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 F. Statistical methods.

I. Assessing MCMC convergence

We simulated three MCMC chains in parallel with the Metropolis-Hastings algorithm to generate samples from the joint posterior distribution for each model (Gilks et al. 1996). We considered the joint posterior to have converged when the marginal posteriors for all parameters of interest had converged according to 'potential scale reduction factors' (see below) (Brooks and Gelman 1998, Gelman et al. 2004). The 'parameters of interest' include all model parameters except for latent variables (Appendix F, Section III), although preliminary analyses where we monitored the marginal posterior for each latent variable showed that these also tended to converge by the time we terminated the MCMC simulation.

For each parameter, we calculated two potential scale reduction factors. The first, $\hat{R}_V$ (page 297 in Gelman et al. 2004), is based on a comparison of within-chain to pooled-chain sample means and variances. The second, $\hat{R}_{80}$, is the empirical ratio of the widths of the central 80% intervals of the pooled- and within-chain samples (Brooks and Gelman 1998). $\hat{R}_V$ and $\hat{R}_{80}$ converge to 1 as $n \to \infty$, where n is the number of posterior samples. We considered the posterior for a parameter to have converged if both $\hat{R}_V$ and $\hat{R}_{80}$ were < 1.1 (Brooks and Gelman 1998, Gelman et al. 2004).

Convergence was assessed for each parameter every 50,000 MCMC steps. To help guard against the possibility that our joint convergence criterion ($\hat{R}_V$ and $\hat{R}_{80}$ < 1.1 for all parameters) had been fulfilled before the full posterior space had been explored, we required that the criterion be met for 10 consecutive 50,000-step intervals before terminating the MCMC simulation. Due to the large number of parameters and the large number of simulation steps (> 10^6) required for some of the models to converge, the following procedure was adopted to limit the computer memory requirements: A posterior sample was saved every k MCMC steps, with $k = 1$ initially. Once 10^4 posterior samples had been saved, the first half of the samples were discarded, the remaining sample was 'thinned' by discarding every alternate sample, and k was doubled. This procedure was repeated until the convergence criterion was met for 10 consecutive 50,000-step intervals. This procedure ensured that (after the first thinning) the first half of the MCMC steps were never used to assess convergence (Brooks and Gelman 1998, Gelman et al. 2004), and that the memory requirements never exceeded 10^4 samples. Gelman et al. (2004) suggest that 10^3 samples should be adequate to assess convergence in most cases.

II. Proposal distributions

Values for all parameters were restricted to a finite range that included the biologically reasonable values. New parameter values were proposed on a modified logistic scale:

$$\tau = \ln\left(\frac{\theta - \theta_{\min}}{\theta_{\max} - \theta}\right) \tag{F1}$$

where τ is the current parameter value on the transformed scale; θ is the current parameter value on the original scale; and $\theta_{\min}$ and $\theta_{\max}$ are, respectively, the lower and upper prior bounds for θ (see Supplement 2). Because τ can take on any real number, new values τ' were assumed to be normally distributed with mean τ (i.e., the current value in the MCMC chain) and standard deviation σ_τ:

$$\tau' = \tau + \sigma_\tau z \tag{F2}$$

where z is a standard normal random variable. Typically, a new value was proposed for each parameter at each MCMC step with probability 0.5, but in some cases, groups of correlated parameters were changed in 'blocks' to speed convergence (Gilks et al. 1996). For each parameter (within each block separately), σ_τ was adjusted during an initial 'burn-in' period (prior to posterior sampling) to achieve an approximate Metropolis-Hastings acceptance rate of 0.25 (Chib and Greenberg 1995). Combining (F1) and (F2) and solving for the proposed value θ' yields:

$$\theta' = \frac{\theta_{\min} + \theta_{\max}\exp(\tau')}{1 + \exp(\tau')}. \tag{F3}$$

Specifying a noninformative prior on the transformed scale (τ) introduces prior influence on the θ scale. For example, a diffuse normal prior on τ results in relatively high prior weights for values of θ near $\theta_{\min}$ and $\theta_{\max}$ compared to intermediate values (i.e., a 'U' shaped prior). Therefore, in order to obtain a uniform prior for θ between $\theta_{\min}$ and $\theta_{\max}$, we used a transformation of variables to determine the corresponding prior on τ:

$$p(\tau) = p\big(\theta(\tau)\big)\left|\frac{d\theta(\tau)}{d\tau}\right| = p\big(\theta(\tau)\big)\frac{\big(\theta_{\max} - \theta(\tau)\big)\big(\theta(\tau) - \theta_{\min}\big)}{(\theta_{\max} - \theta_{\min})}$$

where $\theta(\tau)$ is given by (F3); and $p\big(\theta(\tau)\big)$ is the prior for θ on the original scale. With the uniform priors in our analysis, $p\big(\theta(\tau)\big)$ is simply a constant: $p(\theta) = (\theta_{\max} - \theta_{\min})^{-1}$.

III. Latent-variable integration

Here, we show that a latent-variable approach – in which we assign a free parameter (λ) for each unknown light level – can be used to effectively integrate over f_L (i.e., integrate over the uncertainty in light) in equation 12 for each growth observation at each MCMC step. Note that integrating over f_L for each growth observation at each MCMC step is conceptually distinct from simply marginalizing over the posterior distributions of the

latent variables λ. Below, however, we show that these operations are mathematically equivalent if the λ are conditionally independent. Readers who are familiar with 'latent variables' and 'marginal posteriors' may wish to skip the next two paragraphs, which provide context for readers who are unfamiliar with these concepts.

We base our inferences on the marginal posterior distributions of the model parameters, a common practice in Bayesian analysis (Gelman et al. 2004). The marginal posterior density of each parameter of interest is obtained by integrating over the densities of all other parameters; i.e., the marginal posterior density for parameter θ_i with respect to all other parameters θ_{i-} is

$$p(\theta_i \mid y) = \int p(\theta \mid y) d\theta_{i-}$$

where $p(\theta \mid y)$ is the joint density of all model parameters, θ, given the data, y, and the integral has as many dimensions as θ_{i-}. For complex models that cannot be solved analytically, samples from the joint posterior density can be generated with MCMC methods (Gilks et al. 1996). The properties of $p(\theta_i \mid y)$ (e.g., its percentiles) are then quantified from the values of θ_i in the samples without regard to θ_{i-}; i.e., after generating samples from $p(\theta \mid y)$, marginalizing over θ_{i-} amounts to nothing more than making a histogram of the values of θ_i.

The above 'automatic' integration that occurs with MCMC makes it relatively simple to integrate over random effects by introducing a latent variable, or 'nuisance parameter,' for each effect (pages 73-4 in Gelman et al. 2004). Thus, in our 'uncertain-light' growth analysis, we can introduce a latent variable (λ_i) for each unobserved light level (L_i) and integrate over the uncertainty in L by marginalizing over the posteriors for λ (the vector of latent variables). Each λ_i is treated in the MCMC as a free parameter, and new values are proposed on the interval (0,1), the upper limit corresponding to full sunlight. The model now has more parameters than data, but this poses no problem in principle because each λ_i is constrained by its sampling distribution, $f_L(\lambda_i \mid x_i, \theta_L)$ (the conditional probability density function for light given covariates), which acts as an informative prior. Although this latent-variable approach (which is an application of Bayesian hierarchical modeling) involves many more parameters than the 'direct' integration approach (i.e., numerically integrating equation 12 for each growth observation at each MCMC step), it is computationally more efficient. Because we marginalize over the joint posterior density of λ, we do not need to estimate a posterior for each λ_i (which is computationally demanding, although still less so than direct integration); i.e., we simply calculate the marginal posteriors for the other parameters of interest, ignoring λ.

We now show that marginalizing over λ (i.e., treating the λ_i as free parameters in the MCMC, but ignoring them when making posterior inferences for the parameters of interest) is equivalent to integrating over f_L in equation 12 for each growth observation at each MCMC step. For simplicity, we first demonstrate this fact assuming that θ_L (the parameters that determine f_L) is known. We then generalize the result to the case where θ_L is unknown.

The marginal posterior density for the growth parameters of interest, θ_G, is

$$p(\theta_G \mid G, X_G, \theta_L) = \int \cdots \int \cdots \int p(\theta_G, \lambda_1, \cdots, \lambda_i, \cdots, \lambda_{N_G} \mid G, X_G, \theta_L) d\lambda_1 \cdots d\lambda_i \cdots d\lambda_{N_G}$$

where G is the vector of observed growth rates; X_G is the matrix of covariates with N_G rows (one for each element of G); and the integrand is the joint posterior density of θ_G and λ, which we can write as

$$\begin{aligned} p(\theta_G, \lambda \mid G, X_G, \theta_L) &\propto p(G \mid \lambda, \theta_G) \cdot f_L(\lambda \mid X_G, \theta_L) \cdot p(\theta_G) \\ &= \left[\prod_i^{N_G} p(G_i \mid \lambda_i, \theta_G) \cdot f_L(\lambda_i \mid x_i, \theta_L) \right] \cdot p(\theta_G) \end{aligned}$$

where $p(G \mid \lambda, \theta_G)$ is the likelihood of the growth data, given λ; $f_L(\lambda \mid X_G, \theta_L)$ is the sampling distribution (prior) for λ; and $p(\theta_G)$ is the prior for θ_G. We now write the marginal posterior of θ_G as

$$\begin{aligned} p(\theta_G \mid G, X_G, \theta_L) &\propto \int \cdots \int \cdots \int \prod_i^{N_G} p(G_i \mid \lambda_i, \theta_G) \cdot f_L(\lambda_i \mid x_i, \theta_L) \cdot p(\theta_G) d\lambda_1 \cdots d\lambda_i \cdots d\lambda_{N_G} \\ &= \left[\prod_i^{N_G} \int p(G_i \mid \lambda_i, \theta_G) \cdot f_L(\lambda_i \mid x_i, \theta_L) d\lambda_i \right] \cdot p(\theta_G) . \end{aligned}$$

Above, we can move the integration inside the product because the latent variables are assumed to be conditionally independent. Also, $p(\theta_G)$ does not depend on λ and, therefore, can be moved outside of the integrals. The above shows, assuming known θ_L, that marginalizing over λ is equivalent to integrating over f_L for each growth observation. Because, by definition, an MCMC algorithm samples from its target distribution (Gilks et al. 1996), the latent-variable approach is equivalent to integrating over f_L for each growth observation at each MCMC step.

To generalize the above result to the case with unknown θ_L, we write the joint posterior density of θ_G and θ_L, marginalized over λ, as

$$\begin{aligned} &p(\theta_G, \theta_L \mid G, L, X) = \int \cdots \int \cdots \int p(\theta_G, \theta_L, \lambda_1, \cdots, \lambda_i, \cdots, \lambda_{N_G} \mid G, L, X) d\lambda_1 \cdots d\lambda_i \cdots d\lambda_{N_G} \\ &\propto \int \cdots \int \cdots \int \prod_i^{N_G} p(G_i \mid \lambda_i, \theta_G) \cdot f_L(\lambda_i \mid x_i, \theta_L) d\lambda_1 \cdots d\lambda_i \cdots d\lambda_{N_G} \\ &\cdot \left[\prod_i^{N_L} f_L(L_i \mid x_i, \theta_L) \right] \cdot p(\theta_G, \theta_L) \\ &= \left[\prod_i^{N_G} \int p(G_i \mid \lambda_i, \theta_G) \cdot f_L(\lambda_i \mid x_i, \theta_L) d\lambda_i \right] \cdot \left[\prod_i^{N_L} f_L(L_i \mid x_i, \theta_L) \right] \cdot p(\theta_G, \theta_L) \end{aligned}$$

where L is the vector of N_L light observations; and X is the matrix of covariates with N_L rows (one for each element in L, including the N_G saplings with growth data). Note that $\prod_i f_L(L_i \mid x_i, \theta_L)$, the likelihood for the light observations (i.e., the calibration data), and

$p(\theta_G, \theta_L)$, the joint prior, are independent of λ and, therefore, can be moved outside of the integrals.

IV. Direct integration

As a check on the above latent-variable integration approach, we also performed the uncertain-light analysis (table 3) with a direct approach by numerically integrating (12) for each growth observation at each MCMC step. We used the following simple, albeit inefficient, algorithm: We discretized the light distribution, f_L, into 50 equally-spaced intervals: L = 0.0-0.02, 0.02-0.04, …, 0.98-1.0. We then approximated (12) as

$$p(G_i \mid x_i, \theta_L, \theta_G, Q) \approx \sum_{j=1}^{50} p(G_i \mid L_j, \theta_G, Q) \times f_L(L_j \mid x_i, \theta_L) / C$$

where L_j is the midpoint of light-interval j, and $C = \sum_{j=1}^{50} f_L(L_j \mid x_i, \theta_L)$ is a normalization constant that depends on θ_L due to errors associated with the discretization of f_L.

V. Stand quality (Q) random effects

A latent variable ('nuisance parameter') was introduced for the random effect Q for each stand. These latent variables were treated like all other parameters in the MCMC algorithm (e.g., Appendix F, Section II). Marginalizing over the posterior distributions of these latent variables is equivalent to integrating over the random effects (Appendix F, Section III). The sampling distribution for Q was assumed independent and normal with mean zero. Because r and Q only appear as a product (see equation 10), one of their variances must be fixed (otherwise, the model is underdetermined). We assumed a variance of one for Q, although any other positive constant would lead to equivalent results. To prevent stand effects from being confounded with region, we imposed the constraint that within each region (ENA or WOR), Q sums to zero. Furthermore, because > 99% of the individuals at the Metolius River sites (Appendix A) belonged to a single species (*Pinus ponderosa*) that was absent from all other stands (i.e., species was confounded with site), we set Q = 0 for these stands.

VI. Predicted values

We explored two different approaches to calculating prediction vectors, $\tilde{y}$, corresponding either to vectors of predicted light, $\tilde{L}$, or predicted growth, $\tilde{G}$. (*i*) Method I involved calculating the expectation (mean) of y given θ^{50}, the posterior medians of the model parameters:

$$\tilde{y} = \bar{y} \mid \theta^{50}. \qquad \text{(F3)}$$

For $\tilde{L}$, this involved calculating $\bar{\rho} \mid \theta_L^{50}$ and $V_\rho \mid \theta_L^{50}$ from equations 3 and 5, and then calculating $\bar{L} \mid (\bar{\rho}, V_\rho)$ from equation 2. In analyses involving heights (Models CRN and CRNS; table 2), we used the expected heights calculated from the posterior median allometric parameters (Appendix E and Supplement 1) when calculating $\tilde{L}$. For $\tilde{G}$,

equation F3 is obtained directly from equation 10 using the posterior medians of θ_G and the posterior medians of the latent variables Q (Appendix F, Section V). As an alternative to θ^{50}, we could use $\hat{\theta}$, the posterior means, as point estimates for θ. These two point estimates yielded very similar $\tilde{y}$, because in our analyses, θ^{50} and $\hat{\theta}$ were very similar. (*ii*) Method II involved calculating the posterior mean of the expectation of y; i.e., for each posterior sample j from the MCMC simulation, we calculated the expectation $\bar{y}_j$ given the current parameter values (θ_j) and then calculated

$$\tilde{y} = (1/n)\sum_j \bar{y}_j , \qquad \text{(F4)}$$

where n is the total number of posterior samples. In principle, it might be preferable to obtain a posterior distribution for $\bar{y}$, and then to take its median or mode as $\tilde{y}$. This is cumbersome, however, because it entails numerical estimation of one distribution for each individual (over 2000 in our analysis, or perhaps > 10^5 in an analysis of inventory data). In contrast, it is computationally 'cheap' to monitor $\sum_j \bar{y}_j$ for each individual and then apply (F4). In preliminary analysis of a subset of our data in which we estimated posterior distributions for each individual prediction, the medians of these distributions and the means from (F4) were nearly identical to each other.

Unlike Method I, Method II averages over $p(\theta)$ in the appropriate way; i.e., Method II propagates uncertainty in θ. Note, however, the large difference between the methods in ease of implementation: Method I involves inserting numbers into equations, whereas Method II involves MCMC simulation. We used Method I for $\tilde{L}$ in the predicted-L method for estimating θ_G (table 3), because these are the predictions that would be most practical to implement when analyzing large inventory datasets. In all other cases (e.g., R^2 values in table 4), we used Method II. The decision to use Method I vs. II has little impact on our results, however, as both methods yield very similar predictions for our analyses.

References for Appendix F

Brooks, S. P., and A. Gelman. 1998. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7:434-455.

Chib, S., and E. Greenberg. 1995. Understanding the Metropolis-Hastings algorithm. American Statistician 49:327-335.

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.