

# Modelling Seedling Mortality

Michael Lavine  
Duke University  
ISDS

Brian Beckage  
Duke University  
Dept. of Biology

James Clark  
Duke University  
Dept. of Biology

September 12, 2000

## Abstract

Seedling mortality in tree populations is important because it limits population growth rates and controls the diversity of forests. To learn about seedling mortality, ecologists return regularly to forest quadrats to count 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 therefore 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 paper 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 results presented here, 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, Experimental Design, Fisher information, Gibbs sampling, Ecological statistics.

## 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 as 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 ecologists.

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

This paper considers the alternative of simply censusing seedlings and not flagging each individual. Our goals are two-fold. First, we want to 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. And second, we want to quantify the information lost by not flagging so ecologists can make well informed decisions about whether to flag. The model is laid out in Section 2 and illustrated on a five year data set from the southern Appalachians (Clark et al., 1998) in Section 4. Information loss is quantified in Section 3.

Because mortality is highly variable in space and time (Beatty, 1984), it is important that ecologists' studies have wide coverage both spatially and temporally. But flagging seedlings is highly labor intensive and is the principal limitation on the spatial and temporal coverage of the current body of studies in the published literature. The vast majority of studies last a year or less and tend to be confined to a single stand of trees. There is a need for data analytic techniques that can extract survivorship information from less labor-intensive data (Clark et al., 1999). Our method, which extracts survivorship information from unflagged census data, will allow existing unflagged census data to be brought to bear on questions of survivorship and will enable ecologists to design future survivorship studies with greater spatial and temporal coverage.

The advantage of not flagging is, of course, ease, which permits the ecologist to census many more quadrats. The disadvantages are three-fold. First, unflagged data are uninformative about age-specific mortality rates. Our methods categorize seedlings as either first year (New) or older (Old) seedlings and provides age-specific mortality rates only for those two cat-

egories. More detailed age-specific rates are often unnecessary for many ecological objectives.

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 even when seedlings are flagged initially, many flags are lost subsequently. Our calculations assume that no flags are lost and therefore overstate the advantage of flagging.

And third, elimination of flags complicates the analysis. Without flags, when we observe a seedling in year  $j$ , we do not know how old it was in year  $j - 1$ . 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 class in year  $j - 1$ . Of course, such an analysis could also be used in a flagged study when flags are lost.

We apply the method to a five year data set from the southern Appalachians (Clark et al., 1998) where we conducted annual censuses of New and Old seedlings in  $1\text{m}^2$  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 First 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 does not depend further on age (Streng et al., 1989; Jones et al., 1994). Therefore we adopt a model with two parameters of interest —  $p_{\text{Old}}$  and  $p_{\text{New}}$  — 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; we take  $\{N_{i,j}\}$  as given. The numbers of Old seedlings recorded in year 1,  $\{O_{i,1}\}$ , are also given. Seedlings are typically small enough so as not to be limited by mutual competition for physical resources; their survival can reasonably be modelled 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

$$X_{i,j} \sim \text{Bin}(O_{i,j}, p_{\text{Old}}) \quad (1)$$

$$Y_{i,j} \sim \text{Bin}(N_{i,j}, p_{\text{New}}). \quad (2)$$

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  $p_{\text{Old}}$  and  $p_{\text{New}}$ ; but see Section 4.3 for year and quadrat effects which 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

$$\begin{aligned} & \Pr[\{X_{i,j}, Y_{i,j}\}_{i=1,\dots,Q; j=1,\dots,T-1} | p_{\text{Old}}, p_{\text{New}}] \\ &= \prod_{i=1}^Q \left\{ \Pr[X_{i,1} | p_{\text{Old}}, p_{\text{New}}] \Pr[Y_{i,1} | p_{\text{Old}}, p_{\text{New}}] \right. \\ & \quad \left. \times \prod_{j=2}^{T-1} \Pr[X_{i,j} | X_{i,j-1}, Y_{i,j-1}, p_{\text{Old}}, p_{\text{New}}] \right\} \\ &= \prod_{i=1}^Q \left\{ \text{Bin}(X_{i,1}; O_{i,1}, p_{\text{Old}}) \text{Bin}(Y_{i,1}; N_{i,1}, p_{\text{New}}) \right. \\ & \quad \left. \times \prod_{j=2}^{T-1} \text{Bin}(X_{i,j}; O_{i,j}, p_{\text{Old}}) \text{Bin}(Y_{i,j}; N_{i,j}, p_{\text{New}}) \right\}. \end{aligned} \quad (3)$$

If seedlings are flagged then, when  $O_{i,j}$  is observed in year  $j$ , we also learn  $X_{i,j-1}$  and  $Y_{i,j-1}$ , 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  $j-1$ . That is, we do not observe  $X_{i,j-1}$  and  $Y_{i,j-1}$ . Some information is lost.

Knowing  $O_{i,j}$  places bounds on  $X_{i,j-1}$  and  $Y_{i,j-1}$ . The lower bound on  $X_{i,j-1}$  is  $X_{i,j-1}^{\min} \equiv \max(0, O_{i,j} - N_{i,j-1})$ ; the upper bound is  $X_{i,j-1}^{\max} \equiv \min(O_{i,j-1}, O_{i,j})$ . Then  $X_{i,j-1} \in [X_{i,j-1}^{\min}, X_{i,j-1}^{\max}]$  and  $Y_{i,j-1} = O_{i,j} - X_{i,j-1}$ .

The model says

$$\begin{aligned}
& \Pr[O_{i,j} | O_{i,j-1}, N_{i,j-1}, p_{\text{Old}}, p_{\text{New}}] \\
&= \sum_{x=X_{i,j-1}^{\min}}^{X_{i,j-1}^{\max}} \left\{ \Pr[X_{i,j-1} = x | O_{i,j-1}, N_{i,j-1}, p_{\text{Old}}, p_{\text{New}}] \right. \\
&\quad \left. \times \Pr[O_{i,j} | X_{i,j-1} = x, O_{i,j-1}, N_{i,j-1}, p_{\text{Old}}, p_{\text{New}}] \right\} \quad (4) \\
&= \sum_{x=X_{i,j-1}^{\min}}^{X_{i,j-1}^{\max}} \text{Bin}(x; O_{i,j-1}, p_{\text{Old}}) \text{Bin}(O_{i,j} - x; N_{i,j-1}, p_{\text{New}}).
\end{aligned}$$

The probability of all the data is the product over quadrats and years of terms like Equation 4:

$$\begin{aligned}
& \Pr[\{O_{i,j} : j \geq 2\} | \{O_{i,1}\}, \{N_{i,j}\}, p_{\text{Old}}, p_{\text{New}}] \\
&= \prod_{i=1}^Q \prod_{j=2}^T \Pr[O_{i,j} | O_{i,j-1}, N_{i,j-1}, p_{\text{Old}}, p_{\text{New}}] \\
&= \prod_{i=1}^Q \prod_{j=2}^T \sum_{x=X_{i,j-1}^{\min}}^{X_{i,j-1}^{\max}} \text{Bin}(x; O_{i,j-1}, p_{\text{Old}}) \text{Bin}(O_{i,j} - x; N_{i,j-1}, p_{\text{New}}). \quad (5)
\end{aligned}$$

To judge our ability to estimate parameters in this model we simulated data from 100 experiments with  $Q = 10$ ,  $T = 5$ ,  $N_{i,j} \sim \text{Poisson}(1)$  and either  $(p_{\text{Old}}, p_{\text{New}}) = (0.8, 0.3)$ , values that we judge plausible for many real forests, or  $(p_{\text{Old}}, p_{\text{New}}) = (0.3, 0.3)$ , values that seem plausible for the data set to be analyzed in Section 4 and that are important for other reasons explained in Section 3. For each simulated data set we found mle's from analyses performed either with or without flags. Figure 1 displays the results. The top and bottom rows show mle's from analyses with and without flags, respectively. Note that the mle's are clustered around the true values but there is more scatter without flags due to loss of information. Section 3 describes the loss of information in more detail. For now we look at mle's to make sure the likelihood function is not leading us astray. Beginning in Section 4 we introduce more parameters and look at posterior distributions. Note that  $Q = 10$  quadrats and  $QT = 50$  quadrat-years are small compared to most experiments in practice, so the simulation is likely a conservative estimate of the practical ability to estimate  $p_{\text{Old}}$  and  $p_{\text{New}}$ .

**Figure 1 about here**

### 3 Information

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

The likelihood function in Equation 5, along with a prior distribution for  $p_{\text{Old}}$  and  $p_{\text{New}}$  yields a joint posterior for  $p_{\text{Old}}$  and  $p_{\text{New}}$ . To understand whether flagging is worth the effort, we compare that posterior to the one we would get from the likelihood function in Equation 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  $p_{\text{Old}}$ ,  $p_{\text{New}}$  and  $\lambda$  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 what values of parameters turned out to be plausible *a posteriori* for our data set. We used independent Beta(1.5,1.5) priors for  $p_{\text{Old}}$  and  $p_{\text{New}}$ ,

$$p(p_{\text{Old}}, p_{\text{New}}) \propto p_{\text{Old}}^5 (1 - p_{\text{Old}})^5 p_{\text{New}}^5 (1 - p_{\text{New}})^5,$$

and calculated lengths of approximate 90% highest posterior density (HPD) regions, that is, the smallest possible intervals containing 90% of the posterior mass for  $p_{\text{Old}}$  and  $p_{\text{New}}$ . Results, jittered for legibility, are in Figure 2. (Figure 1 shows mle's and hence approximate posterior modes under the gentle priors we use here.) 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 2 about here

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  $(p_{\text{Old}}, p_{\text{New}})$  is

$$\begin{bmatrix} \frac{-1}{p_{\text{Old}}(1-p_{\text{Old}})} & 0 \\ 0 & \frac{-1}{p_{\text{New}}(1-p_{\text{New}})} \end{bmatrix}.$$

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

$$\frac{-1}{p_{\text{Old}} + p_{\text{New}} - 2p_{\text{Old}}p_{\text{New}}} \begin{bmatrix} \frac{p_{\text{Old}} + p_{\text{New}}^2 - 2p_{\text{Old}}p_{\text{New}}}{p_{\text{Old}}(1-p_{\text{Old}})} & 1 \\ 1 & \frac{p_{\text{Old}}^2 + p_{\text{New}} - 2p_{\text{Old}}p_{\text{New}}}{p_{\text{New}}(1-p_{\text{New}})} \end{bmatrix}$$

with inverse

$$\frac{1}{(p_{\text{Old}} - p_{\text{New}})^2} \times \begin{bmatrix} -p_{\text{Old}}(1-p_{\text{Old}})(p_{\text{Old}}^2 - 2p_{\text{Old}}p_{\text{New}} + p_{\text{New}}) & p_{\text{Old}}(1-p_{\text{Old}})p_{\text{New}}(1-p_{\text{New}}) \\ p_{\text{Old}}(1-p_{\text{Old}})p_{\text{New}}(1-p_{\text{New}}) & -p_{\text{New}}(1-p_{\text{New}})(p_{\text{New}}^2 - 2p_{\text{Old}}p_{\text{New}} + p_{\text{Old}}) \end{bmatrix}.$$

Large sample standard deviations for  $p_{\text{Old}}$  and  $p_{\text{New}}$ , whether classical or Bayesian, will be roughly proportional to the square roots of the diagonal elements of the inverses of the Fisher information matrices. To help decide whether to flag, we examine the ratio of the standard deviations from flagged and unflagged experiments. That is,

$$\frac{\text{SD}_{\text{flag}}(p_{\text{Old}})}{\text{SD}_{\text{no flag}}(p_{\text{Old}})} \approx \sqrt{\frac{p_{\text{Old}}^2 - 2p_{\text{Old}}p_{\text{New}} + p_{\text{New}}^2}{p_{\text{Old}}^2 - 2p_{\text{Old}}p_{\text{New}} + p_{\text{New}}}} \quad (6)$$

Figure 3 displays this function, which, by symmetry, applies to  $p_{\text{New}}$  as well. Panel (a) is a contour plot; panel (b) displays the ratio as a function of  $p_{\text{Old}}$  for five different values of  $p_{\text{New}}$  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 0 and 1. 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.

**Figure 3 about here**

Sundberg (1974) contains relevant theory.  $(X_{i,j}, Y_{i,j})$  follow an exponential family of distributions. Observing only their sum  $O_{i,j+1}$  is a special case of Sundberg's Example 3 on page 50. Paraphrasing from Sundberg, page 54:

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

Our large sample SD ratio goes to 0 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

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

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,  $p_{\text{New}}$  is generally quite small and less than  $p_{\text{Old}}$ , so the SD ratio under realistic conditions is bounded away from 0. 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 2 but

this time with  $p_{\text{Old}} = p_{\text{New}} = 0.3$ . Figure 4 shows the results. (Figure 1 shows that unflagged experiments are less informative than flagged but doesn't show the loss in relative terms.) Not much has changed for  $p_{\text{Old}}$ . But for  $p_{\text{New}}$  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  $p_{\text{New}}$  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. The ecologists among us (Beckage and Clark) assert that this penalty is a small price to pay for the convenience of not flagging.

The relative efficiency of flagged and unflagged experiments depends on the true values of  $p_{\text{Old}}$  and  $p_{\text{New}}$  and on whether  $p_{\text{New}}$  is a parameter of interest. The efficiency is worst when  $p_{\text{Old}} \approx p_{\text{New}}$ , in which case, of course, a model with just a single survival probability will work well and lose no information. Even so, the relative efficiency observed in our simulations is better than asymptotic theory suggests. We think that's because when  $p_{\text{Old}} \approx p_{\text{New}}$  a very large sample size, much larger than practicable, is required for the asymptotic theory to be a good approximation. Therefore we recommend simulation as the preferred method for calculating relative efficiency.

**Figure 4 about here**

## 4 Analysis

This section presents our analysis of data collected annually on red maple (*Acer rubrum*) seedlings from  $1\text{m}^2$  quadrats arranged along transects in the southern Appalachians (Clark et al., 1998). The goal is to demonstrate a Bayesian method for finding posterior distributions for  $p_{\text{Old}}$  and  $p_{\text{New}}$ . Because our main purpose is to show a method, we will concentrate on what is unique to our application and will deemphasize technical aspects of computations. For example, we don't try to develop the best possible MCMC transition scheme or delve into details of monitoring convergence but refer the reader to books such as Gelman et al. (1995) and Robert and Casella (1999).

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 five year period in quadrat 2 at the Coweeta Experimental Station in western North Carolina.

**Table 1 about here**

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. Here are some of the ways in which such errors might arise.

1. The species is misidentified.
2. A New seedling is misclassified as Old in year 3.
3. A seedling is near the quadrat boundary and is recorded as inside the quadrat in some years and outside the quadrat in other years.
4. A seedling was overlooked in year 2.
5. A New seedling emerged later in the season of year 2, after the July census.

We judge errors of types 5 to be the most likely in practice. Errors of type 4 are much more likely to happen to New seedlings than to Old, in which case they are indistinguishable from errors of type 5. We 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 us 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  $f_{\text{New}}$  is the probability that a given New seedling is present and recorded at the time of the census. That is,  $N_{i,j}$  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, for example, the discussion in Kahn (1987), Raftery (1988) or Lavine and Wasserman (1992). We also change Equation 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

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

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:

$$\begin{aligned} p_{\text{Old}} &\sim \text{Beta}(\alpha_{p_{\text{Old}}}, \beta_{p_{\text{Old}}}) \\ p_{\text{New}} &\sim \text{Beta}(\alpha_{p_{\text{New}}}, \beta_{p_{\text{New}}}) \\ f_{\text{New}} &\sim \text{Beta}(\alpha_{f_{\text{New}}}, \beta_{f_{\text{New}}}) \\ \{N_{i,j}^T\} | \lambda &\sim \text{i.i.d. Poi}(\lambda) \\ \lambda &\sim \text{Gam}(\alpha_{\lambda}, \beta_{\lambda}). \end{aligned}$$

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

1.  **$p_{\text{Old}}$** : Each quadrat in each year yields a binomial experiment with  $O_{i,j}$  number of trials,  $X_{i,j}$  number of successes and  $p_{\text{Old}}$  probability. Therefore the complete conditional distribution for  $p_{\text{Old}}$  is

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

2.  $\mathbf{p}_{\text{New}}$ : 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  $p_{\text{New}}$  probability. Therefore the complete conditional distribution for  $p_{\text{New}}$  is

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

3.  $\mathbf{f}_{\text{New}}$ : Each quadrat in each year yields a binomial experiment with  $N_{i,j}^T$  number of trials,  $N_{i,j}$  number of successes and  $f_{\text{New}}$  probability. Therefore the complete conditional distribution for  $f_{\text{New}}$  is

$$\text{Beta}(\alpha_{f_{\text{New}}} + \sum_1^Q \sum_1^T N_{i,j}, \beta_{f_{\text{New}}} + \sum_1^Q \sum_1^T (N_{i,j}^T - N_{i,j})).$$

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

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

5.  $\mathbf{X}_{i,j}$ : 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 page 4.) 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.  $\mathbf{N}_{i,j}^T$ : Each  $N_{i,j}^T$  is a realization of a Poisson random variable with parameter  $\lambda$  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, so the upper bound makes little practical difference.

To test the method we created 100 artificial data sets with

$$\begin{aligned} Q &= 50 \\ T &= 5 \\ p_{\text{Old}} &= 0.3 \\ p_{\text{New}} &= 0.3 \\ f_{\text{New}} &= 0.6 \\ \lambda &= 1 \end{aligned}$$

and used prior parameters

$$\begin{aligned} (\alpha_{p_{\text{Old}}}, \beta_{p_{\text{Old}}}) &= (1, 1) \\ (\alpha_{p_{\text{New}}}, \beta_{p_{\text{New}}}) &= (1, 1) \\ (\alpha_{f_{\text{New}}}, \beta_{f_{\text{New}}}) &= (1, 1) \\ (\alpha_{\lambda}, \beta_{\lambda}) &= (.4, .1). \end{aligned}$$

For each data set we ran a Gibbs sampler, discarded the first 500 iterations and estimated posterior means from the next 500. (We don't suggest that 1000 iterations are sufficient for a careful analysis. In this simulation we are only trying to get a rough idea of whether our analysis is leading us astray.) For each parameter  $p_{\text{Old}}$ ,  $p_{\text{New}}$ ,  $f_{\text{New}}$  and  $\lambda$ , Figure 5 shows a histogram of the 100 posterior means. The main point to notice is that the posterior means are clustered around the true values and don't seem to be misleading in any systematic way.

**Figure 5 about here**

We judge the results in Figure 5 to be satisfactory and turn next to real data. Table 2 shows data from 60 quadrats at Coweeta. Figure 6 shows posterior density estimates based on output from 2000 iterations of a Gibbs sampler, collected at intervals of 50 after discarding the initial 500. Note that Figure 6, panel (d) shows that our posterior for  $\lambda$  is far from the prior mean of 4, showing that our posterior has picked up a signal from the data and is not preordained to mimic the prior. Figure 6 is important not only for what it says about marginal posterior densities, but also as a basis for comparison later in this paper. Sections 4.2 and 4.3 explore sensitivities and figures in those sections will be compared to Figure 6 when the prior and the model are elaborated.

Figure 7 shows plots of parameter values against iteration number from the Gibbs sampler. Autocorrelation functions, not shown here, decay to approximately 0 within about 10 iterations (500 iterations counting the skipped ones) so the 2000 draws we use contain at least as much information as 200 independent draws from the posterior, enough for our purpose of sketching posterior densities.

**Table 2 about here**

**Figure 6 about here**

**Figure 7 about here**

## 4.2 Sensitivity to the Prior

Figure 8 is a pairs plot of draws from the joint posterior distribution of  $p_{\text{Old}}$ ,  $p_{\text{New}}$ ,  $f_{\text{New}}$ , and  $\lambda$ . The last three parameters are highly dependent on each other but relatively independent of  $p_{\text{Old}}$ . 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  $\lambda$ ; we observe each one with probability  $f_{\text{New}}$ ; therefore our observations  $\{N_{i,j}\}$  are Poisson with rate  $\lambda' = \lambda f_{\text{New}}$ ; and therefore the data can pin down  $\lambda'$  fairly well, but not each parameter separately. In fact, the joint posterior density of  $f_{\text{New}}$  and  $\lambda$  has a ridge along the hyperbola  $\lambda f_{\text{New}} = 0.843$ , the posterior mean of  $\lambda'$ .

The joint posterior of  $p_{\text{New}}$  and  $\lambda$  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  $\lambda''$ , so  $\lambda''$  is roughly determined by the number of New seedling survivals needed for the model to be a good fit for the data. And  $p_{\text{New}}$  and  $f_{\text{New}}$  are positively correlated with each other because they are each negatively correlated with  $\lambda$ .

The structure of the joint posterior suggests that MCMC algorithms will be more efficient if they include steps taken along the hyperbolae. However, the straightforward MCMC presented in Section 4.1 appears to explore the parameter space adequately. Since our concern is with finding the posterior and not with computational efficiency we do not further pursue the idea of alternate MCMC proposals.

**Figure 8 about here**

The foregoing discussion suggests that the joint likelihood of  $p_{\text{New}}$ ,  $f_{\text{New}}$  and  $\lambda$  is sharpest in the two directions corresponding to  $\lambda'$  and  $\lambda''$  but relatively flat in other directions; inspection of Figure 8 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  $p_{\text{Old}}$  and  $p_{\text{New}}$ , the parameters of primary interest, to the prior for  $f_{\text{New}}$ . As an extreme exploration, we calculated the posterior conditional on  $f_{\text{New}} = 0.9$ . Figure 9 shows the marginal posteriors. As feared, this changes the posterior of  $p_{\text{New}}$ . It also changes the posterior of  $\lambda$ , but that is of secondary interest, and makes minor changes to the posterior of  $p_{\text{Old}}$ .

**Figure 9 about here**

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  $f_{\text{New}}$  and  $\lambda$ . 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 that  $f_{\text{New}}$  is close to 1. (The possibility that seedlings have emerged but are simply overlooked is judged to be small by the census takers among us.) But a large value of  $f_{\text{New}}$  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 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 MCMC. Figure 10 shows (a): a contour plot of the posterior calculated this way, in contrast to (b): a scatter plot of Gibbs samples from the posterior with  $f_{\text{New}} = 0.9$ . The two posteriors for  $(p_{\text{Old}}, p_{\text{New}})$  are very similar to each other. In this case at least, 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.

**Figure 10 about here**

### 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,
- Figure 11(a), a plot 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  $1\text{m} \times 60\text{m}$  transect along an elevation contour on a hillside,
- quadrat differences in crucial factors such as the amount of rhododendron cover, and
- spatial and temporal variation in weather, seed rain, and other unmeasured covariates.

While all these considerations suggest that  $\lambda$  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 will explore three models in which each (quadrat, year) combination has its own New seedling arrival rate  $\lambda_{i,j}$ , modelled as the sum of a year effect and a quadrat effect. The three models incorporate unconstrained (except for sum-to-0) 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  $\log \lambda'_{i,j} = \log \lambda' + \alpha_i + \beta_j$ . Considering the  $N_{i,j}$ 's only,  $\lambda$  and  $f_{\text{New}}$  are not identifiable, but the  $\alpha_i$ 's and  $\beta_j$ 's are. Standard glm software yields mle's for the  $\alpha_i$ 's and  $\beta_j$ 's. Figures 11(b), (c) and (d) plot  $\hat{\alpha}_i$  as a function of  $i$  when  $\alpha$  is (b) unconstrained, (c) a linear function of  $i$  and (d) a piecewise constant function of  $i$ , respectively. Figures 12(a), (b) and (c) plot actual versus fitted  $N_{i,j}$  from the three models. Figure 12 shows that unconstrained effects can substantially reduce the prediction error in cases with large numbers of seedlings. The figures and analysis of deviance (not shown here) support the use of unconstrained quadrat effects.

**Figure 11 about here**

**Figure 12 about here**

We could now incorporate unconstrained quadrat and year effects for  $\lambda$  into our full model. A common choice would be, for example, 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  $\sigma_\alpha^2$  and  $\sigma_\beta^2$  and possibly identifiability or sum-to-zero constraints on the  $\alpha_i$ 's and  $\beta_j$ 's. However, almost any sensible, fairly flat prior on the  $\alpha_i$ 's and  $\beta_j$ 's should result in shrinkage that will lead to a posterior compromise between setting all the  $\alpha_i$ 's and  $\beta_j$ 's equal to 0 and setting them all equal to their mle's. As an exploratory measure we set them equal to their mle's, find

the posterior for  $(p_{\text{Old}}, p_{\text{New}})$ , and compare it to the posterior in Figures 6 and 8. The complete conditional distribution for  $\lambda$  must be changed. Item 4 on page 12 should be changed to

4'.  $\lambda$ : The terms in the joint distribution involving  $\lambda$  are

$$\prod_{i=1}^Q \prod_{j=1}^T \exp(-\lambda e^{\alpha_i + \beta_j}) (\lambda e^{\alpha_i + \beta_j})^{N_{i,j}^T} \propto e^{-\lambda \sum \sum e^{\alpha_i + \beta_j}} \lambda^{\sum \sum N_{i,j}^T}.$$

Therefore the complete conditional distribution for  $\lambda$  is

$$\text{Gam} \left( \alpha_\lambda + \sum \sum N_{i,j}^T, \beta_\lambda + \sum \sum e^{\alpha_i + \beta_j} \right).$$

Figure 13 shows marginal posterior distributions when the  $\alpha_i$ 's and  $\beta_j$ 's are set equal to their mle's. They are very similar to the marginal posteriors in Figures 6 and 8. We believe that almost any sensible prior on the  $\alpha_i$ 's and  $\beta_j$ 's that would be chosen in practice would yield a similar posterior. In this case, better modelling of the emergence process does not strongly affect our posterior for the parameters of interest.

**Figure 13 about here**

### 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 mle's. 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  $X_{i,j}^{\text{max}}$ , not for any scientific reason but simply to make this averaging possible. So, for example, 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  $p_{\text{Old}} \gg p_{\text{New}}$ , which we believe to be reasonable for many species in many forests, though not necessarily for *Acer rubrum* in this particular plot at Coweeta. Table 3 shows the numbers of Old seedlings and  $X_{i,j}^{\text{max}}$  in each of the first four years.

### Table 3 about here

Let  $p_{\text{Old},j}$  be the survival rate of Old seedlings in year  $j$ . The four estimates are  $p_{\text{Old},1} \approx 14/46 = .30$ ,  $p_{\text{Old},2} \approx 9/18 = .5$ ,  $p_{\text{Old},3} \approx 12/21 = .57$  and  $p_{\text{Old},4} \approx 16/47 = .34$ . 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 14(a), along with the likelihood functions for the other three 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 0.4. Figure 14(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:

1. Figure 10 suggests that small values of  $p_{\text{New}}$  are consistent with  $p_{\text{Old}} \approx 0.35$ , a value which is also consistent with most of the likelihood curves in Figure 14.
2. Only 40 of the 60 quadrats had at least one Old seedling. The other 20 had no Old seedlings in the first four 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 14(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 14(b) is from quadrats 1 and 8 which both have New seedlings in every year. The plots are based on the assumption  $X_{i,j} = X_{i,j}^{\text{max}}$ . Properly accounting for uncertainty in  $X_{i,j}$  would tend to flatten the curves in Figure 14, making them all more consistent with each other and with values of  $p_{\text{Old}} \approx 0.4$ .

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  $p_{\text{Old}}$ , we chose to work with year effects because we believe them to be biologically more important. We fit a model with four different values of  $p_{\text{Old}}$ , one for each year, making appropriate changes to items 1 and 5 on page 11. Posterior densities are in Figure 15 and show that inferences about  $p_{\text{Old}}$  and  $p_{\text{New}}$  do not depend strongly on whether we model year effects for  $p_{\text{Old}}$ .

**Figure 14 about here**

**Figure 15 about here**

## 5 Discussion

We have presented a data collection and modelling 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. 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  $p_{\text{New}}$  appears to be sensitive to the prior. For classical statisticians, we think inference about  $p_{\text{New}}$  will be sensitive to how well  $f_{\text{New}}$  and  $\lambda$  can be estimated. One solution is to schedule censuses late enough in the growing season so  $f_{\text{New}}$  is close to 1. Another is to incorporate external information about  $f_{\text{New}}$  and  $\lambda$ . 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  $f_{\text{New}}$  and improve the inference about  $p_{\text{New}}$ .

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*.

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. Modern computational tools make it possible to find the posterior distribution from such a model, although that would require writing code in a low level language. (We could not make the model run in BUGS (Spiegelhalter et al., 1999) and it runs too slowly in S-Plus (Chambers, 1998).) However, we believe it is good practice to explore refinements before adopting them. For our data we found that more elaborate modelling 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 accumulated 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 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  $p_{\text{Old}}$ ,  $p_{\text{New}}$  and  $\lambda$  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.

## References

Beatty, S. W. (1984). Influence of microtopography and canopy species on spatial patterns of forest understory plants. *Ecology*, 65:1406–1419.

- 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*. PhD thesis, Duke University.
- Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. Institute of Mathematical Statistics, Hayward, CA.
- Casella, G. and George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*, 46:167–174.
- Chambers, J. M. (1998). *Programming with Data*. Springer-Verlag, New York.
- Clark, J. S., Beckage, B., Camill, P., Cleveland, B., HilleRisLambers, J., Lichter, 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.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman & Hall, London.
- 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(3):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.
- 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.
- Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer-Verlag, New York.
- Spiegelhalter, D. J., Thomas, A., and Best, N. G. (1999). *WinBUGS Version 1.2 User Manual*. MRC Biostatistics Unit, Cambridge, UK.
- Streng, D., Glitzenstein, J., and Harcombe, P. (1989). Woody seedling dynamics in an east Texas floodplain forest. *Ecological Monographs*, 59(2):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.

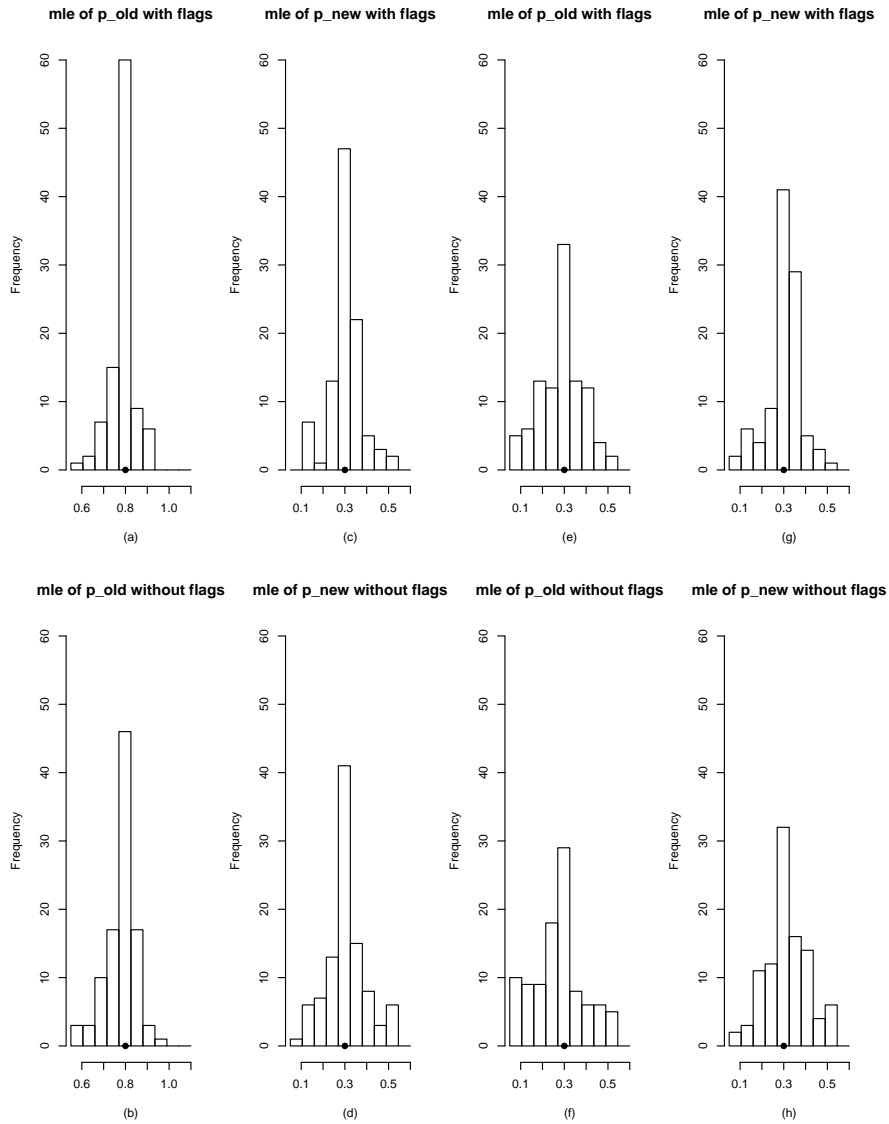


Figure 1: mle's of  $p_{Old}$  and  $p_{New}$  from analyses of simulated data with and without flags. True values are marked by dots on the abscissa. Panels (a)-(d) were generated with  $(p_{Old}, p_{New}) = (0.8, 0.3)$ ; (e)-(h) were generated with  $(p_{Old}, p_{New}) = (0.3, 0.3)$

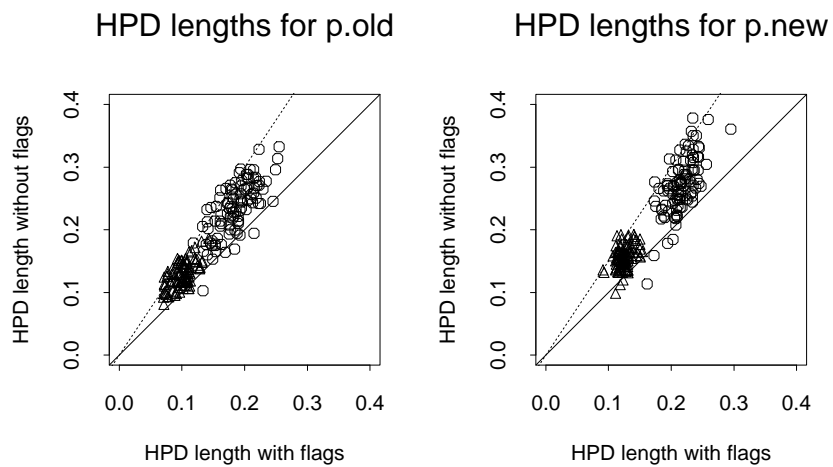


Figure 2: Lengths of HPD regions from simulated experiments. Octagons are from experiments with  $Q = 10$ ; triangles are  $Q = 30$ . The solid line has slope 1; the dotted line has slope 1.5. Data are jittered for legibility.

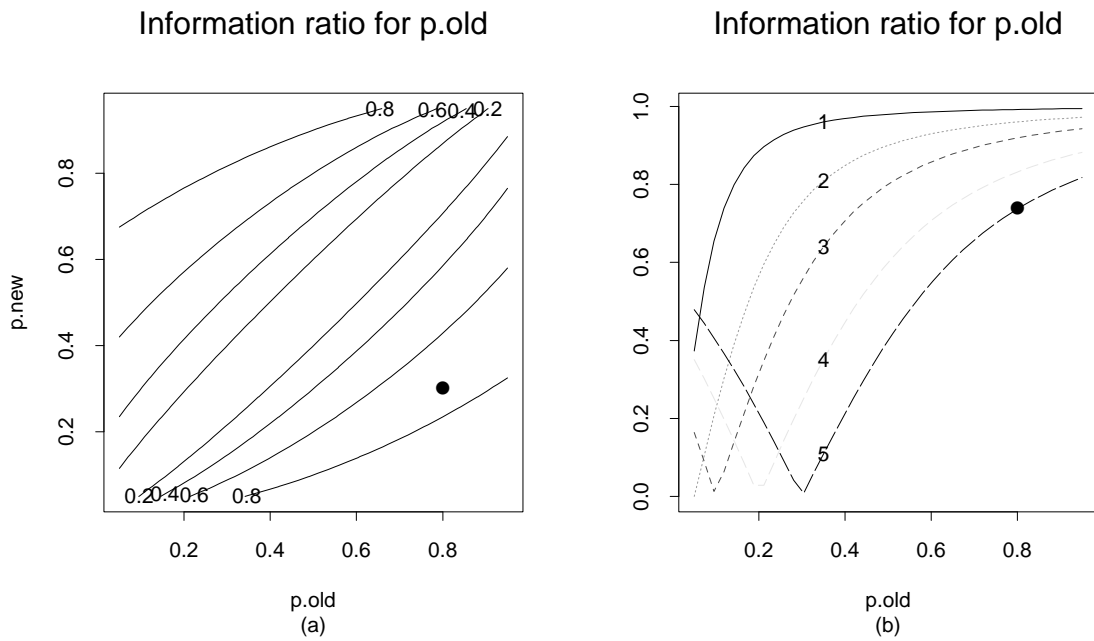


Figure 3:  $SD_{\text{flag}}(p_{\text{Old}})/SD_{\text{no flag}}(p_{\text{Old}})$  as in Eq. 6

(a): as a function of  $p_{\text{Old}}$  and  $p_{\text{New}}$

(b): as a function of  $p_{\text{Old}}$  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.  $p_{\text{New}} = 0.30$ .

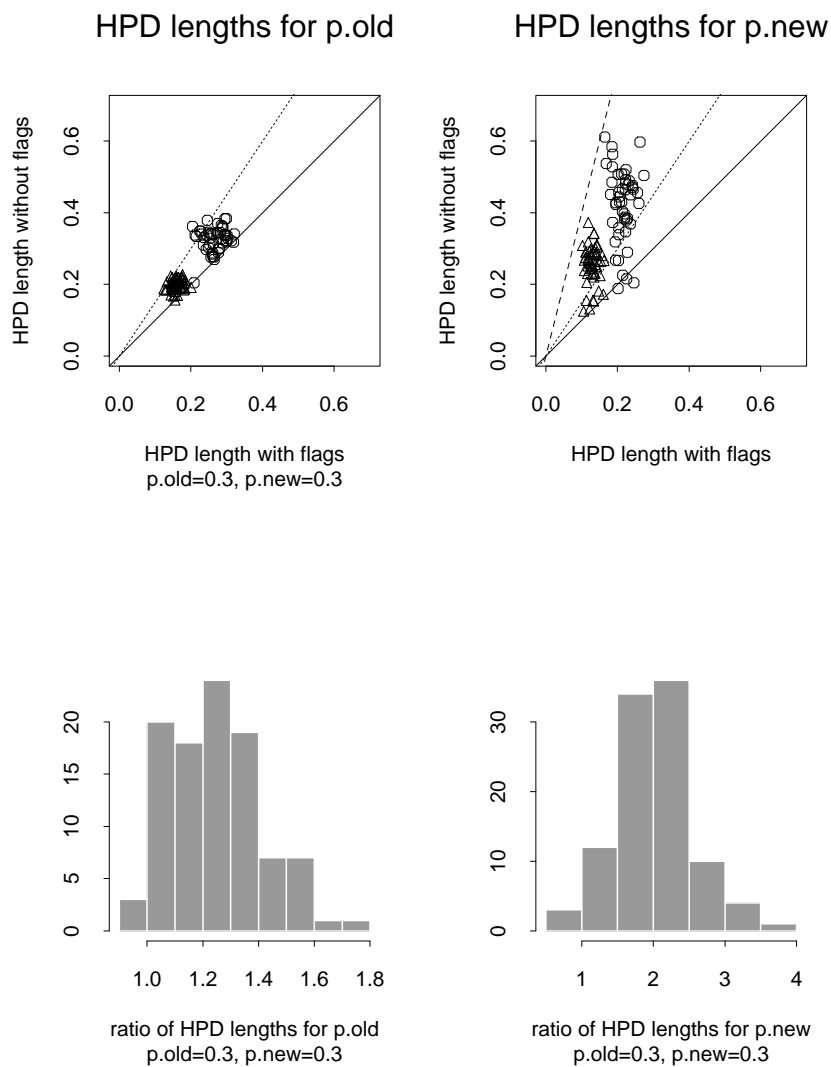


Figure 4: Lengths of HPD regions from simulated experiments. Octagons are from experiments with  $Q = 10$ ; triangles are  $Q = 30$ . The solid line has slope 1; the dotted line has slope 1.5; the dashed line has slope 4. Data are jittered for legibility. Histograms of HPD length from unflagged analysis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

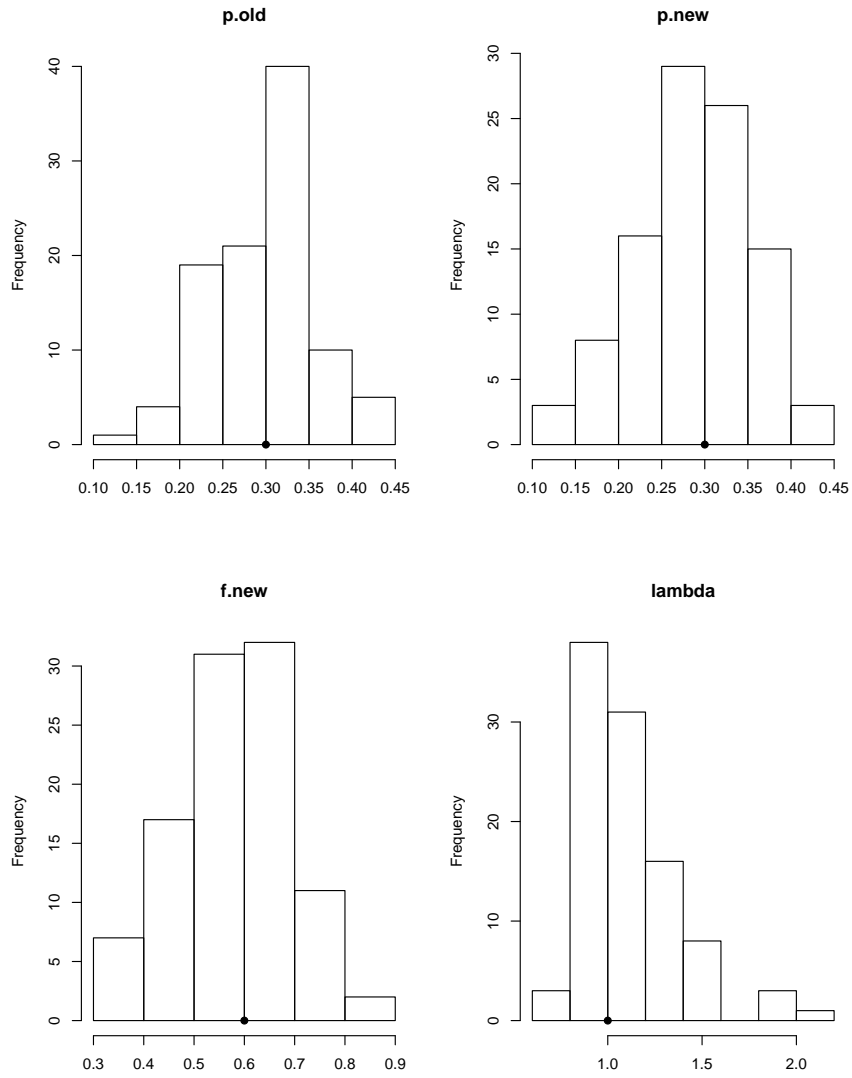


Figure 5: Posterior means from 100 artificial data sets. (a):  $p_{\text{Old}}$ ; (b):  $p_{\text{New}}$ ; (c):  $f_{\text{New}}$ ; (d):  $\lambda$ . Dots on the abscissae are the values used for simulation.

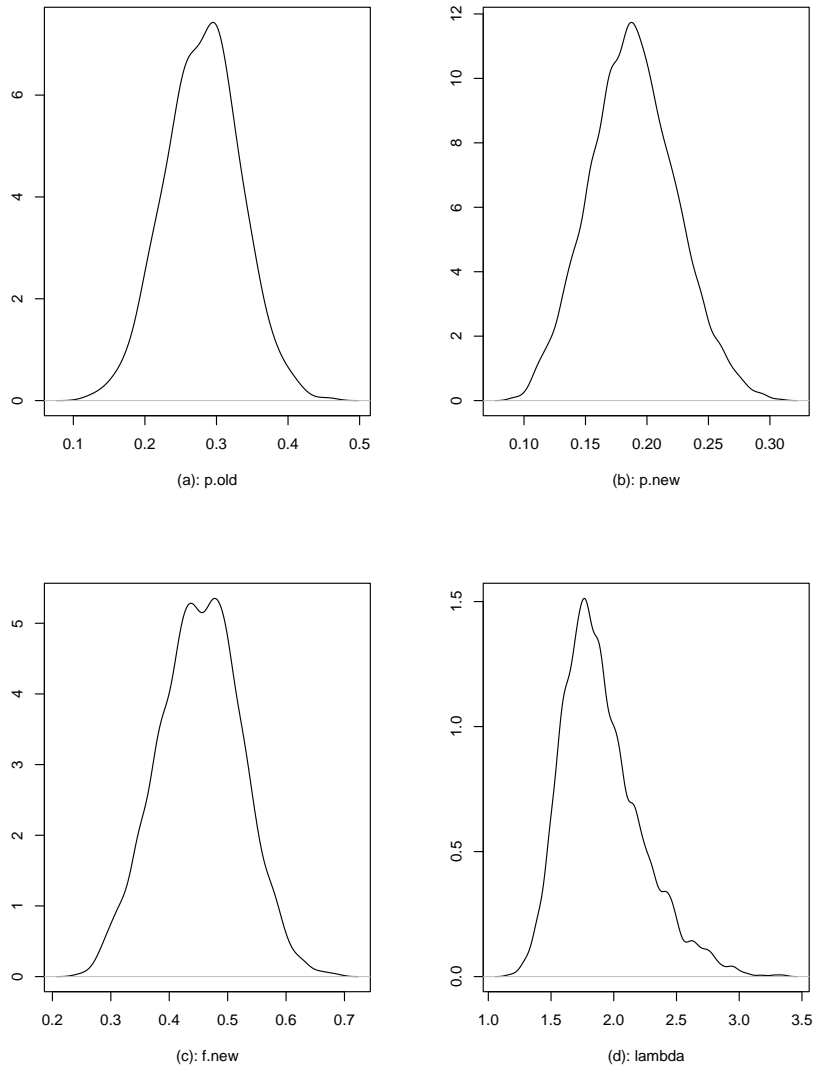


Figure 6: Posterior densities from Table 2 data. (a):  $p_{Old}$ ; (b):  $p_{New}$ ; (c):  $f_{New}$ ; (d):  $\lambda$ .

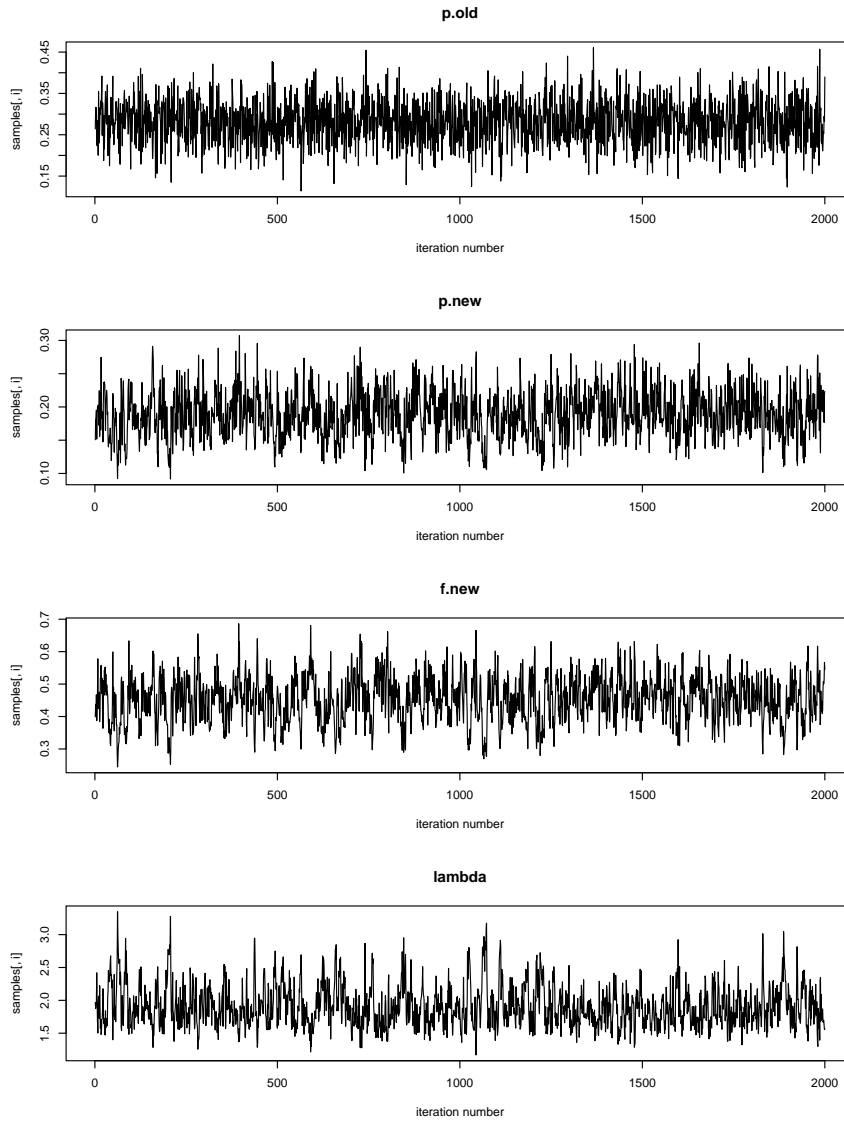


Figure 7: Trace plots of parameter values versus iteration numbers. (a):  $p_{Old}$ ; (b):  $p_{New}$ ; (c):  $f_{New}$ ; (d):  $\lambda$ .

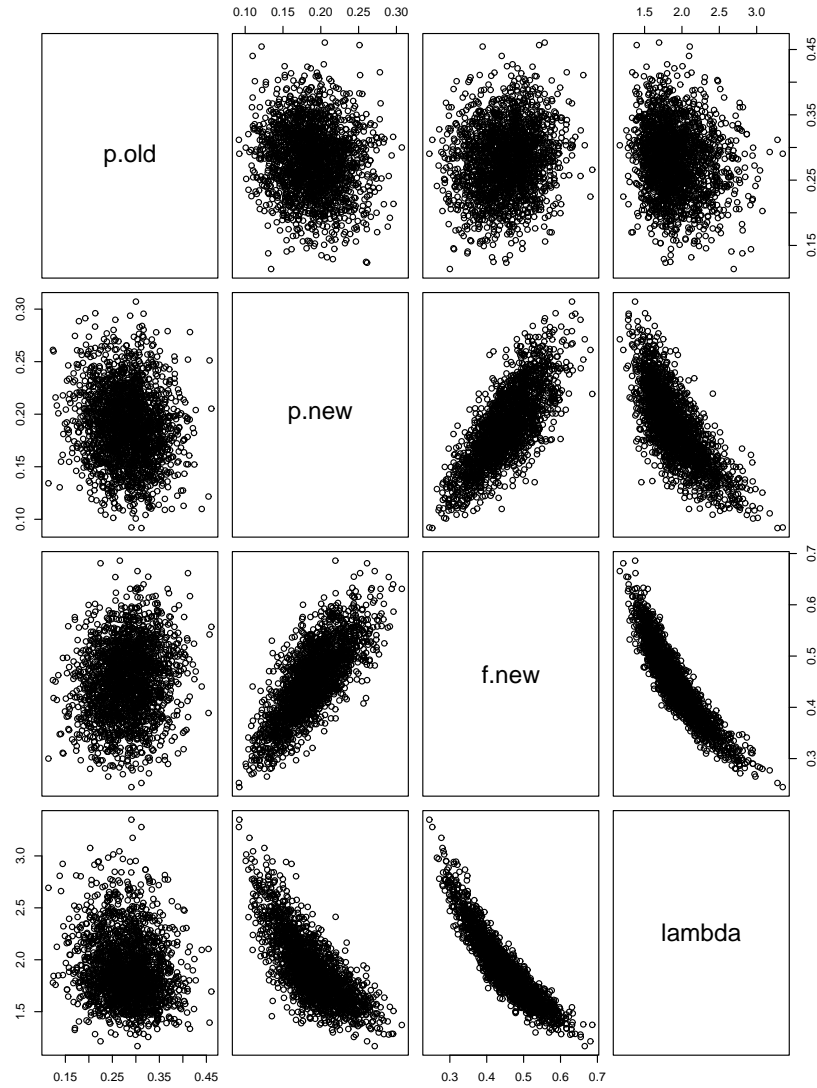


Figure 8: Pairs plot of draws from posterior densities using data from Table 2 data. Each plot represents one pair of variables. Each point on the plot represents one draw from the posterior.

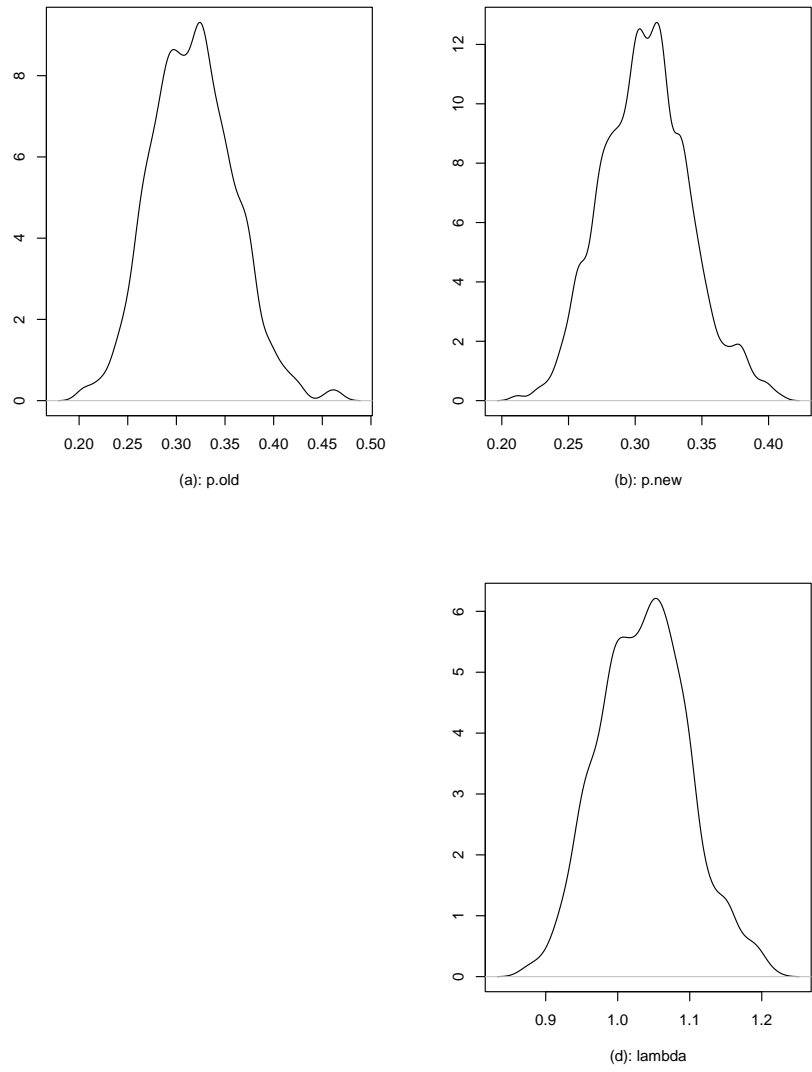


Figure 9: Posterior densities from Table 2 data, setting  $f_{\text{New}} = 0.9$ . (a):  $p_{\text{Old}}$ ; (b):  $p_{\text{New}}$ ; (d):  $\lambda$ .

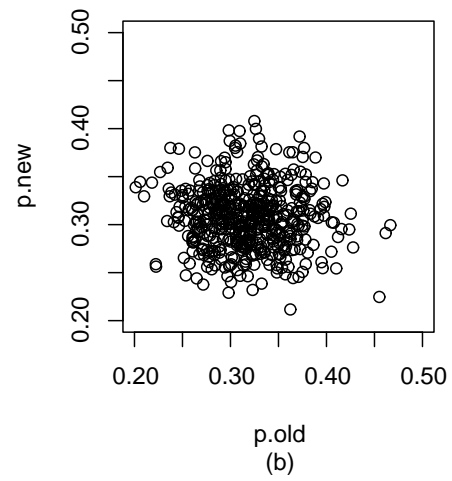
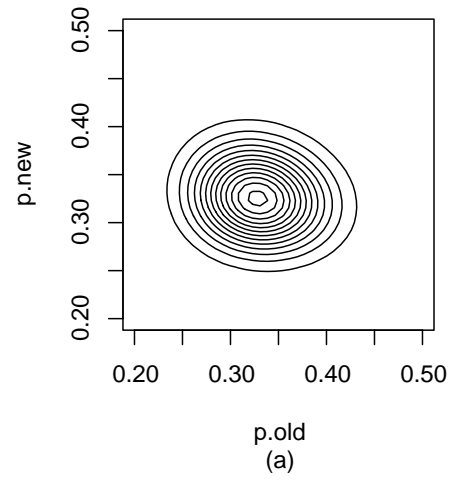


Figure 10: (a): Contour plot of posterior density from exact calculation. (b): Gibbs samples of  $(p_{\text{Old}}, p_{\text{New}} | f_{\text{New}} = 0.9)$

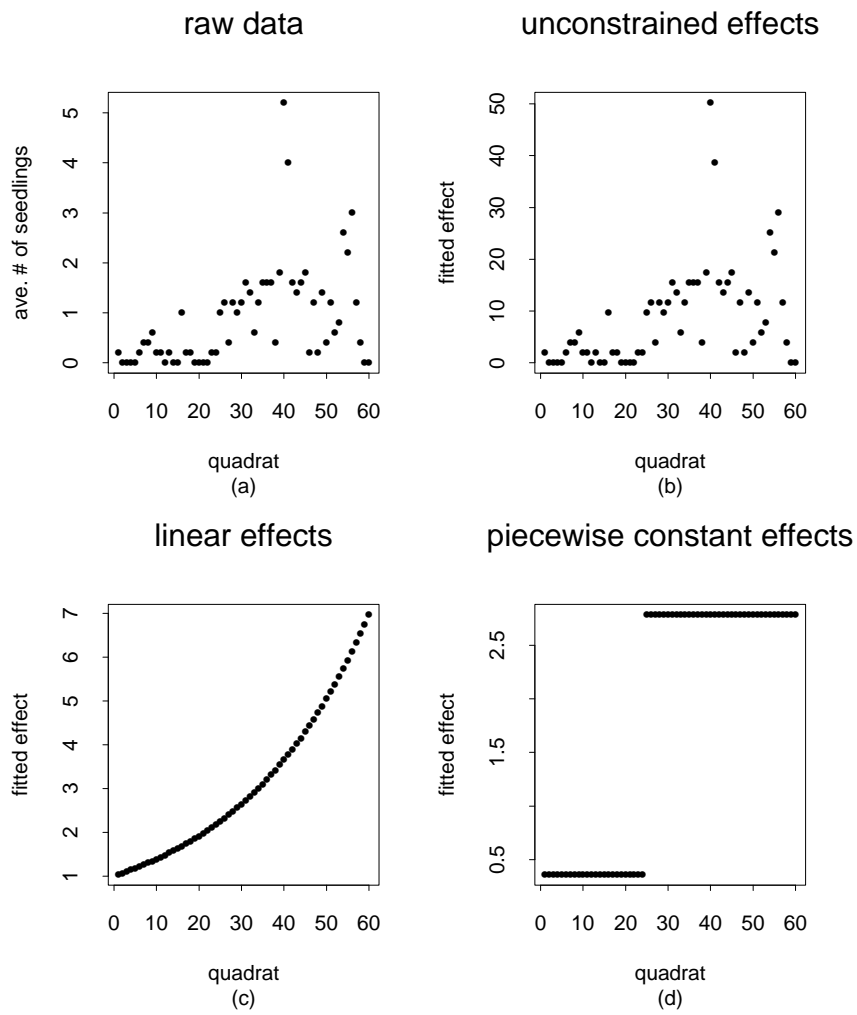


Figure 11: (a): average number of New seedlings per plot; (b), (c), (d): fitted effects from glm's

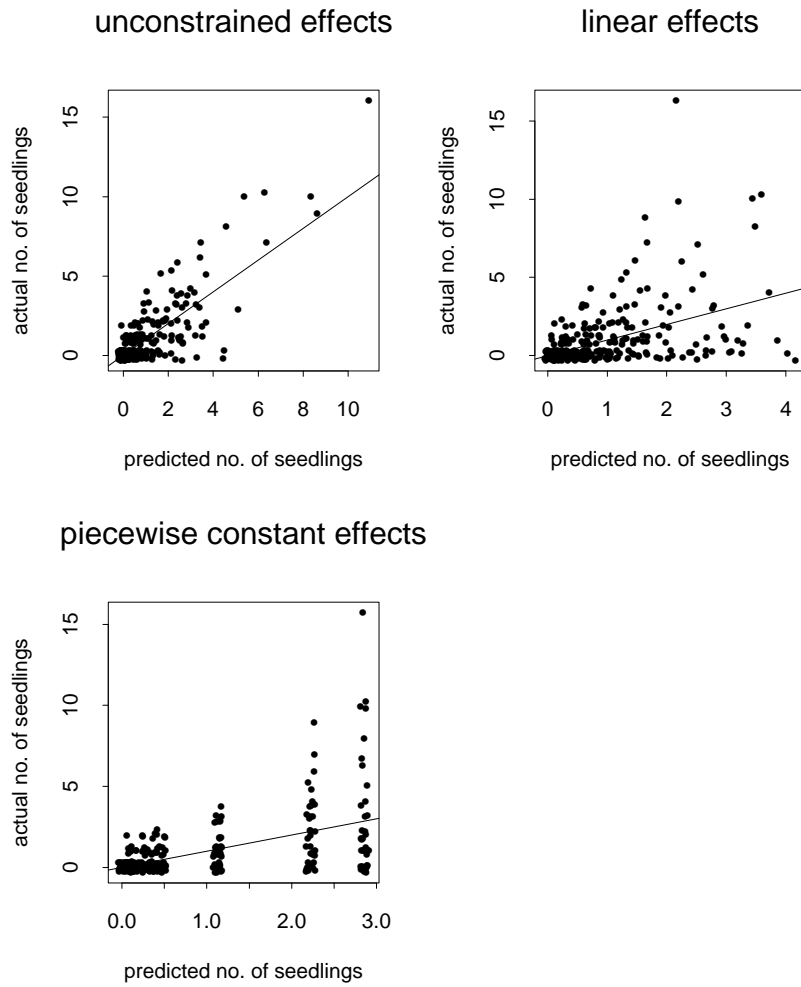


Figure 12: actual versus predicted numbers of New seedlings from three glm's. (a): unconstrained quadrat effects; (b): linear quadrat effects; (c): piecewise constant quadrat effects

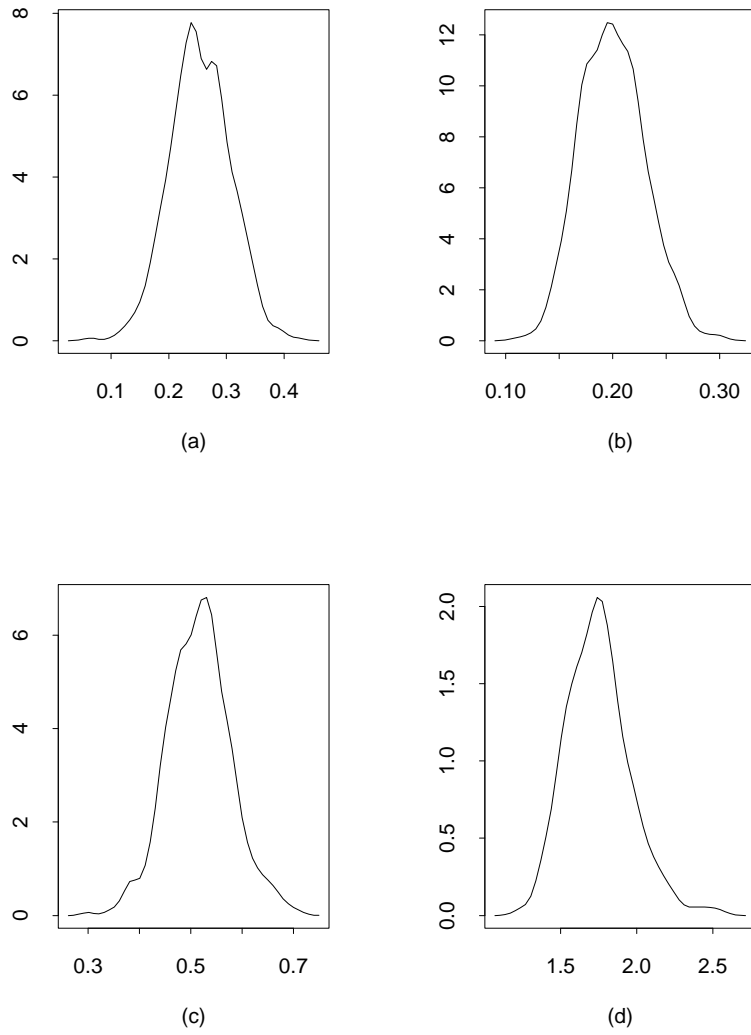
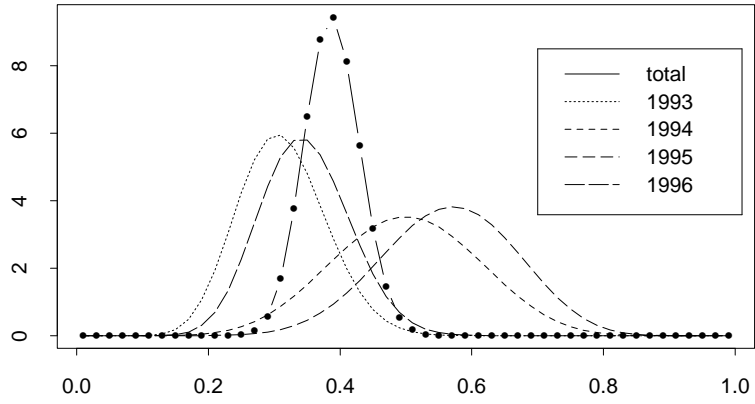
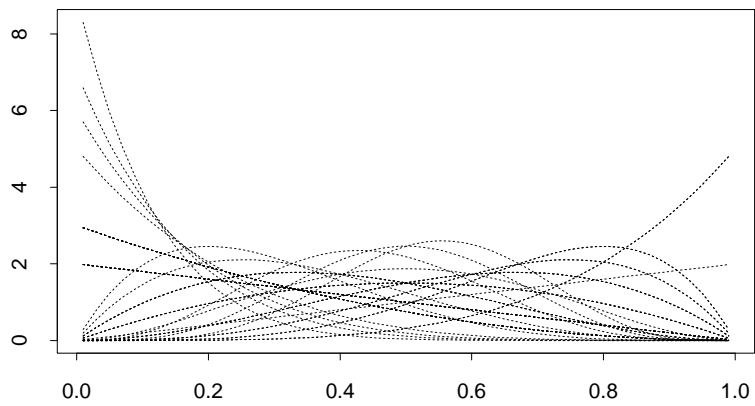


Figure 13: posteriors from model with anova effects for  $\lambda$ . (a):  $p_{Old}$ ; (b):  $p_{New}$ ; (c):  $f_{New}$ ; (d):  $\lambda$ .

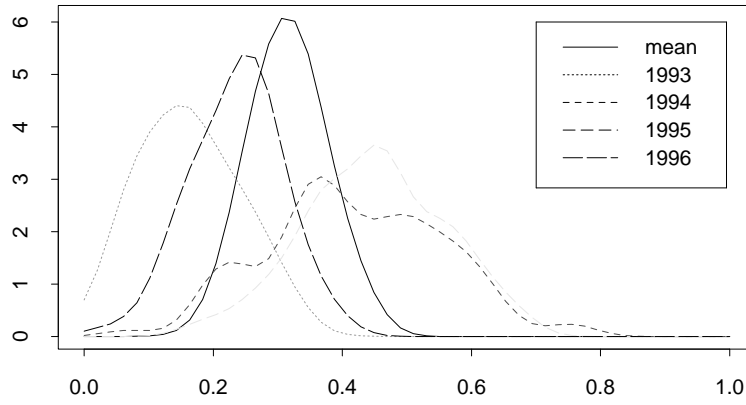


a

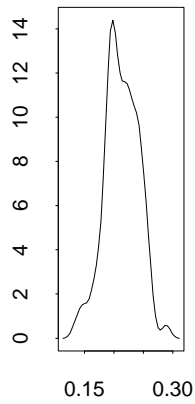


b

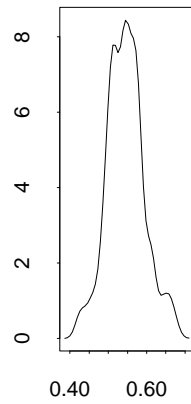
Figure 14: (a): Likelihood plot of  $p_{Old}$  for different years (b): Likelihood plot of  $p_{Old}$  for different quadrats



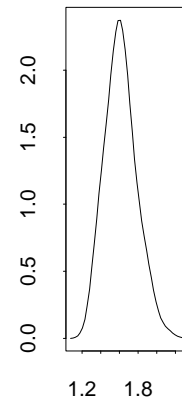
(a):  $p_{old}$



(b):  $p_{new}$



(c):  $f_{new}$



(d):  $\lambda$

Figure 15: (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  $\lambda$ .

<b>Year:</b>	1	2	3	4	5
$O_{2,j}$ :	1	0	1	0	0
$N_{2,j}$ :	0	0	0	0	0

Table 1: *Numbers of Old and New maple (acer) seedlings recorded in quadrat 2 at Coweeta.*

	Old Seedlings					New Seedlings				
	1993	1994	1995	1996	1997	1993	1994	1995	1996	1997
quadrat 1	1	1	1	1	1	0	0	1	0	0
quadrat 2	1	0	1	0	0	0	0	0	0	0
quadrat 3	0	0	1	1	0	0	0	0	0	0
quadrat 4	0	0	0	0	0	0	0	0	0	0
quadrat 5	2	1	1	1	1	0	0	0	0	0
quadrat 6	0	0	0	0	0	1	0	0	0	0
quadrat 7	3	2	1	2	0	1	0	0	0	1
quadrat 8	1	1	1	1	1	0	0	0	0	2
quadrat 9	1	1	1	1	0	1	0	2	0	0
quadrat 10	0	0	0	0	0	0	0	0	1	0
quadrat 11	0	0	2	1	2	0	0	0	0	1
quadrat 12	0	0	0	0	0	0	0	0	0	0
quadrat 13	1	0	1	1	1	0	0	0	0	1
quadrat 14	0	0	0	0	0	0	0	0	0	0
quadrat 15	0	0	0	0	0	0	0	0	0	0
quadrat 16	1	0	1	1	1	0	0	2	1	2
quadrat 17	1	0	0	0	0	0	0	0	0	1
quadrat 18	0	0	0	0	0	0	0	0	0	1
quadrat 19	0	0	0	0	0	0	0	0	0	0
quadrat 20	0	0	0	0	0	0	0	0	0	0
quadrat 21	0	0	0	0	0	0	0	0	0	0
quadrat 22	0	0	0	0	0	0	0	0	0	0
quadrat 23	0	0	0	0	0	0	0	1	0	0
quadrat 24	0	0	0	0	0	0	0	0	0	1
quadrat 25	1	1	2	1	1	0	0	2	1	2
quadrat 26	3	2	1	3	1	3	0	1	1	1
quadrat 27	1	0	1	0	0	0	0	0	0	2
quadrat 28	3	0	1	0	0	1	0	1	0	4
quadrat 29	1	1	1	1	0	0	0	0	0	5
quadrat 30	1	0	0	1	0	3	0	0	1	2
quadrat 31	0	0	0	0	0	3	0	1	2	2
quadrat 32	0	0	0	0	1	0	0	2	0	5
quadrat 33	1	0	0	0	2	1	0	1	1	0
quadrat 34	2	1	1	0	0	4	0	0	0	2
quadrat 35	0	0	0	0	0	2	0	2	1	3
quadrat 36	1	0	1	0	0	2	0	0	0	6
quadrat 37	0	0	0	2	1	1	0	3	0	4
quadrat 38	0	0	0	0	0	0	0	0	1	1
quadrat 39	1	0	0	1	0	1	0	2	2	4
quadrat 40	0	0	0	1	0	0	0	16	1	9
quadrat 41	1	0	0	7	0	3	0	10	0	7
quadrat 42	2	0	0	4	0	1	0	6	0	1
quadrat 43	0	0	0	2	1	0	0	4	0	3
quadrat 44	0	2	0	5	3	0	0	7	0	1
quadrat 45	1	0	0	0	0	0	0	5	0	4
quadrat 46	0	1	0	0	0	1	0	0	0	0
quadrat 47	2	1	0	0	1	1	0	3	2	0
quadrat 48	0	0	1	0	1	1	0	0	0	0
quadrat 49	2	1	0	1	0	0	2	2	0	3
quadrat 50	2	0	0	0	0	1	0	1	0	0
quadrat 51	2	0	0	1	1	3	0	1	1	1
quadrat 52	0	0	0	1	1	1	1	1	0	0
quadrat 53	0	0	0	0	0	2	0	2	0	0
quadrat 54	3	0	0	2	0	3	0	10	0	0
quadrat 55	2	0	0	1	1	0	0	8	2	1
quadrat 56	1	2	0	2	0	1	1	10	0	3
quadrat 57	0	0	0	1	0	1	0	4	0	1
quadrat 58	0	0	0	0	0	1	0	1	0	0
quadrat 59	0	0	0	0	0	0	0	0	0	0
quadrat 60	1	0	1	0	0	0	0	0	0	0

Table 2: Data from Coweeta: 60 plots, 5 years

	Old Seedlings					$X_{i,j}^{\max}$				
	93	94	95	96	Tot	93	94	95	96	Tot
quadrat 1	1	1	1	1	4	1	1	1	1	4
quadrat 2	1	0	1	0	2	0	0	0	0	0
quadrat 3	0	0	1	1	2	0	0	1	0	1
quadrat 4	0	0	0	0	0	0	0	0	0	0
quadrat 5	2	1	1	1	5	1	1	1	1	4
quadrat 6	0	0	0	0	0	0	0	0	0	0
quadrat 7	3	2	1	2	8	2	1	1	0	4
quadrat 8	1	1	1	1	4	1	1	1	1	4
quadrat 9	1	1	1	1	4	1	1	1	0	3
quadrat 10	0	0	0	0	0	0	0	0	0	0
quadrat 11	0	0	2	1	3	0	0	1	1	2
quadrat 12	0	0	0	0	0	0	0	0	0	0
quadrat 13	1	0	1	1	3	0	0	1	1	2
quadrat 14	0	0	0	0	0	0	0	0	0	0
quadrat 15	0	0	0	0	0	0	0	0	0	0
quadrat 16	1	0	1	1	3	0	0	1	1	2
quadrat 17	1	0	0	0	1	0	0	0	0	0
quadrat 18	0	0	0	0	0	0	0	0	0	0
quadrat 19	0	0	0	0	0	0	0	0	0	0
quadrat 20	0	0	0	0	0	0	0	0	0	0
quadrat 21	0	0	0	0	0	0	0	0	0	0
quadrat 22	0	0	0	0	0	0	0	0	0	0
quadrat 23	0	0	0	0	0	0	0	0	0	0
quadrat 24	0	0	0	0	0	0	0	0	0	0
quadrat 25	1	1	2	1	5	1	1	1	1	4
quadrat 26	3	2	1	3	9	2	1	1	1	5
quadrat 27	1	0	1	0	2	0	0	0	0	0
quadrat 28	3	0	1	0	4	0	0	0	0	0
quadrat 29	1	1	1	1	4	1	1	1	0	3
quadrat 30	1	0	0	1	2	0	0	0	0	0
quadrat 31	0	0	0	0	0	0	0	0	0	0
quadrat 32	0	0	0	0	0	0	0	0	0	0
quadrat 33	1	0	0	0	1	0	0	0	0	0
quadrat 34	2	1	1	0	4	1	1	0	0	0
quadrat 35	0	0	0	0	0	0	0	0	0	0
quadrat 36	1	0	1	0	2	0	0	0	0	0
quadrat 37	0	0	0	2	2	0	0	0	1	1
quadrat 38	0	0	0	0	0	0	0	0	0	0
quadrat 39	1	0	0	1	2	0	0	0	0	0
quadrat 40	0	0	0	1	1	0	0	0	0	0
quadrat 41	1	0	0	7	8	0	0	0	0	0
quadrat 42	2	0	0	4	6	0	0	0	0	0
quadrat 43	0	0	0	2	2	0	0	0	1	1
quadrat 44	0	2	0	5	7	0	0	0	3	3
quadrat 45	1	0	0	0	1	0	0	0	0	0
quadrat 46	0	1	0	0	1	0	0	0	0	0
quadrat 47	2	1	0	0	3	1	0	0	0	1
quadrat 48	0	0	1	0	1	0	0	0	0	0
quadrat 49	2	1	0	1	4	1	0	0	0	1
quadrat 50	2	0	0	0	2	0	0	0	0	0
quadrat 51	2	0	0	1	3	0	0	0	1	1
quadrat 52	0	0	0	1	1	0	0	0	1	1
quadrat 53	0	0	0	0	0	0	0	0	0	0
quadrat 54	3	0	0	2	5	0	0	0	0	0
quadrat 55	2	0	0	1	3	0	0	0	1	1
quadrat 56	1	2	0	2	5	1	0	0	0	1
quadrat 57	0	0	0	1	1	0	0	0	0	0
quadrat 58	0	0	0	0	0	0	0	0	0	0
quadrat 59	0	0	0	0	0	0	0	0	0	0
quadrat 60	1	0	1	0	2	0	0	0	0	0
total	46	18	21	47	132	14	9	12	16	51

Table 3: Old seedlings and  $X_{i,j}^{\max}$  from Coweeta