

Ecological Archives A/E/M000-000-A#

Jeremy W. Lichstein, Jonathan Dushoff, Kiona Ogle, Anping Chen, Drew W. Purves, John P. Caspersen, and Stephen W. Pacala. 2009. Unlocking the forest inventory data: relating individual-tree performance to unmeasured environmental factors. *Ecological Applications* VOL:pp–pp.

Appendix E. Estimating height-diameter allometries and the shade-tolerance index (*ST*) for all U.S. tree species.

I. Height allometries

Since 1999, FIA data include height measurements. However, pre-1999 FIA data (which currently comprise most of the available plot re-measurements) do not include height data for most of the U.S. Therefore, incorporating tree height into analyses of pre-1999 FIA data requires height-dbh allometries, which we estimated using the post-1999 FIA data. Because we wished to treat our field data the same as we would treat FIA data, we used these same allometries in the analyses presented in this paper.

In December 2006, we downloaded all available FIA data from www.fia.fs.fed.us/, and we estimated height-dbh allometries for all species listed in FIA documentation. We excluded data from plantations and plots with any documented harvesting. To reduce the computational requirements for the analysis, we randomly subsampled from the data for species with unnecessarily large sample sizes (see *Subsampling algorithm* below).

To constrain the height allometry parameters for rare species, we adopted a Bayesian hierarchical modeling approach (Gelman et al. 2004, Clark 2005, Ogle and Barber 2008) with the following nested taxonomic levels: species, genus, family, order, division, and base, where the base is the ‘root’ of the taxonomic hierarchy. Our approach is ‘hierarchical’ both in the statistical sense (i.e., each allometric parameter was assumed to be a sample from a probability distribution) and in the taxonomic sense (i.e., we used the taxonomic hierarchy to structure the analysis). This approach allows rare taxa to borrow support from common taxa, particularly those that are closely related.

We assumed a power law relationship between height and dbh, which implies a linear model in log-space:

$$\ln(\text{height}) = a + b \cdot \ln(\text{dbh}) + \varepsilon \quad (\text{E1})$$

where a and b are species-specific parameters, $\varepsilon \sim \text{normal}(0, \sigma^2)$, and σ^2 is assumed constant across species. Prior to analysis, diameter at root collar measurements (for multi-stemmed woodland species in the western U.S.) were converted to dbh following Chojnacky and Rodgers (1999). Allometric parameters were assumed to be samples from normal distributions with taxon-specific means and level-specific variances. For example, parameters a and b for species i in genus j were assumed to be distributed as $a_{s,i} \sim \text{normal}(a_{g,j}, \sigma_{g,a}^2)$, and $b_{s,i} \sim \text{normal}(b_{g,j}, \sigma_{g,b}^2)$, where the subscripts ‘s’ and ‘g’ denote, species- and genus-level parameters, respectively; and the parameters $\sigma_{g,a}^2$ and $\sigma_{g,b}^2$ are, respectively, the variance in a and b among species within genera. Similarly, genus-level

a and b were assumed to be samples from family-level distributions, family-level a and b were assumed to be samples from order-level distributions, etc.

All priors were noninformative. The base-level (taxonomic root) priors for a and b were normal with mean zero and variance 1000. The only other priors that must be specified concern the variance parameters. Following Gelman (2006), we assumed uniform priors for the standard deviations, with $\sigma \sim \text{uniform}(0,100)$ at all taxonomic levels.

We used WinBUGS (Spiegelhalter et al. 2004) to generate samples from the posterior distribution of the parameters (Gilks et al. 1996). Three MCMC chains were run in parallel, and convergence was assessed with the ‘bgr’ tool in WinBUGS, which compares the within- to pooled-chain central 80% posterior intervals (Brooks and Gelman 1998). After the marginal posteriors for all parameters had converged, we simulated an additional 100,000 samples from each chain, saving every 20th sample to eliminate any serial autocorrelation. Thus, we saved 5000 independent samples from each of the three chains. Each sample can be viewed as a vector drawn at random from the joint posterior distribution of the allometric parameters. The species-level results are summarized in Supplement 1.

Error propagation

We used the above samples to propagate uncertainty in individual heights through all light-model analyses (including the uncertain- L growth analysis; table 3) involving height (Models CRN and CRNS; see table 2 and equation 4). Each of the three MCMC chains in our light-model analysis was paired with one of the three sets of 5000 samples from the WinBUGS height-allometry analysis. We randomly selected (with replacement) one of the 5000 height-allometry parameter vectors every 100th MCMC step in our light-model analysis, resulting in ≥ 5000 random draws from the height-allometry posterior (our convergence criterion required $\geq 500,000$ MCMC steps; see Appendix F, Section I). Given the new allometry vector, we assigned each individual in our light model (i.e., each focal sapling and its neighbors) a new height by drawing a random value from its height distribution, which is determined by its dbh, its species’ allometric parameters, and the variance in individual height (σ^2). Specifically, the new height is $e^a \text{dbh}^b e^{\sigma z}$, where a and b are the new values for the allometric parameters, σ is the new individual-level standard deviation, and z is a random draw from a standard normal distribution. Given the new heights for all individuals, we recalculated the joint posterior density for our model parameters (θ , which included θ_L and, in the uncertain- L growth analysis, θ_G) before proposing new values for θ . Thus, the Metropolis-Hastings accept/reject decision in our MCMC applied only to θ (conditional on the new heights), and the resulting posterior distribution of θ is its marginal distribution with respect to the posterior distributions of the individual heights.

Subsampling algorithm

We used the following algorithm to sample from the available FIA data: If ≤ 60 individuals (i.e., height measurements) were available for a species, we used all of them. If > 60 individuals were available, we sought to create a sample that contained 20 individuals in each of 18 groups representing combinations of the following categories: three dbh categories (lower, middle, and upper third of the dbh distribution within each

species) \times three soil moisture categories (individuals on plots classified as xeric, mesic, or hydric) \times two crown class categories (overtopped vs. sun-exposed). If ≤ 20 individuals were available in a group, we used all of them. If > 20 individuals were available, we accepted each individual with probability $20/n$, where n is the number of individuals in the group. The above algorithm yields about 360 individuals for a species with ≥ 20 individuals in each of the 18 group. Some states have relatively few height data for saplings (dbh 2.5-12.7 cm), so that for many species, the above algorithm produced samples containing few saplings. Therefore, for each species, if fewer than a third of the total individuals sampled by the above algorithm were saplings, we sampled additional saplings (if any were unused and available) to obtain a maximum sapling proportion of one third of the individuals per species. The dataset that we analyzed had $n \geq 30$ individuals for 260 species and 53,691 total individuals.

II. Shade-tolerance index (ST)

We define ST as the proportion of saplings (dbh 2.5-12.7 cm) of a given species that are in the understory (as opposed to the canopy). We estimated ST from FIA data based on the FIA's crown class categories (USDA 2005, 2006): we considered saplings with an overtopped or intermediate crown-class to be in the understory, and saplings with a co-dominant, dominant, or open-grown crown-class to be in the canopy. To constrain estimates of ST for rare species, we used a hierarchical approach as in the height analysis (see above). We used the same FIA dataset and plot filtering as in the height analysis above, and we used all FIA saplings with a reported crown class. (Sub-sampling the data, as in the height analysis, was not necessary due to the modest computational demands of the ST analysis.) The dataset that we analyzed had ≥ 30 saplings for 240 species and 1,215,987 total saplings.

We assumed that the number of understory saplings (y) for species i was distributed as $y_i \sim \text{binomial}(ST_i, n_i)$, where ST_i is the probability that a sapling is in the understory, and n_i is the total number of saplings of species i . For species i belonging to genus j , we assumed $\text{logit}(ST_i) \sim \text{normal}(\text{logit}(ST_{g,j}), \sigma_g^2)$, where $\text{logit}(x) \equiv \ln[x/(1-x)]$, and the subscript 'g' denotes genus-level parameters. Similarly, we assumed $\text{logit}(ST_{g,j}) \sim \text{normal}(\text{logit}(ST_{f,k}), \sigma_f^2)$, where k is the family of genus j , and the subscript 'f' denotes family-level parameters. As in the height-allometry analysis, one variance parameter was assigned for each taxonomic level. The structure of the model continued through the taxonomic hierarchy on the logit scale as illustrated above until the base of the hierarchy, where we specified a noninformative prior: $ST_{\text{base}} \sim \text{uniform}(0,1)$. As in the height-allometry analysis, we specified noninformative priors for the variances according to Gelman (2006). Specifically, we assumed uniform priors for the standard deviations, with $\sigma \sim \text{uniform}(0,10)$ at each taxonomic level. Increasing the prior range from (0,10) to (0,100) had little impact on the results.

We used WinBUGS (Spiegelhalter et al. 2004) to generate 5000 independent Bayesian posterior samples for each of three MCMC chains (see height analysis above for details). The species-level results are summarized in Supplement 1 and are compared to the *Silvics of North America* (Burns and Honkala 1990) shade-tolerance classification in figure E1. To propagate uncertainty in species-level ST through all analyses involving ST (Models CRN, CRS, and CRNS), we randomly sampled from the WinBUGS output every 100th step of our MCMC simulations, as explained above for the height analysis.

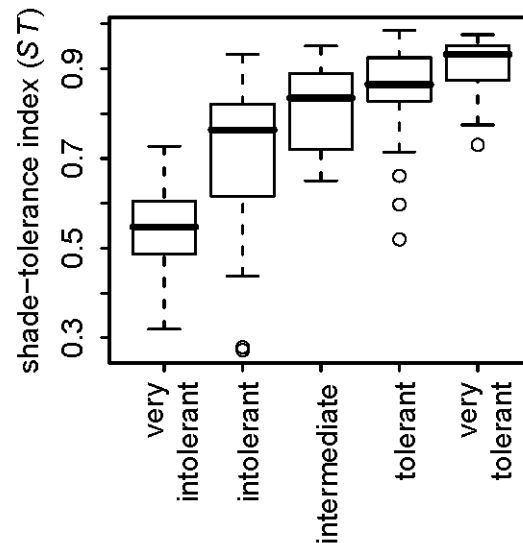


Figure E1. Correspondence between ST (species-level posterior medians) and the categorical shade-tolerance classification in *Silvics of North America* (Burns and Honkala 1990) for the 184 species included in *Silvics*. Similar results are obtained for the posterior means of ST . The figure displays standard box-plots: bold bars are medians, boxes indicate the first and third quartiles, error bars are the most extreme points within 1.5 interquartile ranges of the first and third quartiles, and circles are outliers (all points outside of the error bars).

References for Appendix E

- Brooks, S. P., and A. Gelman. 1998. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7:434-455.
- Burns, R. M., and B. H. Honkala. 1990. *Silvics of North America*: 1. Conifers; 2. Hardwoods. vol. 2. Agriculture Handbook 654. U.S. Department of Agriculture, Forest Service, Washington, DC.
- Chojnacky, D. C., and P. Rogers. 1999. Converting tree diameter measured at root collar to diameter at breast height. *Western Journal of Applied Forestry* 14:14-16.
- Clark, J. S. 2005. Why environmental scientists are becoming Bayesians. *Ecology Letters* 8:2-14.
- Flora of North America Editorial Committee, e. 1993+. *Flora of North America North of Mexico*. 12+ vols, New York and Oxford.
- Gelman, A. 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 1:515-533.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. *Bayesian data analysis*, second edition. Chapman & Hall/CRC, Boca Raton, Florida.
- 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.

- Ogle, K., and J. J. Barber. 2008. Bayesian data-model integration in plant physiological and ecosystem ecology. *Progress in Botany* 69:281-311.
- Spiegelhalter, D., A. Thomas, N. Best, and D. Lunn. 2004. WinBUGS 1.4.1.
<http://www.mrc-bsu.cam.ac.uk/bugs>.
- USDA. 2005. Forest Inventory and Analysis national core field guide. Volume 1: Field data collection procedures for phase 2 plots, version 3.0. Forest Inventory and Analysis Program, USDA, Forest Service.
- USDA. 2006. Users guide to the Forest Inventory Snapshot Database Version 2.1. Forest Inventory and Analysis Program, USDA, Forest Service.