# Predictive Posterior Distributions from a Bayesian Version of a Slash Pine Yield Model

Edwin J. Green and William E. Strawderman

**ABSTRACT:** We formulate a traditional slash pine diameter distribution yield model in a Bayesian framework. We attempt to introduce as few new assumptions as possible. We generate predictive posterior samples for a number of stand variables using the Gibbs sampler. The means of the samples agree well with the predictions from the published model. In addition, our model delivers distributions of outcomes, from which it is easy to establish measures of uncertainty, e.g., Bayesian credible regions. For. Sci. 42(4):456–464.
**Additional Key Words:** Gibbs sampler, Weibull distribution.



In this report, we detail the results of a study undertaken to produce a Bayesian version of an existing diameter distribution yield model for slash pine plantations. Our motivation arises from the fact that most, if not all, published yield models deliver only point estimates for a given set of input parameters, i.e., no statement of expected variability is provided. This may be viewed as a major limitation of such models. In contrast, a Bayesian model would deliver predictive posterior densities which could then be summarized by any statistic(s) the user desired. As an alternative to the Bayesian approach, one could employ error propagation techniques (e.g., see Gertner and Guan 1991). However, these procedures are approximate and tedious to develop. In addition, error propagation is problematic when some or all of the distributions involved are not normal, while the Bayesian approach described herein is amenable to nearly all distributional forms.

We chose to develop a Bayesian version of the model published by Zarnoch et al. (1991), which describes the yield of slash pine plantations in the West Gulf region of the United States. Our goal is to put the model in a Bayesian framework and produce marginal predictive posterior distributions of the quantities of interest, *not* to evaluate the suitability of the original model. We will clearly identify any area where we deviate from the published model. We want to produce a Bayesian version of Zarnoch et al.'s model while introducing as few new assumptions as possible. We implicitly accept the equations built by Zarnoch et al., following much modeling effort, as true.

## Predictive Posterior Distributions

The merits of Bayesian inference have been reported on at length in countless articles, as have its supposed shortcomings. The interested reader is referred to Chapter 10 in Robert 1994 as a good starting place. For some forestry examples, see Burk and Ek 1982, Green et al. 1994, Green and Strawderman 1992, Green et al. 1992, and Swindel 1972. The feature of Bayesian inference that is relevant to this study is the provision for predictive posterior distributions. These are proper probability distributions which summarize all the available information about a quantity which is to be predicted, and can be used to determine the probability that the measured value of the quantity will fall in any specified interval. Note the essential difference between intervals established based on predictive posterior distributions and the commonly used confidence intervals. It is well known that the latter are not proper probability intervals, and thus require one to step out of the probability calculus and accept the vague notion of *confidence*. The foregoing seems to us to be a powerful reason, albeit not the only one, for adopting the Bayesian viewpoint.

## Description of Original Model

The Zarnoch et al. model (hereafter simply referred to as the published model) is a probability density function (pdf) based model. It is based on predicting the distribution of tree diameters in a plantation with specified attributes. The relative frequency function of diameters is

Edwin J. Green is with the Department of Natural Resources and William E. Strawderman is with the Department of Statistics, both from Rutgers University, New Brunswick, NJ 08903-0231. Paper of the Journal Series, NJ Agricultural Experiment Station, Cook College, Rutgers University, New Brunswick, NJ. This work was funded by NJAES Project Number D-17305-96, supported by NJAES and U.S. McIntire-Stennis Act funds.

Manuscript received January 10, 1995. Accepted November 17, 1995. 

assumed to be adequately characterized by a Weibull density. This model is one of a class of models commonly referred to as *parameter-recovery* models because the parameters of the pdf are obtained by predicting functions of diameter, and then solving for the parameters.

Simultaneously, one predicts the surviving number of trees per acre $(N)$ in the plantation. Integration of the Weibull density over diameter classes of interest yields the percentage of trees in the those classes. Multiplication by $N$ yields an estimate of the number of trees in those classes.

A height model is supplied which predicts tree height from diameter and associated plantation attributes. Finally, a volume equation is given which predicts tree volume as a function of diameter and height.

For each diameter class of interest, one may predict the height and volume of a tree with, say, the midpoint diameter of the class. Multiplication of the volume of the tree with the midpoint diameter by the number of trees in the class yields an estimate of the volume in the class. Summing over classes of interest yields an estimate of the total volume in the desired interval.

## Equations

Below we present the equations given in the original model. Note that the equations are of two types: plantation or stand level, and individual tree level.

### *Plantation-Level Models*

$$B = \beta_{11} H_D^{\beta_{12}} N^{\beta_{13}} \exp(\beta_{14} A^{-1}) \tag{1}$$

$$D_{0.93} = \beta_{21} H_D^{\beta_{22}} N^{\beta_{23}} \exp(\beta_{24} A^{-1}) \tag{2}$$

$$D_{\min} = \beta_{31} H_D^{\beta_{32}} N^{\beta_{33}} \exp(\beta_{34} A^{-1}) \tag{3}$$

$$N = T^{1-\beta_{41} A} \tag{4}$$

$$H_D = S \exp\{\beta_{52}(A^{-0.5} - 0.2)\} \tag{5}$$

where

$B$ = basal area ($m^2$/ha),

$H_D$ = height of dominant and/or codominant trees in plantation (m),

$D_{\min}$ = minimum diameter in plantation (cm),

$D_{0.93}$ = 93rd percentile of diameter distribution (cm),

$A$ = age of plantation (yr),

$T$ = number of trees planted per hectare,

$S$ = site index (m), base age 25.

In the published model, (4) is only one of three possible survival curves. We chose (4) because it was the one which related $N$ to $T$. Equation (5) is derived by fitting the model $\ln H_D = \beta_{51} + \beta_{52} A^{-0.5}$ and then imposing the constraint $\ln S = \beta_{51} + \beta_{52}(25)^{-0.5}$ and solving for $\beta_{51}$ in terms of $S$ and $\beta_{52}$.

### *Tree-Level Models*

$$H = \theta_{11} H_D^{\theta_{12}} B^{\theta_{13}} N^{\theta_{14}} \exp(\theta_{15} + \theta_{16} D) \tag{6}$$

$$V = \theta_{21} + \theta_{22} D^2 H + \theta_{23} (D^2 H)^2 \tag{7}$$

where

$H$ = individual tree height (m)

$D$ = individual tree diameter (cm)

$V$ = individual tree volume ($m^3$)

### *Model Operation*

The model operates as follows: the user specifies $A$, $S$, and $T$. Equation (5) is invoked to predict $H_D$. Then, using Equations (1)–(3), predictions for $B$, $D_{\min}$, and $D_{93}$ are obtained. Recall it is assumed that the diameter distribution follows the Weibull density

$$f(D) = \frac{c}{b}\left[\frac{D-a}{b}\right]^{c-1} \exp\left[\frac{D-a}{b}\right]^{c}, D \geq a, a \geq 0, b > 0, c > 0 \tag{8}$$

The three parameters of the Weibull are calculated deterministically from $B$, $D_{\min}$, and $D_{0.93}$.

$$a = D_{\min}/2, \tag{9}$$

$c$ is the solution to

$$a^2 + 2a(D_{0.93} - a)\frac{\Gamma(1+\frac{1}{c})}{-(\ln(1-p))^{1/c}} + (D_{0.93} - a)^2 \frac{\Gamma(1+\frac{2}{c})}{-(\ln(1-p))^{2/c}} - \frac{B}{kN} = 0 \tag{10}$$

and

$$b = \frac{D_{93} - a}{-(\ln(1-p))^{1/c}} \qquad (11)$$

where $p = 0.93$ and $k = 0.0000785$.

The percentage of trees between diameters $D_1$ and $D_2$ ($D_2 > D_1$) is found by $F(D_2 \mid a,b,c) - F(D_1 \mid a,b,c)$, where $F$ denotes the cumulative distribution function of the Weibull density, which exists in closed form.

The published model also includes a provision for predicting the yield of thinned stands. For simplicity, we will only be concerned with unthinned stands here.

## Data

We used the same data as Zarnoch et al. (1991), and interested readers are referred there. We only provide a brief description here. The data come from 507 unthinned plots in slash pine plantations on cutover lands. The plots varied in size from 0.1 to 0.25 ac (0.04 to 0.10 ha). Among the variables measured on each plot were height of dominants and diameter and height of each tree. This resulted in 6,326 individual tree observations. Age and number of trees planted were known from plantation records. It is worth mentioning that observations on individual trees from the same plot are clearly not independent. However, they were treated as such in the published model, and hence we also adopt this assumption.[1]

## Bayes Model

### *Overview*

We generate a joint predictive posterior sample for ($B$, $N$, $D_{min}$, $D_{0.93}$, and $H_D$) given the model inputs $A$, $T$, and $S$. Then, for each observation in the posterior sample, we solve for $(a,b,c)$ using (9)–(11). We assume that (9)–(11) are deterministic, as was done in the published model [note that we do not endorse Equations (9)–(11). Indeed, we believe that there is compelling evidence that (9) in particular is suspect (Green et al. 1994)]. Let $N_j$ be the value of $N$ for the $j$th observation in the posterior sample. We generate $N_j$ independent random numbers (diameters) from the Weibull distribution derived from the $j$th observation. Then for each diameter, we stochastically generate a height and a volume. The volumes are then summed over the $N_j$ trees to yield an estimate of total volume ($TV$). Repeating this for every observation in the posterior sample results in a predictive posterior sample for $TV$.

We used the Gibbs sampler to generate the predictive posterior sample for ($B$, $N$, $D_{min}$, $D_{0.93}$, and $H_D$). Gibbs sampling was popularized by Gelfand and Smith (1990) and has since become a standard method for performing heretofore intractable Bayesian analyses. For details on Gibbs sampling see the above paper, and also Gelfand et al. (1990), Smith and Roberts (1993), and George and Casella (1992), among others. For examples of Gibbs sampling in forestry problems see Green and Strawderman (1992) and Green et al. (1994). To use the Gibbs sampler, one must be able to sample from the *full* conditional distribution for each parameter. In the present context, we need the following distributions [we borrow the Gelfand and Smith (1990) bracket notation to denote pdf's)]:

- $[B \mid \ldots]$
- $[N \mid \ldots]$
- $[D_{min} \mid \ldots]$
- $[D_{0.93} \mid \ldots]$
- $[H_D \mid \ldots]$

where "..." is to be understood as "all the other variables in the model."

In the published model, *no* distributional assumptions were presented for $B$, $N$, $D_{min}$, $D_{0.93}$ and $H_D$. In accord with our stated goal of introducing as few assumptions as possible, we avoid specifying error distributions for Equations (1)–(5), and we also do not impose prior distributions. Instead, we follow Zellner's (1994) Bayesian Method of Moments/Instrumental Variable (BMOM/IV) approach We now summarize the relevant concepts of the BMOM/IV method.

### *BMOM/IV*

The BMOM/IV method is predicated on conditioning the first two moments of the predictive posterior to equal the moments observed in the sample data. Following this, *maximum entropy* or *maxent* (e.g., see Jaynes 1982) is used to find the most conservative choice of density that agrees with the side conditions on the moments. Zellner presents results for the scalar mean problem, and the regression problem. Our interest is in the latter. Suppose we entertain the following model:

$$Y = \mathbf{X}\alpha + u \qquad (12)$$

where $Y$ is an $n \times 1$ vector of observations, $\mathbf{X}$ is an $n \times p$ matrix of regressor variables, $\alpha$ is a $p \times 1$ vector of regression coefficients with unknown values, and $u$ is an $n \times 1$ vector of realized (but unknown) error terms. Premultiplication of both sides of (12) by $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and taking posterior expectations yields[2]

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y = \mathbf{E}\alpha + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{E}u$$

Now we impose the conditions on the moments. Specifically we assume

1. $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}\mathbf{E}u = 0$. This leads directly to $\mathbf{E}\alpha = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y = \hat{\alpha}$.
2. $\mathbf{E}(u - \hat{u})(u - \hat{u})' = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\tau^2$ where $\hat{u} = Y - X\hat{\alpha}$, and $\tau^2$ is a positive constant.

[1] This is a common assumption in growth and yield modeling, and is not without merit (Green et al. 1994).

[2] After $Y$ has been observed, it is known without error. Hence the posterior expectation of $Y$ is $Y$.

Using the above assumptions, Zellner shows

1. The proper maxent posterior density for $\alpha$, given $\tau^2$ and the data is a multivariate normal density with mean $\hat{\alpha}$, and covariance matrix $(\mathbf{X}'\mathbf{X})^{-1}\tau^2$.
2. The proper maxent posterior density for $\tau^2$ with $\mathbf{E}\tau^2 = s^2 = \hat{u}'\hat{u}/(n-p)$ is the exponential density $g_e(\tau^2 \mid \mathbf{X}, Y, s^2) = (1/s^2)\exp(-\tau^2/s^2), 0 < \tau^2 < \infty$.
3. If we denote a future observation $y_f = x_f\alpha + u_f$, then the proper maxent predictive density for $y_f$ given $\tau^2, x_f, \mathbf{X}$, and $Y$ is a normal density $h_N(y_f \mid \tau^2, x_f, \mathbf{X}, Y) \sim \mathbf{N}[\hat{y}_f, (1 + x_f(\mathbf{X}'\mathbf{X})^{-1}x_f')\tau^2)]$, where $\hat{y}_f = x_f\hat{\alpha}$.

Clearly, the means and variances of the above distributions are similar to and/or exactly equal to the usual least squares estimates. Yet notice that here they are derived without specification of a likelihood function. Furthermore, when viewed as predictive posterior distributions, they permit access to the Bayesian inference machine, without refernce to a likelihood or prior distribution.

## Modifications to Published Model

We make two substantive modifications to the published model. First, we linearize models (1)–(6) by taking logarithms. This permits use of the BMOM/IV method. Note that we now write model (4) as

$$(\ln T - \ln N)/\ln T = \beta_{41} A$$

The second modification arises from our reluctance to accept the implicit assumption in the published model that $B$, $D_{\min}$, and $D_{0.93}$ are independent. We "build in" interdependence among these variables by altering Equations (1)–(3) in a hierarchical fashion to:

$$\ln B = \ln\beta_{11} + \beta_{12}\ln H_D + \beta_{13}\ln N + \beta_{14}A^{-1} \quad (13)$$

$$\ln D_{0.93} = \ln\beta_{21} + \beta_{22}\ln H_D + \beta_{23}\ln N + \beta_{24}A^{-1} + \beta_{25}\ln B \quad (14)$$

$$\ln D_{\min} = \ln\beta_{31} + \beta_{32}\ln H_D + \beta_{33}\ln N + \beta_{34}A^{-1} + \beta_{35}\ln B + \beta_{36}\ln D_{0.93} \quad (15)$$

The forms of (13)–(15) were selected after brief experimentation. The PRESS statistic (e.g., see Green 1983) was used to evaluate model forms. A more thorough investigation might well result in different model forms. As stated earlier, our intent is to modify the published model as little as possible, while putting it in a Bayesian framework. In the published model, no error terms were included, although standard errors from the fitting process were reported. We will explicitly include error terms in the model. Specifically, we will assume the RHS of each model should be augmented (after linearization) by "+ $u_i$," $i = 1,2,\ldots,5$. The list of full conditionals from which we must generate samples is then augmented by the following five distributions: $[\sigma_i^2 \mid \ldots], i = 1,2,\ldots,5$, where $\sigma_i^2$ is the error variance for model $i$.

## Implementation of Gibbs Sampler

The full conditional distributions are presented in the Appendix. Sampling from each of these distributions was accomplished with the ratio-of-uniforms method (Wakefield et al. 1991). We simulated $\ln\sigma_i^2$ and logit [(ln T–ln $N$)/ln $T$] since ratio-of-uniforms works best for variables defined over the whole real line.

For each observation in the sample, $N$ diameters were generated as mentioned above. Heights and then volumes were stochastically generated for each diameter using the OLS estimates for $\underline{\theta}_i, i = 1,2$, where $\underline{\theta}_i$ is the vector of parameters for equation $i$, and the predictive variances MSE(1 + $z(Z'Z)^{-1}z'$)), with $Z$ the appropriate design matrix, and $z$ the vector of independent variables for which a prediction is sought. Here we introduce the mild assumption that the error terms for models (6) and (7) are normally distributed. Collecting the $N$ volumes yielded an observation on total volume.

When running a Gibbs sampler, it is now common to run one chain (i.e., from one set of starting values) for a long time until convergence is attained. The preliminary iterations are then discarded. However, diagnosing convergence is problematic. For discussion on convergence issues, see Gelman and Rubin (1992), Geyer (1992), Geweke (1992), Roberts (1992), Smith and Roberts (1993), and Zellner and Min (1994), among others. Here we used a diagnostic approach developed by Raftery and Lewis (1992a, 1992b, 1995). In the Raftery and Lewis algorithm, one must first specify which quantile $q$ of the posterior distribution is of interest. One must specify two additional quantities $r$ and $s$, whose definitions are best defined by example. Suppose we wish to estimate $q$ to within $\pm r$ with probability $s$. For instance, we may specify $q = 0.025$, $r = 0.0125$, $s = 0.95$$. This amounts to requiring that the 0.025 quantile be estimated to within ±0.0125 with probability 0.95, which is roughly equivalent to requiring that nominal 95% intervals have actual posterior probability between 0.925 and 0.975. Raftery and Lewis provide an algorithm which will, among other things, determine the number of iterations required for convergence of the Gibbs sampler, given the specified values for $q$, $r$, and $s$. Their algorithm is coded in the FORTRAN program gibbsit which is freely available from StatLib. We used gibbsit in our study.

We simulated the yields from two hypothetical slash pine plantations; one with $A = 20$, $T = 1483$, $S = 18$ (equivalent to 600 trees/ac and a site index of 60 ft in english units) and one with $A = 15$, $T = 1977$, $S = 15$ (800 and 50 in english units). The gibbsit parameters were $q = 0.025$, $r = 0.0125$, $s = 0.95$.

For each plantation we used gibbsit to determine the number of iterations required for convergence for each of the 10 variables in the Gibbs sampler (see Appendix). In no case was the required number of iterations >8,000 and in most cases the number was <1,000. In any event, we ran the sampler for 50,000 iterations for each plantation. Given the recommended number of iterations from gibbsit, ours was a very conservative procedure. In order to make the posterior distributions manageable, we discarded the first 30,000 observations, and then accepted every 40th iteration thereafter. This resulted in posterior distributions of size 500. This is wasteful of information and would not be recommended in practice. Also, we did not subsample the output in order to produce *iid* samples, although that is the effective result.

## Results

Once samples are obtained from the marginal predictive posterior distributions for the desired variables, either the means or modes of these samples are sensible point estimates. We chose the means, primarily due to computational ease. The means of the samples seem to agree well with the point estimates available from the published model. In Table 1 we present results for both plantations. We also present approximate 95% credible regions for the Bayes estimates. These were obtained by simply recording the 0.025 and 0.975 percentiles of each marginal predictive posterior sample.

Marginal posterior densities are also available from the Bayes model. In Figures 1 and 2 we display the marginal posteriors for $B$, $N$, $H_D$, and $TV$ for plantations 1 and 2, respectively. These histograms indicate substantial variability about the mean or mode for $B$, $N$, and $TV$ (there is much less variablity in $H_D$ because this variable is constrained to be equal to site index at age 25). The information in the histograms, as well as the 95% credible regions, is *unavailable* in the original model.

## Conclusions

It is difficult to evaluate the variability implied in our model (Figures 1 and 2). To do so, we would need large numbers of plots on the same site quality land, planted at the same density, and currently at the same age. Such data are rare at best. In any event, it is apparent that the point estimates available from growth and yield models are subject to substantial error. Our methodology provides a simple method for constructing predictive posterior distributions which appear reasonable. In addition, our results reaffirm the danger in accepting point estimates without some measure of variability. As mentioned in the introduction, we implicitly accept the models constructed by Zarnoch et al. (1994) as true, apart from the modifications noted. The major modification was the linearization of equations (1)–(6), done to permit use of Zellner's BMOM/IV method. Should linearization prove unacceptable, one could still use our general methodology, but one would then have to specify the likelihoods and prior distributions. Our results are conditional on the model forms specified. In the event the models are badly misspecified, then perhaps more modeling effort is called for.

## Literature Cited


Burk, T.E., and A.R. Ek. 1982. Application of empirical Bayes/James-Stein procedures to simultaneous estimation problems in forest inventory. For Sci. 28:753–771.

Gelfand, A.E., and A.F.M. Smith. 1990. Sampling-based approaches to calculating marginal densities. J. Am. Stat. Assoc. 85:398–409.

Gelfand, A.E., S.E. Hills, A. Racine-Poon, and A.F.M. Smith. 1990 Illustration of Bayesian inference in normal data models using Gibbs sampling. J. Am. Stat. Assoc. 85:972–985.

Gelman, A., and D. Rubin. 1992. Inference form iterative simulation using multiple sequences. Stat. Sci. 7:457–472.

George, E.I., and G. Casella. 1992. Explaining the Gibbs sampler. Am. Stat 46:167–174.

Gertner, G.Z., and B. Guan. 1991. Using an error budget to evaluate the importance of component models within a large scale simulation model *In* Proc. of Conf. on Mathematical modelling of forest ecosystems.

Geyer, C.J. 1992. Practical Markov chain Monte Carlo. Stat. Sci.7:473–482

Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. P. 169–194 *in* Bayesian Statistics 4, Proc. of 4th Valencia Internat. Meet., Berger et al. (eds.). Oxford University Press.

Green, E.J. 1983. Evaluating the predictive abilities of regressions with PRESS. For. Sci. 29:712–714.


**Table 1. Predicted values for growth and yield model variables from published model and Bayes model for plantation 1 ($A = 20$, $T = 1483$, $S = 18$) and plantation 2 ($A = 15$, $T = 1977$, $S = 15$). Lower and upper 95% credible bounds for Bayes estimates in parentheses.**

| | Plantation 1 | | Plantation 2 | |
|---|---|---|---|---|
| Variable | Published | Bayes | Published | Bayes |
| $N$ | 957.7 | 973.3 (678.3, 1,336) | 1,405.5 | 1,393.1 (907.6, 1,829.8) |
| $H_D$ | 15.6 | 15.9 (15.7, 15.9) | 10.3 | 10.6 (10.4, 10.5) |
| $B$ | 22.3 | 22.5 (16.4, 28.8) | 16.7 | 16.6 (11.8, 20.9) |
| $D_{min}$ | 6.8 | 7.4 (3, 14.2) | 3.6 | 4.1 (1.5, 8.4) |
| $D_{0.93}$ | 22.4 | 22.4 (20.1, 25.6) | 16 | 16.2 (14.2, 18.3) |
| $TV$ | 177 | 182 (134, 236) | 91 | 100 (70, 129) |
| $a$ | 1.3 | 1.4 (0.6, 2.8) | 0.7 | 0.8 (0.3, 1.7) |
| $b$ | 5.8 | 5.7 (4.6, 6.7) | 4.4 | 4.3 (3.4, 4.9) |
| $c$ | 4 | 4 (2.8, 5.6) | 4.2 | 4 (2.8, 6.1) |

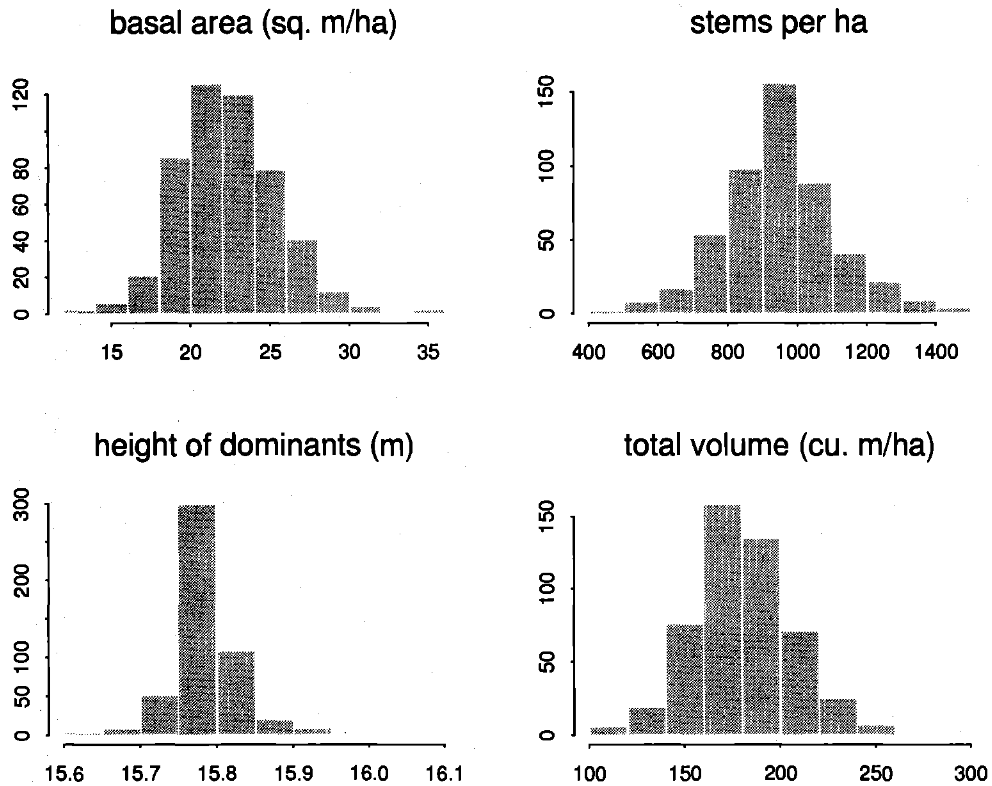


**Figure 1. Histograms from marginal posterior samples of size 500 for *B, N, HD,* and *TV* for stand 1 (*A* = 20, *T* = 1483, *S* = 18).**

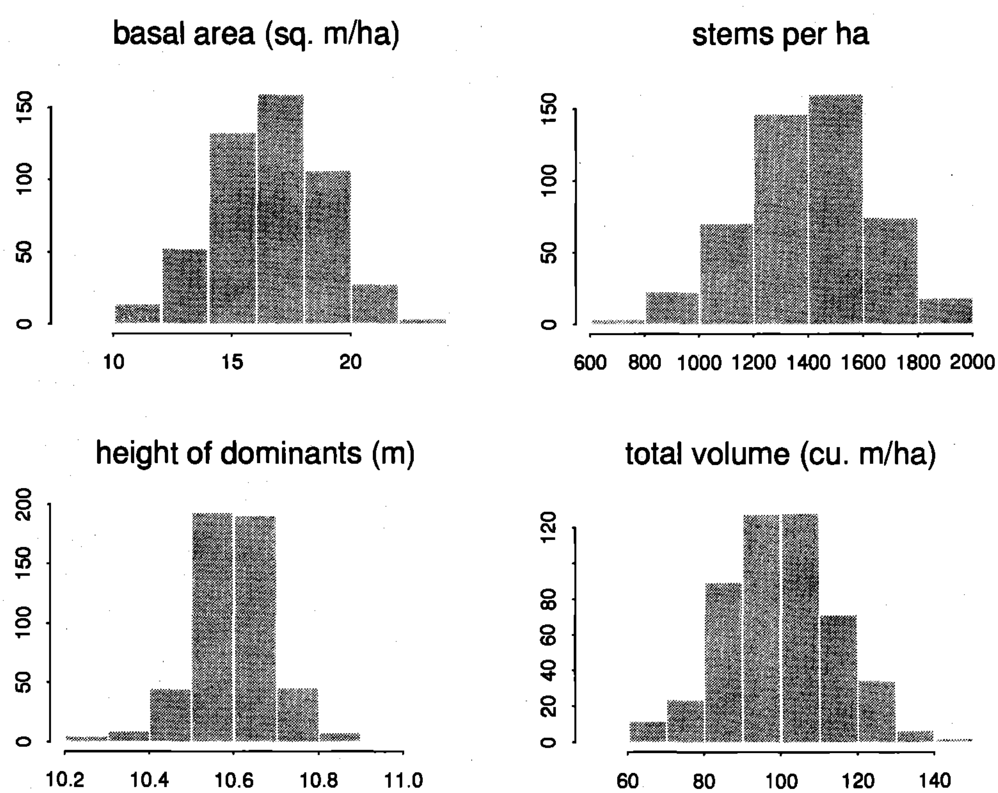


**Figure 2. Histograms from marginal posterior samples of size 500 for *B, N, HD,* and *TV* for stand 2 (*A* = 15, *T* = 1977, *S* = 15).**

GREEN, E.J., AND W.E. STRAWDERMAN. 1992. A comparison of hierarchical Bayes and empirical Bayes methods with a forestry application. For. Sci. 38:350–366.

GREEN, E.J., W.E. STRAWDERMAN, AND C.E. THOMAS. 1992. Empirical Bayes development of Honduran pine yield models. For. Sci. 38:21–33.

GREEN, E.J., F.A. ROESCH, JR., A.F.M. SMITH, AND W.E. STRAWDERMAN. 1994. Bayesian estimation of the three-parameter Weibull distribution with tree diameter data. Biometrics 50:254–269.

JAYNES, E.T. 1982. On the rationale of maximum-entropy methods. Proc. IEEE 70:939–952.

RAFTERY, A.E., AND S.M. LEWIS. 1992a. How many iterations in the Gibbs sampler? Bayesian Statistics 4. P. 763–774 *in* Proc. of 4th Valencia Internat. Meet., Berger et al. (eds.). Oxford University Press.

RAFTERY, A.E., AND S.M. LEWIS. 1992b. One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo. Stat. Sci. 7:493–498.

RAFTERY, A.E., AND S.M. LEWIS. 1995. Implementing MCMC. P. 115–130 *in* Markov chain Monte Carlo in practice. W.R. Gilks et al. (eds). Chapman and Hall, New York.

ROBERT, C.P. The Bayesian choice. Springer-Verlag, New York. 436 p.

ROBERTS, G.O. 1992. Convergence diagnostics of the Gibbs sampler. Bayesian Statistics 4. P. 775–782 *in* Proc. of 4th Valencia Internat. Meet , Berger et al. (eds.). Oxford University Press.

SMITH, A.F.M., AND G.O. ROBERTS. 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo. J. Roy. Stat. Soc. B:3–23.

SWINDEL, B.F. 1972. The Bayesian controversy. USDA For. Serv. Res. Pap SE-95. 12 p.

WAKEFIELD, J.C., A.E. GELFAND, AND A.F.M. SMITH. 1991. Efficient generation of random variates via the ratio-of-uniforms method. Stat. Comput 1:129–133.

ZARNOCH, S.J., D.P. FEDUCCIA, V.C. BALDWIN, AND T.R. DELL. 1991. Growth and yield predictions for thinned and unthinned slash pine plantations on cutover sites in the West Gulf region. USDA For. Serv. Res. Pap. S0-264 32 p.

ZELLNER, A. 1994. Bayesian method of moments/instrumental variable (BMOM/IV) analysis of mean and regression models. P. 58–63 *in* Proc of Am. Stat. Assoc. Sec. on Bayesian Stat. Sci.

ZELLNER, A., AND C. MIN. 1994. Gibbs sampler convergence criteria (GSC$^2$) Unpubl. rep. HGB Alexander Res. Found., Graduate School of Business, University of Chicago.

## APPENDIX: Full Conditional Distributions

Let $Y_1$, $Y_2$, $Y_3$, and $Y_4$ in Zellner's notation be $n \times 1$ vectors containing the $n = 507$ observations on $\ln B$, $\ln D_{\min}$, $\ln D_{0.93}$, and $\{(\ln T - \ln N)/\ln T\}$, respectively. For brevity, let $Q = (\ln T - \ln N)/\ln T$. Furthermore, let $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$, and $\mathbf{X}_4$ in Zellner's notation be matrices of dimension $(n \times 4)$, $(n \times 5)$, $(n \times 6)$, and $(n \times 1)$, respectively, with a column of ones for the intercept (except for $\mathbf{X}_4$), and observations on the appropriate stand attributes in the remaining columns. Finally, let $\hat{\beta}_i, i = 1, 2, \ldots, 4$ be the least squares estimates of $b_i$, and let $s_i^2, i = 1, 2, \ldots, 4$ be the usual mean square errors. Then, using the BMOM/IV method, and exploiting the hierarchical structure of (14)–(16) as in Wakefield (1994), we find the full conditional predictive distributions for $\ln B$, $\ln D_{0.93}$, $\ln D_{\min}$, and $Q$ to be (up to proportionality)[3]:

$$\left[\ln B \mid (\ln D_{\min}, \ln D_{0.93}, \underline{\sigma}^2, \underline{x}^*, \underline{\hat{\beta}})\right] \propto \left(\prod_{i=1}^{3} \delta_i\right)^{-1}$$
$$\exp\left[-\left\{\frac{1}{2}\left(\ln B - x_1^*\hat{\beta}_1\right)^2 / \delta_1^2\right\} - \left\{\frac{1}{2}\left(\ln D_{0.93} - x_2^*\hat{\beta}_2\right)^2 / \delta_2^2\right\} - \left\{\frac{1}{2}\left(\ln D_{\min} - x_3^*\hat{\beta}_3\right)^2 / \delta_3^2\right\}\right] \tag{17}$$

$$\left[\ln D_{0.93} \mid \left(\ln D_{\min}, B, \underline{\sigma}^2, \underline{x}^*, \underline{\hat{\beta}}\right)\right] \propto \left(\prod_{i=2}^{3} \delta_i\right)^{-1}$$
$$\exp\left[-\left\{\frac{1}{2}\left(\ln D_{0.93} - x_2^*\hat{\beta}_2\right)^2 / \delta_2^2\right\} - \left\{\frac{1}{2}\left(\ln D_{\min} - x_3^*\hat{\beta}_3\right)^2 / \delta_3^2\right\}\right] \tag{18}$$

$$\left[\ln D_{\min} \mid \left(\ln D_{\min}, B, \underline{\sigma}^2, \underline{x}^*, \underline{\hat{\beta}}\right)\right] \propto \frac{1}{\delta_3} \exp\left\{-\frac{1}{2}\left(\ln D_{\min} - x_3^*\hat{\beta}_3\right)^2 / \delta_3^2\right\} \tag{19}$$

[3] For the Gibbs sampler, all we require are the conditional distributions up to proportionality. The normalizing constant need never be calculated. This is the primary reason why Gibbs sampling and related procedures have become widespread.

$$\left[Q \mid \left(\ln B, \ln D_{0.93} \ln D_{\min}, A, \underline{\sigma}^2, \underline{x}^*, \hat{\beta}\right)\right] \propto \left(\prod_{i=1}^{4} \delta_i\right)^{-1}$$
$$\exp\left[-\left\{\frac{1}{2}\left(Q - x_4^*\hat{\beta}_4\right)^2 / \delta_4^2\right\} - \left\{\frac{1}{2}\left(\ln B - x_1^*\hat{\beta}_1\right)^2 / \delta_1^2\right\} - \left\{\frac{1}{2}\left(\ln D_{0.93} - x_2^*\hat{\beta}_2\right)^2 / \delta_2^2\right\} - \left\{\frac{1}{2}\left(\ln D_{\min} - x_3^*\hat{\beta}_3\right)^2 / \delta_3^2\right\}\right] \quad (20)$$

where $x_i^*$ represents the level of the independent variables for the new observation, i.e.,

$$x_1^* = \left(1, \ln, H_D, \ln N, A^{-1}\right),$$

$$x_2^* = \left(1, \ln H_D, \ln N, A^{-1}, \ln B\right),$$

$$x_3^* = \left(1, \ln H_D, \ln N, A^{-1}, \ln B, D_{0.93}\right), \text{ and}$$

$$x_4^* = A;\ \underline{\sigma}^2 = \left(\sigma_i^2, i = 1, 2, \ldots, 5\right);$$

$$\underline{x}^* = \left(x_1^{*'}, x_2^{*'}, x_3^{*'}, x_4^{*'}\right)';$$

$$\underline{\hat{\beta} = \left(\hat{\beta}_1', \hat{\beta}_2', \hat{\beta}_3', \hat{\beta}_4'\right)'};\ \delta_i^2 = \sigma_i^2\left(1 + x_i^*\left(\mathbf{X}_i'\mathbf{X}_i\right)^{-1} x_i^{2'}\right), i = 1, 2, \ldots, 4.$$

There is a slight complication with the full conditional for $\ln H_D$. As previously mentioned, we do not fit Equation (5), but rather the linear model $\ln H_D = \beta_{51} + \beta_{52}A^{-0.5}$. The intercept is removed by imposing the constraint $\ln S = \beta_{51} + \beta_{52}(25^{-0.5})$. Hence $\ln H_D$ is a function of the unknown coefficient $\beta_{52}$ and we have $\mathbf{E}(\ln H_D) = \ln S + (A^{-0.5} - 0.2)\mathbf{E}(\beta_{52})$ and $\mathbf{VAR}(\ln H_D) = (A^{-0.5} - 0.2)^2\,\mathbf{VAR}(\beta_{52})$. By applying the BMOM/IV method, we know that the conditional posterior distribution of $\beta_{52}$ is normal.[4]

Hence we include the full conditional distribution for $\beta_{52}$ in the Gibbs sampler. For each generated value of $\beta_{52}$, we can determine a value for $H_D$ by simply multiplying the generated value by $(A^{-0.5} - 0.2)$ and adding $\ln S$. This is proper because we know that if the Gibbs sampler is used to produce a marginal posterior sample of a given variable, then a marginal posterior sample of a function of that variable is available by simply computing the appropriate functional values for each observation in the original marginal posterior sample (Gelfand et al. 1990). The full conditional for $\beta_{52}$ is (up to proportionality):

$$\left[\beta_{52} \mid \ln B, \ln D_{0.93}, D_{\min}, A, \underline{\sigma}^2, \underline{x}^*, \underline{\hat{\beta}}\right] \propto \left(\prod_{i=1}^{5} \delta_i\right)^{-1}$$
$$\exp\left[\begin{array}{l} -\left\{\frac{1}{2}\left(\beta_{52} - \hat{\beta}_{53}\right)^2 / \delta_5^2\right\} - \left\{\frac{1}{2}\left(Q - x_4^*\hat{\beta}_4\right)^2 / \delta_4^2\right\} \\ -\left\{\frac{1}{2}\left(\ln B - x_1^*\hat{\beta}1\right)^2 / \delta_1^2\right\} - \left\{\frac{1}{2}\left(\ln D_{0.93} - x_2^*\hat{\beta}_2\right)^2 / \delta_2^2\right\} - \left\{\frac{1}{2}\left(D_{\min} - x_3^*\hat{\beta}_3\right)^2 / \delta_3^2\right\} \end{array}\right] \quad (21)$$

4 We ignore the error variance $\sigma_5^2$ in the distribution for $\ln H_D$ because we will always assume that it is the mean $\ln H_D$ that is desired. Hence only the variation in the slope of the regression line is important (the intercept having been removed via constraint).

where $\delta_5^2 = c_{22}\sigma_5^2$, and $c_{22}$is the (2,2) element in $(\mathbf{X}_5'\mathbf{X}_5)^{-1}$.

We also need the full conditionals for the error variances. These are (up to proportionality):

$$\left[\sigma_1^2 \mid \ln B, \ln D_{0.93}, \ln D_{\min}, A, \underline{x}^*, \underline{\hat{\beta}}, s_1^2\right] \propto (1/\delta_1)(1/s_1^2)\exp\left[\left(-\sigma_1^2/\delta_1^2\right) + \left\{-\frac{1}{2}\left(\ln B - x_1^*\hat{\beta}_1\right)^2/\delta_1^2\right\}\right] \tag{22}$$

$$\left[\sigma_2^2 \mid \ln B, \ln D_{0.93}, \ln D_{\min}, A, \underline{x}^*, \underline{\hat{\beta}}, s_2^2\right] \propto (1/\delta_2)(1/s_2^2)\exp\left[\left(-\sigma_2^2/\delta_2^2\right) + \left\{-\frac{1}{2}\left(\ln D_{0.93} - x_2^*\hat{\beta}_2\right)^2/\delta_2^2\right\}\right] \tag{23}$$

$$\left[\sigma_3^2 \mid \ln B, \ln D_{0.93}, \ln D_{\min}, A, \underline{x}^*, \underline{\hat{\beta}}, s_3^2\right] \propto (1/\delta_3)(1/s_3^2)\exp\left[\left(-\sigma_3^2/\delta_3^2\right) + \left\{-\frac{1}{2}\left(\ln D_{\min} - x_3^*\hat{\beta}_3\right)^2/\delta_3^2\right\}\right] \tag{24}$$

$$\left[\sigma_4^2 \mid \ln B, \ln D_{0.93}, \ln D_{\min}, A, \underline{x}^*, \underline{\hat{\beta}}, s_4^2\right] \propto (1/\delta_4)(1/s_4^2)\exp\left[\left(-\sigma_4^2/\delta_4^2\right) + \left\{-\frac{1}{2}\left(Q - x_4^*\hat{\beta}_4\right)^2/\delta_4^2\right\}\right] \tag{25}$$

$$\left[\sigma_5^2 \mid \ln B, \ln D_{0.93}, \ln D_{\min}, A, \underline{x}^*, \underline{\hat{\beta}}, s_5^2\right] \propto (1/\delta_5)(1/s_5^2)\exp\left[\left(-\sigma_5^2/\delta_5^2\right) + \left\{-\frac{1}{2}\left(\beta_{52} - \hat{\beta}_{52}\right)^2/\delta_5^2\right\}\right] \tag{26}$$

where $s_i^2, i = 1, 2, \ldots, 5$ is the usual least squares mean square error (MSE) for the $i$th model.