

Statistical Modeling of Seedling Mortality

Michael LAVINE, Brian BECKAGE, and James S. CLARK

Seedling mortality in tree populations limits population growth rates and controls the diversity of forests. To learn about seedling mortality, ecologists use repeated censuses of forest quadrats to determine the number of tree seedlings that have survived from the previous census and to find new ones. Typically, newly found seedlings are marked with flags. But flagging is labor intensive and limits the spatial and temporal coverage of such studies. The alternative of not flagging has the advantage of ease but suffers from two main disadvantages. It complicates the analysis and loses information. The contributions of this article are (i) to introduce a method for using unflagged census data to learn about seedling mortality and (ii) to quantify the information loss so ecologists can make informed decisions about whether to flag. Based on presented results, we believe that not flagging is often the preferred alternative. The labor saved by not flagging can be used to better advantage in extending the coverage of the study.

Key Words: Bayesian inference; Ecological statistics; Experimental design; Fisher information; Gibbs sampling.

1. INTRODUCTION

Forest diversity is limited by recruitment of new tree seedlings (Watt 1947; Grubb 1977; Pacala and Tilman 1994; Clark et al. 1999; Hubbell et al. 1999). Information on seedling mortality is therefore critical to understanding how tree diversity is maintained in forests. Furthermore, spatial and temporal variability in seedling mortality caused by such factors as light levels, understory shrubs, and resource limitation is particularly important because it contributes to patterns of forest composition. Therefore, a method that simplifies the collection and estimation of seedling mortality in varied microenvironmental conditions would be valuable.

To learn about seedling mortality, ecologists conduct regular censuses of forest quadrats to count seedlings that survived from the previous census and to locate new ones. Standard practice is to flag (mark with a unique identifier) newly found seedlings and record which previously found seedlings are still alive.

Michael Lavine is Associate Professor, Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708 (E-mail: michael@stat.duke.edu). James S. Clark is Professor, Department of Biology, Duke University, Durham, NC 27708 (E-mail: jimclark@duke.edu). Brian Beckage is Postdoctoral Research Associate, Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803.

©2002 American Statistical Association and the International Biometric Society
Journal of Agricultural, Biological, and Environmental Statistics, Volume 7, Number 1, Pages 21–41

This article considers the alternative of simply censusing seedlings and not flagging each individual. Our goals are twofold. First, we develop a method for statistical analysis of unflagged data and show what useful information can be extracted from the data in the absence of flags. Second, we quantify the information lost by not flagging so ecologists can make well-informed decisions about whether to flag. The model is explained in Section 2 and illustrated on a 5-year data set from the southern Appalachians (Clark, Macklin, and Wood 1998) in Section 4. Information loss is quantified in Section 3.

The advantage of not flagging is, of course, ease, which permits analysis of many more quadrats. The disadvantages are threefold. First, unflagged data are uninformative about age-specific mortality rates. Our methods categorize seedlings as either first year (new) or older (old) seedlings and provide age-specific mortality rates only for those two categories. More detailed age-specific rates are often unnecessary.

Second, elimination of flags results in the loss of some information, even for our old and new mortality rates. The simulations and calculations of Section 3 quantify the information loss and show that not flagging is often a sensible alternative because the information loss in each quadrat is well compensated by the gain in the number of quadrats that can be censused. Moreover, experience shows that flags are often lost from one census to the next. Our calculations assume that no flags are lost and therefore overstate the advantage of flagging.

Third, elimination of flags complicates the analysis because a seedling's age is unknown. This is precisely the point where information is lost and the analysis becomes more complex. We use a Bayesian approach to account for the unknown age. Of course, such an analysis could also be used in a flagged study when flags are lost.

We apply the method to a 5-year data set from the southern Appalachians (Clark et al. 1998), where we conducted annual censuses of new and old seedlings in 1-m² quadrats arranged along transects. Individuals were not flagged; we simply counted individuals of each species in the two age classes. We use counts of red maple (*Acer rubrum*) to demonstrate our analysis.

2. A PROCESS MODEL

A seedling's survival probability is the chance that it survives from one year to the next. Old and new seedlings can be identified by the presence or absence, respectively, of bud scale scars. Old and new seedlings have different survival probabilities but, roughly speaking, the survival probability of old seedlings is not explained by age (Streng, Glitzenstein, and Harcombe 1989; Jones, Sharitz, Dixon, Segal, and Schneider 1994). Therefore, we adopt a model with two parameters of interest— p_{old} and p_{new} —survival probabilities for old and new seedlings, respectively.

Let $O_{i,j}$ and $N_{i,j}$ be the numbers of old and new seedlings recorded in quadrat i in year j . $N_{i,j}$ represents new recruitment, which, for the moment, we do not model but take as given. The numbers of old seedlings recorded in year 1, $\{O_{i,1}\}$, are also given. Seedlings are typically small enough not to be limited by mutual competition for physical resources;

their survivals can reasonably be modeled as independent within a quadrat. Let $X_{i,j}$ be the number of old seedlings and $Y_{i,j}$ the number of new seedlings in quadrat i that survive from year j to year $j + 1$. Let $\text{bin}(n, p)$ be the binomial distribution with n trials and probability p and $\text{bin}(x; n, p)$ the probability mass function of $\text{bin}(n, p)$ evaluated at x . Our model is

$$X_{i,j} \sim \text{bin}(O_{i,j}, p_{\text{old}}) \quad (2.1)$$

$$Y_{i,j} \sim \text{bin}(N_{i,j}, p_{\text{new}}). \quad (2.2)$$

When new seedlings survive, they become old seedlings. Thus, for $j \geq 2$, $O_{i,j} = X_{i,j-1} + Y_{i,j-1}$. For now, data from different quadrats and different years are assumed conditionally independent of each other given p_{old} and p_{new} , but see Section 4.3 for year and quadrat effects that induce dependence. Let Q be the number of quadrats and T the number of years for which we have data. The full model for all the X 's and Y 's is

$$\begin{aligned} & \Pr[\{X_{i,j}, Y_{i,j}\}_{i=1,\dots,Q; j=1,\dots,T-1} \mid p_{\text{old}}, p_{\text{new}}] \\ &= \prod_{i=1}^Q \left\{ \Pr[X_{i,1} \mid p_{\text{old}}, p_{\text{new}}] \Pr[Y_{i,1} \mid p_{\text{old}}, p_{\text{new}}] \right. \\ & \quad \left. \times \prod_{j=2}^{T-1} \Pr[X_{i,j} \mid X_{i,j-1}, Y_{i,j-1}, p_{\text{old}}, p_{\text{new}}] \right\} \\ &= \prod_{i=1}^Q \left\{ \text{bin}(X_{i,1}; O_{i,1}, p_{\text{old}}) \text{bin}(Y_{i,1}; N_{i,1}, p_{\text{new}}) \right. \\ & \quad \left. \times \prod_{j=2}^{T-1} \text{bin}(X_{i,j}; O_{i,j}, p_{\text{old}}) \text{bin}(Y_{i,j}; N_{i,j}, p_{\text{new}}) \right\}. \quad (2.3) \end{aligned}$$

If seedlings are flagged then, when $O_{i,j}$ is observed in year j , we also learn $X_{i,j-1}$ and $Y_{i,j-1}$ and the experiment is just a collection of many independent binomial observations and statistical analysis is simple. But if seedlings are not flagged then, when $O_{i,j}$ is observed, we don't know how many of those old seedlings in year j were old and how many were new in year $j - 1$, i.e., we do not observe $X_{i,j-1}$ and $Y_{i,j-1}$. Some information is lost.

Although $\{X_{i,j}\}$ and $\{Y_{i,j}\}$ are not observed, we can still draw inferences about p_{old} and p_{new} from observations of $\{O_{i,j}\}$ and $\{N_{i,j}\}$. To account for $O_{i,j}$ seedlings in quadrat i in year j , $X_{i,j-1}$ must be at least $O_{i,j} - N_{i,j-1}$, so the lower bound on $X_{i,j-1}$ is $X_{i,j-1}^{\min} \equiv \max(0, O_{i,j} - N_{i,j-1})$. Similarly, the upper bound is $X_{i,j-1}^{\max} \equiv \min(O_{i,j-1}, O_{i,j})$. Thus, $X_{i,j-1} \in [X_{i,j-1}^{\min}, X_{i,j-1}^{\max}]$ and $Y_{i,j-1} = O_{i,j} - X_{i,j-1}$. Then for a given i and j ,

$$\begin{aligned} & \Pr[O_{i,j} \mid O_{i,j-1}, N_{i,j-1}, p_{\text{old}}, p_{\text{new}}] \\ &= \sum_{x=X_{i,j-1}^{\min}}^{X_{i,j-1}^{\max}} \left\{ \Pr[X_{i,j-1} = x \mid O_{i,j-1}, N_{i,j-1}, p_{\text{old}}, p_{\text{new}}] \right. \\ & \quad \left. \times \Pr[O_{i,j} \mid X_{i,j-1} = x, O_{i,j-1}, N_{i,j-1}, p_{\text{old}}, p_{\text{new}}] \right\} \end{aligned}$$

$$= \sum_{x=X_{i,j-1}^{\min}}^{X_{i,j-1}^{\max}} \text{bin}(x; O_{i,j-1}, p_{\text{old}}) \text{bin}(O_{i,j} - x; N_{i,j-1}, p_{\text{new}}). \quad (2.4)$$

The probability of all the data is the product over quadrats and years of terms like Equation (2.4),

$$\begin{aligned} & \Pr[\{O_{i,j} : j \geq 2\} \mid \{O_{i,1}\}, \{N_{i,j}\}, p_{\text{old}}, p_{\text{new}}] \\ &= \prod_{i=1}^Q \prod_{j=2}^T \Pr[O_{i,j} \mid O_{i,j-1}, N_{i,j-1}, p_{\text{old}}, p_{\text{new}}] \\ &= \prod_{i=1}^Q \prod_{j=2}^T \sum_{x=X_{i,j-1}^{\min}}^{X_{i,j-1}^{\max}} \text{bin}(x; O_{i,j-1}, p_{\text{old}}) \text{bin}(O_{i,j} - x; N_{i,j-1}, p_{\text{new}}). \quad (2.5) \end{aligned}$$

We verified that this model produces satisfactory posterior distributions on simulated data and then turned our attention to the question of flags.

3. INFORMATION

To flag or not to flag, that is the question!

The likelihood function in Equation (2.5), along with a prior distribution for p_{old} and p_{new} , yields a joint posterior for p_{old} and p_{new} . To understand whether flagging is worth the effort, we compare that posterior with the one we would obtain using the likelihood function in Equation (2.3).

One way to portray the value of flagging is to simulate many data sets, analyze each one with and without flags, and calculate regions of high posterior density for the parameters of interest. For this purpose, we simulated data from 100 experiments with $T = 5$, $p_{\text{old}} = 0.8$, $p_{\text{new}} = 0.3$, $O_{i,1} = 1$, $N_{i,j}$ having a Poisson distribution with rate $\lambda = 1$, and either $Q = 10$ or $Q = 30$. The values of p_{old} , p_{new} , and λ are ones we considered plausible for real forests a priori and were used only to generate data, not to analyze it. Later in this section, we consider a more malicious choice of parameters; Section 4 shows the values of parameters that turned out to be plausible a posteriori for our data set. We used independent beta(1.5, 1.5) priors for p_{old} and p_{new} and calculated lengths of approximate 90% highest posterior density (HPD) regions, i.e., the smallest possible intervals containing 90% of the posterior mass for p_{old} and p_{new} . Results, jittered for legibility, are in Figure 1. Each point represents one simulated data set. The horizontal axis shows the length of the 90% HPD interval that would have been obtained had the experiment been run with flags. The vertical axis shows the length of the 90% HPD interval that would have been obtained had the experiment been run without flags. As expected, HPD regions from nonflagged analyses are longer than HPD regions from flagged analyses. The plots also show that nonflagging typically results in a length increase of less than 50%. The usual asymptotics imply that lengths will decrease in proportion to $Q^{-0.5}$. Thus, assuming conditions similar to those of the simulation, an experiment without flags will need roughly twice as many quadrats to obtain HPD intervals of similar length.

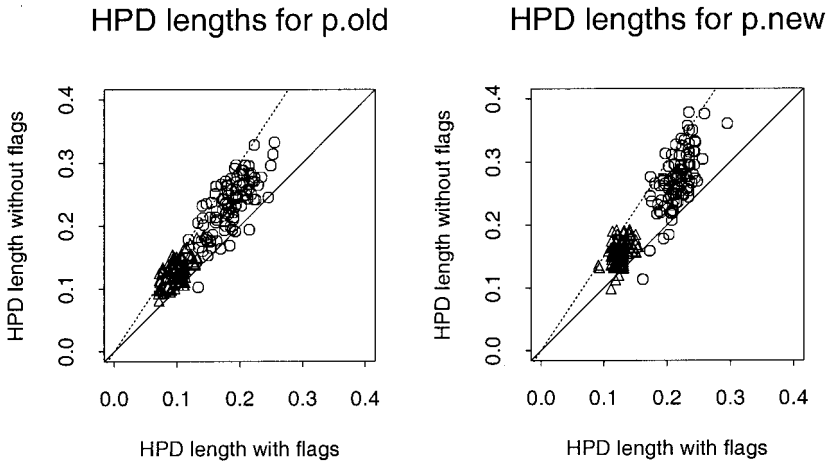


Figure 1. Lengths of HPD Regions From Simulated Experiments. Circles are from experiments with $Q = 10$; triangles are $Q = 30$. The solid line has slope 1; the dotted line has slope 1.5. Data are jittered for legibility.

The simulation can be supported by calculations of the Fisher information in flagged and unflagged experiments. As an example, consider a single quadrat and suppose that $O_{1,1} = N_{1,1} = 1$. In year 2 of a flagged experiment, we observe $X_{1,1}$ and $Y_{1,1}$. The Fisher information matrix for (p_{old}, p_{new}) is

$$\begin{bmatrix} \frac{-1}{p_{old}(1-p_{old})} & 0 \\ 0 & \frac{-1}{p_{new}(1-p_{new})} \end{bmatrix}.$$

In contrast, in year 2 of an unflagged experiment, we observe $O_{1,2}$. The Fisher information matrix is

$$\frac{-1}{p_{old} + p_{new} - 2p_{old}p_{new}} \begin{bmatrix} \frac{p_{old} + p_{new}^2 - 2p_{old}p_{new}}{p_{old}(1-p_{old})} & 1 \\ 1 & \frac{p_{old}^2 + p_{new} - 2p_{old}p_{new}}{p_{new}(1-p_{new})} \end{bmatrix}$$

with inverse

$$\frac{1}{(p_{old} - p_{new})^2} \times \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

where

$$\begin{aligned} a &= -p_{old}(1-p_{old})(p_{old}^2 - 2p_{old}p_{new} + p_{new}) \\ b &= p_{old}(1-p_{old})p_{new}(1-p_{new}) \\ c &= p_{old}(1-p_{old})p_{new}(1-p_{new}) \\ d &= -p_{new}(1-p_{new})(p_{new}^2 - 2p_{old}p_{new} + p_{old}). \end{aligned}$$

Large-sample standard deviations for p_{old} and p_{new} , whether classical or Bayesian, will be roughly proportional to the square roots of the diagonal elements of the inverses of the

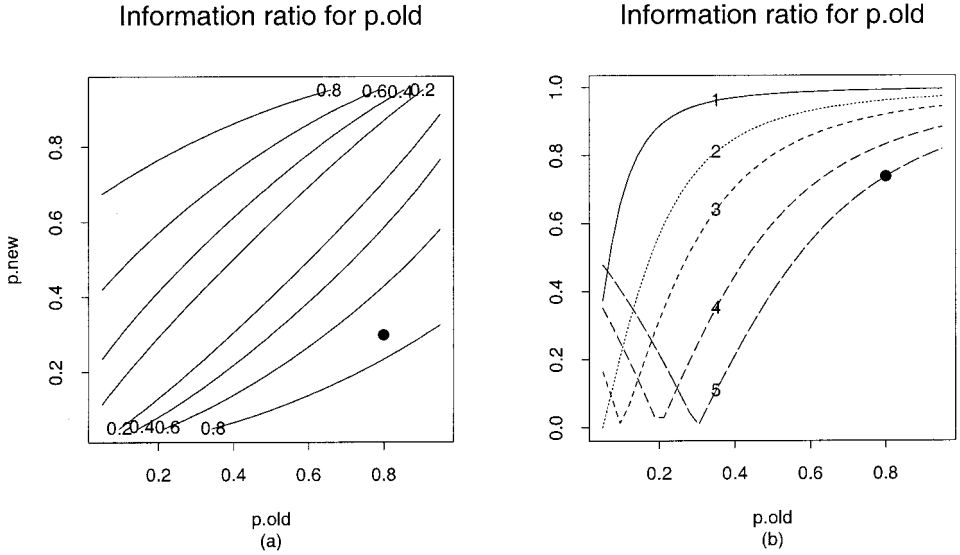


Figure 2. $SD_{\text{flag}}(p_{\text{old}})/SD_{\text{noflag}}(p_{\text{old}})$ as in Equation (3.1) (a) as a Function of p_{old} and p_{new} , (b) as a Function of p_{old} for (1) $p_{\text{new}} = 0.01$, (2) $p_{\text{new}} = 0.05$, (3) $p_{\text{new}} = 0.10$, (4) $p_{\text{new}} = 0.20$, and (5) $p_{\text{new}} = 0.30$.

Fisher information matrices. To help decide whether to flag, we examine the ratio of the standard deviations from flagged and unflagged experiments, i.e.,

$$\frac{SD_{\text{flag}}(p_{\text{old}})}{SD_{\text{noflag}}(p_{\text{old}})} \approx \sqrt{\frac{p_{\text{old}}^2 - 2p_{\text{old}}p_{\text{new}} + p_{\text{new}}^2}{p_{\text{old}}^2 - 2p_{\text{old}}p_{\text{new}} + p_{\text{new}}}}. \quad (3.1)$$

Figure 2 displays this function, which, by symmetry, applies to p_{new} as well. Panel (a) is a contour plot; panel (b) displays the ratio as a function of p_{old} for five different values of p_{new} ranging from 0.01 to 0.30, values we consider plausible in real forests. In each panel, the black spot is the point at which the HPD simulations were done. This ratio is bounded between zero and one. Low values say that a flagged experiment is much more informative than an unflagged experiment; high values say the opposite. The ratio goes to zero along the diagonal, suggesting that, when $p_{\text{new}} = p_{\text{old}}$, a flagged experiment is infinitely more informative than an unflagged experiment. Sundberg (1974) contains relevant theory. $(X_{i,j}, Y_{i,j})$ follows an exponential family of distributions. Observing only their sum $O_{i,j+1}$ is a special case of Sundberg's Example 3 (1974, p. 50). Paraphrasing from Sundberg (p. 54),

Let $\alpha = (p_{\text{old}}, p_{\text{new}})$ be the true parameter. All we need impose in order to obtain $n^{1/2}$ consistency is that the Fisher information matrix be strictly positive definite, ensuring that the observable statistic provides information on all of the parameter α . This condition is not necessary for consistency, but when the condition is not satisfied, we cannot expect a maximum likelihood estimate to be $n^{1/2}$ consistent, i.e., $n^{1/2}(\hat{\alpha} - \alpha)$ to be bounded in probability as $n \rightarrow \infty$.

Our large sample SD ratio goes to zero along the diagonal because the Fisher information matrix in the unflagged case is not positive definite when $p_{\text{old}} = p_{\text{new}}$. To see for which function of $(p_{\text{old}}, p_{\text{new}})$ we have direct information, it is useful to rewrite the likelihood, dropping subscripts on $X_{1,1}$ and $Y_{1,1}$, as

$$\begin{aligned} p(x, y \mid p_{\text{old}}, p_{\text{new}}) &= p_{\text{old}}^x (1 - p_{\text{old}})^{1-x} p_{\text{new}}^y (1 - p_{\text{new}})^{1-y} \\ &= \exp \left\{ x \ln \frac{p_{\text{old}}}{1 - p_{\text{old}}} + y \ln \frac{p_{\text{new}}}{1 - p_{\text{new}}} + \ln(1 - p_{\text{old}})(1 - p_{\text{new}}) \right\} \\ &= \exp \left\{ (x + y) \frac{1}{2} \ln \frac{p_{\text{old}} p_{\text{new}}}{(1 - p_{\text{old}})(1 - p_{\text{new}})} \right. \\ &\quad \left. + (x - y) \frac{1}{2} \ln \frac{p_{\text{old}}(1 - p_{\text{new}})}{(1 - p_{\text{old}})p_{\text{new}}} + \ln(1 - p_{\text{old}})(1 - p_{\text{new}}) \right\}, \end{aligned}$$

from which we see that $\ln[(p_{\text{old}} p_{\text{new}})/(1 - p_{\text{old}})(1 - p_{\text{new}})]$ is the parameter with direct information when we observe $x + y$. Brown (1986, pp. 8–13) contains more details about marginal distributions from exponential families.

In real forests, p_{new} is generally quite small and less than p_{old} , so the SD ratio under realistic conditions is bounded away from zero. Nonetheless, there are realistic values of $(p_{\text{old}}, p_{\text{new}})$ for which the ratio is quite small. Apparently, under these conditions, one could lose quite a bit of information, relatively speaking, by not flagging. To see just how much, without relying on asymptotics, we conducted a simulation similar to that in Figure 1 but this time with $p_{\text{old}} = p_{\text{new}} = 0.3$. Figure 3 shows the results. Not much has changed for p_{old} . But for p_{new} , the situation is quite different. HPD regions from unflagged experiments can be up to four times as long as those from flagged experiments. If estimation of p_{new} were the goal and length of HPD region were the utility, then unflagged experiments would need up to about 16 times as many quadrats to achieve the same utility as flagged experiments. This penalty is generally a small price to pay for the convenience of not flagging.

4. ANALYSIS

This section presents our analysis of data collected annually on red maple (*Acer rubrum*) seedlings from 1-m² quadrats arranged along transects in the southern Appalachians (Clark et al. 1998). To carry out the analysis, we have to account for errors in the data (Section 4.1). We also examine some aspects of sensitivity to the prior (Section 4.2) and to the assumption that parameters are constant through space and time (Section 4.3).

4.1 ERRORS IN DATA

Table 1 shows the recorded counts of maple seedlings over a 5-year period in quadrat 2 at the Coweeta Experimental Station in western North Carolina.

The point of interest is that $O_{2,3} > O_{2,2} + N_{2,2}$. According to our model, that's not possible and can only be the result of a data error. In the case of red maple, such errors are

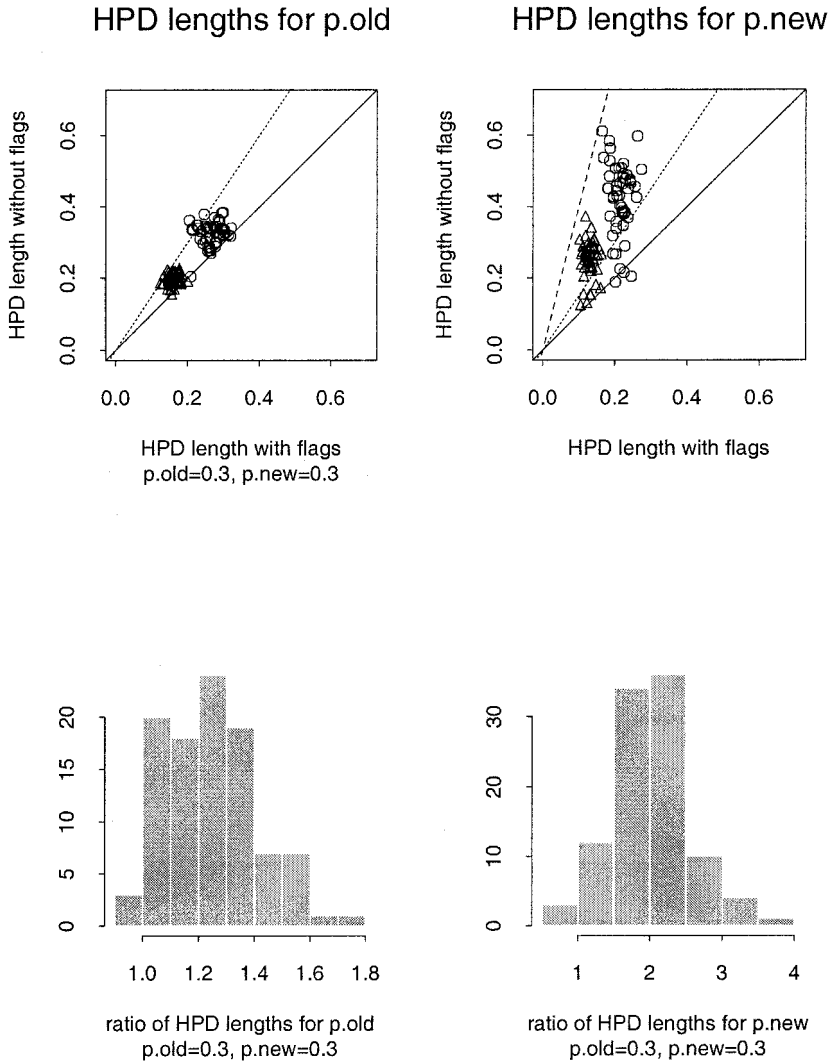


Figure 3. Lengths of HPD Regions From Simulated Experiments. Octagons are from experiments with $Q = 10$; triangles are $Q = 30$. The solid line has slope 1; the dotted line has slope 1.5; the dashed line has slope 4. Data are jittered for legibility. Histograms of HPD length from unflagged analysis divided by HPD length from flagged analysis.

most likely due to new seedlings emerging late in the year, after the July census. We therefore refine the model to account for such errors. In effect, the true number of new seedlings should be the number that have emerged before winter, which might be more than the number that have emerged and been found when the data are recorded. These considerations lead to a model in which we assume that the number of old seedlings has been correctly recorded but the number of new seedlings might be in error.

Let $N_{i,j}^T$ be the true number of new seedlings in quadrat i in year j . We model $N_{i,j} \sim \text{bin}(N_{i,j}^T, f_{\text{new}})$, where f_{new} is the probability that a given new seedling is present

Table 1. Numbers of Old and New Maple (*Acer*) Seedlings Recorded in Quadrat 2 at Coweeta.

	Year				
	1	2	3	4	5
$O_{2,j}$	1	0	1	0	0
$N_{2,j}$	0	0	0	0	0

and recorded at the time of the census, i.e., $N_{i,j}$ comes from a binomial distribution with both parameters unknown. Inference in such a problem is notoriously difficult because of an ill-behaved likelihood function. (See, e.g., the discussion in Kahn (1987), Raftery (1988), or Lavine and Wasserman (1992).) We also change Equation (2.2) to

$$Y_{i,j} \sim \text{bin}(N_{i,j}^T, p_{\text{new}})$$

and set

$$X_{i,j-1}^{\min} = \max(0, O_{i,j} - N_{i,j-1}^T).$$

The full model is now

$$\begin{aligned} & \Pr [\{O_{i,j} : j \geq 2\}, \{N_{i,j}\} \mid \{O_{i,1}\}, \{N_{i,j}^T\}, p_{\text{old}}, p_{\text{new}}, f_{\text{new}}] \\ &= \prod_{i=1}^Q \prod_{j=2}^T \sum_{x=X_{i,j-1}^{\min}}^{X_{i,j-1}^{\max}} \text{bin}(x; O_{i,j-1}, p_{\text{old}}) \text{bin}(O_{i,j} - x; N_{i,j-1}^T, p_{\text{new}}) \\ & \times \prod_{i=1}^Q \prod_{j=1}^T \text{bin}(N_{i,j}; N_{i,j}^T, f_{\text{new}}). \end{aligned}$$

Let $\text{beta}(\cdot, \cdot)$, $\text{Poi}(\cdot)$, and $\text{gam}(\cdot, \cdot)$ be the beta, Poisson, and gamma distributions, respectively. For the purpose of Bayesian analysis, we adopt the following priors:

$$\begin{aligned} p_{\text{old}} &\sim \text{beta}(\alpha_{p_{\text{old}}}, \beta_{p_{\text{old}}}) \\ p_{\text{new}} &\sim \text{beta}(\alpha_{p_{\text{new}}}, \beta_{p_{\text{new}}}) \\ f_{\text{new}} &\sim \text{beta}(\alpha_{f_{\text{new}}}, \beta_{f_{\text{new}}}) \\ \{N_{i,j}^T\} &\mid \lambda \sim \text{i.i.d. Poi}(\lambda) \\ \lambda &\sim \text{gam}(\alpha_{\lambda}, \beta_{\lambda}). \end{aligned}$$

We take $\alpha_{p_{\text{old}}} = \beta_{p_{\text{old}}} = \alpha_{p_{\text{new}}} = \beta_{p_{\text{new}}} = 1$ for flat priors on p_{old} and p_{new} and $\alpha_{\lambda} = 0.4$, $\beta_{\lambda} = 10$ for a fairly flat prior on λ with a mean = 4. Including $\{X_{i,j}\}$ as auxiliary variables makes the posterior amenable to Gibbs sampling (Gelfand and Smith 1990; Casella and George 1992), with the following complete conditionals:

- (1) p_{old} : Each quadrat in each year yields a binomial experiment with $O_{i,j}$ number of trials, $X_{i,j}$ number of successes, and p_{old} probability. Therefore, the complete

conditional distribution for p_{old} is

$$\text{beta} \left(\alpha_{p_{\text{old}}} + \sum_1^Q \sum_1^{T-1} X_{i,j}, \quad \beta_{p_{\text{old}}} + \sum_1^Q \sum_1^{T-1} (O_{i,j} - X_{i,j}) \right).$$

- (2) p_{new} : Each quadrat in each year yields a binomial experiment with $N_{i,j}^T$ number of trials, $O_{i,j+1} - X_{i,j}$ number of successes, and p_{new} probability. Therefore, the complete conditional distribution for p_{new} is

$$\text{beta} \left(\alpha_{p_{\text{new}}} + \sum_1^Q \sum_1^{T-1} (O_{i,j+1} - X_{i,j}), \beta_{p_{\text{new}}} + \sum_1^Q \sum_1^{T-1} (N_{i,j}^T - O_{i,j+1} + X_{i,j}) \right).$$

- (3) f_{new} : Each quadrat in each year yields a binomial experiment with $N_{i,j}^T$ number of trials, $N_{i,j}$ number of successes, and f_{new} probability. Therefore, the complete conditional distribution for f_{new} is

$$\text{beta} \left(\alpha_{f_{\text{new}}} + \sum_1^Q \sum_1^T N_{i,j}, \quad \beta_{f_{\text{new}}} + \sum_1^Q \sum_1^T (N_{i,j}^T - N_{i,j}) \right).$$

- (4) λ : Each quadrat in each year yields a Poisson experiment with parameter λ and $N_{i,j}^T$ number of occurrences. Therefore, the complete conditional distribution for λ is

$$\text{gam} \left(\alpha_\lambda + \sum_1^Q \sum_1^T N_{i,j}^T, \beta_\lambda + QT \right).$$

- (5) $X_{i,j}$: Each quadrat in each year yields two binomial experiments with parameters $(O_{i,j}, p_{\text{old}})$ and $(N_{i,j}^T, p_{\text{new}})$. We observe their sum $O_{i,j+1}$, which gives bounds $X_{i,j}^{\min}$ and $X_{i,j}^{\max}$ on $X_{i,j}$. (See last paragraph of page 23 and the second display equation on page 29.) Therefore, the complete conditional distribution for $X_{i,j}$ is discrete on the integers from $X_{i,j}^{\min}$ to $X_{i,j}^{\max}$ with probabilities proportional to

$$\text{bin}(X_{i,j}; O_{i,j}, p_{\text{old}}) \times \text{bin}(O_{i,j+1} - X_{i,j}; N_{i,j}^T, p_{\text{new}}).$$

- (6) $N_{i,j}^T$: Each $N_{i,j}^T$ is a realization of a Poisson random variable with parameter λ and is also the number-of-trials parameter in two binomial experiments, one in which $N_{i,j}$ seedlings are observed and one in which $O_{i,j+1} - X_{i,j}$ seedlings survive. Therefore, the complete conditional distribution for $N_{i,j}^T$ is discrete on the integers greater than or equal to $\min(O_{i,j+1} - X_{i,j}, N_{i,j})$ with probabilities proportional to

$$\text{Poi}(N_{i,j}^T; \lambda) \times \text{bin}(O_{i,j+1} - X_{i,j}; N_{i,j}^T, p_{\text{new}}) \times \text{bin}(N_{i,j}; N_{i,j}^T, f_{\text{new}}).$$

One could sample $N_{i,j}^T$ with a random walk Metropolis step, but we find it simpler to fix an upper bound on $N_{i,j}^T$, calculate all the probabilities between the lower and upper bounds, and use a Gibbs step. This works well in practice because the Poisson part of the distribution prevents $N_{i,j}^T$ from taking very large values. Therefore, the upper bound makes little practical difference.

After verifying that this model yields satisfactory posteriors on simulated data, we turned our attention to real data. Table 2 shows data from 60 quadrats at Coweeta. Figure 4 shows posterior density estimates based on output from 2,000 iterations of a Gibbs sampler, collected at intervals of 50 after discarding the initial 500. Figure 4 is important not only for what it says about marginal posterior densities but also as a basis for comparison later in this article. Sections 4.2 and 4.3 explore sensitivities, and figures in those sections will be compared with Figure 4 when the prior and the model are elaborated.

4.2 SENSITIVITY TO THE PRIOR

Figure 5 is a pairs plot of draws from the joint posterior distribution of p_{old} , p_{new} , f_{new} , and λ . The last three parameters are highly dependent on each other but relatively independent of p_{old} . The dependence is easiest to understand in the case of $(f_{\text{new}}, \lambda)$. Our model says that new seedlings arise as a Poisson process with rate λ ; we observe each one with probability f_{new} ; therefore, our observations $\{N_{i,j}\}$ are Poisson with rate $\lambda' = \lambda f_{\text{new}}$ and, therefore, the data can pin down λ' fairly well but not each parameter separately. In fact, the joint posterior density of f_{new} and λ has a ridge along the hyperbola $\lambda f_{\text{new}} = 0.843$, the posterior mean of λ' .

The joint posterior of p_{new} and λ also follows a hyperbola, $\lambda p_{\text{new}} = 0.352$, the posterior mean of $\lambda'' = \lambda p_{\text{new}}$, for similar reasons. Our model says the number of new seedlings that survive from one year to the next in a given quadrat is Poisson with rate λ'' , so λ'' is roughly determined by the number of new seedling survivals needed for the model to be a good fit for the data. And p_{new} and f_{new} are positively correlated with each other because they are each negatively correlated with λ .

The foregoing discussion suggests that the joint likelihood of p_{new} , f_{new} , and λ is sharpest in the two directions corresponding to λ' and λ'' but is relatively flat in other directions; inspection of Figure 5 shows that the conditional posterior distribution of any one of these three variables depends strongly on the values of the others. These considerations suggest that prior elicitation is likely to be an important issue, one that could have a large influence on the posterior.

It is worth exploring, then, the sensitivity of the posterior density of p_{old} and p_{new} , the parameters of primary interest, to the prior for f_{new} . As an extreme exploration, we calculated the posterior conditional on $f_{\text{new}} = 0.9$. Figure 6 shows the marginal posteriors. As feared, this changes the posterior of p_{new} . It also changes the posterior of λ , but that is of secondary interest, and makes minor changes to the posterior of p_{old} .

There are several possible solutions to this problem. One is to incorporate other information about at least one dimension of $(p_{\text{new}}, f_{\text{new}}, \lambda)$. For example, frequent revisiting of a small number of quadrats during the course of a growing season could provide good information and hence strong priors for f_{new} and λ . Such priors would cut across the flat direction of the likelihood function and yield sharper inferences.

Another possible solution is to conduct future censuses later in the growing season. A census timed after most seedlings have emerged is almost equivalent to an a priori belief

Table 2. Data From Coweeta: 60 Plots, 5 Years

Quadrat	<i>Old seedlings</i>					<i>New seedlings</i>				
	1993	1994	1995	1996	1997	1993	1994	1995	1996	1997
1	1	1	1	1	1	0	0	1	0	0
2	1	0	1	0	0	0	0	0	0	0
3	0	0	1	1	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	2	1	1	1	1	0	0	0	0	0
6	0	0	0	0	0	1	0	0	0	0
7	3	2	1	2	0	1	0	0	0	1
8	1	1	1	1	1	0	0	0	0	2
9	1	1	1	1	0	1	0	2	0	0
10	0	0	0	0	0	0	0	0	1	0
11	0	0	2	1	2	0	0	0	0	1
12	0	0	0	0	0	0	0	0	0	0
13	1	0	1	1	1	0	0	0	0	1
14	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0
16	1	0	1	1	1	0	0	2	1	2
17	1	0	0	0	0	0	0	0	0	1
18	0	0	0	0	0	0	0	0	0	1
19	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	1	0	0
24	0	0	0	0	0	0	0	0	0	1
25	1	1	2	1	1	0	0	2	1	2
26	3	2	1	3	1	3	0	1	1	1
27	1	0	1	0	0	0	0	0	0	2
28	3	0	1	0	0	1	0	1	0	4
29	1	1	1	1	0	0	0	0	0	5
30	1	0	0	1	0	3	0	0	1	2
31	0	0	0	0	0	3	0	1	2	2
32	0	0	0	0	1	0	0	2	0	5
33	1	0	0	0	2	1	0	1	1	0
34	2	1	1	0	0	4	0	0	0	2
35	0	0	0	0	0	2	0	2	1	3
36	1	0	1	0	0	2	0	0	0	6
37	0	0	0	2	1	1	0	3	0	4
38	0	0	0	0	0	0	0	0	1	1
39	1	0	0	1	0	1	0	2	2	4
40	0	0	0	1	0	0	0	16	1	9
41	1	0	0	7	0	3	0	10	0	7
42	2	0	0	4	0	1	0	6	0	1
43	0	0	0	2	1	0	0	4	0	3
44	0	2	0	5	3	0	0	7	0	1
45	1	0	0	0	0	0	0	5	0	4
46	0	1	0	0	0	1	0	0	0	0
47	2	1	0	0	1	1	0	3	2	0
48	0	0	1	0	1	1	0	0	0	0
49	2	1	0	1	0	0	2	2	0	3
50	2	0	0	0	0	1	0	1	0	0
51	2	0	0	1	1	3	0	1	1	1
52	0	0	0	1	1	1	1	1	0	0

Table 2. (Cont'd.)

Quadrat	Old seedlings					New seedlings				
	1993	1994	1995	1996	1997	1993	1994	1995	1996	1997
53	0	0	0	0	0	2	0	2	0	0
54	3	0	0	2	0	3	0	10	0	0
55	2	0	0	1	1	0	0	8	2	1
56	1	2	0	2	0	1	1	10	0	3
57	0	0	0	1	0	1	0	4	0	1
58	0	0	0	0	0	1	0	1	0	0
59	0	0	0	0	0	0	0	0	0	0
60	1	0	1	0	0	0	0	0	0	0

that f_{new} is close to one. (The possibility that seedlings have emerged but are simply overlooked is small.) But a large value of f_{new} suggests that $N_{i,j}^T$ is likely to be very close to its logical minimum $\max(N_{i,j}, O_{i,j+1} - O_{i,j})$ and therefore that an accurate approximation to the posterior can be had by conditioning on each $N_{i,j}^T$ being equal to its logical minimum. If valid, this leads back to the likelihood in Equation (2.5) but where $N_{i,j}$ is replaced by $\max(N_{i,j}, O_{i,j+1} - O_{i,j})$. The posterior can be computed exactly, with no need for Markov chain Monte Carlo. For our dataset, the exact conditional posterior is an excellent approximation to the unconditional posterior, is very much easier to compute, and is well worth considering for this species when the census can be conveniently timed and when the ecologists are willing to assert a strong prior. If these conditions don't hold, then setting $N_{i,j}^T$ as small as possible is not appropriate.

4.3 CHANGES THROUGH TIME AND SPACE

4.3.1 Arrivals

Several considerations suggest that new seedling arrival rates may vary either spatially or temporally:

- visual inspection of Table 2,
- a plot, not shown here, of $\sum_{j=1}^5 N_{i,j}/5$, the annual average value of $N_{i,j}$ as a function of i ,
- the physical layout of the 60 quadrats in a single 1-m \times 60-m transect along an elevation contour on a hillside,
- spatial and temporal variation in weather, seed rain, rhododendron cover, and other unmeasured covariates.

While all these considerations suggest that λ may vary spatially and temporally, it is not at all clear what form the effects should take or even whether the variation will be smooth or rough. We explored three models in which each (quadrat, year) combination has its own new seedling arrival rate $\lambda_{i,j}$, modeled as the sum of a year effect and a quadrat effect.

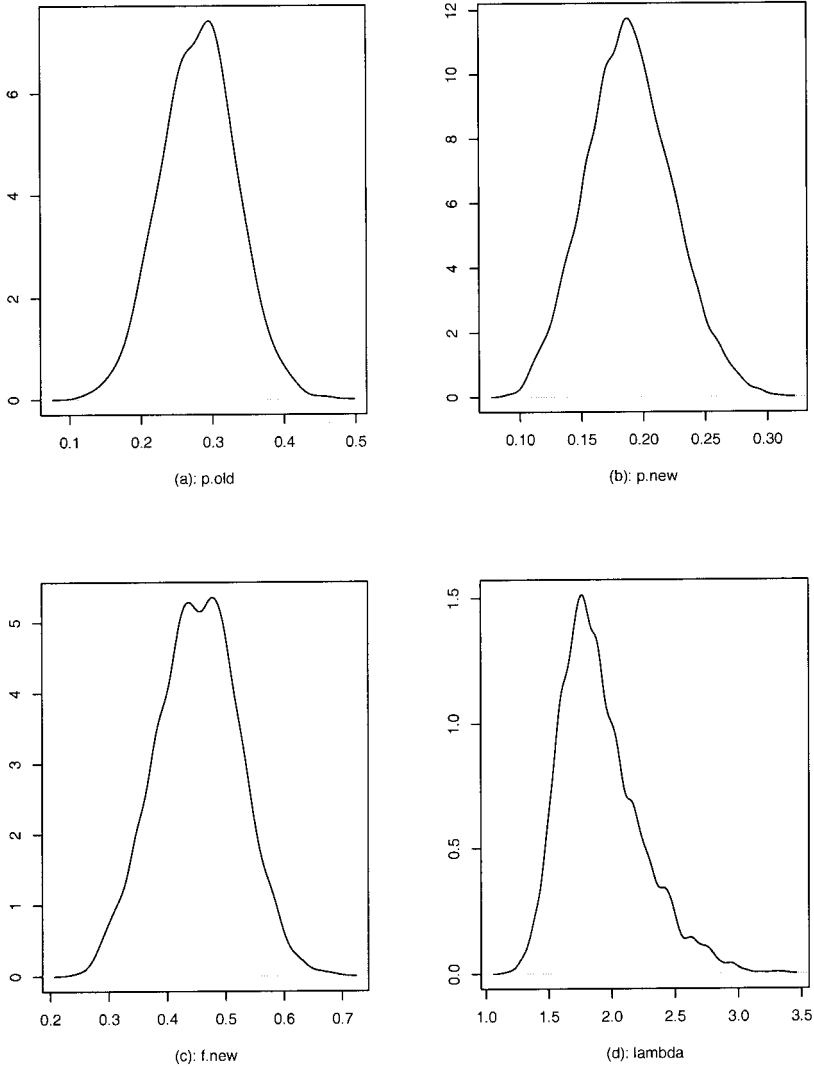


Figure 4. Posterior Densities From Table 2 Data. (a) p_{old} ; (b) p_{new} ; (c) f_{new} ; (d) λ .

The three models incorporate unconstrained (except for sum-to-zero) year effects and have quadrat effects that are either

- (1) unconstrained,
- (2) increasing linearly from left to right, or
- (3) piecewise constant with a low value on quadrats 1–24 and a high value on quadrats 25–60.

Let $\lambda_{i,j}$ be the Poisson arrival rate of new seedlings in plot i and year j . We suppose that $N_{i,j}^T$ is distributed $\text{Poisson}(\lambda_{i,j})$ and adopt a generalized linear model with $\log \lambda_{i,j} = \log \lambda + \alpha_i + \beta_j$. With this refinement, $N_{i,j}$ is distributed $\text{Poisson}(\lambda'_{i,j})$ with $\log \lambda'_{i,j} =$

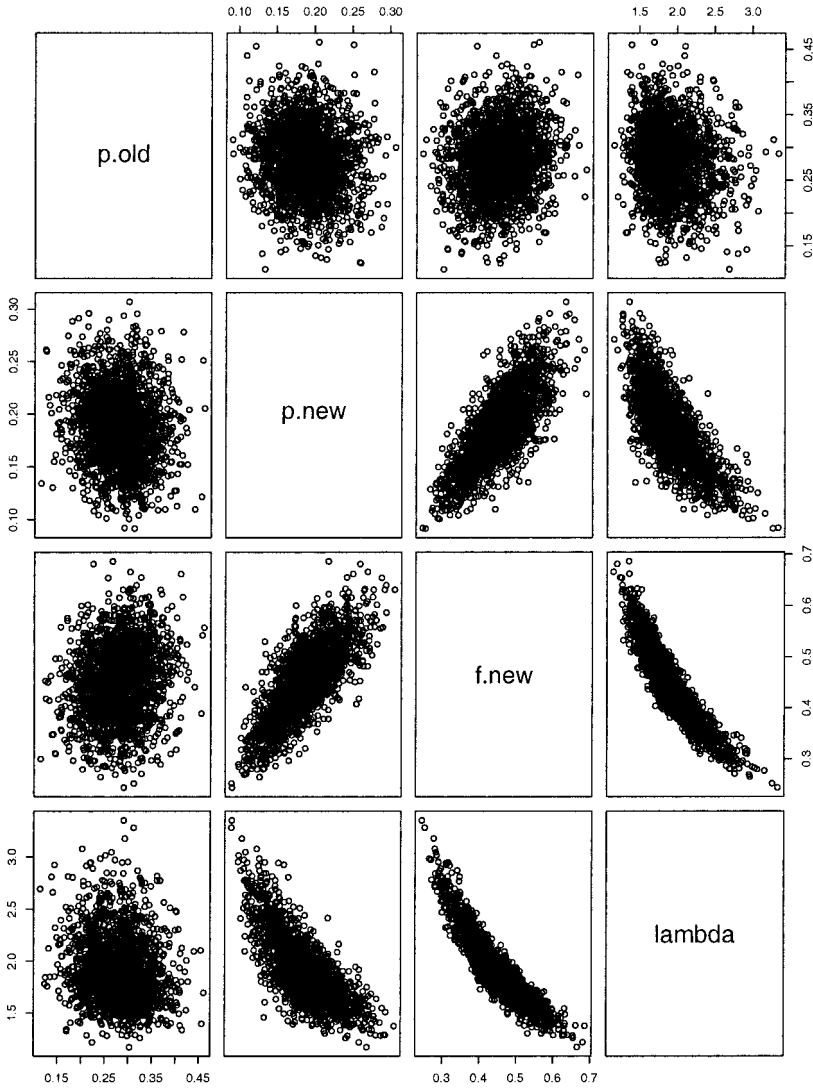


Figure 5. Pairs Plot of Draws From Posterior Densities Using Data From Table 2. Each plot represents one pair of variables. Each point on the plot represents one draw from the posterior.

$\log \lambda' + \alpha_i + \beta_j$. Considering the $N_{i,j}$'s only, λ and f_{new} are not identifiable but the α_i 's and β_j 's are. Standard generalized linear model (GLM) software yields maximum likelihood estimates (MLEs) for the α_i 's and β_j 's. Plots of fitted effects and predicted versus actual numbers of seedlings and analyses of deviance (not shown here) support the use of unconstrained quadrat effects.

We could now incorporate unconstrained quadrat and year effects for λ into our full model. A common choice would be, e.g., a hierarchical structure such as $\log \lambda \sim N(m, s^2)$, $\alpha_i \sim N(0, \sigma_\alpha^2)$, $\beta_j \sim N(0, \sigma_\beta^2)$, suitable priors for σ_α^2 and σ_β^2 , and possibly identifiability or sum-to-zero constraints on the α_i 's and β_j 's. However, almost any sensible, fairly flat

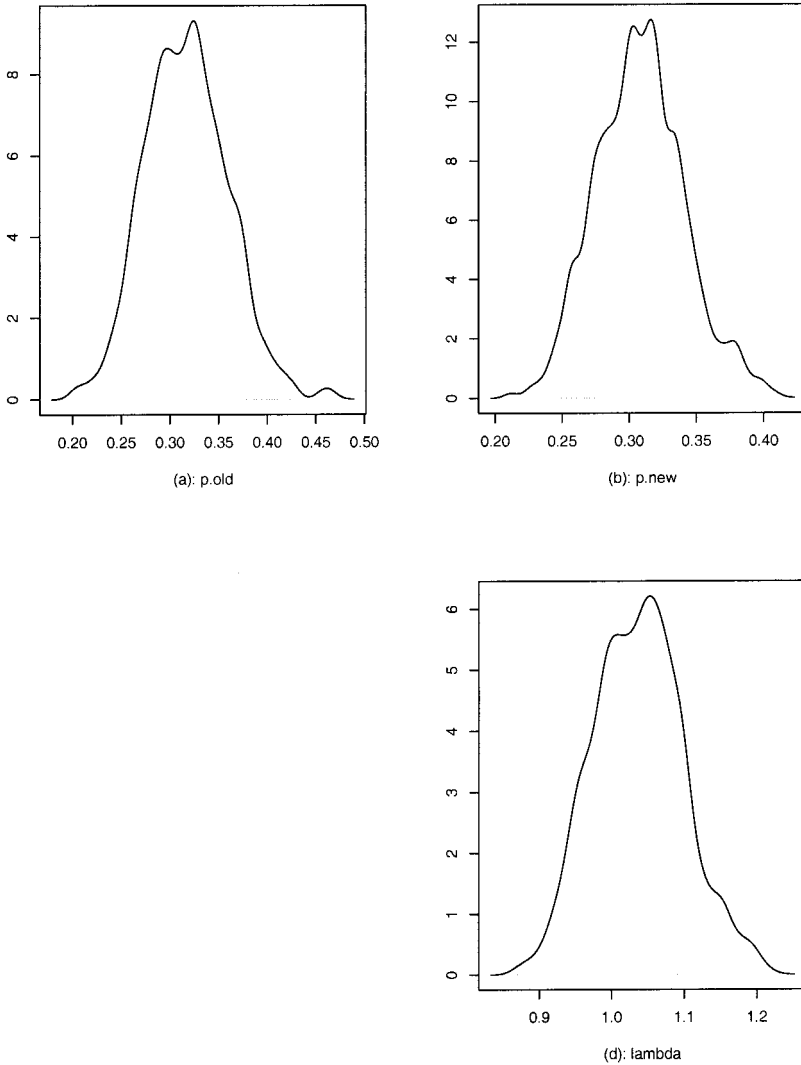


Figure 6. Posterior Densities From Table 2 Data Setting $f_{\text{new}} = 0.9$. (a) p_{old} ; (b) p_{new} ; (d) λ .

prior on the α_i 's and β_j 's should result in shrinkage that will lead to a posterior compromise between setting all the α_i 's and β_j 's equal to zero and setting them all equal to their MLEs. As an exploratory measure, we set them equal to their MLEs, find the posterior for $(p_{\text{old}}, p_{\text{new}})$, and compare it with the posterior in Figures 4 and 5.

Figure 7 shows marginal posterior distributions when the α_i 's and β_j 's are set equal to their MLEs. They are very similar to the marginal posteriors in Figures 4 and 5. We believe that almost any sensible prior on the α_i 's and β_j 's that would be chosen in practice would yield a similar posterior. In this case, better modeling of the emergence process does not strongly affect our posterior for the parameters of interest.

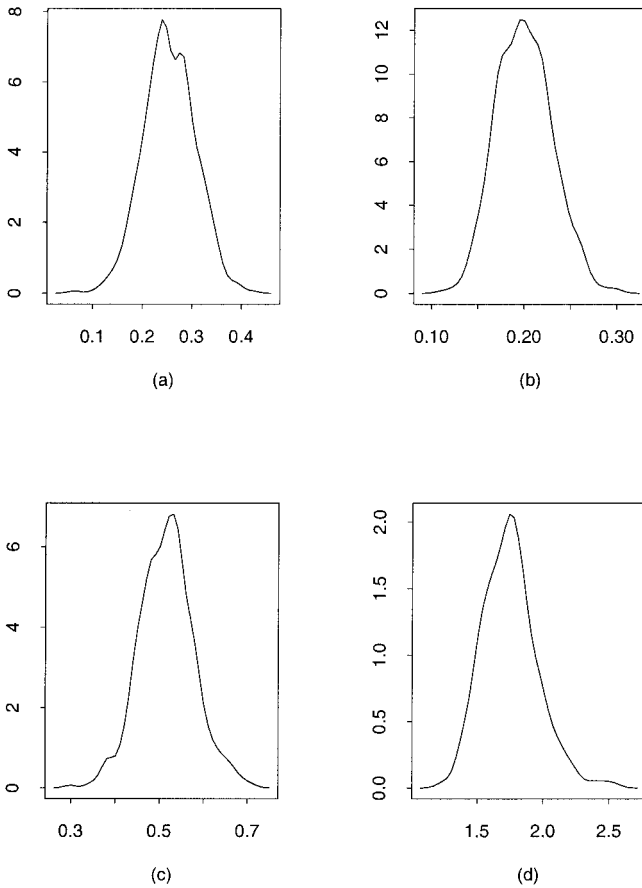
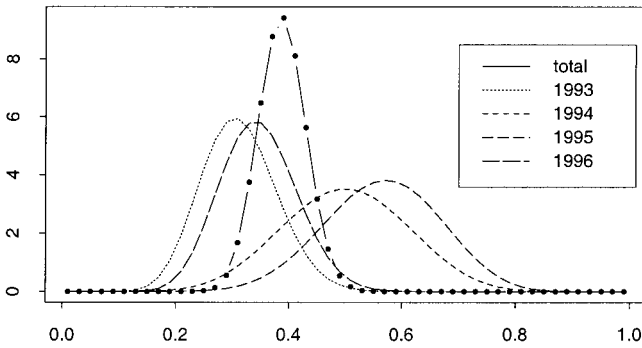


Figure 7. Posteriors From Model With ANOVA Effects for λ . (a) p_{old} ; (b) p_{new} ; (c) f_{new} ; (d) λ .

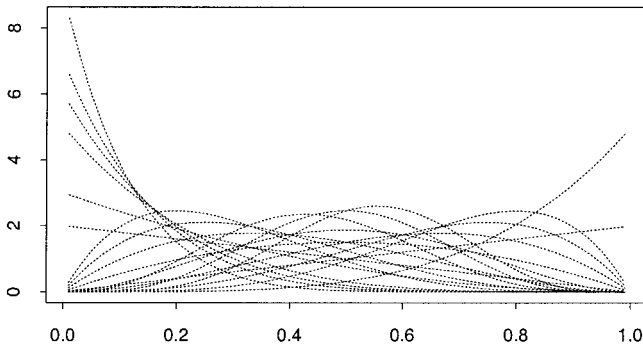
4.3.2 Survival

Handling quadrat and year effects in survival rates is more difficult because we don't know the $X_{i,j}$'s and there is no off-the-shelf software for estimating MLEs. If we did know the $X_{i,j}$'s, then we could average over quadrats (or years) to get crude estimates of survival rates by years (or quadrats) and a rough sense of whether refining the model is likely to be important. To pursue this idea, we temporarily set each $X_{i,j}$ equal to $X_{i,j}^{max}$ simply to make this averaging possible. So, e.g., we attribute the old seedling in quadrat 1, 1996, to the old seedling in quadrat 1, 1995, not the new seedling. As before, we attribute the old seedling in quadrat 2, 1995, to an unobserved new seedling in quadrat 2, 1994. This is roughly equivalent to conditioning on $p_{old} \gg p_{new}$, which we believe to be reasonable for many species in many forests, though not necessarily for *Acer rubrum* in this particular plot at Coweeta.

Let $p_{old,j}$ be the survival rate of old seedlings in year j . The four estimates are $p_{old,1} \approx 14/46 = .30$, $p_{old,2} \approx 9/18 = .5$, $p_{old,3} \approx 12/21 = .57$, and $p_{old,4} \approx 16/47 = .34$.



a



b

Figure 8. (a) Likelihood Plot of p_{old} for Different Years; (b) Likelihood Plot of p_{old} for Different Quadrats.

Ignoring quadrat effects for the moment, we can treat the 14 survivals out of 46 seedlings in year 1 like data from a binomial experiment and calculate a “likelihood” function. That likelihood function is plotted in Figure 8(a), along with the likelihood functions for the other 3 years. The overall impression is that 1993 and 1996 had lower survival rates than 1994 and 1995, although the data from all years is consistent, with a single survival rate near .4. Figure 8(b) shows likelihood functions for quadrat effects. A first impression is of greater differences between quadrats than between years, with most quadrats being consistent with $p_{\text{old}} \approx .4$, but see points 3 and 4 below.

Some points to note are as follows:

- (1) Most of the likelihood curves in Figure 8 are consistent with a single value of $p_{\text{old}} \approx 0.35$.
- (2) Only 40 of the 60 quadrats had at least one old seedling. The other 20 had no old seedlings in the first 4 years and consequently no likelihood function.

- (3) The quadrat likelihood functions most different from the others are those that rise toward the lefthand side of Figure 8(b). The three most extreme are for quadrats 41, 42, and 54, which had 8, 6, and 5 old seedlings, respectively, and no survivors. The old seedlings in these three quadrats all emerged in 1993 and 1996, years with low survival rates. Therefore, an analysis accounting for year and quadrat effects simultaneously would attenuate the apparent quadrat effects.
- (4) The curve that rises most toward the righthand side of Figure 8(b) is from quadrats 1 and 8, which both have new seedlings in every year. The plots are based on the assumption $X_{i,j} = X_{i,j}^{\max}$. Properly accounting for uncertainty in $X_{i,j}$ would tend to flatten the curves in Figure 8, making them all more consistent with each other and with values of $p_{\text{old}} \approx .4$.

We conclude that there is no clear indication whether to include year and quadrat effects on seedling survival; the data seem compatible with the hypothesis of no effects. To see whether their inclusion makes a difference in our ultimate inference about p_{old} , we chose to work with year effects. We fit a model with four different values of p_{old} , one for each year, making appropriate changes to the model. Posterior densities are in Figure 9 and show that inferences about p_{old} and p_{new} do not depend strongly on whether we model year effects for p_{old} .

5. DISCUSSION

We have presented a data collection and modeling approach that allows us to estimate seedling survival probabilities with a fraction of the field labor required by the common practice of flagging all seedlings. Unfortunately, when seedlings are not flagged, the statistical labor increases. Fortunately, for the analysis presented here, the labor does not increase dramatically. Field labor for flagging is the biggest factor limiting the size and scope of data collection and analysis (Clark et al. 1999). By substantially reducing the labor cost, our approach allows for more studies to be undertaken in the future and for those to be more comprehensive.

Inference about p_{new} , the mortality of first year seedlings, appears to be sensitive to the prior. For classical statisticians, we think inference about p_{new} will be sensitive to how well f_{new} and λ can be estimated. One solution is to schedule censuses late enough in the growing season so f_{new} is close to one. Another is to incorporate external information about f_{new} and λ . If external information is not available and if the full census must be done early in the growing season, then revisiting a fraction of the quadrats late in the season would yield information about f_{new} and improve the inference about p_{new} .

Our analysis led to one unanticipated finding: our posterior did not strongly show $p_{\text{old}} > p_{\text{new}}$. There are several possible explanations including data errors, insufficient understanding of biology, and unreasonable prior expectations. We would like to be able to shed more light but our current state of understanding forces us to resort to cliché: more work needs to be done.

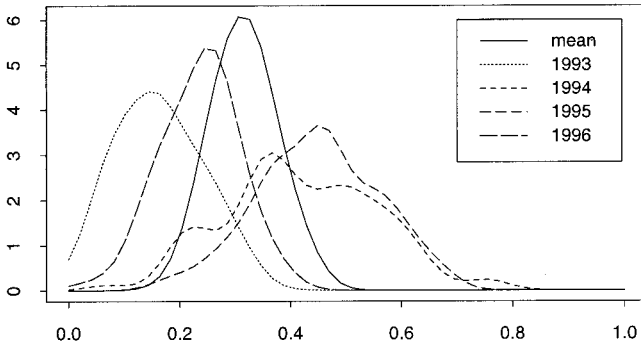
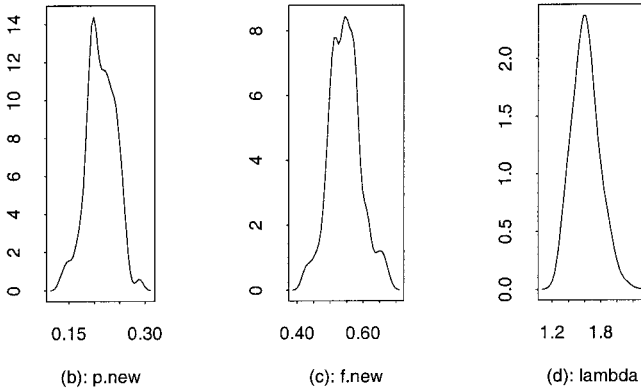
(a): p_{old} (b): p_{new} (c): f_{new} (d): λ

Figure 9. (a) Posterior Densities of p_{old} 's for Four Different Years and Their Mean; (b) Posterior Density of p_{new} ; (c) Posterior Density of f_{new} ; (d) Posterior Density of λ .

We refined the model to account for spatial and temporal effects on arrival and survival probabilities. A common approach is to use a generalized linear model having ANOVA-like quadrat and year effects in the linear predictor. However, we believe it is good practice to explore refinements before adopting them. For our data, we found that more elaborate modeling of arrivals and survivals has little effect on the marginal posteriors of the parameters of interest. We would not have discovered this and would have had less confidence in our results had we simply adopted the refinements without exploring their influence.

By considering only two age classes, our approach does not provide a full age-specific survivorship schedule. A long-term census study could, by flagging each seedling, eventually accumulate age-specific information that would enhance our understanding of seedling mortality. However, based on past experience and the labor limitations inherent in flagging, we do not expect many such data sets in the near future. Our method provides a quick alternative that is more likely to yield data sets of greater spatial and temporal

coverage that will increase our understanding of seedling survival and ultimately our understanding of the survival and spread of past, present, and future populations of trees.

Our method is being used as the basis for studies investigating factors that affect seedling mortality. These studies typically use our model as a base but model p_{old} , p_{new} , and λ with generalized linear models having other factors as covariates. For example, Beckage (2000) studies the effects of canopy gaps and understory shrubs on seedling survival, and work in progress investigates variability in survival across species as well as across elevation gradients.

ACKNOWLEDGMENTS

This research was supported by NSF grants BSR-9444146, DEB 9453498, and DEB-9632854.

[Received February 2000. Accepted January 2001.]

REFERENCES

- Beckage, B. (2000), "A Long-Term Study of Red Maple (*Acer rubrum* L.) Seedling Survival in Southern Appalachian Forests: The Effects of Canopy Gaps and Shrub Understories," Ph.D. thesis, Duke University, Durham, NC.
- Brown, L. D. (1986), *Fundamentals of Statistical Exponential Families With Applications in Statistical Decision Theory*, Hayward, CA: Institute of Mathematical Statistics.
- Casella, G., and George, E. I. (1992), "Explaining the Gibbs Sampler," *The American Statistician*, 46, 167–174.
- Clark, J. S., Beckage, B., Camill, P., Cleveland, B., HilleRisLambers, J., Lichten, J., MacLachlan, J., Mohan, J., and Wyckoff, P. (1999), "Interpreting Recruitment Limitation in Forests," *American Journal of Botany*, 86, 1–16.
- Clark, J. S., Macklin, E., and Wood, L. (1998), "Stages and Spatial Scales of Recruitment Limitation in Southern Appalachian Forests," *Ecological Monographs*, 68, 213–235.
- Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.
- Grubb, P. J. (1977), "The Maintenance of Species-Richness in Plant Communities: The Importance of the Regeneration Niche," *Biological Reviews*, 52, 107–145.
- Hubbell, S. P., Foster, R. B., O'Brien, S. T., Harms, K. E., Condit, R., Wechsler, B., Wright, S. J., and de Lao, S. L. (1999), "Light-Gap Disturbances, Recruitment Limitation, and Tree Diversity in a Neotropical Forest," *Science*, 283, 554–557.
- Jones, R., Sharitz, R., Dixon, P., Segal, D., and Schneider, R. (1994), "Woody Plant Regeneration in Four Floodplain Forests," *Ecological Monographs*, 64, 345–367.
- Kahn, W. D. (1987), "A Cautionary Note for Bayesian Estimation of the Binomial Parameter n ," *The American Statistician*, 41, 38–39.
- Lavine, M., and Wasserman, L. (1992), "Can We Estimate N ?" Discussion Paper 92A-08, Institute of Statistics and Decision Sciences, Duke University, Durham, NC.
- Pacala, S. W., and Tilman, D. (1994), "Limiting Similarity in Mechanistic and Spatial Models of Plant Competition in Heterogeneous Environments," *The American Naturalist*, 143, 222–257.
- Raftery, A. E. (1988), "Inference for the N Parameter: A Hierarchical Bayes Approach," *Biometrika*, 75, 223–228.
- Streg, D., Glitzenstein, J., and Harcombe, P. (1989), "Woody Seedling Dynamics in an East Texas Floodplain Forest," *Ecological Monographs*, 59, 177–204.
- Sundberg, R. (1974), "Maximum Likelihood Theory for Incomplete Data From an Exponential Family," *Scandinavian Journal of Statistics*, 1, 49–58.
- Watt, A. S. (1947), "Pattern and Process in the Plant Community," *Journal of Ecology*, 35, 1–22.