

Nathanael Lichti, Michael A. Steele, Hao Zhang, and Robert K. Swihart. 2014. Mast species composition alters seed fate in North American rodent-dispersed hardwoods. *Ecology* 95:1746–1758.

APPENDIX B: Modeling seed dispersal with imperfect detection.

One of the principal challenges in estimating seed dispersal arises from the fact that the area which must be searched to relocate a dispersed seed increases proportionally to the square of the dispersal radius. For methods such as metal-tagging, which use inconspicuous tags that cannot be detected from long distances, the probability that a searcher will encounter a particular seed during a fixed time period declines with the area searched during that period. Even when the seed's exact location is encountered, the likelihood of detecting a tag may vary with substrate, vegetation, tag burial depth, the equipment used, or the experience of the searcher.

For the study design and tags used in our study, detection depends on both survival status and dispersal (Fig. B1). Moore et al. (2007) estimated that the probability of detection for tags cached 2-3 cm below the soil surface was approximately half of that for tags in the leaf litter. In addition, tags may be overlooked because they have been dispersed outside of the search area, in which case $P(\text{recovery}) = 0$. Consequently, survival, dispersal distance, and detection cannot be treated as independent random variables unless all tags are recovered; a change in the parameters that determine one value will necessarily influence estimates for the other two by reallocating the number of seeds estimated to be outside the search area, inside the search and surviving, or inside and not surviving. Moreover, methods that fail to properly account for non-detection are likely to produce biased parameter estimates for both seed survival and dispersal (Lichti 2012).

We used a maximum-likelihood approach to obtain simultaneous parameter estimates for seed survival, dispersal, and detection, given data on search area geography and recovered seeds. Because many dispersed seeds are never recovered, and therefore have unknown dispersal distances and survival statuses, we fit the model using an EM algorithm (Wei & Tanner 1990). Lichti (2012) provides a detailed discussion of the model derivation, the specific algorithm used to solve the model, and comparisons to other estimators in terms of statistical bias and efficiency. Relative to the typical naïve approaches based on recovered tags alone, the model presented here provides substantially less biased estimates for both survival and dispersal (Lichti 2012).

The model discussed here concerns only dispersed seeds; non-removed seeds are not considered. Begin by assuming that all tags are recovered. If seeds survive the period from presentation to recovery with probability ψ and dispersal is isotropic with a marginal density function for distance of $f(r | \alpha)$, the likelihood of a parameter vector $\theta = (\psi, \alpha, d_1, d_0, \lambda)$ for a given design stratum is,

$$L(\theta | \text{data}) = \prod_{j=1}^N \prod_{i=1}^{R_j} \psi^{s_{ij}} (1-\psi)^{1-s_{ij}} f(r_{ij} | \alpha) \left[g(r_{ij}, A_j) d_1^{s_{ij}} d_0^{1-s_{ij}} e^{-\lambda |A_j|} \right] \quad (\text{B.1})$$

where N is the number of boxes in the stratum, R_j is the number of seeds removed from box j , s_{ij} equals 1 if seed i from box j remains cached when recovered, and 0 otherwise, and r_{ij} is the dispersal distance for seed i . By definition, $r_{ij} > 0$ because eqn. B.1 concerns only the seeds that were removed from box j . The variable A_j indicates a 2-dimensional spatial region where searches around box j were conducted; it does not need to be contiguous, symmetrical, or centered on j , but its geometry must be known. The region A_j has area $|A_j|$ and the fraction of a circle of radius r centered on box j that intersects A_j is given by the function $g(r, A_j)$. Because the location of box j and the geometry of A_j are both known, $g(r, A_j)$ is also known.

In eqn. B.1, the term $\left[g(r_{ij}, A_j) d_1^{s_{ij}} d_0^{1-s_{ij}} e^{-\lambda|A_j|} \right]$ describes the probability of tag recovery conditional on A_j , dispersal distance r_{ij} , and survival status s_{ij} . Assuming isotropic dispersal and no data on dispersal direction, the probability that a seed dispersed to distance r lands in A_j is given by $g(r, A_j)$. The term $e^{-\lambda|A_j|}$ describes the probability that a seed's exact location in A_j will be encountered by a searcher given that the seed lands in A_j . Parameters d_1 and d_0 give the probabilities that a tag will be detected, given that its exact location is encountered and that it is buried in a cache or lying on the surface, respectively. We assume that the probability of actually encountering a given seed's location declines exponentially at rate λ with the area searched. That is, for a tag that falls inside A_j , the probability of recovery is equal to $d_1^{s_{ij}} d_0^{1-s_{ij}} e^{-\lambda|A_j|}$. The exponential structure is predicated on the assumption that the total area actually encountered by searchers during a standardized effort is a fixed function of search time, and that this effort is uniformly distributed across $|A_j|$.

If all search regions have the same area, then d_1 , d_0 , and λ cannot be separately estimated without additional data. However, if the relationship between search intensity and search area remains constant, then the detection parameters can be separated by searching regions of different sizes or by including pilot data for known search areas and recovery rates, such as those provided by Moore *et al.* (2007). Our analysis uses the latter approach.

When only $n_j \leq R_j$ tags are recovered, eqn. B.1 must be modified to account for the missing seeds:

$$L(\boldsymbol{\theta} | \text{data}) = \prod_{j=1}^N \prod_{i=1}^{n_j} \psi^{s_{ij}} (1-\psi)^{1-s_{ij}} f(r_{ij} | \boldsymbol{\alpha}) \left[g(r_{ij}, A_j) d_1^{s_{ij}} d_0^{1-s_{ij}} e^{-\lambda|A_j|} \right] \times \prod_{j=1}^N \left\{ W_j \left[\psi \left(1 - d_1 e^{-\lambda|A_j|} \right) + (1-\psi) \left(1 - d_0 e^{-\lambda|A_j|} \right) \right] + (1 - W_j) \right\}^{R_j - n_j} \quad (\text{B.2})$$

where $W_j = \iint_{A_j} f(r | \boldsymbol{\alpha}) dA$ is the probability that a seed with unknown dispersal distance lands in A_j

given the geometry of A_j and dispersal parameters $\boldsymbol{\alpha}$. To estimate the parameters for eqn. B.2, we used a Monte Carlo EM algorithm (Wei and Tanner 1990) to maximize the complete data likelihood,

$$\begin{aligned}
L_c(\boldsymbol{\theta} | \mathbf{X}, \tilde{\mathbf{X}}, \text{pilot data}) = & \prod_{j=1}^N \left\{ \prod_{i=1}^{n_j} \left[\psi^{s_{ij}} (1-\psi)^{1-s_{ij}} f(r_{ij} | \boldsymbol{\alpha}) \left[g(r_{ij}, A_j) e^{-\lambda|A_j|} d_1^{s_{ij}} d_0^{1-s_{ij}} \right] \right]^{F_{ij}} \times \right. \\
& \left. \prod_{i=n_j+1}^{R_j} \left[\psi^{\tilde{s}_{ij}} (1-\psi)^{1-\tilde{s}_{ij}} f(\tilde{r}_{ij} | \boldsymbol{\alpha}) \left[1 - g(\tilde{r}_{ij}, A_j) d_1^{\tilde{s}_{ij}} d_0^{1-\tilde{s}_{ij}} e^{-\lambda|A_j|} \right] \right]^{1-F_{ij}} \right\} \\
& \times L(d_1, d_0, \lambda | \text{pilot data})
\end{aligned} \tag{B.3}$$

where $\mathbf{X} = (s, \mathbf{r}, \mathbf{F}, \mathbf{A}, \mathbf{n})$ and $\tilde{\mathbf{X}}$ contains the hidden (i.e., unobserved) survival statuses and dispersal distances for the unrecovered tags, indicated by \tilde{s}_{ij} and \tilde{r}_{ij} , respectively. The indicator variable F_{ij} equals 1 if a seed was found, and 0 otherwise. With the inclusion of pilot data, eqn. B.3 is equivalent to a Bayesian posterior with flat priors on ψ and $\boldsymbol{\alpha}$ and an informative prior for the detection process. The hidden data were simulated from their conditional distributions:

$$\begin{aligned}
f(r | \boldsymbol{\theta}, A_j, F=0) = & \frac{f(r | \boldsymbol{\alpha}) \left\{ (1 - g[r, A_j]) + g(r, A_j) (\psi [1 - d_1 p_j] + [1 - \psi] [1 - d_0 p_j]) \right\}}{1 - \left\{ p_j (\psi d_1 + [1 - \psi] d_0) \int f(r | \boldsymbol{\alpha}) g(r, A_j) dr \right\}}
\end{aligned} \tag{B.4}$$

where $p_j = e^{-\lambda|A_j|}$, and $f(s | \boldsymbol{\theta}, r, A_j, F=0) = \zeta^s (1-\zeta)^{1-s}$ where

$$\zeta = \frac{\psi \left\{ g(r, A_j) (1 - d_1 p_j) + (1 - g[r, A_j]) \right\}}{g(r, A_j) (\psi [1 - d_1 p_j] + [1 - \psi] [1 - d_0 p_j]) + (1 - g[r, A_j])}. \tag{B.5}$$

For the pilot data from Moore et al. (2007),

$$\begin{aligned}
\log L(d_1, d_0, \lambda | \text{pilot data}) = & 77 \log(p_{\text{pilot}} d_1) + 100 \log(1 - p_{\text{pilot}} d_1) + 156 \log(p_{\text{pilot}} d_0) + 26 \log(p_{\text{pilot}} d_0)
\end{aligned} \tag{B.6}$$

where $p_{\text{pilot}} = e^{-125\lambda}$.

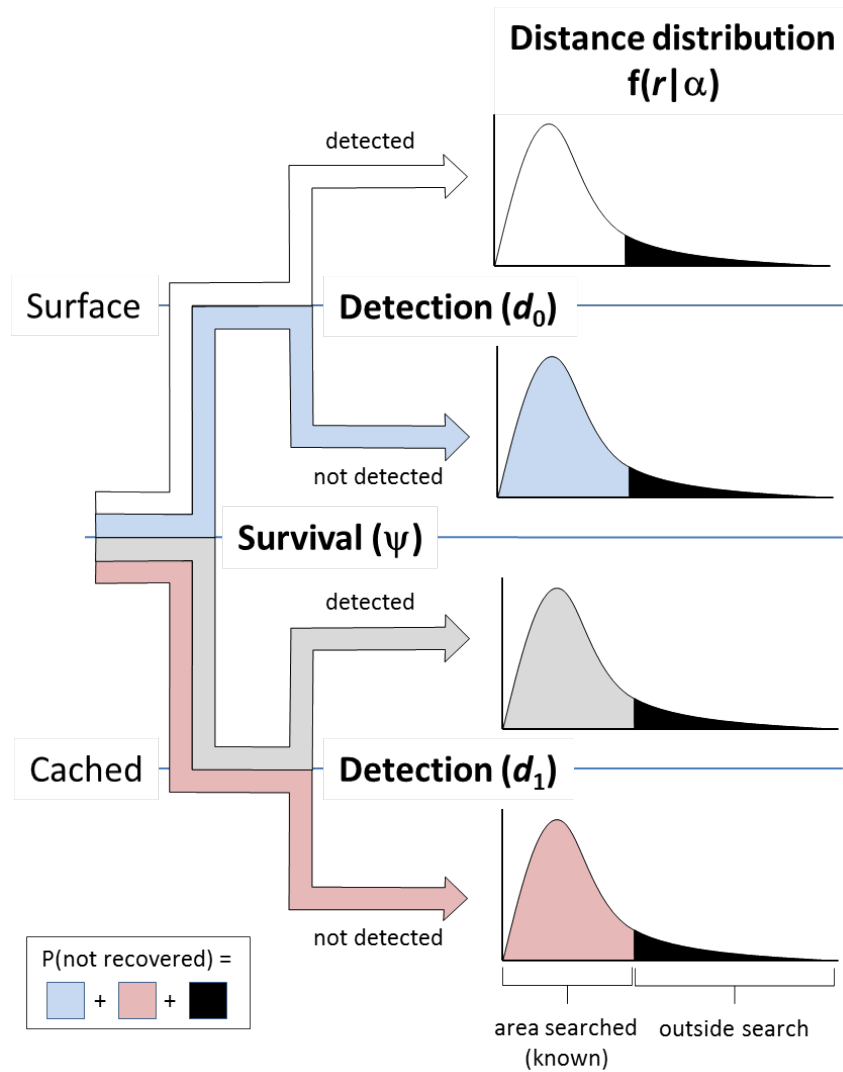


FIG. B1. Schematic representation of observed seed fate, conditional on removal from a source. Branch labels (regular type) indicate the cache and observation status of tagged seeds. Model parameters are shown in bold. Surface seeds are assumed to have been eaten, so cache status is determined by the survival probability, ψ . Seed status, including tag recovery, is defined by a Bernoulli branching process in which the probability of recovery depends on a tag's location and cache status. All fate categories are assumed to share a common, isotropic dispersal distribution, $f(r|\alpha)$, where r is dispersal distance and α is a parameter vector that defines dispersal. For a fixed-radius, circular search area around the seed source, the probability of falling outside the search area is the integral of $f(r|\alpha)$ from the maximum search radius to infinity. Detection of tags that fall within the search area depends on their cache survival status (detection probabilities conditional on encountering the seed, d_1 and d_0 for cached and eaten, respectively) and the probability of encountering the seed, given the search area, $\exp(-\lambda|A|)$, where $|A|$ is the area searched and $\lambda > 0$ controls the relationship between area and encounter probability.

LITERATURE CITED

- Lichti, N.I. 2012. Implications of context-dependent scatterhoarding for seed survival and dispersal in North American oaks (*Quercus*). Dissertation. Purdue University, West Lafayette, Indiana, USA. 298 pages.
- Moore, J. E., A. B. McEuen, R. K. Swihart, T. A. Contreras, and M. A. Steele. 2007. Determinants of seed removal distance by scatter-hoarding rodents in deciduous forests. *Ecology* 88:2529–2540.
- Wei, G. C. G., and M. A. Tanner. 1990. A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association* 85:699–704.