

Linking models and data on vegetation structure

G. C. Hurtt,^{1,2} J. Fisk,¹ R. Q. Thomas,³ R. Dubayah,⁴ P. R. Moorcroft,⁵ and H. H. Shugart⁶

Received 15 January 2009; revised 30 September 2009; accepted 5 November 2009; published 8 June 2010.

[1] For more than a century, scientists have recognized the importance of vegetation structure in understanding forest dynamics. Now future satellite missions such as Deformation, Ecosystem Structure, and Dynamics of Ice (DESDynI) hold the potential to provide unprecedented global data on vegetation structure needed to reduce uncertainties in terrestrial carbon dynamics. Here, we briefly review the uses of data on vegetation structure in ecosystem models, develop and analyze theoretical models to quantify model-data requirements, and describe recent progress using a mechanistic modeling approach utilizing a formal scaling method and data on vegetation structure to improve model predictions. Generally, both limited sampling and coarse resolution averaging lead to model initialization error, which in turn is propagated in subsequent model prediction uncertainty and error. In cases with representative sampling, sufficient resolution, and linear dynamics, errors in initialization tend to compensate at larger spatial scales. However, with inadequate sampling, overly coarse resolution data or models, and nonlinear dynamics, errors in initialization lead to prediction error. A robust model-data framework will require both models and data on vegetation structure sufficient to resolve important environmental gradients and tree-level heterogeneity in forest structure globally.

Citation: Hurtt, G. C., J. Fisk, R. Q. Thomas, R. Dubayah, P. R. Moorcroft, and H. H. Shugart (2010), Linking models and data on vegetation structure, *J. Geophys. Res.*, *115*, G00E10, doi:10.1029/2009JG000937.

1. Introduction

[2] From the original development of “Nachhaltigkeit,” or in English “sustainability,” with the Germans in the mid-eighteenth century, forest structure has been recognized as an essential component of understanding forest dynamics. Scientific forestry has as its basis the application of “yield tables” compiled from empirical measurement of, in some cases, centuries of observations on the complex relationships among tree volumes, tree numbers, tree sizes, tree growth rates, and stand basal area for forest stands arranged by their canopy height at a particular age [Shugart, 2008]. The historically early recognition of structure as a significant control on the dynamic responses of stand volume or biomass has continued in a number of the quantitative approaches used in forest models [also see Porté and Bartelink, 2002]. Early models of individual based forest dynamics [Newnham, 1964; Mitchell, 1969, 1975; Hegyi, 1974] soon found application for mixed

species forests [Ek and Monserud, 1974; Ranney et al., 1981] and evolved into so-called “gap” models with many world-wide applications [Shugart, 1984, 1998]. Today, gap models are increasingly emphasizing the spatial interactions in forests [Busing and Maily, 2004] and are moving back in the direction of their roots in the earlier spatially explicit forestry models. In Japan, the mechanisms underlying the functioning of yield tables were explored using partial differential equation models of the dynamics of the number and sizes of forest trees and stand thinning [Yoda et al., 1963; Shinozaki et al., 1964]. These models were initially developed for even-aged, monospecies plantations [Suzuki and Umemura, 1967, 1974] but were soon pushed into applications in species-rich, mixed-aged rain forest [Kohyama, 1993]. An approach much like that of Kohyama was developed by Russian ecologists [Korzukhin and Antonovski, 1992] using size-structured integro-differential equations. Thus, detailed models including vegetation structure have existed for decades. But due to the high level of detail and resolution required to capture relevant heterogeneity, these models have largely been confined to relatively small spatial/temporal scales (e.g., forest gap models operating at scales of 1 km or less). As a consequence, nearly all large-scale global carbon models have by necessity been highly aggregated, omitting much of this detail, and been silent on emerging key questions such as the role of disturbance/recovery in the global carbon balance [Hurtt et al., 1998].

[3] From the ground, structural properties of vegetation such as DBH, height, etc. are routinely measured in localized field plots using field methods. For some regions these measurements have been coordinated into regional or

¹Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, New Hampshire, USA.

²Department of Natural Resources and the Environment, University of New Hampshire, Durham, New Hampshire, USA.

³Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York, USA.

⁴Department of Geography, University of Maryland, College Park, Maryland, USA.

⁵Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA.

⁶Department of Environmental Sciences, University of Virginia, Charlottesville, Virginia, USA.

national inventories of large numbers of sample plots (e.g., USFIA), but to date and for the foreseeable future there is no consistent global coverage available based on ground measurements. In addition, estimates for large areas with limited ground-based sampling are potentially uncertain or biased [Fisher *et al.*, 2008]. Optical remote sensing has literally revolutionized the characterization of key properties of the land surface resulting in the mapping of forest/nonforest areas and important vegetation indices at high spatial and temporal resolution. However, vegetation structure is difficult or impossible to get at with optical remote sensing alone. In many respects we know more now about the extent of our forests, than their content.

[4] In 2007, the preface to the NRC Decadal Survey stated the importance of a foundation of integrated observations on which to build forecast models and other tools for making informed decisions [Anthes *et al.*, 2007]. It also gives high priority to a new satellite mission, Deformation, Ecosystem Structure, and Dynamics of Ice (DESDynI), intended in part to provide unprecedented global data on vegetation structure that can be used to improve estimates of terrestrial carbon stocks, fluxes, and the mechanisms controlling them. Purves and Pacala [2008] recently suggested that the explosion of ground-based inventory data on individual trees may preface the beginning of a new generation of models. Here we provide a brief review of recent uses of data on vegetation structure in forest models, present new theoretical studies designed to investigate and quantify general model-data requirements for vegetation structure, and describe recent results using data on vegetation structure in a mechanistic model with an integrated scaling strategy to improve predictions. Taken together, this work provides a background, theoretical, and mechanistic basis for a robust framework linking data on vegetation structure to models to improve predictions of terrestrial carbon dynamics.

2. Recent Studies Linking Models and Remote-Sensing Data on Vegetation Structure

[5] Remote-sensing data can be used as inputs to ecosystems models, to test model predictions, and to update or adjust models [Plummer, 2000]. The role of remote sensing in providing inputs to models can be further divided into initialization and parameterization. Whether a particular quantity is a variable or a constant is model specific. For example, a model simulating instantaneous carbon exchange with the atmosphere may treat aboveground biomass as a constant, while biomass is a key dynamic variable in a forest succession model that predicts changes over decades to centuries. Here, we present examples of studies that use lidar and/or radar remote sensing of forest structure as inputs to, validation of, and for updating ecosystem models. Models can also be used to interpret remote sensing data [Plummer, 2000], although this is not addressed here. For more general reviews of remote sensing and model synthesis, see Turner *et al.* [2004], Nightingale *et al.* [2004], Plummer [2000], and Lucas and Curran [1999].

[6] Initialization establishes the starting point of model simulations, which corresponds to the starting point of modeled canopy height and/or biomass when lidar or radar data are used for initialization. Defining the starting point for simulations is critical because future projections of forest

dynamics are highly dependent on initial state. Small trees have different future dynamics than large ones. At the patch scale, during regrowth the biomass and canopy height tend to increase through time, while the net flux of carbon into the system likely slows. From natural disturbances, to land use history [Hurt *et al.*, 2006], the entire landscape is a mosaic of patches in some stage of recovery. While land use history reconstructions can provide some of the information that is needed [Hurt *et al.*, 2006], comprehensive spatially explicit data are not available globally at sufficient resolution. Remote sensing of vegetation structure can provide relevant data.

[7] Ranson *et al.* [2001] used airborne radar measurements to initialize the biomass in a forest gap model, Zelig [Urban, 1990], at a northern forest in Maine. They found that radar data on biomass and species composition along with soil data at the 30×30 m resolution allow initialization that generated biomass in the model that compared well to field data and generated expected successional trajectories.

[8] Data on vegetation structure have also been used in model testing. Le Toan *et al.* [2004] used radar measurements of forest structure to assess where the Sheffield Dynamic Global Vegetation Model [Woodward *et al.*, 1995] was underpredicting or overpredicting biomass across a landscape in Siberia. Their study focused on two scales: $0.5^\circ \times 0.5^\circ$ and a “local” scale that corresponded to individual stands. At both scales they initialized using the best known land use history at $0.5^\circ \times 0.5^\circ$ resolution [Goldewijk, 2001] used in the coarser scale study, and data on stand establishment at the local scale study. Radar data aided in identifying where the model predicted erroneous biomass values. At coarse resolution, erroneous predictions were either errors in predicting environmental influences on growth or areas where the land use or disturbance history was not included in the coarse resolution land history. At finer resolution, the land use history was well constrained and differences between the predictions and the radar data were due to inadequate representation of fine-scale variation in environmental conditions.

[9] Vegetation structure data have also been used to parameterize models. For example, Kotchenova *et al.* [2004] used lidar data on vertical height profiles from the SLICER sensor to parameterize a canopy photosynthesis model. While canopy height profiles are a dynamic variable in models of succession that simulate canopy change at multiyear timescales, the photosynthesis model used static canopy height profiles to simulate the GPP at the daily timescale where successional changes are less important. The authors compared the simulated GPP between a model parameterized with a uniform vertical canopy distribution and a model parameterized using the lidar measured vertical distribution. The latter corresponded to the observed distribution of sun and shade leaves and improved GPP simulations by over 50%. Patenaude *et al.* [2008] used three sources of remote sensing: radar data on biomass, lidar data on tree height, and hyperspectral data on LAI, to aid in parameterizing nondynamic variables in the 3-PG (Physiological Processes Predicting Growth [Landsberg and Waring, 1997; Sands and Landsberg, 2002]) model at a forestry plantation in England. They used a Bayesian approach to estimate the likelihood and distribution of parameters in the model given the tree height, biomass, and LAI data that were collected. While Patenaude *et al.* [2008] used the Bayesian framework for model calibration before

prognostic simulations, the study gives an example how remote sensing data on forest structure can be used to update parameters in ecosystem models after comparison between predicted values of forest structure and remotely sensed measurements.

3. Theoretical Studies to Quantify Resolution Requirements

[10] For accurate predictions, models generally require estimates of the land surface to be used for initial conditions. Errors in estimation and/or limited sampling of the state of the land surface will create errors in model initialization, which in turn can propagate as errors in model predictions.

[11] We developed a simple theoretical model to investigate generalized data requirements of vegetation structure for forest models. The modeling framework is a simple descendent of the gap model paradigm and consists of extensions to the model of *Fisher et al.* [2008], developed originally to assess issues of field-plot sampling requirements for large-scale estimates of forest carbon stocks and fluxes. Following *Fisher et al.* [2008], we imagine a forested landscape on a large grid on a horizontal plane where each grid cell is the approximate size of an adult canopy tree (e.g., 10×10 m). Each cell accumulates biomass b at constant rate g , and dies (is disturbed) with probability μ . Disturbance events are distributed across the landscape following a power law size-frequency relationship

$$n_z = Az^{-\alpha} \quad (1)$$

where n is the number of gaps of size z , and A is a constant [*Pascual and Guichard*, 2005]. The scaling exponent α describes the clustering of disturbance events. Small values of α indicate a relatively flat power law distribution in which large clustered disturbance events are relatively common, whereas large values of α indicate a relatively steep distribution dominated by smaller events. Over sufficiently large spatial scales, or timescales, the expectation of the above stochastic process is simply

$$dB(t)/dt = G - \mu B(t), \quad (2)$$

where $B(t)$ is the domain mean biomass, and G is the domain mean growth rate ($G = g$, in this simple homogeneous case). This model has the time-dependent solution $B(t) = G/\mu (1 - e^{-\mu t})$ and the dynamic equilibrium $B^* = G/\mu$.

[12] For the analyses described below, we ran the above model on a 1000×1000 grid cell domain for a 250 year spin-up period to achieve dynamic equilibrium, and then for an additional 1000 years for use in analyses. Following *Fisher et al.* [2008], for each time step (year) we use a pseudorandom process to choose the number of gaps in each size class from the power law distribution, and then place these gaps on randomly drawn center points chosen to prevent overlap. Biomass in gaps is set to zero, and biomass in all grid cells is increased by g . Figure 1 illustrates the evolution of the simulated landscape through the spin-up period in a range of disturbance regimes from highly clustered ($\alpha = 1.5$) to well distributed ($\alpha = 2.5$). We also adapted the model to address issues of environmental gradients, and nonlinear forest growth rates, explained in greater detail below. Using the model as the reference, we then calculated the effects of

limited sampling and coarse resolution averaging estimates of the state of the system. Next, we quantified how resulting errors propagated in model estimates of biomass and biomass flux.

3.1. Incomplete Observations of the State of the Land Surface

[13] The starting point in our analyses was to quantify the effects of limited sampling and coarse resolution averaging on estimates of the state of the land surface. To do this, we first produced a set of reference cases using the above model. To produce the set, values for g and μ were chosen to be on the order of values reported in the literature on forests ($g = 1$, $\mu = 0.02$), and the parameter α was varied from 1.1 to 3.0 in increments of 0.1 to represent a range of spatial scales of heterogeneity caused by disturbance events. For each reference case, we then produced two sets of corresponding scenarios. In the first set, we simulated the effects of resolution by averaging the state of the reference case at a range of resolutions from tree level (i.e., grid-cell level, no averaging) to 1000×1000 cell (i.e., domain average). In the second set, we simulated the effects of sampling by sampling the reference case at a range of intensities from 100% (i.e., complete coverage) to 0.01% (i.e., 1 in 1000 cells) and filling in gaps using bilinear interpolation between samples. For both sets, we then computed the tree level (grid-cell level) average absolute error between each reference case-scenario pair averaged over all time steps.

[14] Figure 2a illustrates the average tree level (grid-cell level) error in the state of the land surface that results from coarse measurements (averaging) as a function of α . As expected, accurate wall-to-wall measurements of the domain at the resolution of individual trees introduced no errors. However, measurements at coarser resolution averaged over tree level variability and translated into errors in the estimated state of the land surface. The errors increased with degree of averaging over a range of plausible α s, and were relatively less for landscapes with extremely high degrees of spatial clumping (extremely low α s). For all except the smallest α s, errors increased from 0 with tree level measurements to over 30, or greater than 60% of B^* , at resolutions of 1 ha or coarser. Cases with α s less than 1.5 had lower tree level error for a given resolution due to the relatively clumped spatial structure resulting from relatively clumped disturbances. Figure 2b illustrates analogous results for limited sampling. Both coarse resolution averaging and limited sampling translated into errors in the estimated state of the land surface at the tree scale.

3.2. Model Propagation of Errors: Homogeneous Land Surface

[15] Errors in the state of the land surface have the potential to propagate in model estimates of carbon stocks and fluxes at various scales. At the same time, in larger spatial scale averages, tree-level errors are expected to compensate to some degree. To investigate these phenomena, we quantified the effects of coarse resolution averaging on model estimates of biomass stocks and fluxes across a range of modeling scales. Here, modeling scale refers to the spatial scale of interest and is the scale at which comparisons with reference cases are made. It is also the spatial resolution at which the location of individual canopy trees (cells) are known.

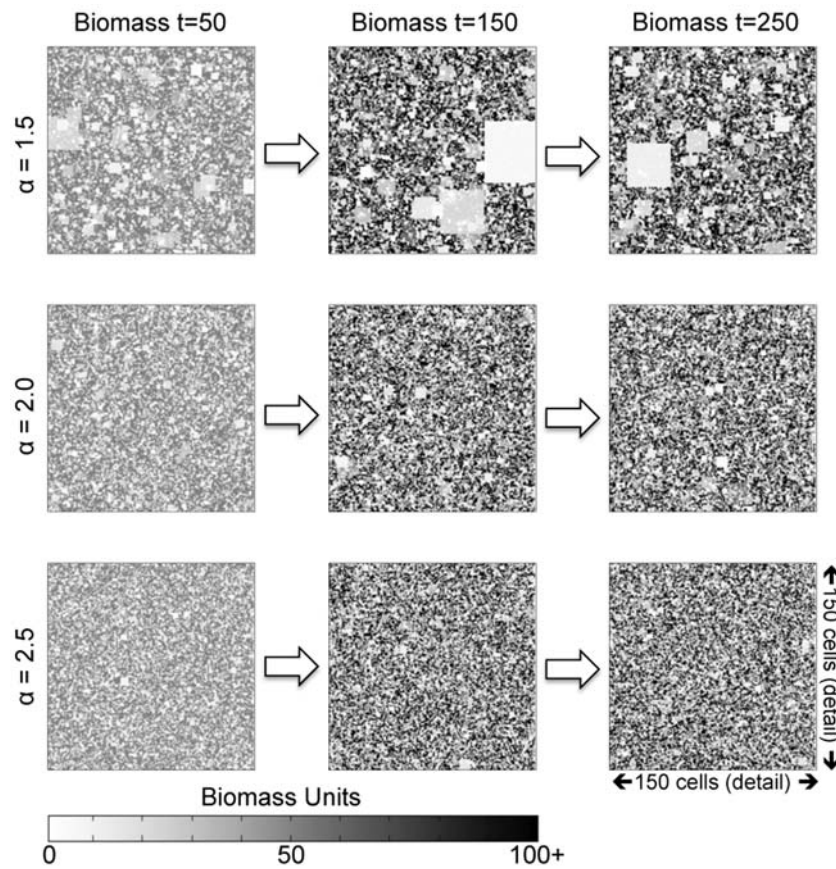


Figure 1. Time series snapshots of simulated forest biomass (150×150 cell detail) and a map of simulated forest biomass at spin-up years 50, 150, and 250. Different disturbance regimes from highly clustered ($\alpha = 1.5$) to well distributed ($\alpha = 2.5$) are shown.

[16] Specifically, we extended the methods described above and calculated the average absolute error in biomass and biomass flux between each reference case, and corresponding model scenarios of the same system, across a range

of both measurement and modeling resolutions. Model scenarios of biomass were calculated at each modeling scale as the average biomass of member cells. Model scenarios of biomass flux were calculated at each modeling resolution

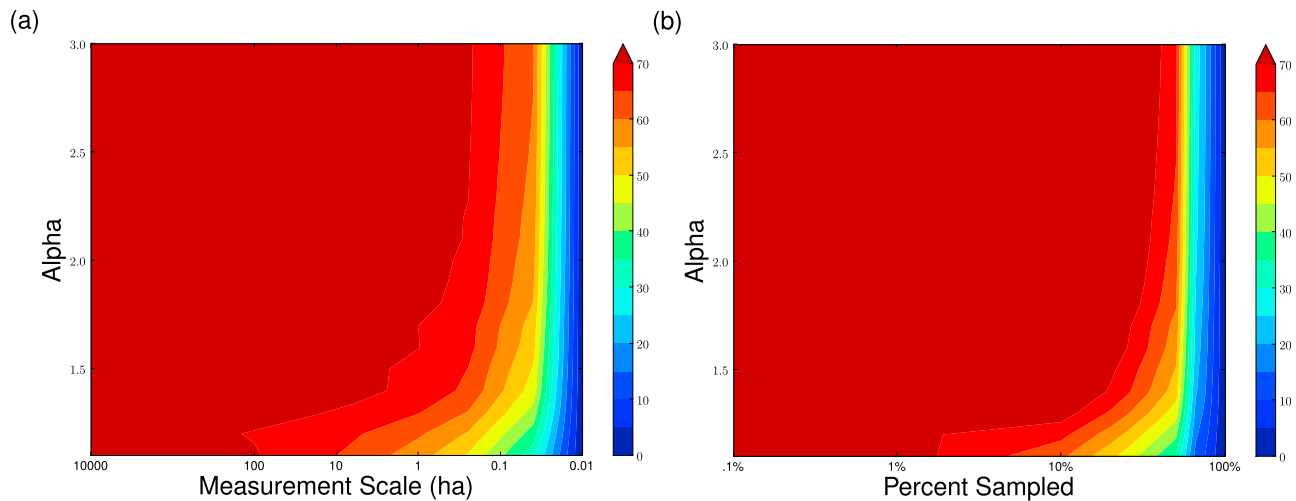


Figure 2. Tree-level average absolute errors in the state of the land surface introduced through (a) averaging and (b) limited sampling for systems with α s from 1.1 (clustered disturbance) to 3.0 (distributed disturbance).

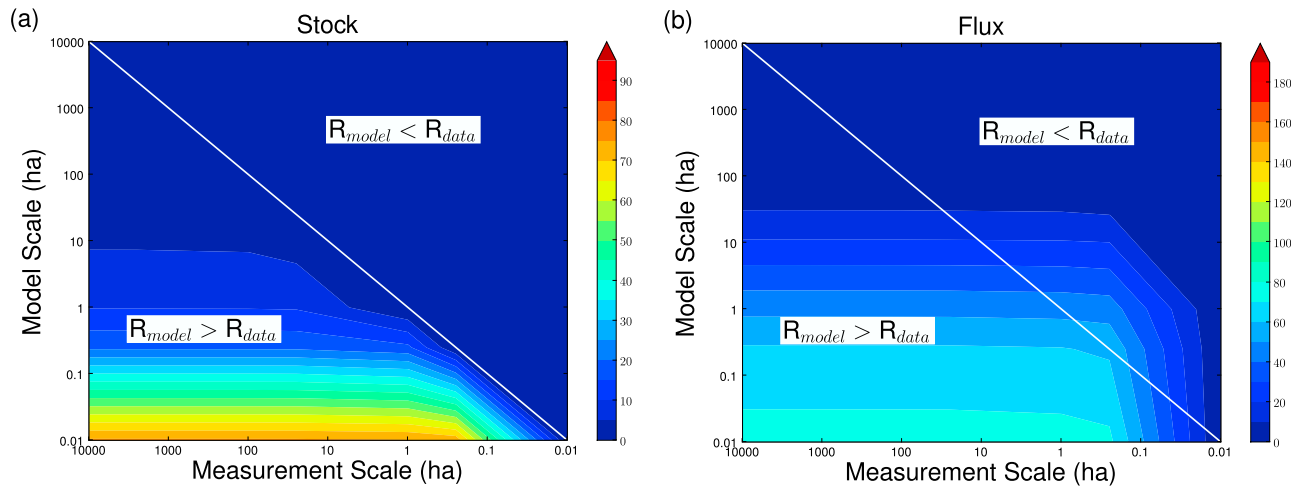


Figure 3. Average absolute error in biomass stocks and predicted fluxes for constant g in a uniform environment across all combinations of model and measurement resolution. (a) Absolute error in biomass stocks as a percent of equilibrium biomass, B^* . (b) Absolute error in predicted flux as a percent of annual growth rate, g . R refers to resolution which gets smaller as scale gets larger. In both cases, error increases as model resolution exceeds measurement resolution.

as the difference between modeled gains due to growth and losses due to disturbance. Gains of biomass were calculated at each modeling scale as the average growth of member cells, simply $G = g$ in this case. Losses of biomass were calculated at each modeling scale using the actual disturbance rates from the reference case applied to the estimated biomass of member cells, to isolate potential errors introduced from initialization.

[17] Figure 3 illustrates the errors that resulted from poor mismatches between model resolution and data resolution in estimates of biomass stocks and fluxes for the reference case $\alpha = 2.5$ (chosen to be representative of a wide range of α s 1.5–3, see above). In Figure 3a, average absolute error in biomass at the modeling resolution is shown as a percent of equilibrium biomass, B^* . Above the 1:1 line in Figure 3, data resolution is finer (higher) than model resolution, and thus errors at the modeling scale were minimal or nonexistent. Errors may exist at finer spatial scales, but these compensated at the coarser modeling scale. However, in cases where the data resolution was much coarser (lower) than the modeling resolution (below the 1:1 line), errors increased due to relatively large-scale averages of biomass incorrectly being applied at smaller scales. Modeling at tree-scale resolution with coarse resolution data created errors as large or larger than B^* at that scale. However, at modeling resolutions of 100 ha or coarser, the errors reduced to near zero regardless of the measurement resolution as these patches are large enough to always be near equilibrium in this system.

[18] Figure 3b illustrates corresponding results for model estimates of biomass flux. In Figure 3b the average absolute error in predicted biomass flux at the model resolution is shown as a percent of annual growth rate, g . The pattern of errors for biomass flux was qualitatively analogous to that for biomass stocks (Figure 3a). Modeling at tree-scale resolution with coarse resolution data created errors in flux greater than 100% of the annual growth rate, falling off to near zero at 100 ha or coarser modeling resolutions. However, unlike

biomass, errors in biomass flux also existed at relatively fine (high) model resolutions even without errors in biomass stock at that scale, when disturbances generally do not take average stock at that scale.

3.3. Model Propagation of Errors: Nonlinear Growth Rates

[19] The above analyses were based on the simplest possible forest gap-type model. To make the model more realistic, we investigated the effects of local nonlinear growth rates as forest gaps fill and trees age. Specifically, we replaced the assumption of constant g , with a local nonlinear growth rate (g_i) in which the growth rate is a nonlinear function of the biomass in the grid cell (b_i).

$$g_i(t) = -(b_i(t) + K_1)(b_i(t) - K_2)/K_3 \quad (3)$$

This equation has slow initial growth, fast midlife growth, and a slow approach to maximum biomass. For the following simulations, we set $K_1 = 2$, $K_2 = 100$, $K_3 = 1000$ to create a system with a dynamic equilibrium (B^*) approximately equal to that of the linear formulation described above. Using this new parameterization of the model, we then repeated the analyses for model propagation of errors described above.

[20] Figure 4 illustrates the results. Biomass error was unaffected by the nonlinearity because model estimates of biomass were based on direct observations of structure. However, errors in predicted flux increased dramatically across all simulations that relied on averaging above the tree scale. The increased error in model prediction of flux resulted from the misrepresentation of tree-level heterogeneity by averages, and the fact that the growth function of the average is not equal to the average of the growth function. In these simulations, the lack of information on tree-level heterogeneity is debilitating to model predictions of dynamics.

[21] Could the loss of tree-level information from coarse measurements be mitigated with a fusion of averaging and

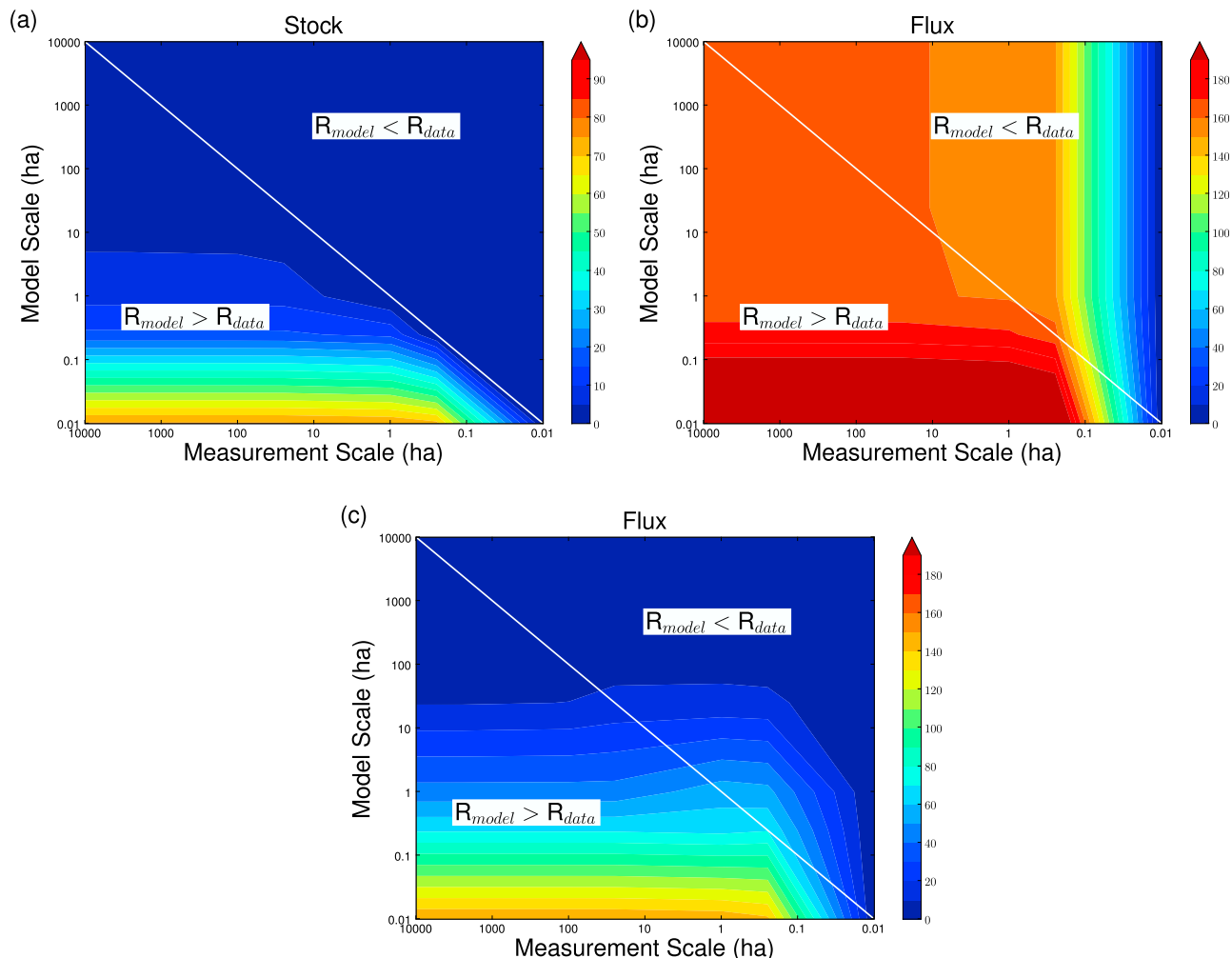


Figure 4. Average absolute error in biomass stocks and predicted fluxes for nonlinear G in a uniform environment across all combinations of model and measurement resolution. (a) Average absolute error in biomass stocks as a percent of equilibrium biomass, B^* . (b) Average absolute error in predicted flux as a percent of domain mean growth rate using mean measured tree biomass per model grid cell. Error increases rapidly as model resolution becomes coarser due to the difference between the growth rate of an average tree, and the average growth rate of all trees. (c) Average absolute error in predicted flux as a percent of domain mean growth rate using a distribution of trees in each model grid cell. Knowledge of the distribution of trees reduces errors to similar to constant growth case.

high-resolution sampling technologies? To investigate this, we recomputed flux errors assuming both average biomass and the subgrid-scale distribution of biomass were known at the modeling scale. As Figure 4c illustrates, knowing both the average biomass and distribution of tree level biomass values largely eliminated the errors in growth described above. As with the linear case, coarse measurements lead to large errors at the tree scale, but as model resolution approaches 100 ha, errors drop to near zero.

3.4. Model Propagation of Errors: Nonlinear Growth Rates and Environmental Gradients

[22] To make the model still more realistic, we next considered the case in which forest growth rates are locally nonlinear, and the environment is nonhomogeneous and has an environmental gradient that strongly affects growth rates.

For example, consider a mountainside on which potential tree growth rates are high in the valley and low toward the summit. How does the combination of nonlinear tree-level growth rates and a strong environmental gradient affect model-data requirements? To investigate this, we adapted the above model to be scaled by an environmental factor affecting plant growth rates

$$g_i(t) = -(b_i(t) + K_1 e_i)(b_i(t) - K_2 e_i)/K_3 e_i. \quad (4)$$

For simplicity, the local environment, e_i , and thus the initial growth rate, maximum growth rate, and maximum biomass, varied linearly from 0.2 at one edge of the domain to 2.0 at the other in 10 tree (1 ha) stripes. The overall domain average maximum growth rate and equilibrium biomass were all approximately equal to the above cases. Using this new parameterization of the model, we then repeated the analyses

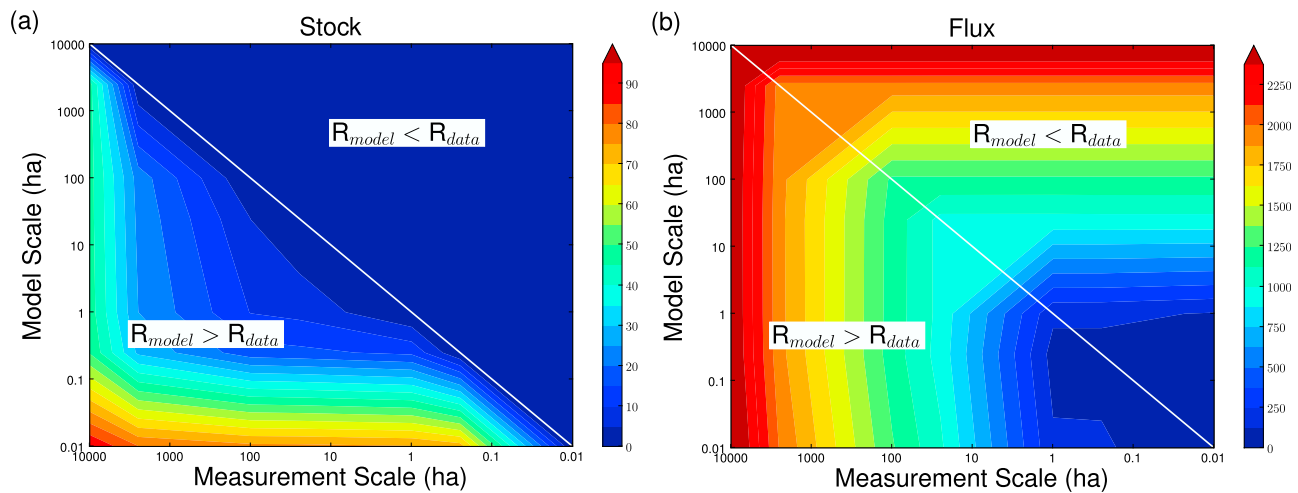


Figure 5. Average absolute error in biomass stocks and predicted fluxes for nonlinear growth rates on an environmental gradient. (a) Average absolute error in measured biomass stocks as a percent of equilibrium biomass B^* . Errors are larger than previous cases where measurements are coarser than the scale of the environmental gradient. (b) Average absolute error in predicted flux as a percent of domain mean growth rate. Coarse resolution measurements lead to large errors in predicted flux due to modeling trees in inappropriate environmental conditions. Coarse resolution modeling leads to large errors in predicted flux due to inability to capture the environmental gradient. To reduce errors, both the model and measurements need to be done at the scale of the environmental heterogeneity.

described above simulating the fusion of technologies described there in which both the average biomass and distribution of biomass values were assumed to be accurately known for each modeling unit.

[23] Figure 5 illustrates the results. Model biomass error was analogous to previous results, but greater in cases where measurements were substantially coarser than the scale of the environmental gradient. In these cases, relatively coarse biomass averages were inaccurate at the resolution over which biomass varied due to the underlying environmental gradient. This led to some cells being initialized outside of the range of possibility, and propagated as large biomass flux errors due to mortality.

[24] Figure 5b illustrates that a potentially optimal/efficient scale exists for model-data combinations at which both the model and measurements are sufficient. If measurement resolution is too coarse, errors in initial biomass and resulting flux are large. If model resolutions are too coarse, the effect of the environmental gradient on growth rates is missed. Errors were minimized when both the modeling resolution and measurement resolution approached the scale of the underlying environmental heterogeneity, in this case at ≤ 1 ha.

4. Recent Results From a Mechanistic Model

[25] Both the importance of forest structure to forest dynamics, and the theoretical studies on scale/resolution described above, provide a sound basis for including fine-scale data on vegetation structure in mechanistic models of forests. While the above theoretical studies are based on the simplest forms of forest gap models, over the last decade an advanced mechanistic model of forest ecosystem dynamics has been developed in which individual-based forest dynamics can be efficiently modeled over large scales (Ecosystem Demography (ED) model [Hurtt *et al.*, 1998; Moorcroft

et al., 2001]). Studies using this model have illustrated the importance of and potential for incorporating data on vegetation structure to improving mechanistic model predictions, and are described below in the context of linking models and data on vegetation structure.

[26] The ED model is an individual-based model of vegetation dynamics with integrated submodels of plant growth, mortality, phenology, biodiversity, disturbance, hydrology, and soil biogeochemistry [Moorcroft *et al.*, 2001]. Individual plants of different functional types compete mechanistically in ED under local environmental conditions for light, water, and nutrients. ED differs from most other terrestrial models by formally scaling up physiological processes through individual-based vegetation dynamics to ecosystem scales, while simultaneously modeling natural disturbances, land use, and the dynamics of recovering lands. ED has recently been implemented in South and Central America [Moorcroft *et al.*, 2001], the U.S. [Hurtt *et al.*, 2002; Albani *et al.*, 2006; Medvigy *et al.*, 2009], and is now a global model. Of particular relevance to this study is the fact that all plants in ED have an explicit height, a property that allows for a direct connection to data on vegetation structure.

[27] Recent studies using ED have used lidar remote sensing data of vegetation structure to initialize and test predictions of carbon stocks and fluxes at a range of experimental study sites in North and South America with aircraft data. Hurtt *et al.* [2004] used lidar measurements of canopy height to initialize the ED model at the La Selva Biological Station in Costa Rica. The method of initialization used a look-up table approach (Figure 6), in which 1 ha resolution mean canopy height data from the LVIS sensor were used to index precomputed ED-based projections of how individual-based forest structure can be expected to change through succession at that site. Lidar-initialized ED estimates of aboveground biomass were within 1.2% of regression-based

Regression-Based Approach

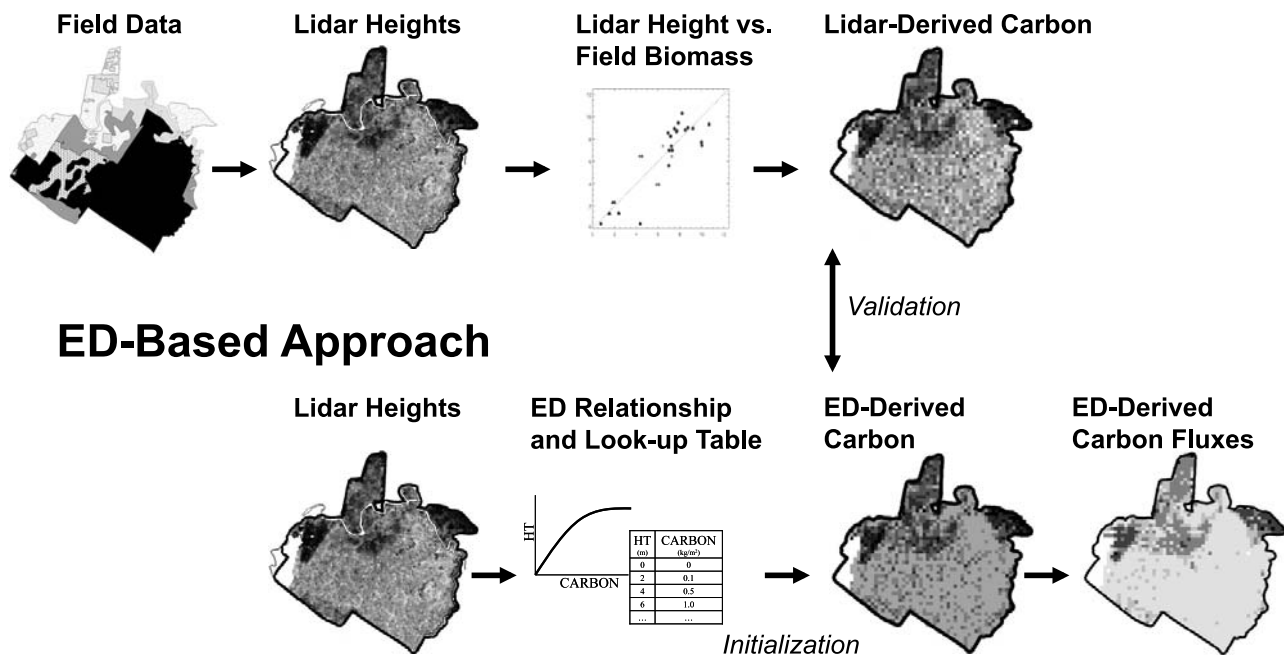


Figure 6. Regression- and ED-based approaches for estimating stocks and fluxes. In the regression-based approach, field data and lidar data are used to produce a regression model for estimating biomass across the domain. In the ED-based approach, lidar data are used to initialize the ED model using a look-up table approach. Resulting lidar-initialized estimates of biomass are validated using the regression-based approach, and the ED model is then used to predict carbon fluxes. From *Hurt et al.* [2004].

approaches using field data, and the resulting predictions of carbon flux were tightly constrained relative to bracketing alternatives that lacked data on vegetation structure (Figure 7).

[28] In a follow-on study, *Hurt et al.* [2007] used repeat lidar data on canopy height at La Selva Biological Station in Costa Rica to initialize and test ED model predictions. Airborne lidar remote sensing was used to measure spatial heterogeneity in the vertical structure of vegetation in 1998 and 2005. Using the approach described above, 1998 lidar data were first used to initialize the ED model. Lidar data from 2005 were then used to test model predictions of canopy height change during the interval. Lidar-initialized ED estimates of changes in maximum canopy height were comparable to but lower than observed over the whole domain (0.53 ± 0.4 m modeled versus 0.85 ± 0.9 m observed). Most of the model-data difference was due to growth of primary forest trees that exceeded model estimates (0.04 ± 0.31 m modeled versus 0.44 ± 0.9 m observed). The model-data comparison was significantly better over secondary forest areas (1.71 ± 0.9 m modeled versus 1.84 ± 0.18 m observed). Model predictions of change were also close to observations of change at finer spatial scales, with a model-data RMSE of < 0.5 m at scales > 20 ha, and < 0.25 m at scales > 50 m.

[29] In some systems, patterns or gradients in environmental conditions are known to exert strong influences on patterns of vegetation structure. Building on the study of *Hurt et al.* [2004], *Thomas et al.* [2008] used lidar canopy height data to initialize the ED model at the mountainous Hubbard Brook Experimental Forest (HBEF) in NH. At

HBEF, spatial patterns in forest structure, including lidar measurements of canopy height are strongly dependent on the environmental and disturbance that varied over the 700 m elevation gradient. In the study, lidar canopy height initialized aboveground biomass to within 6% of the field value and yielded carbon flux predictions that compared well to ground-based carbon inventory measurements. In a sensitivity analysis, the study demonstrated that accurate predictions of

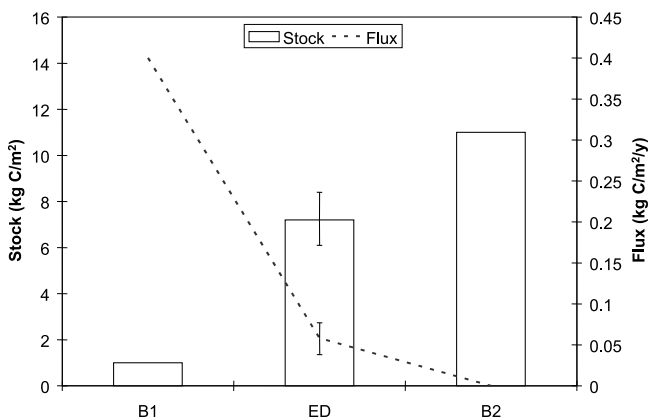


Figure 7. ED model estimates of domain mean above-ground carbon stocks (bars) and fluxes (dashed line). “ED” indicates the lidar-initialized model estimates. Bars indicate uncertainty. For comparison, B1 and B2 are bracketing scenarios from ED that result from the absence of data on vegetation structure. From *Hurt et al.* [2004].

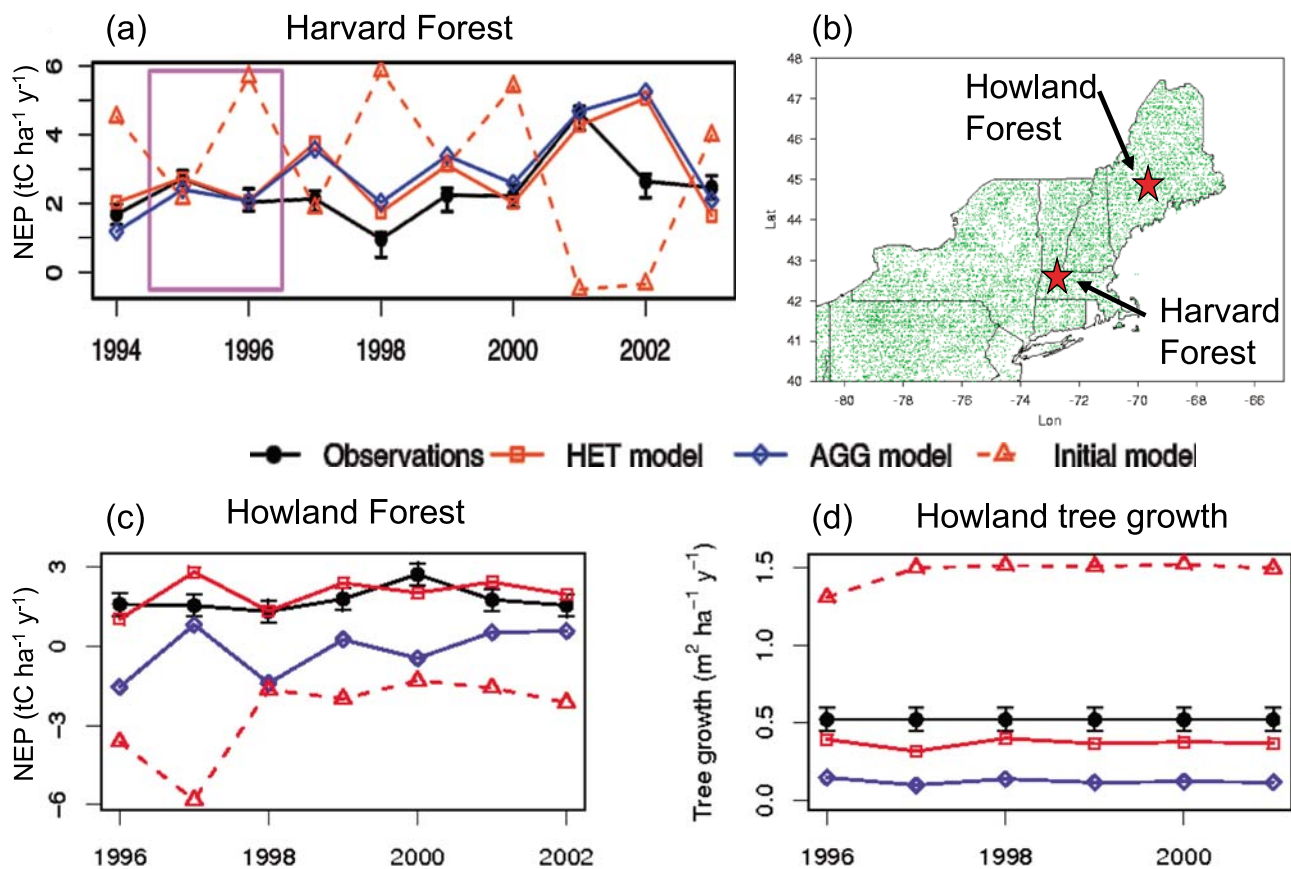


Figure 8. Results from ED2 using data on vegetation structure to improve model predictions at Harvard Forest and Howland Forest. (a) Predicted and observed patterns of Net Ecosystem Productivity (NEP) at Harvard Forest. Units are $\text{tC ha}^{-1} \text{yr}^{-1}$. The black line indicates the flux-tower measurements, and the error bars indicate the 2σ error estimates for the observations. The solid red line indicates the predictions of the ED2 biosphere model that incorporates fine scale ecosystem heterogeneity following optimization against 2 years of flux tower and forest demography observations at Harvard Forest (HET). The blue line indicates the predictions of a modified version of the ED2 model that was similarly optimized against the observations, but used a conventional, aggregated “big-leaf” representation of the plant canopy (AGG). The dashed red line indicates the predictions from an initial, unoptimized version of the ED2 biosphere model (initial). The magenta box indicates the 2 years of observations used to constrain the two optimized model formulations. (b) Map showing the locations of Harvard and Howland Forest research sites. Green dots show the locations of forest inventory measurements in the region. Patterns of annual (c) net ecosystem productivity (NEP, $\text{tC ha}^{-1} \text{yr}^{-1}$) and (d) tree growth rates (basal area increment ($\text{m}^2 \text{ha}^{-1} \text{yr}^{-1}$)) at Howland Forest for the period 1996–2002. Note that there was no change in model parameters between the simulations shown in Figure 8a and those shown in Figures 8c and 8d. While the AGG model is able to capture the pattern of NEP variability at Harvard Forest, it strongly underpredicts the observed rates of NEP and tree growth of Howland. In contrast, the predictions of the HET model closely match both the observed pattern of NEP at Harvard, and the observed rates of NEP and tree growth rates at Howland. For further details, see text and Medvigy *et al.* [2009].

carbon fluxes were highly sensitive to both model resolution and the resolution of lidar inputs, in order to appropriately account for elevation-dependent influences on forest dynamics.

[30] Most recently, Medvigy *et al.* [2009] illustrated how fine-scale measurements of forest structure can be combined with eddy-flux measurements to improve the predictive abilities of terrestrial biosphere models. The ED2 terrestrial biosphere model was initialized with the observed ecosystem structure in the footprint of the Harvard Forest eddy-flux tower, and then fitted to the 1995 and 1996 hourly, monthly

and yearly CO_2 and ET flux data, and to observed basal area growth and mortality in these years (pink box in Figure 8a). Prior to optimization, the model significantly underestimated the seasonal cycle of net ecosystem productivity measured by the flux tower, and significantly overestimated measured rates of individual tree growth and mortality. After fitting, the model accurately captured the observed fluxes of CO_2 and H_2O , canopy growth, and mortality over timescales spanning hours to decades (Figure 8a). The performance of the optimized ED2 biosphere model was then evaluated at a different site, Howland Forest (Figure 8b). The model was initialized

with the observed canopy composition in the tower footprint, but model parameters were not reoptimized. Despite the markedly different forest composition between the Howland and Harvard Forest sites (conifer-dominated as opposed to mixed hardwood), there was a substantial improvement in model predictions of the 5 year CO₂ flux record, and measured tree growth dynamics at Howland (Figures 8c and 8d). All optimized parameter values fell within a priori acceptable ranges. The parameters most responsible for the improved goodness of fit were an increased maximum photosynthetic rate of hardwoods, a marked increase in the rate of fine root turnover, and a decrease in the carbon allocation to fine roots in conifer species. The transferability between very different ecosystems provides confidence that the optimization of the model actually tests the hypotheses embodied in its formulation, rather than being a trivial exercise in site-specific model tuning. A key conclusion of this study was that the inclusion of forest structure and growth measurements into the model optimization was essential for constraining plant carbon allocation. This is particularly significant since this aspect of ecosystem models play a critical role in determining rates of plant growth and thus rates of aboveground biomass accumulation, but are nearly impossible to measure directly. The study also shows that the improvements in the model's ability to capture regional variation in ecosystem carbon fluxes and biomass dynamics is contingent on having measurements of fine subgrid-scale variation in canopy structure, and an ability to explicitly represent this heterogeneity within the ecosystem model formulation.

5. Discussion

[31] Decades of research have established the importance of vegetation structure to forest dynamics. Forecast models of terrestrial ecosystem (carbon) dynamics require data on vegetation structure for accurate initialization and testing. The studies described here using both simple theoretical models, and advanced mechanistic models, combined with the primacy of forecasting for policy decisions, suggest that the requirements of models may actually drive data requirements for future missions. Because vegetation dynamics are generally local, nonlinear, and depend strongly on environmental conditions, both models and data must track fine-scale (tree-level) heterogeneity in vegetation structure at scales determined by underlying environmental gradients. Generally, limited sampling and/or coarse resolution create errors in the estimated initial state of the land surface, which are then propagated in model initialization and prediction errors.

[32] Both the theoretical and mechanistic approaches described here suggest that within these guidelines, the quantitative model-data requirements depend strongly the patterns and scales of underlying environmental heterogeneity in terrestrial systems. In principle, if one knew important characteristics of the distribution of disturbance events (i.e., α , see above), and the underlying environmental gradients that affect plant growth rates (e.g., elevation, soils, etc.), then an efficient modeling resolution and set of data requirements for vegetation structure could be determined. However, these fundamental characteristics vary and are not adequately known globally. Over relatively homogeneous terrain, relatively coarse model-data resolutions that provided accurate estimates of both average structure and information on the

subgrid-scale distribution of structure would likely suffice. But over complex terrain with steep environmental gradients, such as elevation gradients, soil gradients, and the like, much higher model-data resolutions would be required for accurate predictions.

[33] Both the theoretical and mechanistic studies described here suggest that ~1 ha data on average vegetation structure, and its subgrid-scale (tree-level) heterogeneity, would be sufficient to drive accurate model predictions over complex forested systems on steep environmental gradients. Data on vegetation structure with these properties could potentially be obtained by a fusion of the radar and lidar technologies envisioned for DESDynI, and this very high spatial resolution would be useful for localizing biomass stocks and fluxes and interpreting forest dynamics. However, ~1 ha spatial resolution may prove impractical and/or unachievable globally. Functionally, the driver of this scale is heterogeneity, and the general need to accurately measure and model forest structure at the scales that determine dynamics in order to minimize prediction errors. In addition to new global studies on relevant environmental heterogeneity, new modeling schemes that include consistent subgrid-scale parameterizations of environmental heterogeneity should be investigated as a means of potentially easing this resolution requirement.

[34] Current data from comprehensive ground-based forest inventories are impressive in some regions (e.g., U.S. Forest Inventory), but nonstandard internationally and nearly absent from some vast, remote, and important regions such as the Amazon. Data from the current ICESAT mission are proving invaluable for their consistency and global coverage of vegetation structure, but sampling is limited. Over important regions such as the domain of hurricane Katrina, > 25% of 0.25° × 0.25° grid cells are entirely unsampled (K. Dolan, personal communication, 2010). DESDynI as currently conceived will provide a qualitative leap over present data availability on vegetation structure globally. In preparation, new modeling studies are also needed to quantify and develop strategies to anticipate and minimize errors that may result where limited sampling and/or coarse resolution averaging remain important, assess the importance of potential data/sensor errors, and build off the studies described here to develop and implement a robust global model-data framework for assimilating future data in mechanistic forecast models of terrestrial carbon dynamics.

[35] **Acknowledgments.** We gratefully acknowledge helpful comments in developing this paper from members of the NASA Vegetation Structure Working Group. We are also grateful for the constructive comments of two anonymous reviewers. This work was supported by research grants from the NASA Terrestrial Ecology and Interdisciplinary Science Programs.

References

- Albani, M., D. Medvigy, G. C. Hurtt, and P. R. Moorcroft (2006), The contributions of land-use change, CO₂ fertilization, and climate variability to the eastern U.S. carbon sink, *Global Change Biol.*, 12, 2370–2390, doi:10.1111/j.1365-2486.2006.01254.x.
- Anthes, R. A., et al. (2007), *Earth Science and Applications From Space: National Imperatives for the Next Decade and Beyond*, Natl. Acad. Press, Washington, D. C.
- Busing, R. T., and D. Maily (2004), Advances in spatial, individual-based modeling of forest dynamics, *J. Veg. Sci.*, 15, 831–842.

- Ek, A. R., and R. A. Monserud (1974), FOREST: Computer model for the growth and reproduction simulation for mixed species forest stands, *Res. Rep. A2635*, Coll. of Agric. and Life Sci., Univ. of Wis., Madison.
- Fisher, J., G. C. Hurtt, R. Q. Thomas, and J. Q. Chambers (2008), Clustered disturbances lead to bias in large-scale estimates based on forest sample plots, *Ecol. Lett.*, *11*, 554–563, doi:10.1111/j.1461-0248.2008.01169.x.
- Goldewijk, K. K. (2001), Estimating global land use change over the past 300 years: The HYDE database, *Global Biogeochem. Cycles*, *15*, 417–433, doi:10.1029/1999GB001232.
- Hegy, F. (1974), A simulation model for managing jack-pine stands, in *Growth Models for Tree and Stand Simulation*, edited by J. Fries, pp. 74–90, R. Coll. of For., Stockholm, Sweden.
- Hurt, G. C., P. R. Moorcroft, S. W. Pacala, and S. A. Levin (1998), Terrestrial models and global change: Challenges for the future, *Global Change Biol.*, *4*(5), 581–590, doi:10.1046/j.1365-2486.1998.t01-1-00203.x.
- Hurt, G. C., S. W. Pacala, P. R. Moorcroft, J. Caspersen, E. Shevliakova, R. Houghton, and B. Moore III (2002), Projecting the future of the U.S. carbon sink, *Proc. Natl. Acad. Sci. U. S. A.*, *99*(3), 1389–1394(PNAS), doi:10.1073/pnas.012249999.
- Hurt, G. C., R. Dubayah, J. Drake, P. Moorcroft, S. Pacala, and M. Fearon (2004), Beyond potential vegetation: Combining lidar remote sensing and a height-structured ecosystem model for improved estimates of carbon stocks and fluxes, *Ecol. Appl.*, *14*(3), 873–883, doi:10.1890/02-5317.
- Hurt, G. C., S. Frolking, M. G. Fearon, B. Moore III, E. Shevliakova, S. Malyshev, S. W. Pacala, and R. A. Houghton (2006), The underpinnings of land-use history: Three centuries of global gridded land-use transitions, wood harvest activity, and resulting secondary lands, *Global Change Biol.*, *12*, 1208–1229, doi:10.1111/j.1365-2486.2006.01150.x.
- Hurt, G. C., R. Q. Thomas, and R. Dubayah (2007), Beyond potential vegetation II: Using repeat lidar data on changes in vegetation height to test model predictions of ecosystem dynamics, *Eos Trans. AGU*, *88*(52), Fall Meet. Suppl., Abstract B42B-06.
- Kohyama, T. (1993), Size-structured tree populations in gap dynamic forests—the forest architecture hypothesis for stable coexistence of species, *J. Ecol.*, *81*, 131–143, doi:10.2307/2261230.
- Korzukhin, M. D., and M. Y. Antonovski (1992), Population-level models of forest dynamics, in *A Systems Analysis of the Global Boreal Forest*, edited by H. H. Shugart, R. Leemans, and G. B. Bonan, pp. 334–372, Cambridge Univ. Press, Cambridge, U. K.
- Kotchenova, S. Y., X. D. Song, N. V. Shabanova, C. S. Potter, Y. Knyazikhin, and R. B. Myneni (2004), Lidar remote sensing for modeling gross primary production of deciduous forests, *Remote Sens. Environ.*, *92*, 158–172, doi:10.1016/j.rse.2004.05.010.
- Landsberg, J. J., and R. H. Waring (1997), A generalized model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning, *For. Ecol. Manage.*, *95*, 209–228, doi:10.1016/S0378-1127(97)00026-1.
- Le Toan, T., S. Quegan, I. Woodward, M. Lomas, N. Delbart, and G. Picard (2004), Relating radar remote sensing of biomass to modeling of forest carbon budgets, *Clim. Change*, *67*, 379–402, doi:10.1007/s10584-004-3155-5.
- Lucas, N. S., and P. J. Curran (1999), Forest ecosystem simulation modeling: The role of remote sensing, *Prog. Phys. Geogr.*, *23*, 391–423.
- Medvigy, D. M., 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: The Ecosystem Demography model version 2, *J. Geophys. Res.*, *114*, G01002, doi:10.1029/2008JG000812.
- Mitchell, K. J. (1969), Simulation of growth of even-aged stands of white spruce, *Yale Univ. Sch. For. Bull.*, *75*, 1–48.
- Mitchell, K. J. (1975), Dynamics and simulated yield of Douglas-fir, *For. Sci. Monogr.*, *17*, 1–39.
- Moorcroft, P. R., G. C. Hurtt, and S. W. Pacala (2001), A method for scaling vegetation dynamics: The ecosystem demography model (ED), *Ecol. Monogr.*, *71*(4), 557–585.
- Newnam, R. M. (1964), The development of a stand model for douglas-fir, Ph.D. thesis, Univ. of B. C., Vancouver, B. C., Canada.
- Nightingale, J. M., S. R. Phinn, and A. A. Held (2004), Ecosystem process models at multiple scales for mapping tropical forest productivity, *Prog. Phys. Geogr.*, *28*, 241–281, doi:10.1191/0309133304pp411ra.
- Pascual, M., and F. Guichard (2005), Criticality and disturbance in spatial ecological systems, *Trends Ecol. Evol. Biol.*, *20*, 88–95, doi:10.1016/j.tree.2004.11.012.
- Patenaude, G., R. Miline, M. Van Oigen, C. S. Rowland, and R. A. Hill (2008), Integrating remote sensing datasets into ecological modeling: A Bayesian approach, *Int. J. Remote Sens.*, *29*, 1295–1315, doi:10.1080/01431160701736414.
- Plummer, S. E. (2000), Perspectives on combining ecological process models and remotely sensed data, *Ecol. Modell.*, *129*, 169–186, doi:10.1016/S0304-3800(00)00233-7.
- Porté, A., and H. H. Bartelink (2002), Modelling mixed forest growth: A review of models for forest management, *Ecol. Modell.*, *150*, 141–188, doi:10.1016/S0304-3800(01)00476-8.
- Purves, D., and S. W. Pacala (2008), Predictive models of forest dynamics, *Science*, *320*(5882), 1452–1453, doi:10.1126/science.1155359.
- Ranney, J. W., M. C. Bruner, and J. B. Levenson (1981), The importance of edge in the structure and dynamics of forest islands, in *Forest Island Dynamics in Man-Dominated Landscapes*, edited by R. L. Burgess and D. M. Sharpe, pp. 67–92, Springer, New York.
- Ranson, K. J., G. Sun, R. G. Knox, E. R. Levine, J. F. Weishampel, and S. T. Fifer (2001), Northern forest ecosystem dynamics using coupled models and remote sensing, *Remote Sens. Environ.*, *75*, 291–302, doi:10.1016/S0304-4257(00)00174-7.
- Sands, P. J., and J. J. Landsberg (2002), Parameterisation of 3-PG for plantation grown Eucalyptus globus, *For. Ecol. Manage.*, *163*, 273–292, doi:10.1016/S0378-1127(01)00586-2.
- Shinozaki, K., K. Yoda, K. Hozumi, and T. A. Kira (1964), Quantitative analysis of plant form: The pipe model theory: I. Basic analysis, *Jpn. J. Ecol.*, *14*, 97–105.
- Shugart, H. H. (1984), *A Theory of Forest Dynamics: The Ecological Implications of Forest Succession Models*, Springer, New York.
- Shugart, H. H. (1998), *Terrestrial Ecosystems in Changing Environments*, 537 pp., Cambridge Univ. Press, Cambridge, U. K.
- Shugart, H. H. (2008), Forestry management, in *Population Dynamics*, vol. 2, *Encyclopedia of Ecology*, edited by S. E. Jorgensen and B. D. Fath, pp. 1680–1686, Elsevier, New York.
- Suzuki, T., and T. Umemura (1967), Forest transition as a stochastic process. III, *J. Jpn. For. Soc.*, *49*, 208–210.
- Suzuki, T., and T. Umemura (1974), Forest transition as a stochastic process. V, *J. Jpn. For. Soc.*, *56*, 195–204.
- Thomas, R. Q., G. C. Hurtt, R. Dubayah, and M. Schilz (2008), Using lidar data and a height structured ecosystem model to estimate forest carbon stocks and fluxes over mountainous terrain, *Can. J. Remote Sens.*, *34*(2), S351–S363.
- Turner, D. P., S. V. Ollinger, and J. S. Kimball (2004), Integrating remote sensing and ecosystem process models for landscape- to regional-scale analysis of the carbon cycle, *BioScience*, *54*, 573–584, doi:10.1641/0006-3568(2004)054[0573:IRSAEP]2.0.CO;2.
- Urban, D. L. (1990), A versatile model to simulate forest pattern: A user guide to ZELIG, version 1.0, Environ. Sci. Dep., Univ. of Va., Charlottesville.
- Woodward, F. I., T. M. Smith, and W. R. Emanuel (1995), A global land primary productivity and phytogeography model, *Global Biogeochem. Cycles*, *9*, 471–490, doi:10.1029/95GB02432.
- Yoda, K., T. Kita, H. Ogawa, and K. Hozumi (1963), Self-thinning in over-crowded pure stands under cultivated and natural conditions, *J. Biol.*, *14*, 107–129.

R. Dubayah, Department of Geography, University of Maryland, College Park, MD 20740, USA.

J. Fisk and G. C. Hurtt, Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, NH 03824, USA. (george.hurt@unh.edu)

P. R. Moorcroft, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA.

H. H. Shugart, Department of Environmental Sciences, University of Virginia, Charlottesville, VA 22903, USA.

R. Q. Thomas, Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA.