

Unlocking the forest inventory data: relating individual tree performance to unmeasured environmental factors

JEREMY W. LICHSTEIN,^{1,6} JONATHAN DUSHOFF,² KIONA OGLE,³ ANPING CHEN,¹ DREW W. PURVES,⁴ JOHN P. CASPERSEN,⁵
AND STEPHEN W. PACALA¹

¹*Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544 USA*

²*Department of Biology, McMaster University, Hamilton, Ontario L8S4K1 Canada*

³*Departments of Botany and Statistics, University of Wyoming, Laramie, Wyoming 82071 USA*

⁴*Computational Ecology and Environmental Science Group, Microsoft Research, Cambridge CB30FP United Kingdom*

⁵*Faculty of Forestry, University of Toronto, Toronto, Ontario M5S 3B3 Canada*

Abstract. Geographically extensive forest inventories, such as the USDA Forest Service's Forest Inventory and Analysis (FIA) program, contain millions of individual tree growth and mortality records that could be used to develop broad-scale models of forest dynamics. A limitation of inventory data, however, is that individual-level measurements of light (L) and other environmental factors are typically absent. Thus, inventory data alone cannot be used to parameterize mechanistic models of forest dynamics in which individual performance depends on light, water, nutrients, etc. To overcome this limitation, we developed methods to estimate species-specific parameters (SG) relating sapling growth (G) to L using data sets in which G , but not L , is observed for each sapling. Our approach involves: (1) using calibration data that we collected in both eastern and western North America to quantify the probability that saplings receive different amounts of light, conditional on covariates x that can be obtained from inventory data (e.g., sapling crown class and neighborhood crowding); and (2) combining these probability distributions with observed G and x to estimate SG using Bayesian computational methods. Here, we present a test case using a data set in which G , L , and x were observed for saplings of nine species. This test data set allowed us to compare estimates of SG obtained from the standard approach (where G and L are observed for each sapling) to our method (where G and x , but not L , are observed). For all species, estimates of SG obtained from analyses with and without observed L were similar. This suggests that our approach should be useful for estimating light-dependent growth functions from inventory data that lack direct measurements of L . Our approach could be extended to estimate parameters relating sapling mortality to L from inventory data, as well as to deal with uncertainty in other resources (e.g., water or nutrients) or environmental factors (e.g., temperature).

Key words: *Bayesian hierarchical model; dynamic global vegetation model; forest inventory; latent variable; Markov chain Monte Carlo (MCMC); neighborhood analysis; numerical integration; shade tolerance; understory light.*

INTRODUCTION

Efforts to understand interactions between the carbon cycle, land use, and climate rely on geographically extensive models of vegetation dynamics (Foley et al. 1996, Cramer et al. 2001, Hurtt et al. 2002, Sitch et al. 2005, Albani et al. 2006, Friedlingstein et al. 2006, Moorcroft 2006). A number of "dynamic global vegetation models" (DGVMs) have been developed that can reproduce observed broad scale patterns in plant biomass and net primary production (NPP) (e.g., Foley et al. 1996, Sitch et al. 2003, Krinner et al. 2005). However, such models lead to widely different predictions regarding the future state of the biosphere, even when forced with the same climate and CO₂ scenarios (Cramer et al.

2001). Allowing for dynamic climate and CO₂ leads to even greater uncertainty due to feedbacks between climate and the carbon cycle (Friedlingstein et al. 2006). Vegetation plays a major role in this uncertainty. For example, in a comparison of 11 coupled climate-carbon cycle models, terrestrial NPP was the largest source of uncertainty in predicting future atmospheric CO₂ (Friedlingstein et al. 2006, Denman et al. 2007).

Dynamic global vegetation models require numerous parameters to describe the physiology of plant functional types (e.g., Foley et al. 1996, Sitch et al. 2003, Krinner et al. 2005). Typically, these parameters are assigned one of many possible values from the literature. Rarely, however, are the implied whole-plant performances (i.e., growth, mortality, and reproduction) compared to individual, whole-plant data. In fact, most DGVMs bypass the individual level altogether and attempt to scale directly from finer-scale processes (e.g., leaf photosynthesis) to ecosystems (e.g., forest NPP)

Manuscript received 18 December 2008; revised 5 June 2009; accepted 2 July 2009. Corresponding Editor: A. D. McGuire.

⁶ E-mail: JWL@princeton.edu

(Purves and Pacala 2008). While this approach is appealing in its simplicity, it is problematic due to the nonlinear effects of individual-level interactions (e.g., height-structured competition for light) on ecosystem dynamics (Moorcroft et al. 2001, Strigul et al. 2008).

Forest inventories provide a vast source of data on individual tree performance that could be harnessed to develop broad scale forest dynamics models that are parameterized at the individual level using formal, quantitative methods. Together, inventory data sources, including systematic government inventories, such as the USDA Forest Service's Forest Inventory and Analysis (FIA) program, as well as other networks of permanent plots such as those monitored by the Center for Tropical Forest Science, include on the order of 10^7 individual growth and mortality records spanning all major forest types.

Recently, Purves et al. (2008) showed that a computationally and analytically tractable forest dynamics model (Adams et al. 2007, Strigul et al. 2008), parameterized with FIA data, can accurately predict long-term (>100 years) community dynamics on different soil types in the U.S. Lake States. The model is mechanistic at the community level, in that it explicitly accounts for height-structured competition in scaling up from individual performance. The version implemented by Purves et al. (2008), however, is phenomenological at the individual level; i.e., growth and mortality rates of understory vs. canopy individuals were estimated for different species on different soil types, without attempting to explain variation in these rates in terms of resources (e.g., light, water, nutrients) or other environmental factors (e.g., temperature). This is an important limitation, because an explicit treatment of physiological mechanisms is needed to predict vegetation response to climate change and elevated CO_2 (Reynolds et al. 2001).

An obstacle to using forest inventory data to estimate resource-dependent growth and mortality functions is that inventories typically lack measurements of individual-level resource availabilities. In the absence of such measurements, resource-dependent performance functions could still be estimated if the probability distribution of resource availabilities were known (or could be estimated) for each individual. In this paper, we present a statistical model that estimates the probability that individual saplings receive different amounts of light, conditional on covariates (e.g., crown class and neighborhood crowding) that can be extracted from widely available inventory data. We then show how this light model can be used to estimate species-specific parameters relating sapling growth to light from data sets in which growth, but not light, is observed for each sapling. We test the methods using data from sites in eastern and western North America.

OVERVIEW OF METHODS

Our goal is to estimate parameters θ_G relating sapling growth rate (G) to light availability (L). For example, if

G increases with L according to a two-parameter curve, then θ_G might include a high-light growth rate and an initial slope. Unlike the standard case, in which G and L are observed for each sapling, our methods are designed to estimate θ_G using data sets in which G , but not L , is observed. Our approach relies on the availability, for each sapling, of one or more covariates (x), such as a neighborhood crowding index, that have information about unobserved L .

First, consider the standard "observed- L " case in which G and L are observed for $i = 1, 2, \dots, N$ individuals. The likelihood of the growth data is $L = \prod_{i=1}^N p(G_i | L_i, \theta_G)$, where $p(\cdot)$ is the probability density of observed G_i given observed L_i and particular values of θ_G . We quantify the values of θ_G that maximize L (maximum likelihood analysis) or we describe the probability distribution of θ_G (Bayesian analysis).

Now, consider the "unobserved- L " case in which G and x are observed for each sapling, but L is not. Although we have no direct observations of L , suppose we know from a previous study the values of parameters θ_L that relate L to x . Specifically, suppose we know the probabilities associated with all possible values of each unobserved L , given particular values of x . These probabilities are described by the density function $f_L(L_i | x_i, \theta_L)$. To account for uncertainty in unobserved L when estimating θ_G , we define the likelihood as $L = \prod_{i=1}^N \int_0^1 p(G_i | x_i, \theta_G, \theta_L) f_L(L_i | x_i, \theta_L) dL_i$, where $p(G_i | x_i, \theta_G, \theta_L) = \int_0^1 p(G_i | L_i, \theta_G) f_L(L_i | x_i, \theta_L) dL_i$. Thus, we integrate each sapling's likelihood component over the probability density of unobserved L from zero to full sunlight ($L = 1$). We call this the "uncertain- L ," method. Alternatively, we could ignore uncertainty in L and take the expectation of $f_L(L_i | x_i, \theta_L)$ as a point estimate for each unobserved L_i . We could then treat these point estimates as if they were observed values and estimate θ_G as in the observed- L case. We call this the "predicted- L ," method.

Above, we assume θ_L is known, but in practice, we must estimate θ_L from field data and (in the uncertain- L , method) propagate uncertainty in θ_L when estimating θ_G . Thus, our approach involves the following steps: (1) measure L and x for saplings in stands spanning a range of forest types and disturbance histories; (2) use observed L and x from step 1 to estimate θ_L ; (3) using inventory data sets in which G and x , but not L , are observed for each sapling, estimate θ_L by combining θ_L with observed G and x .

To test our methods, we measured G for a subset of the saplings used to estimate θ_L . These saplings constitute a test data set for which G , L , and x were all observed. This test data set allowed us to compare estimates of θ_G obtained from (1) the observed- L method, which uses observed G and L for each sapling, and (2) our unobserved- L methods (uncertain- L and predicted- L), which use observed L and x to estimate θ_L and then combine θ_L with observed G and x to estimate θ_G .

In the following sections, we present the details of the field methods used to measure G , x , and L , and the

TABLE 1. Glossary of terms.

Term	Definition
α_N	coefficient in Nbr index (Eq. 4); α_N is related to ST by Eq. 6
β_N	exponent in Nbr index (Eq. 4); β_N is assumed constant across species
θ_G	vector of species-specific growth parameters
θ_L	vector of light model parameters that determine f_p (and thus f_L)
ρ	effective leaf area above a focal sapling, defined as $-\ln(L)/k$
$\bar{\rho}$	mean (expectation) of ρ above a focal sapling; $\bar{\rho}$, along with V_p , specifies f_p (and thus f_L)
Φ_G	vector of hyperparameters that specify the lognormal sampling distributions for species-specific growth parameters (θ_G)
b_0, b_1	intercept and slope of relationship between $\ln(\alpha_N)$ and ST (Eq. 6) in models CRN and CRNS (see Table 2 for an explanation)
c_0, c_1	intercept and slope of relationship between $\ln(M)$ and ST (Eq. 7) in models CRS and CRNS (see Table 2 for an explanation); this relationship applies to overtopped saplings only
CC	sapling crown class: overtopped (OT) or sun-exposed (SE)
ENA	eastern North America
f_p	probability density function for ρ ; f_p for sapling i depends on parameters θ_L and covariates $\mathbf{x}_i$
f_L	probability density function for L , obtained by transforming f_p (Eq. 1)
FIA	U.S. Forest Service's Forest Inventory and Analysis program
G	annual dbh growth increment (cm/yr)
$\mathbf{G}$	vector of G values
$\bar{G}$	mean (expectation) of G (Eq. 10)
$\bar{G}_{\max}$	maximum value of $\bar{G}$
k	extinction coefficient in the Beer-Lambert equation relating light at a given point in space to the leaf area index (LAI) above that point: $\text{light} = e^{-k\text{LAI}}$
L	proportion of above-canopy light that is incident on a sapling's crown; observed L was quantified for focal saplings using hemispherical photographs; we assume that L depends on the "effective leaf area" (ρ) above a sapling according to the Beer-Lambert equation, $L = e^{-k\rho}$, with an extinction coefficient (k) of 0.5
$\mathbf{L}$	vector of L values
$\bar{L}$	expectation (mean) of L (Eq. 2), which depends on covariates $\mathbf{x}$ and parameters θ_L
$\hat{L}$	predicted value of L
M	minimum value of $\bar{\rho}$ (Eq. 5)
Nbr	neighborhood crowding index (Eq. 4)
Q	random effect for stand "quality" (Eq. 10) that accounts for stand-level differences in G
$\mathbf{Q}$	vector of Q values
r	species-specific responsiveness of G to Q (Eq. 10)
R^2	proportion of observed variation that is explained (Eq. 16)
S_p	slope of $\bar{\rho}$ as a function of Nbr (Eq. 5)
S_G	initial slope of $\bar{G}$ as a function of L (Eq. 8)
ST	shade tolerance index: proportion of saplings that are in understory (Appendix E: section II)
u	gamma-scale parameter for ρ
V_p	variance of ρ , which is assumed to be proportional to $\bar{\rho}$: $V_p = u\bar{\rho}$
WOR	western Oregon, USA
$\mathbf{x}$	vector of covariates (CC and Nbr) for a focal sapling
$\mathbf{X}$	matrix of covariates for all focal saplings

statistical methods used to estimate θ_L and θ_G . Our methods allow for complete freedom in the functional forms that relate G to L and L to $\mathbf{x}$, as well as the choice of covariates $\mathbf{x}$. Thus, the details presented below should only be viewed as one possible realization of a broad approach to dealing with unmeasured environmental factors. Our paper primarily aims to demonstrate that the approach can be useful, rather than to advocate for specific functional forms or other methodological details.

FIELD METHODS

We estimated light availability (L ; Table 1) using hemispherical photographs for 2128 saplings at sites in eastern North America (ENA) and western Oregon, USA (WOR), spanning a range of forest types and disturbance histories. The stands are described in Appendix A, and sample sizes by species are in Appendix B. For each "focal sapling" (i.e., each sapling for which light was measured), we recorded the spatial

coordinates and diameter at breast height (dbh; measured at 1.37 m above the ground surface) of neighboring saplings (dbh 2.5–12.7 cm) and trees (dbh > 12.7 cm). We define the neighbors of focal sapling i as all stems (excluding i) that would be included in an FIA subplot that includes i . (Each FIA plot contains 4–10 subplots.) Thus, the analyses presented in this paper are tailored to FIA data, but minor modifications would make the methods applicable to other inventories. We mapped neighbors according to the current fixed-radius FIA subplot design (Bechtold and Scott 2005), as well as pre-1999 variable-radius designs (Doman et al. 1981, USDA 1992) in which the limiting distance that determines whether a tree is sampled is proportional to its dbh (Beers and Miller 1964). All analyses in this paper are based on pre-1999 designs, as this period includes most of the currently available FIA remeasurement data (i.e., plots with growth and mortality data). The subplot designs and their implementation at our field sites are described in Appendix C.

Light availability

We estimated L for each focal sapling using WINSCANOPY (Regent Instruments, Sainte-Foy, Quebec, Canada), which includes a remotely triggered camera (Nikon Coolpix 4500, Melville, New York, USA), a hemispherical lens, a self-leveling camera mount, and image analysis software. We used an extendable pole to position the camera over the center of sapling crowns <11 m in height ($>90\%$ of the saplings at our sites). For taller saplings, we used an extendable tower constructed from aluminum ladder segments. This larger device was used at all sites except for Highlands and Haliburton (Appendix A), where we did not estimate L for saplings taller than 11 m.

We used the gap light index (GLI) (Canham 1988, Canham et al. 1994), referred to in WINSCANOPY as "Total Site Factor," to estimate L (the proportion of above-canopy radiation, integrated over the growing season, that reaches the top of a sapling's crown). Estimates of GLI from WINSCANOPY were calibrated with a set of photon sensor measurements (J. W. Lichstein, unpublished data) to minimize bias in L .

Growth

For 579 of the 2128 focal saplings, we estimated G (annual dbh increment) from increment cores or stem cross-sections (see Appendix D for details). Nine species were targeted, including shade-intolerant and -tolerant species in ENA and WOR: *Pinus ponderosa*, *P. strobus*, *P. taeda*, *Pseudotsuga menziesii*, *Tsuga canadensis*, *T. heterophylla*, *Acer rubrum*, *Liquidambar styraciflua*, and *Liriodendron tulipifera*. Sample sizes for G by site and species are in Appendices A and B, respectively.

STATISTICAL METHODS

Modeling sapling light availability

The focal-sapling data from our field sites were used to develop a model for unobserved L in forest inventories. Covariates (x) can provide information about unobserved L , but considerable uncertainty about the true value may remain. Thus, rather than estimating a single value for each unobserved L , we quantify the probabilities associated with all possible values of L . These probabilities are specified by the conditional probability density function $f_L(L|x)$ (Fig. 1). Because L depends on both the density and spatial arrangement of canopy leaves (Kira et al. 1969, Horn 1971, Breda 2003, Mansi and Saeki 2005), we define the "effective leaf area index" above a sapling as $p = -\ln(L)/k$, which we obtain from the Beer-Lambert equation, $L = e^{-kp}$. We assume $k = 0.5$, which is close to many empirical estimates (White et al. 2000); but our results do not depend in any important way on k , as long as it is a positive constant. In this paper, we are concerned only with the proportion of above-canopy radiation (I_0) that reaches a sapling's crown, but it would be straightforward to account for variation in I_0 via the product I_0L .

We model L via p because (1) the logarithmic scale emphasizes differences between small values of L , which are biologically important (Bazzaz 1979, Givnish 1988, Kobe et al. 1995); (2) the transformation requires no extra parameters (assuming fixed k) and (3) in our experience with this problem, parameter estimates converge more quickly when we work on the transformed scale. We assume that p is gamma-distributed with density function $f_p(x)$, and we obtain f_L by a standard change of variables (Fig. 2):

$$f_L = \left| \frac{dp}{dL} \right| f_p = \frac{f_p}{kL}. \quad (1)$$

It is straightforward to show that, if p is gamma-distributed, then the mean (expectation) of L is as follows (Lichstein 2007):

$$\bar{L} = (1 + ku)^{-v} \quad (2)$$

where $u = \sqrt{p}/p$ and $v = p/\sqrt{p}$ are the gamma scale and shape parameters, and p and $\sqrt{p}$ are the mean and variance of p .

We assume that the variance of p is proportional to its mean:

$$V_p = u\bar{p}. \quad (3)$$

Allowing for a power law relationship between V_p and p yields similar results. We considered four candidate models for p that included different combinations of the following covariates (Table 2): neighborhood crowding, crown class, geographic region, and species identity of focal saplings and their neighbors.

1. *Neighborhood.*—We adopt the following neighborhood index for sapling i :

$$\text{Nbr}_i = \sum_{\{h_j > h_i\}} \alpha_N h_j^{\beta_N} / \text{area}_i \quad (4)$$

where h_i is the height of individual i ; $\{h_j > h_i\}$ is the set of individuals in i 's subplot taller than i ; α_N depends on i 's species identity; β_N for simplicity, is assumed constant across species; and area_i , the sample area for i , reflects the variable-radius FIA subplot designs (Appendix C). Measured heights are often unavailable, so we estimated species-specific height-dbh allometries for all U.S. tree species using FIA data (Supplement 1), and we propagated uncertainty in the individual heights through our light model (Appendix E: section 1).

We assume that β_N increases linearly with Nbr :

$$\bar{\beta} = M + S_p \text{Nbr} \quad (5)$$

where M is the minimum and S_p is the slope. We allow for $M > 0$ to account for sampling and process errors; e.g., saplings may be shaded by trees not included in Eq. 4, so $\text{Nbr} = 0$ does not imply $p = 0$ (full sunlight). Replacing Eq. 5 with a saturating curve yields qualitatively similar results (Lichstein 2007).

2. *Crown class.*—We assigned each focal sapling in our field sample to one of two crown classes (CC)

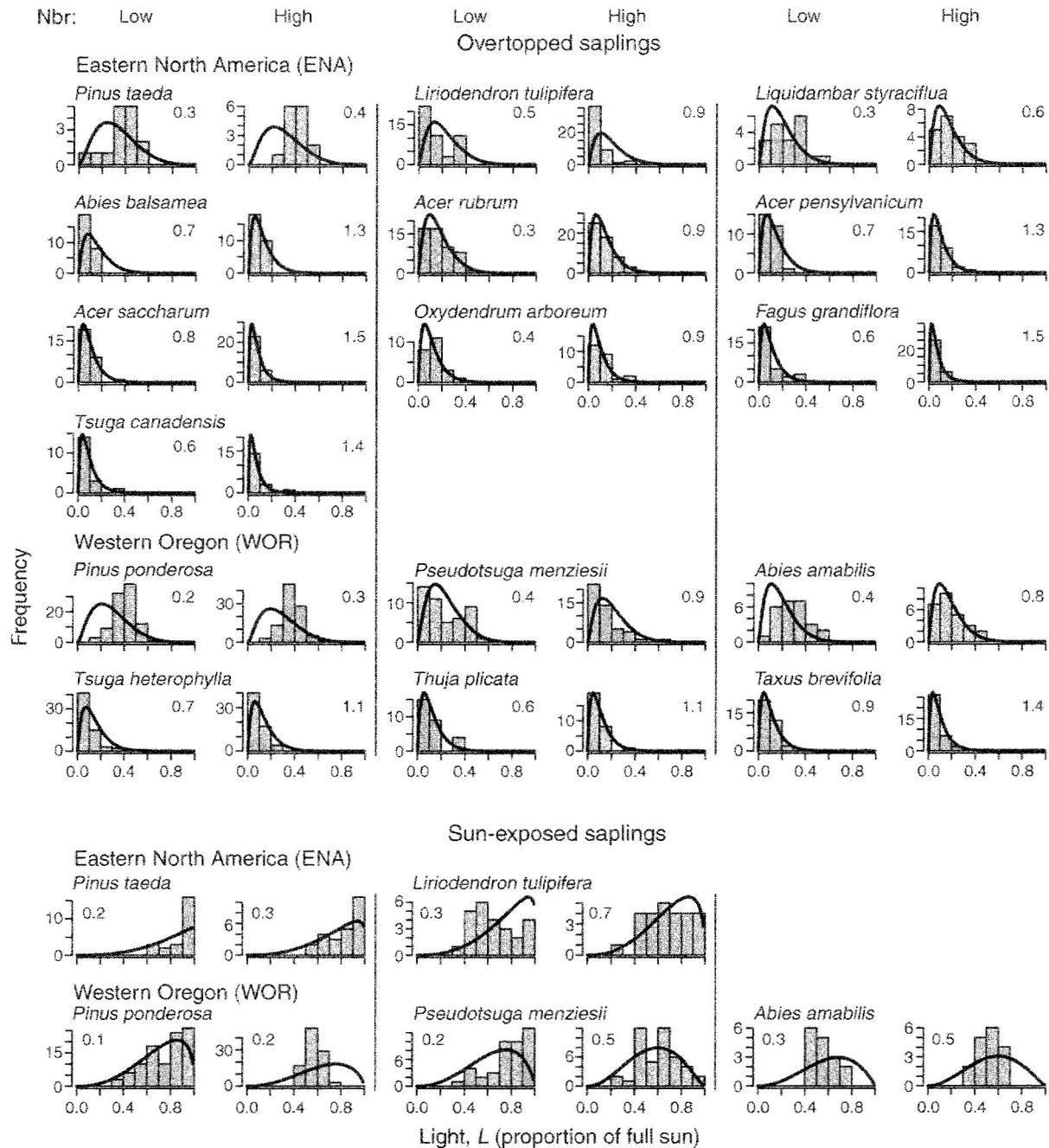


FIG. 1. Probability distributions of light availability, f_L , and histograms of observed light for overtopped and sun-exposed saplings in eastern North America (ENA) and western Oregon, USA (WOR). Results are shown for all species $\times$ crown class combinations with $n \geq 30$ observations. Within each combination, separate histograms are shown for saplings with relatively low (below median) vs. high (above median) values of the neighborhood crowding index (Nbr). The number at the top of each histogram is its median Nbr value, at which f_L is plotted. Lack of fit occurs because the model (model CRNS; see Table 2) was not optimized for each species separately. In each panel, f_L is scaled so that its integral equals the histogram area.

according to FIA definitions (USDA 2005, 2006): "overtopped" (which includes the FIA's overtopped and intermediate classes) or "sun-exposed" (all other FIA classes combined).

3. *Region.*—We allow for region-specific (ENA or WOR) parameters in Eq. 5 to account for differences in

sapling light environments that are not captured by Nbr and CC. Compared to our ENA sites, our WOR sites have greater tree stature and a higher proportion of conifers. We did not account for latitudinal differences in understory light (e.g., Canham et al. 1990) due to the limited latitudinal range of our sites.

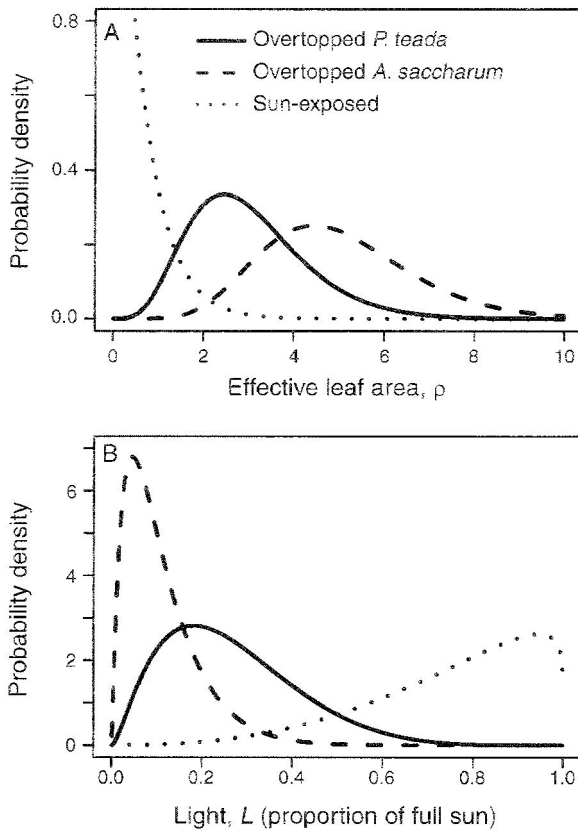


FIG. 2. Examples of probability density functions f_p and f_L for effective leaf area (p) and light (L), respectively. Curves with the same style (solid, dashed, or dotted) in panels (A) and (B) are related by Eq. 1. The distributions are for eastern species from model CRNS (see Table 2) and are plotted at the median Nbr value among all eastern overtopped or sun-exposed saplings. In model CRNS, f_L for overtopped saplings depends on the species shade tolerance index (ST) via the parameter M (Eqs. 5 and 7). *Pinus taeda* and *Acer saccharum* have relatively low and high ST, respectively (see Fig. 3 and Appendix B). See Table 1 for an explanation of variable abbreviations.

4. *Species identity*.—Species may differ in the amount of shade cast by individuals of a given size (i.e., different values of a_N or B_N in Eq. 4), as well as the amount of light received by saplings given CC and Nbr (i.e., different values of M or S_p in Eq. 5). This latter effect

arises if CC and Nbr only partially explain interspecific differences in L so that a species-level bias remains.

Because inventories include species not included in our field sample, we used a non-species-specific approach to account for species differences in shade cast and light received. Our approach is based on the observation that both effects are correlated with species differences in shade tolerance: The amount of shade cast by a tree of a given size tends to increase with shade tolerance (Horn 1971, Canham et al. 1994, 1999). Conversely, the amount of light received by a sapling tends to decrease with shade tolerance (Clark and Clark 1992, Davies 2001, Poorter and Arets 2003). Thus, if an appropriate shade-tolerance index (ST) were available for each species, we would need only to estimate a few parameters that relate the parameters in Eqs. 4 and 5 to ST.

We define ST as the proportion of saplings (2.5–12.7 cm dbh) that are in the understory (as opposed to the canopy). We estimated ST for each U.S. tree species using FIA data (Supplement 1), and we propagated uncertainty in ST through our light model (Appendix E: section II). Shade-tolerant species with high understory survivorship (Kitajima 1994, Kobe et al. 1995) should have proportionately more understory saplings (higher ST) than less tolerant species (Davies 2001, Wright et al. 2003). The shade-tolerance index ST corresponds closely to a widely used shade tolerance classification (Burns and Honkala 1990) (Appendix E: Fig. E 1).

For simplicity, we allowed only one parameter in each of Eqs. 4 and 5 to vary with ST. Specifically, we assumed that $\ln(C(v))$ and $\ln(M)$ are linearly related to ST:

$$\ln(\alpha_N) = b_0 + b_1 ST \quad (6)$$

and

$$\ln(M) = c_0 + c_1 ST. \quad (7)$$

Modeling sapling growth

Ultimately, we intend to combine our light model with inventory data to fit mechanistic (i.e., physiological) whole-tree growth models (e.g., Friend et al. 1997). In this paper, however, we restrict our analysis to a simple, phenomenological growth model in order to focus on

TABLE 2. Candidate light models.

Model	Description
CR	Four values of $\bar{p}$, one for each CC $\times$ region combination.
CRN	$\bar{p}$ depends on Nbr according to Eq. 5. There are four values of M and S_p in Eq. 5, one for each CC $\times$ region combination. To set the scale for Nbr (which is poorly constrained), S_p was arbitrarily fixed at 1 for overtopped saplings in ENA.
CRS	For overtopped saplings, $\bar{p} = M$, which depends on ST according to Eq. 7. There are two values of $\bar{p}$ for sun-exposed saplings, one for each region.
CRNS	Same as model CRN except that M for overtopped saplings depends on ST (Eq. 7), rather than on region.

Notes: The probability distribution of sapling light availability depends on the mean effective leaf area, $\bar{p}$. The letters C, R, N, and S, respectively, indicate inclusion of the following effects on $\bar{p}$: crown class of focal saplings (CC; overtopped or sun-exposed), region (eastern North America or western Oregon, USA), neighborhood crowding (Nbr), and shade tolerance (ST) of focal saplings. See Table 1 for explanations of variable abbreviations.

TABLE 3. Summary of methods used to estimate parameters θ_G relating growth (G) to light (L).

Method	Description
Observed L	uses observed L for each sapling, as in standard analyses
Unobserved L	
Uncertain L	accounts for uncertainty in L by integrating over f_L for each sapling (Eq. 12)
Uncertain- L validation	same as uncertain- L method, but excludes light data for <i>Liriodendron</i> and <i>Pseudotsuga</i> when estimating f_L
Predicted L	replaces each observed L with a point estimate, $\hat{L}$, which is treated as if it were an observed value
Predicted- L validation	same as predicted- L method, but excludes light data for <i>Liriodendron</i> and <i>Pseudotsuga</i> when estimating $\hat{L}$

Note: All of the methods require observed G for each sapling. See Table 1 for explanations of variable abbreviations.

the problem of unobserved L . Due to the limited dbh range in our sample (2.5–12.7 cm), we assumed that G (in centimeters of dbh per year) was size independent. Our function relating expected growth, G , to L is derived from the Michaelis-Menten (M-M) function (e.g., Pacala et al. 1994, Wright et al. 1998): $G = AL/(A/S_G + L)$, where A is the asymptote and S_G is the slope at $L = 0$. Because G at $L = 1$ (full sunlight) may be much less than A , A may be poorly constrained by the data. Therefore, we solved the M-M function for $G_{\max} = G$ when $L = 1$ and re-parameterized the function in terms of $G_{\max}$

$$\tilde{G} = \frac{\tilde{G}_{\max} S_G L}{\tilde{G}_{\max} + L(S_G - \tilde{G}_{\max})}. \quad (8)$$

We assume that G is normally distributed with mean G and variance proportional to the mean:

$$V_G = \alpha_G \tilde{G}. \quad (9)$$

Assuming that V_G is a power law of G leads to similar results. To account for unmeasured site effects (e.g., edaphic conditions), we introduced a random effect (Q) for each stand in Appendix A and modified Eq. 8 as

$$\tilde{G} = \left(\frac{\tilde{G}_{\max} S_G L}{\tilde{G}_{\max} + L S_G - \tilde{G}_{\max}} \right) \left(\frac{\exp(rQ)}{1 + \exp(rQ)} \right) \quad (10)$$

where r is a species-specific parameter that determines responsiveness to variation in Q . The values of Q were assumed to be independent and normally distributed (Appendix F: section V). Growth variance (Eq. 9) was calculated from G after applying Eq. 10. Including stand effects in our analysis shifts the growth-light curve up or down by a species- and stand-specific multiplier that ranges from zero to one. Thus, $\tilde{G}_{\max}$ is now mean growth rate in full sunlight ($L = 1$) on good sites ($Q \gg 0$, assuming $r > 0$). For typical stands ($Q = 0$), $\tilde{G} = G_{\max}/2$ when $L = 1$. For poor stands ($Q < 0$), $\tilde{G} = 0$ for all L .

For each species, we estimated $\theta_G = (G_{\max}, S_G, (a_G, r))$ using the "observed- L ," method and two "unobserved- L " methods (Table 3), which were implemented using each of the candidate light models (Table 2).

1. *Observed- L method*-- The probability density of observed G_i for sapling i , given observed L_i and

unknown θ_G and $\mathbf{Q}$ (the vector of stand effects) is

$$p(G_i | L_i, \theta_G, \mathbf{Q}) = \text{normal}(G_i | \tilde{G}_i, V_{G,i}) \quad (11)$$

where G and V_G are given by Eqs. 9 and 10.

2. *Uncertain- L method*-- Here, G and $\mathbf{x}$ are observed for one set of saplings ("inventory data"), and L and $\mathbf{x}$ for another set of saplings ("calibration data") that constrain the light model parameters θ_L . (In the test case presented here, the first set of saplings is a subset of the second set.) The covariates $\mathbf{x}$ include the dbh measurements of a sapling's neighbors (which combine with height allometry parameters and θ_L to determine Nbr) and CC. The joint density for observed G , and unobserved L , for sapling i is

$$p(G_i, L_i | \mathbf{x}_i, \theta_L, \theta_G, \mathbf{Q}) = p(G_i | L_i, \theta_G, \mathbf{Q}) \times f_L(L_i | \mathbf{x}_i, \theta_L)$$

where $p(G_i | L_i, \theta_G, \mathbf{Q})$ is Eq. 11, and f_L is defined by Eq. 1. The marginal density of G_i is then

$$p(G_i | \mathbf{x}_i, \theta_L, \theta_G, \mathbf{Q}) = \int_0^1 p(G_i | L_i, \theta_G, \mathbf{Q}) \times f_L(L_i | \mathbf{x}_i, \theta_L) dL_i. \quad (12)$$

3. *Predicted- L method*-- The uncertain- L method accounts for uncertainty in L by integrating over f_L . A simpler alternative is to use a point estimate for unobserved L . Thus, the predicted- L method proceeds exactly as in (1) (*Observed- L method*), but we replace observed L with $\hat{L}$, the expectation of L (Eq. 2) evaluated at the posterior medians of θ_L . Defining $\hat{L}$ as the posterior mean of the expectation of L (Appendix F: section VI) yielded very similar results.

Bayesian posterior estimation

Because there is no analytical solution to Eq. 12, we used Markov chain Monte Carlo (MCMC) methods (Gilks et al. 1996) to effectively integrate over f_L without having to repeatedly evaluate Eq. 12 (Appendix F: section III). We used the Metropolis-Hastings algorithm (executed with a program written in C) to simulate three parallel MCMC chains until the marginal posteriors for θ_L and θ_G converged according to standard criteria (Appendix F: section I). For all parameters, we assumed uniform priors on a finite range that included the biologically reasonable values (Supplement 2).

All species-specific parameters (i.e., the elements of $\theta_G, G_{\max}, S_G, x_G$, and r) were modeled hierarchically (Gelman et al. 2004, Clark 2005, Ogle and Barber 2008). We assumed a lognormal sampling distribution for each of the four populations of parameters. This assumption implies nonnegative values, as must be the case for $G_{\max}, S_G$, and x_G . In principle, r could be positive or negative, but visual examination of growth data indicated positive r for all species (i.e., a good site for a given species was a good site for all co-occurring species).

We now outline the posteriors for each analysis. We refer to parameters that specify the sampling distribution of other parameters as "hyperparameters," and we refer to their priors as "hyperpriors" (Clark 2005, Ogle and Barber 2008). We assumed uniform hyperpriors with range sufficiently wide so as not to affect the posteriors. For simplicity, the presentation below ignores uncertainty in individual heights and ST, but our analysis does account for these sources of uncertainty (Appendix E).

1. *Light model.*— The likelihood of the light observations is

$$\mathcal{L}_L = \prod_i^{N_L} f_L(L_i | \mathbf{x}_i, \theta_L) \quad (13)$$

where $N_L = 2128$ and other terms are defined as in Eq. 12. The joint posterior is

$$p(\theta_L | \mathbf{L}, \mathbf{X}) \propto \mathcal{L}_L \times p(\theta_L)$$

where $\mathbf{L}$ is the vector of N_L light observations, $\mathbf{X}$ is the covariates matrix for the N_L focal saplings, $\mathbf{L}$ is as in Eq. 13, and $p(\theta_L)$ is the joint prior.

2. *Growth: observed-L method.*— The likelihood of the growth observations is

$$\mathcal{L}_G = \prod_i^{N_G} p(G_i | L_i, \theta_G, \mathbf{Q}) \quad (14)$$

where $N_G = 579$ and $p(G_i | \cdot)$ is as in Eq. II. The joint posterior is

$$p(\theta_G, \Phi_G, \mathbf{Q} | \mathbf{G}, \mathbf{L}) \propto \mathcal{L}_G \times p(\theta_G | \Phi_G) \times p(\Phi_G) \times p(\mathbf{Q})$$

where $\hat{\theta}_G$ is the vector of hyperparameters for θ_G ; $\mathbf{G}$ is the vector of N_G growth observations; $\mathbf{L}$, here, is the vector of N_G light observations for the saplings with observed $\mathbf{G}$; $\mathcal{L}_G$ is the likelihood (Eq. 14); $p(\theta_G | \Phi_G)$ is the joint prior for θ_G given Φ_G ; $p(\Phi_G)$ is the joint hyperprior for Φ_G ; and $p(\mathbf{Q})$ is the joint prior for the stand effects, $\mathbf{Q}$.

3. *Growth: uncertain-L method.*— The likelihood of the growth observations is

$$\mathcal{L}_G = \prod_i^{N_G} p(G_i | \mathbf{x}_i, \theta_L, \theta_G, \mathbf{Q}) \quad (15)$$

where $p(G_i | \cdot)$ is as in Eq. 12 and other terms are as defined in subsections 1 (*Light model*) and 2 (*Growth:*

observed-L method). The joint posterior is

$$p(\theta_G, \Phi_G, \mathbf{Q}, \theta_L | \mathbf{G}, \mathbf{X}, \mathbf{L}) \propto \mathcal{L}_G \times p(\theta_G | \Phi_G) \times p(\Phi_G) \times p(\mathbf{Q}) \times \mathcal{L}_L \times p(\theta_L)$$

where $\mathbf{L}$, here, is the vector of N_L light observations used to parameterize the light model; $\mathcal{L}_G$ is as in Eq. 15; $\mathcal{L}_L$ is as in Eq. 13; and other terms are as previously defined. Estimating the posteriors for θ_L and θ_G simultaneously is a convenient way to propagate uncertainty in θ_L . A potential drawback of this approach is that the growth data could influence θ_L , but this effect was weak in our analysis (see *Results*).

4. *Growth: predicted-L method.*— This analysis is identical to that in subsection 2 (*Growth: observed-L method*), except observed $\mathbf{L}$ are replaced by predictions obtained from subsection 1 (*Light model*).

Explained variation

Residuals for each $\mathbf{G}$ and $\mathbf{L}$ observation i were calculated as $e_i = Y_i - \hat{Y}_i$, where y and $\hat{Y}$ represent observations and predicted means, respectively. Methods for calculating y are described in Appendix F: section VI. We calculated the proportion of explained variation in light and growth resulting from each candidate light model as

$$R^2 = \frac{SS_{\text{explained}}}{SS_{\text{total}}} = \frac{SS_{\text{total}} - SS_{\text{residual}}}{SS_{\text{total}}} = 1 - \frac{\sum_i \varepsilon_i^2}{\sum_i (y_i - \bar{y})^2} \quad (16)$$

where SS is sum of squares, and $\bar{y}$ is the sample mean of observed L or G . Because R^2 is based on predicted means, it may be misleadingly low for skewed distributions (e.g., Fig. 2B). Also, large residual variance does not necessarily indicate lack of fit in our light model, where the aim is to quantify the distribution of $\mathbf{L}$, rather than to precisely estimate L . Nevertheless, R^2 provides an easily interpretable index of model performance that we use for informal model comparisons.

Model validation

In future applications, our light model will be used to estimate growth-light parameters from inventory data in which G , but not L , is observed for each sapling. In this case, the saplings in the growth analysis (the inventory data) would be distinct from those in the light analysis (our calibration data), and we must rely on our light model (parameterized from a limited set of species and stand conditions) to estimate $\mathcal{L}$ for each sapling in the larger inventory data set. In contrast, in the test case presented in this paper, the saplings in the growth analysis are a subset of those in the light analysis. Therefore, we performed a validation exercise in which we excluded light data for one eastern (*Liriodendron*) and one western (*Pseudotsuga*) species.

TABLE 4. Explained variation (R^2 ; Eq. 16) in light (L) and growth data for candidate light models (Table 2).

Model	Light R^2			Mean growth R^2			
	All ($n = 2128$)	OT ($n = 1510$)	SE ($n = 618$)	UL ($n = 579$)	ULV ($n = 579$)	PL ($n = 579$)	PLV ($n = 579$)
CR	0.72	0.09	0.02	0.40	0.38	0.17	0.11
CRN	0.78	0.38	0.16	0.45	0.46	-0.05	-0.30
CRS	0.75	0.30	0.02	0.48	0.50	0.11	0.01
CRNS	0.79	0.39	0.16	0.53	0.53	0.02	-1.17

Notes: Light model R^2 values are the proportion of explained variation among all saplings, among overtopped (OT) saplings only, and among sun-exposed (SE) saplings only. Mean growth R^2 (explained variation in growth data, averaged across species) is given for the following methods (see Table 3): uncertain L (UL), uncertain- L validation (ULV), predicted L (PL), and predicted- L validation (PLV). Mean growth R^2 for the observed- L growth analysis (Table 3) is 0.63.

Thus, we estimated O_G for these species using O_L estimated from other species' light data, as would be necessary in analyses of inventory data. We chose *Liriodendron* and *Pseudotsuga* because their parameters in the observed- L growth analysis are well constrained, which makes comparisons with parameter estimates from the unobserved- L methods relatively straightforward. In addition to validating the growth analysis, this exercise provides a test of the robustness of the light model itself, because a substantial fraction of the light data is comprised by *Liriodendron* and *Pseudotsuga* (146/1183 and 180/945 saplings in ENA and WOR, respectively).

RESULTS

Posterior 95% credible intervals (Supplement 2) spanned a small part of the prior ranges for all parameters except for two (S_G for *Pinus strobus* and *Tsuga canadensis*), indicating that the analyses were well-constrained by the data.

Light-model parameters (θ_L)

Estimates of O_L were similar whether θ_L was fit in isolation or simultaneously with O_G (Supplement 2). Thus, the growth data exerted little leverage on O_L . Estimates of O_L presented in this section were obtained in isolation from the growth data.

All four candidate models (Table 2) explained a large proportion of the variation in L ($R^2 = 0.72$ - 0.79 ; Table 4). The high R^2 values were due primarily to CC (i.e., the large difference in L between overtopped and sun-exposed saplings; Fig. 1). The R^2 values were lower when calculated within CC ($R^2 = 0.09$ - 0.39 and 0.02 - 0.16 for overtopped and sun-exposed saplings, respectively; Table 4).

The amount of shade cast by neighbors and the amount of shade received by overtopped saplings both increased with ST (Figs. 2 and 3). These effects were significant (i.e., 95% credible intervals for b_p and c_1 excluded zero) in models eRN and CRS (which only included one of the two effects) and in model CRNS (which included both effects). Together, ST and Nbr accounted for much of the interspecific variation in L (Figs. 1 and 2).

Growth parameters (θ_G)

Among the four candidate light models (Table 2) and two unobserved- L methods (Table 3), model CRNS with the uncertain- L method was most similar to the observed- L method in terms of estimated SG (Fig. 4), predicted growth curves (Fig. 5), and growth R^2 values (Table 4). Uncertainty in BG tended to be higher in the unobserved- L methods compared to the observed- L method, and some parameter estimates differed considerably between the unobserved- L and observed- L methods (Fig. 4 and Supplement 2).

The predicted- L method yielded negative R^2 values for some species; i.e., the predicted curves did not fit the data as well as the mean. In most cases, these negative R^2 values were associated with concave-up growth curves that overestimated high-light growth (e.g., Fig. 5; *Pinus ponderosa*, *Pseudotsuga*, and *Liquidambar*).

Excluding light data for *Liriodendron* and *Pseudotsuga* had little impact on the uncertain- L analysis (compare uncertain- L vs. uncertain- L validation in Fig. 5, Table 4, and Supplement 2), but had a noticeable impact on the predicted- L analysis. The tendency for the predicted- L method to overpredict high-light growth (resulting in negative R^2) was exaggerated in the predicted- L validation analysis (e.g., *Pseudotsuga* in Fig. 5).

A latent-variable integration approach to the uncertain- L method (Appendix F: section III) yielded very similar results (Fig. 5 and Supplement 2) to a more computationally demanding, direct integration approach (Appendix F: section IV), in which the integral in Eq. 12 is estimated numerically for each growth observation at each MCMC step.

DISCUSSION

Geographically extensive forest inventories typically lack the individual-level measurements of resource availabilities needed to parameterize mechanistic (i.e., physiological) forest models. To overcome this limitation, we developed an approach to estimate parameters relating individual performance to resource availability (or other environmental factors) by combining field-calibrated models of resource availability with inventory data sets on individual performance.

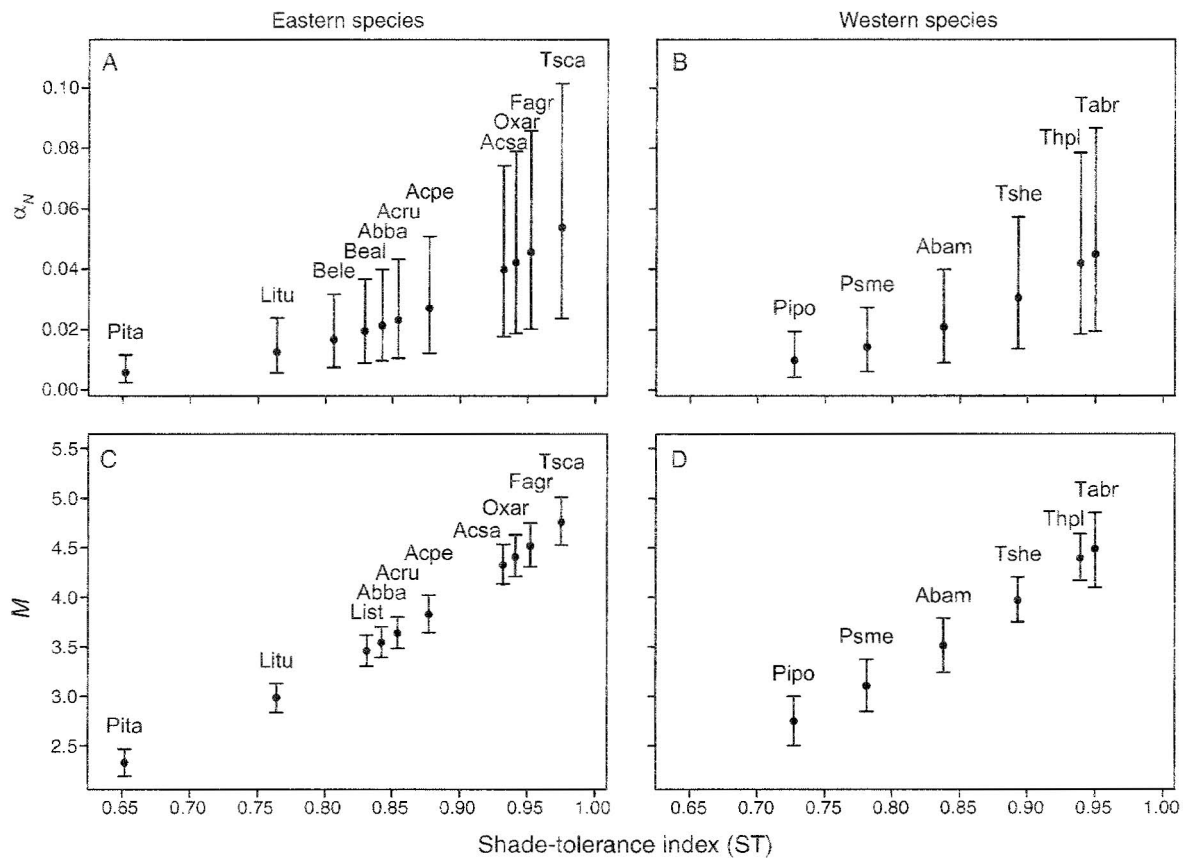


FIG. 3. Posterior medians and 95% credible intervals from model CRNS (see Table 2) for α_N and M vs. the shade tolerance index, ST (see Table 1 for definitions of terms). The amounts of shade cast by neighbors and shade received by focal saplings increase with α_N and M , respectively. The relationship between α_N and ST is determined by b_0 and b_1 (Eq. 6), and the relationship between M for overtopped saplings and ST is determined by c_0 and c_1 (Eq. 7); i.e., α_N and M are not free parameters for each species. (A, B) Values of α_N for species with $n \geq 30$ neighbor trees or saplings. (C, D) M for species with $n \geq 30$ overtopped focal saplings. Abbreviations are the first two letters of the genus and species names (see Appendix B for full names). Uncertainty in ST was propagated through the analysis, but for simplicity, ST is plotted at its posterior median value.

The test case presented in this paper focuses on estimating parameters θ_G relating sapling growth (G) to light availability (L) from data sets in which G , but not L , is observed for each sapling. Estimates of θ_G from the uncertain-L method (which accounts for uncertainty in L) were similar to those from the observed-L method (see Table 3 for a summary of methods). Furthermore, the uncertain-L method was robust to excluding light data for two common species. This validation exercise required that we estimate θ_G for species not represented in the calibration data, as would be necessary when applying our methods to inventory data. Compared to the uncertain-L method, the predicted-L method (which ignores uncertainty in L) yielded estimates of θ_G that were less robust to excluding data and less similar to the observed-L method.

The predicted-L method sometimes produced unrealistic, concave-up growth curves that overpredicted high-light growth (e.g., Fig. 5, *Pseudotsugai*). This is likely an extrapolation problem. The point estimates used in the predicted-L method are expected values and, therefore, are never as high as the highest L observations. Thus,

when estimating θ_G , the predicted-L method never "sees" the highest L values for which growth is overpredicted. In contrast, the uncertain-L method integrates over all possible values of L and is therefore less prone to this type of error.

The uncertain-L method performed well overall, but there were some differences when compared to the observed-L analysis. For example, growth curves for several species were more concave-up (or less concave-down) in the uncertain-L than in the observed-L analysis (e.g., Fig. 5, *Pseudotsuga*). Nevertheless, the uncertain-L curves always appeared reasonable compared to the data, even in cases in which R^2 was low (e.g., Fig. 5, *Liquidambar*). The ultimate test of the uncertain-L method will be whether or not it leads to improved broadscale models of forest dynamics. At this stage, we can only conclude that the method appears promising. The computational demands of the method are substantially reduced by a latent-variable approach to numerical integration that yields very similar results to a direct integration approach (Appendix F: sections III-IV).

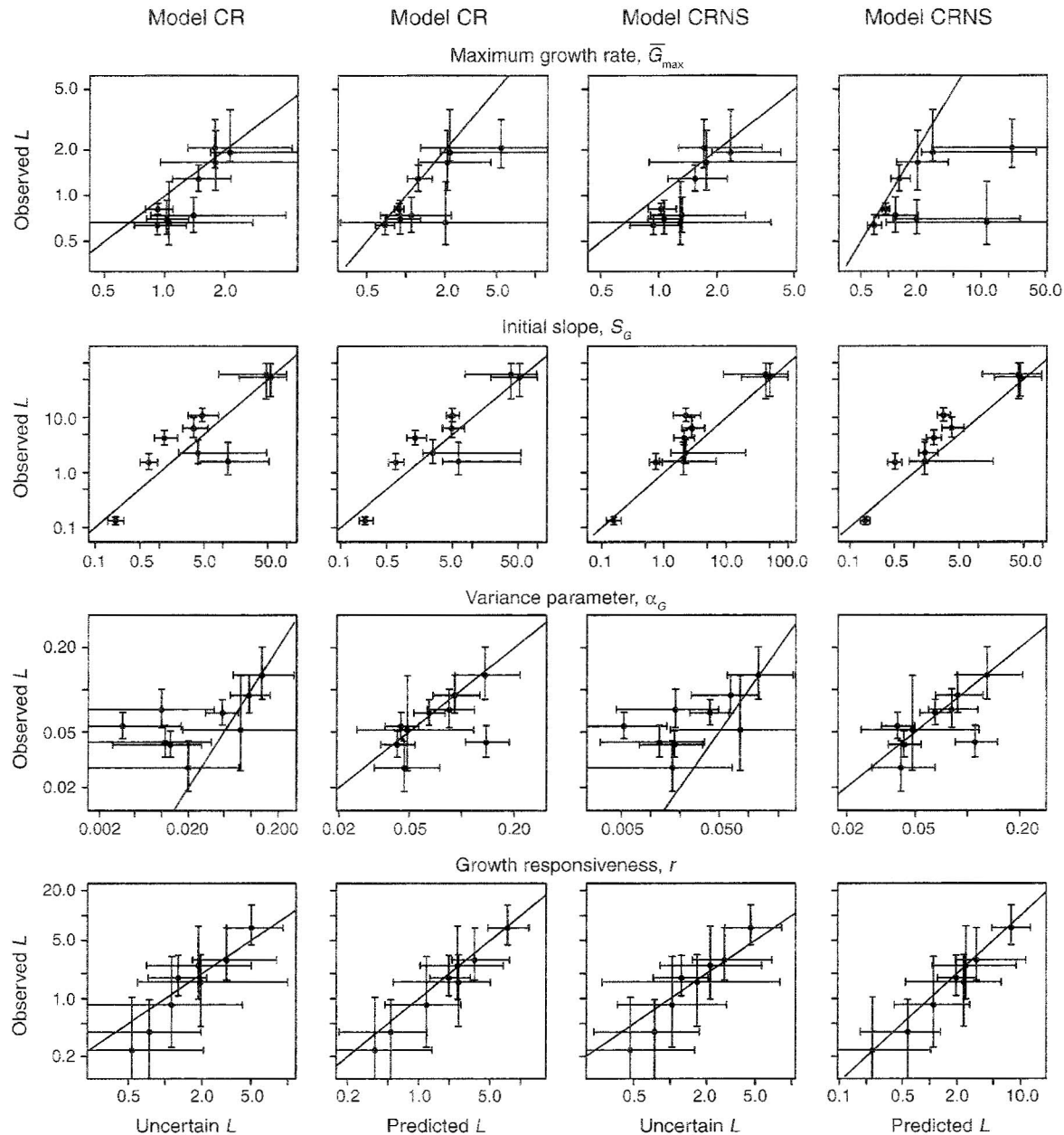
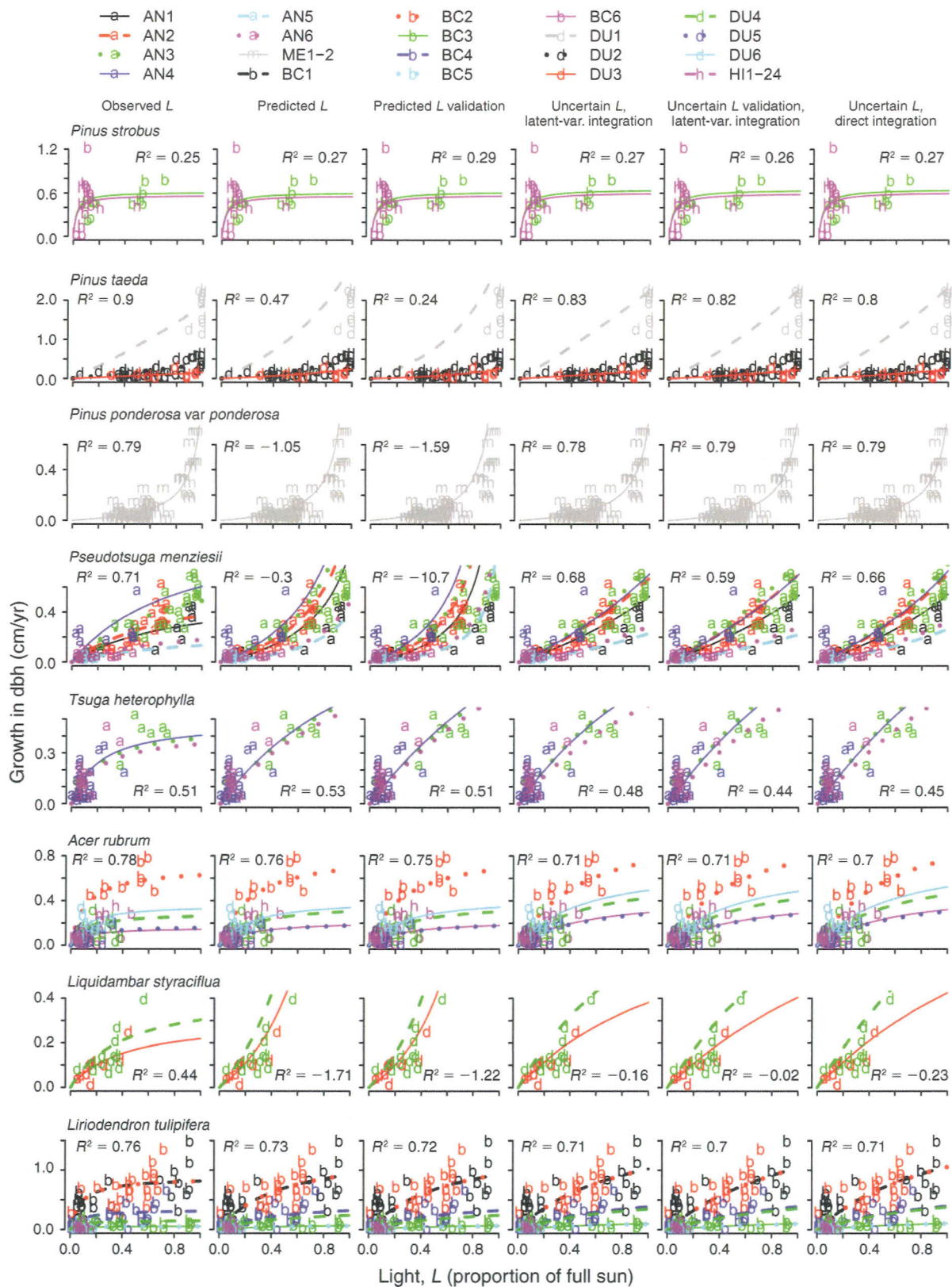


FIG. 4. Posterior medians and 95% credible intervals for the species-specific growth parameters (θ_G) from models CR and CRNS (Table 2) using observed- L and unobserved- L methods (Table 3). In each graph, the nine points are the posterior medians for each of nine species from the observed- L method (y -axis) and an unobserved- L method (uncertain- L or predicted- L ; x -axes). Abbreviations are as follows (see Eqs. 8–10): $\bar{G}_{\max}$, maximum expected growth rate; S_G , initial slope of growth vs. light curve; α_G , proportionality constant between growth mean and variance; and r , responsiveness of growth to stand effects. The uncertain- L results use latent-variable integration (Appendix F: section III). The plotted values, along with their species identities, are in Supplement 2.

FIG. 5. Scatterplots of observed growth (diameter at breast height, dbh) vs. light (L) and estimated curves from the observed- L method (column 1) and the unobserved- L methods (columns 2–6; see Table 3) implemented with model CRNS (see Table 2). Observed L values are shown in each graph for reference but are used directly only in the observed- L method. For each species, estimated curves are shown for stands with $n \geq 5$ growth observations. Stand abbreviations are: AN, H. J. Andrews Experimental Forest; ME, Metolius Research Natural Area; BC, Bent Creek Experimental Forest; DU, Duke Forest; HI, Highlands Biological Station. See Appendix A for a description of the stands. Curves and R^2 reflect the posterior medians of the parameters. Observed L



for *Pseudotsuga* and *Liriodendron* were excluded from the validation analyses (columns 3 and 5). The uncertain-L analysis was performed with both latent-variable integration (columns 4–5) and direct integration (column 6) (see Appendix F: sections III and IV). Results for *Tsuga canadensis* (with only 10 growth observations) are not shown.

Factors affecting sapling light availability

We considered four covariates, available from most inventory data sets, in modeling L : crown class (CC, the classification of focal saplings as overtopped or sun-exposed), neighborhood crowding (Nbr), geographic region, and species identity (via the shade tolerance index, ST). The candidate models (all of which accounted for CC and region) explained 72–79% of the variation in L , with much of the explanatory power due to CC. Nevertheless, Nbr, region, and ST also had important effects.

Including Nbr increased R^2 from 0.09 to 0.38 for overtopped saplings and from 0.02 to 0.16 for sun-exposed saplings (Table 4; compare models CR and CRN). The Nbr effects within species were not dramatic (Fig. 1) because the effects are conditional on CC. These conditional effects should be weaker than those in studies in which neighborhood effects were quantified without controlling for canopy status of focal individuals (e.g., Uriarte et al. 2004).

The amount of shade cast by neighbors and the amount of shade received by saplings both increased with ST (Fig. 3). These results are consistent with observations of increased shading by canopy trees (Horn 1971, Canham et al. 1994, 1999) and decreased light received by understory saplings (Clark and Clark 1992, Davies 2001, Poorter and Arets 2003) as shade tolerance increases across species. Importantly, ST can be estimated directly from inventory data without the need for additional calibration data. Other widely available indices may also prove useful in developing a broadly applicable light model and should be explored in future analyses.

The region-specific parameters in our analysis differed significantly between the eastern and western sites. These parameter differences do not translate into simple regional differences in light environments, because M , S_p , and typical Nbr values (Eq. 5) are all region-dependent. More sophisticated approaches to neighborhood analysis (e.g., Canham et al. 1994, Beaudet et al. 2002) may account for regional differences in light environments, which would obviate the need for region-specific parameters, but would require species-specific parameters (e.g., canopy transparencies) that are not widely available. Another potential strategy for developing a general light model would be to account for the dependence of leaf area (and thus canopy light-interception) on site productivity (Grier and Running 1977, Gholz 1982). Until a general, yet simple, approach to modeling understory light is available, it may be necessary to calibrate models separately for different forest types.

Toward a new generation of global vegetation models

A number of dynamic global vegetation models have been developed that include physiological mechanisms (e.g., dependence of photosynthesis on light, water, temperature, and CO_2) expected to play an important

role in determining vegetation response to climate change and elevated atmospheric CO_2 levels (Cramer et al. 2001, Friedlingstein et al. 2006). However, to our knowledge, only two such models (Friend et al. 1997, Moorcroft et al. 2001) account for the individual-level height-structured competition that is a key driver of forest dynamics, and none of the models have assimilated the vast inventory databases on individual, whole-plant performance. Existing methods that scale from individuals to ecosystems (Moorcroft et al. 2001, Strigul et al. 2008) make global implementation of individual-based models computationally practical. However, parameterizing a mechanistic, individual-based forest model using rigorous, quantitative methods remains a formidable challenge. The methods presented in this paper are an important step toward meeting this challenge.

In simple terms, our approach is to use field-calibrated statistical models to translate covariates that are available for individuals in inventories (e.g., crown class) into resources that are not (e.g., light). The translation from covariates to resources is a necessary step to using inventories to inform mechanistic models of individual performance, which would be components of a mechanistic forest model (e.g., Friend et al. 1997, Moorcroft et al. 2001). For example, a mechanistic growth model might entail: (1) calculating environmental conditions, including ambient temperature, soil water potential, and light incident on an individual's crown; (2) calculating whole-plant photosynthesis by integrating a leaf photosynthesis model (Farquhar et al. 1980, Ball et al. 1987, Collatz et al. 1991, Lenning 1995) through the crown; and (3) allocating assimilated carbon to stem, roots, leaves, etc. Estimating the parameters that determine carbon assimilation (step 2) and allocation (step 3) would require embedding the mechanistic growth model within a statistical framework, e.g., our phenomenological growth model (Eq. 10) would be replaced by the above mechanistic model. Assuming inventory data are used to constrain the analysis (i.e., by comparing the predicted growth rates to those observed in the inventory), uncertainty in light and other environmental inputs could be dealt with using the methods presented here.

Our methods are designed to interpret inventory data when estimating growth parameters, not to predict light availability when using the growth parameters in a dynamic model. Therefore, the utility of our light model does not depend on whether or not its parameter values (i.e., the specific relationships between light and covariates, such as crown class) hold in the future. In contrast, the utility of the growth parameters does depend on their generality under changing climate and CO_2 conditions. This generality can only be achieved by accounting for physiological mechanisms (Reynolds et al. 2001), which requires explicit treatments of light and other resources. Non-mechanistic treatments of growth (e.g., modeling growth as a function of crown class)

may lead to erroneous predictions in a changing environment (e.g., because understory light availability may change). A global vegetation model that includes the right mechanisms should accurately predict how individual- and ecosystem-level properties, including canopy leaf area and understory light, respond to environmental change. Developing such a model will require novel statistical approaches to assimilating a variety of data sources, including forest inventories, eddy covariance towers (Baldocchi et al. 2001, Friend et al. 2007), and plant trait databases (Wright et al. 2004). Our methods open the door to using individual-level inventory data to inform mechanistic models of individual tree performance, which could be important components of a new generation of global vegetation models.

ACKNOWLEDGMENTS

We thank two anonymous reviewers for helpful comments. We thank April Melvin, John Kim, Nathan Senn, Autumn Lear, Iris Koski, Genevieve Becker, and Tanguy Daufresne for outstanding assistance in the field. We are grateful to many individuals and institutions that graciously provided access to field sites, housing, logistical support, and knowledge of local natural history: Karie O'Connell and Howard Bruner at H. J. Andrews Experimental Forest; Sarah Greene at Metolius Research Natural Area (permanent plot data provided courtesy of the Oregon State University Forest Science Databank); Richard Broadwell at Duke Forest; David Loftis at Bent Creek Experimental Forest; Peter Schleifenbaum at Haliburton Forest and Wildlife Preserve; and Highlands Biological Station. Funding was provided by the National Science Foundation (K. Ogle supported under grant 0630119 awarded in 2006) and the Carbon Mitigation Initiative (CMI) of the Princeton Environmental Institute. CMI is sponsored by BP and Ford.

LITERATURE CITED

- Adams, T. P., D. W. Purves, and S. W. Pacala. 2007. Understanding height-structured competition in forests: Is there an R^* for light? *Proceedings of the Royal Society B* 274: 3039-3047.
- Albani, M., D. Medvigy, G. C. Hurtt, and P. R. Moorcroft. 2006. The contributions of land-use change, CO_2 fertilization, and climate variability to the Eastern US carbon sink. *Global Change Biology* 12:2370-2390.
- Baldocchi, D., et al. 2001. FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bulletin of the American Meteorological Society* 82:2415-2434.
- Ball, J. T., I. E. Woodrow, and J. A. Berry. 1987. A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions. Pages 221-224 in J. Biggens, editor. *Progress in photosynthesis research*. Martinus Nijhoff, Dordrecht, The Netherlands.
- Bazzaz, F. A. 1979. The physiological ecology of plant succession. *Annual Review of Ecology and Systematics* 10: 351-371.
- Beaudet, M., C. Messier, and C. D. Canham. 2002. Predictions of understory light conditions in northern hardwood forests following parameterization, sensitivity analysis, and tests of the SORTIE light model. *Forest Ecology and Management* 165:235-248.
- Bechtold, W. A., and C. T. Scott. 2005. The Forest Inventory and Analysis plot design. Pages 27-42 in W. A. Bechtold and P. L. Patterson, editors. *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, CSA.
- Beers, T. W., and C. I. Miller. 1964. Point sampling: research results, theory, and applications. Research Bulletin 786. Purdue University Agricultural Experiment Station. West Lafayette, Indiana, CSA.
- Breda, N. J. J. 2003. Ground-based measurements of leaf area index: a review of methods, instruments and current controversies. *Journal of Experimental Botany* 54:2403-2417.
- Burns, R. M., and B. H. Honkala. 1990. *Silvics of North America: 1. Conifers; 2. Hardwoods*. Volume 2. Agriculture Handbook 654. U.S. Department of Agriculture, Forest Service, Washington, D.C., USA.
- Canham, C. D. 1988. An index for understory light levels in and around canopy gaps. *Ecology* 69: 1634-1638.
- Canham, C. D., K. D. Coates, P. Bartemucci, and S. Quaglia. 1999. Measurement and modeling of spatially explicit variation in light transmission through interior cedar-hemlock forests of British Columbia. *Canadian Journal of Forest Research* 29:1775-1783.
- Canham, C. D., J. S. Denslow, W. J. Platt, J. R. Runkle, T. A. Spies, and P. S. White. 1990. Light regimes beneath closed canopies and tree-fall gaps in temperate and tropical forests. *Canadian Journal of Forest Research* 20:620-631.
- Canham, C. D., A. C. Finzi, S. W. Pacala, and D. H. Burbank. 1994. Causes and consequences of resource heterogeneity in forests: interspecific variation in light transmission by canopy trees. *Canadian Journal of Forest Research* 24:337-349.
- Clark, D. A., and D. B. Clark. 1992. Life history diversity of canopy and emergent trees in a Neotropical rain forest. *Ecological Monographs* 62:315-344.
- Clark, J. S. 2005. Why environmental scientists are becoming Bayesians. *Ecology Letters* 8:2-14.
- Collatz, G. J., J. T. Bali, 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.
- Cramer, W., et al. 2001. Global response of terrestrial ecosystem structure and function to CO_2 and climate change: results from six dynamic global vegetation models. *Global Change Biology* 7:357-373.
- Davies, S. J. 2001. Tree mortality and growth in 11 sympatric *Macaranga* species in Borneo. *Ecology* 82:920-932.
- 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.
- Doman, A. P., R. Ennis, and D. Weigel. 1981. Resource evaluation field instructions for Wisconsin 1981. USDA Forest Service, North Central Forest Experiment Station, St. Paul, Minnesota, USA.
- 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.
- 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.
- Friend, A. D., A. Arneth, N. Y. Kiang, M. Lomas, J. Ogee, C. Rodenbeck, S. W. Running, J. D. Santaren, S. Sitch, N.

- Viovy, F. I. Woodward, and S. Zaehle. 2007. FLUXNET and modelling the global carbon cycle. *Global Change Biology* 13:610-633.
- Friend, A. D., A. K. Stevens, R. G. Knox, and M. G. R. Cannell. 1997. A process-based, terrestrial biosphere model of ecosystem dynamics (Hybrid v3.0). *Ecological Modelling* 95:249-287.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. Bayesian data analysis. Second edition. Chapman and Hall/CRC, Boca Raton, Florida, USA.
- Gholz, H. L. 1982. Environmental limits on aboveground net primary production, leaf area, and biomass in vegetation zones of the Pacific Northwest. *Ecology* 63:469-481.
- Gilks, W. R., S. Richardson, and D. J. Spiegelhalter. 1996. Introducing Markov chain Monte Carlo. Pages 1-19 in W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC, Boca Raton, Florida, USA.
- Givnish, T. J. 1988. Adaptation to sun and shade: a whole-plant perspective. *Australian Journal of Plant Physiology* 15:63-92.
- Grier, C. C., and S. W. Running. 1977. Leaf area of mature northwestern coniferous forests: relation to site water balance. *Ecology* 58:893-899.
- Horn, H. S. 1971. The adaptive geometry of trees. Princeton University Press, Princeton, New Jersey, USA.
- Hurtt, G. C., S. W. Pacala, P. R. Moorcroft, J. Caspersen, E. Shevliakova, R. A. Houghton, and B. Moore. 2002. Projecting the future of the U.S. carbon sink. *Proceedings of the National Academy of Sciences (USA)* 99:1389-1394.
- Kira, T., K. Shinokawa, and K. Hozumi. 1969. Structure of forest canopies as related to their primary productivity. *Plant and Cell Physiology* 10:129-142.
- Kitajima, K. 1994. Relative importance of photosynthetic traits and allocation patterns as correlates of seedling shade tolerance of 13 tropical trees. *Oecologia* 98:419-428.
- Kobe, R. K., S. W. Pacala, J. A. Silander, and C. D. Canham. 1995. Juvenile tree survivorship as a component of shade tolerance. *Ecological Applications* 5:517-532.
- 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.
- Leuning, R. 1995. A critical appraisal of a combined stomatal-photosynthesis model for C_3 plants. *Plant, Cell and Environment* 18:339-355.
- Lichstein, J. W. 2007. Theoretical and empirical perspectives on the dynamics of U.S. forests. Dissertation. Princeton University, Princeton, New Jersey, USA.
- Monsi, M., and T. Saeki. 2005. On the factor light in plant communities and its importance for matter production. *Annals of Botany* 2005:2549-2567.
- Moorcroft, P. R. 2006. How close are we to a predictive science of the biosphere? *Trends in Ecology and Evolution* 21:400-407.
- Moorcroft, P. R., G. C. Hurtt, and S. W. Pacala. 2001. A method for scaling vegetation dynamics: the ecosystem demography model (ED). *Ecological Monographs* 71:557-585.
- Ogle, K., and J. J. Barber. 2008. Bayesian data-model integration in plant physiological and ecosystem ecology. *Progress in Botany* 69:28-311.
- Pacala, S. W., C. D. Canham, J. A. Silander, and R. K. Kobe. 1994. Sapling growth as a function of resources in a north temperate forest. *Canadian Journal of Forest Research* 24:2172-2183.
- Poorter, L., and E. J. M. M. Arets. 2003. Light environment and tree strategies in a Bolivian tropical moist forest: an evaluation of the light partitioning hypothesis. *Plant Ecology* 166:295-306.
- Purves, D. W., J. W. Lichstein, N. Strigul, and S. W. Pacala. 2008. Predicting and understanding forest dynamics using a simple tractable model. *Proceedings of the National Academy of Sciences (USA)* 105:17018-17022.
- Purves, D. W., and S. W. Pacala. 2008. Predictive models of forest dynamics. *Science* 320:1452-1453.
- Reynolds, J. F., H. Bugmann, and L. F. Pitelka. 2001. How much physiology is needed in forest gap models for simulating long-term vegetation response to global change? Challenges, limitations, and potentials. *Climatic Change* 51:541-557.
- Sitch, S., V. Brovkin, W. von Bloh, D. van Vuuren, B. Eickhout, and A. Ganopolski. 2005. Impacts of future land cover changes on atmospheric CO_2 and climate. *Global Biogeochemical Cycles* 19:GB2013.
- Sitch, S., B. Smith, I. C. Prentice, A. Arneth, A. Bondeau, W. Cramer, J. O. Kaplan, S. Levis, W. Lucht, M. T. Sykes, K. Thonicke, and S. Venevsky. 2003. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Global Change Biology* 9:161-185.
- Strigul, N., D. Pristinski, D. Purves, J. Dushoff, and S. Pacala. 2008. Scaling from trees to forests: tractable macroscopic equations for forest dynamics. *Ecological Monographs* 78:523-545.
- Uriarte, M., C. D. Canham, J. Thompson, and J. K. Zimmerman. 2004. A neighborhood analysis of tree growth and survival in a hurricane-driven tropical forest. *Ecological Monographs* 74:591-614.
- USDA. 1992. Field manual, Michigan 1991. Second printing, April 1992. USDA Forest Service, North Central Forest Experiment Station, St. Paul, Minnesota, USA.
- USDA. 2005. Forest Inventory and Analysis national core field guide. Volume I. Field data collection procedures for phase 2 plots. Version 3.0. Forest Inventory and Analysis Program, USDA, Forest Service, Arlington, Virginia, USA.
- USDA. 2006. Users guide to the Forest Inventory Snapshot Database. Version 2.1. Forest Inventory and Analysis Program, USDA, Forest Service, Arlington, Virginia, USA.
- White, M. A., P. E. Thornton, S. W. Running, and R. R. Nemani. 2000. Parameterization and sensitivity analysis of the BIOME-BGC terrestrial ecosystem model: net primary production controls. *Earth Interactions* 4:1-85.
- Wright, E. F., K. D. Coates, C. D. Canham, and P. Bartemucci. 1998. Species variability in growth response to light across climatic regions in northwestern British Columbia. *Canadian Journal of Forest Research* 28:871-886.
- Wright, I. J., et al. 2004. The worldwide leaf economics spectrum. *Nature* 428:821-827.
- Wright, S. J., H. C. Muller-Landau, R. Condit, and S. P. Hubbell. 2003. Gap-dependent recruitment, realized vital rates, and size distributions of tropical trees. *Ecology* 84:3174-3185.

APPENDIX A

Description of study sites (*Ecological Archives* A020-020-A1).

APPENDIX B

Sample sizes and shade tolerance indices for species in light and growth analyses (*Ecological Archives* A020-020-A2).

APPENDIX C

Forest Inventory and Analysis (FTA) subplot designs and implementation (*Ecological Archives* A020-020-A3).

APPENDIX D

Growth data: sampling protocol and growth measurements (*Ecological Archives* A020-020-A4).

APPENDIX E

Estimating height-diameter allometries and the shade tolerance index for all U.S. tree species (*Ecological Archives* A020-020-A5).

APPENDIX F

Statistical methods (*Ecological Archives* A020-020-A6).

SUPPLEMENT 1

Posterior means and percentiles for height allometry and shade tolerance parameters for U.S. tree species (*Ecological Archives* A020-020-S 1).

SUPPLEMENT 2

Priors and posterior means and percentiles for light model and growth parameters (*Ecological Archives* A020-020-S2).