 for comparison with the following FIA stand-age classes: 40–60, 60–80, and 80–100 years old. The 1948–2006 Sheffield meteorology was recycled prior to 1948 as needed. Because LM3V vegetation appeared overly sensitive to the baseline LaD-derived AWC map (see *Results*), we also considered LM3V experiments with all grid cells assigned the eastern United States mean AWC of 146 kg/m^2. The uniform AWC case reduces spatial variation in soil moisture, but does not eliminate

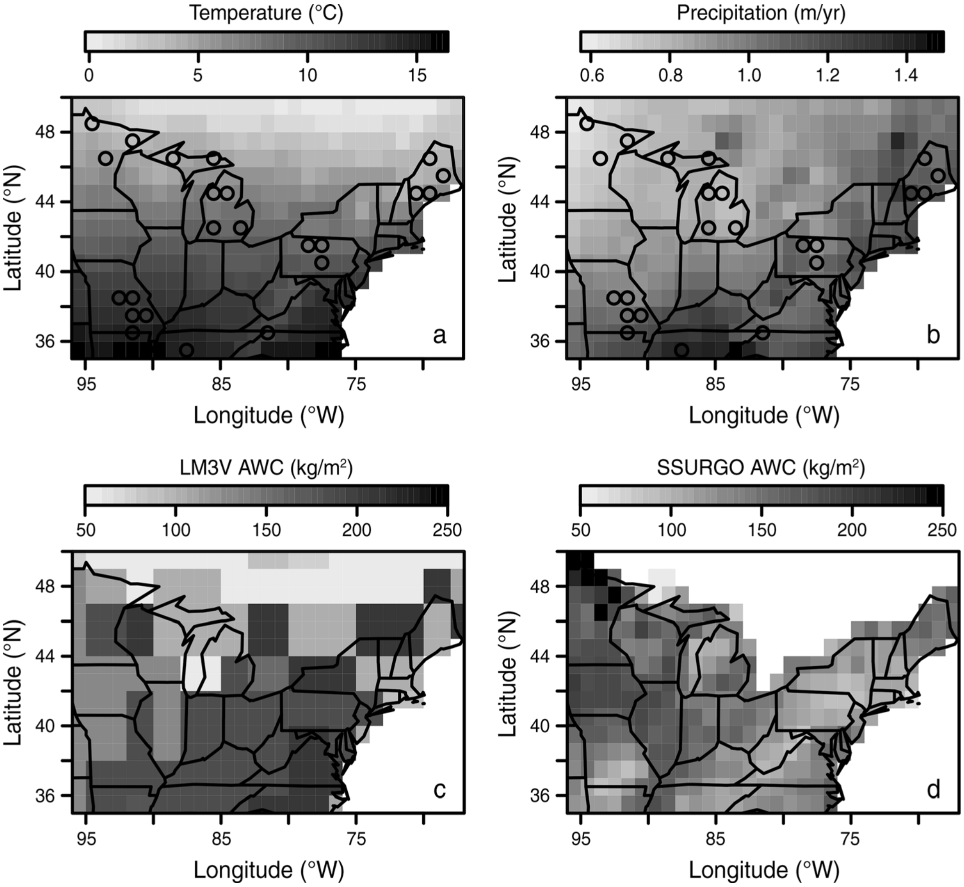

FIG. 2. Climate and soil available water capacity (AWC) maps used as explanatory variables in regression analyses of Geophysical Fluid Dynamics Laboratory LM3V land model output and Forest Inventory and Analysis (FIA) data aggregated to 1° grid cells (FIA-grid). (a–b) Mean annual temperature and precipitation calculated from the LM3V three-hourly forcing data (Sheffield et al. 2006). (c) AWC map used in LM3V simulations and in regressions of LM3V output. (d) SSURGO-derived AWC map, upscaled to 1°, used in FIA-grid regressions. SSURGO is the U.S. Department of Agriculture Soil Survey Geographic Database. Circles in panels a and b show the locations of the 23 grid cells used in the LM3V optimization; the grid cells were selected to span gradients of mean annual temperature and precipitation.

precipitation-induced spatial or temporal variation in soil moisture. For each of the two AWC maps, we performed LM3V simulations using the parameter values from Shevliakova et al. (2009) and also optimized parameter values estimated in the present study, yielding four versions of LM3V (Table 1).

FIA data set

FIA systematically samples U.S. forests on both public and private land with one inventory plot every ~2400 ha (Bechtold and Patterson 2005). Each plot samples trees with a minimum diameter at breast height (dbh) of 2.54 cm in subplots distributed over a 0.4 ha area (Bechtold and Patterson 2005). We downloaded all publicly available FIA data from the FIA website in October 2010.[7] We restricted our analysis to inventory plots that were measured and remeasured since 1999 under FIA's National Sampling Design (Bechtold and Patterson 2005). Additional plot selection criteria included (see also Appendix A) (1) reported stand age (mean age of trees in dominant size class) in one of the three age classes used in this study (40–60, 60–80, or 80–100 years); (2) no trees harvested during the remeasurement interval, for consistency with our no-harvest LM3V simulations; and (3) naturally regenerating (non-plantation) forest.

Calculating growth from FIA data and LM3V output

We estimated growth from FIA data at two spatial scales: individual FIA plots (FIA-plot) and FIA plots in a given age class (40–60, 60–80, or 80–100 years) averaged within 1° grid cells (FIA-grid; Fig. 3; Appendix B: Figs. B1 and B2). To estimate growth at FIA plots, we first estimated aboveground live biomass (AGB; kg C/m^2) at each plot at time 0 (first plot measurement in our data set) and time *t* (remeasurement, *t* years later) by

---

[7] http://www.fia.fs.fed.us/

TABLE 1. Data sources for regression analyses of U.S. Forest Inventory and Analysis (FIA) data and Geophysical Fluid Dynamics Laboratory LM3V land model output.

Source	Description
FIA-plot	Individual FIA inventory plots (0.4 ha). Exact plot coordinates were obtained from the U.S. Forest Service. Climate and soil available water capacity (AWC) were extracted from the PRISM (4-km resolution) and SSURGO (~1.6-ha resolution) data sets, respectively.
FIA-grid	FIA plots aggregated to 1° grid cells. Each value is the mean of $\geq$10 FIA plots in a given stand-age class (40–60, 60–80, or 80–100 years). Climate data were extracted from Sheffield et al. (2006), and soil AWC was upscaled to 1° from SSURGO.
LM3V-baseline	Baseline LM3V model with parameter values and soil AWC as in Shevliakova et al. (2009) and meteorological drivers from Sheffield et al. (2006).
LM3V-SbasePopt	Same as LM3V-baseline, but with optimized vegetation parameters.
LM3V-SmeanPbase	LM3V with baseline parameter values and a uniform soil AWC map (every grid cell assigned the eastern U.S. mean, 146 kg/m^2).
LM3V-SmeanPopt	Same as LM3V-SmeanPbase, but with optimized vegetation parameters.

combining reported dbh values with published biomass allometries (Jenkins et al. 2003) and assuming a 2:1 ratio of biomass to C. We then estimated growth (kg C·m^{-2}·yr^{-1}) at each plot as

$$\text{growth} = (\text{AGB}_t - \text{AGB}_0 + M_0)/t \qquad (1)$$

where AGB$_0$ and AGB$_t$ are, respectively, AGB at times 0 and t; M_0 is the total time-0 AGB of trees that were alive at time 0 but dead at time t; and t is the reported remeasurement interval (about five years, on average, for eastern United States FIA plots). Eq. 1 is also straightforward to calculate from LM3V output (over the years 2002–2006 in our analysis) by replacing M_0/t with $\mu \times \text{AGB}_0$, where $\mu = 0.015\ \text{yr}^{-1}$ is the LM3V mortality rate for temperate deciduous trees (Shevliakova et al. 2009). Eq. 1 ignores growth of trees that died between times 0 and t, and therefore has a small negative bias in both the FIA and LM3V calculations, which should not bias the model–data comparison. Note that growth as defined here excludes production of leaves, roots, and reproductive tissues and is roughly 30% of NPP.

Quantifying LM3V and FIA growth responses to climate and soil

We quantified the growth response in LM3V output and FIA data (both FIA-plot and FIA-grid; Table 1) to spatial variation in mean annual temperature (which is strongly correlated with growing season length within the eastern United States), mean annual precipitation, mean July plus August precipitation, and soil AWC using linear and nonlinear regression models. Plot- and grid-scale analyses of FIA data each have pros and cons. Errors in climate and AWC are more easily quantified at the plot scale, but FIA-plot growth data are noisy, which makes visual comparison to LM3V inconvenient. In contrast, FIA-grid data allow for easy visual comparison to LM3V, but climate and AWC error estimates are not available at the 1° grid scale.

For the regression analysis, climate variables were averaged over 1948–2006. We obtained similar regres-

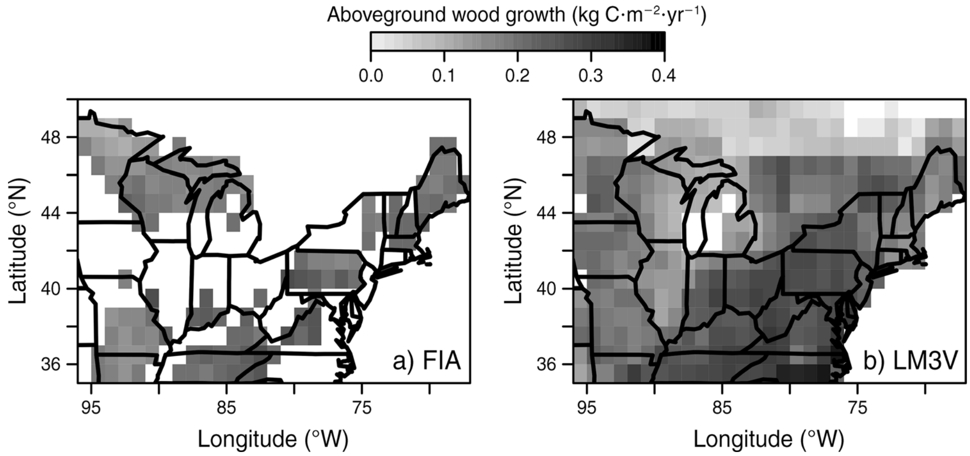

FIG. 3. Growth in (a) 60–80 year-old FIA plots and (b) 70-year-old vegetation in the baseline LM3V model (Shevliakova et al. 2009). Each FIA value represents the mean aboveground wood growth rate of $\geq$10 remeasured inventory plots; cells with <10 remeasured plots are white. LM3V values are shown only for grid cells where the temperate deciduous tree plant functional type was present in $\geq$20 of the last 40 simulation years; other cells are white. Maps for younger and older stand-age classes are shown in Appendix B: Figs. B1 and B2.

sion results when climate was averaged over 2002–2006 (results not reported). Across the three sources (FIA-plot, FIA-grid, and LM3V) and three age classes, neither of the two precipitation variables (mean annual or mean July plus August) was consistently a better predictor of growth. We only report results for mean annual precipitation, which does not qualitatively affect our conclusions. Below, we describe in sequence (1) data sources; (2) OLS regressions that ignore driver errors at FIA plots; (3) MEMs applied to FIA-plot data; and (4) a net growth response index that facilitated comparisons among regression models with different functional forms and error structures.

To quantify growth responses of FIA-plot data, we obtained the true locations of FIA plots through an agreement with the U.S. Forest Service (FIA reports approximate locations in their public database), and we overlaid these locations on high-spatial-resolution climate and soils data: the 4-km monthly PRISM climate data set and the ~1:24 000 scale U.S. Department of Agriculture Soil Survey Geographic (SSURGO) database, which has a minimum mapping unit of ~2 ha (Soil Survey Division Staff 1993; data *available online*).[8,9] The high spatial resolution of these climate and soils data should reduce, but not eliminate, errors at the scale of individual FIA plots (0.4 ha). We did not run LM3V at this spatial resolution, because PRISM lacks sufficient temporal resolution (monthly, compared to the three-hourly inputs required by LM3V) and does not include the full suite of meteorological drivers required by land models (e.g., radiation and wind speed). We calculated AWC from SSURGO as available water supply to 1 m depth (similar to the LM3V AWC rooting depth), with units converted to kg/m^2 as in LM3V. We obtained qualitatively similar results (not reported) by estimating AWC from SSURGO as available water supply to 0.5 m depth or available water capacity averaged over the entire soil depth. Preliminary analysis showed that FIA-plot growth was unimodally related to AWC, declining beyond AWC of ~200 kg/m^2, which likely reflects growth inhibition under waterlogged soil conditions. Because LM3V includes no mechanisms that can reproduce this pattern (i.e., for a given meteorological forcing, NPP in LM3V is non-decreasing with increasing AWC), we excluded FIA plots with AWC >200 kg/m^2 from the FIA-plot regressions.

To quantify growth responses of LM3V and FIA-grid data using the identical set of grid cells, we restricted both analyses to grid cells where (1) SSURGO AWC, upscaled to 1°, was less than 200 kg/m^2 (as in FIA-plot regressions); (2) at least 10 FIA plots were available to estimate FIA-grid growth; and (3) the PFT in LM3V was temperate deciduous forest for at least 20 of the last 40 simulation years (the LM3V configuration we used allowed for a single PFT per grid cell per year). We used the 1° LM3V meteorological drivers (Sheffield et al. 2006) to quantify climate in both LM3V and FIA-grid regressions. Because these drivers are constrained to match monthly CRU temperature and precipitation statistics, climate errors in FIA-grid regressions should be small. We used LM3V AWC and 1° SSURGO AWC in LM3V and FIA-grid regressions, respectively.

We estimated separate growth responses for each of the three age classes (50-, 70-, and 90-year-old LM3V vegetation; and 40–60-, 60–80-, and 80–100-year-old FIA plots). For each age class, we fit linear and nonlinear regression models of growth as a function of mean annual temperature, mean annual precipitation, and soil AWC, including OLS multiple regressions with linear terms for temperature, precipitation, and AWC; OLS multiple regressions with the three linear terms plus one or more quadratic terms for temperature, precipitation, and/or their interaction (quadratic AWC terms were unnecessary, because we excluded locations with SSURGO AWC >200 kg/m^2); and nonlinear models following Luyssaert et al. (2007) of the form: E[growth] $= b_0 T^{b_1} P^{b_2} S^{b_3}$, where E[growth] is expected growth, and T, P, and S are temperature, precipitation, and soil AWC, respectively. We fit these nonlinear models assuming that errors in growth were (1) additive normal random variables with constant variance, (2) additive normal random variables with variance proportional to expected growth, or (3) multiplicative lognormal random variables. Within a given data source (LM3V, FIA-grid, or FIA-plot) and age class, we compared the above models using the Akaike information criterion (AIC). In most cases, models that were linear in parameters (including those with quadratic terms) and that lacked interaction terms had AIC values that were similar to or better (lower) than the other models. Therefore, we restricted subsequent analyses to linear-in-parameters models with no interaction terms. Examination of model residuals indicated approximate normality in all cases, as assumed by OLS.

To explore how driver errors may affect estimates of FIA growth responses to climate and soil, we estimated errors in PRISM climate data and in SSURGO AWC at the scale of individual FIA plots, and we used these error estimates to fit MEMs to FIA-plot data. We did not fit MEMs to FIA-grid data because we lacked a straightforward means of estimating climate and soil errors at the 1° scale. The MEMs we fit use a variance-covariance matrix of the error structure of response and explanatory variables to yield unbiased estimates of regression coefficients and their uncertainties (Fuller 1987). These MEMs assume normally distributed errors and are restricted to functional forms that are linear in the parameters, including the quadratic models we fit. These restrictions allow the MEM to be solved using linear algebra (see Appendix A for details). MEMs may be extended to non-normal and nonlinear cases using more computationally demanding approaches (e.g., Lichstein et al. 2010). As noted above, the assumptions

---

[8] http://prism.oregonstate.edu

[9] http://soildatamart.nrcs.usda.gov/

of normality and linearity are not overly restrictive in the present analysis. We fit one MEM for each age class using the functional form with the lowest AIC among the candidate OLS models. It is important to note that multiple regression coefficients obtained from MEMs are not necessarily larger in magnitude than those obtained from methods such as OLS that ignore errors in x. This is in contrast to the example in Fig. 1 with a single explanatory variable, in which the regression slope is always biased toward zero by errors in x.

We quantified errors in PRISM climate data using two different approaches. First, we used temperature and precipitation data from eastern U.S. AmeriFlux sites included in the North American Carbon Program (NACP) site-level interim synthesis, which were gap-filled using a standard protocol (Schaefer et al. 2012). We included 15 non-irrigated sites with reliable rain gage data: US-ARM (Fischer et al. 2007), US-Dk2 (Stoy et al. 2008), US-Dk3 (Oren et al. 2006), US-Ha1 (Urbanski et al. 2007), US-Ho1 (Hollinger et al. 1999), US-Ib2, US-Los (Sulman et al. 2009), US-MMS (Schmid et al. 2000), US-MOz (Gu et al. 2006), US-Ne3 (Verma et al. 2005), US-PFa (Sulman et al. 2009), US-Shd (Suyker et al. 2003), US-Syv (Desai et al. 2005), US-UMB (Gough et al. 2008), and US-WCr (Cook et al. 2004). For each of the 15 AmeriFlux sites, we calculated mean annual temperature and precipitation over the available time period (mean period of 7.2 years; range 3–16 years) for comparison with PRISM data for the same location and time. Second, we used point-scale error statistics reported by Daly et al. (2008) for the 800-m-scale PRISM climate normals (1971–2000 means). The AmeriFlux-based approach probably overestimates errors in the 4-km monthly PRISM product, because AmeriFlux meteorologies include measurement error (especially for precipitation), and because only a few years of data were available at some sites to estimate mean climate. In contrast, the 800-m-scale error estimates reported by Daly et al. (2008) probably underestimate errors in mean climate derived from the 4-km monthly PRISM product due to the difference in spatial resolution and temporal coverage. Thus, the true errors in the 4-km PRISM data, when applied to point locations such as FIA plots, likely lie somewhere between the two estimates described above. In Appendix A, we explain in detail how we estimated the error covariance matrix for the MEMs.

In contrast to temperature and precipitation, we did not obtain quantitative error estimates for SSURGO AWC. Several studies have assessed at fine spatial scales the accuracy of soil texture and other attributes reported in SSURGO (Thomas et al. 1989, Rogowski and Wolf 1994, Drohan et al. 2003) and have found good overall agreement despite considerable spatial heterogeneity within SSURGO mapping units. However, to our knowledge, there is no quantitative information available on errors in SSURGO AWC. Therefore, we considered two scenarios to bracket the likely range of errors in SSURGO AWC: we assumed that either 10% or 50% of the total variance in SSURGO AWC across FIA plots was due to errors, including any errors due to scale mismatches between SSURGO mapping units and FIA plots.

We defined a net growth response index (GR) to compare the response of growth to climate and soil variables across statistical models with different functional forms (with vs. without quadratic terms) and error structures (OLS vs. MEMs). We define GR_x as the estimated net change in growth across a fixed range of an explanatory variable x, holding other explanatory variables constant. The ranges, chosen to span most of the variation in each x, were 3–14°C for temperature, 0.65–1.3 m/yr for precipitation, and 100–200 kg m^{-2} for AWC. For example, $\mathrm{GR}_{\mathrm{pr}}$ is the growth rate estimated by a given model at 1.3 m/yr precipitation minus the corresponding growth rate at 0.65 m/yr precipitation. For models without a quadratic x term, GR_x is simply the regression coefficient (slope) times the range of x. Note that because the models include no interaction terms, GR_x does not depend on the values of the other explanatory variables. To quantify uncertainty in GR_x, we calculated GR_x for each of 10 000 random samples drawn from the distribution of regression model parameters, which was assumed to be multivariate normal (Bolker 2008).

Optimization

To explore potential pitfalls that may arise in model–data fusion in the presence of errors in x, we implemented an optimization scheme to minimize the sum of squared deviations between growth predicted by LM3V (run at 1° spatial resolution) and growth observed at FIA plots (FIA-plot; Table 1). This optimization approach ignores any differences between LM3V drivers and the actual meteorology and AWC at FIA plots. Therefore, we would expect the optimized LM3V model to be less sensitive than real forests to environmental variation (see *Introduction*).

We optimized three parameters for the temperate deciduous tree PFT: fraction of NPP allocated to leaves and fine roots (with the remainder allocated to wood) and maximum photosynthetic rate at 15°C (V_{Cmax}). These three parameters were selected because they directly affect growth and should be well-constrained by forest inventory data.

To minimize computational costs, we used the Gauss-Newton method (Fletcher 1987), a "gradient-descent" algorithm (Raupach et al. 2005, Williams et al. 2009), and we based the optimization on a small subset ($n = 23$ grid cells) of eastern United States 1° grid cells (see Appendix A for selection criteria). The optimization involved iteratively running LM3V for the 23 grid cells, comparing the grid-cell-scale model output to individual FIA plots in these grid cells, and then rerunning the model with new parameter values until the algorithm described in Fletcher (1987) converged (see Appendix A

TABLE 2. Vegetation parameter values for the temperate deciduous tree plant functional type in different versions of the LM3V model and empirical estimates.

Source	Leaf allocation	Fine-root allocation	V_{Cmax} (μmol $CO_2 \cdot m^{-2} \cdot s^{-1}$)
LM3V-baseline	0.520	0.250	40
LM3V-SbasePopt	0.264; CI, 0.249–0.279	0.466; CI, 0.459–0.472	17.7; CI, 16.9–18.4
LM3V-SmeanPopt	0.369; CI, 0.331–0.401	0.305; CI, 0.278–0.332	11.5; CI, 11.3–11.7
Observations	0.28; range, 0.24–0.35†	0.36; range, 0.22–0.44†	40.7; CI, 39.2–42.2‡

Notes: Leaf and fine-root allocation are fractions of net primary production (NPP). V_{Cmax} is the maximum photosynthetic rate (carboxylase velocity) at 15°C. CI are 95% confidence intervals. LM3V-baseline values are from Shevliakova et al. (2009). LM3V-SbasePopt and LM3V-SmeanPopt values (maximum likelihood estimates and 95% confidence intervals) were estimated using the baseline and uniform soil AWC maps, respectively (Table 1). Observations of leaf and fine-root allocation (mean and range) are biometric estimates from FLUXNET sites (see details in footnotes). V_{Cmax} observations (mean and 95% confidence interval) are from Kattge et al. (2009).

† Biometric estimates ($n = 3$ estimates) from unmanaged, eastern U.S. temperate broad-leaved forest sites in the FLUXNET synthesis data set (Luyssaert et al. 2007), including unmanaged secondary forests. More than three unmanaged eastern U.S. sites are reported by Luyssaert et al. (2007), but only three include biometric estimates of NPP components. Including all global unmanaged forests older than 30 years that report biometric estimates of allocation yields 25 sites, with mean (and SE) leaf and fine-root allocations of 0.238 (0.015) and 0.296 (0.027), respectively.

‡ Kattge et al. (2009) report V_{Cmax} values at 25°C. We converted these values to 15°C using the Arrhenius-type temperature sensitivity in Shevliakova et al. (2009). The uncertainty we report is for the mean V_{Cmax} of temperature deciduous trees, based on the standard error in Kattge et al. (2009) and assuming the mean is normally distributed.

for details). Local optimization, such as the Gauss-Newton algorithm used here, is less robust than some other estimation methods, such as Markov chain Monte Carlo (Gilks et al. 1996) and simulated annealing (Goffe et al. 1994). However, these alternative methods are computationally demanding, typically requiring $>10^4$ model runs to converge compared to $<10^2$ for many local optimizations (Fletcher 1987). To increase our confidence that we had identified the global optimum, we repeated the optimization a second time with alternative initial parameter values, which yielded similar results. Convergence of the Gauss-Newton algorithm and the modest width of confidence intervals (Table 2) imply that the parameters were well constrained by the data.

Power analysis

We performed a power analysis to estimate the number of intensive study sites (N_{site}) needed to constrain the relationship between growth and mean annual precipitation; similar results were obtained for mean July plus August precipitation (not reported). To simplify the analysis, we assumed a linear response between growth and precipitation, we ignored the effects of other variables (e.g., temperature and soil AWC), and we assumed that precipitation could be measured without error at "intensive study sites." These assumptions make our power analysis conservative; i.e., the true value of N_{site} is likely larger than the estimates we report.

We quantified the minimum N_{site} needed to have at least an 80% chance of (1) correctly rejecting the null hypothesis, H_0, that the growth–precipitation slope is zero; and (2) estimating a slope within 20% of the true value. Because the true slope is unknown, we considered six different hypothetical values obtained from the FIA-grid and LM3V-baseline OLS regressions for each of the three age classes. For each of the six cases, we used the precipitation slope from the OLS model with the lowest AIC among models with a linear precipitation response. We did not use FIA-plot regressions as targets for power analysis, because the high plot-scale growth variance is not representative of intensive study sites and would artificially inflate N_{site} (which depends on the residual standard error).

The first power analysis (N_{site} needed to reject H_0 with 80% probability) is standard and was executed with R software (R Development Core Team 2012) using the ss.power.reg.coef function in the MBESS package. The second power analysis (N_{site} needed to estimate a slope within 20% of the true slope with 80% probability) is nonstandard and was executed with a simulation approach in which we performed OLS regressions on randomly generated data sets and numerically solved for the N_{site} of interest. Each randomly generated data set had randomly generated x (precipitation) values drawn from a uniform distribution with range equal to the actual precipitation range in our study; and y (growth) values equal to $bx + \varepsilon$, where b is the "true" slope (i.e., the slope from one of the six multiple regressions described above), and ε is a normal random variable with mean zero and standard deviation equal to the residual standard error from the source regression.

RESULTS

LM3V appears significantly more sensitive than 1° grid-scale FIA data to soil available water capacity (AWC), but not to temperature or precipitation.—Here, we describe the functional response of the LM3V model relative to the apparent functional response estimated from FIA-grid data (Table 1) when measurement errors are ignored (OLS regression). Plotting growth vs. climate and AWC for LM3V-baseline and FIA-grid reveals different shapes of temperature responses (unimodal in FIA-grid, monotonic in LM3V-baseline) and greater sensitivity of LM3V-baseline to both precipitation and AWC (see Fig. 4 and Appendix B: Figs. B3 and

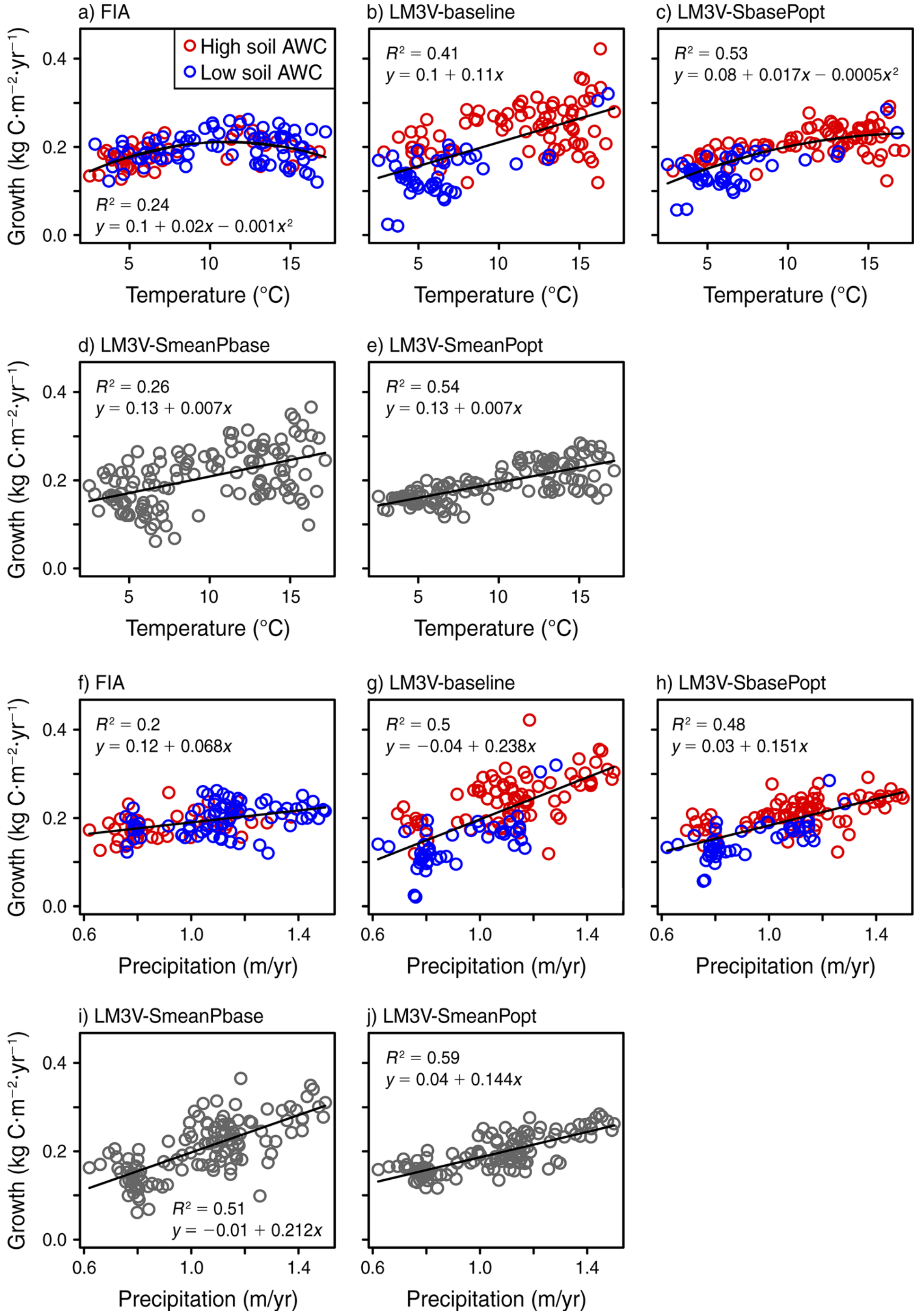

FIG. 4. Growth in FIA plots and the LM3V model vs. climate and soil available water capacity (AWC). Aboveground wood growth rates of 60–80 year-old FIA inventory plots (1° grid-cell means) and 70-year-old vegetation in four versions of the LM3V model (Table 1) are plotted against (a–e) mean annual temperature and (f–j) mean annual precipitation. Blue/red points refer to grid cells where AWC is below/above its median value. Points are gray in cases where LM3V was run with a uniform AWC map. Black lines and curves show ordinary least squares regressions: R^2 is the proportion of variance explained; y in the equations is growth, and x is (a–e) temperature or (f–j) precipitation. Nonsignificant ($P > 0.05$) quadratic terms were excluded. Results for younger and older stand-age classes are shown in Appendix B: Figs. B3 and B4.

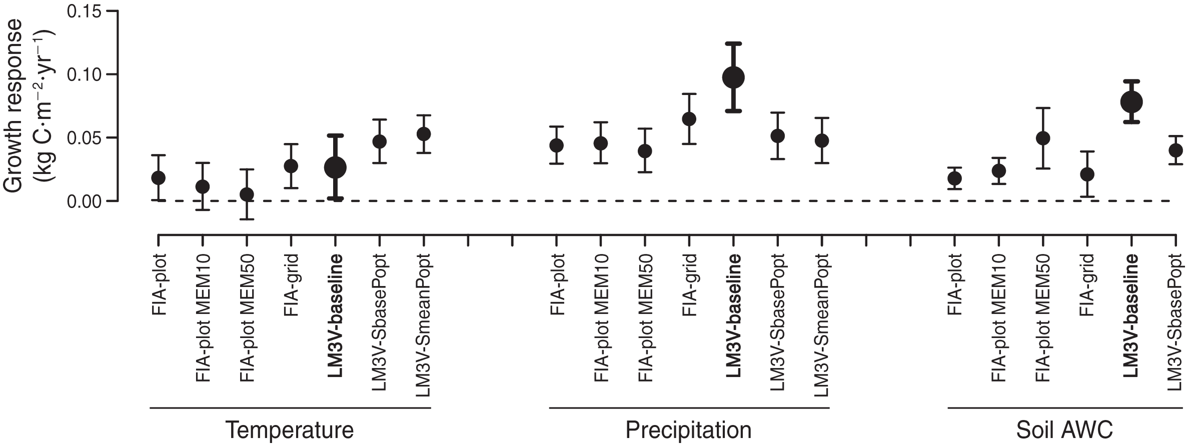

FIG. 5. Growth response of FIA plots and the LM3V model to climate and soil available water capacity (AWC). Response of aboveground wood growth to mean annual temperature, mean annual precipitation, and AWC were estimated from 60–80 year-old FIA inventory plots and output from different versions of the LM3V model. From left to right, the seven estimates for temperature and precipitation responses are from (1) ordinary least squares (OLS) analysis of plot-scale FIA data; (2 and 3) measurement error model (MEM) analysis of plot-scale FIA data using temperature and precipitation error estimates from Daly et al. (2008) and assuming 10% (MEM10) or 50% (MEM50) of AWC variance is due to measurement error; (4) OLS analysis of FIA data aggregated to 1° grid cells; (5) the baseline LM3V model (shown in boldface type); and (6 and 7) LM3V with baseline or uniform AWC maps and optimized parameter values (Table 1). For soil AWC, only the first six estimates are shown, because no AWC response is available for the uniform-AWC LM3V-SbasePopt case (Table 1). The *y*-axis (growth response) quantifies the expected change in growth across the following ranges: 3–14°C for temperature, 0.65–1.3 m/yr for precipitation, and 100–200 kg/m^2 for AWC. Error bars are 95% confidence limits. Results for younger and older stand-age classes, and for MEMs based on an alternative set of temperature and precipitation errors estimates, are shown in Appendix B: Figs. B5 and B6.

B4). In OLS multiple regressions, the greater sensitivity of LM3V-baseline than FIA-grid to AWC persisted (Figs. 5 and B5), but other model–data differences were diminished: LM3V-baseline and FIA-grid both had unimodal temperature responses with maxima near 10°C (Table B1) and were not significantly different from each other in their net growth responses (GR) to temperature and precipitation (Figs. 5 and B5).

LM3V appears significantly more sensitive than plot-level FIA data to precipitation and soil AWC if driver errors are ignored or assumed small, but not if driver errors are assumed large.—Here, we describe the functional response of the LM3V model relative to (1) the apparent functional response estimated from FIA-plot data (Table 1) when measurement errors are ignored (OLS models), and (2) the functional response estimated from FIA-plot data using measurement error models (MEMs). OLS multiple regressions of plot-level FIA data (FIA-plot) yielded similar temperature responses to LM3V-baseline, but precipitation and AWC responses that were significantly weaker than LM3V-baseline (Figs. 5 and B5 and Table B1). The MEM results depended on the assumed error variances. FIA-plot precipitation responses were significantly weaker than LM3V-baseline when the relatively small climate errors reported by Daly et al. (2008) were used in MEMs (Figs. 5 and B5), but not when the larger climate errors estimated from AmeriFlux sites were used (Fig. B6). FIA-plot had weaker AWC responses than LM3V-baseline when AWC errors were assumed small at FIA plots (10% of total AWC variance), but this difference became nonsignificant when AWC errors were assumed large (50% of total AWC variance; Figs. 5, B5, B6).

Ignoring driver errors when optimizing LM3V vegetation parameter values yields unrealistic parameter estimates.—Optimizing LM3V vegetation parameters by minimizing the sum of squared deviations between LM3V and FIA growth in 23 selected grid cells substantially reduces model–data deviations (Figs. 6 and Appendix B: Figs. B7 and B8), but results in unrealistic parameter values (see LM3V-SbasePopt in Table 2): optimized fine-root allocation is significantly higher than biometric estimates from eastern U.S. FLUXNET sites, and optimized V_{Cmax} is significantly lower than the mean V_{Cmax} reported for temperate deciduous forest by Kattge et al. (2009). In contrast, if we eliminate AWC heterogeneity in LM3V (by assigning all model grid cells the eastern United States mean AWC), then optimized leaf and fine-root allocation are consistent with FLUXNET biometric estimates, but optimized V_{Cmax} still appears too low (see LM3V-SmeanPopt in Table 2).

At least 40 intensive study sites would be needed to accurately characterize forest growth response to precipitation within the eastern United States.—Regardless of which data source (LM3V-baseline or FIA-grid) or forest age class is used to specify the hypothetical "true" response of growth to precipitation, at least 40 intensive study sites (with little or no driver error) would be needed to accurately characterize the growth–precipitation response within the eastern United States (Table 3).

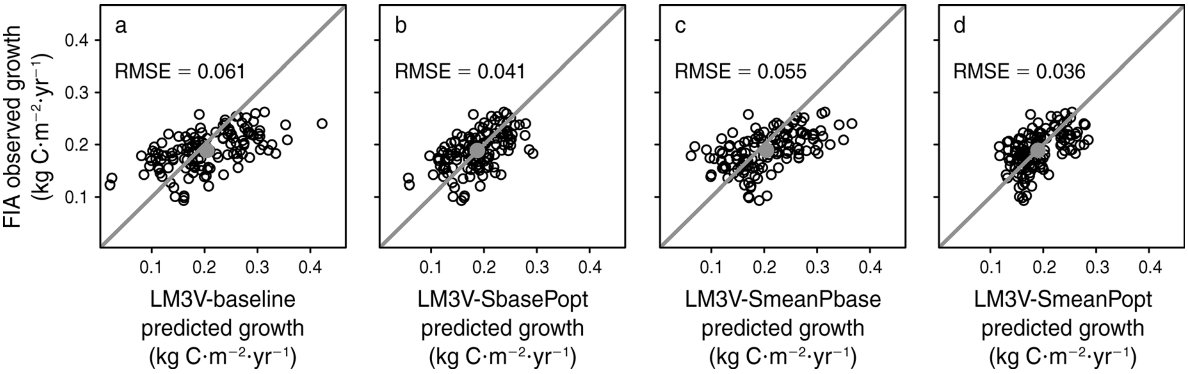

FIG. 6. Observed (FIA) vs. predicted (LM3V model) aboveground wood growth (kg C·m^{-2}·yr^{-1}). FIA observations (mean of $\geq$10 FIA plots in 1° grid cells) are plotted against predictions from four versions of the LM3V model (Table 1). RMSE is the root mean squared error of predictions. The 1:1 line and overall means (solid gray circles) are shown. LM3V vegetation parameter values in panels a and c are from Shevliakova et al. (2009), whereas those in panels b and d were optimized using FIA data in 23 grid cells (see locations in Fig. 2a, b). Results are for 60–80 year-old FIA plots and 70-year-old LM3V vegetation. Younger and older stand-age classes yielded similar results (Appendix B: Figs. B7 and B8).

Here, "accurately characterize" means having at least an 80% chance of estimating a precipitation response that is within 20% of the true response. At least 23 sites would be needed merely to detect a significant precipitation response with 80% probability, assuming a Type I error rate of 0.05 (Table 3).

DISCUSSION

Our analysis explores how errors in driver variables (e.g., meteorology and soil physical properties) can affect the functional response of forest growth (i.e., the dependence of aboveground wood production on climate and soil) estimated from geographically extensive forest inventory data sets, such the U.S. Forest Inventory and Analysis (FIA) database. Our results show that accounting for driver errors at inventory plots can reduce apparent differences between functional responses of terrestrial biosphere models (TBMs) and real forests. Ignoring these errors can bias parameter estimates obtained from standard data-assimilation methods. Our results apply to geographically extensive forest inventories where the large number of plots (e.g., ~100 000 FIA plots) precludes detailed environmental measurements at each location. Our results do not necessarily apply to flux-tower or other intensive study sites where driver errors may be relatively small. We now revisit the questions posed at the end of *Introduction*.

Does the functional response of the LM3V model appear similar to that of real forests? Does accounting for driver errors at FIA plots affect the perceived similarity between model and data?—Responses of LM3V and real forests (eastern U.S. FIA data) to spatial variation in mean annual temperature were similar in most of our comparisons. The main exception was in univariate regressions that ignored precipitation and soil available water capacity (AWC; Fig. 4). However, this difference largely disappeared in multiple regressions that included precipitation and AWC, in which case both LM3V and FIA data showed a temperature optimum around 10°C (Appendix B: Table B1).

In contrast, LM3V appeared more sensitive than real forests to mean annual precipitation and AWC in most comparisons, particularly in cases where driver errors were potentially large yet ignored in ordinary least squares (OLS) regressions. When driver errors at FIA plots were accounted for in measurement error models (MEMs), the apparent oversensitivity of LM3V to precipitation and AWC was reduced. At the scale of 1° grid cells, we expect climate errors to be small due to the high density of meteorological towers in the eastern United States (New et al. 1999, Sheffield et al. 2006). Thus, precipitation responses estimated

TABLE 3. Number of intensive study sites needed to quantify forest growth (woody biomass increment; kg C·m^{-2}·yr^{-1}) response to mean annual precipitation in the eastern U.S.

Source	Age (yr)	Slope	N_{signif}	N_{20}
FIA-grid	40–60	0.120***	30	52
LM3V-baseline	50	0.136***	26	49
FIA-grid	60–80	0.010***	24	40
LM3V-baseline	70	0.142***	23	42
FIA-grid	80–100	0.062*	79	191
LM3V-baseline	90	0.109***	37	85

Notes: Results reported in the table refer to the hypothetical (unrealistic) case where intensive study sites have no error in precipitation, and so the minimum number of sites reported is conservative. Each row in the table represents a different hypothetical "true" precipitation response, based on multiple regressions of growth vs. temperature, precipitation, and soil available water capacity (see Table 1 for source definitions). Age is the FIA-grid stand-age range or the LM3V-baseline vegetation age. Slope is the multiple regression precipitation coefficient. N_{signif} is the number of intensive study sites needed to have an 80% chance of correctly rejecting the null hypothesis (H_0: slope = 0; two-tailed test; Type-I error rate = 0.05), assuming that the value in the slope column is the true slope. N_{20} is the number of intensive study sites needed to have an 80% chance of estimating a slope within 20% of the value in the slope column.

* $P < 0.05$; *** $P < 0.001$.

from OLS regressions of 1° grid-scale FIA data (FIA-grid) should be approximately unbiased. FIA-grid precipitation responses were weaker than LM3V precipitation responses, but not significantly so. Precipitation responses estimated from MEMs of plot-scale FIA data (FIA-plot) were also weaker than in LM3V, but these differences were not statistically significant when precipitation errors at FIA plots were assumed to be large (i.e., the upper limit of the plausible range).

In contrast to 1° climate errors, we expect relatively large AWC errors because soil properties are highly variable at fine spatial scales (see Plate 1) and therefore difficult to quantify. One indication of large AWC errors at the 1° scale is that LM3V AWC and 1° SSURGO AWC are not significantly correlated with each other (Fig. 2c, d). Given the potentially large 1° AWC errors, AWC responses estimated from FIA-grid OLS regressions may be substantially biased toward zero, and thus may be a poor indicator of the sensitivity of real forests to soil properties. FIA-plot MEMs yielded weaker AWC responses than LM3V, but again, these differences were not statistically significant when AWC errors were assumed large. Because quantitative estimates of AWC errors are not currently available, we considered arbitrary (but plausible) lower and upper bounds for these errors. Clearly, quantitative error estimates for regional- to global-scale soil data sets (e.g., Dunne and Willmott 1996) would improve our capacity to test TBMs.

Despite the uncertainty in our results, which largely reflects uncertainty in our estimates of driver errors, our analyses provide at least weak evidence that LM3V vegetation is too sensitive to water availability. This oversensitivity was not unexpected, given that the model lacks important slow-timescale mechanisms that may moderate ecosystem response to environmental drivers (Leuzinger et al. 2011, Luo et al. 2011*b*). For example, C–N feedbacks may introduce time lags into ecosystem dynamics that dampen vegetation response to interannual variation in temperature and precipitation, as observed in some coupled C–N TBMs (Thornton et al. 2007, Gerber et al. 2010). Inflexible allocation (to roots, leaves, etc.) in LM3V also likely contributes to its oversensitivity to water availability. Incorporating flexible allocation in response to water stress, as in some TBMs (e.g., Friedlingstein et al. 1999, Ostle et al. 2009, Zaehle and Friend 2010), would likely dampen the LM3V vegetation response to water availability. Finally, the crude representation of biodiversity in LM3V and most other TBMs likely fails to capture important dampening mechanisms in real ecosystems (Rastetter 1996, Leuzinger et al. 2011, Luo et al. 2011*b*), such as adaptations of different species or populations to different hydrological regimes (Cavender-Bares et al. 2004, Engelbrecht et al. 2007, McDowell et al. 2008). Despite the above reasons to expect TBMs to be overly sensitive to environmental variation, limitations of current TBMs may in some cases lead to under-sensitivity (Powell et al. 2013).

How is TBM parameter optimization affected if driver errors are ignored?—Leaf allocation in the baseline LM3V model appears too high compared to biometric estimates from FLUXNET sites (compare LM3V-baseline to Observations in Table 2), which likely makes the model too sensitive to soil moisture. Increasing the fraction of NPP allocated to fine roots should decrease model sensitivity to soil moisture, thereby bringing the LM3V functional response into better agreement with that of real forests. However, if driver errors are ignored during optimization, then the sensitivity of the model to soil moisture may be reduced too much. If driver errors are severe, then parameter optimization may completely (and erroneously) eliminate model sensitivity to environmental variation (see *Introduction*).

The data-assimilation exercise we performed, using standard optimization methods that ignore driver errors, yielded estimates for fine-root allocation that are higher than biometric estimates at eastern U.S. FLUXNET sites (compare LM3V-SbasePopt to Observations in Table 2). Although there is clearly a scale mismatch between our 1° LM3V configuration and individual FLUXNET sites, the distribution of stand-scale allocation estimates across FLUXNET sites (e.g., means and variances) should be directly comparable with coarse-scale land models. The FLUXNET synthesis data set (Luyssaert et al. 2007) reports biometric allocation estimates for only three unmanaged eastern U.S. sites (including unmanaged secondary forests), which precludes rigorous model–data comparisons. Nevertheless, the optimized parameter estimates appear unrealistic compared to the available data, and the direction of mismatch is consistent with the bias expected from ignoring driver errors. In an alternative data-assimilation experiment, in which we reduced spatial variation in soil moisture by substituting a uniform AWC map for the baseline map, optimized estimates for leaf and fine-root allocation were consistent with biometric estimates (compare LM3V-SmeanPopt to Observations in Table 2). Furthermore, LM3V with uniform AWC yielded the best model–data fits (Fig. 6d, Appendix B: Figs. B7d, and B8d). These results do not imply that real vegetation is insensitive to soil moisture. The results do, however, suggest the presence of soil driver errors, the presence of model errors (e.g., in hydrological parameters or processes), and/or the absence of important mechanisms in LM3V that moderate the functional response of real vegetation to environmental variation.

LM3V optimization also yielded V_{Cmax} estimates that were inconsistent with (lower than) empirical leaf-level estimates (Table 2). Systematic driver errors (e.g., soil AWC being too high, on average, in the model) may contribute to this mismatch. The absence of nutrient feedbacks in LM3V is unlikely to explain the V_{Cmax} discrepancy, because nutrient effects are already included in the empirical V_{Cmax} estimates (Kattge et al. 2009,

PLATE 1. Physical soil properties are highly variable across space and have a strong effect on vegetation. The photographs show two forests in northern Wisconsin, USA, with contrasting soil texture (left, 84% sand; right, 57% sand) but similar climate. The stands differ markedly in leaf area index (left, 2.25; right, 6), the percent of sunlight penetrating to the ground surface (left, 17%; right, 1%), and fine-root biomass (left, 1.6 mg/cm^3; right, 0.6 mg/cm^3; J. Lichstein, *unpublished data*). Both stands are ~50 years old with an approximately even-aged upper canopy. Photo credits: J. W. Lichstein.

see also Bonan et al. 2012). Bonan et al. (2012) concluded that explicitly representing multiple canopy layers in the CLM4 model produced more realistic levels of within-canopy shading and thereby reduced the discrepancy between empirical leaf-level estimates of V_{Cmax} and "effective V_{Cmax}" in CLM4. Thus, the low V_{Cmax} values obtained from our optimization likely reflect, at least in part, the fact that LM3V does not explicitly represent multiple canopy layers. Additional biases in LM3V that may be compensated for by reducing V_{Cmax} include: errors in model structure and parameters related to hydrology and photosynthesis, and failing to account for partial disturbances (e.g., edge effects and legacies of past selective logging) that affect the real landscapes sampled by FIA (Van Tuyl et al. 2005, Brown et al. 1997).

Is the existing network of intensive study sites adequate to constrain the functional response of TBMs within biomes?—Given the complications that arise from driver errors in geographically extensive forest inventories, an alternative approach to testing/improving TBMs might be to rely on intensive study sites where measurement errors are small. How many intensive sites are needed to quantify TBM functional responses? We found that at least 40 sites are needed to have an 80% chance of estimating a growth–precipitation slope within 20% of the true value. Forty sites is probably an underestimate, because our power analysis addressed the simplistic case of a single environmental gradient with zero measurement error. Even this conservative estimate is larger than the number of intensive study sites typically available to estimate within-biome functional responses. For example, if the FLUXNET synthesis data set (Luyssaert et al. 2007) is restricted to "unmanaged" sites, then none of the seven climatic regions (boreal humid, boreal semiarid, Mediterranean warm, temperate humid, temperate semiarid, tropical humid, tropical semiarid) includes more than 24 sites. Thus, while FLUXNET may be sufficient to characterize functional responses across global PFTs (Luyssaert et al. 2007), it appears inadequate to constrain the within-PFT functional responses that have been largely ignored by the global modeling community. It may be possible to constrain within-PFT functional responses by pooling data from multiple networks, but this would require that soil properties are quantified with little error and in a manner that is both consistent across networks and comparable to hydrological sub-models in TBMs. Given these stringent requirements, we suggest that a more practical approach to quantifying within-PFT functional responses is to leverage the vast amount of data available from geographically extensive forest inventories by quantifying and accounting for driver errors.

Accounting for driver errors in model–data fusion

Obtaining unbiased estimates of ecosystem functional responses is an important step in benchmarking TBMs, but a further challenge is to develop data-assimilation methods that account for driver errors while estimating TBM parameter values. One strategy would be to employ a hierarchical Bayesian (HB) modeling framework, in which each driver variable is treated as an unknown parameter ("latent variable") from a probability distribution (Clark 2005, Ogle and Barber 2008, Lichstein et al. 2010). This is a potentially high-dimensional problem because there are many driver variables (e.g., sub-daily meteorological drivers in some TBMs), but it may be possible to reduce the dimensionality greatly, e.g., using a principal component analysis of meteorological drivers. Although rigorous and robust, the HB approach may be impractical, because the Markov chain Monte Carlo algorithms used to solve these problems (Gilks et al. 1996, Zobitz et al. 2011) typically require $>10^5$ iterations (TBM runs) to converge, and possibly orders of magnitude more with high-dimensional problems.

Another option for data assimilation with error propagation is to use ensemble filtering (EF) methods, such as the ensemble Kalman filter (Reichle et al. 2002, Evensen 2003) or particle filters for nonlinear systems with non-normal errors (e.g., Hill et al. 2011). EF methods may be used to estimate model state variables and/or parameters (e.g., Williams et al. 2005, Mo et al. 2008, Gao et al. 2011), and could accommodate driver errors by using a different randomly generated driver data set (from an estimated multivariate distribution, as in the HB approach above) for each ensemble member. The computational expense of the EF approach may be less than that of the HB approach and can be distributed across a parallel computing system, but the computational costs would still be large (e.g., $>10^4$ TBM runs).

An alternative to data assimilation in the presence of driver errors that avoids the computational expense of the above approaches would be to use the driver-error-corrected functional response estimated from observations (e.g., inventory data), rather than the observations themselves, as the optimization target. Again, "functional response" here means the relationship between ecosystem-level properties (e.g., NPP) and one or more environmental variables (e.g., climate or soil physical properties). Consider a simplistic example, in which NPP responds linearly to mean annual precipitation. Suppose we estimate from observations the slope of NPP vs. precipitation using appropriate methods to account for precipitation errors (e.g., the MEMs used in this paper). We seek an optimization algorithm that adjusts the TBM model parameters so that a regression of modeled NPP vs. precipitation yields the same slope as the driver-error-corrected slope estimated for the real ecosystem. Again, standard optimization algorithms that ignore driver errors and adjust TBM parameter values so as to fit modeled NPP to observed NPP will not yield the correct slope.

To implement the proposed functional-response-based approach, let θ be a parameter vector describing the functional response; e.g., the regression intercept and slopes if the functional response is linear, with one slope for each driver variable (temperature, precipitation, soil, and so on). Note that θ is comprised of parameters from a statistical model, not process-level TBM parameters such as those in Table 2. Estimates and uncertainties for θ are quantified by joint probability density functions $f_{\mathrm{D}}(\theta)$ and $f_{\mathrm{M}}(\theta)$ for the real ecosystem and TBM, respectively. Suppose we estimate $f_{\mathrm{D}}(\theta)$ from observations, using MEMs to account for driver errors. The optimization goal then is to adjust the TBM parameters (e.g., the process-level parameters in Table 2 that determine the TBM's functional response) so as to bring $f_{\mathrm{M}}(\theta)$ into agreement with $f_{\mathrm{D}}(\theta)$. This requires specifying an optimization cost function in terms of the distributions $f_{\mathrm{D}}(\theta)$ and $f_{\mathrm{M}}(\theta)$, rather than in terms of predictions and observations, as is typically done (Raupach et al. 2005, Williams et al. 2009). A natural choice for the cost function is the Kullback-Leibler divergence (Kullback and Leibler 1951, Burnham and Anderson 2004): $D_{\mathrm{KL}}(f_{\mathrm{D}}, f_{\mathrm{M}}) = \int f_{\mathrm{D}}(\theta)\ln[f_{\mathrm{D}}(\theta)/f_{\mathrm{M}}(\theta)]d\theta$, which is a measure of the information lost when f_{M} is used to approximate f_{D}, and which can be expressed in closed form if f_{M} and f_{D} are multivariate normal (as in OLS regression, and as is asymptotically so in likelihood and Bayesian contexts).

The approaches outlined above to account for driver errors in model–data fusion are preliminary proposals. These, and other, candidate approaches should be studied in detail; e.g., by analyzing artificial data sets (created from model output, with driver errors added to the "data") to assess each method's capacity to recover known parameter values. We speculate that the functional-response-based approach will allow for rapid progress in improving TBMs because its modest computational requirements make it accessible to a large group of researchers.

CONCLUSIONS

The C-cycle modeling community has made great progress over the last decade in developing data sets and protocols to rigorously test TBMs with a variety of data sources (Raupach et al. 2005, Randerson et al. 2009, Williams et al. 2009, Luo et al. 2011*a*). However, TBMs have yet to incorporate key slow timescale processes, such as within-PFT shifts in plant species composition, that are likely to moderate C-cycle responses to global change (Rastetter 1996, Leuzinger et al. 2011, Luo et al. 2011*b*). Geographically extensive forest inventories include data from many thousands of inventory plots that could be used to constrain slow timescale processes in TBMs. Such data sets present special challenges for model–data fusion, due to the potentially large errors in driver variables at inventory plots. Ignoring these errors can lead to (1) biased estimates of ecosystem response to environmental gradients, thereby exaggerating the discrepancy between TBMs and real ecosystems; and (2) biased TBM parameter estimates that impose too weak of a response to environmental gradients. Statistical methods are available to account for driver errors when estimating ecosystem response to environmental gradients. Data-assimilation methods that account for driver errors when optimizing TBM parameters need to be developed.

Acknowledgments

We thank the USDA Forest Service (09-JV-11242306-051 and 11-JV-11242306-059), the U.S. DOE (DE-FC02-06ER64157), the Princeton Carbon Mitigation Initiative (sponsored by BP and Ford), and the NOAA GFDL for funding. We thank Richard McCullough and Liz LaPoint for facilitating access to FIA plot coordinates. We thank Ensheng Weng, Mark Vanderwel, and two anonymous reviewers for helpful comments on an earlier draft, and P. C. D. Milly for useful conversations about LM3V hydrology. We thank the North American Carbon Program Site-Level Interim Synthesis team and the Oak Ridge National Laboratory Distributed Active Archive Center for organizing and distributing the meteorological observations used to quantify errors in the PRISM climate data.

Literature Cited

Barrett, D. J. 2002. Steady state turnover time of carbon in the Australian terrestrial biosphere. Global Biogeochemical Cycles 16:1108.

Bechtold, W. A., and P. L. Patterson. 2005. The enhanced Forest Inventory and Analysis program. National sampling design and estimation procedures. General Technical Report SRS-80. USDA Forest Service, Southern Research Station, Asheville, North Carolina, USA.

Bolker, B. M. 2008. Ecological models and data in R. Princeton University Press, Princeton, New Jersey, USA.

Bonan, G. B., K. W. Oleson, R. A. Fisher, G. Lasslop, and M. Reichstein. 2012. Reconciling leaf physiological traits and canopy flux data: Use of the TRY and FLUXNET databases in the Community Land Model version 4. Journal of Geophysical Research 117:G02026.

Brown, S., P. Schroeder, and R. Birdsey. 1997. Aboveground biomass distribution of US eastern hardwood forests and the use of large trees as an indicator of forest development. Forest Ecology and Management 96:37–47.

Burnham, K. P., and D. R. Anderson. 2004. Multimodel inference: understanding AIC and BIC in model selection. Sociological Methods and Research 33:261–304.

Cavender-Bares, J., K. Kitajima, and F. A. Bazzaz. 2004. Multiple trait associations in relation to habitat differentiation among 17 Floridian oak species. Ecological Monographs 74:635–662.

Chesher, A. 1991. The effect of measurement error. Biometrika 78:451–462.

Clark, J. S. 2005. Why environmental scientists are becoming Bayesians. Ecology Letters 8:2–14.

Collatz, G. J., J. T. Ball, C. Grivet, and J. A. Berry. 1991. Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer. Agricultural and Forest Meteorology 54:107–136.

Cook, B. D., K. J. Davis, W. G. Wang, A. Desai, B. W. Berger, R. M. Teclaw, J. G. Martin, P. V. Bolstad, P. S. Bakwin, C. X. Yi, and W. Heilman. 2004. Carbon exchange and venting anomalies in an upland deciduous forest in northern Wisconsin, USA. Agricultural and Forest Meteorology 126:271–295.

Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris. 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology 28:2031–2064.

Denman, K. L., et al. 2007. Couplings between changes in the climate system and biogeochemistry. Pages 499–587 *in* S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, editors. Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK.

Desai, A. R., P. V. Bolstad, B. D. Cook, K. J. Davis, and E. V. Carey. 2005. Comparing net ecosystem exchange of carbon dioxide between an old-growth and mature forest in the upper Midwest, USA. Agricultural and Forest Meteorology 128:33–55.

Drohan, P. J., E. J. Ciolkosz, and G. W. Petersen. 2003. Soil survey mapping unit accuracy in forested field plots in northern Pennsylvania. Soil Science Society of America Journal 67:208–214.

Dunne, K. A., and C. J. Willmott. 1996. Global distribution of plant-extractable water capacity of soil. International Journal of Climatology 16:841–859.

Engelbrecht, B. M. J., L. S. Comita, R. Condit, T. A. Kursar, M. T. Tyree, B. L. Turner, and S. P. Hubbell. 2007. Drought sensitivity shapes species distribution patterns in tropical forests. Nature 447:80–83.

Evensen, G. 2003. The ensemble Kalman filter: Theoretical formulation and practical implementation. Ocean Dynamics 53:343–367.

Farquhar, G. D., S. V. Caemmerer, and J. A. Berry. 1980. A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species. Planta 149:78–90.

Fischer, M. L., D. P. Billesbach, J. A. Berry, W. J. Riley, and M. S. Torn. 2007. Spatiotemporal variations in growing season exchanges of CO_2, H_2O, and sensible heat in agricultural fields of the Southern Great Plains. Earth Interactions 11:1–22.

Fletcher, R. 1987. Practical methods of optimization. John Wiley, Chichester, UK.

Foley, J. A., I. C. Prentice, N. Ramankutty, S. Levis, D. Pollard, S. Sitch, and A. Haxeltine. 1996. An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics. Global Biogeochemical Cycles 10:603–628.

Friedlingstein, P., et al. 2006. Climate–carbon cycle feedback analysis: Results from the C^4MIP model intercomparison. Journal of Climate 19:3337–3353.

Friedlingstein, P., G. Joel, C. B. Field, and I. Y. Fung. 1999. Toward an allocation scheme for global terrestrial carbon models. Global Change Biology 5:755–770.

Fuller, W. A. 1987. Measurement error models. John Wiley, New York, New York, USA.

Gao, C., H. Wang, E. Weng, S. Lakshmivarahan, Y. Zhang, and Y. Luo. 2011. Assimilation of multiple data sets with the ensemble Kalman filter to improve forecasts of forest carbon dynamics. Ecological Applications 21:1461–1473.

Gerber, S., L. O. Hedin, M. Oppenheimer, S. W. Pacala, and E. Shevliakova. 2010. Nitrogen cycling and feedbacks in a global dynamic land model. Global Biogeochemical Cycles 24:GB1001.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter. 1996. Markov chain Monte Carlo in practice. Chapman and Hall/CRC, Boca Raton, Florida, USA.

Goffe, W. L., G. D. Ferrier, and J. Rogers. 1994. Global optimization of statistical functions with simulated annealing. Journal of Econometrics 60:65–99.

Gough, C. M., C. S. Vogel, H. P. Schmid, H.-B. Su, and P. S. Curtis. 2008. Multi-year convergence of biometric and meteorological estimates of forest carbon storage. Agricultural and Forest Meteorology 148:158–170.

Griliches, Z., and V. Ringstad. 1970. Error-in-the-variables bias in nonlinear contexts. Econometrica 38:368–370.

Gu, L., T. Meyers, S. G. Pallardy, P. J. Hanson, B. Yang, M. Heuer, K. P. Hosman, J. S. Riggs, D. Sluss, and S. D. Wullschleger. 2006. Direct and indirect effects of atmospheric conditions and soil moisture on surface energy partitioning revealed by a prolonged drought at a temperate forest site. Journal of Geophysical Research 111:D16102.

Hill, T. C., T. Quaife, and M. Williams. 2011. A data assimilation method for using low-resolution Earth observation data in heterogeneous ecosystems. Journal of Geophysical Research 116:D08117.

Hollinger, D. Y., S. M. Goltz, E. A. Davidson, J. T. Lee, K. Tu, and H. T. Valentine. 1999. Seasonal patterns and environmental control of carbon dioxide and water vapour exchange in an ecotonal boreal forest. Global Change Biology 5:891–902.

Hollinger, D. Y., and A. D. Richardson. 2005. Uncertainty in eddy covariance measurements and its application to physiological models. Tree Physiology 25:873–885.

Jackson, R. B., J. Canadell, J. R. Ehleringer, H. A. Mooney, O. E. Sala, and E. D. Schulze. 1996. A global analysis of root distributions for terrestrial biomes. Oecologia 108:389–411.

Jenkins, J. C., D. C. Chojnacky, L. S. Heath, and R. A. Birdsey. 2003. National-scale biomass estimators for United States tree species. Forest Science 49:12–35.

Kalnay, E., et al. 1996. The NCEP/NCAR 40-year reanalysis project. Bulletin of the American Meteorological Society 77:437–471.

Kattge, J., W. Knorr, T. Raddatz, and C. Wirth. 2009. Quantifying photosynthetic capacity and its relationship to leaf nitrogen content for global-scale terrestrial biosphere models. Global Change Biology 15:976–991.

Keenan, T. F., M. S. Carbone, M. Reichstein, and A. D. Richardson. 2011. The model–data fusion pitfall: assuming certainty in an uncertain world. Oecologia 167:587–597.

Kistler, R., et al. 2001. The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. Bulletin of the American Meteorological Society 82:247–267.

Krinner, G., N. Viovy, N. de Noblet-Ducoudre, J. Ogee, J. Polcher, P. Friedlingstein, P. Ciais, S. Sitch, and I. C. Prentice. 2005. A dynamic global vegetation model for studies of the coupled atmosphere–biosphere system. Global Biogeochemical Cycles 19:GB1015.

Kucharik, C. J., J. A. Foley, C. Delire, V. A. Fisher, M. T. Coe, J. D. Lenters, C. Young-Molling, N. Ramankutty, J. M. Norman, and S. T. Gower. 2000. Testing the performance of a Dynamic Global Ecosystem Model: Water balance, carbon balance, and vegetation structure. Global Biogeochemical Cycles 14:795–825.

Kullback, S., and R. A. Leibler. 1951. On information and sufficiency. Annals of Mathematical Statistics 22:79–86.

Leuning, R. 1995. A critical appraisal of a combined stomatal-photosynthesis model for C_3 plants. Plant, Cell and Environment 18:339–355.

Leuzinger, S., Y. Luo, C. Beier, W. Dieleman, S. Vicca, and C. Körner. 2011. Do global change experiments overestimate impacts on terrestrial ecosystems? Trends in Ecology and Evolution 26:236–241.

Lichstein, J. W., J. Dushoff, K. Ogle, A. P. Chen, D. W. Purves, J. P. Caspersen, and S. W. Pacala. 2010. Unlocking the forest inventory data: relating individual tree performance to unmeasured environmental factors. Ecological Applications 20:684–699.

Luo, Y., K. Ogle, C. Tucker, S. Fei, C. Gao, S. LaDeau, J. S. Clark, and D. S. Schimel. 2011*a*. Ecological forecasting and data assimilation in a data-rich era. Ecological Applications 21:1429–1442.

Luo, Y. Q., et al. 2011*b*. Coordinated approaches to quantify long-term ecosystem dynamics in response to global change. Global Change Biology 17:843–854.

Luo, Y. Q., L. W. White, J. G. Canadell, E. H. DeLucia, D. S. Ellsworth, A. C. Finzi, J. Lichter, and W. H. Schlesinger. 2003. Sustainability of terrestrial carbon sequestration: A case study in Duke Forest with inversion approach. Global Biogeochemical Cycles 17:1021.

Luyssaert, S., et al. 2007. CO_2 balance of boreal, temperate, and tropical forests derived from a global database. Global Change Biology 13:2509–2537.

McDowell, N., et al. 2008. Mechanisms of plant survival and mortality during drought: why do some plants survive while others succumb to drought? New Phytologist 178:719–739.

Medvigy, D., S. C. Wofsy, J. W. Munger, D. Y. Hollinger, and P. R. Moorcroft. 2009. Mechanistic scaling of ecosystem function and dynamics in space and time: Ecosystem Demography model version 2. Journal of Geophysical Research 114:G01002.

Milly, P. C. D., and A. B. Shmakin. 2002. Global modeling of land water and energy balances. Part I: the land dynamics (LaD) model. Journal of Hydrometeorology 3:283–299.

Mitchell, T. D., and P. D. Jones. 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. International Journal of Climatology 25:693–712.

Mo, X., J. M. Chen, W. Ju, and T. A. Black. 2008. Optimization of ecosystem model parameters through assimilating eddy covariance flux data with an ensemble Kalman filter. Ecological Modelling 217:157–173.

New, M., M. Hulme, and P. Jones. 1999. Representing twentieth-century space-time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology. Journal of Climate 12:829–856.

New, M., M. Hulme, and P. Jones. 2000. Representing twentieth-century space-time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. Journal of Climate 13:2217–2238.

Ogle, K., and J. J. Barber. 2008. Bayesian data-model integration in plant physiological and ecosystem ecology. Progress in Botany 69:281–311.

Oren, R., C. I. Hseih, P. Stoy, J. Albertson, H. R. McCarthy, P. Harrell, and G. G. Katul. 2006. Estimating the uncertainty in annual net ecosystem carbon exchange: spatial variation in turbulent fluxes and sampling errors in eddy-covariance measurements. Global Change Biology 12:883–896.

Ostle, N., et al. 2009. Integrating plant-soil interactions into global carbon cycle models. Journal of Ecology 97:851–863.

Powell, T. L., et al. 2013. Confronting model predictions of carbon fluxes with measurements of Amazon forests subjected to experimental drought. New Phytologist 200:350–364.

R Development Core Team. 2012. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Randerson, J. T., et al. 2009. Systematic assessment of terrestrial biogeochemistry in coupled climate–carbon models. Global Change Biology 15:2462–2484.

Rastetter, E. B. 1996. Validating models of ecosystem response to global change. BioScience 46:190–198.

Raupach, M. R., P. J. Rayner, D. J. Barrett, R. S. DeFries, M. Heimann, D. S. Ojima, S. Quegan, and C. C. Schmullius. 2005. Model–data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. Global Change Biology 11:378–397.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi. 2002. Hydrologic data assimilation with the ensemble Kalman filter. Monthly Weather Review 130:103–114.

Richardson, A. D., et al. 2010. Estimating parameters of a forest ecosystem C model with measurements of stocks and fluxes as joint constraints. Oecologia 164:25–40.

Rogowski, A. S., and J. K. Wolf. 1994. Incorporating variability into soil map unit delineations. Soil Science Society of America Journal 58:163–174.

Schaefer, K., et al. 2012. A model–data comparison of gross primary productivity: Results from the North American Carbon Program site synthesis. Journal of Geophysical Research 117:G03010.

Schennach, S. M. 2004. Estimation of nonlinear models with measurement error. Econometrica 72:33–75.

Schmid, H. P., C. S. B. Grimmond, F. Cropley, B. Offerle, and H. B. Su. 2000. Measurements of CO_2 and energy fluxes over a mixed hardwood forest in the mid-western United States. Agricultural and Forest Meteorology 103:357–374.

Sheffield, J., G. Goteti, and E. F. Wood. 2006. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. Journal of Climate 19:3088–3111.

Sheffield, J., A. Ziegler, E. Wood, and Y. Chen. 2004. Correction of the high-latitude rain day anomaly in the NCEP-NCAR reanalysis for land surface hydrological modeling. Journal of Climate 17:3814–3828.

Shevliakova, E., S. W. Pacala, S. Malyshev, G. C. Hurtt, P. C. D. Milly, J. P. Caspersen, L. T. Sentman, J. P. Fisk, C. Wirth, and C. Crevoisier. 2009. Carbon cycling under 300 years of land use change: Importance of the secondary vegetation sink. Global Biogeochemical Cycles 23:GB2022.

Shinozaki, K., K. Yoda, K. Hozumi, and T. Kira. 1964. A quantitative analysis of plant form—the pipe model theory. I. Basic analyses. Japanese Journal of Ecology 14:97–105.

Sitch, S., 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.

Soil Survey Division Staff. 1993. Soil survey manual. Soil Conservation Service, U.S. Department of Agriculture, Washington, D.C., USA.

Spadavecchia, L., M. Williams, and B. E. Law. 2011. Uncertainty in predictions of forest carbon dynamics: separating driver error from model error. Ecological Applications 21:1506–1522.

Stoy, P. C., G. G. Katul, M. B. S. Siqueira, J.-Y. Juang, K. A. Novick, H. R. McCarthy, A. C. Oishi, and R. Oren. 2008. Role of vegetation in determining carbon sequestration along ecological succession in the southeastern United States. Global Change Biology 14:1409–1427.

Sulman, B. N., A. R. Desai, B. D. Cook, N. Saliendra, and D. S. Mackay. 2009. Contrasting carbon dioxide fluxes between a drying shrub wetland in Northern Wisconsin, USA, and nearby forests. Biogeosciences 6:1115–1126.

Suyker, A. E., S. B. Verma, and G. G. Burba. 2003. Interannual variability in net CO_2 exchange of a native tallgrass prairie. Global Change Biology 9:255–265.

Thomas, P. J., J. C. Baker, and T. W. Simpson. 1989. Variability of the Cecil Map Unit in Appomattox County, Virginia. Soil Science Society of America Journal 53:1470–1474.

Thornton, P. E., J. F. Lamarque, N. A. Rosenbloom, and N. M. Mahowald. 2007. Influence of carbon-nitrogen cycle coupling on land model response to CO_2 fertilization and climate variability. Global Biogeochemical Cycles 21:GB4018.

Urbanski, S., C. Barford, S. Wofsy, C. Kucharik, E. Pyle, J. Budney, K. McKain, D. Fitzjarrald, M. Czikowsky, and J. W. Munger. 2007. Factors controlling CO_2 exchange on timescales from hourly to decadal at Harvard Forest. Journal of Geophysical Research 112:G02020.

Van Tuyl, S., B. E. Law, D. P. Turner, and A. I. Gitelman. 2005. Variability in net primary production and carbon storage in biomass across Oregon forests—an assessment integrating data from forest inventories, intensive sites, and remote sensing. Forest Ecology and Management 209:273–291.

Verma, S. B., et al. 2005. Annual carbon dioxide exchange in irrigated and rainfed maize-based agroecosystems. Agricultural and Forest Meteorology 131:77–96.

Williams, M., et al. 2009. Improving land surface models with FLUXNET data. Biogeosciences 6:1341–1359.

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

Zaehle, S., and A. D. Friend. 2010. Carbon and nitrogen cycle dynamics in the O-CN land surface model: 1. Model description, site-scale evaluation, and sensitivity to parameter estimates. Global Biogeochemical Cycles 24:GB1005.

Zhou, T., and Y. Q. Luo. 2008. Spatial patterns of ecosystem carbon residence time and NPP-driven carbon uptake in the conterminous United States. Global Biogeochemical Cycles 22:GB3032.

Zobitz, J. M., A. R. Desai, D. J. P. Moore, and M. A. Chadwick. 2011. A primer for data assimilation with ecological models using Markov Chain Monte Carlo (MCMC). Oecologia 167:599–611.

Supplemental Material

Appendix A

Supplementary methods for Forest Inventory and Analysis (FIA) data filtering and analysis, measurement error models, and Geophysical Fluid Dynamics Laboratory LM3V dynamic global vegetation model optimization (*Ecological Archives* A024-041-A1).

Appendix B

Supplementary table (regression results for FIA data and LM3V output) and figures (*Ecological Archives* A024-041-A2).