Appendix S1:

Foodwebs based on unreliable foundations: spatio-temporal masting merged with consumer movement, storage, and diet

JAMES S. CLARK[1,2], CHASE NUÑEZ[1], and BRADLEY TOMASEK[1]

February 25, 2019

[1] *Nicholas School of the Environment, Duke University, Durham NC 27708, USA*
[2] *Department of Statistical Science, Duke University, Durham NC 27708, USA*

Contents

S1   GENERAL THEORY

The spatially-explicit analysis for field data can be generalized to an arbitrary landscape, where explicit tree locations are unspecified. Consider a stand where trees that differ in fecundity are specified only in terms of their spatial contagion and fecundity class, perhaps summarized by a distance-size distribution. Fecundity differences may be linked to the distribution of tree-to-tree distances, as when fecund trees tend to be large and also sparse. Let r represent distance, and $r_1, r_2, \ldots$ represent the ranked distances from a randomly selected tree to others in the forest. Spatial contagion is described by density functions $p_k(r) \cdot dr$ for the probability that the kth closest tree occurs within distance interval $(r, r + dr]$ from a random tree. For simplicity, we assume second-order stationarity. In other words, the landscape is homogeneous in the sense that spatial aggregation exists at a fine scale relative to the area of the entire landscape. In this sense, 'homogeneous' means that density $p_k(r)$ does not depend on location, despite the possibility of clumped tree distributions.

The probability that the kth closest neighbor will be at least distance r is $P_k(r) = [r_k \geq r]$. The corresponding density for nearest neighbors is $p_1(r) = -dP_1/dr$. The density for nearest trees from a random spatial location is

$$q_1(r) = \frac{1}{\mu} P_1(r)$$

where $\mu = \int_0^\infty P_1(r) dr$ is the mean distance between nearest neighbors. This is the translation of tree-to-tree distances to location-to-tree distances.

The density of distances to the kth neighbor is translated to a renewal intensity for the expected number of trees at distance r. From a randomly selected location, the renewal intensity is

$$m(r) = \sum_{k=0}^{\infty} q_k(r)$$

where $m(r)dr$ is the expected number of trees in distance interval $(r, r + dr]$. For second-order stationarity (density $q_k(r)$ does not depend on location), $m(r)$ is a non-decreasing function that converges to an asymptote as $q_k(r)$ tends to zero with higher values of k.

Thus far, the model takes no account of the fact that fecundity may increase with tree size, and large and fecund trees tend to be sparse. To account for this effect, the renewal intensity can be determined for each fecundity class g, or m_g. Let $E(f_g)$ represent expected seed production for trees in fecundity class g. The expected intensity for seed arrivals emanating from trees in class g at all distances is

$$E(\lambda) = \sum_g E(f_g) \int_0^\infty m_g(r) S(r) dr$$

The landscape (conditional) variance is

$$Var(E(y|f)) = Var(\lambda|f) = \sum_g E^2(f_g) \int_0^\infty m_g(r) S(r) dr - (E(\lambda))^2$$

The variance in seeds is

$$Var(y) = Var(E(y|f)) + E(Var(y|f))$$

$$= \sum_g E^2(f_g) \int_0^\infty m_g(r)S(r)dr - E(\lambda)\,(E(\lambda) - 1)$$

the last term incorporating the Poisson sampling variance. The resource score in the section *Variance for a generalist consumer* is available for this generic landscape based on these expressions for mean and variance. The basic elements of this analysis are detailed in the next section.

S2   GENERATIVE MODEL AND PREDICTION

S2.1   *Model elements*

Elements summarized in MODEL DEVELOPMENT of the main text are described with additional detail here, with variable definitions from table 3:

- An *observation* consists of covariates and responses for trees and seed traps, including $\{\mathbf{X}_{j,t}, \mathbf{W}_{j,t}, \mathbf{V}_{j,t}, \mathbf{Y}_{j,t}, \mathbf{z}_{j,t}, \{\mathbf{c}_{ij,t}\}_{i=1}^{n_j}, \{A_{sj,t}\}_{s=1}^{S_{j,t}}\}$. There are $i = 1, \ldots, n_{j,t}$ trees on plot $j, \ldots, J$ in years $t, \ldots, T_j$. There are $s = 1, \ldots, S_{j,t}$ seed traps in plot j in year t. The observation matrices are:

  - *Predictors* that explain maturation occupy the $n_{j,t} \times q^v$ design matrix $\mathbf{V}_{j,t}$. Predictors that explain fecundity occupy the $n_{j,t} \times q^x$ design matrix $\mathbf{X}_{j,t}$. If there are random individual effects, they occupy the random effects design $\mathbf{W}_{j,t}$. The different species h are treated as factor levels in $\mathbf{V}_{j,t}, \mathbf{X}_{j,t}$, and $\mathbf{W}_{j,t}$ including all interactions with predictors. Thus, species labels are absorbed into design matrices.

  - The length-$n_{j,t}$ *maturation* vector $\mathbf{z}_{j,t}$ holds observed maturation states (often unknown).

  - The length-3 observed fecundity vector $\mathbf{c}_{ij,t} = (c_{ij,t}^n, c_{ij,t}^f, c_{ij,t}^s)$ holds the seed count for a tree, the fraction of the crown that was viewed, and the standard deviation assigned to the crown fraction. If fruiting structures containing multiple seeds are counted (e.g., cones), then $c_{ij,t}^n$ is expressed on a seed basis (cones $\times$ seeds per cone) (LaDeau and Clark, 2006). The observation model becomes $betaBinom(c_{ij,t}^n|\psi_{ij,t}, a_{ij,t}, b_{ij,t}) = \int_0^1 binom(c_{ij,t}^n|\psi_{ij,t}, c_{ij,t}^f)beta(c_{ij,t}^f|a_{ij,t}, b_{ij,t})dc_{ij,t}^f$.

  - The $S_{j,t} \times M$ *response* matrix $\mathbf{Y}_{j,t}$ holds seed counts. It has one row for each seed-trap-year and one column for each seed type m.

- *Sample effort* $A_{sj,t}$ is the trap area (m^{-2}).

- Trap-to-tree distances enter the $S_{j,t} \times n_{j,t}$ *redistribution kernel* matrix $\mathbf{S}_{j,t}$. The distance from seed trap s, j to tree i, j is $d_{sij,t} = |\mathbf{s}_{sj,t} - \mathbf{s}_{ij,t}|$. The t subscript allows for ingrowth of new individuals, for mortality loss, and for the addition or loss of

seed traps over time. $\mathbf{S}$ incorporates dispersal parameters $u_g, g = 1, \ldots G$, where g can correspond to species h or to random groups. The redistribution kernel matrix $\mathbf{S}_{j,t}$ has elements

$$\mathbf{S}_{sij,t} = \frac{u_{i[h]}}{\pi \left(u_{i[h]} + d_{sij,t}^2 \right)^2} \tag{S2.1}$$

for fitted dispersal parameter $u_{i[h]}$, where subscript $i[h]$ references the parameter value for the species h to which tree i belongs. This is a two-dimensional Student's t distribution (Clark et al., 1999).

S2.2   *Censored and missing seed-trap data*

Seed trap counts can be censored, either intentionally or not. Counts are intentionally censored when they are recorded in bins. Let $\mathcal{P}$ represent a partition of the non-negative integers $\{0, 1, \ldots\}$. For example, Bourg et al. (2013) use the interval scale $\mathcal{P} = \{0, 1, 2, 5, 20, \infty\}$. For this study, censored values $y_{smj,t}^c \in \{0, 1, 2, 3, 4\}$ represent counts that are partitioned as $\mathbf{p}_{smj,t} = \left[\mathcal{P}_{y_{sjm,t}^c}, \mathcal{P}_{y_{sjm,t}^c} + 1, \ldots, \mathcal{P}_{y_{sjm,t}^c + 1} \right]$. To get specific, a recorded value $y_{sjm,t}^c = 3$ indicates counts in the range $6 \leq y_{sjm,t} \leq 20$. When there are multiple seed collections $k = 1, \ldots$ during a year t, then the censored interval is bounded by the sums of lower and upper values of intervals for all censored values within that year,

$$\mathbf{p}_{smj,t} = \left[\sum_{k \in t} \mathcal{P}_{y_{sjm,k}^c}, \sum_{k \in t} \mathcal{P}_{y_{sjm,k}^c + 1} \right] \tag{S2.2}$$

Counts are censored unintentionally when traps are damaged at some point during the collection interval, in which case the count is a minimum for the year, with the 'true' value lying in the censored interval

$$\mathbf{p}_{smj,t} = \left[\sum_{k^u \in t} y_{sjm,k}, \infty \right] \tag{S2.3}$$

where uncensored intervals k^u are the undamaged collections. In either case, there is a censored likelihood based on the cumulative Poisson distribution,

$$CPoi \left(y_{smj,t}^c | A_{sj,t} \lambda_{smj,t}, \mathbf{p}_{smj,t} \right) = \sum_{k \in \mathbf{p}_{smj,t}} Poi \left(k | A_{sj,t} \lambda_{smj,t} \right) \tag{S2.4}$$

For censored counts, this likelihood substitutes for eq. (4).

Seed trap counts can be missing, in which case they are imputed from the likelihood eq. (4).

S2.3   *Dynamic interpretation*

The process is a multivariate dynamic (state-space) model for conditional fecundity and maturation, with a joint distribution of these latent states, $[\psi_{ij,t}, \rho_{ij,t}]$. Both state variables respond to environmental conditions.

The maturation observation model recognizes uncertainty in the assignment of maturation (fruits are often unobservable in crowded canopies) and the fact that trees are not observed in many years. Let $z_{ij,t}$ be the observed status, which can be mature (fruits observed, $z_{ij,t} = 1$), uncertain (fruits not observed, canopy obscure), and immature (entire canopy visible in the fruiting season and fruits not observed, $z_{ij,t} = 0$). Observations on trees can be missing in many years (Clark et al., 2004). $t_{ij,l}$ is the last year in which individual ij was observed to be immature. $t_{ij,m}$ is the first year ij was observed in the mature state. True maturation status is the indicator $\rho_{ij,t} \in \{0, 1\}$, with $\rho_{ij,t} = 1$ being the event that individual ij is mature in year t. Maturation is a one-way process, $[\rho_{ij,t+1} = 1|\rho_{ij,t} = 1] = 1$, and $[\rho_{ij,t} = 1|\rho_{ij,t+1} = 0] = 0$. Status is known to be mature any time after first observed to be mature and to be immature any time before the last time it is established to have been immature. Between these times, the status is unknown and modeled with a probit:

$$
\begin{aligned}
z_{ij,t_m} = 0 &\rightarrow & \rho_{ij,t} = 0, \forall t \leq t_m \\
z_{ij,t_l} = 1 &\rightarrow & \rho_{ij,t} = 1, \forall t \geq t_l \\
t_l < t < t_m &\rightarrow & [\rho_{ij,t} = 1|\rho_{ij,t-1} = 0] = \Phi\left(\mathbf{v}'_{ijt}\beta^v\right)
\end{aligned}
\tag{S2.5}
$$

where t_l is an observation year earlier than t, and t_m is an observation year after year t. Where status observations are unavailable, eq. (4) disappears from the model.

In some data sets there are observations of seeds, fruits, or cones counted from the ground or canopy. The count $c_{ij,t}^n$ is obtained for the fraction of the tree's canopy that is observed $c_{ij,t}^f$. Where counts are unavailable, eq. (6) disappears from the model.

Conditional fecundity is continuous, $\psi_{ij,t} \in (0, \infty)$, and depends on maturation status,

$$
\psi_{ij,t} \in \begin{cases} (1, \infty) & \rho_{ij,t} = 1 \\ (0, 1] & \rho_{ij,t} = 0 \end{cases}
$$

Note that fecundity here means the capacity to produce at least one seed.

S2.4 *Multiple seed types*

The observation model includes the uncertain assignment of seeds to species. Seed collections are classified as those that can be confidently assigned to a species and those that can only be assigned to genus or even family. There are $h = 1, \ldots, H$ species that might potentially contribute to $m = 1, \ldots, M$ seed types in $H \times M$ matrix $\mathbf{M}$. For example, seed types in a data set might include three *Pinus* species plus a larger category, *Pinus* spp., which includes all seeds that could not be confidently identified to species. A seed-type composition matrix for $H = 3$ species might be organized like this:

$$
\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3]'
$$

	P. echinata	P. taeda	P. virginiana	P. spp.	
	0.1	0	0	0.9	$h = P.\ echinata$
$=$	0	0.2	0	0.8	$h = P.\ taeda$
	0	0	0.2	0.8	$h = P.\ virginiana$

A row $\mathbf{m}_h$ of matrix $\mathbf{M}$ represents the fraction of seeds produced by species h that are counted in each seed-type class m. The rows accommodate observation errors, the fraction of species-h seed that is misclassified as seed type m. The vector corresponding to individual i is designated with the notation $\mathbf{m}_{i[h]}$. The length-M fecundity vector on the seed-type basis is

$$\mathbf{F}_{ij,t}|\psi_{ij,t}, \rho_{ij,t} = \mathbf{m}_{i[h]}\rho_{ij,t}\psi_{ij,t} \tag{S2.6}$$

where $i[h]$ indicates the species for individual i. Of course, $\mathbf{M}$ must be estimated.

By Bayes' theorem we obtain from the fitted model the inverse $M \times H$ probability matrix $\mathbf{H}$ that an unknown seed of type m was produced by species h. The posterior estimate of total seed produced on the plot is $f_{hj,t} = \sum_{i \in h} \rho_{ij,t}\psi_{ij,t}$. Row m of the inverse matrix is

$$\mathbf{h}_{mj,t} = [h = 1, \ldots, H|\mathbf{m}_h]_{j,t} = \int \int \frac{\mathbf{m}_h f_{hj,t}}{\sum_h \mathbf{m}_h f_{hj,t}}[f_{hj,t}, \mathbf{m}_h]df_{hj,t}d\mathbf{m}_h \tag{S2.7}$$

where the integral is taken over the posterior distribution of $f_{hj,t}$ and $\mathbf{m}_h$.

Table S2.1: Additional prior parameter values.

	$\tilde{u}$	U^*	u_l	u_h	d_l	d_h	f_{max}
pinuEchi	162	30	92	365	15	40	200000
pinuRigi	162	30	92	253	18	40	50000
pinuStro	162	30	92	365	18	40	80000
pinuTaed	162	30	92	365	15	40	200000
pinuVirg	131	30	40	318	8	25	500000

S2.5 *Prior parameter values*

The prior parameter distribution depends on which components are included in the model. *Maturation* and *conditional fecundity* parameters have the prior distributions $MVN(\boldsymbol{\beta}^v|\mathbf{0}, 10 \times \mathbf{I}_{q^v})$, and $MVN(\boldsymbol{\beta}^x|\mathbf{0}, 10 \times \mathbf{I}_q)$, both non-informative–design matrices are centered and standardized, so the prior is 10 standard deviations. Random effects have the prior distribution $\boldsymbol{\beta}_i^w \sim MVN(\mathbf{0}, \mathbf{B}_w)$ and $\mathbf{B}_w \sim IW(\mathbf{I}_{q^w}, df)$ with degrees of freedom $df = q^w + \sqrt{n}_+$, the second term being rounded to the next integer. Fecundity has maximum value f_{max}.

Year effects have the prior distribution $\prod_t N(\gamma_t|0, 10)$. For random groups, $\gamma_{g,t} \sim N(\gamma_t, \tau_t^2)$, with group variance $\tau_t^2 \sim IG(2, 1)$. *Autoregressive lag terms* have the prior distribution $(\alpha_1, \ldots, \alpha_L) \sim MVN(\mathbf{0}, \mathbf{I}_p)$. If there are G random groups, they have the prior $(\alpha_{g,1}, \ldots, \alpha_{g,L}) \sim MVN(\mathbf{0}_G, \mathbf{A}_\alpha)$, with $IW(\mathbf{A}_\alpha|\mathbf{I}_p, p + 1)$.

Seed composition vectors $\mathbf{m}_h$ in the observation error matrix $\mathbf{M}$ have a Dirichlet prior distribution.

Dispersal parameters have the prior distribution $N(u_h|\tilde{u}, \tilde{U})I(u_l < u < u_h)$ for upper and lower bounds (u_l, u_h). The next stage is $\tilde{u} \sim N(u^*, U^*)$ and $\tilde{U} \sim IG(v_1, v_2)$,

where parameter values (u^*, U^*, v_1, v_2) are selected based on understanding of the species dispersal properties.

Additional prior parameter values are given in table S2.1. These include d_l, the diameter below which a tree of unknown maturation status is assumed to be immature; d_h, the diameter above which a tree of unknown status is assumed mature; and f_{max}, maximum fecundity. U^* is the prior variance in u between random groups, in this case species.

S2.6   *In- and out-of-sample prediction*

Seed data are generated by dispersal from mature and fecund trees. Data can be predicted either from the posterior estimates of these latent states and dispersal or from the posterior distribution of parameters. Let $\boldsymbol{\theta}_1 = \{\boldsymbol{\psi}, \boldsymbol{\rho}, \mathbf{u}, \mathbf{M}\}$ be the set of latent states, the dispersal kernel, and error matrix. The predictive distribution of seed data $\mathbf{Y}^*$ is

$$[\mathbf{Y}^*] = \int [\mathbf{Y}^*|\mathbf{F}(\boldsymbol{\psi}, \boldsymbol{\rho}), \mathbf{S}(\mathbf{u}), \mathbf{M}] \, [\boldsymbol{\theta}_1|\mathbf{Y}, \mathbf{X}, \mathbf{V}, \mathbf{z}, \mathbf{c}] d\boldsymbol{\theta}_1 \tag{S2.8}$$

From right to left, the interpretation of the integrand is i) the posterior estimates of states, dispersal, and observation error, and ii) the likelihood (variables are defined in table 3). This prediction is available for trees and years that are part of the model fitting. It can be used to predict seed at locations where seed rain was not observed on observed plots and years. It cannot be used to predict outside sample plots and years, because it requires estimates for the states of the trees themselves.

Predicting out-of-sample plots and years requires the posterior distribution of parameters $\boldsymbol{\theta}_2 = \{\boldsymbol{\beta}^x, \boldsymbol{\beta}^v, \sigma^2, \mathbf{u}, \mathbf{M}\}$, marginalizing the latent states $\{\boldsymbol{\psi}, \boldsymbol{\rho}\}$,

$$
\begin{aligned}
[\mathbf{Y}^*|\mathbf{X}^*, \mathbf{V}^*] = \int & [\mathbf{Y}^*|\mathbf{F}(\boldsymbol{\psi}, \boldsymbol{\rho}), \mathbf{S}^*(\mathbf{u}), \mathbf{M}] \\
& \times [\psi|\boldsymbol{\beta}^x, \sigma^2, \mathbf{X}^*] \, [\rho|\boldsymbol{\beta}^v, \mathbf{V}^*] \\
& \times [\boldsymbol{\theta}_2|\mathbf{Y}, \mathbf{X}, \mathbf{V}, \mathbf{z}, \mathbf{c}] \, d\boldsymbol{\theta}_2 d\boldsymbol{\psi} d\boldsymbol{\rho}
\end{aligned} \tag{S2.9}
$$

From right to left in the integrand we have the posterior distribution of parameters, maturation status, conditional fecundity, and likelihood. Predictions from eq. (S2.8) will be more accurate than eq. (S2.9), but, again, eq. (S2.8) is only available for in-sample prediction, which relies on the capacity to predict maturation and fecundity from environmental variables as opposed to simply estimating them and on scenarios for $\mathbf{X}^*$ and $\mathbf{V}^*$. The fully generative eq. (S2.9) can be used for fore- and backcasting. Comparing these predictions identifies when a combination of latent states may predict the data, but environmental predictors do not predict the states themselves.

S3   Data simulation

The simulator takes numbers of plots, trees, traps, and years as mean values for stochastic generation of sample size. The simulator follows these steps:

- Generate random species identities, random diameters, and random locations for trees.

- Draw coefficients for maturation, conditional fecundity, and dispersal parameters from a range that can generate patterns like observed data. Both maturation and conditional fecundity have an intercept and a slope for log diameter.

- Draw maturation status of each tree year from the probit submodel, subject to the constraint that this is a one-way transition (eq. (5), eq. (S2.5)).

- For mature individuals, draw conditional fecundity matrix $\mathbf{F} \sim MVN(\boldsymbol{\psi}, \mathbf{C}_n)$ from eq. (7).

- Assuming that only a fraction of seeds can be assigned to species, construct the $\mathbf{m}_h$ vector for each species h and distribute seed production for each individual seed types that will be recognized in traps.

- Generate a distance matrix from the tree locations to the trap locations and evaluate a corresponding kernel matrix $\mathbf{S}(|\mathbf{s}^{\mathcal{S}} - \mathbf{s}^{\mathcal{N}}|; u)$ with dispersal parameter u.

- Evaluate intensities $\mathbf{\Lambda}_{\mathcal{S}}|\mathbf{f}_{\mathcal{N}}$ from eq. (1) and draw counts from $y_{s,t} \sim Poi(\lambda_{s,t})$ from eq. (4). The variance in simulated $y_{s,t}$ is the horizontal axis in $fig.$ 5. The vertical axis shows the prediction from eq. (15) and its two terms.

Code is provided as part of the package MASTIF (http://rpubs.com/jimclark/457399). For fig. 5, the specific algorithm is:

S4   MEANS AND COVARIANCES

This section summarizes means and covariances observed by a consumer in the canopy and on the forest floor.

S4.1   *Space-lag covariance*

The *Tree-time and space-time covariance* of the main text starts with a space-time covariance matrix and moves immediately to summaries thereof. Here we provide additional explanation that is relevant to the interpretation of fig. 4 and fig. 11. In this section a matrix $\mathbf{L}$ can represent either host tree covariance $\mathbf{C}$ or spatial covariance $\mathbf{G}$.

We can estimate a lag matrix $\mathbf{L}$ consisting of vector blocks and elements $L_{ii',l}$ for lag $l = 1, l = -p, \dots, 0, \dots, p$. For the simplest example of two host trees or locations and one lag ($i = 1, i' = 2, p = 1$), this structure is

$$\mathbf{L} = \left[\begin{array}{cc} \text{AC}_1 & \text{CC}_{12} \\ \text{CC}_{21} & \text{AC}_2 \end{array} \right] = \left[\begin{array}{ccc|ccc} L_{11,-1} & L_{11,0} & L_{11,+1} & L_{12,-1} & L_{12,0} & L_{12,+1} \\ L_{21,-1} & L_{21,0} & L_{21,+1} & L_{22,-1} & L_{22,0} & L_{22,+1} \end{array} \right] \tag{S4.10}$$

Blocks contain elements $L_{ii',l}$. Blocks along the diagonal are the symmetric auto-covariance (AC) vectors, $L_{ii,-l} = L_{ii,+l}$. Off-diagonal blocks are cross-covariance vectors (CC), which are asymmetric, because they depend on which individual leads the other. However, each element in $\text{CC}_{ii'}$ has a counterpart in its mirror block $\text{CC}_{i'i}$, where $L_{ii',-l} = L_{i'i,+l}$. In

other words, $\mathbf{C}_L$ becomes block symmetric upon reversing the individual labels and the signs of lag indices in the upper (or lower) block.

Matrix $\mathbf{L}$ offers all combinations of tree-tree-lag covariances. Tree-to-tree covariances for a given lag are held in off-diagonal blocks (CC_{21} in eq. (S4.10)). Center elements of the off-diagonal blocks hold the covariances between trees in the same year ($L_{21,0}$) (fig. 4, left panel). These elements determine the resource heterogeneity that consumers abide by moving between trees, depending on their foraging ambits. The masting phenomenon is identified with episodic seed production in lagged year l', ($L_{ii,l'} > 0, l' > 1$) and synchronicity ($L_{ii',0} > 0$). The plots in fig. 11 show columns of $\mathbf{L}$ for sequential lags $0, 1, \ldots$.

Tree-to-lag covariances determine the advantages of movement between years, as when a good year for this tree might be followed by a good year for a different tree next year ($L_{ii',\pm 1} > 0$). These are shown for lag 1 in the right panel of fig. 4. A consumer might abide pulsed availability ($L_{ii,\pm 1} < 0$) if it has the capacity to switch hosts, and hosts are not synchronized ($L_{ii',0} < 0$). The same principles apply to ground foragers with limited spatial ambit. We use the lag matrix to evaluate the advantages of movement between host trees between years.

S4.2 *Entropy*

To quantify the heterogeneity experienced by a consumer across trees, space, years, and hosts we evaluated entropy,

$$\frac{1}{2}\left[\log(2\pi) + \frac{\log |\mathbf{V}|}{d}\right] \tag{S4.11}$$

where $|\mathbf{V}|$ is the determinant of a $d \times d$ matrix $\mathbf{V}$ that represents one of the covariance matrices $\mathbf{C}_L$ (tree-to-tree over years), $\mathbf{C}_n$ (year-to-year over trees), $\mathbf{G}_T$ (spatial over years) or $\mathbf{G}_S$ (year-to-year over space). This entropy calculation is reported on a per-dimension basis (per d) due to the multiplicative effect of integrating volume over d dimensions.

S5 STABILITY OF THE $\mathrm{AR}(p)$ PROCESS

The sequence of fecundities for each tree can be evaluated for its stability properties. Eigenvalues for time series quantify the tendency for quasi-periodic behavior. For the $\mathrm{AR}(p)$ model, we construct the matrix

$$\mathbf{E} = \begin{bmatrix} \psi_1 & \psi_2 & \ldots & \psi_{p-1} & \psi_p \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}$$

(West and Harrison, 1997). If there are random groups, then there is a matrix $\mathbf{E}_g$ for each group, combining the (global) fixed effect and the random effect for that group. A quasi-periodic process has complex eigenvalues. A stationary $\mathrm{AR}(p)$ process has all eigenvalues

of less than unit modulus (real plus imaginary parts). Eigenvalues are evaluated for the fecundity of each mature tree and summarized by plot and species. Eigenvalues for the AR(4) model are shown in fig. 10.

S6   Dispersal kernel

Seeds arriving in seed traps need not all derive from within the inventory plot (Clark et al., 1998; Muller-Landau et al., 2008; Clark et al., 2004). This possibility suggests an intercept proportional to basal area in eq. (1) as a rough accommodation of long-distance dispersal. We do not include an intercept here, because it can have a large impact on estimates, without actually being sensitive to trees outside the plot. An intercept can have a large impact, because it provides an alternative to the dispersal kernel anytime data are noisy, which is always the case when seed recovery is low. It is insensitive to LDD otherwise, because the tail of the kernel has no impact on estimates when seed except in cases where seed is rare (Clark et al., 1999). Distant trees do not affect estimates, because they do not influence the likelihood, as demonstrated with experiments where we incrementally increased the sizes of sample plots and showed these diminishing effects with distance (Clark et al., 1998).

The lack of effect on the likelihood is further demonstrated by integrating the kernel (eq. (S2.1)) arc-wise and with distance. Here is the mean distance associated with dispersal parameter u,

$$\bar{d} = \frac{\pi\sqrt{u}}{2} \tag{S6.12}$$

Seed arriving at a location from beyond distance R is

$$S(R) = \int_R^\infty \oint_{2\pi} \frac{u}{\pi(u+r^2)^2} dr = 2u \int_R^\infty \frac{r}{(u+r^2)^2} dr = \frac{u}{u+2R^2} \tag{S6.13}$$

(Clark et al., 1999), and a distance-specific intensity

$$l(r) = -\frac{d\log S}{dr} = \frac{4r}{u+2r^2} \tag{S6.14}$$

These kernel-derived estimates are multiplied by an assumed constant seed production per area of forest. For perspective, if the average seed of a well-dispersed species moves 10 m from the parent in a forest of infinite area that is everywhere the same in composition, then $< 1\%$ of all seed derives from beyond 50 m. Again, insensitivity to distant sources is shown in the analysis of estimates in (Clark et al., 1998). It is consistent with the high uncertainty in estimated fecundity for trees that are distant from all seed traps.

The likelihood does respond to seeds from outside the plot when there are no nearby sources. However, in that case there is no information to estimate the kernel. In cases where fleshy-fruits are dispersed by songbirds, there can still be substantial dispersal from nearby sources.

In addition to an undesirable effect on estimates, the intercept is costly. If we want to estimate how much seed might be arriving at a location from outside a plot, we would not integrate arcwise, but rather would require Reimann sums for areas that differ in

every direction from every seed trap. They would necessarily be truncated at some distance beyond which inputs would be assumed negligible. For MCMC, this would require reevaluation at each iteration, an extremely costly algorithm.

S7 DIAGNOSTICS

S7.1 *Simulated data*

Plots for simulation examples include coefficients, latent states, and predictions. An MCMC chain for the dispersal parameter in the simulation example described in the main text is shown in fig. S7.1a. If there is no learning from data, then chains would return the prior (shown at left in this figure). Instead, chains are concentrated on the true value.
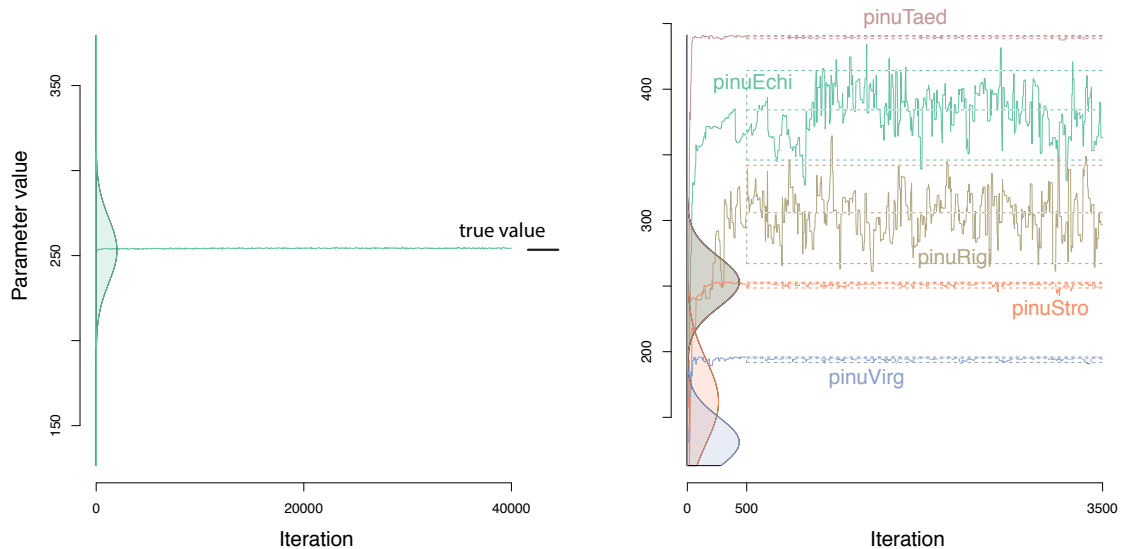

Figure S7.1: MCMC chains for the dispersal parameter fitted to data simulated with $u = 253$ and a non-informative prior (left) and fitted to *Pinus* data (right) with prior distributions shown on the vertical axis. Dashed lines show the post-burnin values used for estimates. For simulated data (a), the chain converges to the true value (blue line). For *Pinus* data species have prior distributions consonant with previous studies, but with low weight relative to the data–the posterior departs from the prior distribution as expected with Bayesian updating.

Maturation, fecundity, and seed composition parameters are plotted against true values in fig. S7.2. Tabular form is presented in table S7.2 and table S7.3. Wide credible intervals for intercepts in fig. S7.2a, b reflect the small numbers of trees on each plot, the fact that the intercept at diameter $= 0$ is outside the data, the product form of these two functions in the likelihood is not identified at small diameters, and, for fecundity, the log scale for near-zero values cannot be observed (fractions of a seed). Near zero diameter the model may not distinguish between the immature state versus low conditional fecundity, because both result in no seeds. However, the slope parameters used to interpret sensitivity are accurate.
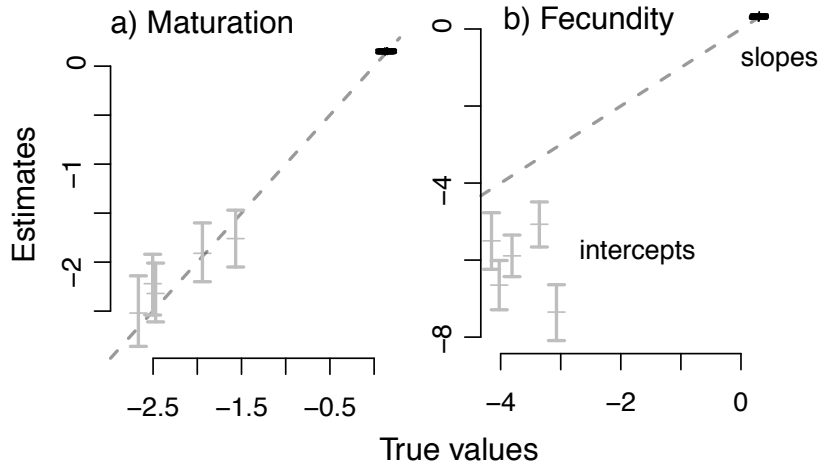
Figure S7.2: Parameter estimates plotted against values used to simulate data for maturation (a) and conditional fecundity (b) of four species. In both panels, low values are intercepts and high values are slopes. Also shown is the dashed line of agreement.

Table S7.2: Posterior estimates for maturation coefficients and true values for the simulated example.

	estimate	SE	CI 0.025	CI 0.975	true
pinuEchi	-1.760	0.147	-2.05	-1.47	-1.57
pinuRigi	-2.320	0.155	-2.61	-2.01	-2.47
pinuStro	-2.220	0.157	-2.54	-1.92	-2.50
pinuTaed	-2.520	0.185	-2.86	-2.14	-2.67
pinuVirg	-1.910	0.151	-2.20	-1.60	-1.94
pinuEchi:diam	0.152	0.007	0.138	0.165	0.120
pinuRigi:diam	0.153	0.006	0.141	0.164	0.144
pinuStro:diam	0.153	0.006	0.140	0.164	0.152
pinuTaed:diam	0.154	0.008	0.137	0.168	0.144
pinuVirg:diam	0.147	0.006	0.134	0.159	0.125

The error matrix $\mathbf{M}$ is recovered, with 1/3 of seeds identified to species and 2/3 to the unknown class (fig. S7.3). The poorest estimates come from plots where species-seed type combinations are missing or rare. These specific parameters do not affect the likelihood (because they are not represented in data) and, thus, do not affect the rest of the posterior distribution.

The simulator was not constructed to generate data containing lag effects. However, fitting an AR(3) model to data simulated without autoregressive terms returns 95% CIs that include zero for lag-3 (fig. S7.4), as expected for a time series that lacks lag effects, but slightly above zero for lag-1 and below for lag-2. This suggests caution in interpreting AR estimates that are near zero. The full MASTIF package is available in R for experiments with simulated data, showing posterior estimates, predictive diagnostics, DIC, root mean square predictive error (RMSPE), and mapped estimates of latent states. The summary diagnostics provided here can be supplemented with direct experiments by the user.

Table S7.3: Posterior estimates for fecundity coefficients and true values for the simulated example.

|	estimate	SE	CI 0.025	CI 0.975	true
pinuEchi	-5.890	0.280	-6.43	-5.35	-3.81
pinuRigi	-6.650	0.331	-7.29	-6.01	-4.02
pinuStro	-5.500	0.374	-6.24	-4.77	-4.15
pinuTaed	-7.350	0.361	-8.09	-6.64	-3.07
pinuVirg	-5.070	0.300	-5.66	-4.49	-3.36
pinuEchi:diam	0.309	0.007	0.296	0.323	0.3021
pinuRigi:diam	0.320	0.007	0.305	0.334	0.297
pinuStro:diam	0.296	0.008	0.280	0.312	0.290
pinuTaed:diam	0.354	0.008	0.338	0.371	0.304
pinuVirg:diam	0.297	0.008	0.282	0.312	0.296
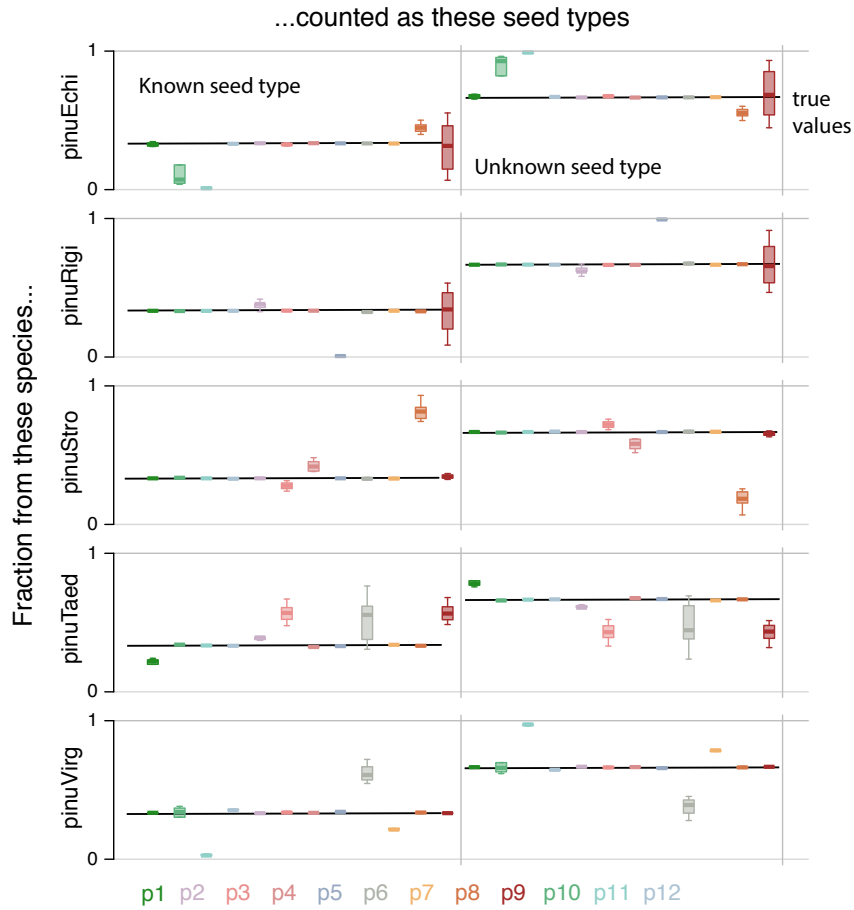
Figure S7.3: Estimates of the composition vectors $\mathbf{m}_h$ in simulation, where $h = 1, \ldots, 5$ species that can be classified as $M = 6$ seed types on plots $p1, \ldots, p12$. For each species, only 33% were recorded to the correct species, the remainder counted as "unknown". The model fitting correctly recovers 2/3 of seed in the undifferentiated genus class, shown as horizontal black lines, except in cases where reproductive trees are missing or rare.
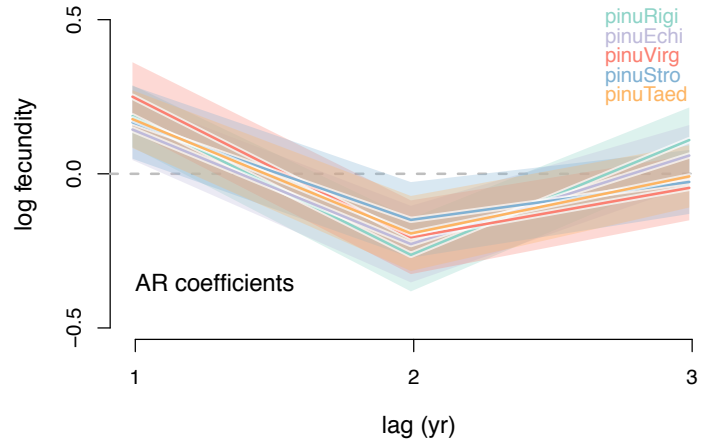
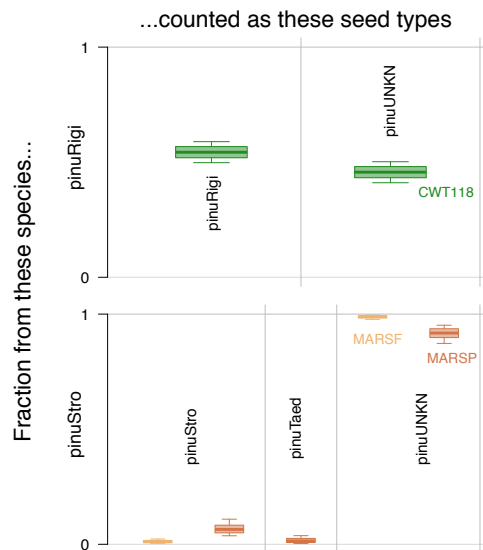Figure S7.4: Estimates of lag terms for simulated data generated without them for five species with 95% CIs.

Figure S7.5: The error matrix $\mathbf{M}$ for the *Pinus* example shows the probability that a seed produced by each species (rows) will be counted as each seed type (column).

S7.2   *Application to* Pinus

The fitted model included diameter as a predictor for maturation and fecundity, random effects for trees, and fixed and random effects for region-species groups in the AR(4) model (table 3). The $H \times M$ error matrix $\mathbf{M}$ has only a few fitted entries (fig. S7.5), because most are unambiguous, as when a plot comes from a location where only one species occurs or all seeds are identified only to genus level (DUKE BW, EE, EW, HW). Exceptions are CWT118, where *P. rigida* occurs, but there are also *P. strobus* in the region, and MARS HF and HP, where *P. strobus* occurs, but other pines occur in the region. The estimates in fig. S7.5 are the elements $\mathbf{M}_{h,m}$, the probability that a seed produced by species h (rows) is counted as species m (columns).

As for the simulated example, seed prediction from the fitted model shows the additional uncertainty when predicting from parameter estimates (fig. S7.6a) rather than from estimates of maturation and fecundity (fig. S7.6b). The former requires that covariates can be found to accurately predict maturation and fecundity. In cases where covariates do not explain fecundity, there can still be estimates of fecundity that predict seed trap

data.
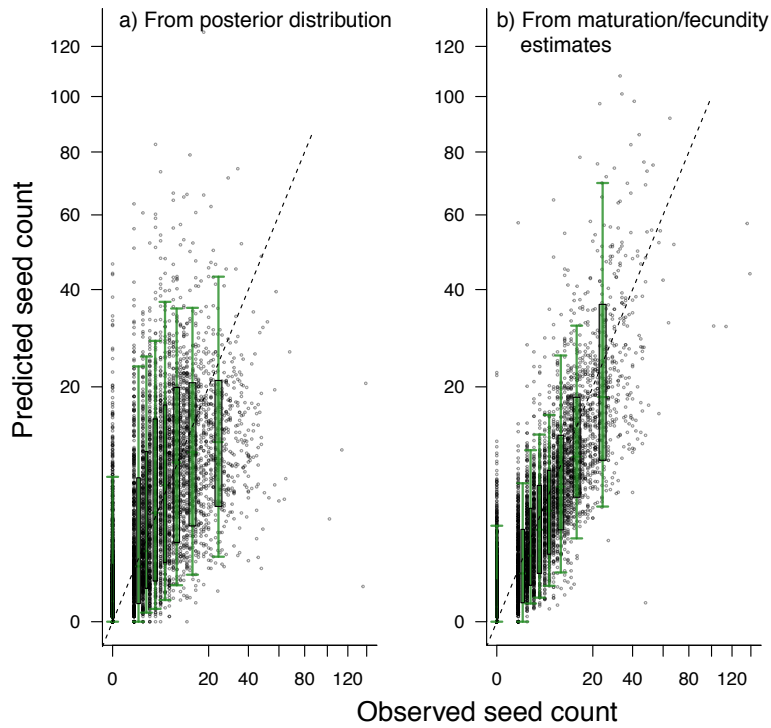
Figure S7.6: Seed prediction from the AR(3) model fitted to *Pinus*. Predictions are from the full posterior distribution (a) and from the fecundity estimates (b). Also shown is the dashed line of agreement.

Random coefficients associated with individual trees are shown in fig. S7.7. The random effects are not centered on zero, because they have been added to the fixed effect, thus allowing more informative species comparisons. The range of variation between individual trees differs by species and is greatest for *P. rigida*.

Figure S7.7: Random intercepts (added to fixed effects for each species) for trees from the coefficient vector $\boldsymbol{\beta}^w$ in the AR(4) model.

Model comparisons and variable selection by DIC, RMSPE, and prediction scores

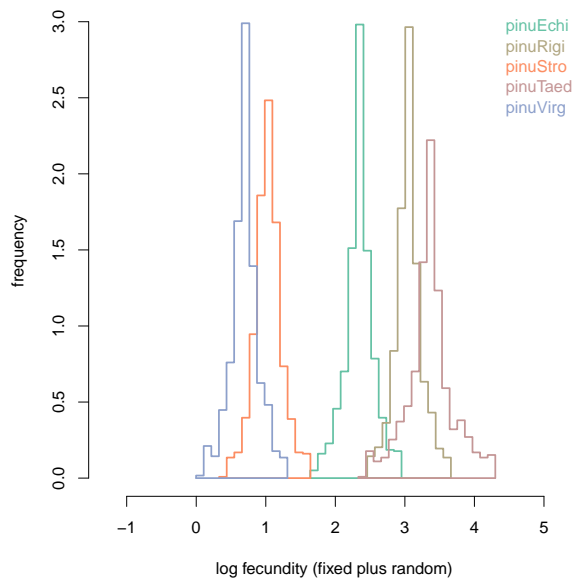available in the MASTIF package are not the focus here, where the attention centers on description of the basic model. They are provided as a table (table S7.4). Full MCMC chains are available for convergence diagnostics.

Table S7.4: Residual variance σ^2, RMSPE, and DIC.

	estimate	SE	CI_025	CI_975
σ^2	10.70	0.14	10.50	11.00
rmspe	3.76	0.04	3.68	3.83
deviance	-71700	2500	-76900	-67500

The partial autocorrelation in fecundity (table S7.5, standard errors in table S7.6) are provided as output from MASTIF.

Table S7.5: Partial autocorrelation in log fecundity.

	lag-1	lag-2	lag-3	lag-4	lag-5	lag-6	lag-7	lag-8
pinuEchi-DUKEBW	0.486	0.182	0.177	-0.042	0.063	0.011	-0.030	0.040
pinuEchi-DUKEEW	0.573	0.153	0.169	0.043	-0.020	-0.081	0.045	-0.000
pinuRigi-CWT118	0.370	0.123	0.121	-0.008	0.028	0.057	0.022	0.044
pinuStro-HARVBW	0.176	0.070	0.043	0.007	-0.017	-0.003	0.000	0.001
pinuStro-HARVS	0.182	0.072	0.055	0.001	-0.019	-0.004	0.001	0.002
pinuStro-MARSF	0.178	0.072	0.147	0.021	0.088	-0.001	-0.019	0.026
pinuStro-MARSP	0.191	0.140	0.099	0.012	0.032	0.037	-0.021	0.018
pinuStro-SCBILFDP	0.226	0.106	0.036	-0.047	-0.005	0.005	0.004	-0.001
pinuTaed-DUKEBW	0.494	0.213	0.075	0.049	0.039	0.040	-0.029	0.002
pinuTaed-DUKEEW	0.492	0.268	0.080	0.046	0.033	-0.017	-0.026	-0.019
pinuTaed-DUKEFACE1	0.395	0.156	0.147	0.007	0.042	-0.012	0.009	0.019
pinuTaed-DUKEFACE5	0.382	0.169	0.096	0.049	0.035	-0.001	0.015	0.030
pinuTaed-DUKEFACE6	0.361	0.161	0.097	0.072	0.019	0.024	0.001	0.010
pinuTaed-DUKEHW	0.457	0.164	0.179	0.024	0.042	0.004	-0.115	0.026
pinuVirg-DUKEBW	0.404	0.090	0.206	0.037	0.184	-0.016	0.015	-0.137
pinuVirg-DUKEEW	0.518	0.228	0.071	-0.006	0.073	-0.008	-0.103	0.062
pinuVirg-SCBILFDP	0.592	0.148	-0.319	0.070	0.108	-0.072	-0.018	0.034

S8 Posterior simulation: algorithm notes

S8.1 *Maturation updates*

Where available, maturation status is determined by observations $z_{ij,t}$. The sampling distribution for unobserved maturation states in eq. (5) conditions on previous and current state,

$$\rho_{ij,t} \sim Bernoulli(p_{ij,t}) \tag{S8.15}$$

Table S7.6: Partial autocorrelation standard error in log fecundity

	lag-1	lag-2	lag-3	lag-4	lag-5	lag-6	lag-7	lag-8
pinuEchi-DUKEBW	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
pinuEchi-DUKEEW	0.000	0.000	0.001	0.000	0.000	0.001	0.001	0.001
pinuRigi-CWT118	0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.001
pinuStro-HARVBW	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
pinuStro-HARVS	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
pinuStro-MARSF	0.001	0.002	0.001	0.001	0.000	0.000	0.000	0.000
pinuStro-MARSP	0.004	0.003	0.003	0.003	0.002	0.002	0.001	0.001
pinuStro-SCBILFDP	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
pinuTaed-DUKEBW	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
pinuTaed-DUKEEW	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
pinuTaed-DUKEFACE1	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000
pinuTaed-DUKEFACE5	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
pinuTaed-DUKEFACE6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
pinuTaed-DUKEHW	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
pinuVirg-DUKEBW	0.001	0.001	0.000	0.000	0.000	0.000	0.001	0.000
pinuVirg-DUKEEW	0.005	0.005	0.006	0.002	0.004	0.004	0.002	0.007
pinuVirg-SCBILFDP	0.024	0.042	0.015	0.022	0.017	0.003	0.008	0.003

where

$$p_{ij,t} = \begin{cases} \rho_{ij,t-1} + (1 - \rho_{ij,t-1})\rho_{ij,t+1}\Phi\left(\mathbf{v}'_{ij,t}\boldsymbol{\beta}^v\right) & \text{missing } z_{ij,t} \\ z_{ij,t} & \text{otherwise} \end{cases}$$

and $\boldsymbol{\Phi}(\cdot)$ is the standard normal distribution function.

S8.2 *Joint fecundity and maturation updates*

To impute states, fecundity and maturation status are either proposed jointly (this section) or conditional fecundity of currently imputed mature individuals is updated separately (next section). For joint updates, we factor maturation and conditional fecundity as

$$[\psi^*_{ij,t}, \rho^*_{ij,t}] = [\psi^*_{ij,t}|\rho^*_{ij,t}][\rho^*_{ij,t}] \tag{S8.16}$$

Maturation year is a random walk, centered on the currently imputed maturation year and subject to constraints imposed by observed or currently imputed states (mature individuals cannot become immature, mature individuals remain mature). Possible maturation years range from the last year in which an individual was observed in the immature state to the first year in which that individual was observed in the mature state (section S2.3). If there are no maturation/fecundity observations, then all information on maturation status comes through the seed data. Maturation is proposed and accepted jointly with fecundity for all trees on a plot year, with proposals subject to this one-way constraint on maturation (section S2.3). This blocking is necessitated by the fact that the likelihood for each trap-year conditionally depends on all tree-years for that plot (Clark et al., 2004). Sampling of a plot-year block is summarized this way:

$$[\boldsymbol{\psi}_{j,t}, \boldsymbol{\rho}_{j,t}|\mathbf{Y}_{j,t}, \mathbf{z}_{j,t}, \{\mathbf{c}_{ij,t}\}_{i=1}^{n_j}, \mathbf{M}] \propto P_1 \times P_2 \times P_3 \times P_4 \tag{S8.17}$$

where

$$P_1 = \prod_{s=1}^{S_j} \prod_{m=1}^{M} Poi\left(y_{smj,t}|A_{sj,t}\lambda_{smj,t}(\boldsymbol{\psi}_{j,t}, \boldsymbol{\rho}_{j,t}, \mathbf{u}, \mathbf{M})\right)$$

$$P_2 = \prod_{i=1}^{n_{ij}} N(\log \psi_{ij,t}|\mathbf{x}'_{ij,t}\beta^x + \ldots, \sigma^2)I(\psi_{ij,t} \leq 1)^{1-\rho_{ij,t}}I(\psi_{ij,t} > 1)^{\rho_{ij,t}}$$

$$P_3 = \prod_{i=1}^{n_{ij}} Bernoulli(\rho_{ij,t}|p_{ij,t})$$

$$P_4 = \prod_{i=1}^{n_{ij}} binom(c_{ij,t}^n|\psi_{ij,t}, c_{ij,t}^f)$$

where $p_{ij,t}$ is given in eq. (S8.15).

Two methods are used, depending on the process model, both beginning with proposed $\{\rho_{ij,t}^*|\rho_{j,t-1}, \rho_{j,t+1}\}_{i=1}^{n_j}$. Only cases where $\rho_{ij,t-1} = 0$ and $\rho_{ij,t+1} = 1$ can change maturation state. A new maturation state is proposed with probability 0.5. Fecundity is then proposed for all proposed mature ij from a normal distribution, with censoring imposed by the proposed $\rho_{ij,t}^*$. Here are the two methods that follow:

Method 1:

1. The proposed $\psi_{ij,t}^*|\rho_{ij,t}^*$ come from a normal distribution centered on the currently imputed $\psi_{ij,t}$ and censored as in eq. (7).

2. Accept/reject $\boldsymbol{\rho}_{j,t}^*, \boldsymbol{\psi}_{j,t}^*$ as a block with probabilities for current and proposed from $\prod_{k=1}^{4} P_k$.

Method 2:

1. The proposed $\psi_{ij,t}^*|\rho_{ij,t}^*$ come from the conditional normal distribution obtained from P_2. Details for this sampling distribution are given for the most complex case of the AR(p) model in section S8.8.5.

2. Accept/reject $\boldsymbol{\rho}_{j,t}^*, \boldsymbol{\psi}_{j,t}^*$ as a block with probability $P_1 \times P_3 \times P_4$.

S8.3 *Hamiltonian updates of conditional fecundity*

Hamiltonian updates accelerate mixing for continuous states, but they cannot be used with discrete states, in this case maturation status. However, mixing of fecundity is accelerated with Hamiltonian updates for currently imputed mature individuals. Each observation is a length-M vector $\mathbf{y}_{sj,t}$. Element m has Poisson intensity $A_{sj}\lambda_{smj,t} = A_{sj}\sum_{i=1}^{n_j,t} \mathbf{S}_{[sjt,i]}e^{\psi_{ij,t}}\mathbf{m}_{i[h]}$, where $\mathbf{m}_{i[h],m}$ is the row vector $\mathbf{m}_h$ corresponding to the species of individual i and the column for seed type m. The Hamiltonian can be written as

$$H(\psi_{ij,t}, p) = B(\psi_{ij,t}) + C(p) \tag{S8.18}$$

where $C(p) = \sum_{i=1}^{n_j} \frac{p_i^2}{2m_i}$ is the kinetic energy, taken as a quadratic function of momentum variables m_i, which are tuned to optimize performance (Neal, 2011). The first term incorporates the conditional distribution,

$$B(\psi_{ij,t}) = -\log[\psi_{ij,t}|\mathbf{y}_{j,t}, \mu_{ij,t}, \sigma^2)] \tag{S8.19}$$

$$\propto \sum_{m,s} (-y_{smj,t} \log \lambda_{smj,t} + A_{sj}\lambda_{smj,t}) + \frac{1}{2\sigma^2}(\psi_{ij,t} - \mu_{ij,t})^2 \tag{S8.20}$$

The gradient is used to direct proposals efficiently:

$$\frac{\partial B}{\partial \psi_{ij,t}} = e^{\psi_{ij,t}} \sum_m \mathbf{m}_{i[h],m} \sum_s \mathbf{S}_{[sjt,i]} \left(-\frac{y_{smj,t}}{\lambda_{smj,t}} + A_{sj}\right) + \frac{1}{\sigma^2}(\psi_{ij,t} - \mu_{ij,t}) \tag{S8.21}$$

Hamiltonian updates are individually slow, but affect larger steps then a Metropolis random walk, especially with large data sets. The two methods are mixed stochastically in the Gibbs sampler.

S8.4 *Fecundity coefficients*

Direct sampling of coefficients in $\boldsymbol{\beta}^x$ and $\boldsymbol{\beta}^v$ is available from Gaussian conditional posterior distributions. Gaussian prior distributions are non-informative. For $\boldsymbol{\beta}^x$ conditional distributions marginalize random effects (see below). The variance σ^2 has an inverse gamma prior – and is sampled directly from the conjugate inverse gamma posterior.

S8.5 *Random individual effects*

Let $\mathbf{w}_{ij,t}$ be a design vector holding all or some of the columns in $\mathbf{x}_{ij,t}$. There is an individual-effects coefficient matrix $\boldsymbol{\beta}_{ij}^w$,

$$\psi_{ij,t} \sim N\left(\mathbf{x}'_{ij,t}\boldsymbol{\beta}^x + \mathbf{w}'_{ij,t}\boldsymbol{\beta}_{ij}^w, \sigma^2\right) \tag{S8.22}$$

The prior distribution includes

$$\boldsymbol{\beta}_{ij}^w|\mathbf{B}_w \sim MVN(\mathbf{0}, \mathbf{B}_w) \tag{S8.23}$$

$$\mathbf{B}_w \sim IW\left(\mathbf{B}^w, df\right) \tag{S8.24}$$

where $df = Q^w + 2$, Q^w is the number of columns in $\mathbf{w}_{ij,t}$, and $\mathbf{B}^w = \mathbf{I}_r$ is a prior diagonal matrix. The conditional posterior matrix is

$$\boldsymbol{\beta}_{ij}^w|\boldsymbol{\beta}^x, \mathbf{B}_w \sim MVN(\mathbf{V}_{ij}\mathbf{v}_{ij}, \mathbf{V}_{ij}) \tag{S8.25}$$

where

$$\mathbf{v}_{ij} = \frac{1}{\sigma^2} \sum_{t \in \{t_i\}} \mathbf{w}_{ijt}(\psi_{ijt} - \mathbf{x}'_{ijt}\boldsymbol{\beta}^x) \tag{S8.26}$$

$$\mathbf{V}_{ij} = \frac{1}{\sigma^2} \sum_{t \in \{t_i\}} \mathbf{w}_{ijt}\mathbf{w}'_{ijt} + \mathbf{B}_w^{-1} \tag{S8.27}$$

The summations are taken over all observation years for an individual i, the set $\{t_i\}$ for which the individual is mature. Here is the conditional for the covariance,

$$\mathbf{B}_w | \{\boldsymbol{\beta}_{ij}^w\} \sim IW \left(\sum \boldsymbol{\beta}_{ij}^w \boldsymbol{\beta}_{ij}^{w'} + df \times \tilde{\mathbf{B}}, \sum_j n_j + df \right) \tag{S8.28}$$

where $\tilde{\mathbf{B}}$ is the prior covariance.

S8.6   *Random groups in the year and AR(p) models*

The year and AR(p) models allow for group random effects on year and lag coefficients, respectively–if groups are defined by the user, they will be treated as random. This is done because year and lag terms across groups are highly unbalanced. Plot-species groups can hold different numbers of plots and trees in different years. Plots can be established at different times, have different plot areas, and support very different communities of species. For a given species, abundance across plots may range from zero to high. Within plots, numbers of mature individuals vary across years with recruitment, maturation, and death. Within posterior simulation, their imputed maturation statuses change by tree and year. The sizes of design matrices are thus dynamic.

Given this imbalance, treating groups as random provides the advantage that no arbitrary rules are needed to catch computation errors that would result from plot-years that are at some iterations imputed to have mature trees and other iterations not.

S8.7   *Year effects*

For a single group, years effects are fixed, drawn from the conditional

$$\gamma_t \sim N(V_t v_t, V_t) \tag{S8.29}$$

$$V_t^{-1} = \frac{n_t}{\sigma^2} + 1/\tau^2 \tag{S8.30}$$

$$v_t = \frac{1}{\sigma^2} \sum_i (\psi_{i,t} - \mu_{i,t}) \tag{S8.31}$$

With multiple groups, there are random year effects across groups:

$$\gamma_{g,t} \sim N(V_t v_t, V_t) \tag{S8.32}$$

$$V_{g,t}^{-1} = \frac{n_{g,t}}{\sigma^2} + 1/\tau_t^2 \tag{S8.33}$$

$$v_{g,t} = \frac{1}{\sigma^2} \sum_{i \in g} (\psi_{i,t} - \mu_{i,t} - \gamma_t) \tag{S8.34}$$

Years have a sum-to-zero constraint imposed in Gibbs sampling. The intercept for a given year is the overall intercept plus the year effect for that year. The variance for random effects is

$$\tau_t^2 \sim IG \left(2 + \frac{G_t}{2}, 1 + \frac{1}{2} \sum_g \gamma_{g,t}^2 \right) \tag{S8.35}$$

where G_t is the number of groups available in year t, i.e., those having mature individuals in that year.

S8.8   *AR(*p*) model*

The AR(p) model allows for the dependence of the current states of ψ_t on p previous states. The process is homogeneous in time, because the lag coefficients α_p are constant. We start with a few words on structure.

S8.8.1   *Imputed past, predicted future*

AR models handle the early years in different ways. There is no AR(p) estimate for years $t \in \{1, \ldots, p\}$. One of the more common ways to deal with these years is to simply condition on them. This seems like a big price to pay. Because MASTIF is a state-space model, and we are imputing fecundity and maturation anyway, it makes sense to impute fecundity/maturation for years $t - p, \ldots, t - 1$.

So while we are imputing the past, it makes sense to predict the future. Conditionally, fecundity in the final year T_i depends on the future, up to year $T_i + p$. To accommodate past and future, MASTIF imputes backward p years from the first observation and predicts forward p years beyond the last observed year.

A consistent treatment would appear to demand that AR effects be restricted to individuals that have been mature for the past $t - p$ years. Note that this would not be a concern if maturation state was known. We adopt this rule, so lag effect estimates are not biased downward by the inclusion of trees that might have immature during one or more of years $t - p, \ldots, t - 1$. Although fecundity is imputed for all years, including before observations began, sampling of coefficients for fixed effects and lag effects is restricted to years in which trees were observed and mature.

S8.8.2   *AR(*p*) model structure*

To avoid further notation, the description that follows applies only to tree-years for which the mature state extends back to $t - p$ years. Also to simplify notation we omit the subscript j. Note that multiple plots j might fall within a group g.

Conditionally, the model for an individual i in group $g \in \{1, \ldots, G\}$ can be written as

$$\psi_{ig,t} | \mu_{ig,t}, \boldsymbol{\alpha}, \boldsymbol{\alpha}_{g[i]}, \tilde{\boldsymbol{\psi}}_{ig,t} \sim N\left(m_{ig,t}, \sigma^2\right) \tag{S8.36}$$

where

$$m_{ig,t} = \mu_{ig,t} + \sum_{l=1}^{p} (\alpha_l + \alpha_{g[i]l}) \psi_{ig,t-l} \tag{S8.37}$$

$$= \mu_{ig,t} + (\boldsymbol{\alpha} + \boldsymbol{\alpha}_{g[i]})' \tilde{\boldsymbol{\psi}}_{ig,t} \tag{S8.38}$$

$\mu_{ig,t} = \mathbf{x}'_{ig,t} \boldsymbol{\beta}^x$ is the fixed effect, $\tilde{\boldsymbol{\psi}}_{ig,t} = (\psi_{ig,t-1}, \ldots, \psi_{ig,t-p})'$ is the vector of lagged fecundities for (ig, t), $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)'$ is the vector of fixed effects for lag $l = 1, \ldots, p$, and $\boldsymbol{\alpha}_g = (\alpha_{g1}, \ldots, \alpha_{gp})$ is the random effect for group g with prior distribution

$$\boldsymbol{\alpha}_g \sim MVN(\mathbf{0}, \mathbf{A}_\alpha) \tag{S8.39}$$

$$\mathbf{A}_\alpha \sim IW\left(\tilde{\mathbf{A}}_\alpha, df\right) \tag{S8.40}$$

To facilitate sampling, the fecundity values are organized into a vector $\boldsymbol{\psi} = \{\psi_{ig,t}|i = 1, \ldots, n, g = 1, \ldots, G, t = 1, \ldots, T_i\}$ and a corresponding matrix of p lag terms. For example, a vector with these subscripts

$$\boldsymbol{\psi} = (\psi_{i,g,t}, \psi_{i,g,t+1}, \ldots, \psi_{i,g,T_i}, \psi_{i+1,g,t}, \ldots,) \tag{S8.41}$$

has the lag matrix with matching rows and p columns,

$$\tilde{\boldsymbol{\Psi}} = \begin{pmatrix} \psi_{i,g,t-1} & \cdots & \psi_{i,g,t-p} \\ \psi_{i,g,t} & \cdots & \psi_{i,g,t+1-p} \\ \vdots & \vdots & \vdots \\ \psi_{i,g,T_i-1} & \cdots & \psi_{i,g,T_i-p} \\ \psi_{i+1,g,t} & \cdots & \psi_{i+1,g,t+1-p} \\ \vdots & \vdots & \vdots \end{pmatrix} \tag{S8.42}$$

S8.8.3 *Sample fixed effects*

To sample fixed effects, we move a few terms to the left,

$$\mathbf{m} = \tilde{\boldsymbol{\Psi}}\boldsymbol{\alpha} \tag{S8.43}$$

where $\mathbf{m}$ has elements $\psi_{ig,t} - \mu_{ig,t} - \boldsymbol{\alpha}'_{g[i]}\tilde{\boldsymbol{\psi}}_{ig,t}$, and, again, $\boldsymbol{\alpha}_{g[i]}$ indicates the vector of lags for the group to which individual i belongs. The conditional posterior matrix for fixed effects is

$$\boldsymbol{\alpha}|\{\boldsymbol{\alpha}_g\} \sim MVN(\mathbf{V}\mathbf{v}, \mathbf{V}) \tag{S8.44}$$

$$\mathbf{v} = \sigma^{-2}\tilde{\boldsymbol{\Psi}}'\mathbf{m} \tag{S8.45}$$

$$\mathbf{V}^{-1} = \sigma^{-2}\tilde{\boldsymbol{\Psi}}'\tilde{\boldsymbol{\Psi}} + 0.001 \times \mathbf{I}_p \tag{S8.46}$$

S8.8.4 *Random group effects*

For random effects, make a slight change in the mean vector and write

$$\boldsymbol{\alpha}_g \sim MVN(\mathbf{V}_g\mathbf{v}_g, \mathbf{V}_g) \tag{S8.47}$$

$$\mathbf{v}_g = \frac{1}{\sigma^2}\sum_{i,t}\tilde{\boldsymbol{\psi}}'_{ig,t}m_{ig,t} \tag{S8.48}$$

$$\mathbf{V}_g = \frac{1}{\sigma^2}\sum_{i,t}\tilde{\boldsymbol{\psi}}_{ig,t}\tilde{\boldsymbol{\psi}}'_{ig,t} + \mathbf{A}_\alpha^{-1} \tag{S8.49}$$

where $m_{ig,t} = \psi_{ig,t} - \mu_{ig,t} - \boldsymbol{\alpha}' \tilde{\boldsymbol{\psi}}_{ig,t}$. The summations are taken over all observation years in which individual i has been in the mature state for the previous p years, for all individuals in group g. Here is the conditional for the $p \times p$ covariance matrix,

$$\mathbf{A}_\alpha | \{\boldsymbol{\alpha}_g\} \sim IW \left(\sum_{g=1}^{G} \boldsymbol{\alpha}_g \boldsymbol{\alpha}'_g + df \times \mathbf{I}_p, G + df \right) \tag{S8.50}$$

S8.8.5   *Latent states*

Latent states in the $AR(p)$ model are sampled by proposing from the conditional posterior for the fecundity/maturation submodel $\psi_t, z_t | \psi_{\{-t\}}, z_{t-1}, z_{t+1}$, where $\psi_{\{-t\}}$ is the set of all fecundity values except t, and accepting from the likelihood (Method 2 in section S8.2). To reduce clutter, we now omit subscripts ijg. If there are random groups in the model, then everything below is handled at the group level, with lag coefficient α_l being replaced with $\alpha_l + \alpha_{gl}$ for group g.

To isolate the terms in ψ_t the $AR(p)$ model can be written as

$$\psi_t \sim N \left(m_t, \sigma^2 \right) \tag{S8.51}$$

where

$$m_t = \mu_t + \sum_{l=1}^{p} \alpha_l \psi_{t-l} \tag{S8.52}$$

The exponent of the conditional distribution $\psi_t | \psi_{\{-t\}}$ can be factored this way:

$$\frac{1}{\sigma^2} \left[(\psi_t - m_t)^2 + \sum_{k=1}^{p} (n_{t,k} - \alpha_k \psi_t)^2 \right] \tag{S8.53}$$

where

$$n_{t,k} = \psi_{t+k} - \mu_{t+k} - \sum_{l=1}^{p} \alpha_l \psi_{t+k-l} I(l \neq k) \tag{S8.54}$$

To sample latent states, we propose from

$$\psi_t \sim N \left(V v_t, V \right) \tag{S8.55}$$

where

$$v_t = \frac{1}{\sigma^2} \left(m_t + \sum_{k=1}^{p} n_{t,k} \alpha_k \right) \tag{S8.56}$$

$$V^{-1} = \frac{1}{\sigma^2} \left(1 + \sum_{k=1}^{p} \alpha_k^2 \right) \tag{S8.57}$$

Proposals are accepted as a block for each plot-year in the data set, based on the likelihood for seed data (see Method 2).

S8.9 *Other parameters*

The error variance σ^2 is sampled from the conditional inverse gamma posterior distribution.

If there are no random groups, the dispersal parameter u is sampled with Metropolis, with an adaptive proposal variance and truncated normal prior distribution,

$$[u] \propto L \times N(u|u_0, U_0)I(u_{min} < u < u_{max}) \tag{S8.58}$$

where the likelihood L is $\prod_{r,s,j,t} Poi\left(y_{rsj,t}|A_{sj}\lambda_{smj,t}(u, \boldsymbol{\psi}_{j,t}, \boldsymbol{\rho}_{j,t}, \mathbf{m})\right)$, u_0 and U_0 are the prior mean and variance dispersal parameters, and (u_{min}, u_{max}) is an interval selected with prior knowledge.

If there are random groups, then there is an additional stage for the global mean. The previous distribution applies to u_g for group g,

$$[u_g] \propto L \times N(u_g|u, U) \tag{S8.59}$$

The u_g are proposed and accepted as a block. The global mean and variance have conditional distributions:

$$u|u_1, \ldots, u_G, u_0, U, U_0 \sim N\left(Vv, V\right) \tag{S8.60}$$

$$V^{-1} = \frac{G}{U} + \frac{1}{U_0} \tag{S8.61}$$

$$v = \frac{1}{U}\sum_g u_g + \frac{u_0}{U_0} \tag{S8.62}$$

$$U|u_1, \ldots, u_G, u \sim IG\left(2 + \frac{G}{2}, 1 + \frac{1}{2}\sum_g (u_g - u)^2\right) \tag{S8.63}$$

References

Bourg, N. A., W. J. McShea, J. R. Thompson, J. C. McGarvey, and X. Shen. 2013. Initial census, woody seedling, seed rain, and stand structure data for the SCBI SIGEO Large Forest Dynamics Plot. Ecology **94**:2111–2112. URL https://doi.org/10.1890/13-0010.1.

Clark, J., S. LaDeau, and I. Ibanez. 2004. Fecundity of trees and the colonization-competition hypothesis. Ecological Monographs **74**:415–442.

Clark, J., E. Macklin, and L. Wood. 1998. Stages and spatial scales of recruitment limitation in southern Appalachian forests. Ecological Monographs **68**:213–235.

Clark, J., M. Silman, R. Kern, E. Macklin, and J. HilleRisLambers. 1999. Seed dispersal near and far: Patterns across temperate and tropical forests. Ecology **80**:1475–1494.

LaDeau, S. L., and J. S. Clark. 2006. Elevated CO2 and tree fecundity: the role of tree size, interannual variability, and population heterogeneity. Global Change Biology **12**:822–833. URL `http://dx.doi.org/10.1111/j.1365-2486.2006.01137.x`.

Muller-Landau, H. C., S. J. Wright, O. Calderon, R. Condit, and S. P. Hubbell. 2008. Interspecific variation in primary seed dispersal in a tropical forest. Journal of Ecology **96**:653–667. URL `<GotoISI>://WOS:000256635800010`.

Neal, R., 2011. MCMC Using Hamiltonian Dynamics, Book section 5, pages 1–49 . CRC Press, New York, New York.

West, M., and J. Harrison. 1997. Bayesian Forecasting and Dynamic Models. 2nd edition. Springer.