

# INDIVIDUAL-BASED INTEGRAL PROJECTION MODELS: THE ROLE OF SIZE-STRUCTURE ON EXTINCTION RISK AND ESTABLISHMENT SUCCESS

SEBASTIAN J. SCHREIBER AND NOAM ROSS

ABSTRACT.

- (1) Matrix models or integral projection models (IPMs) are commonly used to study the dynamics of structured populations, where discrete or continuous traits influence survival, growth, or reproduction.
- (2) When a population's size is small, as is often the case for threatened species or potentially invasive species arriving in novel habitats, extinction risk may be substantial due to demographic stochasticity. Branching processes, which are individual-based counterparts to matrix models and IPMs, allow one to quantify these risks of extinction.
- (3) For discretely structured populations, the theory of multi-type branching processes provides analytic methods to compute how extinction risk changes over time and how it depends on the size and composition of the population. Building on prior work on continuous-state branching processes, we extend these analytic methods to individual-based models accounting for any mixture of discrete and continuous population structure.
- (4) The individual-based IPMs are defined by probabilistic update rules at the level of the individual which determine how each individual with a given trait value dies, changes trait values (e.g. grows in size), or produces individuals with the same or other trait values. We show that probabilities of extinction can be analytically determined by probability-generating functionals associated with the

individual-based IPMs. In particular, we present analytical expressions for how extinction probabilities change over time and depend on the initial abundance and trait distribution of the population.

- (5) We illustrate how to numerically implement these methods using data from the short-lived desert shrub species *Cryptantha flava*, and provide a more general discussion of how to implement these methods to other data sets including those involving fluctuating environmental conditions.
- (6) As most IPM studies have the necessary data to parameterize individual-based IPMs, these methods provide a computationally efficient means to explore how continuously structured populations differing in their evolutionary history and environmental context may differ in their vulnerability to extinction or ability to colonize new habitats.

## INTRODUCTION

Computations of extinction probabilities or likelihoods of establishment success lie on opposing sides of a theoretician's coin and have been used to address theoretical and practical issues in conservation biology, restoration ecology, biological invasions and population genetics. Risks of extinction or establishment failure stem from populations consisting of a finite number of individuals, each of which faces a non-zero risk of mortality on any given day. These extinction risks are shaped, in part, by the size and composition of a population whose individuals may differ in age, size, geographical location, or other important characteristics influencing demography. When population structure is finite-dimensional (e.g. a finite number of age classes, stages, geographical locations), multi-type branching processes can model these extinction risks and, thereby, serve as the stochastic, finite-population counterpart of matrix models [Harris, 1963, Athreya and Ney, 2004, Caswell, 2001, Haccou et al., 2005]. These stochastic models have been used successfully to address a

49 diversity of questions concerning fixation probabilities of beneficial alleles [Patwa and Wahl, 2008],  
50 evolutionary emergence of pathogens [Antia et al., 2003, Park et al., 2013], extinction risk of small  
51 populations [Boyce, 1992, Gosselin and Lebreton, 2000, Fujiwara and Caswell, 2001, Erickson et al.,  
52 in press], and establishment success in heterogeneous environments [Haccou and Iwasa, 1996, Haccou  
53 and Vatunin, 2003, Schreiber and Lloyd-Smith, 2009].

54 To parameterize matrix models or multi-type branching processes, individuals must be discretely  
55 categorized into a finite number of types. However, when collecting demographic data, researchers  
56 commonly measure continuous traits (e.g. mass, length, geographical location) about individuals  
57 and use continuum-based statistics to approximate “fine-grained” discrete-traits. Integral projec-  
58 tion models (IPMs) allow one to account for this continuous population structure [Easterling et al.,  
59 2000]. These IPMs can be viewed as infinite-dimensional matrix models which can be numerically  
60 approximated by finite-dimensional matrix models. Consequently, many of the standard demo-  
61 graphic concepts and methods for matrix models (e.g. stable state distributions, reproductive  
62 values, life table response experiments, sensitivity analysis) exist for IPMs [Easterling et al., 2000,  
63 Ellner and Rees, 2006, 2007, Rees and Ellner, 2009, Coulson, 2012, Ellner and Schreiber, 2012,  
64 Metcalf et al., 2013, Rees et al., 2014, Merow et al., 2014].

65 Here, we describe individual-based counterparts of IPMs using continuous-state branching pro-  
66 cesses [Harris, 1963]. For these finite population, stochastic models, we present an analytical method  
67 for computing extinction probabilities. As these methods are easily implemented numerically, they  
68 circumvent the need to use individual-based simulations and allow one to efficiently study how ex-  
69 tinction probabilities or establishment failure depend on continuous as well as discrete population  
70 structure. We illustrate the application of these methods with an individual-based IPM of the short-  
71 lived desert shrub *Cryptantha flava* from Utah, USA [Salguero-Gómez et al., 2012, Salguero-Gómez  
72 et al., 2014].

## THE GENERAL MODELS AND METHODS

**The Individual-Based IPM.** We consider an individual-based model where the set of all possible individual states (e.g. age, size, geographical location, etc.) lies in a compact metric space  $X$ . For a standard size-structured IPM,  $X = [a, b]$  corresponds to the range of sizes measured in the field where  $a$  is the minimal size and  $b$  is the maximal size. For models with a mixture of age and size structure,  $X$  could be given by  $\{1, \dots, T\} \times [a, b]$  where  $T$  corresponds to the maximal age of an individual. For spatially structured populations,  $X = [a, b]$  may correspond to locations along a transect or  $X = [a, b] \times [c, d]$  may correspond to latitude and longitude coordinates of individuals.

Following Harris [1963], we consider finite populations in which the state of the population at any point in time is characterized by the different states  $(x_1, x_2, \dots, x_k)$  of individuals within the population and the number of individuals in each state  $(n_1, n_2, \dots, n_k)$ . Specifically, if there  $n_1$  individuals in state  $x_1$ ,  $n_2$  individuals in state  $x_2$ , ...,  $n_k$  individuals in state  $x_k$ , then the state of the population is given by  $\mathbf{s} = (n_1, n_2, \dots, n_k; x_1, x_2, \dots, x_k)$ . The set of all possible population states is

$$\mathcal{S} = \{(\mathbf{n}; \mathbf{x}) = (n_1, \dots, n_k; x_1, \dots, x_k) : k, n_i \in \mathbb{N}, x_i \in X\} \cup \{0\}$$

where  $\mathbb{N} = \{1, 2, 3, \dots\}$  denotes the natural numbers and 0 is the extinction state corresponding to no individuals in the population.

Let  $\mathbf{s}_t = (\mathbf{n}; \mathbf{x}) \in \mathcal{S}$  be the population state at time  $t$ . The dynamics of  $\mathbf{s}_t$  are determined by a set of probabilistic rules that determine the contribution of each individual in the population to the population in next time step  $t + 1$ . These “contributions” may correspond to an individual surviving and changing state (e.g. growing in size, getting older, dispersing to another geographical location), or an individual having offspring. Consistent with standard branching process theory, each individual updates independently of all other individuals in the population.

The update rule for an individual in state  $x$  is given by a probability measure  $m(x, d\mathbf{s})$  on the state space  $\mathcal{S}$ . Specially, the probability an individual in state  $x$  contributes  $\mathbf{s}$  individuals to the

97 population in the next time step where  $\mathbf{s}$  lies in a subset  $A \subset \mathcal{S}$  equals

$$\mathbb{P}[\mathbf{s}_1 \in A | \mathbf{s}_0 = (1; x)] = \int_A m(x, d\mathbf{s})$$

98 where the left hand side reads “the probability the population state lies in  $A$  at time 1 after initially  
99 having only one individual in state  $x$  at time 0.”

100 If the population state is currently  $\mathbf{s}_t = (n_1, \dots, n_k; x_1, \dots, x_k)$ , then the state  $\mathbf{s}_{t+1}$  is determined  
101 as follows:

- 102 (1) for each of the  $n_1$  individuals in state  $x_1$ , randomly and independently choose the number  
103 of replacement individuals from distribution  $m(x_1, d\mathbf{s})$ ,
- 104 (2) repeat step (1) for the states  $x_2, \dots, x_k$ , and
- 105 (3) determine the new population state  $\mathbf{s}_{t+1}$  by identifying the states of all individuals and  
106 counting the total number of individuals in each of these states.

107 This iterative algorithm can be used to create individual-based simulations of the individual-based  
108 IPM. As with any branching process, stochastic realizations of this process, with probability one,  
109 either go to extinction in finite time, or the population abundance grows without bound. This latter  
110 event is typically interpreted as a population becoming established or persisting.

111 **Probability-generating functionals and extinction probabilities.** We can characterize the  
112 probabilistic state of the system using probability-generating functionals  $\Psi$  (pgfs). Unlike moment-  
113 generating functionals as used by Harris [1963], the pgfs allow us to directly compute how extinction  
114 probabilities change in time as well as compute asymptotic extinction probabilities.

115 To define the pgf  $\Psi$ , we introduce some notation. To motivate this notation, suppose  $h(x)$  equals  
116 the probability that an individual of size  $x$  dies and has no offspring over the next year. If the  
117 current population state is  $\mathbf{s} = (x_1, \dots, x_k; n_1, \dots, n_k)$ , then independence of individuals implies  
118 that  $\prod_{i=1}^k h(x)^{n_i}$  is the probability that population goes extinct by the next time step i.e., all  
119 individuals die and have no offspring. With this motivation in mind, given *any* continuous function

120  $h : X \rightarrow \mathbb{R}$  and population state  $\mathbf{s} \in \mathcal{S}$ , define

$$h^{\mathbf{s}} = \begin{cases} \prod_{i=1}^k h(x_i)^{n_i} & \text{if } \mathbf{s} = (n_1, \dots, n_k; x_1, \dots, x_k) \\ 1 & \text{if } \mathbf{s} = 0. \end{cases}$$

121 The probability-generating functional  $\Psi$  is defined by

$$(1) \quad \Psi(h)(x) = \int h^{\mathbf{s}} m(x, d\mathbf{s})$$

122 and corresponds to the expected value of  $h^{\mathbf{s}_1}$  where  $\mathbf{s}_1$  is the population state at time 1, given the  
 123 population initially consisted of a single individual in state  $x$ . This expectation requires integrating  
 124 over all possible states  $\mathbf{s}$  that the population may take on in the next time step due to the different  
 125 possible contributions of the individual in state  $x$ . Appendix A describes how this definition of  $\Psi$   
 126 simplifies to the pgf for the classical multi-type branching processes.

127 Two facts allow us to use  $\Psi$  to calculate extinction probabilities. First, when  $h_0$  is the zero  
 128 function (i.e.,  $h_0(x) = 0$  for all  $x$ ),  $\Psi(h_0)(x)$  is the probability the population goes extinct in one  
 129 time step, given that initially it consisted of one individual in state  $x$ . That is,

$$(2) \quad \Psi(h_0)(x) = \int_{\{0\}} 1 \times m(x, d\mathbf{s}) + \int_{\mathcal{S} \setminus \{0\}} 0 \times m(x, d\mathbf{s}) = \mathbb{P}[\mathbf{s}_1 = 0 | \mathbf{s}_0 = (1; x)].$$

130 using the fact that the integral over  $\mathcal{S}$  equals the integral over the extinction state  $\{0\}$  plus the  
 131 integral over the non-extinction states  $\mathcal{S} \setminus \{0\}$ . Secondly, we claim that the  $t$ -fold composition of  $\Psi$   
 132 applied to  $h_0$  and evaluated at the individual state  $x$ ,

$$\Psi^t(h_0)(x) = \underbrace{\Psi(\Psi(\dots \Psi(h_0) \dots))}_{t \text{ times}}(x)$$

133 is the probability of extinction by time  $t$  given that the population initially consisted of one indi-  
 134 vidual in state  $x$ :

$$(3) \quad \Psi^t(h_0)(x) = \mathbb{P}[\mathbf{s}_t = 0 | \mathbf{s}_0 = (1; x)] =: E_t(x).$$

135 To verify this claim, we argue by induction. Equation (2) implies that equation (3) holds for  
 136  $t = 1$ . Now suppose that equation (3) holds at time  $t$ ; we will show it holds at time  $t + 1$ . On the  
 137 event that  $\mathbf{s}_1 = (n_1, \dots, n_k; x_1, \dots, x_k)$  is the population state at time 1, extinction occurs by time  
 138  $t + 1$  only if each of the lineages of the  $n_1 + \dots + n_k$  individuals go extinct in the next  $t$  time steps.  
 139 As the fates of these lineages are independent of one another, it follows that

$$\begin{aligned} \mathbb{P}[s_{t+1} = 0 | \mathbf{s}_0 = (1; x), \mathbf{s}_1 = (n_1, \dots, n_k; x_1, \dots, x_k)] &= \prod_{i=1}^k \mathbb{P}[\mathbf{s}_{t+1} = 0 | \mathbf{s}_1 = (1; x_i)]^{n_i} \\ &= \prod_{i=1}^k (\Psi^t(h_0)(x_i))^{n_i} = \Psi^t(h_0)^{\mathbf{s}_1} \end{aligned}$$

140 where the second equality follows from our inductive hypothesis. By the law of total probability, we  
 141 can integrate the conditional probabilities of extinction over all possible population states at time  
 142  $t = 1$  (i.e.  $\mathbf{s}_1$ ) to get

$$E_{t+1}(x) = \mathbb{P}[s_{t+1} = 0 | \mathbf{s}_0 = (1; x)] = \int [\Psi^t(h_0)]^{\mathbf{s}} m(d\mathbf{s}, x)$$

143 which, by definition, equals  $\Psi^{t+1}(h_0)(x)$  as claimed.

144 Equation (3) can be used to compute extinction probabilities iteratively. Furthermore, as indi-  
 145 viduals update independent of one another, the probability of the population going extinct by time  
 146  $t$  for any initial condition  $\mathbf{s} = (n_1, \dots, n_k; x_1, \dots, x_k)$  equals

$$(4) \quad \mathbb{P}[\mathbf{s}_t = 0 | \mathbf{s}_0 = \mathbf{s}] = E_t^{\mathbf{s}}.$$

These analytic expressions allow us to efficiently compute extinction probabilities by constructing a numerical approximation of the pgf  $\Psi$  and iterating it with an initial condition of a zero vector which corresponds to the numerical approximation of the zero function.

As  $E_0(x) \leq E_1(x) \leq E_2(x) \dots$  for any  $x \in X$  and  $E_t(x) \leq 1$  for all  $t$ , there is a well defined limit corresponding to the probability of eventual extinction:

$$E_\infty(x) := \lim_{t \rightarrow \infty} E_t(x).$$

Using moment-generating functionals with suitable technical hypotheses, Harris [1963] showed that  $E_\infty(x) < 1$  for all  $x$  if the dominant eigenvalue of the mean-field IPM is greater than one, and  $E_\infty(x) = 1$  for all  $x$  otherwise.

#### AN ILLUSTRATION WITH A SHORT-LIVED PERENNIAL

We illustrate these general methods using an individual-based IPM for the yellow-flowered perennial plant, Plateau Yellow Miner's Candle (*Cryptantha flava*), of the borage family (*Boraginaceae*). For populations growing along the Colorado Plateau, USA, Salguero-Gómez et al. [2012] developed an IPM using data collected from 2004 to 2010 [Salguero-Gómez et al., 2014]. Here, we use a subset of this data available in an R package, IPMpack [Metcalf et al., 2014]. All code to for this example is archived at Zenodo [Schreiber and Ross, 2015].

[Figure 1 about here.]

In the model, the state  $x$  of an individual is the size, which equals the total number of vegetative and flowering rosettes. If  $N_t(x)$  denotes the density of individuals of size  $x$  in year  $t$ , then Salguero-Gómez et al. [2012] used an IPM of the form

$$N_{t+1}(x) = \int_a^b [s(y)G(y, x) + e(x)p(y)f(y)N_t(y)] dy$$



166 where  $s(y)$  is the probability of surviving to the next year for individuals of size  $y$ ,  $G(y, x) dy$  is the  
 167 infinitesimal probability that a surviving individual of size  $y$  is size  $x$  in the next year,  $p(y)$  is the  
 168 probability that an individual of size  $y$  flowers,  $f(y)$  is the mean number of offspring produced by  
 169 an individual of size  $y$ , and  $e(x)$  is the infinitesimal probability that an offspring is of size  $x$  at the  
 170 time of the annual census.

171 Following Salguero-Gómez et al. [2012], we use generalized linear models (GLMs) for most of the  
 172 functional forms of the demographic kernels. Computations were performed using the base GLM  
 173 function in R [R Core Team, 2015]. We used logistic regression (i.e. a GLM with the binomial  
 174 family) for determining  $s(y)$  and  $p(y)$ , and a GLM with the Poisson family for modelling  $f(y)$ .  
 175 For the growth kernel, linear regression determined the expected size of an individual in the next  
 176 year and the actual size was assumed to be normally distributed about this mean. The variance  
 177 of this normal distribution, for simplicity, was assumed to be independent of the current size of  
 178 an individual. We modeled  $e(y)$  with a gamma distribution fit to the the empirical distribution of  
 179 germinants. Figure 1 shows the data and fits for  $s(y)$ ,  $G(y, x)$ ,  $p(x)$  and  $f(x)$ .

180 The kernels  $s(x)$ ,  $G(y, x)$ ,  $p(x)$  and  $e(x)$  provide us with all the information that the individual-  
 181 based IPM requires for probabilistic updating individuals for survival, growth, flowering, reproduc-  
 182 tion, and size of germinating individuals at first census. The fecundity kernel  $f(y)$ , however, only  
 183 specifies the mean number of offspring produced, but for an individual-based IPM, we need the  
 184 distribution of the number of offspring produced by an individual. Fortunately, this information is  
 185 built into the structure of the GLMs due to the assumptions in our model choice. As the fecundity  
 186 data was modeled using a Poisson family for the GLM, the mean number  $f(x)$  of offspring also  
 187 specifies the distribution. More generally, one might use multi-parameter distributions such as a  
 188 zero-inflated Poisson or a negative binomial, in which case parameters in addition to the mean are  
 189 needed to specify the distribution of offspring number.

190 **Deriving the pgf  $\Psi$ .** To define  $\Psi$ , we observe that the contributions of an individual of size  $x$  to  
 191 the population in the next time step involves the sum of two independent random variables: the  
 192 contribution due to survival and growth and the contribution due to reproduction. We will identify  
 193 two pgfs,  $\Psi_g$  and  $\Psi_f$ , for each of these processes separately. Then, we make use of a fundamental  
 194 property of pgfs

195 **Fundamental Property 1:** The pgf for a sum of independent random variables is the product of  
 the pgfs of these random variables.

196 to get that

$$\Psi = \Psi_g \times \Psi_f.$$

197 To write down each of these pgfs, we make use of two other fundamental property of pgfs:

198 **Fundamental Property 2:** The pgf for a sum of  $N$  independent, identically distributed random  
 variables  $X_i$  is the composition of the pgf for  $N$  with the pgf for the  $X_i$ .

199 **Fundamental Property 3:** Let  $X_1, X_2, \dots, X_n$  be  $n$  independent, random variables with pgfs  $\Psi_1,$   
 $\Psi_2, \dots, \Psi_n$ . If  $Y = X_i$  with probability  $p_i$  where  $p_1 + p_2 + \dots + p_n = 1$ , then the pgf of  $Y$  is  
 $\sum_{i=1}^n p_i \Psi_i$ .

200 For survival and growth,  $\Psi_g(h)(x)$  corresponds to integrating  $h^{\mathbf{s}}$  over all possible contributions  
 201  $\mathbf{s}$  from an individual of size  $x$  surviving and growing. These contributions are of two types:  $\mathbf{s} = 0$   
 202 when the individual dies, and  $\mathbf{s} = (1; y)$  when the individual survives and grows to size  $y$ . The  
 203 first event occurs with probability  $1 - s(x)$  and the infinitesimal probability of the second event is  
 204  $s(x)G(y, x)dy$ . As  $h^{\mathbf{s}} = 1$  when  $\mathbf{s} = 0$  and  $h^{\mathbf{s}} = h(y)$  when  $\mathbf{s} = (1; y)$ , integrating over all possible  
 205 contributions due to survival and growth yields

$$\Psi_g(h)(x) = (1 - s(x)) \times 1 + s(x) \int h(y)G(y, x)dy.$$

206 For fecundity,  $\Psi_f(h)(x)$  is given by integrating  $h^{\mathbf{s}}$  over all possible states  $\mathbf{s}$  corresponding to the  
 207 offspring produced by an individual of size  $x$ . To write this down, we begin by conditioning on the

event that an individual of size  $x$  flowers. On this event, the individual produces a Poisson number  
 $N$  of offspring with mean  $f(x)$ . The pgf for  $N$  is given by  $\phi(x, \xi) = \exp(-f(x)(\xi - 1))$  where  $\xi$  is  
a dummy variable. The size of each of these offspring is drawn interdependently from the common  
offspring distribution  $e(y)dy$ . Hence, the contribution of a flowering individual of size  $x$  is the sum  
of  $N$  independent random variables with distribution  $e(y)dy$ . By **Fundamental Property 2** of  
pgfs, we need to take the composition of the pgf  $\phi$  for  $N$  with the pgf  $\Psi_e$  for a single offspring,  
namely

$$\Psi_e(h)(x) = \int h(y)e(y)dy.$$

Thus, we get the pgf associated with a flowering individual is the following composition of pgfs

$$\Psi_{flowering}(h)(x) = \phi(x, \int h(y)e(y)dy)$$

To get the pgf for flowering and non-flowering contributions to fecundity, we observe that fecundity  
contributions of an individual of size  $x$  is given by 0 with probability  $1 - p(x)$  and by a flowering  
individuals with probability  $p(x)$ . By the **Fundamental Property 3**, we have the pgf for fecundity  
contributions equals

$$\Psi_f(h)(x) = (1 - p(x)) \times 1 + p(x)\phi(x, \int h(y)e(y)dy)$$

where 1 is the pgf for the zero random variable.

Using **Fundamental Property 1**, we multiply  $\Psi_g$  and  $\Psi_f$  to get  $\Psi$ , the pgf for the individual-  
based IPM:

$$\Psi(h)(x) = \left( (1 - s(x)) \times 1 + s(x) \int h(y)G(y, x)dy \right) \left( 1 - p(x) + p(x)\phi(x, \int h(y)e(y)dy) \right).$$

In Appendix B, we illustrate how to derive pgf for two alternative forms of this model including  
monocarpic species.

225 **Numerically implementing and using the pgf  $\Psi$ .** To approximate  $\Psi$  numerically, we discretize  
 226 the size interval  $[0, 2b]$  using  $n = 500$  equal sized intervals of width  $\Delta x = 2b/n$ . We used the larger  
 227 size interval of  $[0, 2b]$ , where  $b$  is the size of the largest observed individual, to minimize the effects  
 228 of eviction (see below). We created a vector  $\vec{x}$  corresponding to the midpoints of these intervals.  
 229 Using this vector we discretized the survival function as a vector  $\vec{s} = s(\vec{x})$ , the growth kernel as  
 230 a matrix using the outer product  $G$  of the growth kernel  $g$  with  $\vec{x}$ , the probability of flowering  
 231 function as a vector  $\vec{p} = p(\vec{x})$ , the fecundity function as a vector  $\vec{f} = f(\vec{x})$ , and the offspring size  
 232 distribution as a vector  $\vec{e} = e(\vec{x})$ .

233 For the previously described methods to work, it is critical that column sums for  $G$  and the sum  
 234 of  $\vec{e}$  equal one. For most IPMs, this will not occur automatically due to individuals being evicted  
 235 from the size interval  $[0, 2b]$ . There are a variety of ways to handle this issue [Williams et al., 2012].  
 236 As the offspring size vector nearly summed to one, we simply re-normalized it so that it summed  
 237 to one. For the growth matrix  $G$ , we treated eviction as mortality. To do this, we took one minus  
 238 the column sums of  $G$ , subtracted these sums from the survival vector, and then normalized the  
 239 column sums of  $G$  so that they added to one. When taking the product of survival and growth, the  
 240 resulting mean-field IPM is unaffected by these changes.

241 Using these discretized demographic components and the pgf  $\phi$  for fecundity, we get the discretized  
 242 pgf  $\Psi$ , given by

$$\Psi_{discrete}(\vec{h}) = (1 - \vec{s} + \vec{s} \circ (G^T \vec{h})) \circ (1 - \vec{p} + \vec{p} \circ \phi(\vec{x}, \vec{e}^T \vec{h}))$$

243 where  $\vec{h}$  corresponds to a discretized function i.e. a vector of length  $n$ ,  $^T$  denotes the transpose of  
 244 a matrix or vector, and  $\circ$  denotes element by element multiplication.

245 Iterating  $\Psi$  yields how extinction probabilities  $E_t(x)$  vary with time for a population initiated with  
 246 a single individual (Figure 2). Intuitively, this figure illustrates that the probability of extinction  
 247 decreases with the size of founding individual, and that extinction probabilities increase over time.  
 248 Furthermore,  $E_t(x)$  as  $t$  increases are approaching limiting extinction probabilities  $E_\infty(x)$  which

always equals one (Figure 2A). This stems from the fact that the dominant eigenvalue of the mean-field IPM is less than one, consistent with the results of Salguero-Gómez et al. [2012]. By increasing seed survivorship by a factor of three, the dominant eigenvalue of the mean-field IPM becomes greater than one and the asymptotic extinction probabilities become less than one (Figure 2B). We approximated these asymptotic extinction probabilities by iterating  $\Psi$  until the difference between  $E_t$  and  $E_{t+1}$  were below a specified error tolerance.

[Figure 2 about here.]

To scale things up to an entire population initially in state  $\mathbf{s}_0$ , recall that the extinction probability at time  $t$  is  $E_t^{\mathbf{s}_0}$ . Figure 3 illustrates how the extinction probabilities over time vary for a population with initially 100 individuals of the smallest size  $x = 1$  (i.e.  $\mathbf{s}_0 = (1; 100)$ ) versus a population with 5 or 8 individuals of the largest size  $x = 60$  typically observed in the field (i.e.,  $\mathbf{s}_0 = (60; 5)$  or  $(60; 8)$ ). Figure 3 suggests that, from the extinction risk perspective, about 6 or 7 larger individuals are equivalent to 100 of the smallest individuals. These types of comparisons may be particularly useful when trying to assess whether planting small or larger individuals are more effective for establishment success.

[Figure 3 about here.]

We also used the data to see how extinction risk over different time frames depends on the size and composition of the population. Specifically, for different founding population abundances  $N$ , we randomly sampled  $N$  individuals from the data and computed extinction risk of this sampled population over 5 and 10 year periods (Figure 4). For each founding population abundance  $N$ , we considered 500 random samples of size  $N$ . Figure 4A illustrates that, on average, log extinction risk decreases with the founding population size  $N$  and is greater for the 10 year period than the 5 year period. For smaller founding population sizes  $N$ , there is substantial overlap in the distributions of extinction times for the 5 and 10 year time frames. This overlap stems from founding populations of mostly large individuals being more likely to persist at least 10 years than founding populations

274 of mostly small individuals persisting at least 5 years. Consistent with this explanation, Figure 4B  
 275 illustrates that mean size of an individual within a founding population has a strong negative  
 276 correlation with (log) extinction probability, and the slope of this correlation is steeper over shorter  
 277 time frames than longer time frames.

278 [Figure 4 about here.]

## 279 RECOMMENDATIONS, EXTENSIONS, AND FUTURE CHALLENGES

280 To implement the methods presented here, there are two main steps. First, one needs to identify  
 281 the main demographic processes of the population, the order in which these processes occur relative  
 282 to the censuses used for data collection, and develop the statistical models for the each of the  
 283 demographic processes. Rees et al. [2014] and Merow et al. [2014] provide excellent reviews on the  
 284 philosophical and methodological issues associated with this step. Typically, whenever a study has  
 285 sufficient data for constructing the mean-field IPM, there is no need to collect any additional data  
 286 to build the individual-based IPM. Unlike the mean-field IPMs, though, the individual-based IPMs  
 287 make use of the complete distributional information associated with fecundity. As with other areas  
 288 of stochastic demography, the shapes, not just the means, of these distributions may have significant  
 289 effects on the likelihood of extinction or establishment success [Lloyd-Smith et al., 2005]. Hence,  
 290 it is best to examine several options (e.g. Poisson versus negative Binomial versus zero-inflated  
 291 distributions) to identify which distribution does a better job of describing the fecundity data. As  
 292 in all areas of modeling, if there is significant uncertainty about the “best” choice of the model,  
 293 one should perform the analyses with each of the alternative fecundity distributions to identify the  
 294 sensitivity of predictions of extinction risk to these alternatives.

295 The second step, the focus of this paper, involves constructing the probability-generating func-  
 296 tional  $\Psi$ . For the uninitiated, this step may be intimidating. However, there are a three basic  
 297 principles that simplify this construction. First, while this pgf  $\Psi$  takes functions to functions, one

should focus on writing down  $\Psi(h)(x)$ , which involves understanding the contributions of a single individual of size  $x$  to the next census. Second, one can often break up these contributions into a sum of independent contributions, find the pgfs associated with these simpler contributions, and then use the two fundamental properties of pgfs to “stitch” together  $\Psi$ . Third, the distributions used to describe the number of offspring produced by an individual typically involve random variables (e.g. Poisson, negative binomial) for which the associated pgfs are well-known. Finally, whenever in doubt, find a collaborator that you trust to help put the pieces together correctly.

For the individual-based models considered here, we assumed the environment remains constant over time. However, IPMs have and continue to be used to study the effects of fluctuating environmental conditions on population demography and life history evolution [Childs et al., 2004, Dahlgren and Ehrlén, 2011, Rees and Ellner, 2009]. The methods described here easily extend to fluctuating environments. Specifically, if  $\Psi_t$  is the pgf of the individual-based IPM associated with year  $t$ , then the probability of extinction by year  $t + 1$  for a population initially in state  $\mathbf{s}$  is

$$\Psi_0(\Psi_1(\dots \Psi_t(h_0)))^{\mathbf{s}}$$

where  $h_0$ , as before, is the zero function (see Appendix C for justification). Note that the composition here is in the reverse order of what one does when iterating the mean-field IPM model forward in time. While we know of no formal proof, in the case of a stationary and ergodic environment with suitable technical assumptions, we conjecture the following limit theorem holds: the asymptotic extinction probabilities are strictly less than one if and only if the stochastic growth rate (aka dominant Lyapunov exponent) of the mean-field IPM is positive. For multi-type branching processes, Athreya and Karlin [1971, Theorem 12] proved this result under suitable technical hypotheses.

Despite this relatively straightforward extension to temporally variable environments, many challenges remain. From the computational perspective, finding the efficient methods to deal with

multi-dimensional states variables (e.g. size and location, or multi-dimensional traits) continues to be a challenge, as it is for mean-field IPMs. While the analytical methods presented here cover multi-dimensional state variables, their numerical implementation involves approximating multi-dimensional integrals which can be computationally expensive. From the analytical perspective, accounting for temporal correlations in individual growth or reproductive rates (e.g. individuals that grew larger than expected in one year being more likely to grow larger than expected in the next year) or correlations among individuals are particularly important challenges as strong correlations likely have large effects on extinction risk. Finally, and perhaps most importantly, the manners in which size-structured demography may shape extinction risk for real-world population remains to be understood. One might hope that by applying these methods to the many data sets for which IPMs have been developed, as well as future data sets, might provide a computationally efficient means to explore how size-structure for populations differing in their evolutionary history and environmental context influences their vulnerability to extinction.

**Acknowledgements.** We thank William Cuello, Eric Eager, Richard Erickson, Vadim Karatayev, Jacob Moore, and an anonymous reviewer for providing comments on an earlier draft of this manuscript.

## REFERENCES

- R. Antia, R.R. Regoes, J.C. Koella, and C.T. Bergstrom. The role of evolution in the emergence of infectious diseases. *Nature*, 426:658–661, 2003.
- K. B. Athreya and P. E. Ney. *Branching processes*. Dover Publications Inc., Mineola, NY, 2004.
- K.B. Athreya and S. Karlin. On branching processes with random environments: I: Extinction probabilities. *The Annals of Mathematical Statistics*, pages 1499–1520, 1971.
- M.S. Boyce. Population viability analysis. *Annual review of Ecology and Systematics*, pages 481–506, 1992.



- 345 H. Caswell. *Matrix Population Models*. Sinauer, Sunderland, Massachusetts, 2001.
- 346 D.Z. Childs, M. Rees, K.E. Rose, P.J. Grubb, and S.P. Ellner. Evolution of size-dependent flowering  
347 in a variable environment: construction and analysis of a stochastic integral projection model.  
348 *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 271:425–434, 2004.
- 349 T. Coulson. Integral projections models, their construction and use in posing hypotheses in ecology.  
350 *Oikos*, 121:1337–1350, 2012.
- 351 J.P. Dahlgren and J. Ehrlén. Incorporating environmental change over succession in an integral  
352 projection model of population dynamics of a forest herb. *Oikos*, 120:1183–1190, 2011.
- 353 M.R. Easterling, S.P. Ellner, and P.M. Dixon. Size-specific sensitivity: applying a new structured  
354 population model. *Ecology*, 81:694–708, 2000.
- 355 S. P. Ellner and S. J. Schreiber. Temporally variable dispersal and demography can accelerate the  
356 spread of invading species. *Theoretical Population Biology*, 82:283–298, 2012.
- 357 S.P. Ellner and M. Rees. Integral projection models for species with complex demography. *American*  
358 *Naturalist*, 167:410–428, 2006.
- 359 S.P. Ellner and M. Rees. Stochastic stable population growth in integral projection models: theory  
360 and application. *Journal of Mathematical Biology*, 54:227–256, 2007.
- 361 R.A. Erickson, J.C. Eager, E.A. and Stanton, J.A. Beston, and Thogmartin W.E. Diffendorfer,  
362 J.E. Assessing local population vulnerability with branching process models. *Ecosphere*, in press.
- 363 M. Fujiwara and H. Caswell. Demography of the endangered north atlantic right whale. *Nature*,  
364 414:537–541, 2001.
- 365 F. Gosselin and J. Lebreton. Potential of branching processes as a modeling tool for conservation  
366 biology. In *Quantitative methods for conservation biology*, pages 199–225. Springer, 2000.
- 367 P. Haccou and Y. Iwasa. Establishment probability in fluctuating environments: a branching process  
368 model. *Theoretical Population Biology*, 50:254–280, 1996.

- 369 P. Haccou and V. Vatutin. Establishment success and extinction risk in autocorrelated environ-  
370 ments. *Theoretical Population Biology*, 64:303–314, 2003.
- 371 P. Haccou, P. Jagers, and V.A. Vatutin. *Branching processes: variation, growth, and extinction of*  
372 *populations*. Cambridge University Press, 2005.
- 373 T.E. Harris. *The theory of branching processes*. Springer-Verlag, 1963.
- 374 J. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz. Superspreading and the impact of  
375 individual variation on disease emergence. *Nature*, pages 355–359, 2005.
- 376 C. Merow, J.P. Dahlgren, C.J.E. Metcalf, D.Z. Childs, M.E.K. Evans, E. Jongejans, S. Record,  
377 M. Rees, R. Salguero-Gómez, and S.M. McMahon. Advancing population ecology with integral  
378 projection models: a practical guide. *Methods in Ecology and Evolution*, 5:99–110, 2014.
- 379 C.J.E. Metcalf, S.M. McMahon, R. Salguero-Gómez, and E. Jongejans. Ipmpack: an r package for  
380 integral projection models. *Methods in Ecology and Evolution*, 4:195–200, 2013.
- 381 C.J.E. Metcalf, S.M. McMahon, R. Salguero-Gomez, E. Jongejans, and C. Merow.  
382 *IPMpack: Builds and analyses Integral Projection Models (IPMs).*, 2014. URL  
383 <http://CRAN.R-project.org/package=IPMpack>. R package version 2.1.
- 384 M. Park, C. Loverdo, S.J. Schreiber, and J.O. Lloyd-Smith. Multiple scales of selection influence  
385 the evolutionary emergence of novel pathogens. *Philosophical Transactions of the Royal Socety.*  
386 *B*, Vol.:368, 2013.
- 387 Z Patwa and Lindi M Wahl. The fixation probability of beneficial mutations. *Journal of The Royal*  
388 *Society Interface*, 5:1279–1289, 2008.
- 389 R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for  
390 Statistical Computing, Vienna, Austria, 2015. URL <http://www.R-project.org/>.
- 391 M. Rees and S.P. Ellner. Integral projection models for populations in temporally varying environ-  
392 ments. *Ecological Monographs*, 79:575–594, 2009.

- 393 M. Rees, D.Z. Childs, and S.P. Ellner. Building integral projection models: a user's guide. *Journal*  
394 *of Animal Ecology*, 83:528–545, 2014.
- 395 R. Salguero-Gómez, W. Siewert, B.B. Casper, and K. Tielbörger. A demographic approach to  
396 study effects of climate change in desert plants. *Philosophical Transactions of the Royal Society*  
397 *B: Biological Sciences*, 367:3100–3114, 2012.
- 398 R. Salguero-Gómez, H. Kempenich, I. Forseth, and B.B. Casper. Long-term demography of a desert  
399 perennial species. *Ecology*, 95:577, 2014.
- 400 S.J. Schreiber and J.O. Lloyd-Smith. Invasion dynamics in spatially heterogenous environments.  
401 *American Naturalist*, 174:490–505, 2009.
- 402 S.J. Schreiber and N. Ross. *ibipm: Code for Individual-Based Integro-*  
403 *Projection Models*. Zenodo. <http://dx.doi.org/10.5281/zenodo.32086>, 2015. URL  
404 <http://dx.doi.org/10.5281/zenodo.32086>.
- 405 J.L. Williams, T.E.X. Miller, and S.P. Ellner. Avoiding unintentional eviction from integral projec-  
406 tion models. *Ecology*, 93:2008–2014, 2012.

## APPENDIX A: THE PGF FOR MULTI-TYPE BRANCHING PROCESSES

To see how this definition of  $\Psi$  recovers the pgf for the classical multi-type branching processes (see, for example, Caswell [2001]), assume there are  $k$  individual states with  $\mathcal{S} = \{1, 2, \dots, k\}$ . In this special case, functions  $h : \mathcal{S} \rightarrow \mathbb{R}$  can be viewed as vectors:  $h = (h(1), h(2), \dots, h(k))$ . The probability update rules,  $m(i, d\mathbf{s})$ , are determined by the probabilities  $p(i; n_1, n_2, \dots, n_k)$  that an individual in state  $i$  is “replaced” in the next time step by  $n_1$  individuals of type 1,  $n_2$  individuals of type 2,  $\dots$ ,  $n_k$  individuals of type  $k$ . Then the pgf  $\Psi$  is defined by

$$\begin{aligned} \Psi(h)(i) &= \int h^{\mathbf{s}} m(x, d\mathbf{s}) \\ &= \sum_{n_1, \dots, n_k} \prod_{i=1}^k h(i)^{n_i} p(i; n_1, n_2, \dots, n_k). \end{aligned}$$

where the sum is taken over all  $k$ -tuples  $n_1, n_2, \dots, n_k$  of non-negative integers.

## APPENDIX B: PROBABILITY GENERATING FUNCTIONALS FOR OTHER COMMON IPMS

Here, we illustrate how to write down probability generating functions (pgfs) for two common alternative forms of the IPM considered in the main text. The first alternative form considers the case where survival occurs before flowering, and the second alternative considers semelparous populations.

For the first alternative, if  $N_t(x)$  denotes the density of individuals of size  $x$  in year  $t$ , then we get an IPM of the form

$$N_{t+1}(x) = \int_a^b [s(y)G(y, x) + s(y)e(x)p(y)f(y)N_t(y)] dy$$

where  $s(y)$  is the probability of surviving to the next year for individuals of size  $y$ ,  $G(y, x) dy$  is the infinitesimal probability that a surviving individual of size  $y$  is size  $x$  in the next year,  $p(y)$  is the probability that an individual of size  $y$  flowers,  $f(y)$  is the mean number of offspring produced by an individual of size  $y$ , and  $e(x)$  is the infinitesimal probability that an offspring is of size  $x$  at the time of the annual census. As in the main text, we assume that a flowering individual produces a Poisson number  $N$  of offspring with mean  $f(x)$ . The pgf for  $N$  is given by  $\phi(x, \xi) = \exp(-f(x)(\xi - 1))$  where  $\xi$  is a dummy variable.

To write down the pgf in this case, we begin by writing down the pgf for a surviving individual of size  $x$ . An individual of this type provides two independent contributions to the next time step: the contribution to its own growth and the contribution due the production of offspring. As discussed in the main text, the pgfs for these contributions are, respectively,

$$\Psi_{growth}(h)(x) = \int h(y)G(y, x) dy \text{ and } \Psi_{fecundity} = 1 - p(x) + p(x)\phi(x, \int h(y)e(y)dy).$$

433 As these contributions are independent of one another, the pgf for a surviving individual  $x$  is the  
 434 product of these pgfs

$$\Psi_{surv}(h)(x) = \int h(y)G(y, x) dy \times \left(1 - p(x) + p(x)\phi(x, \int h(y)e(y)dy)\right).$$

435 To get the pgf for surviving and non-surviving individuals, we observe that the contributions of  
 436 an individual of size  $x$  is given by 0 with probability  $1 - s(x)$  and by the contributions of a surviving  
 437 individual with probability  $s(x)$ . By the **Fundamental Property 3**,

$$\begin{aligned} \Psi(h)(x) &= 1 - s(x) + s(x)\Psi_{surv}(h)(x) \\ &= 1 - s(x) + s(x) \times \int h(y)G(y, x) dy \times \left(1 - p(x) + p(x)\phi(x, \int h(y)e(y)dy)\right). \end{aligned}$$

438 Next we consider a semelparous (monocarpic) population where flowering is followed by death.  
 439 In this case, the IPM is given by

$$N_{t+1}(x) = \int_a^b [s(y)(1 - p(y))G(y, x) + s(y)e(x)p(y)f(y)N_t(y)] dy.$$

440 The main difference for the pgf of this IPM is that surviving individuals don't have independent  
 441 contributions through growth and fecundity. Instead, these contributions are mutually exclusive  
 442 and **Fundamental Property 3** of pgfs implies

$$\Psi_{surv}(h)(x) = (1 - p(x)) \int h(y)G(y, x) dy + p(x)\phi(x, \int h(y)e(y)dy).$$

443 Hence,

$$\begin{aligned} \Psi(h)(x) &= 1 - s(x) + s(x)\Psi_{surv}(h)(x) \\ &= 1 - s(x) + s(x) \left( (1 - p(x)) \int h(y)G(y, x) dy + p(x)\phi(x, \int h(y)e(y)dy) \right). \end{aligned}$$

## APPENDIX C: EXTINCTION PROBABILITIES IN A TIME VARYING ENVIRONMENT

Let  $m_0(x, d\mathbf{s}), m_1(x, d\mathbf{s}), m_2(x, d\mathbf{s}), \dots$  be a sequence of probability measures defining the population update rules at times  $0, 1, 2, \dots$  i.e.  $\mathbb{P}[\mathbf{s}_{t+1} \in A | \mathbf{s}_t = (1; x)] = \int_A m_t(x, d\mathbf{s})$  for any  $A \subset \mathcal{S}$ . For each of these probability measures, define the pgf

$$\Psi_t(h)(x) = \int h^{\mathbf{s}} m_t(x, d\mathbf{s})$$

for any continuous function  $h : X \rightarrow \mathbb{R}$ . We claim that for  $t \geq 0$  and  $\tau \geq 0$

$$(5) \quad \Psi_t(\Psi_{t+1}(\dots \Psi_{t+\tau}(h_0)))(x)$$

is the probability of extinction by time  $t + \tau + 1$  for a population consisting of one individual in state  $x$  at time  $t$ . As before, we argue by induction on  $\tau$ . For  $\tau = 0$  and any  $t \geq 0$ , the definition of  $\Psi_t$  implies

$$(6) \quad \Psi_t(h_0)(x) = \int_{\{0\}} 1 \times m_t(x, d\mathbf{s}) + \int_{\mathcal{S} \setminus \{0\}} 0 \times m_t(x, d\mathbf{s}) = \mathbb{P}[\mathbf{s}_{t+1} = 0 | \mathbf{s}_t = (1; x)].$$

Now suppose that equation (5) holds for some  $\tau \geq 0$  and all  $t$ ; we will show it holds  $\tau + 1$  and all  $t$ . Fix  $t \geq 0$  and assume  $\mathbf{s}_t = (1; x)$ . On the event that  $\mathbf{s}_{t+1} = (n_1, \dots, n_k; x_1, \dots, x_k)$  is the population state at time  $t + 1$ , extinction occurs by time  $t + \tau + 2$  only if each of the lineages of the  $n_1 + \dots + n_k$  individuals go extinct in the next  $\tau + 1$  time steps. As the fates of these lineages are independent of one another, it follows that

$$\begin{aligned} \mathbb{P}[\mathbf{s}_{t+\tau+2} = 0 | \mathbf{s}_t = (1; x), \mathbf{s}_{t+1} = (n_1, \dots, n_k; x_1, \dots, x_k)] &= \prod_{i=1}^k \mathbb{P}[\mathbf{s}_{t+\tau+2} = 0 | \mathbf{s}_{t+1} = (1; x_i)]^{n_i} \\ &= \prod_{i=1}^k (\Psi_{t+1}(\Psi_{t+2}(\dots \Psi_{t+1+\tau}(h_0)))(x_i))^{n_i} \\ &= \Psi_{t+1}(\Psi_{t+2}(\dots \Psi_{t+\tau+1}(h_0)))^{(n_1, \dots, n_k; x_1, \dots, x_k)} \end{aligned}$$

457 where the second equality follows from our inductive hypothesis that (5) holds for  $\tau$  and any  $t$ . By  
 458 the law of total probability and conditioning over all possible states at time  $t + 1$ , we get

$$\begin{aligned}\mathbb{P}[s_{t+\tau+2} = 0 | \mathbf{s}_t = (1; x)] &= \int \Psi_{t+1}(\Psi_{t+2}(\dots \Psi_{t+\tau+1}(h_0)))^{\mathbf{s}} m_t(x, d\mathbf{s}) \\ &= \Psi_t(\Psi_{t+1}(\dots \Psi_{t+\tau+1}(h_0)))(x).\end{aligned}$$

459 As  $t \geq 0$  was arbitrary, this completes the inductive proof.

460 DEPARTMENT OF EVOLUTION AND ECOLOGY, ONE SHIELDS AVENUE, UNIVERSITY OF CALIFORNIA, DAVIS,  
 461 CALIFORNIA 95616

462 *E-mail address:* `sschreiber@ucdavis.edu`

463 ECOHEALTH ALLIANCE, 460 WEST 34TH STREET, 17TH FLOOR, NEW YORK, NY 10001

464 *E-mail address:* `ross@ecohealthalliance.org`



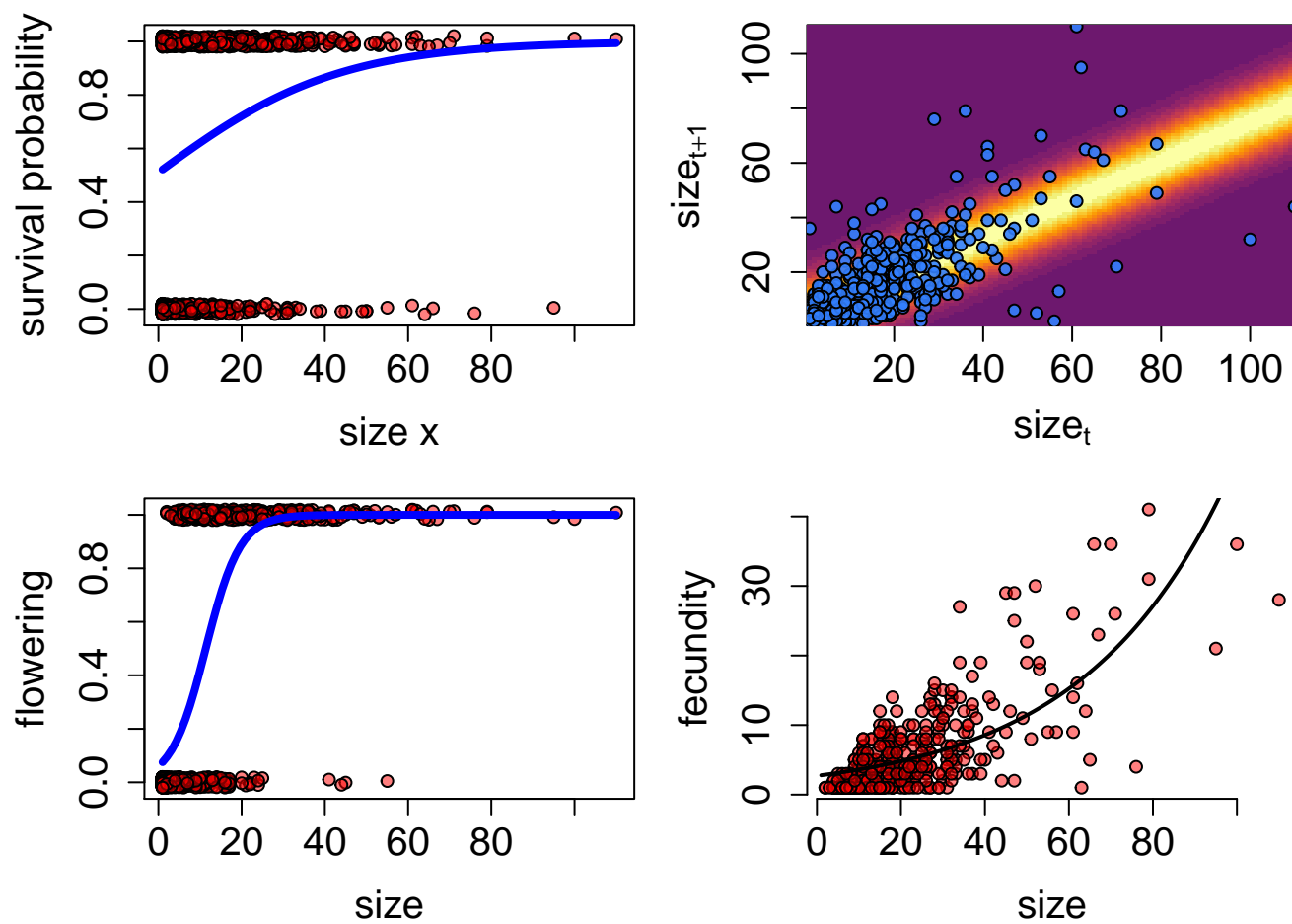


FIGURE 1. The demographic kernels for Plateau Yellow Miner's Candle with the corresponding data from Salguero et al. 2012.

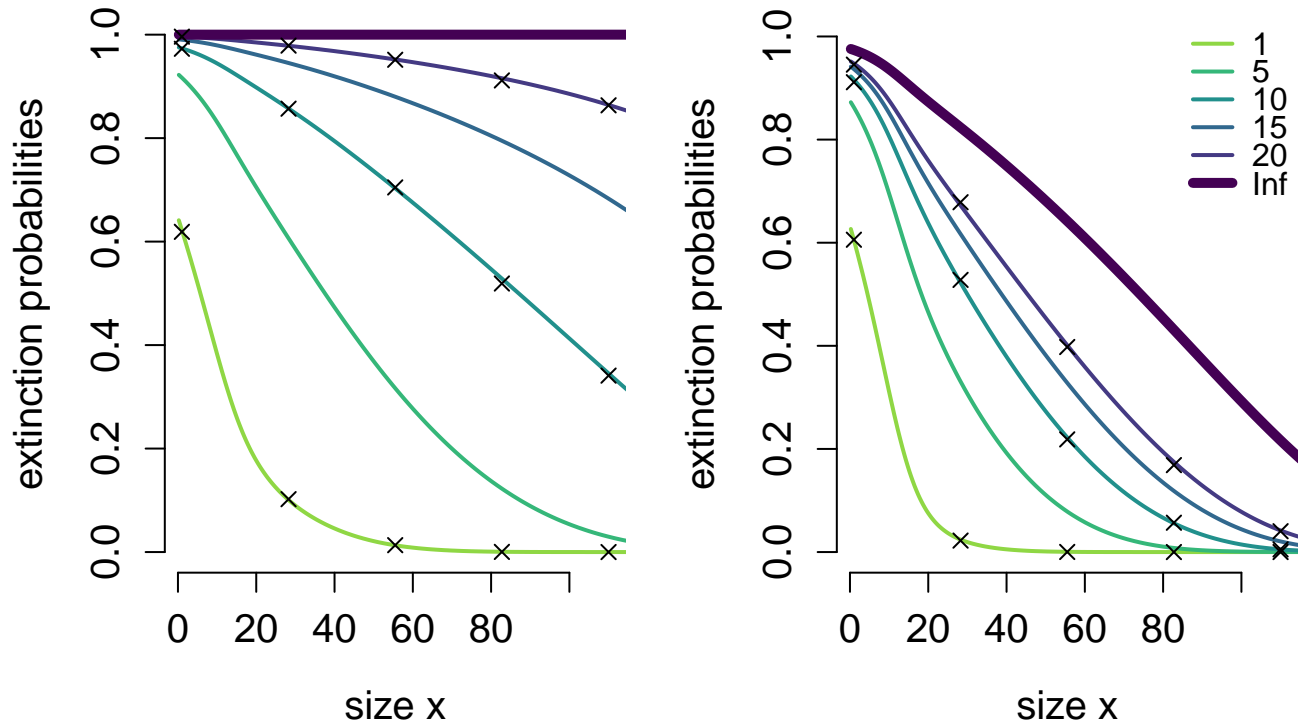


FIGURE 2. Extinction probabilities for a population initially consisting of a single individual of size  $x$ . Different curves correspond to extinction occurring in 1, 5, 10, 15 or 20 years. Asymptotic extinction probabilities are shown by the thicker curve. In A, the extinction curves for the baseline individual-based IPM. In B, extinction curves for the case when seed survival is increased by a factor of three. Crosses indicate estimates of extinction probabilities from 10,000 simulations of a full individual-based model (see code in [Schreiber and Ross, 2015]).

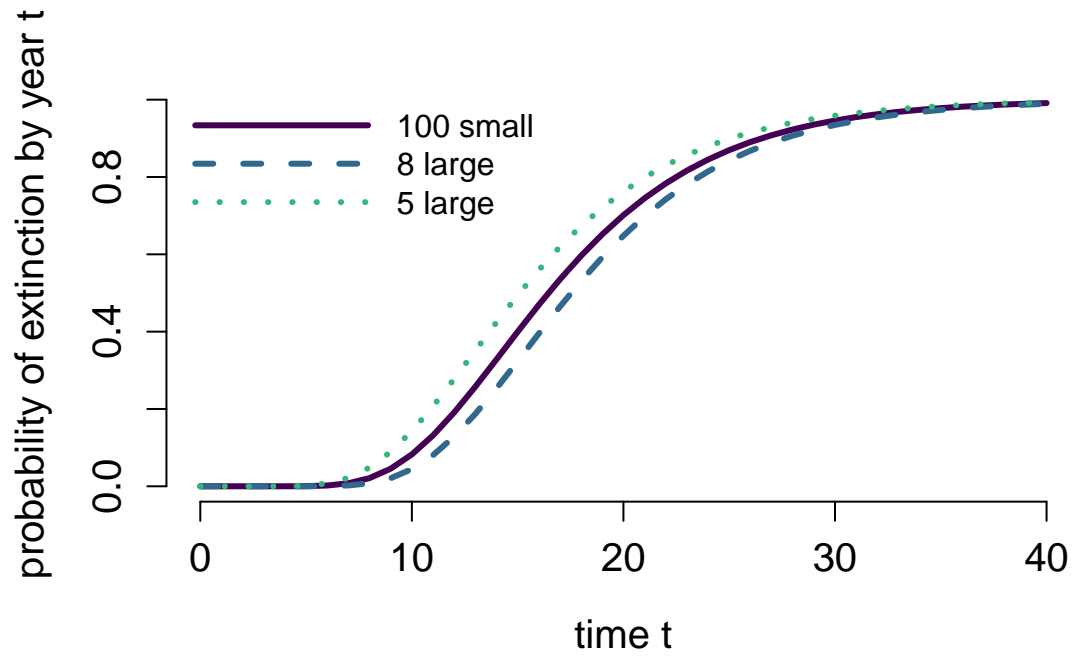


FIGURE 3. Extinction probabilities as a function of time for populations with 100 individuals of size  $x = 1$ , and populations with 5 or 8 individuals of size  $x = 60$ .

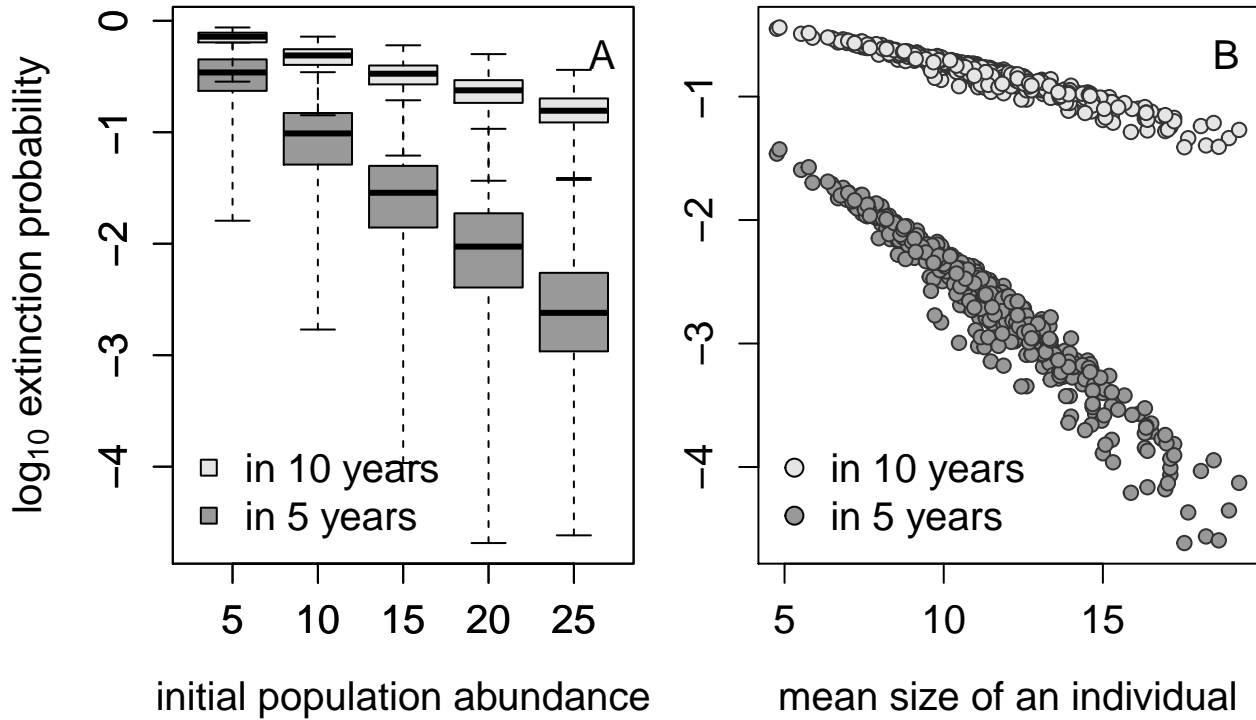


FIGURE 4. Extinction probabilities for founding populations of different abundance. For each founding population abundance  $N$ , 500 samples consisting of  $N$  randomly chosen individuals from the data set were used to create a founding population of  $N$  individuals. Extinction probabilities by year 5 and year 10 were calculated for each of these sample populations. In A, extinction probability is plotted as a box plot for different  $N$  values. In B, extinction probability is plotted against the mean size of an individual for a population abundance  $N = 25$ .