Tree fecundity
James S. Clark, Duke University

This summary describes a process model for the fecundity schedule that accommodates the time series character of seed production by individuals. Fecundity is a latent process—we do not directly observe seed production. Thus, we embed the fecundity model within a structure that links the latent process to the two types of data ecologists typically collect, seed collections and observations of whether or not individual trees are reproductive. The model is hierarchical Bayes, with the fecundity process being a mixed model (fixed and random effects) with AR(1) error structure (Clark et al., in prep.). The latent fecundity process is estimated tree-by-tree together with ‘population level’ variables, such as sex ratios, covariates (including size effects), and autocorrelation. The individual time series, together with their uncertainties, are then available for more extensive investigations of spatial and temporal effects. As part of this analysis we demonstrate how to assimilate different types of evidence as basis for inference at the individual and population levels.

Figure 1. Classical model for a fecundity model (eqn 2). The upper box contains ‘data’. The process model contains parameters that are estimated with error. A stochastic sampling distribution might be process error, observation error, or some combination.

<table>
<thead>
<tr>
<th>PROCESS MODEL</th>
</tr>
</thead>
<tbody>
<tr>
<td>The conditional fecundity schedule is an ‘allometric’ function of tree diameter,</td>
</tr>
<tr>
<td>$y_{it} = a_0 d_{it}^{a_1}$</td>
</tr>
<tr>
<td>or, as a linear regression,</td>
</tr>
<tr>
<td>$Y_{it} = \alpha_0 + \alpha_1 D_{it} + \beta_i + \varepsilon_{it}$</td>
</tr>
</tbody>
</table>

1

2
where \( Y_{it} = \log(y_{it}) \), \( D_{it} = \log(d_{it}) \), and \( \alpha_0 = \log(a_0) \)(Fig. 1). We use the term ‘conditional fecundity’ to denote seed production of individuals that are mature (LaDeau and Clark 2001) and (for dioecious species) female. If fecundity is simply proportional to basal area (e.g., Ribbens et al. 1994, Clark et al. 1998, 1999), then \( \alpha_1 = 2 \) (Fig. 2). If fecundity increases less rapidly than diameter, as expected when trees are large, then \( 0 < \alpha_1 < 1 \) (**). This slowing rate of increase with diameter is not realistic for young trees (e.g., Figure 2), but the increasing probability of reaching reproductive maturity with diameter (e.g., LaDeau and Clark 2001) provides for this rapid increase when trees are young (see below).

![Figure 2. Two sources of fecundity information. Felling trees and counting all seeds (Abbott 1974) is impractical and provides no information on year effects. Tradition maximum likelihood estimates based on seed rain data (Ribbens et al. 1994, Clark et al. 1999) cannot identify sufficient parameters to accurately characterize the size effect, covariates, or other sources of variation. This allometric equation has one fitted parameter, \( a_0 \).](image)

**DATA**

Our study involves mapped stands that span a range of environmental settings in the southern Appalachians and in the Piedmont of central North Carolina. Sampled stands range in area from 0.64 to ** ha. Southern Appalachian sites include mixed oak (at several elevations), cove hardwoods, mixed hemlock, ridgetop pitch pine, and northern hardwoods. Piedmont sites are mixed pine and hardwood. Two types of data inform fecundity schedules. Seed rain data are derived from seed traps deployed in mapped stands. Traps 0.42 m x 0.42 m are supported by pvc pipes 1.2 m above the ground. Traps have a chicken wire cover (to minimize rodent losses) and hardware cloth that suspends seeds above the trap base (Clark et al. 1998). Holes in traps provide drainage. Collection periods began in 1991 (five Coweeta stands), 1999 (Blackwood Division at Duke Forest), and 2000 (Experimental gap sites at Duke Forest and Coweeta). Seeds are recovered from traps 3 to 12 times annually, sorted, identified, and counted.

The second data type consists of observations of tree status. To determine reproductive status, trees were visited during flowering, fruiting, and in winter. For most trees, reproductive status is uncertain, because failure to observe seeds does not mean they are necessarily absent. In our study regions we have several dioecious species, including *Acer rubrum*, *Nyssa sylvatica*, and *Fraxinus americana*. On each occasion, individual trees were scored as reproductive, not reproductive, or uncertain. Dioecious species were identified as males if they flowered but did not bear fruit and female
otherwise. For *A. rubrum*, flower morphology was used as an additional check, but could only be used if flowers could be reached. There are uncertainties in all identifications, quantification of which is one goal of this analysis.

Trees and traps were mapped using a total station. Trees were identified to species and diameters were remeasured at 2 to 4 yr intervals. For this analysis observations consist of a total of ** traps on ** for a total of ** trap-years and ** tree-years.

**CLASSICAL APPROACH**

To motivate our approach, we begin by describing classical methods (Fig. 1). The only study of which we are aware that reports seed production of forest-grown trees based on direct observation (Abbott 1974) illustrates the challenges (Fig. 2). Seven *Acer rubrum* trees were felled, and all seeds were removed and weighed. Subsamples were counted. Total fecundity was then extrapolated to whole trees based on sample weights.

Most of the components of fecundity needed for basic demographic analysis, evolutionary considerations, and stand dynamics cannot be determined from this destructive harvest. It provides no information on sex ratio for this dioecious species, because seeds without observable seeds might have been males, immature, or simply not recognized. Of course, all trees selected for this study were reproductive; they were selected because they had seeds. We do not know the onset of maturity, because estimation requires a sample of individuals of different sizes that is blind as to their reproductive statuses. Counts show a relationship between fecundity and diameter, with the smallest individual supporting approximately 12,000 seeds and the largest individual (33 cm diameter) having nearly 1 million (Fig. 2). The snapshot view gives no indication of interannual effects. We do not know if the harvest year 1971 is representative, if there is autocorrelation, and if there are higher order effects (e.g., masting). We cannot separate size effects from individual effects—if size effects explain Figure 2, then we infer a dramatic increase in fecundity at 25 cm diameter. Individual differences might explain the scatter in Figure 2. If present, individual effects and size effects are confounded. Because there is a single year of data, we cannot assess correlations among individuals.

Indirect estimates of fecundity pose a different set of limitations. Seed-rain studies provide indirect, stand-level estimates of seed production, but not tree-level estimates. Recent inverse methods for estimating fecundity provide for simplistic assumptions of size effects (e.g., fecundity is proportional to tree basal area (Ribbens et al. 1994, Clark et al. 1998)), but such assumptions do not allow for the many sources of variability that affect fecundity. For instance, due to high interannual variability, Clark et al. (1999a) found that nearly 5 yr of data, pooled across years, were needed to estimate a single fecundity parameter. Because more parameters are not identifiable, the fecundity schedule is unrealistic over much of the size range (Fig. 2). Using such traditional methods, there is not much promise for estimating the full range of size and covariate effects, sex ratios, and the stochastic structure, including autocorrelation and synchronicity. Moreover, because these components are all part of the fecundity schedule, they should be estimated together, i.e., from the same data sets. Given that we now struggle to estimate a single parameter, a full characterization that involves size and covariate effects, autocorrelation periodicity, synchronicity, and tree-to-tree correlation, seems remote.
Hierarchical Bayes

Our model relates seed production to the two types of data ecologists typically collect. Seed trap data provide information on fecundity provided we link those data (via a process model) to the individual trees. Status observations constrain estimates based on seed traps, because they provide explicit information on individual trees. The full model that allows us to combine these different data types is hierarchical Bayes, consisting of i) process models (middle levels of Figure 3), ii) data models (upper level of Figure 3), and iii) parameter models (lower levels of Figure 3). The full hierarchical structure can be represented in terms of these three stages,

\[
p(\text{parameters}|\text{data}, \text{priors}) \propto p(\text{data}|\text{process}, \text{data parameters}) \\
\times p(\text{process}|\text{process parameters}) \\
\times p(\text{parameters})
\]

(Gelfand and Smith 1990, Wikle 2002). The left-hand side of eqn 3 is a joint distribution of all parameters, including those related to the process of interest and those that allow for additional sources of stochasticity (uncertainty and variability).

Figure 3. Hierarchical Bayes model for fecundity that links seed production by trees to observations of seed rain (upper left) and tree status (upper right).

Process models (eqn 3b) describe conditional fecundity, reproductive status, and dispersal (“Process” level of Figure 3). The conditional fecundity schedule is the seed production of a tree, conditioned on the events that it is mature, correctly recognized as reproductive, and (for dioecious species) female. It is a non-linear mixed model with longitudinal structure. The model is ‘mixed’, because there is a population level relationship between diameter and seed production (fixed effects) and a superimposed effect that varies among individuals (random effects). There is process error (model misspecification) in the diameter:fecundity relationship (log normal ‘error’), and there is
autocorrelation in seed production. The longitudinal structure means that each schedule is a time series.

Due to the nature of observations, there are two additional process models that link seed production of trees to observations of tree status (maturation schedule) and seed counts in traps (dispersal). The maturation schedule describes the increasing probability of maturation with diameter and the sex ratio. The dispersal process distributes tree production to seed traps; it is spatial and based on maps of trees and seed traps.

Data models (eqn 3a) relate the processes of conditional fecundity, maturation, and dispersal to observations (“Data” level of Figure 3). The data model for tree status observations assigns probabilities to each type of observation (seeds not observed, seeds present, male flowers present and no seeds) conditioned on the true state of the tree (immature, female and mature, male and mature). This is a model for recognition error. The data model for seed rain relates the seed rain predicted for a spatial location to seed counts based on the collecting area of traps and sampling stochasticity.

Parameter models (eqn 3c) consist of priors and hyperpriors that establish structure for parameters and various sources of stochasticity (“Parameter” and “Hyperparameter” levels of Figure 3). We constrain several aspects of the model as basis for assimilating prior knowledge. Most are noninformative. The elements of eqn 1 follow.

**The process model: fecundity**

To account for the range of stochasticity sources expected to influence fecundity, we develop a mixed model with fixed and random effects for repeated measures with autocorrelated error. We treat the full time series of fecundity estimates for an individual tree as an ‘observation’. Let $T_i$ be the number of years in the fecundity time series for tree $i$. The likelihood for the $i^{th}$ tree is

$$N_{T_i} \left( Y_i | X_i \alpha + 1 \beta_i, \Sigma_i \right).$$

Here, $Y_i$ is the $T_i \times 1$ vector of log conditional seed production for the $i^{th}$ tree, $X_i$ is the $T_i \times 2$ design matrix,

$$X_i = \begin{bmatrix}
1 & D_{i1} \\
\vdots & \vdots \\
1 & D_{iT_i}
\end{bmatrix},$$

$\alpha = [\alpha_0, \alpha_1]^T$ is the vector of parameters for ‘fixed effects’, $1$ is the $T_i \times 1$ vector of ones, $\beta_i$ are random effects associated with individuals, and $\Sigma_i$ is the covariance matrix. For a balanced design (data for all individuals of equal duration $T$) we can write the full likelihood compactly as $N_{mT} \left( Y | X \alpha + 1 \beta, I_m \otimes \Sigma \right)$, where the Kronecker product gives us a block diagonal covariance structure. Because large studies will often be unbalanced (trees may enter or leave the analysis at different times), we do not limit ourselves to balanced designs and work directly with eqn 4. Trees can enter and leave the study at any time.
The parameter vector \( \alpha \) applies to the whole population and operates on the ‘fixed effects’ in \( X_i \). The ‘random effect’ of individual \( i \) is described by \( \beta_i \). Thus, we assume that the size effects are consistent in shape, but can vary individually in strength. To see this, note that the conditional fecundity for an individual is
\[
y_{it} = (a_0 + a_i) \alpha_i
\]
where \( a_i = 10^{\beta_i} \).

The correlation among individuals is largely addressed by the random effects (Laird and Ware 1982, Lange et al. 1992), but we additionally provide for autocorrelation through the covariance matrix (e.g., Lindsey 1999). For individual \( i \), the covariance matrix with AR(1) structure is
\[
\Sigma_i = \frac{z}{1 - \rho^2} \begin{bmatrix}
1 & \rho & \rho^2 & \cdots & \rho^{r-1} \\
\rho & 1 & \rho & \cdots & \rho \\
\rho^2 & \rho & 1 & \cdots & \rho \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
\rho^{r-1} & \cdots & \cdots & \rho & 1
\end{bmatrix},
\]
with total variance \( z \), and autocorrelation \( \rho \). Note that the covariance matrix depends on individual \( i \) only in dimension. In our implementation, estimates of \( z \) and \( \rho \) conditionally depend on the full population. This assumption is not essential in our MCMC implementation. Together, the random effects and autoregression allow us to capture variability and to meet the assumptions that fecundities are conditionally independent (they are marginally dependent).

Parameter models fill out the fecundity submodel. The random effects \( \beta_i \) are normally distributed with mean zero and variance \( \tau^2 \), \( \prod_{i=1}^{m} N(\beta_i | 0, \tau^2) \), the prior for which, \( IG(\tau^2 | a_\beta, b_\tau) \), defines the third stage and makes the fecundity submodel hierarchical. Parameters for fixed effects have prior \( N_2(\alpha | c, V_\alpha) \), with mean vector \( c \) and parameter covariance matrix \( V_\alpha \). The total variance \( z \) has prior \( IG \). The autocorrelation parameter \( \rho \) has a prior that is uniform on \((-1, 1)\).

The tree status model

The second process model describes the probability of being mature. Let \( R_i \) be the status of tree \( i \), assuming values of zero (not yet reproductive) or one (mature). The probability of maturity
\[
p(R_i) = \text{Bernoulli}(R_i | \theta_i)
\]
increases with tree diameter. A gamma cumulative distribution function describes this schedule,
\[
\theta_i = CGam(d_i ; a_\theta, b_\theta)
\]
There is a 50% probability of maturity at diameter \( a_\theta b_\theta \). The sex ratio is relevant for dioecious species, described by parameter \( \phi \).
The third process model is dispersal. The density of seed expected to arrive at location \( j \) is determined by the seeds contributed by each tree. Each tree contributes the fraction of its total reproduction that disperses to location \( j \). We use a two-dimensional Student’s t distribution (2Dt),

\[
f(r_{ij}) = \frac{1}{\pi u \left( 1 + \frac{r_{ij}^2}{u} \right)^{\frac{3}{2}}}
\]

because it describes well the pattern of dispersal for a large number of species (Clark et al. 1999).

**The data models**

The two data models involve different sets of observation errors. Like other sources of stochasticity, proper accounting of observation errors is necessary to assure realistic parameter estimates. Data models translate the underlying processes to the data that are observed.

*Tree status*—Tree status is observed with error; if seeds are observed, then maturity is certain. But seeds may be missed, because they are obscure in the canopy, seed production can be low, or the timing of observations may not coincide with visible seeds on trees. Begin with a monocious species; the population consists of mature and immature individuals. Let \( q_i \) be an indicator variable representing the event that an individual is identified as bearing seeds \((q_i = 1)\) or not \((q_i = 0)\). \( Q_i \) be the corresponding true status of the tree (mature and immature, respectively), and \( v \) be the probability that a mature individual will be recognized as such. The probabilities associated with true status are \( p(Q_i = 1) = \theta \) and \( p(Q_i = 0) = 1 - \theta \) (eqn 8). Then

\[
p(q_i = 1|Q_i = 1) = \text{Bernoulli}(q_i | v). \quad \text{We assume that } p(q_i = 1|Q_i = 0) = 0 \text{ and } p(q_i = 0|Q_i = 0) = 1 \text{ (observers do not invent seeds). Thus, we are concerned with the probabilities for correct identifications } p(q_i = 1|Q_i = 1) = v \theta \text{ and ‘mistakes’ } p(q_i = 0|Q_i = 1) = 1 - v. \quad \text{With these assumptions we have probabilities for observations}
\]

\[
p(q_i = 1) = p(q_i = 1|Q_i = 1)p(Q_i = 1) = v \theta
\]

and

\[
p(q_i = 0) = p(q_i = 0|Q_i = 0)p(Q_i = 0) + p(q_i = 0|Q_i = 1)p(Q_i = 1)
\]

\[
= 1 \cdot (1 - \theta) + (1 - v) \theta = 1 - v \theta
\]

For a monoecious species, the likelihood for the status data set is

\[
p(q) = \prod_{i, q_i = 1} v \theta_i \prod_{i, q_i = 0} (1 - v \theta_i)
\]

\[
= \prod_{i = 1}^m \text{Bernoulli}(q_i | v \theta_i)
\]

Dioecious plants can be recognized as male and reproductive \((q = 2)\), female and reproductive \((q = 1)\), or unknown \((q = 0)\). The fraction of females is \( \phi \) and the sex ratio (females:males) is \( \phi/(1 - \phi) \). The likelihood is taken over the three classes,
\[ p(q) = \prod_{i,q_i=1} v\phi\theta_i \prod_{i,q_i=2} (1-\phi)\theta_i \prod_{i,q_i=0} [1-v\theta_i] \]

\[ = \prod_{i=1}^{m} \text{Multinom}(q_i|I,w) \]

with the vector of probabilities \( w \) having elements \( w_1 = v\phi\theta_i, w_2 = v(1-\phi)\theta_i, \) and \( w_3 = 1 - w_1 - w_2 \). Note that if males cannot be recognized as such, the likelihood collapses to \( p(q) = \prod_{i,q_i=1} v\phi\theta_i \prod_{i,q_i=0} (1-v\phi\theta_i) \), and the \( v \) and \( \phi \) are not independently identifiable. We assume that the same recognition errors apply to males and females.

**Seed rain**—The annual seed density observed in seed trap \( j \) in yr \( t \) is a Poisson variate

\[ s_{jt} \sim \text{Pois}(A_j\lambda_{jt}) \]

with expected seed rain \( A_j\lambda_{jt} \) determined by the full set of \( m \) trees that contribute seed to trap \( j \) in year \( t \). This expectation is the product of the collection area of the trap, a known constant \( A_j \), and the expected seed arrival per unit area, \( \lambda_{jt} \). The expected contribution from tree \( i \) to seeds in trap \( j \) depends on the probability that it is mature, which increases with diameter, described by \( \theta_{it} \), the probability that it is correctly identified as such \( v \), (for dioecious species) the probability that it is female \( \phi \), and the production of seed conditioned on reproductive status, \( y_{it} \). The fraction of that seed that falls in seed trap \( j \) is approximately \( f(r_{ij})A_j \), where \( f(r_{ij}) \) is a dispersal kernel (a pdf) for distance \( r_{ij} \) between seed trap \( j \) and tree \( i \). The expected seed per unit area in a given year is

\[ \lambda_{jt} = \sum_{i=1}^{m} y_{it} \delta_i f(r_{ij}) \]

\[ j = 1, \ldots, n \]

Note that seed traps need not have the same collecting areas. \( \delta_i \) integrates uncertainty in tree status,

\[ \delta_i = p(Q_i = 1|q_i) = \begin{cases} \theta_i \phi(1-v) & q_i = 0 \\ 1 & q_i = 1 \\ 1-v & q_i = 2 \end{cases} \]

In other words, the expected contribution from tree \( i \) is \( y_{it} \) if seeds are observed on the tree \( (q_i = 1) \), and \( y_{it} (1-v) \) if it is identified as a male \( (q_i = 2) \), the factor \( (1-v) \) representing the probability that seeds were present despite not having been observed (in which case the tree is actually female). For unknown trees \( (q_i = 0) \), the expectation is the product of \( y_{it} \) and the events that it is mature \( \theta_{it} \), it is female \( \phi \), and it was not recognized as seed-bearing, despite being so \( (1-v) \).

**The full model**

Combining both types of data with process and parameter models we have the joint posterior

\[ p(\alpha, \beta, z, \rho, Y, \tau^2, a_o, b_o | X, s, q, \ldots) \]

\[ \propto \prod_{j=1}^{n} \prod_{i=1}^{r_j} \text{Pois}(s_{jt}|A_j,\lambda_{jt}) \left[ \prod_{i=1}^{m} \text{Multinom}(q_i|I,w_i) \right]^k \]

likelihoods
\[
\times \prod_{i=1}^{m} N_N(Y_i | X_i \alpha + 1 \beta_i, \Sigma_i) \\
\quad \text{(latent fecundity process)}
\]

\[
\times N_2(\alpha | c, V_\alpha) \prod_{i=1}^{m} N(\beta | 0, \tau^2) \text{Gam}(z | a_z, b_z) \text{Unif}(\rho|1,1)
\]

\[
\times \text{Gam}(a_\alpha | a_0, b_0) \text{Gam}(b_\alpha | a_0, b_0) \text{Gam}(u | a_u, b_u) \text{Unif}(\nu|0.1) \text{Unif}(\phi|0.1)
\]

\[
\times \text{IG}(\tau^2 | a_\tau, b_\tau)
\]

hyperprior

combined with eqns 8, 9, 11, 12, 13, and 14. The monoecious case uses the Bernoulli likelihood 11a and the dioecious, 11b. The model is integrated using a Gibbs sampler (Appendix).

Fig. 4. Fecundity schedules for every individual of two species. Year effects (a) show production by year. Serial autocorrelation is positive for *Liriodendron* and negative for *Acer* (Fig. 6). Size effects (b) show the contribution of each individual (short lines are posterior modes) together with the full population summary (posterior mode and 95%CI for the population effects). The two are combined in part (c).

**THE ELEMENTS OF THE FECUNDITY SCHEDULE**

The model simultaneously estimates the effects of tree size on fecundity, sex ratio, and dispersal, together with stochasticity that affects sampling, process errors, and serial autocorrelation. It estimates the fecundity schedule of every individual tree in the data.
set (Fig. 4). The advantages over a traditional approach include the detailed information that is provided and ability to incorporate stochasticity that, if ignored, would result in biased estimates and unrealistically narrow confidence intervals. Here we provide some example results for *Acer rubrum* and *Liriodendron*.

The error in identifying individuals is sufficiently large that its incorporation is necessary (Fig. 5). Failure to accommodate it would undoubtedly affect estimates of many parameters. Approximately 60% of reproductive individuals can be correctly identified as such. The estimates of recognition success are informed by seeds that accumulate near trees.

Serial autocorrelation is similar in these two species, being slightly (but significantly) negative (Fig. 5). Higher order effects are now being analyzed for these and other species. The dioecious *Acer rubrum* is 20% female (Fig. 5).

The dramatic range of size effects across individuals (Fig. 4) results in large parameter bias in the classical model (Fig. 2). The classical ML approach does not provide a realistic structure for stochasticity. For example, it allows for no process error in seed production. The process model is assumed exact, and all ‘scatter’ must be soaked up by the Poisson sampling distribution for seed rain. By defining a process error that applies directly to fecundity, we can allow for the skewed distribution, with most years having low seed production, and a few being disproportionately high. The Poisson sampling distribution for the classical model is a catch-all that does not allow for these infrequent events, with the result being a large overestimate of the size effects on fecundity (Fig. 2). Our process error is designed for this error structure (it is log normal), such that log seed production is normally distributed about the log diameter effect (Fig. 6).

While year-to-year seed production for individual trees can be estimated with some precision, the seed production for a hypothetical tree cannot be predicted with much confidence (Fig. 6). For the sample trees we can estimate seed production, because we have seed trap data for each year. The size effect itself has a confidence interval that spans seven orders of magnitude. The analysis clearly demonstrates that interannual variability is the basis for precise estimates and noninformation predictions. The process error and autocorrelation describe large scatter about the diameter regression. Thus, prediction of future seed production will require time series analyses of the estimated series themselves.
Fig. 6. Example individual posterior modes and 95% CIs for 15 *Acer* and 15 *Liriodendron* trees (note log scale). Jagged, black curves are year-specific estimates. Smooth, colored curves are predictions for a given diameter.

Realistic estimates probably also depend on admitting the large, individual differences. Individuals are random effects. Forests are always dominated by small individuals. If they are ignored, their dominance disproportionately controls parameter
estimates. We allow for random effects of individuals, and we can examine how different size individuals affect parameterization (Fig. 4). Thus, the model simply allows for stochasticity in realistic ways, the ways in which ecologists believe that is affects the process. The result is a much different response than would be obtained under classical methods (Fig. 2).

By providing estimates for every tree in every year, we can determine levels of synchronicity, which determines masting. For example, Acer rubrum trees tend to have positive synchronicity, but the tree-to-tree correlations are rather uniformly distributed between zero and one (Fig. 7). Thus, trees of this species have been weakly correlated over the last decade.

Fig. 7. Correlations among individuals in 4 different stands.

Additional references

The general structure for the fecundity process follows Laird and Ware (1982)(see Lange et al. (1992) for an MCMC implementation). Classical treatments include the ‘random coefficients’ discussion of Lindsey (1999), who also discusses serial autocorrelation. Treatment of random (individual) effects does not differ appreciably from that of Clark et al. (2002) for seedling growth.

The dispersal process and its contribution to the sampling distribution for seed rain follows Clark et al. (1999). The sampling distribution (recognition error) for tree status extend a simple formulation described in LaDeau and Clark (2001), it is are akin to error distributions applied to mark-recapture data sets (Seber **, Jolly **, Lavine et al. 2002).

APPENDIX

The Gibbs sampler for our model is based on iterative sampling from the full conditional density for each parameter. We provide those full conditionals here.

Fecundity parameters

Regression parameters are conditionally independent of all relationships outside of the fecundity submodel. Direct sampling is based on distributions derived from theory of linear models. Parameters for fixed effects \( \alpha \) conditionally depend on all tree fecundity schedules and on a bivariate normal prior,
\[ p(\mathbf{a}|\mathbf{Y}, \mathbf{X}, \mathbf{b}, \Sigma, \ldots) \propto \prod_{i=1}^{m} N_{f_i}(\mathbf{y}_i|\mathbf{x}_i, \mathbf{a} + 1\beta_i, \Sigma_i) N_2(\mathbf{a}|\mathbf{c}, \mathbf{V}_a) \]

\[ = N(\mathbf{a} | \mathbf{v}, \mathbf{V}) \]

where \( \mathbf{V}^{-1} = \sum_{i=1}^{n} \mathbf{X}_i^T \Sigma_i^{-1} \mathbf{X}_i + \mathbf{V}_a^{-1} \) and \( \mathbf{v} = \sum_{i=1}^{n} \mathbf{X}_i^T \Sigma_i^{-1} (\mathbf{y}_i - 1\beta_i) + \mathbf{V}_a^{-1} \mathbf{c} \). To avoid the inefficient matrix inversions, we complete some tedious algebra to obtain expressions for the matrix elements. For the diagonal prior covariance matrix \( \mathbf{V}_a = \text{Diag}(\sigma_a, \sigma_a) \) we have

\[ \mathbf{v}_1 = \frac{1}{z} \sum_{i=1}^{m} \left[ (1 - \rho)(Y_i - \beta_i) + (Y_{i+1} - \beta_i) \right] + \left( 1 - \rho \right)^2 \sum_{i=2}^{T-1} (Y_i - \beta_i) \]

\[ \mathbf{v}_2 = \frac{1}{z} \sum_{i=1}^{m} \left[ (D_i - \rho D_{i+1}) (Y_i - \beta_i) + (D_{i+1} - \rho D_i) (Y_{i+1} - \beta_i) \right] \]

\[ + \sum_{i=2}^{T-1} \left[ D_i \left( 1 + \rho^2 \right) - \rho (D_{i+1} - \rho D_{i+1}) (Y_i - \beta_i) \right] + \frac{c_1}{\mathbf{v}_a} \]

This 2 x 2 matrix is readily inverted analytically. For the mean vector, we obtain

\[ \mathbf{v}_1 = \frac{1}{z} \sum_{i=1}^{m} \left[ (1 - \rho)(Y_i - \beta_i) + (Y_{i+1} - \beta_i) \right] + \left( 1 - \rho \right)^2 \sum_{i=2}^{T-1} (Y_i - \beta_i) \]

\[ \mathbf{v}_2 = \frac{1}{z} \sum_{i=1}^{m} \left[ (D_i - \rho D_{i+1}) (Y_i - \beta_i) + (D_{i+1} - \rho D_i) (Y_{i+1} - \beta_i) \right] \]

\[ + \sum_{i=2}^{T-1} \left[ D_i \left( 1 + \rho^2 \right) - \rho (D_{i+1} - \rho D_{i+1}) (Y_i - \beta_i) \right] + \frac{c_2}{\mathbf{v}_a} \]

The posterior conditional for the \( \beta_i \) contains only the fecundity schedule for the \( i^{th} \) tree,

\[ p(\beta_i|\mathbf{Y}, \mathbf{X}, \mathbf{a}, \Sigma, \ldots) \propto N_{f_i}(\mathbf{y}_i|\mathbf{x}_i, \mathbf{a} + 1\beta_i, \Sigma_i) N_2(\beta_i|0, \tau^2) \]

\[ = N(\beta_i | \mathbf{v}, \mathbf{V}) \]

where

\[ \mathbf{V}^{-1} = \mathbf{1}^T \Sigma_i^{-1} \mathbf{1} + \tau^{-2} = \frac{2m(1 - \rho) + (T_i - 2)(1 - \rho)^2}{z} + \frac{1}{\tau^2} \]

and

\[ \mathbf{v} = \mathbf{1}^T \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{x}_i, \mathbf{a}) \]

\[ = \frac{1}{z} \left[ (1 - \rho)(Y_i - Y_{i+1} - 2\alpha_0 - \alpha_1 (D_i + D_i)) + (1 - \rho)^2 \sum_{i=2}^{T-1} (Y_i - \alpha_0 - \alpha_1 D_i) \right] \]

The variance on hyperparameters has conditional posterior

\[ p(\tau^2 | \ldots) \propto \prod_{i=1}^{m} N(\beta_i|0, \tau^2) \text{IG}(\tau^2 | a_r, b_r) \]
Imputed seed production from the \(i\)th tree conditionally depends on the likelihood for all seed traps in all years (because there is dispersal and autocorrelation), and on its’ own fecundity schedule. For the \(i\)th tree we have the full posterior conditional

\[
p(y_i|x_i, \alpha, \beta_i, \Sigma, \ldots) \propto \prod_{i=1}^{T_i} \prod_{j=1}^{n} \text{Pois}(s_{ji} | A_j \lambda_{ji}) \; N_{T_i}(y_i | x_i + 1 \beta_i, \Sigma_i)
\]

We use a Metropolis step based on the proposal density \(N_{T_i}(y_i, 0.1 \Sigma_i)\). Thus, we propose a full time series for each tree, and we accept or reject on the basis of the full series, not on specific years.

Total variance \(z\) was drawn with a M-H step. We used a gamma proposal for \(z\) centered on the current value and used it to construct the covariance matrices \(\Sigma_i\) for the conditional

\[
p(z_i \ldots) \propto \prod_{i=1}^{m} N_{T_i}(y_i | x_i + 1 \beta_i, \Sigma_i) \; \text{Gam}(z | a_z, b_z)
\]

For \(\rho\) we used the symmetric proposal density \(\text{Unif}(\rho^{(k)} - h, \rho^{(k)} - h)\) and a Metropolis step, where \(\rho^{(k)}\) is the current value of \(\rho\), and \(h = \min(0.2(1 - |\rho^{(k)}|), 0.05)\) to sample from the conditional posterior

\[
p(\rho | \ldots) \propto \prod_{i=1}^{m} N_{T_i}(y_i | x_i + 1 \beta_i, \Sigma_i) \; \text{Unif}(\rho | -1, 1)
\]

Unless near –1 or 1, the proposal density has width 0.1. Near either limit, it narrows. Because correlations are always bounded away from limits, the algorithm does not bog down there. As with \(z\), the covariance matrix is assembled as basis for evaluating the conditional posterior.

**Dispersal parameter**

The dispersal parameter conditionally depends on all of the seed trap data,

\[
p(u | \ldots) \propto \prod_{i=1}^{T_i} \prod_{j=1}^{n} \text{Pois}(s_{ji} | A_j \lambda_{ji}) \; \text{Gam}(u | a_u, b_u)
\]

We used a M-H step with a Gamma proposal density.

**Tree status parameters**

Conditional posteriors for tree status parameters contain both full likelihoods, because statuses \(q_i\) directly influence all seed traps through expected seed rain (eqn 13, 14). For \(a_{\theta}\), the conditional is

\[
p(a_{\theta} | \ldots) \propto \prod_{i=1}^{T_i} \prod_{j=1}^{n} \text{Pois}(s_{ji} | A_j \lambda_{ji}) \left[ \prod_{i=1}^{m} \text{Multinom}(q_i l, w_i) \right]^k \; \text{Gam}(a_{\theta} | a_0, b_0)
\]

with that for \(b_0\) having the appropriate prior. For both we used Gamma proposals and M-H. Likewise, the female fraction full conditional is
\[
p(\phi|\ldots) \propto \prod_{i=1}^{T_i} \prod_{j=1}^{n_i} \text{Pois}(s_{ji} | A_{ji} \lambda_{ji}) \left[ \prod_{i=1}^{m} \text{Multinom}(q_{j} | l_i, w_i) \right]^k \text{Unif}(\phi|0,1)
\]

Recognition error
\[
p(v|\ldots) \propto \prod_{i=1}^{T_i} \prod_{j=1}^{n_i} \text{Pois}(s_{ji} | A_{ji} \lambda_{ji}) \left[ \prod_{i=1}^{m} \text{Multinom}(q_{j} | l_i, w_i) \right]^k \text{Unif}(v|0,1)
\]

**Prior parameter ranges**

With one exception, all priors are noninformative. We use prior knowledge that fecundity must eventually decline with size to inform the population-level allometric parameter $\alpha_1$. We center the prior mean at 0.5 and use the variance $**$. Note that the hierarchical structure still allows for individual departures from the population estimate.